UNIVERSITÄT DES SAARLANDES
Lehrstuhl für Elektronik und Schaltungstechnik

Prof. Dr.-Ing. Michael Möller

Skriptum zur

Vorlesung

Elektronik I

– Halbleiterbauelemente –

WS 2005/06

Letzte Korrektur: 11. Februar 2015

1 Festkörperphysik für Elektroniker

1.1 Grundlagen des Atomaufbaus

1. Das Atom in der klassischen Teilchenvorstellung besteht aus einem Atomkern und einer um den Kern liegenden Elektronenhülle. Atome liegen in der Größenordnung von einem bis zwei Angström (Å) = 0,1 nm ... 0,2 nm.

2. Der Kern besteht aus einer Anzahl Z Protonen und in etwa genauso viel Neutronen. Die Anzahl Z heißt Ordnungszahl und ist charakteristisch für ein Element. Die verschiedenen Arten eines Elements mit unterschiedlicher Anzahl von Neutronen heißen Isotope. Der Kern besitzt nur ca. $\frac{1}{10.000}$ der Atomgröße und liegt in der Größenordnung 10^{-6} nm.

3. Die Hülle eines nicht ionisierten, elektrisch neutralen Atoms besteht aus Elektronen, die den Kern auf bestimmten ausgezeichneten Bahnen (Orbitalen) umlaufen. Die Größe eines Elektrons liegt in der Größenordnung des Atomkerns $\approx 5,6 \cdot 10^{-6}$ nm. Daher wird die Größe eines Atoms in guter Näherung nur durch die Größe der Elektronenhülle bestimmt.

4. Die Masse der Protonen und Neutronen im Atomkern ist sehr viel größer als die der Elektronen ($m_e = 9,1 \cdot 10^{-31}$ kg). Daher bestimmt der Atomkern im wesentlichen die Masse des Atoms.

5. Die Ladung jedes Protons im Kern beträgt $+e = 1,6 \cdot 10^{-19}$ C. Neutronen sind elektrisch neutral. Ein Elektron trägt die Ladung $-e$. Daher ist ein nicht ionisiertes Atom von außen gesehen (makroskopisch) neutral.

1.2 Atommodelle

1.2.1 Planetenmodell

Das klassische Rutherfordsche Atommodell, bei dem die Elektronen den Kern auf Kreis- oder Ellipsenbahnen umlaufen, wie die Planeten die Sonne, dient heute (nur) noch als Hilfe bei der Vorstellung des Atomaufbaus. Physikalisch ist dieses Modell nicht haltbar. Die Elektronen erfahren auf ihren Kreisbahnen eine Beschleunigung. Die Ladung der Elektronen, die diese Beschleunigung erfährt, müsste demnach eine elektromagnetische

Welle abstrahlen, deren Energie jedoch dem System verloren geht. Aufgrund der geringeren Energie müsste dann aber ein Elektron auf einer Kreisbahn mit niedrigem Radius fliegen, was letztendlich dazu führt, dass das Elektron in den Kern stürzt.

1.2.2 Bohrsches Modell

Abhilfe gegen diese planetarische Katastrophe liefert das von Niels Bohr 1913 postulierte Axiom, wonach es nur noch bestimmte stabile Bahnen der Elektronen um den Kern gibt, auf denen ein Elektron keine Energie verliert. Die Gesamtenergie eines solchen stabilen Zustandes ist konstant und besteht aus kinetischer und potentieller Energie. Der Unterschied zwischen den Energien der Bahnen ist aufgrund der einzelnen erlaubten Bahnen für die stabilen Zustände quantisiert. Die Energieaufnahme oder -abgabe erfolgt demnach in quantisierten Mengen, in sogenannten Quanten. Zum Wechsel von einem Energieniveau auf ein anderes muss Energie z. B. in Form von Licht (Photonen) aufgenommen oder abgegeben werden. Bohr postulierte stabile Bahnen auf Basis einer Quantelung des Drehimpulses $m \cdot v \cdot r$ eines Elektrons der Masse m, das sich mit der Geschwindigkeit v auf einer Kreisbahn mit dem Radius r um den Kern befindet. Danach darf der Drehimpuls nur ganzzahlige Vielfache von $\hbar = \frac{h}{2\pi}$ (h = Plancksches Wirkungsquantum $= 6,626 \cdot 10^{-34}$ Js betragen:

$$m \cdot v \cdot r = n \cdot \hbar \quad . \tag{1.1}$$

Mit dieser Quantelung des Drehimpulses lassen sich quantisierte Energiezustände berechnen, die bei Anwendung auf das Wasserstoffatom zu einer Übereinstimmung der Ergebnisse von Modell und Experiment für das Absorptions- und Emissionsspektrum (Balmer-Serie) führen. Bei Anwendung auf komplexere Atome mit mehr als einem Elektron liefert das Bohrsche Atommodell falsche Ergebnisse.

1.2.3 De Broglies-(Wellen-)Modell

Die Lösung, die zugleich unserer heutigen Modellvorstellung entspricht, beruht auf der 1924 von de Broglie eingeführten Welleneigenschaft von bewegten Teilchen.

Ein Photon (Lichtquant) mit der Frequenz f hat als bewegtes Teilchen den Impuls

$$p = \frac{h \cdot f}{c} = \frac{h}{\lambda} \quad \text{mit } \lambda f = c \quad . \tag{1.2}$$

Die Wellenlänge des Photons ist demnach

$$\lambda = \frac{h}{p} \quad . \tag{1.3}$$

De Broglies Welleneigenschaft von bewegten Teilchen postuliert, dass diese Beziehung sowohl für Photonen als auch für alle anderen Teilchen gilt. D. h. jedes Teilchen kann auch als Welle mit einer Wellenlänge λ aufgefaßt werden. Der Impuls eines Teilchens der Masse m und Geschwindigkeit v ist $p = mv$. Die de Broglie-Wellenlänge dieses Teilchens beträgt dann

$$\lambda = \frac{h}{m \cdot \mathrm{v}} \, . \tag{1.4}$$

Die kinetische Energie eines Teilchens der Ruhemasse m_0 beträgt

$$W_{kin} = \frac{1}{2} m_0 v^2 \, .$$

Die Geschwindigkeit v des Teilchens ist nicht identisch mit der Phasengeschwindigkeit $\mathrm{v}_p = f \cdot \lambda$ der Welle ($y = A \cdot \cos 2\pi f(t - \frac{x}{\mathrm{v}_p})$). Dies sieht man, wenn man die quantentheoretische Darstellung der Energie

$$W = h \cdot f \tag{1.5}$$

mit der relativistischen Gesamtenergie eines Teilchens

$$W = m \cdot c^2 = m_0 c^2 + W_{kin} \tag{1.6}$$

gleichsetzt und nach der Frequenz umstellt

$$f = \frac{mc^2}{h} \, . \tag{1.7}$$

Die Phasengeschwindigkeit der de Broglie-Welle ist daher mit Gl. (1.4)

$$\mathrm{v}_{ph} = f \cdot \lambda = \frac{mc^2}{h} \cdot \frac{h}{mv} = \frac{c^2}{v} \, . \tag{1.8}$$

Da die Geschwindigkeit des Teilchens immer kleiner als die Lichtgeschwindigkeit ist, ist die Phasengeschwindigkeit der de Broglie Wellenlänge immer größer als die Lichtgeschwindigkeit. Dies klingt zunächst spektakulär, ist aber bei genauerer Betrachtung ein Resultat der gewählten mathematischen Beschreibung. Die physikalische, die Geschwindigkeit eines Teilchens im Raum beschreibende Größe der Gruppengeschwindigkeit, liegt immer unter der Lichtgeschwindigkeit und entspricht daher den Vorhersagen der Relativitätstheorie.
Je größer die Masse des Teilchens, umso kleiner ist die Wellenlänge seiner de Broglie-Welle. Zu beachten ist, dass m die relativistische Masse

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{1.9}$$

bezeichnet (m_0 ist die Ruhemasse).

1.2.4 Anwendung auf das Bohr-Modell des Wasserstoffatoms

Abb. 1.1 zeigt ein Bahnmodell des Wasserstoffatoms.

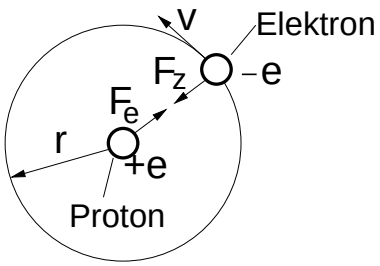

Abb. 1.1: Kräftegleichgewicht zwischen Zentripetalkraft F_Z und Anziehungskraft F_e aufgrund der Ladungen $\pm e$.

Auf einer stabilen Bahn gilt

$$F_Z = \frac{m\mathrm{v}^2}{r} = F_e = \frac{e^2}{4\pi\epsilon_0 r^2} \; . \tag{1.10}$$

Daraus bestimmt sich die Geschwindigkeit des Elektrons zu

$$\mathrm{v} = \frac{e}{\sqrt{4\pi\epsilon_0 m r}} \; . \tag{1.11}$$

Einsetzen in Gl. (1.4) ergibt die de Broglie-Wellenlänge

$$\lambda = \frac{h}{e}\sqrt{\frac{4\pi\epsilon_0 r}{m}} \; . \tag{1.12}$$

De Broglies Postulat für einen stabilen Orbit des Elektrons beruht auf der Welleneigenschaft des Teilchens. Danach muss das Elektron auf sog. stabilen Kreisbahnen mit Radien r_n laufen, für deren Umfang gilt:

$$2\pi r_n = n\lambda \; . \tag{1.13}$$

Darin ist $n = 1, 2, 3, \ldots$ die Hauptquantenzahl des Orbits. Die Einhaltung von Gl. (1.13) gewährleistet, dass die Wellenfunktion (Schwingung) des Elektrons für jeden Umlauf gleichphasig (konstruktiv) überlagern. Die ortsabhängige Amplitude der Wellenfunktion ist in diesem Fall zeitunabhängig und man spricht von einer stehenden Welle. In allen anderen Fällen kommt es zu einer Auslöschung der Wellenfunktion aufgrund der Überlagerung mit beliebigen Phasenlagen. Abb. 1.2 zeigt Beispiele für Wellenfunktionen auf stabilen Kreisbahnen mit $n = 1$, 2 und 3.

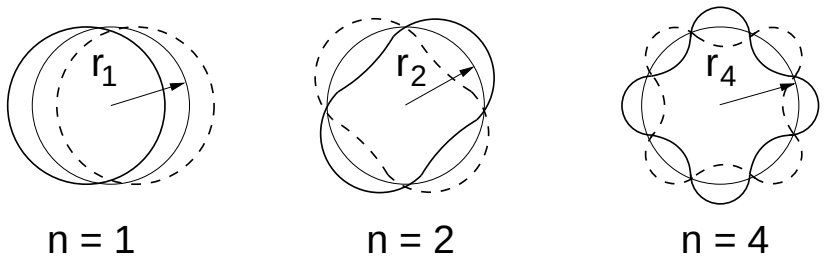

Abb. 1.2: Gedankenmodell für die konstruktive Überlagerung der Wellenfunktion nach der Bedingung in Gl. (1.13). Beachten: Die Abbildungen für $n = 1, 2, 4$ sind in unterschiedlichen Maßstäben.

Mit den nach Gl. (1.13) erforderlichen Wellenlängen ergeben sich aus Gl. (1.12) die Radien des Bohrschen Atommodells[1]

$$r_n = \frac{n^2 h^2 \epsilon_0}{\pi m e^2} \qquad n = 1, 2, 3, \ldots \; . \tag{1.14}$$

Der innerste Radius des Wasserstoffatoms (Bohrradius) lässt sich damit zu $r_1 = 5,3 \cdot 10^{-2}$ nm berechnen.

[1]Die gleichen Radien könnten auch unter Verwendung des Bohrschen Postulats für den Bahndrehimpuls gewonnen werden.

Die Energie W_n des Elektrons im Wasserstoffatom in Abhängigkeit des Bahnradius ergibt sich als Summe seiner kinetischen Energie

$$W_{kin} = \frac{1}{2}mv^2 \tag{1.15}$$

und seiner potentiellen Energie im elektrostatischen Potential seines Atomkerns mit der Ladung eines Protons (vgl. Übung)

$$W_{pot} = \frac{-e^2}{4\pi\epsilon_0 r} \, . \tag{1.16}$$

Abb. 1.3 zeigt den Verlauf der potentiellen Energie in Abhängigkeit vom Abstand r vom Kern.

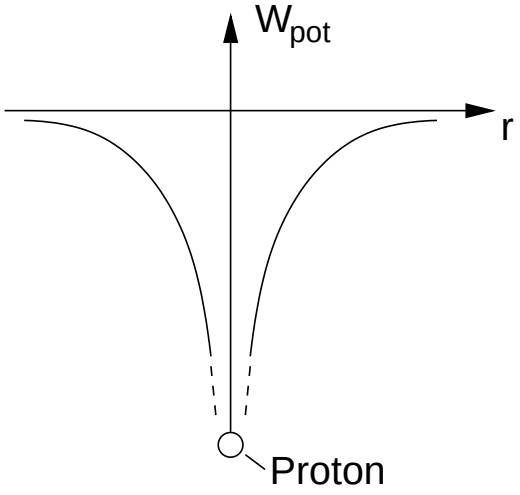

Abb. 1.3: Potentielle Energie W_{pot} eines Elektrons auf einer Kugelfläche mit dem Abstand r vom Atomkern.

Die potentielle Energie bildet demnach einen Potentialtopf, in dem sich das Elektron befindet. Die Gesamtenergie des Elektrons ist

$$W = W_{kin} + W_{pot} = \frac{1}{2}m\mathrm{v}^2 - \frac{e^2}{4\pi\epsilon_0 r} \, . \tag{1.17}$$

Einsetzen der Geschwindigkeit aus Gl. (1.11) liefert allgemein

$$W = \frac{1}{2}m\left(\frac{e}{\sqrt{4\pi\epsilon_0 mr}}\right)^2 - \frac{e^2}{4\pi\epsilon_0 r} = \frac{-e^2}{8\pi\epsilon_0 r} \, . \tag{1.18}$$

Für Radien r_n nach Gl. (1.14), deren Umfang ein ganzzahliges Vielfaches der de Broglie-Wellenlänge entspricht, ergeben sich die quantisierten Energien

$$W_n = \frac{-e^2}{8\pi\epsilon_0 r_n} = \frac{-me^4}{8\epsilon_0^2 h^2} \cdot \frac{1}{n^2} = \frac{W_1}{n^2} \,. \tag{1.19}$$

$n = 1, 2, 3, \ldots$ wird als Quantenzahl bezeichnet. W_1 ist die Grundenergie im nicht angeregten Zustand. Abb. 1.4 zeigt die Energiezustände grafisch.

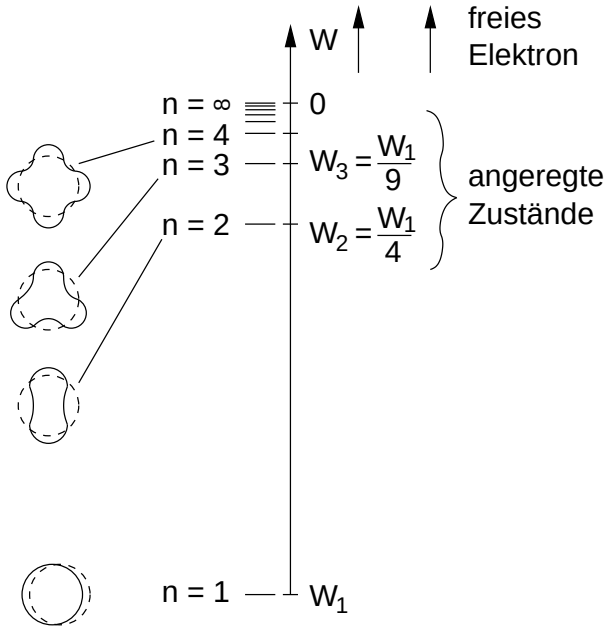

Abb. 1.4: Quantisierte Energien des Wasserstoff-Elektrons.

Für $n \to \infty$ geht die Energie des Elektrons gegen Null und es ist nicht mehr an den Kern gebunden. Man spricht von einem freien Elektron. In diesem Fall ist das Wasserstoffatom ionisiert. Die Energie, die dem Wasserstoffatom zugeführt werden muss, um das Elektron aus dem Grundzustand in den freien Zustand zu bekommen, ist $-W_1$.

1.3 Wellenmodell des Atoms – Schrödingergleichung

Schwierigkeiten bei der Vorstellung der Materiewelle von de Broglie bereitet die Vorstellung, was (welcher Gegenstand / Objekt) genau die Wellenbewegung (oder Schwingung) ausführt. In der von begreifbaren Gegenständen

geprägten menschlichen Vorstellung ist man geneigt, sich das Elektron als Teilchen auf einer gewellten Kreisbahn (vgl. Abb. 1.2) vorzustellen. Diese Vorstellung ist aber falsch. Vielmehr müssen wir in der Wellenvorstellung dem Elektron seinen Teilchencharakter nehmen und es in seiner Gesamtheit als Welle betrachten. (Stellen Sie sich z. B. Licht oder allgemein ein elektromagnetisches Feld Ihrer Handy-Antenne vor.) Wenn wir Elektronen in einem Orbit um den Kern betrachten, handelt es sich um stehende Wellen mit der de Broglie-Wellenlänge. Für freie Elektronen ergeben sich Wellen, die anderen Randbedingungen genügen.

Schrödinger formuliert 1926 eine (die!) bedeutende Wellengleichung, die allen Teilchen Welleneigenschaften zuschreibt und damit die moderne Quantenmechanik begründet. Er definiert hierzu eine komplexe Wellenfunktion $\psi(\vec{r})$, die die Welleneigenschaft eines (oder mehrerer) Teilchen[2] beschreibt. $\psi(\vec{r})$ ist eine abstrakte mathematische Größe, beinhaltet jedoch alle physikalischen und damit anschaulichen und (be-)greifbaren Eigenschaften des beschriebenen Systems. Durch mathematische Operationen können diese aus ihr gewonnen werden.

Die Schrödingergleichung in ihrer allgemeinen zeitabhängigen Form ist eine der fundamentalsten Gleichungen der Physik. Sie enthält die Newtonschen Grundgleichungen und über die Dirac-Erweiterung auch die Maxwellschen Gleichungen und die spezielle Relativitätstheorie. Sie kann nicht aus bisher bekannten physikalischen Prinzipien hergeleitet werden, sondern stellt selbst ein auf Experimente gestütztes physikalisches Axiom dar. Aufgrund ihrer Mächtigkeit sind Lösungen der Schrödingergleichung in der Regel sehr umfangreich.

Wir begnügen uns im Rahmen dieser Vorlesung mit Lösungen zu sehr einfachen Modellen, die jedoch zum Verständnis von Ursachen und deren Wirkungen sowie von Methoden zu deren Beschreibung genügen. Wir betrachten stationäre Systeme, in denen sich also über der Zeit nichts ändert. In diesen Systemen existieren nur stehende Wellen und es gilt die vereinfachte

[2]Hinsichtlich der Teilchengröße gibt es keine Einschränkung. Theoretisch könnten auch Murmeln oder Fußbälle als Welle beschrieben werden.

zeitunabhängige Schrödingergleichung

$$\frac{-\hbar^2}{2m}\Delta\psi(\vec{r}) + (W_{pot}(\vec{r}) - W)\,\psi(\vec{r}) = 0\,. \tag{1.20}$$

Darin ist m die Masse des betrachteten Teilchens, $W_{pot}(\vec{r})$ die potentielle Energie des betrachteten Systems am Ort $\vec{r}$ und $\hbar = \frac{h}{2\pi}$. Die Gesamtenergie W des Systems (in der Regel kinetische plus potentielle Energie) ist wegen des Energieerhaltungssatzes eine feste Zahl und hängt nicht von einer Ortskoordinate ab. Der auf $\psi(\vec{r})$ angewendete Operator Δ ist der Laplace-Operator, der in kartesischen Koordinaten

$$\Delta\psi(\vec{r}) = \Delta\psi(x,y,z) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi(x,y,z) \tag{1.21}$$

lautet. Für einfache eindimensionale Betrachtungen wird nur eine Abhängigkeit von $\psi(\vec{r})$ in eine Richtung angenommen. Für $\psi(\vec{r}) = \psi(x)$ vereinfacht sich Gl. (1.21) zu

$$\Delta\psi(x) = \frac{\partial^2}{\partial x^2}\psi(x) \tag{1.22}$$

und die eindimensionale Schrödingergleichung bekommt die Gestalt:

$$\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) + (W_{pot}(x) - W)\,\psi(x) = 0\,. \tag{1.23}$$

Eine wichtige Eigenschaft der Schrödingergleichung ist ihre Linearität: wenn ψ_1 und ψ_2 Lösungen der Schrödingergleichung sind, dann ist auch $a\psi_1 + b\psi_2$ eine Lösung. Zu beachten ist dabei, dass a und b so gewählt sein müssen, dass die später diskutierte Normierungsbedingung nach Gl. (1.36) erfüllt sein muss.

1.4 Wellenfunktionen

Die Orts- und Zeitabhängigkeit einer in einer Dimension (hier in x-Richtung) fortschreitenden harmonischen Schwingung (z. B. einer Saite) lässt sich durch die Wellenfunktion

$$\psi(x,t) = A\cos 2\pi f\left(t - \frac{x}{\mathrm{v}_p}\right) \tag{1.24}$$

beschreiben. Darin ist A die Amplitude der Welle, f die Frequenz und v_p die Phasengeschwindigkeit. Wenn wir mit v_p die Geschwindigkeit der de Broglie-Welle mit der Wellenlänge $\lambda = \frac{h}{mv}$ (vgl. Gl. (1.4)) bezeichnen, gilt

$$v_p = f\lambda\,, \tag{1.25}$$

d. h. die Welle hat eine Orts-Periode von $x = \lambda$.
Gl. (1.24) kann mit der Definition der Kreisfrequenz $\omega = 2\pi f$ geschrieben werden:

$$\psi(x,t) = A\cos(\omega t - kx) \ . \tag{1.26}$$

Darin ist mit

$$k := \frac{\omega}{\mathrm{v}_p} = \frac{2\pi}{\lambda} \tag{1.27}$$

die Wellenzahl der in x-Richtung fortschreitenden Welle definiert worden.
In drei Dimensionen wird k zum Wellenvektor

$$\vec{k} = \vec{e}_k k = k_x\vec{e}_x + k_y\vec{e}_y + k_z\vec{e}_z \tag{1.28}$$

und die Ortskoordinate x geht über in einen Raumvektor

$$\vec{r} = x\vec{e}_x + y\vec{e}_y + z\vec{e}_z \ . \tag{1.29}$$

Aus der Wellengleichung (1.26) wird dann

$$\psi(\vec{r},t) = A\cos\left(\omega t - \vec{k}\vec{r}\right) \ . \tag{1.30}$$

Gl. (1.30) beschreibt eine in Richtung $\vec{k}$ fortschreitende ebene Welle. D. h. auf Ebenen senkrecht zu $\vec{k}$ ist die Amplitude für t=const. konstant. Abb. 1.5 zeigt einen Ausschnitt dieser in Raum und Zeit fortschreitende Wellenfunktion.

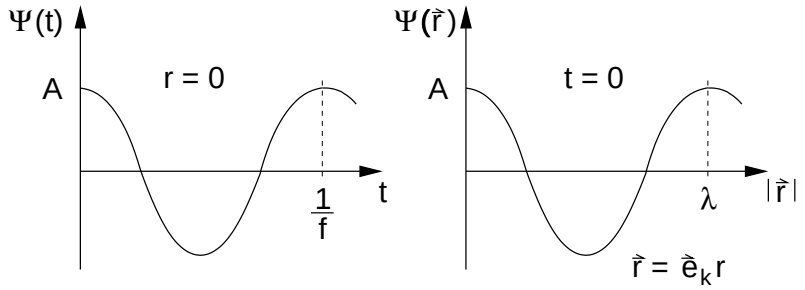

Abb. 1.5: Links: Wellenfunktion an einem festen Ort ($r = 0$) in Abhängigkeit von der Zeit. Rechts: Wellenfunktion zu einem Zeitpunkt bei Bewegung in Richtung $\vec{r} = \vec{e}_k r$.

In Gl. (1.30) wurde willkürlich eine cos-Schwingung verwendet. Um einen allgemeinen Ansatz der Wellenfunktion als Lösung der Schrödingergleichung zu erhalten, ist es sinnvoll, der cos-Schwingung noch eine sin-Schwingung zu

überlagern. Man bedient sich bei der Überlagerung vorteilhaft der komplexen Zahlen und versieht (willkürlich) den sin-Term mit dem imaginären Vorfaktor j. Dadurch ergibt sich die Möglichkeit, beide Schwingungen elegant in der Euler-Schreibweise nach Gl. (1.32) zusammenzufassen:

$$\psi(\vec{r},t) = A\cos\left(\omega t - \vec{k}\vec{r}\right) + jA\sin\left(\omega t - \vec{k}\vec{r}\right) \ , \tag{1.31}$$

$$\psi(\vec{r},t) = Ae^{j(\omega t - \vec{k}\vec{r})} \ . \tag{1.32}$$

Im zeitunabhängigen Fall t=const.=0 geht Gl. (1.32) über in

$$\psi(\vec{k},\vec{r}) = Ae^{j\vec{k}\vec{r}} \ . \tag{1.33}$$

Aus Linearkombinationen von Gl. (1.33) gehen wieder die zuvor verwandten sin- und cos-Schwingungen hervor.

Beispiel/Übung:
Ermitteln Sie zur Übung die resultierende Funktion aus der Überlagerung von $\psi(\vec{k},\vec{r}) + \psi(-\vec{k},\vec{r})$ und $\psi(\vec{k},\vec{r}) - \psi(-\vec{k},\vec{r})$ entsprechend Gl. (1.33) unter Zuhilfenahme der Eulerschreibweise $a(\cos\varphi + j\sin\varphi) = ae^{j\varphi}$.

Wichtig: Für Elektrotechniker ist das Arbeiten mit Phasoren zur Vereinfachung von Berechnungen selbstverständlich geworden. Bei der Verwendung von Phasoren wie z. B.

$$\underline{U} = \underline{I}(R + j\omega C)$$

wird definitionsgemäß der Übergang zur reellen Zeitfunktion durch Realteilbildung vollzogen:

$$u(t) = \mathfrak{Re}\left\{|\underline{U}|e^{j\varphi_u}e^{j\omega t}\right\}$$

$$u(t) = \mathfrak{Re}\left\{|\underline{I}||R + j\omega C|e^{j(\varphi_I + \arctan\frac{\omega C}{R})}e^{j\omega t}\right\}$$

Diese Vorgehensweise ist für die komplexe Wellenfunktion nicht richtig! Sie ist tatsächlich eine komplexe Funktion. Die komplexe Schreibweise dient hier nicht zur Vereinfachung von Berechnungen.

1.5 Physikalische Interpretation der Wellenfunktion

Da die Wellenfunktion letztendlich Lösungen zu realen, von uns mess- und erfahrbaren Problemen enthält, ist eine physikalische Interpretation der Wellenfunktion notwendig. Wir erwarten von dieser Interpretation, dass sie für uns reellwertige und damit in Form von Messergebnissen (Experimenten) begreifbare Ergebnisse liefert.

Die Quantentheorie postuliert, dass das Betragsquadrat $|\psi(\vec{r})|^2$ der Wellenfunktion ein Maß für die Wahrscheinlichkeit ist, ein Teilchen an dem Ort $\vec{r}$ anzutreffen. Die Wahrscheinlichkeit $f(\vec{r})$, dass das Teilchen in einem Volumenelement $dx\ dy\ dz$ an der Stelle $\vec{r} = (x, y, z)'$ angetroffen wird, ist

$$f(\vec{r}) = |\psi|^2\ dx\ dy\ dz = |\psi|^2 dV\ . \tag{1.34}$$

Damit stellt das Betragsquadrat der Wellenfunktion

$$|\psi|^2 = \psi\psi^* = \frac{f(\vec{r})}{dV} \tag{1.35}$$

eine Wahrscheinlichkeitsdichte dar.
Die Wahrscheinlichkeit, das Teilchen irgendwo in einem beliebig großen Volumen zu finden, ist 100%. Daraus ergibt sich die Normierungsbedingung

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\psi|^2 dx\ dy\ dz\ = 1\ . \tag{1.36}$$

1.6 Die Schrödingergleichung und das Wasserstoffatom.

Das Wasserstoffatom ist das am einfachsten aufgebaute Atom. Hier bewegt sich nur ein Elektron im Potential des Atomkerns. Trotz seines einfachen Aufbaus ist die Lösung der Schrödingergleichung wegen der dreidimensionalen Struktur bereits aufwendig. Die einzige Größe, die zur Lösung bekannt sein muss, ist die potentielle Energie des Atomrumpfes, in der sich das Elektron bewegt. Wir haben sie bereits in Gl. (1.16) angegeben. Sie beträgt

$$W_{pot} = \frac{-e^2}{4\pi\epsilon_0 r} \tag{1.16}$$

Die Gesamtenergie des Atoms sowie die Wellenfunktionen ergeben sich als Lösung der Schrödingergleichung. Diese wird zur Lösung in Kugelkoordinaten nach Abb. 1.6 dargestellt, da aufgrund des kugelsymmetrischen Potentials (W_{pot} = const. auf Kugelschalen um den Kern) ebenfalls Lösungen mit Kugelsymmetrie zu erwarten sind.

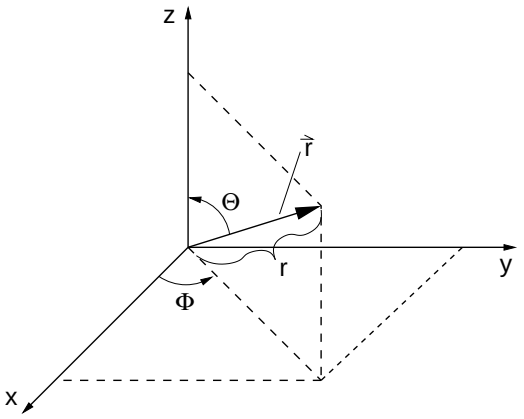

Abb. 1.6: Zusammenhang zwischen Kugelkoordinaten (r, ϕ, θ) und kartesischen Koordinaten (x, y, z).

Die zu lösende Schrödingergleichung in Kugelkoordianten mit $\psi = \psi(r, \phi, \theta)$ lautet mit der potentiellen Energie nach Gl. (1.16)

$$\frac{\hbar^2}{2m}\Delta\psi + \left(\frac{e^2}{4\pi\epsilon_0 r} + W\right)\psi = 0\ , \tag{1.37}$$

wobei der auf $\psi(\vec{r})$ angewendete Laplaceoperator in Kugelkoordinaten

$$\Delta\psi(\vec{r}) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} \tag{1.38}$$

einzusetzen ist.
Die Lösung von Gl. (1.37) ist bekannt und erfolgt, ähnlich wie die Lösung der aus der Elektrostatik bekannten Laplacegleichung, über die Separation der Variablen mit Hilfe des Produktansatzes

$$\psi(r, \phi, \theta) = R(r)\ \varphi(\phi)\ \vartheta(\theta)\ . \tag{1.39}$$

Nach etwas Rechnung ergeben sich die drei entkoppelten Differentialgleichungen:

Für $\varphi(\phi)$

$$\frac{d^2\varphi}{d\phi^2} + m_l^2\ \varphi = 0\ . \tag{1.40}$$

Für $\vartheta(\theta)$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\vartheta}{d\theta}\right) + \left(l(l+1) - \frac{m_l^2}{\sin^2\theta}\right)\vartheta = 0\ . \tag{1.41}$$

Für $R(r)$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + \left(\frac{2m}{\hbar^2}\left(\frac{e^2}{4\pi\epsilon_0 r} + W\right) - \frac{l(l+1)}{r^2}\right)R = 0\ . \tag{1.42}$$

Die Lösungsfunktionen für die Differenzialgleichung in $\varphi(\phi)$ sind bekannt:

$$\varphi(\phi) = Ae^{jm_l\phi}\ . \tag{1.43}$$

Aus der Bedingung der stehenden Welle geht hervor, dass $\varphi(\phi)$ nach jedem Umlauf $\phi + 2\pi$ den gleichen Wert haben muss. Es muss also gelten

$$Ae^{jm_l\phi} = Ae^{jm_l(\phi+2\pi)}\ . \tag{1.44}$$

Hieraus folgt, dass $m_l = 0, \pm1, \pm2, \ldots$ sein muss. m_l wird magnetische Quantenzahl genannt.
Die Differenzialgleichung für $\vartheta(\theta)$ hat Lösungen, vorausgesetzt l ist eine ganze Zahl, die größer gleich $|m_l|$ ist (ohne Beweis). Daraus folgt $m_l = 0, \pm1, \pm2, \ldots \pm l$. Die Konstante l wird Nebenquantenzahl genannt. Häufig werden anstelle der Zahlen Buchstaben verwandt. Es gelten die Definitionen

s für $l = 0$,
p für $l = 1$,
d für $l = 2$ und
f für $l = 3$.

Die Lösung von $R(r)$ für den radialen Anteil erfordert als Nebenbedingung, dass die Gesamtenergie positiv ist oder einen der negativen Werte W_n annimmt, die bereits anhand des Bohrschen Atommodells in Gl. (1.19) ermittelt wurden:

$$W_n = \frac{-m\,e^4}{8\epsilon_0^2 h^2}\frac{1}{n^2}, \quad n = 1, 2, 3, \ldots\ . \tag{1.45}$$

Hierbei darf n nur gleich oder größer $l+1$ sein. Die Konstante n wird als Hauptquantenzahl bezeichnet. Für die Quantenzahlen ergeben sich daher die bekannten Bedingungen:

Hauptquantenzahl $n = 1, 2, 3, \ldots,$
Nebenquantenzahl $l = 0(s), 1(p), 2(d), \ldots, (n-1),$
Magnetische Quantenzahl $m = 0,\ \pm 1,\ \pm 2,\ \ldots \pm l.$

In Analogie zur gebräuchlichen Schreibweise ist der Index l bei m_l weggelassen worden, da es hier nicht zur Konfusion mit der Elektronenmasse kommen kann.

Hinzu kommt noch die vierte bekannte Quantenzahl, die

Spinquantenzahl $s = +\frac{1}{2},\ -\frac{1}{2},$

die unabhängig von allen anderen Quantenzahlen ist. Sie kann zwei verschiedene Werte $+\frac{1}{2}$, $-\frac{1}{2}$ (spin-up, spin-down) annehmen. Sie ergibt sich nicht aus der Lösung der Schrödingergleichung, sondern muss über die erweiterte Form der Diracgleichung ermittelt werden.
Zu jedem Satz von Quantenzahlen ergeben sich Lösungen

$$\psi(\vec{r}) = \psi_{n,l,m,s}(\vec{r}) \tag{1.46}$$

mit einer zugehörigen Energie

$$W = W_{n,l,m,s}\ . \tag{1.47}$$

Beispiel: Für $n = 1$, $l = 0$, $m = 0$ ergeben sich (ohne Beweis) die Lösungen
$\varphi(\phi) = 2\pi^{-\frac{1}{2}}$,
$\vartheta(\theta) = 2^{-\frac{1}{2}}$,
$R(r) = 2\ a_0^{-\frac{3}{2}} e^{-\frac{r}{r_1}}$ mit r_1 als kleinstem Radius des Bohrschen Atommodells nach Gl. (1.14).
Einsetzen in Gl. (1.42) (zur Übung) liefert W_1 nach Gl. (1.45).

Die Energie zu einem bestimmten Satz Quantenzahlen kann, muss sich aber nicht von der Energie eines anderen Satzes Quantenzahlen unterscheiden. Sie wird nach Gl. (1.45) nur durch die Hauptquantenzahl n bestimmt. Sind

die Gesamtenergien für verschiedene Sätze von Quantenzahlen gleich, so spricht man von einer Entartung. Für das Wasserstoffatom (und nur da) sind alle Zustände mit m, l und s für die gleiche Hauptquantenzahl n entartet.

Werden für die verschiedenen Quantenzahlen aus den Lösungen der Wellenfunktion $\psi_{n,l,m,s}(\vec{r})$ die Aufenthaltswahrscheinlichkeit des Elektrons nach Gl. (1.34) bestimmt, so ergeben sich die bekannten Orbitalformen, von denen einige Beispiele in Abb. 1.7 gezeigt sind. Insbesondere zeigen sich die kugelsymmetrischen s-Orbitale und die hantelförmigen p-Orbitale. Sie geben an, wie hoch die Wahrscheinlichkeit ist, das Elektron in einem bestimmten Raumsegment zu finden.

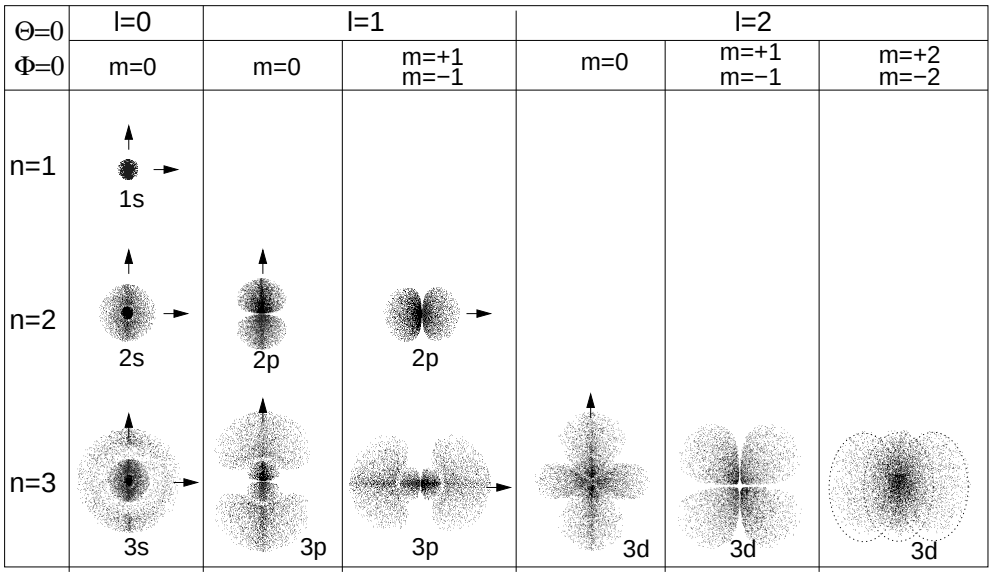

Abb. 1.7: Aufenthaltswahrscheinlichkeiten des Elektrons im Wasserstoffatom bei Betrachtung im Winkel $\Theta = 0$, $\Phi = 0$. Bei rotationssymmetrischen Orbitalen kennzeichnet der Pfeil die Rotationsachse. Die beiden nicht rotationssymmetrischen $3d$-Orbitale besitzen die gleiche Form, jedoch eine unterschiedliche Anordnung im Raum.

Orbital	n	l	m	s	Zustände für jedes l	Zustände für jedes n
1s	1	0	0	$\frac{1}{2}$, $-\frac{1}{2}$	2	2
2s	2	0	0	$\frac{1}{2}$, $-\frac{1}{2}$	2	
2p		1	-1,0,1	$\frac{1}{2}$, $-\frac{1}{2}$	6	8
3s	3	0	0	$\frac{1}{2}$, $-\frac{1}{2}$	2	
3p		1	-1,0,1	$\frac{1}{2}$, $-\frac{1}{2}$	6	
3d		2	-2, -1, 0, 1, 2	$\frac{1}{2}$, $-\frac{1}{2}$	10	18
4s	4	0	0	$\frac{1}{2}$, $-\frac{1}{2}$	2	
4p		1	-1, 0, 1	$\frac{1}{2}$, $-\frac{1}{2}$	6	
4d		2	-2, -1, 0, 1, 2	$\frac{1}{2}$, $-\frac{1}{2}$	10	
4f		3	-3, -2, -1, 0, 1, 2, 3	$\frac{1}{2}$, $-\frac{1}{2}$	14	32

Tabelle 1.1: Mögliche Quantenzustände des Elektrons des Wasserstoffatoms bis $n = 4$.

1.7 Quantenzustände des Elektrons

Für die später benötigte Herleitung der Bändertheorie ist die Kenntnis der möglichen Quantenzustände und deren Besetzung wichtig. Jeder Quantenzustand ergibt sich als Lösung $\psi_{n,l,m,s}$ über die Wahl der Quantenzahlen, wobei die im letzten Kapitel dargestellten Abhängigkeiten und Einschränkungen von n, l, m und s zu beachten sind. Tabelle 1.1 zeigt die möglichen Zustände des Elektrons für die ersten vier Hauptquantenzahlen. Die Gesamtzahl der möglichen Quantenzustände bis zum $4f$-Orbital ist demnach 2+8+18+32 = 60.

1.8 Lösung der Schrödingergleichung für allgemeinen Atomaufbau

Um die Schrödingergleichung für einen allgemeinen Atomaufbau mit einer bestimmten Anzahl Z Elektronen und Protonen zu lösen, ist formal die gleiche Vorgehensweise wie für das Wasserstoffatom gezeigt, möglich.
Die potentielle Energie der Elektronen ergibt sich aus der potentiellen Energie der Elektronen im Feld des Atomkerns sowie der Abstoßung der Z Elektronen untereinander. Die Wellenfunktion $\psi = \psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_Z)$ hängt damit von den Ortsvektoren $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_Z$ ab, wobei jeder Ortsvektor durch

drei Variablen (x, y, z bzw. r, ϕ, θ) beschrieben wird. (vgl. Abb. 1.8).

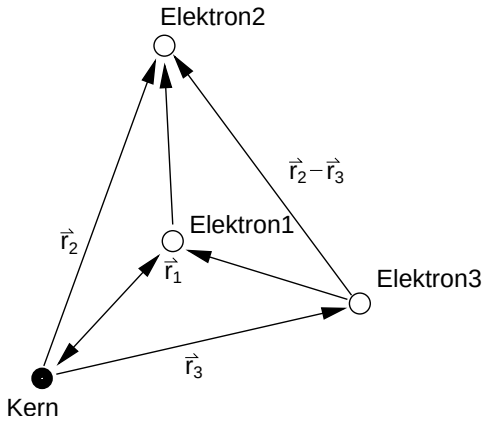

Abb. 1.8: Darstellung eines allgemeinen Atomaufbaus mit Ortsvektoren $\vec{r}_n$, zu den einzelnen Elektronen.

Die Wahrscheinlichkeitsdichte $|\psi|^2$ gibt dann an, wie groß die Wahrscheinlichkeit ist, das Elektron 1 am Ort $\vec{r}_1$ und gleichzeitig die Elektronen $2, 3, \ldots, z$ am Ort $\vec{r}_2, \vec{r}_3, \ldots, \vec{r}_z$ zu finden.
Eine analytische Lösung der Schrödingergleichung selbst für das einfache Heliumatom mit $Z = 2$ ist nicht mehr möglich. Es bietet sich jedoch die Möglichkeit, numerische Verfahren und/oder Näherungsmethoden zur Lösung einzusetzen.
Eine sehr einfache und für den hier angestrebten Verständnis-Anspruch ausreichende Näherung beruht auf der Übertragung der Lösungen des Wasserstoffatoms auf alle Atome. Hierbei wird jeweils nur ein Elektron des Atoms betrachtet. Die restlichen $Z - 1$ Elektronen befinden sich entsprechend ihrer Wahrscheinlichkeit verteilt in ihren Orbitalen um den Atomkern. Die einfache Näherung besteht darin, anzunehmen, dass das betrachtete Elektron als effektive Ladung, die sein Potentialfeld erzeugt, eine positive, verschmierte (verteilte) Ladung e sieht, da die $Z - 1$ negativen verteilten Ladungen der restlichen Elektronen vom Kern kompensiert werden.
Das Elektron dieses allgemeinen Atoms befindet sich daher, ähnlich wie das Elektron im Wasserstoffatom, im Potential einer positiven Elementarladung. Jedoch ist diese Elementarladung nicht mehr punktförmig im Atomkern lokalisiert, sondern verteilt (verschmiert).
Mit dieser Näherung erhalten wir für das eine Elektron Lösungen der Schrödingergleichung, die analog zu den Lösungen des Wasserstoffatoms

gewonnen werden. Aufgrund der Ähnlichkeit der betrachteten Struktur lassen sich die folgenden Aussagen machen:

- Es ergeben sich aufgrund der Näherung Abweichungen in der Orbitalform, jedoch bleibt die Kugelsymmetrie der s-Orbitale und die Hantelform der p-Orbitale erhalten.
- Jede Lösung der Schrödingergleichung $\psi_{n,l,m,s}$ wird wieder durch die vier Quantenzahlen n, l, m, s bestimmt.
- Die Lösungen $\psi_{n,l,m,s}$ werden ähnlich denen des Wasserstoffatoms sein.
- Zu jedem Satz von Quantenzahlen gehören wieder Energien $W_{n,l,m,s}$. Deren Werte unterscheiden sich jedoch von denen im Wasserstoffatom.
- Die Entartung der Energie bezüglich l, m, s wie beim Wasserstoffatom muss nicht mehr vorliegen. D. h. bei der selben Quantenzahl n können durch l, m, s bestimmte verschiedene Energien vorliegen.

Basierend auf diesen einfachen Aussagen lässt sich der Aufbau der Elektronenhülle von beliebigen Atomen verstehen. Es existieren entsprechend den Quantenzuständen Orbitale, die mögliche Zustände für ein Elektron darstellen. Bei einem Atom mit Z Elektronen werden Z dieser Zustände besetzt.

1.9 Pauli-Prinzip

Die Besetzung der möglichen Zustände regelt das Paulische Ausschließungsprinzip. Es besagt:

Fermionen dürfen nie Zustände einnehmen, die in allen Quantenzahlen übereinstimmen.

Fermionen sind Teilchen mit einem halbzahligen Spin. Daher sind auch Elektronen Fermionen und folgen dem Paulische Ausschließungsprinzip. Damit ist jedem Satz Quantenzahlen genau ein Elektron in einem Zustand zugeordnet. Das Pauli-Prinzip gilt insbesondere in allen Atomverbünden (z. B. Molekül, Kristall) in denen sich die Wellenfunktionen der Lösungen der Schrödingergleichung der einzelnen Atome zu einer Gesamtlösung überlagern.

1.10 Elektronischer Aufbau der Elemente

Bei der Besetzung der Orbitale gelten folgen Regeln:

- Die energetisch am tiefsten liegenden Zustände werden zuerst von Elektronen besetzt.
- Jeder Zustand mit halbzahligem Spin enthält nur ein Elektron (Pauli-Prinzip).
- Erst wenn alle Zustände einer Nebenquantenzahl l mit einem Elektron besetzt sind, werden sie durch ein Elektron mit entgegengesetztem Spin ergänzt.

Mit diesen Vorschriften lässt sich das in Abb. 1.9 gezeigte Quantenzahl-Energiediagramm zeichnen.

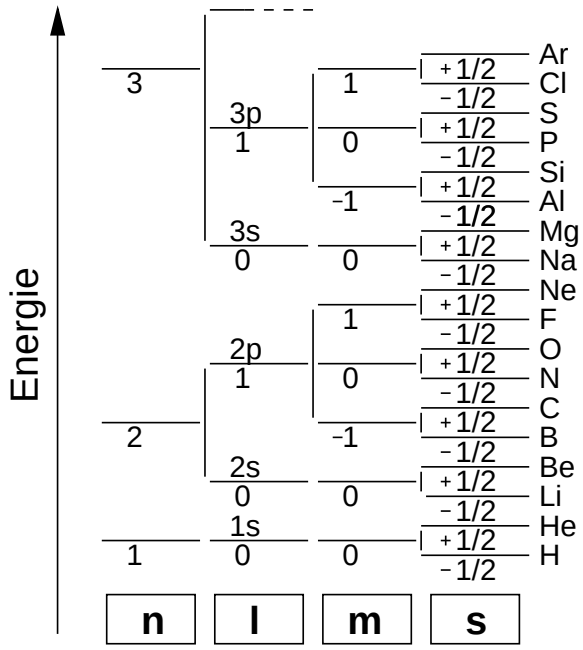

Abb. 1.9: Quantenzahl-Energiediagramm für die ersten 18 Elemente des Periodensystems.

Die Lage der Energiezustände darin ist in der Reihenfolge der Quantenzahlen dargestellt. Diese Anordnung ist als prinzipielles qualitatives Schema zu verstehen. Im Detail kann allein schon wegen möglicher Entartungen das Schema nicht stimmen.

Ein genaueres Bild der sich mit der Ordnungszahl ändernden Bindungsenergien zeigt Abb. 1.10.

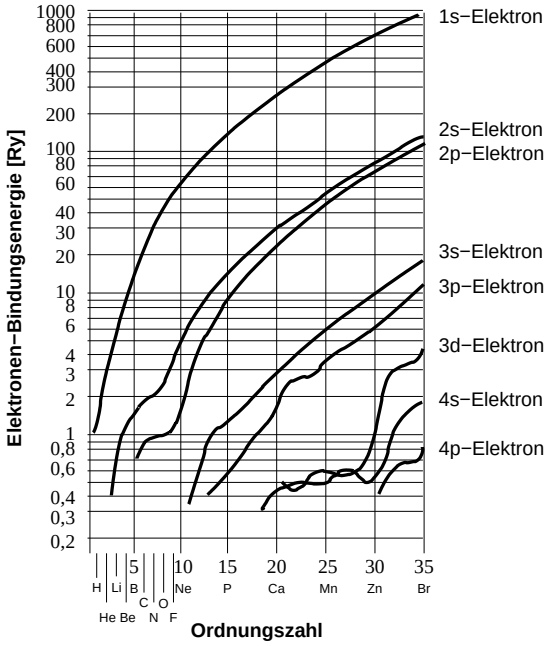

Abb. 1.10: Bindungsenergien der Elektronen für Elemente mit verschiedenen Ordnungszahlen. Die Einheit 1 Ry (Rydberg) entspricht 13,6 eV. Das ist die Energie des Grundzustandes des **H**-Atoms.

Im Großen jedoch behält die vereinfachte Darstellung in Abb. 1.9 ihre Gültigkeit, da Elektronen in Zuständen mit kleinen Quantenzahlen eine hohe Aufenthaltswahrscheinlichkeit in Nähe des Atomkerns haben. Ihnen muss eine hohe Energie zugeführt werden, um sie aus dem Atom zu lösen, so dass sie zu freien Elektronen werden.
Aus Abb. 1.9 liest man für das in der Halbleitertechnik verwendete Material Silizium eine Besetzung der Zustände

$$1s^2,\ 2s^2,\ 2p^6,\ 3s^2,\ 3p^2$$

ab. In dieser Schreibweise zeigt die hochgestellte Zahl die Anzahl der Zustände in der jeweiligen Nebenquantenzahl l an.
Abb. 1.9 zeigt auch, dass die Edelgase **He**, **Ne**, **Ar (, Kr, Xe, Rn)** sich dadurch auszeichnen, dass sie gefüllte Orbitale („Schalen“) besitzen. Bei gefüllten Orbitalen ist die Energie des Elektronensystems minimal im Vergleich zu benachbarten Konfigurationen. Daher besteht bei den Edelgasen auch keine

Tendenz zur Veränderung, denn die findet nur statt, wenn sich dadurch die Systemenergie senken lässt. Dies ist auch der Grund, warum Elemente mit abgeschlossenen Schalen nur in Gasform vorliegen: Die Energie, die aufgrund einer Bindung angenommen wurde, ist höher als die Energie im ungebundenen Zustand. Atome mit nicht abgeschlossenen Schalen gehen dagegen Bindungen ein, um eine abgeschlossene Schale und damit ein energetisch niedrigeres Niveau zu erhalten.

1.11 Potentialtopfdarstellung

Wir können die energetischen Zustände der Elektronen eines Atoms schematisch in einer Potentialtopfdarstellung ähnlich Abb. 1.3 erfassen. Die Potentialtopfdarstellung dient nur zur Veranschaulichung und ist rein qualitativ. Abb. 1.11 zeigt als Beispiel den Potentialtopf eines **Na**-Atoms.

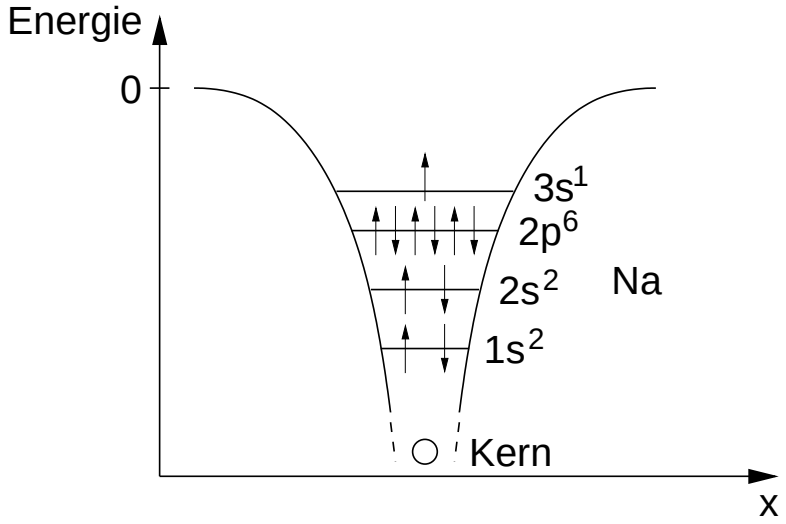

Abb. 1.11: Potentialtopfdarstellung eines **Na**-Atoms. Die Pfeile repräsentieren die Spin-Orientierung des jeweiligen Elektrons.

1.12 Bindungstypen

Atome gruppieren sich dann zu einem Festkörper, wenn sich dadurch ein energetisch günstiger Zustand einstellt. Aus Sicht zweier Atome, die ein Molekül bilden, wirken zwischen den beiden Atomen anziehende und abstoßende Kräfte. Eine stabile in sich geordnete Anordnung liegt im Gleichgewicht dieser Kräfte vor.

1.12.1 Ionenbindung

Ionenbindungen ergeben sich zwischen verschiedenen Atomen, die für eine Edelgaskonfiguration zum einen zuviel und zum anderen zuwenig Elektronen auf der äußeren Schale haben. Durch Transfer der Elektronen von einem zum anderen Atom gewinnen beide eine abgeschlossene äußere Schale. Durch die Abgabe eines Elektrons wird das eine Atom positiv geladen, das aufnehmende Atom wird negativ geladen.
Das anziehende Potential ist das bereits mehrfach verwendete elektrostatische Potential (Coulombanziehung) zwischen den beiden Ladungen.
Das Potential ist entsprechend Gl. (1.16) der Form

$$W_{pot}^{-} = -\frac{\alpha}{r} \, . \tag{1.48}$$

Kommen sich die beiden ionisierten Atome ausreichend nahe, so überwiegen die abstoßenden Kräfte zwischen den vielen Elektronen der inneren Schale gegenüber der anziehenden Kraft der einen Elementarladung. Da sich die abstoßende Kraft sehr viel schneller mit r ändert, lässt sich schreiben

$$W_{pot}^{+} = \frac{\beta}{r^n} \, , \tag{1.49}$$

wobei n im Bereich $6 \ldots 12$ liegt. Aus den Energien der anziehenden und abstoßenden Kräfte ergibt sich das Bindungspotential der Ionenbindung von zwei Atomen

$$W_{pot} = -\frac{\alpha}{r} + \frac{\beta}{r^n} \, . \tag{1.50}$$

Diese lässt sich durch einfache Überlegung auf einen ganzen Festkörper (Kristall bei Ionenbindung) übertragen. In dem Kristall überlagern sich die anziehenden und abstoßenden Potentiale aller benachbarten Ionen. Der abstoßende Teil der Nachbaratome ist jedoch aufgrund der starken Abhängigkeit vom Abstand ($r^{6\ldots12}$) vernachlässigbar. Der anziehende Anteil ergibt sich aus der Überlagerung der Energien aller anziehenden und abstoßenden Kräfte. Dabei sind die für jeden Kristalltyp unterschiedlichen Abstände der Ionen zu berücksichtigen. Es ergibt sich aufgrund der linearen Abhängigkeit das Bindungspotential eines Ionenkristalls

$$W_{pot} = \alpha_m \frac{-e^2}{4\pi\epsilon_0 r} + \frac{\beta}{r^n} \, . \tag{1.51}$$

Darin ist α_m die Mandelung-Konstante. Sie beträgt für den kubisch primitiven **NaCl**-Kristall 1,748.

Für den hexagonalen Kristall $\mathbf{SiO_2}$, der in integrierten Schaltungen als Isolator z. B. für Metallisierungen und für die Gate-Elektrode in MOS-Transistoren dient, ist $\alpha_m = 2,219$.

Beispiel: Im **NaCl**-Kristall ist jedes $\mathbf{Cl}^-$-Ion von 6 $\mathbf{Na}^+$-Ionen im Abstand r umgeben. Darauf folgen im Abstand $2^{\frac{1}{2}}r$ 12 $\mathbf{Cl}^-$ Ionen und darauf im Abstand $3^{\frac{1}{2}}r$ 8 $\mathbf{Na}^+$ Ionen und so weiter. Daraus ergibt sich

$$W_{pot}^{-} = \frac{-e^2}{4\pi\epsilon_0 r}(6 - \frac{12}{2^{\frac{1}{2}}} + \frac{8}{3^{\frac{1}{2}}} - \frac{6}{4^{\frac{1}{2}}} \ldots) \quad ,$$

woraus $\alpha_m = 1,748$ folgt.

Abb. 1.12 zeigt die Verläufe der Energien von abstoßenden und anziehenden Kräften in allgemeiner Form, wobei auch der Radius der anziehenden Energie mit einem Exponenten m versehen wurde. Im Beispiel der Ionenbildung nach Gl. (1.51) ist $m = 1$.

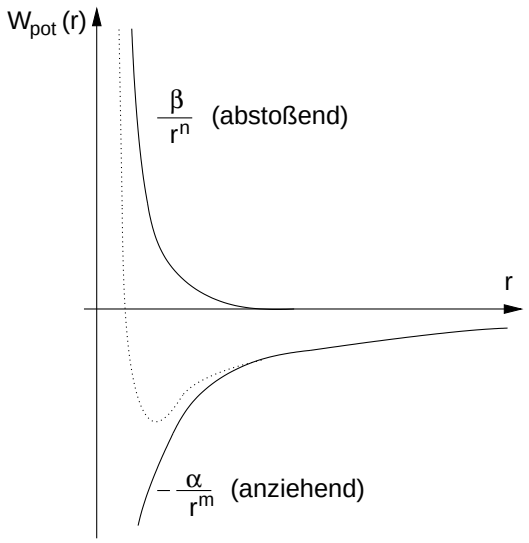

Abb. 1.12: Verläufe der potentiellen Energien von abstoßenden und anziehenden Kräften bei der Bindung von Atomen in Abhängigkeit des Atomabstandes r.

Der für die Bindung stabile Abstand ergibt sich im Gleichgewicht von abstoßender und anziehender Kraft. Die Kraft ergibt sich im Eindimensionalen aus der negativen Ableitung des gesamten Potentials $W = W_{pot}^{+} + W_{pot}^{-}$:

$$\vec{F} = -\frac{d\,W(x)}{d\,x}\vec{e_x} \tag{1.52}$$

bzw. im Dreidimensionalen durch die Bildung des Gradienten

$$\vec{F} = -\,\mathrm{grad}\; W(\vec{r})\;. \tag{1.53}$$

Der Gleichgewichtszustand ist also bei der Entfernung r_0, bei der $\vec{F} = 0$ gilt, also W in Abb. 1.12 ein Minimum besitzt.

1.12.2 Kovalente Bindung

Die kovalente Bindung ist für die Halbleitertechnologie von hohem Interesse, da sie unter anderem die Bindung in den für Halbleiterbauelemente verwendeten Materialien **Si**, **Ge** und **GaAs** bestimmt. Auch **C** in der Diamantstruktur hat eine kovalente Bindung.
Das Prinzip der kovalenten Bindung beruht darauf, dass jedes an der Bindung beteiligte Atom zu wenig Elektronen hat, um eine abgeschlossene Schale zu besitzen. Z. B. sind die äußeren Schalen bei **Si**: $3s^2$, $3p^2$, **Ge**: $4s^2$, $4p^2$. Beide Materialien benötigen also vier Elektronen um eine abgeschlossene $3p$- bzw. $4p$-Schale zu erhalten. Durch eine Anordnung (dreidimensional), bei der sich um jedes Atom herum vier nächste Nachbarn gruppieren, teilt sich jedes Atom mit seinem Nachbarn die vier Elektronen auf seiner äußeren Schale und erhält so die gewünschte abgeschlossene Schale mit 8 Elektronen. Abb. 1.13 links zeigt dies in einem einfachen zweidimensionalen Modell zur Veranschauung.
Im Dreidimensionalen ergibt sich aus dieser Gruppierung die auf der rechten Seite gezeigte Anordnung der Atome in einem Kristallgitter in Diamantstruktur.
In der Vorstellung als Wellenmodell ergibt sich die kovalente Bindung durch die Überlappung (Überlagerung) der Wellenfunktionen der äußeren Orbitale. Es bilden sich dabei aus den s- und p-Orbitalen neue Misch- oder Hybridorbitale. Dies ist möglich aufgrund der Linearität der Schrödingergleichung, die es erlaubt, dass bei Lösungen ψ_1 und ψ_2 ($\psi_3, \psi_4 \ldots$) auch Linearkombinationen $a\psi_1 + b\psi_2$ Lösungen sind. Es bilden sich aus den vier Orbitalwellenfunktionen $\psi_s, \psi_{p_x}, \psi_{p_y}, \psi_{p_z}$ des einzelnen Atoms die Linearkom-

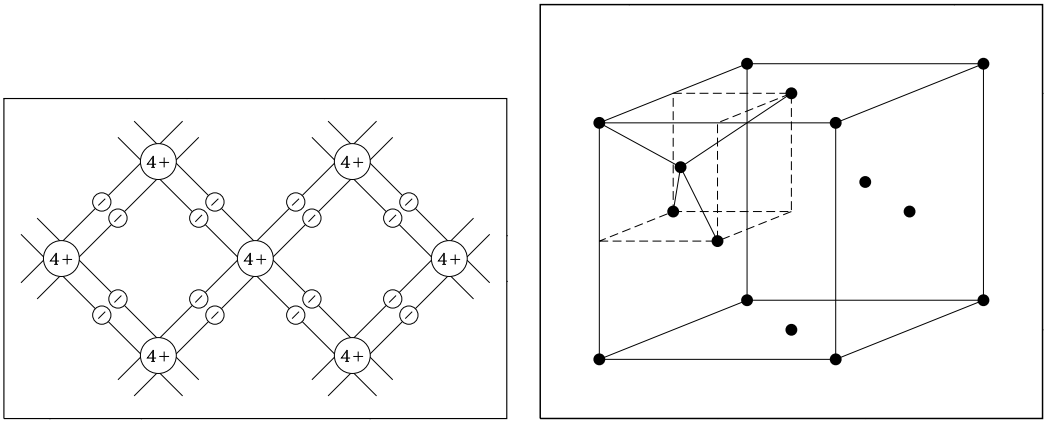

Abb. 1.13: Links: Zweidimensionales Modell zur Veranschaulichung der kovalenten Bindung z. B. bei **Si** oder **Ge**. Die Rümpfe der Atome ohne die vier Elektronen der äußeren Schale besitzen eine Ladung von +4e.
Rechts: Dreidimensionale Anordnung der Atome in einem Kristallgitter mit Diamantstruktur (**Si**, **Ge**).

binationen der Wellenfunktionen

$$\begin{aligned}\psi_1 &= \frac{1}{2}(\psi_s + \psi_{p_x} + \psi_{p_y} + \psi_{p_z}) \\ \psi_2 &= \frac{1}{2}(\psi_s + \psi_{p_x} - \psi_{p_y} + \psi_{p_z}) \\ \psi_3 &= \frac{1}{2}(\psi_s - \psi_{p_x} + \psi_{p_y} - \psi_{p_z}) \\ \psi_4 &= \frac{1}{2}(\psi_s - \psi_{p_x} - \psi_{p_y} - \psi_{p_z})\ .\end{aligned} \tag{1.54}$$

Diese bilden die für die Bindung geeigneten vier sp^3-Orbitale[3] aus, die von jedem an der Bindung beteiligten Atom mit je einem zusätzlichen Elektron besetzt sind.
Abb. 1.14 links zeigt die Darstellung der reinen s- und p-Orbitale, die für eine kovalente Bindung in einem dreidimensionalen Kristallgitter nicht geeignet sind. Dort ist das s-Orbital mit zwei Elektronen voll besetzt. In den 6 Keulen der p-Orbitale, die mit den Orbitalen der Nachbaratome überlappen, befinden sich nur zwei Elektronen. Die rechte Seite zeigt die Form

[3]Entgegen der bisher verwandten Darstellung kennzeichnet bei Hybridorbitalen die hochgestellte Zahl nicht die Anzahl der Elektronen, sondern die Anzahl der an der Hybridisierung beteiligten Orbitale.

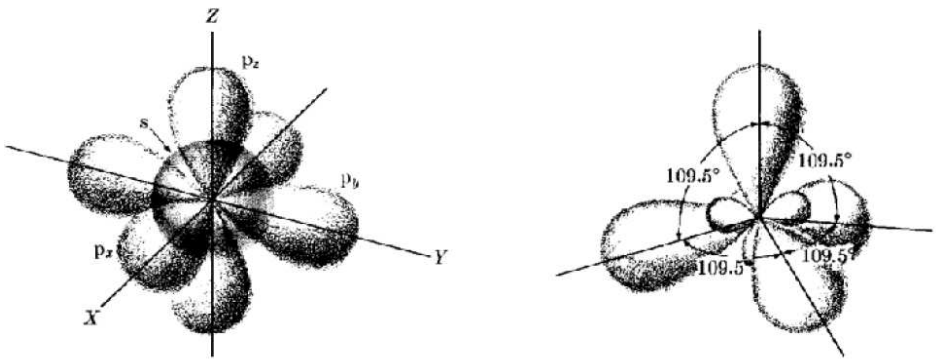

Abb. 1.14: p-Orbitale als Bindungsarme der kovalenten Bindung. Links: vor der Hybridisierung (einzelnes Atom). Rechts: Hybridisierte sp^3-Orbitale.

der sp^3-Hybridorbitale. Hier sind in jeder der Keulen der Hybridorbitale jeweils ein Elektron, das durch die Überlappung mit den Orbitalen der angrenzenden Atome um ein weiteres Elektron ergänzt wird, wodurch sich voll besetzte Orbitale bilden.
Die räumliche Anordnung der Hybridorbitale mit den Winkeln von $109,5°$ stimmt mit der in Abb. 1.13 rechts gezeigten Diamantstruktur überein. Darin sitzt das Atom in der Mitte eines gleichseitigen Tetraeders und die Keulen zeigen vom Zentrum in die Ecke des Tetraeders mit den Tetraederwinkeln $109,5°$.
In diesem Sinne muss man die in Abb. 1.13 links eingezeichneten Verbindungslinien zwischen den Elektronen und den Atomrümpfen sehen. Sie symbolisieren in vereinfachter, zweidimensionaler Darstellung die Bindungsarme in Form der Keulen der Hybridorbitale.

Ähnlich wie bei der Ionenbindung treten auch bei der kovalenten Bindung anziehende und abstoßende Kräfte auf. Die Gesamtenergie dieser Kräfte lässt sich wieder ähnlich wie in Gl. (1.50) darstellen:

$$W = -\frac{\alpha}{r^n} + \frac{\beta}{r^m} \,. \tag{1.55}$$

Auch bei der Energie der anziehenden Kraft gibt es hier jetzt einen Exponenten n, der größere Werte als 1 annimmt.
Auch hier gibt es wieder einen Potentialtopf mit Minimum für ein Kräftegleichgewicht, wie schon in Abb. 1.12 gezeigt.

1.12.3 Metallbindung

Bei der Metallbindung haben die Atome, die sich verbinden wollen, zu viele Elektronen um eine abgeschlossene Schale zu erhalten. In diesem Fall geben die Atome, wie in Abb. 1.15 gezeigt, die überschüssigen Elektronen an den Festkörper ab.

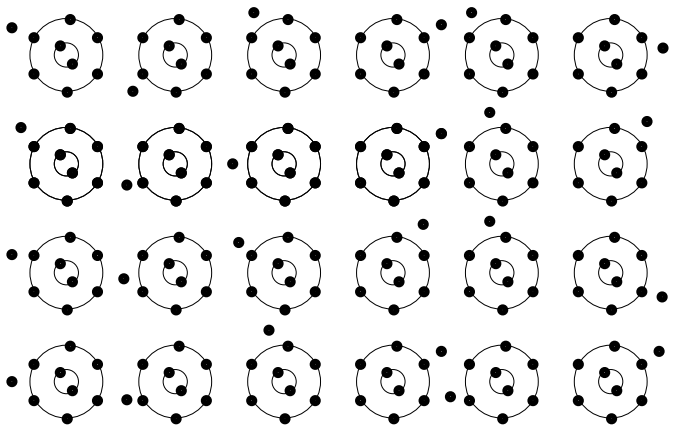

Abb. 1.15: Beispiel zur Metallbindung: $\mathbf{Na}^+$-Ionenrümpfe sind umgeben von negativ geladenen Elektronen, die frei beweglich sind und ein Elektronengas bilden.

Die abgegebenen Elektronen sind frei und bilden ein Elektronengas aus negativ geladenen Elektronen, das innerhalb des Festkörpers beweglich ist. Die Atomrümpfe sitzen ortsfest mit positiver Ladung (das Metall ist makroskopisch elektrisch neutral).
Die Wellenfunktionen von freien Elektronen werden in einem späteren Kapitel berechnet. Sie zeigen, dass ein Elektron überall im Festkörper mit der gleichen Wahrscheinlichkeit zu finden ist. Die Bindungskräfte zwischen dem negativ geladenen Elektronengas und den positiv geladenen Atomrümpfen sind daher völlig ungerichtet. Sie lassen sich dennoch am besten mit der auch schon für die kovalente und Ionenbindung verwandte Gl. (1.55) beschreiben.

1.13 Bindungs-Ionisationsenergie, Elektronenaffinität

Ein Prozess läuft ohne Zwang (Zufuhr von Energie) nur dann ab, wenn bei dem Prozess Energie frei wird, d. h. das System einen energetisch niedrigeren Wert annimmt. Das ist der Grund für Atome, Bindungen einzugehen.
Die bei der Bindung frei werdende Energie heißt Bindungsenergie.

Um eine Ionenbindung einzugehen (z. B. **NaCl**) muss sich das **Na** von seinem äußeren Elektron trennen. Dazu ist die Energie zuzuführen, die notwendig ist, um das Elektron gegen die elektrostatische Anziehung des + geladenen Ions in das Unendliche zu bringen. Dazu muss Arbeit verrichtet werden. Die hierdurch erzeugte potentielle Energie wird in der Einheit

$$1\,\mathrm{eV} = 1\,\text{Elektronenvolt} \tag{1.56}$$

gemessen. Sie entspricht der Energie, die ein Elektron gewinnt, wenn es eine Spannung von 1 V durchläuft. Die für das Abtrennen eines Elektrons notwendige Energie wird Ionisationsenergie genannt. Für die Ionisation des **Na**-Atoms ist z. B. die Zufuhr einer Ionisationsenergie von +5,14 eV (pos. Vorzeichen, da Zufuhr) notwendig.
Das Partneratom einer Ionenverbindung muss ein Elektron aufnehmen. Im Fall des **Cl** wird bei der Anlagerung eine Energie von 3,61 eV frei (negatives Vorzeichen der Energie). Diese freiwerdende Energie wird als Elektronenaffinität bezeichnet. Das ist weniger als die Ionisationsenergie von 13,01 eV des **Cl**.

Beispiel: Die Energiebilanzen bei der Bildung der **NaCl**-Ionenbindung lauten im einzelnen

$\mathbf{Na}$+5,14 eV (Ionisierungsenergie) $\Rightarrow$ $\mathbf{Na}^+ + \mathrm{e}^-$
$\mathbf{Cl} + \mathrm{e}^-$ (Anlagerung) $\Rightarrow$ $\mathbf{Cl}^-$ + 3,61 eV (Elektronenaffinität)
$\mathbf{Na}^+ + \mathbf{Cl}^-$ $\Rightarrow$ $\mathbf{Na}^+\mathbf{Cl}^-$ + 7,9 eV (Bindungsenergie)

Wir schreiben die gesamte Energiebilanz:

	Energie (**Na**-Atom)	+	Ionisierungsenergie (I)
+	Energie (**Cl**-Atom)	-	Elektronenaffinität (A)
=	Energie ($\mathbf{Na}^+\mathbf{Cl}^-$ Molekül)	+	Bindungsenergie (B)

oder kurz $\mathrm{W}(\mathbf{Na})+\mathrm{I}+\mathrm{W}(\mathbf{Cl})-\mathrm{A} = \mathrm{W}(\mathbf{Na}^+\mathbf{Cl}^-)+\mathrm{B}$
umgestellt $\mathrm{W}(\mathbf{Na})+\mathrm{W}(\mathbf{Cl})+\mathrm{I}-\mathrm{A}-\mathrm{B} = \mathrm{W}(\mathbf{Na}^+\mathbf{Cl}^-)$
$\mathrm{W}(\mathbf{Na})+\mathrm{W}(\mathbf{Cl}) \quad +(-7,9-3,61+5,14)\,\mathrm{eV} = \mathrm{W}(\mathbf{Na}^+\mathbf{Cl}^-)$

Die Energie des $\mathbf{Na}^+\mathbf{Cl}^-$-Moleküls ist also um $(-7,9-3,61+5,14)\,\mathrm{eV}$ geringer.

1.14 Energiebänder

Wir fragen uns, welche Änderungen die diskreten Energieniveaus der Elektronen eines einzelnen Atoms erfahren, wenn sie eine Bindung zu einem Festkörper eingegangen sind.
Zur Beantwortung dieser Frage schauen wir uns die Lösungen der Schrödingergleichung für die Wellenfunktion der Elektronen in einem einzelnen Atom genauer an. Speziell betrachten wir die Radial-Komponente, welche die Entfernung zum Atomkern erfasst. Als Näherung betrachten wir die Lösungen für das Wasserstoffatom, die wegen Kap. 1.8 auf alle Atome übertragbar sind. Die ersten sechs Lösungen der Radialkomponente lauten

$$R_{1,0,0}(r) = \frac{2}{a_0^{3/2}} e^{\frac{-r}{a_0}} \tag{1.57}$$

$$R_{2,0,0}(r) = \frac{1}{2\sqrt{2}\, a_0^{3/2}} \left(2 - \frac{r}{a_0}\right) e^{\frac{-r}{a_0}} \tag{1.58}$$

$$R_{2,1,0}(r) = R_{2,1,+1}(r) = R_{2,1,-1}(r) = \frac{1}{2\sqrt{6}\, a_0^{3/2}} \frac{r}{a_0} e^{\frac{-r}{2a_0}} \tag{1.59}$$

$$R_{3,0,0}(r) = \frac{1}{81\sqrt{3}\, a_0^{3/2}} \left(27 - 18\frac{r}{a_0} + 2\frac{r^2}{a_0^2}\right) e^{\frac{-r}{3a_0}} \tag{1.60}$$

Z. B. beschreibt Gl. (1.60) die Abhängigkeit der Wellenfunktion und damit der Aufenthaltswahrscheinlichkeit eines Elektrons im $3s^1$ Zustand. Für ein **Na**-Atom ist dies der letzte besetzte Zustand.
Für die Wellenfunktionen höherer Orbitale ergeben sich für die Radialkomponenten Terme ähnlicher Struktur, immer mit einem Faktor $e^{\frac{-r}{x}}$. Bemerkenswert ist, dass der radius-abhängige Anteil zwar mit dem Faktor $e^{\frac{-r}{x}}$ gegen Null strebt, aber im Endlichen niemals den Wert Null annimmt. Dies führt dazu, dass in einer Anordnung von mehreren Atomen sich deren Wellenfunktionen überlappen. Bei sehr dichten Atomverbänden, wie z. B. in einer Kristallstruktur mit kovalenter Bindung, berühren sich daher die äußeren Orbitale.
Da die Zustände aller Elektronen in einem Kristall Lösungen ein und derselben Schrödingergleichung, aufgestellt für diesen Kristall, sind, kann jedes Elektron aufgrund des Pauli-Prinzips auch nur einen Zustand be-

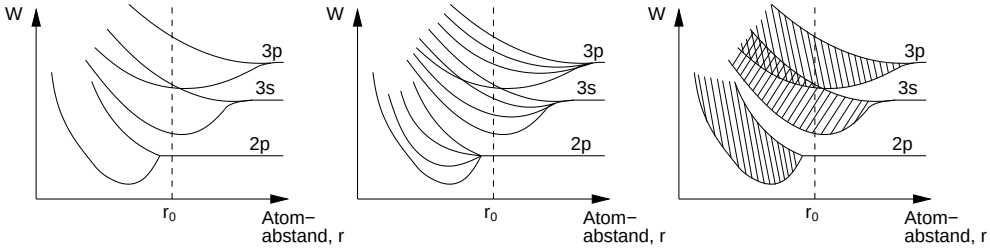

Abb. 1.16: Entstehung von Energiebändern. Zur Konkretisierung der Vorstellung kann angenommen werden, dass es sich um die äußeren Orbitale eines **Na**-Atoms handelt. Links: wenn zwei **Na**-Atome mit zunächst gleichen Energiewerten sich nähern, spalten diese sich auf, da ihre Wellenfunktionen sich überlappen. Mitte: die Anzahl der Aufspaltung stimmt mit der Anzahl der Elektronen in diesem Zustand, also mit der Anzahl der Atome überein (hier 4 Atome). Rechts: sehr große Anzahl von Atomen, durch die die Einzelniveaus zu einem Band kontinuierlich besetzbarer Energiezustände verschmieren.

setzen.[4] Dies bedeutet z. B., dass bei einem Festkörper mit 10^{20} Atomen auch 10^{20} voneinander verschiedene Zustände der Elektronen mit gleichen Quantenzahlen existieren müssen. Es bilden sich also in einem Atomverbund Unterniveaus aus zu den Niveaus eines einzelnen Atoms.
Aus diesem Grund erfahren auch die Energieniveaus der Elektronen des einzelnen Atoms eine Aufspaltung in Unterniveaus.
Dies gilt insbesondere für die Energieniveaus der äußeren Elektronen, da sich deren Wellenfunktionen stark überlappen und damit deren Energieniveaus für die anderen Elektronen im Kristall verfügbar sind.
Da die Unterniveaus der Energie sehr direkt beieinander liegen, werden sie nicht mehr als diskrete Energien wahrgenommen, sondern als Energieband, in dem kontinuierlich alle Energiezustände besetzbar sind. Abb. 1.16 zeigt die Entstehung von Energiebändern in Abhängigkeit des Atomabstandes schematisch. Die Breite eines Bandes ändert sich bei steigender Anzahl von Atomen kaum. Der Zwischenbereich spaltet sich jedoch immer weiter in Unterniveaus auf.
Die tatsächliche Breite des Bandes bestimmt der Atomabstand, der sich aus dem Gleichgewichtszustand der jeweiligen Bindungsart ergibt (vgl.

[4] Aufgrund des von Null verschiedenen Wertes der Wellenfunktion besteht aufgrund der von Null verschiedenen Wahrscheinlichkeit prinzipiell für jedes Elektron im Kristall die Möglichkeit, sich im Orbit eines anderen Atoms zu befinden.

Kap. 1.12). In Abb. 1.16 ist dieser Abstand mit r_0 bezeichnet.
Energieniveaus zu Quantenzuständen mit geringem Abstand zum Atomkern spalten weniger stark auf. Anschaulich lässt sich das durch ihre starke Bindung an den Atomkern und die abschirmende Wirkung der Elektronen auf den äußeren Orbitalen verstehen. Elektronen in den äußeren Orbitalen sind nur noch wenig oder gar nicht mehr an den Atomkern gebunden und reagieren daher stärker auf die Aufspaltung der Energieniveaus. Durch die Aufspaltung der Energieniveaus können diese auch, wie in Abb. 1.16 gezeigt, überlappen, wodurch die Anzahl (Breite) der kontinuierlich besetzbaren Energieniveaus steigt.
Abb. 1.17 zeigt die für eine Metallbindung charakteristischen Energiebänder am Beispiel des **Na**-Kristalls. Dabei sitzen die **Na**-Atome in einem festen Kristallverband und erzeugen ein periodisches Potentialfeld.

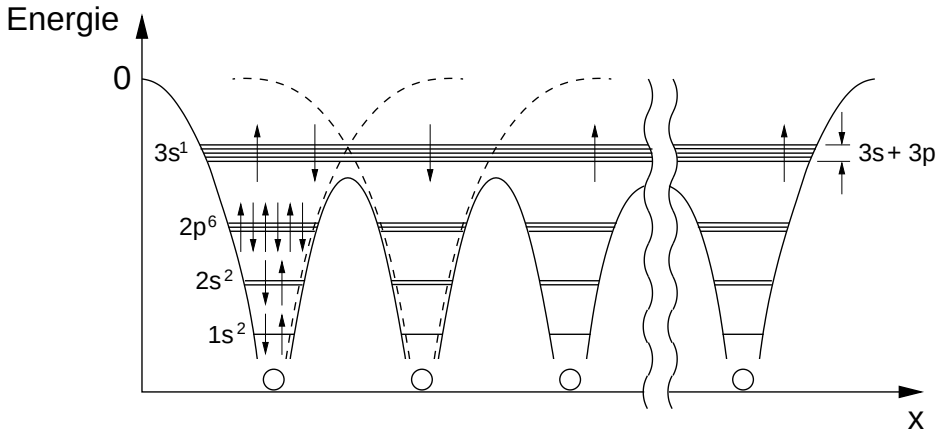

Abb. 1.17: Aufspaltung des $3s^1$ Energieniveaus der Atome eines **Na**-Kristalls in ein Energieband (schwache Aufspaltung auch für das darunter liegende $2p^6$ Niveau). Das periodische Potential entsteht durch Überlagerung der Potentiale der einzelnen Atome (gestrichelte Potentialtöpfe).

Das periodische Potential des Kristalls ergibt sich durch die Überlagerung der Potentialtöpfe der einzelnen Atome (vgl. Abb. 1.11). Außer an den Kristallgrenzen (keine Nachbaratome) ergibt sich dadurch eine Absenkung der Energie der Potentialtöpfe, wodurch die Elektronen des $3s^1$-Niveaus nicht mehr in den Potentialtöpfen liegen. Sie können ihre Energie nicht senken, da es unter ihnen keine freien Quantenzustände mehr gibt. Da sie oberhalb der Potentialberge liegen, können sie leicht von einem Atom

zu einem anderen gelangen. Sie können sich daher in dem Kristallgitter (dreidimensional) frei bewegen. Man nennt sie daher „freie Elektronen“. Aufgrund ihrer Beweglichkeit in allen drei Dimensionen wird ihre Gesamtheit als Elektronengas beschrieben.

Die vereinfachte Modellvorstellung, dass sich die Elektronen als Elektronengas in einem Kristall frei bewegen, heißt jedoch nicht, dass sie wie im freien Raum jede beliebige Energie annehmen können. In dem Kristall gibt es weiterhin nur Wellenfunktionen der einen Schrödingergleichung dieses Kristalls, deren diskrete Quantenzustände nach dem Pauli-Prinzip besetzt werden. Diese Energien sind jedoch so zahlreich und fein abgestuft, dass ein quasi-kontinuierliches Band möglicher Energiezustände angenommen wird.
Das aufgrund der Periodizität des Kristallgitters entstehende periodische Potential erzwingt spezielle Lösungen der Schrödingergleichung, wodurch sich innerhalb des Energiebandes eine Unterteilung in Bereiche mit nicht besetzbarer Energie (Band- oder Energielücke) ergibt. Diese Unterteilung ist außerdem abhängig davon, in welcher Richtung sich die Wellenfunktion durch den dreidimensionalen Kristall ausbreitet.
Wir werden diese Besonderheiten später genauer untersuchen.
Abb. 1.18 zeigt am Beispiel des **Si**-Kristalls die Aufspaltung der Energiebänder in Abhängigkeit des Atomabstandes. Bei der Annäherung der Atome bilden sich, wie in Kap. 1.12.2 erläutert, die sp^3-Hybridorbitale aus. Die Energiebänder der Orbitale entstehen durch Aufspaltung des $3s$- und $3p$-Zustandes in jeweils ein energievergrößerndes und ein energieabsenkendes Teilband. Beide Bänder überlappen bei dem sich im Kristallgitter einstellenden Atomabstand. Zwischen den beiden vermischten Bändern entsteht eine neue Energielücke W_g.

1.15 Elektronenleitung in einem Band

Wir betrachten hier die Elektronenleitung.[5] Um Strom leiten zu können, müssen in einem Festkörper frei bewegliche Elektronen vorhanden sein. Diese Elektronen nehmen Energie auf durch Beschleunigung in einem elektrischen Feld, das von außen durch Anlegen einer Spannung an den Festkörper entsteht.

[5] Im Gegensatz zu Flüssigkeiten und Gasen spielt in Festkörpern die Ionenleitung eine untergeordnete Rolle.

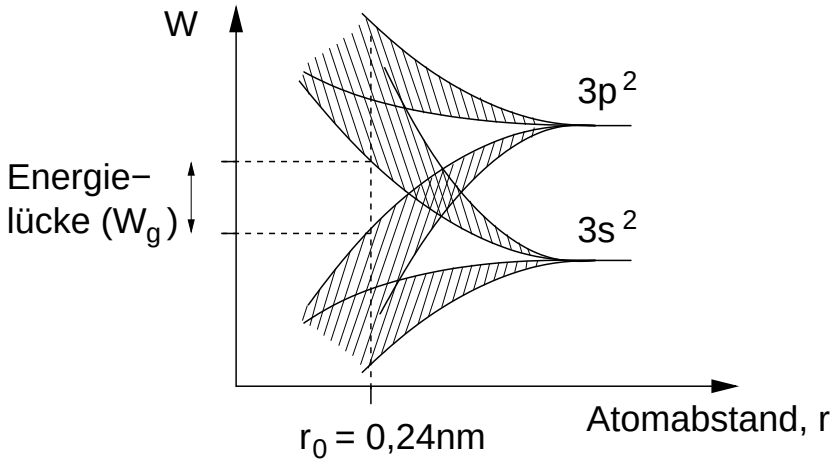

Abb. 1.18: Aufspaltung der $3s^2$- und $3p^2$-Zustände durch Annäherung der **Si**-Atome bei Bildung eines **Si**-Kristalls. Der Atomabstand im **Si**-Kristall beträgt r_0. Die Energiebänder unter- und oberhalb der Energielücke werden als Valenz- und Leitungsband bezeichnet.

Um Energie aufnehmen zu können, muss ein Elektron auf ein höher gelegenes Energieniveau gelangen. Ist dies nicht möglich, weil es in dem Bereich der für das Elektron erreichbaren Energien kein freies Niveau gibt, kann das Elektron nicht beschleunigt werden und steht nicht für die Stromleitung zur Verfügung. Aus dieser Überlagerung lässt sich unmittelbar schlussfolgern:

> *Hat ein Festkörper ein teilbesetztes Energieband, so ist er ein elektronischer Leiter.*

Nach dieser Definition werden sämtliche Festkörper-Elemente in Leiter und Nichtleiter unterteilt. Da jeder Quantenzustand zwei Elektronen mit antiparallelem Spin aufnehmen kann, ist unmittelbar klar, dass Stoffe mit einer ungeraden Anzahl von Elektronen in einem Band Leiter sein müssen. Nichtleiter haben ein vollbesetztes Band mit zweimal (wegen Spin) Anzahl der Zustände (n, l, m) der Elektronen darin. Ein Elektron in diesem Band müsste in einem elektrischen Feld mindestens die Energiedifferenz bis zum nächst höher gelegenen (unbesetzten) Band aufnehmen. Dann wäre es frei beweglich und könnte zum Stromfluss beitragen. Dieser Vorgang nennt man Stoßionisation. Dazu sind hohe Feldstärken im Inneren des Festkörpers nötig. Die Erzeugung freier Elektronen auf diese Weise heißt Zener-Effekt und zählt zu den Durchbruch-Effekten.

Er ist die Grundlage der sogenannten Zener-Diode,[6] die als elektronisches Bauelement zur Erzeugung einer konstanten Spannung dient.

Die vorstehenden Überlegungen machen deutlich, dass am Leitungsmechanismus nur zwei Bänder des Festkörpers beteiligt sind. Wegen ihrer Bedeutung für die Stromleitung bekommen sie eigene Namen:

- Das letzte besetzte Band heißt Valenzband (VB). Dabei ist es egal, ob das Band teil- oder vollbesetzt ist.
- Das Band direkt über dem Valenzband heißt Leitungsband (LB). Ohne zusätzliche Anregung der Elektronen im Leitungsband (z. B. durch Zener-Effekt oder thermische Energiezufuhr) ist es immer leer.

Mit dem bis zu diesem Kapitel erarbeiteten Wissen lässt sich folgende wichtige Erkenntnis formulieren:

> *Das Vermögen eines Festkörpers mit voll besetztem Valenzband, Strom zu leiten, wird dadurch bestimmt, ob und in welchem Umfang sich (freie) Elektronen in seinem Leitungsband befinden.*

1.16 Bändermodell

Durch die Definition von Leitungs- und Valenzband, lässt sich die Darstellung der für die Stromleitung relevanten Eigenschaften deutlich vereinfachen. Aus der bisher verwendeten Energietopf-Darstellung brauchen nur, wie in Abb. 1.19 am Beispiel des **Na**-Kristalls gezeigt, das Leitungs- und Valenzband übernommen werden.
Für manche Überlegungen ist es nützlich, auch das Energieniveau mit in das Bänderdiagramm zu zeichnen, bei dem die Elektronen den Kristall verlassen. Das ist die Energie, die ein freies ruhendes Elektron (kinetische Energie = 0) außerhalb des Kristalls hat. Man nennt diese Energie Vakuumenergie, da dies die Energie ist, die man gewinnt, wenn man ein Elektron aus dem Unendlichen in den Kristall bringt. Aus Richtung des Kristalls gesehen, ist dies die Energie, die benötigt wird, um ein Kristallelektron gegen das Potentialfeld außerhalb des Kristalls, in das „Unendliche“ zu bringen. In dieser Betrachtungsweise wird die Energie häufig in der Form: Potential mal Elementarladung $\Phi_0(-e)$ angegeben. Das Potential Φ_0 nennt man

[6] meist zusammen mit Lawinendioden als Z-Dioden bezeichnet.

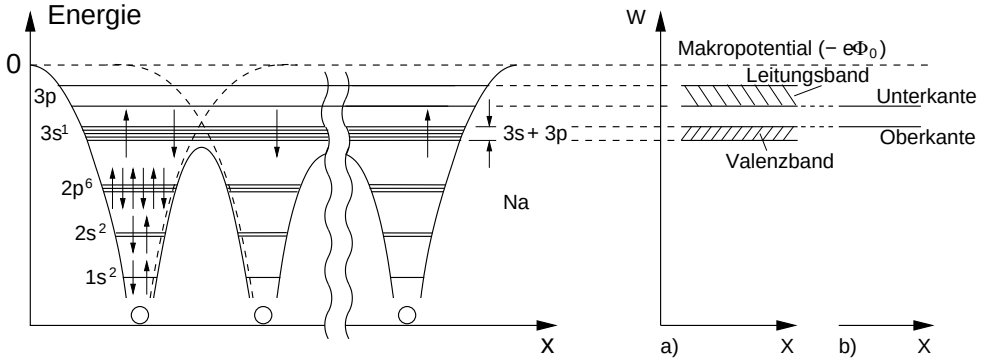

Abb. 1.19: Vereinfachung des Potentialtopfmodells (links) zum Bändermodell (rechts). Im gezeigten Beispiel des **Na**-Kristalls ist das Valenzband teilbesetzt und das Leitungsband leer. Das Bändermodell rechts (a) kann weiter vereinfacht werden zu (b), indem das Koordinatensystem weggelassen und nur die Ober- bzw. Unterkante des Valenz- bzw. Leitungsbandes gezeichnet werden.

Makropotential.

Wir besprechen im Folgenden einige wichtige Definitionen und Konventionen, die bei der Anwendung des Bänderdiagramms nützlich sind. Wir verwenden zur Erläuterung das in Abb. 1.20 gezeigte Bändermodell.

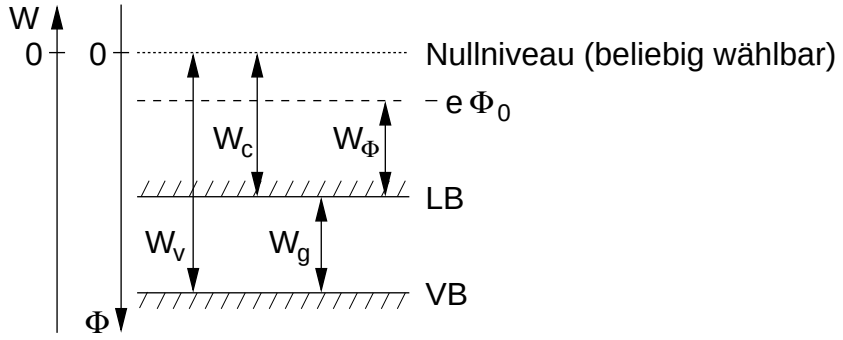

Abb. 1.20: Allgemeines Bänderdiagramm zur Erläuterung der darin ablesbaren Energien und Potentiale.

Zu der Leistungs- und Potentialskala auf der linken Seite gibt es folgendes anzumerken:

- Der Energienullpunkt ist für jedes abgeschlossene System frei wählbar,

da die Gesamtenergie des Systems bis auf eine frei wählbare Konstante bestimmt ist. (Es gibt keinen Energienullpunkt im Universum)

- Die Richtung der Energieskala ist so gewählt, dass eine Zunahme der Elektronenenergie nach oben aufgetragen wird.

- Die Potentialskala hat die entgegengesetzte Richtung zur Energieskala, da für den Zusammenhang zwischen Energie und Potential gilt (vgl. Übung 1)

$$W = -e\Phi \; . \tag{1.61}$$

- Auch der Nullpunkt der Potentialskala ist wie bei der Energieskala frei wählbar (vgl. Integrationskonstante in Übung 2). Zur Vereinfachung werden Energie- und Potentialnullpunkt zusammengelegt.

- In der Literatur wird häufig – anders als in Abb. 1.20 gezeigt – der Nullpunkt an die Oberkante des Valenzbandes gelegt.

- Da letztlich nur Energie- und Potentialdifferenzen die elektrischen Eigenschaften des Festkörpers bestimmen, werden wir, wann immer möglich, ohne Definition eines Nullpunktes arbeiten.

Die in Abb. 1.20 eingezeichneten Energien besitzen folgende Bedeutung:

- W_Φ wird im Bänderdiagramm als Austrittsarbeit bezeichnet. Sie ist die Differenz der Energie eines Elektrons an der LB-Unterkante zum Makropotential.

- W_C und W_V („C“ steht für „conducting“) sind die Energien des Leitungs- und Valenzbandes bezogen auf den Nullpunkt.

- W_g („g“ steht für „gap“) ist eine der wichtigsten Größen für die Halbleitertechnik. Sie gibt die Größe der Energielücke zwischen Leitungs- und Valenzband an. Sie ist in erster Näherung konstant und charakteristisch für das gewählte Material:

$$\begin{aligned} W_g(\mathbf{Ge}) &= 0,68\,\mathrm{eV} \\ W_g(\mathbf{Si}) &= 1,08\,\mathrm{eV} \\ W_g(\mathbf{GaAs}) &= 1,38\,\mathrm{eV} \; . \end{aligned} \tag{1.62}$$

Der Energiebereich W_g wird im Deutschen häufig als „verbotene Zone“ oder „Bandlücke“ bezeichnet, da sich in diesem Bereich wegen fehlender Energiewerte als Lösung der Schrödingergleichung keine Elektronen aufhalten können.
Im internationalen Sprachgebrauch wird dieser Bereich als „Bandgap“ bezeichnet und es steht E_g (Energy gap) für W_g.

1.17 Leiter, Halbleiter, Isolator

Anhand der vorangegangenen Überlegungen zur Elektronenleitung in einem Band, lassen sich mit Hilfe des Bändermodells in Abb. 1.21 die Unterschiede zwischen Leiter und Nichtleiter darstellen.

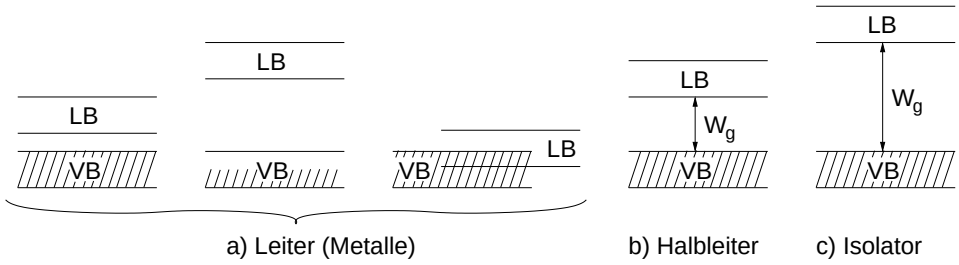

Abb. 1.21: Bändermodelle für Festkörper mit verschiedener Leitfähigkeit.

Ein Festkörper ist ein Leiter und wird als Metall bezeichnet, wenn er entweder

- einen sehr kleinen Bandabstand hat, so dass geringe (thermische) Energien ausreichen, damit Elektronen vom VB in das nicht besetzte LB gelangen können (Halbmetall), oder
- ein teilbesetztes Valenzband besitzt, oder
- ein überlappendes VB mit einem leeren LB, wodurch ein durchgehendes teilbesetztes Band entsteht.

In allen Fällen stehen freie Elektronen in einem teilbesetzten Band zur Stromleitung zur Verfügung.
Spezifische Leitfähigkeiten für in der Elektronik am häufigsten verwendeten Materialien sind in Tab. 1.2 angegeben. Darin ist zu sehen, dass Gold zwar das edelste aller Metalle ist, aber **Cu** und **Ag** besser leiten.
Ein Festkörper ist ein Isolator, wenn er ein vollbesetztes Valenzband und

	Ag	Cu	Au	Al	Sn	Pb
Spez. Widerstand [10^{-6} Ωcm]	1,61	1,7	2,2	2,74	11,0	21,0
Spez. Leitfähigkeit [10^5 $(\Omega\text{cm})^{-1}$]	6,21	5,88	4,55	3,65	0,96	0,48

Tabelle 1.2: Spezifische Leitfähigkeit und spezifischer Widerstand für einige Metalle bei 295 K.

einen großen Bandabstand besitzt (i. e. $> 2\,\text{eV}$). Auch die Zuführung von thermischer Energie reicht nicht aus, damit Elektronen vom VB in das LB gelangen; somit stehen keine freien Ladungsträger im LB zur Stromleitung zur Verfügung.
Der spezifische Widerstand von Isolatoren liegt in der Größenordnung $> 10\,\text{G}\Omega\,\text{cm}$ (10^{10}).
Das in integrierten Schaltungen zur Isolation der Verdrahtungs-Metallebenen eingesetzte $\mathbf{SiO_2}$ hat einen Wert von 10^{19} Ωcm. Wegen seines hohen Widerstandes und seiner damit verbundenen geringen Leckströme wird es zur Isolation der Gate-Elektrode bei Feldeffekt-Transistoren eingesetzt.

Ist der Bandabstand kleiner als bei einem Isolator, aber größer als bei einem (Halb-)Metall und ist das VB vollbesetzt, so spricht man von einem Halbleiter.
Durch die geringere Größe des Bandabstandes als bei einem Isolator gelangen durch thermische Anregung Elektronen in das LB und ermöglichen eine Stromleitung. Die thermische Energie bei Raumtemperatur reicht bei einem Halbleiter aus, so dass „einige wenige“ Elektronen in das LB gelangen. Der spezifische Widerstand bei Raumtemperatur (!) liegt im Bereich kΩcm ... MΩcm. Werte für **Si** liegen bei 0,3 MΩcm, für **GaAs** bei 10 GΩcm. Ohne thermische Anregung ($T = 0$) sind Halbleiter Isolatoren. Aufgrund der Abhängigkeit der freien Elektronen von der Zuführung thermischer Energie ist die Fähigkeit zur Stromleitung (Eigenleitung) stark temperaturabhängig.

Halbleiter entstehen aus kovalenten Bindungen, wodurch die Elemente in der vierten Gruppe des Periodensystems zu finden sind (**Si**, **Ge**). Halbleiter mit kovalenter Bindung sind auch durch Kombination von Elementen der 3. und 5. Gruppe möglich, wodurch z. B. das in der Halbleiterindustrie eingesetzte **GaAs** entsteht.

1.18 Zustände in Leitungs- und Valenzband

Mit dem Blick auf das Ziel, die physikalischen Grundlagen für die Halbleiter-Elektronik zu erarbeiten, verwenden wir im Folgenden auch den Begriff Halbleiter und meinen damit einen (Festkörper-)Kristall.
In den vorangegangenen Kapiteln haben wir gelernt, dass bei Halbleitern nur Elektronen im Leitungsband (LB) zur Stromleitung beitragen können. Das Leitungsband wurde definiert als ein bei fehlender Anregung leeres Energieband, dessen Energien zu einem oder mehreren (Entartung) Zuständen (Lösungen der Schrödingergleichung) der Elektronen gehören. Jeder Zustand ist aufgrund des Pauli-Prinzips genau mit zwei Elektronen mit antiparallelem Spin ($\downarrow\uparrow$) besetzbar. Durch das Prinzip der minimalen Energie besetzen Elektronen, die in das LB gelangen, immer zuerst die Zustände niedrigster Energie. Sie halten sich also an der Unterkante des LB auf. Die energetisch über ihnen liegenden freien Zustände ermöglichen ihnen in einem von außen an den Halbleiter angelegten Feld Energie aufzunehmen, wodurch sie auf diese freien Plätze gelangen. Durch die Kraftwirkung des elektrischen Feldes auf die Elektronen wandern sie durch den Kristall und tragen zur Stromleitung bei. Durch die physikalische Beschreibung dieses Transportvorgangs erhält man die bekannte Beziehung für die spezifische Leitfähigkeit aufgrund der Elektronen im Leitungsband

$$\sigma_{LB} = e\mu n_L \tag{1.63}$$

mit e = Elementarladung, μ = Beweglichkeit der Elektronen im LB und n_L = Konzentration der Elektronen im LB.

Anschaulich ist es einsichtig, dass eine solche Beziehung von der Beweglichkeit und Konzentration der Ladungsträger abhängt. Doch von welchen Eigenschaften des Halbleiters werden sie bestimmt? Das ist die Frage, mit deren Beantwortung wir uns im Folgenden beschäftigen werden.

1.19 Modell freier Elektronen

Wir vergessen zunächst wieder, dass es das aus den vorangegangenen qualitativen Überlegungen entstandene Leitungs- und Valenzband gibt. Wir wollen die Struktur der Energiebänder in diesem Bereich genauer beschreiben, um quantitative Aussagen über das Verhalten der Elektronen machen zu können. Dazu müsste die Schrödingergleichung zumindest für

mehrere Atome im Kristallgitter gelöst werden. Das ist analytisch nicht, und numerisch nur mit großem Aufwand möglich.
Eine Näherung, die eine analytische Berechnung erlaubt, ist das Kronig-Penny-Modell, das ein periodisch moduliertes Potential (z. B. rechteck- oder cos-förmig) einer unendlich ausgedehnten Atomkette berücksichtigt.

Die Berechnung lässt sich mit den im Studium vermittelten mathematischen Kenntnissen leicht durchführen. Sie erfordert jedoch in der kompletten Herleitung einen gehörigen Aufwand, liefert aber verglichen mit dem viel einfacheren Modell für freie Elektronen (unter Einbeziehung der später hergeleiteten Bragg-Bedingung) keine für diese Vorlesung benötigten zusätzlichen Informationen. Im übrigen liefern beide Modelle eher qualitative Ergebnisse und sind nicht in der Lage, die tatsächliche, relativ komplizierte Bandstruktur von Halbleitern quantitativ zu beschreiben.

Das Modell freier Elektronen geht von Elektronen aus, die nur schwach oder gar nicht an die Atome im Kristallgitter gebunden sind. Elektronen, die dieser Bedingung genügen, befinden sich also in etwa in dem in Abb. 1.22 schraffierten Bereich oberhalb des periodischen Potentialverlaufs.

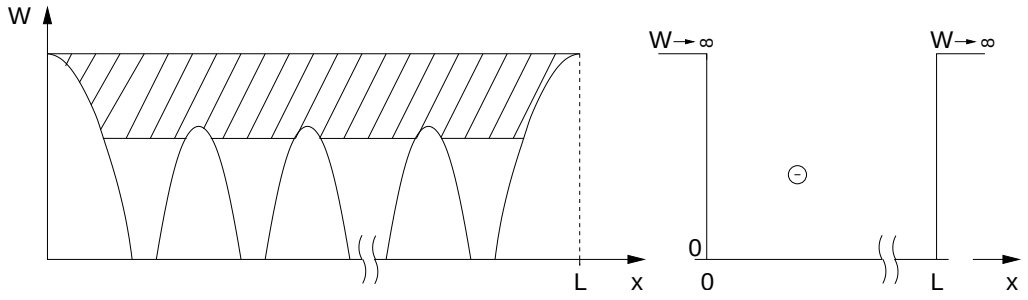

Abb. 1.22: Links: Energiebereich für schwach gebundene und freie Elektronen. Der Kristall soll einer Kantenlänge L haben. Rechts: vereinfachtes Kastenmodell mit einem Elektron innerhalb einer unendlich großen Energiebarriere.

Wir gehen in dem Modell für freie Elektronen von folgenden vereinfachenden Annahmen aus:

1. Der Atomkern und innere Elektronen werden nur ortsunabhängig (keine Modulation) durch ein konstantes Potential berücksichtigt.

2. Da wir frei in der Wahl des Potential-Nullpunktes sind, legen wir dieses konstante Potential aus Punkt 1 als Nullpunkt fest.

3. Die Potentialbarriere an den Kristallgrenzen soll unendlich hoch sein. Für das Potentialfeld, in dem sich das Elektron bewegt, gilt dann

$$W_{pot}(x) = \begin{cases} 0 & \text{für } 0 \leq x \leq L, \\ \infty & \text{sonst.} \end{cases} \tag{1.64}$$

4. Wir betrachten nur ein freies Elektron und unterstellen, dass die dafür erhaltenen Ergebnisse auf eine beliebige Anzahl Elektronen übertragbar ist (keine Wechselwirkung).

5. Wir suchen stationäre Lösungen, also Lösungen, die nicht von der Zeit abhängen.

6. Die Lösungen sollen für Halbleiterkristalle mit beliebiger Größe gelten. Als Variable für die Größe nehmen wir einen Würfel mit der Kantenlänge L an, wobei L beliebig wählbar ist.

Wir machen ein kleines Gedankenexperiment, um eine wichtige Randbedingung für die Lösung der Schrödingergleichung zu erhalten:

a) Wenn die Lösung für beliebige Kantenlänge L gilt (vgl. 6.), dann muss sie auch für $L \to \infty$ gelten.

b) Für $L \to \infty$ lässt sich der Kristall zu einem Kreis mit unendlich großem Radius biegen (Gedankenmodell), dessen Anfang und Ende sich berühren.

c) In dieser Vorstellung läuft das Elektron im Kreis (eindimensional) bzw. in der Kugel (dreidimensional), da Kristall-Anfang und -Ende aufeinander gebogen sind (vgl. Abb. 1.23). Die Potentialschwelle entfällt in dieser Vorstellung, da das Elektron in dem Kasten mit $L \to \infty$ unendlich lange läuft, um an die Kristallgrenzen zu gelangen.

d) Abb. 1.23 zeigt das „Biegen“ in einer Dimension zu einem Kreis. Dadurch fallen Anfang ($x = 0$) und Ende ($x = L$) in einem Punkt zusammen. Da das Elektron (die Elektron-Welle ψ) unendlich lange läuft, bedeutet dies auf der Kreisbahn, dass nur Lösungen existieren, die eine

stetige und periodische Fortsetzung mit der Periode L besitzen. Für Periodizität im Eindimensionalen gilt

$$\psi(x) = \psi(x + L) \; . \tag{1.65}$$

Da alle drei Dimensionen gleichberechtigt sind, muss für den Würfel

$$\psi(x, y, z) = \psi(x+L,\; y+L,\; z+L) \tag{1.66}$$

gelten. Wir nennen diese Randbedingung: „periodische Randbedingung“.

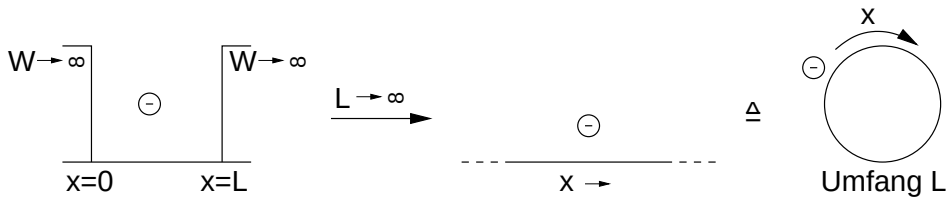

Abb. 1.23: Biegen (gedanklich) einer Kristalldimension (hier x) zu einem Kreis, um einen unendlich ausgedehnten Kristall zu beschreiben. Beachten: Da es sich nicht um die tatsächliche physikalische Anordnung, sondern nur um ein Gedankenexperiment bezüglich des Fortschreitens der Wellenfunktion handelt, treten keine Radialkräfte und damit verbundene Energien auf.

1.20 Lösungen der Schrödingergleichung für freie Elektronen

Wir suchen stationäre Lösungen der Wellenfunktion und können daher die zeitunabhängige Schrödingergleichung Gl. (1.20) verwenden. Mit dem statischen Potential $W_{pot} = 0$ in unserem Modell für freie Elektronen lautet sie

$$-\frac{\hbar^2}{2m_e}\Delta\psi(x, y, z) = W\psi(x, y, z) \; . \tag{1.67}$$

Da das Elektron sich in keinem Potentialfeld bewegt, stellt W die rein kinetische Energie des freien Elektrons dar.

Die Schrödingergleichung (1.67) besitzt die gleiche Form wie die in der Hochfrequenztechnik bekannte homogene Helmholtz-Gleichung, die die Wellenausbreitung im (quell-)freien Raum beschreibt. Sie hat als Lösungen ebene Wellen der Gestalt

$$\psi = a e^{j\vec{k}\vec{r}} \tag{1.68}$$

$$\text{mit } \vec{k} = k_x\vec{e}_x + k_y\vec{e}_y + k_z\vec{e}_z, \quad (k_x, k_y, k_z = const.) \tag{1.69}$$

$$\text{und } \vec{r} = x\,\vec{e}_x + y\,\vec{e}_y + z\,\vec{e}_z\ . \tag{1.70}$$

Gl. (1.68) in die Schrödingergleichung (1.67) eingesetzt ergibt unmittelbar die kinetische Energie des Elektrons

$$W = \frac{\hbar^2 k^2}{2m_e}, \quad \text{mit } k^2 = k_x^2 + k_y^2 + k_z^2\ . \tag{1.71}$$

Sie ist unabhängig von der Kantenlänge L unseres Kristalls.
Die Amplitude a von ψ ergibt sich aus der Normierungsbedingung (1.55) zu $a = L^{-\frac{3}{2}}$, wodurch die Wellenfunktion aus (1.68) lautet

$$\psi = \pm\frac{1}{L^{\frac{3}{2}}} e^{j\vec{k}\vec{r}}\ . \tag{1.72}$$

Rechnen Sie zur Übung doch einmal Gl. (1.71) und (1.72) nach (leicht).
Die Aufenthaltswahrscheinlichkeit des Elektrons an einem beliebigen Ort $\vec{r}$ im Kristall ist

$$|\psi(\vec{r})|^2 = \left|\frac{1}{L^{\frac{3}{2}}} e^{j\vec{k}\vec{r}}\right|^2 = \frac{1}{L^3}\ . \tag{1.73}$$

Das freie Elektron befindet sich also überall im Kristall mit der gleichen Wahrscheinlichkeit. Man sagt, das Elektron ist über den Kristall „ausgeschmiert“.
Der Wellenvektor $\vec{k}$ der Wellenfunktion muss unserer Randbedingung für Periodizität genügen. Dies ist erfüllt für Komponenten des Wellenvektors

$$k_x = \pm\frac{n_x 2\pi}{L}, \quad k_y = \pm\frac{n_y 2\pi}{L}, \quad k_z = \pm\frac{n_z 2\pi}{L} \tag{1.74}$$

mit den natürlichen Zahlen

$$n_x, n_y, n_z = 0, \pm1, \pm2, \pm3 \ldots\ . \tag{1.75}$$

Damit existieren unendlich viele Lösungen für den Wellenvektor $\vec{k}$, die sich aus der Wahl und Kombination der n_x, n_y, n_z ergeben. Da über die

$\vec{k}$-Vektoren Richtung und Wellenlänge der Wellenfunktionen unterschieden werden, stellt jeder $\vec{k}$-Vektor einen möglichen Zustand des Elektrons dar. Wie zuvor für das Elektron des Wasserstoffatoms haben wir auch für das freie Elektron einzelne Zustände der Wellenfunktion gefunden. Hier sind diese über Gl. (1.71) mit der Energie des Elektrons verknüpft. Abb. 1.24 zeigt die diskreten Energien des Elektrons in Abhängigkeit des Betrags seines Wellenvektors. Wir nennen eine solche Darstellung in Abhängigkeit des Betrags eines Wellenvektors auch als Darstellung im k-Raum (auch Zustands- oder Phasenraum). Wir stellen uns dabei vor, dass die Komponenten k_x, k_y, k_z des Wellenvektors ihren eigenen Raum, den k-Raum, aufspannen.

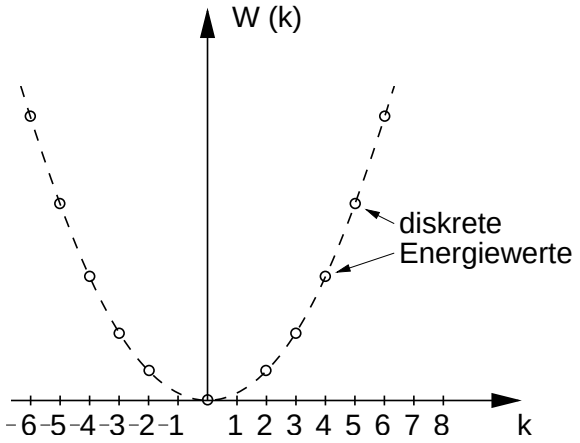

Abb. 1.24: Diskrete Energien (entartet) für die verschiedenen Zustände des freien Elektrons in einem Kristall.

Durch Einsetzen von Gl. (1.74) in (1.71) wird unmittelbar klar, dass die Zustände des Elektrons bezüglich der Energie entartet sind:

$$W(k) = W_{n_x,n_y,n_z} = \frac{\hbar^2}{2m_e}\left(\frac{2\pi}{L}\right)^2 (n_x^2 + n_y^2 + n_z^2)\ . \tag{1.76}$$

Wir können in Analogie an die Lösungen des Wasserstoffatoms die n_x, n_y, n_z als Quantenzahlen interpretieren.
Gleichbedeutend mit den Quantenzahlen kann jedoch auch der Wellenvektor $\vec{k}$ verwendet werden, wovon wir im Folgenden Gebrauch machen werden.

Wir merken uns folgende Eigenschaften der Wellenfunktion des freien Elektrons:

- Das Elektron nimmt nur diskrete Energien $W(k(n_x, n_y, n_z))$ an.

- Die Energien werden durch den Betrag k des Wellenvektors bestimmt.
- Die Energie steigt quadratisch mit k.
- Zu jedem k und damit auch zu jeder Energie existieren mehrere Zustände (Sätze von Quantenzahlen).
- Jedes $\vec{k}$ beschreibt einen Zustand (eindeutig).
- Jeder Zustand kann wegen des Pauli-Prinzips mit maximal zwei Elektronen mit antiparallelem Spin besetzt werden.

Für einen Wellenvektor

$$k^2(n_x, n_y, n_z) = \left(\frac{2\pi}{L}\right)^2 (n_x^2 + n_y^2 + n_z^2) = \left(\frac{2\pi}{L}\right)^2 N \qquad (1.77)$$

ergeben sich für $N = 1$ sechs Zustände, die durch 12 Elektronen besetzt werden können. Für $N = 2$ ergeben sich 12 Zustände, für $N = 3$ bzw. 4 aber nur 8 bzw. 6 Zustände. Überprüfen Sie das zur Übung doch einmal.
Für große N wird die Berechnung von Hand schnell aufwendig. Wollen wir 10^{23} Elektronen unterbringen, benötigen wir $\frac{1}{2} \cdot 10^{23}$ Zustände, die sich auf Kombinationen der n_x, n_y, n_z verteilen.
Wir führen im nächsten Kapitel eine einfache Systematik zur Abschätzung der Anzahl der Zustände ein.
Aus der Vielzahl der Zustände wird deutlich, dass die diskreten Energiewerte so häufig und dicht aufeinander folgen, dass man die Parabel im k-Raum als kontinuierlichen Verlauf ansehen kann.

Zum Abschluss dieses Kapitels berechnen wir noch aus der Lösung der Wellenfunktion einige Größen der Teilchennatur des Elektrons:

Die Energie des freien Elektrons ist rein kinetisch, daher gilt mit Gl. (1.71)

$$W = \frac{\hbar^2 k^2}{2m_e} = \frac{1}{2} m_e v^2 = \frac{p^2}{2m_e}, \qquad (1.77\text{ b})$$

woraus durch Vergleich

$$\vec{p} = \hbar\vec{k} = \hbar\frac{2\pi}{\lambda}\vec{e}_k \qquad (1.77\text{ c})$$

und

$$\vec{v} = \frac{\vec{p}}{m_e} = \frac{\hbar}{m}\vec{k} \qquad (1.77\text{ d})$$

folgt.

1.21 Zustandsdichte freier Kristallelektronen

Wir haben im letzten Kapitel den Begriff des k-Raumes eingeführt. Wir stellen uns dabei vor, dass die drei Komponenten k_x, k_y und k_z der möglichen Wellenzahlen $\vec{k}(n_x, n_y, n_z)$ einen dreidimensionalen Raum aufspannen.
Wir benutzen diese Vorstellung, um die Vielzahl der möglichen Wellenzahlvektoren räumlich darzustellen.
Da sich sämtliche Wellenzahlen in allen Raumrichtungen aus Vielfachen von $\frac{2\pi}{L}$ zusammensetzen, definieren wir eine Einheitszelle aus den drei Vektoren

$$\begin{aligned} \vec{k_x} &= \left(\frac{2\pi}{L}, 0, 0\right)^T \\ \vec{k_y} &= \left(0, \frac{2\pi}{L}, 0\right)^T \\ \vec{k_z} &= \left(0, 0, \frac{2\pi}{L}\right)^T \end{aligned} \tag{1.78}$$

mit dem Volumen

$$V_{EZ} = \left(\frac{2\pi}{L}\right)^3 . \tag{1.79}$$

Diese Einheitszelle kann in allen drei Raumrichtungen durch Addition eines Translationsvektors

$$\vec{k} = \frac{2\pi}{L}(n_x, n_y, n_z)^T \tag{1.80}$$

verschoben werden.
Jeder Endpunkt im dreidimensionalen Raum stellt dann einen möglichen Zustand dar. Abb. 1.25 zeigt die Einheitszelle und den sich daraus ergebenden Gitteraufbau. Jeder Gitterpunkt ist durch ein Kreuz markiert.
Wellenvektoren mit gleichem Betrag haben in dieser Darstellung den gleichen Abstand zum Nullpunkt, liegen also auf einer Kugelschale mit dem Mittelpunkt im Ursprung.
Da über Gl. (1.71) die Energie des Elektrons für einen bestimmten Zustand nur von k abhängt, besitzen sämtliche Zustände, deren k auf der gleichen Kugelfläche liegt, auch die gleiche Energie.

Wir fassen zusammen:

- Kugelschalen im k-Raum sind Flächen mit konstanter Energie.
- Schalen mit kleinerem Radius k gehören zu geringeren Energien.

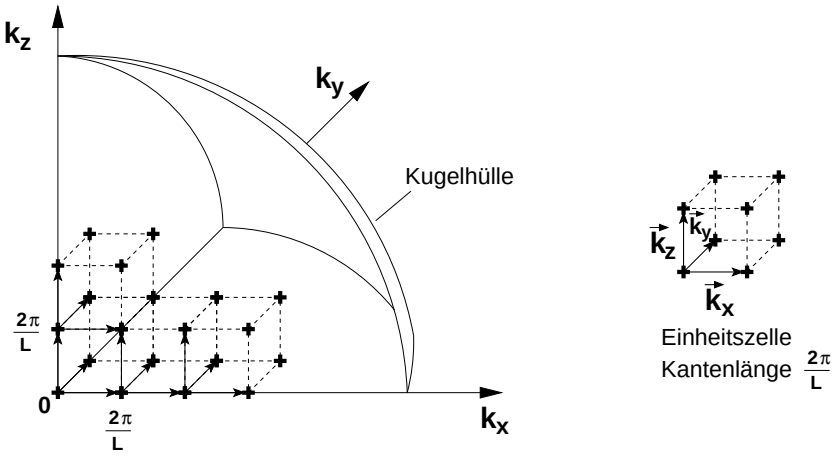

Abb. 1.25: Darstellung der Wellenvektoren sämtlicher Zustände eines freien Elektrons in einem Kristall der Kantenlänge L. Die durch Kreuze markierten Zustände formen eine dreidimensionale Gitterstruktur, die durch Verschiebung der Einheitszelle um $\frac{2\pi}{L}$ in allen Raumrichtungen entsteht.

- Eine Kugel mit dem Radius k beinhaltet demnach nur Zustände mit Energien $W \leq \dfrac{\hbar^2 k^2}{2m_e}$.

Die Anzahl N_Z der Zustände in einer Kugel mit dem Radius k entspricht der Anzahl der, in der Kugel enthaltenen Elementarzellen[7]. Mit dem Volumen der Kugel

$$V_k(k) = \frac{4\pi}{3}k^3 \tag{1.81}$$

ergeben sich mit dem Volumen der Einheitszelle nach Gl. (1.79)

$$N_Z = \frac{V_k(k)}{V_{EZ}} = \frac{\frac{4\pi}{3}k^3}{(\frac{2\pi}{L})^3} = \frac{L^3k^3}{6\pi^2} = \frac{Vk^3}{6\pi^2} \tag{1.82}$$

durch Elektronen besetzbare Zustände. V ist darin das Volumen des Kristalls mit der Kantenlänge L.
Jeder Zustand kann mit zwei Elektronen besetzt werden, wodurch

$$N_{EZ} = 2N_Z = \frac{Vk^3}{3\pi^2} \tag{1.83}$$

[7]Die Zuordnung Anzahl Elementarzellen zu Zuständen fällt einfach, wenn man sich die Einheitszellen um eine halbe Translationsvektorlänge verschoben vorstellt. Dann liegen die k-Werte jeweils in der Mitte einer Zelle.

Elektronen auf den Zuständen in einer Kugel mit dem Radius k untergebracht werden können. Wir bezeichnen N_{EZ} im Folgenden als Anzahl der Elektronenzustände. Die Energie der Elektronen steigt entsprechend Gl. (1.71) von den inneren Kugelschalen mit wachsenden k-Werten kontinuierlich an und erreicht auf der Kugelhülle ihr Maximum.
Da nach Gl. (1.71) jedem k-Wert eine Energie zugeordnet ist, können wir die Anzahl der Elektronen auch über die maximale Energie $W(k)$ auf der Kugelhülle ausdrücken:

$$N_{EZ} = \frac{V}{3\pi^2}\left(\frac{2m_e W}{\hbar^2}\right)^{\frac{3}{2}} \quad . \tag{1.84}$$

Darin hängt die Anzahl der Elektronenzustände von dem Volumen V des betrachteten Kristalls ab. Wir möchten für die weitere Berechnung lieber mit Größen arbeiten, die unabhängig von den Abmessungen eines Kristalls sind. Wie immer in einem solchen Fall bildet man eine spezifische Größe, indem man das Ergebnis auf die Abmessung bezieht. Da in unserem Fall die Abmessung ein Volumen ist, stellt die spezifische Größe eine Dichte dar. Wir erhalten für die spezifische Anzahl der Elektronenzustände die Dichte

$$\frac{N_{EZ}}{V} = \frac{1}{3\pi^2}\left(\frac{2m_e W}{\hbar^2}\right)^{\frac{3}{2}} \quad . \tag{1.85}$$

Bei Änderung der Energie W der Kugelhülle ändert sich die Dichte der Elektronenzustände, da sich über $k(W)$ das Volumen V_k der Kugel im k-Raum ändert, die die möglichen Zustände einschließt.
Durch Ableitung von Gl. (1.85) ergibt sich die Änderung

$$D(W) := \frac{d}{d\,W}\frac{N_{EZ}}{V} = \frac{1}{3\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{\frac{3}{2}} \cdot \frac{3}{2} W^{\frac{1}{2}} = \frac{(2m_e)^{\frac{3}{2}}}{2\pi^2\hbar^3} W^{\frac{1}{2}} = \frac{8\pi\sqrt{2}}{h^3} m_e^{\frac{3}{2}} \sqrt{W} \; . \tag{1.86}$$

$D(W)$ gibt die Änderung der spezifischen Anzahl der eingeschlossenen Elektronenzustände an, für eine Änderung um dW der Energie ($\hat{=}$ Radius) der Kugelhülle bei einem Wert W. $D(W)$ ist eine wichtige Größe. Wir bezeichnen sie als Zustandsdichte.
Da Gl. (1.86) für das Modell der freien Kristallelektronen hergeleitet wurde, beschreibt sie die Zustandsdichte freier Kristallelektronen (bzw. des Elektronengases).

Das Arbeiten mit Zustandsdichten wird uns durch das ganze Skript hindurch begleiten. Daher ist ein Verständnis der Aussage einer Zustandsdichte wichtig. Dabei ist zu merken, dass es sich um die Anzahl der Elektronenzustände in einem Volumen (Dichte) in einem Energieintervall dW handelt, also genau genommen um die (Energie-) Dichte einer (Volumen-) Dichte.
Dies wird klar, wenn wir die Anzahl der Elektronenzustände dN_{EZ} in einem Energieintervall dW ausrechnen:

$$\begin{aligned}\text{Anzahl der Elektronen-}\\ \text{zustände pro Volumen } V\end{aligned} \quad = \frac{dN_{EZ}}{V} = D(W) \cdot dW \tag{1.87}$$

$$\text{Anzahl der Elektronenzustände} \quad = dN_{EZ} = V \cdot D(W) \cdot dW \tag{1.88}$$

1.22 Fermienergie, Fermikugel

Wir nehmen an, wir haben N_e freie Elektronen in unserem Kristall. Unter der Annahme, dass keine Wechselwirkung zwischen den Elektronen stattfindet, stehen alle im vorangegangenen Kapitel berechneten Elektronenzustände zur Besetzung durch die N_e Elektronen zur Verfügung. Nach dem Pauli-Prinzip werden beginnend mit der niedrigsten Energie alle Elektronenzustände nacheinander aufgefüllt, bis das letzte Elektron untergebracht ist. Dann wurden genau $N_{EZ} = N_e$ Elektronenzustände besetzt. Den Wellenvektor mit dem Radius der Kugelhülle, die die N_{EZ} Elektronenzustände umschließt, nennen wir Fermiwellenvektor k_F.
Die zu k_F gehörende Kugel nennen wir Fermikugel und die Energie der letzten Zustände auf der Hülle nennen wir Fermienergie W_F:

$$W_F = \frac{\hbar^2}{2m_e} k_F^2 \underbrace{=}_{\text{Gl. (1.83)}} \frac{\hbar^2}{2m_e} \left(3\pi^2 \frac{N_e}{V}\right)^{\frac{2}{3}} . \tag{1.89}$$

Die Fermienergie ist also eine Funktion der Volumendichte $\frac{N_e}{V}$ der freien Elektronen in dem Kristall.
Ist die Anzahl der freien Elektronen pro Volumen eines Kristalls bekannt, so kann mit Gl. (1.89) dessen Fermienergie abgeschätzt werden.

Beispiel: Für das Beispiel des **Na**-Kristalls steht ein freies $3s^1$ Elektron pro Atom zur Verfügung. **Na** kristallisiert in einer kubisch-raumzentrierten Struktur (*bcc* = body-centered-cubic). Abb. 1.26 zeigt die Anordnung der Atome in einer Elementarzelle, das *bcc*-Raumgitter sowie zwei weitere Kristallgitter, das einfache kubische Gitter (*sc* = simple cubic) und das kubisch flächenzentrierte Gitter (*fcc* = face-centred-cubic).
Jede Elementarzelle teilt sich die Atome auf ihren Ecken mit den benachbarten Elementarzellen.
Im Fall der *bcc*-Struktur entfallen auf eine Elementarzelle zwei Atome, also auch zwei freie Elektronen.
Die Gitterkonstante (Abstand der auf den Ecken sitzenden Atome) beträgt bei **Na** 4,225 Å=0,4225 nm.
Daher beträgt die Elektronendichte

$$\frac{N_e}{V} = \frac{2}{(0,4225 \cdot 10^{-9})^3\,\mathrm{m}^3} = \frac{2 \cdot 10^{27}}{0,075\,\mathrm{m}^3} = 2,65 \cdot 10^{28} \frac{1}{\mathrm{m}^3}$$

mit der Elektronenmasse $m_e = 9,1 \cdot 10^{-31}\,\mathrm{kg}$ und dem Wirkungsquantum $h = 6,63 \cdot 10^{-34}\,\mathrm{Js} = \hbar \cdot 2\pi$ sich aus Gl. (1.89)

$$\begin{aligned} W_F &= \frac{\hbar^2}{2m_e} \left(3\pi^2 \frac{N_e}{V}\right)^{\frac{2}{3}} \\ &= \frac{(\frac{1}{2\pi} 6,63 \cdot 10^{-34}\,\mathrm{Js})^2}{2 \cdot 9,1 \cdot 10^{-31}\,\mathrm{kg}} \left(3\pi^2 \cdot 2,65 \cdot 10^{28} \frac{1}{\mathrm{m}^3}\right)^{\frac{2}{3}} \\ &= 6,1 \cdot 10^{-39} \frac{\mathrm{J}^2\mathrm{s}^2}{\mathrm{kg}} \cdot 8,5 \cdot 10^{19} \frac{1}{\mathrm{m}^2} \\ &= 5,2 \cdot 10^{-19}\,\mathrm{J} = 3,2\,\mathrm{eV} \qquad (1\,\mathrm{eV} = 1,602 \cdot 10^{-19}\,\mathrm{J}) \end{aligned}$$

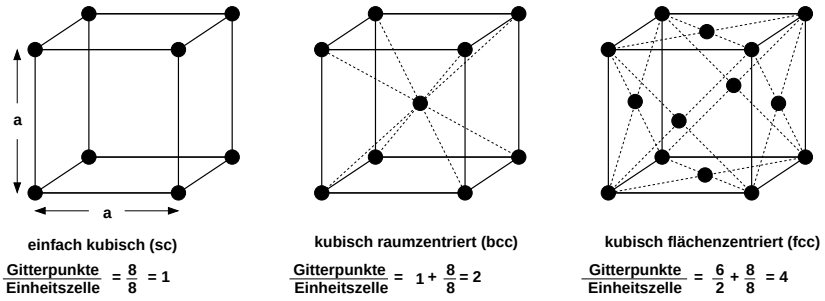

Abb. 1.26: Elementarzellen (Einheitszellen) für die drei einfachsten Kristallgitterstrukturen. a ist die Gitterkonstante

1.23 Streuung am Kristallgitter

Bisher haben wir angenommen, dass sich die nicht oder nur schwach an die Atome im Kristallgitter gebundenen Elektronen frei bewegen können. Als Ergebnis bekamen wir Wellenfunktionen, die zeigen, dass in diesem Fall die Elektronen über den Kristall ausgeschmiert sind. Es ergab sich ein quasi-kontinuierlicher parabelförmiger Verlauf der kinetischen Energien der Elektronen in Abhängigkeit des Betrags des Wellenvektors.
Die Struktur der realen Verläufe von Valenz- und Leitungsband eines Halbleiters werden wesentlich durch die Wechselwirkung der Elektronen mit den Atomen im Kristallgitter bestimmt. Um den qualitativen Einfluss des Gitters zu verstehen, führen wir im Folgenden einige sehr einfache Überlegungen durch.

Wir definieren zunächst einige nützliche Größen des Kristallgitters.

1.24 Definition von Richtungen und Flächen in Kristallgittern

In Abb. 1.26 haben wir drei Elementarzellen kennengelernt, aus denen sich Kristallgitter aufbauen lassen. Sie gehören zu dem kubischen Kristallsystem, das gleiche Kantenlängen und Winkel von 90° aufweist. Das kubische Kristallsystem ist eines von 14 möglichen Kristallsystemen (Bravais-Gitter), aus denen sich alle Kristalle aufbauen lassen. Der Aufbau geschieht durch räumlich identische Fortsetzung der Elementarzelle. Die Fortsetzung erfolgt über Translationsvektoren $(\vec{a}, \vec{b}, \vec{c})$, die mit den drei Achsen der Elementarzelle identisch sind (vgl. Abb. 1.27).
Über einen Vektor

$$\vec{r_n} = (n_1\vec{a} + n_2\vec{b} + n_3\vec{c})$$

läßt sich jeder Gitterpunkt erreichen.

Eine Richtung wird durch Angabe der $[n_1,\ n_2,\ n_3]$ beschrieben. So ist z. B. $[1, 0, 0]$ die mit $\vec{a}$ in Abb. 1.27 zusammenfallende Richtung. $[2, 1, 1]$ ist die Richtung des Vektors $\vec{u}$ in Abb. 1.27.
Eine Fläche wird über Achsenabschnitte mit Hilfe der Miller-Indizierung

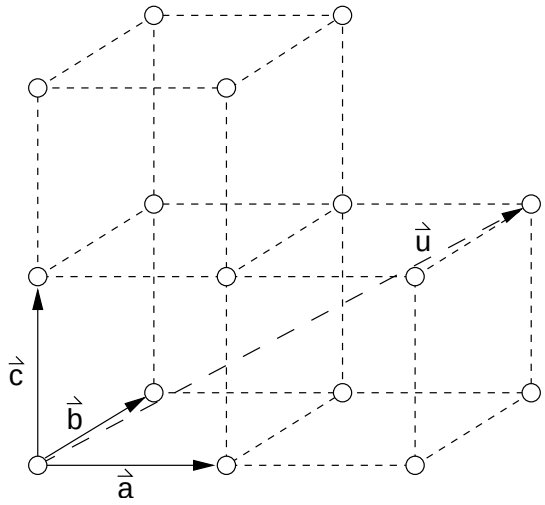

Abb. 1.27: Translationsvektoren einer Elementarzelle. Eingezeichnet ist auch ein Beispielvektor $\vec{u}$ für die Richtung [2,1,1].

angegeben. Abb. 1.28 zeigt ein Beispiel.

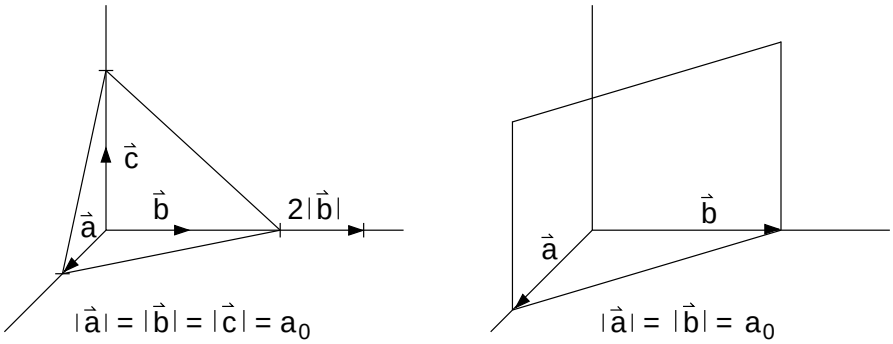

Abb. 1.28: Beispiele für die Miller-Indizierung von Flächen. Links: Schnittpunkte bei den Achsenabschnitten $a_0, 2a_0, 2a_0$ mit den Kehrwerten $\frac{1}{a_0}, \frac{1}{2a_0}, \frac{1}{2a_0}$, woraus sich die Miller-Indizes (2,1,1) ergeben. Rechts: Achsenabschnitte bei a_0, a_0, ∞ mit Miller-Indizes (1,1,0).

Die Vorgehensweise bei der Miller-Indizierung:

1. Bestimmung der kleinsten ganzzahligen Achsenabschnitte in Einheiten von

$$|\vec{a}|, |\vec{b}|, |\vec{c}| \Rightarrow u, v, w$$

2. Verwendung der Kehrwerte

$$\frac{1}{u}, \frac{1}{v}, \frac{1}{w}$$

3. Multiplikation der Kehrwerte mit q als dem kleinsten gemeinsamen Vielfachen von u, v, w, falls eine der Zahlen $\frac{1}{u}, \frac{1}{v}, \frac{1}{w}$ ein Bruch ist, ergibt die Miller-Indizes (h,k,l).

$$\frac{q}{u}, \frac{q}{v}, \frac{q}{w} \Rightarrow (h, k, l) \; .$$

Zwei planparallele Flächen haben also die gleichen Miller-Indizes. Man verwendet geschweifte Klammern h, k, l, um die Gesamtheit aller gleichwertigen Ebenen zu bezeichnen.
Den Abstand zwischen zwei planparallelen Flächen bezeichen wir mit $d_{h,k,l}$, wobei auch Teile davon erlaubt sind. So bezeichnet z. B. d_{110} in Abb. 1.29

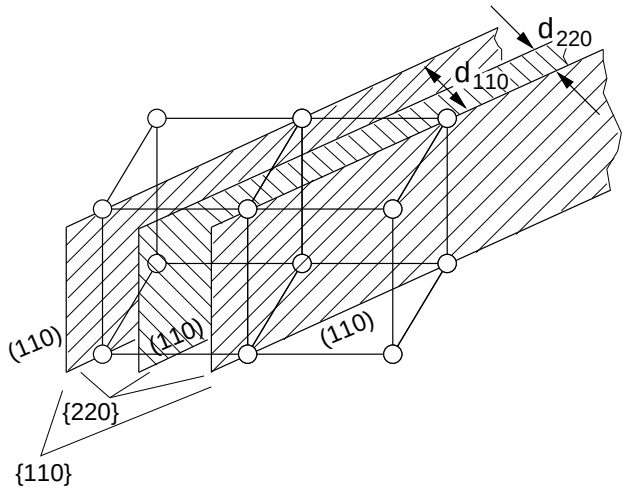

Abb. 1.29: Beispiel zur Definition eines Abstandes zwischen benachbarten planparallelen Flächen.

den Abstand zweier benachbarter (110)-Ebenen und d_{220} den halben Abstand zwischen den (110)-Ebenen.

1.25 Bragg-Reflektion

Die Wellenfunktion der freien Elektronen sind ebene Wellen, die durch Gl. (1.68) beschrieben werden:

$$\psi = a e^{j\vec{k}\vec{r}} \; . \tag{1.68}$$

Sie unterscheiden sich durch verschiedene Wellenvektoren $\vec{k} = k_x\vec{e}_x + k_y\vec{e}_y + k_z\vec{e}_z$, die aufgrund der Abhängigkeit der k_x, k_y, k_z von n_x, n_y, n_z (vgl. Gl. (1.74)) in (nahezu) beliebige Raumrichtungen zeigen und mit Ausnahme der entarteten Zustände auch unterschiedliche Beträge besitzen.
Wir betrachten eine solche ebene Welle, die nach Abb. 1.30 unter einem beliebigen Winkel ϕ auf eine Kristallebenenschar mit einer Orientierung (h, k, l) fällt.

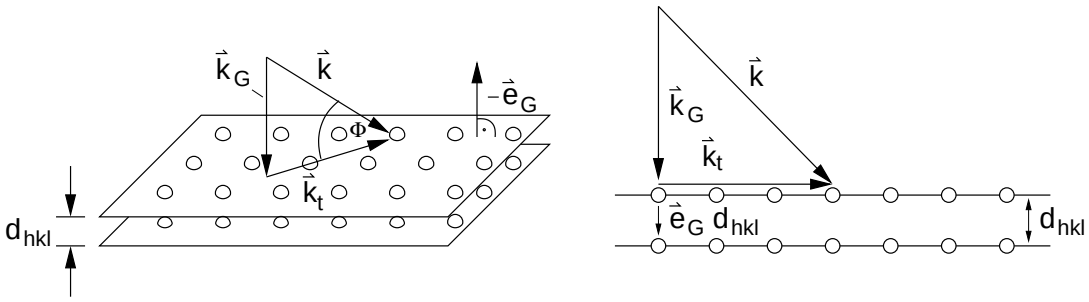

Abb. 1.30: Einfall einer ebenen Welle auf eine Kristallebenenschar (h,k,l). Links: Zerlegung des Wellenvektors in eine Normal- und eine Tangentialkomponente. Rechts: Zweidimensionale Darstellung der Zerlegung mit Blick auf die durch $\vec{k}_G$, $\vec{k}$ und $\vec{k}_t$ aufgespannte Fläche.

Der Wellenvektor lässt sich immer in zwei Komponenten aufteilen: Die Komponente $\vec{k}_G$ steht senkrecht auf der Kristallebene (h, k, l) in Richtung der Flächennormalen $\vec{e}_G$. Die Tangentialkomponente $\vec{k}_t$ liegt in der Kristallebene. Abb. 1.30 rechts zeigt eine zweidimensionale Darstellung der Zerlegung mit Blick auf die, durch die Vektoren aufgespannte Fläche.
Für die Aufteilung gilt

$$\vec{k} = \vec{k}_G + \vec{k}_t \; . \tag{1.90}$$

Der Wellenvektor in Richtung der Flächennormalen

$$\vec{k} \cdot \vec{e}_G = k_G = |\vec{k}| \sin\phi = k \sin\phi \tag{1.91}$$

beschreibt eine ebene Welle, deren Phasenflächen parallel zu den Flächen (h, k, l) sind. Gemäß dem Huygens-Fresnelschen Prinzip entsteht an jeder der Flächen eine Reflektion[8], die in entgegengesetzter Richtung zu $\vec{k}_G$ läuft.

[8]Genauer müsste man sagen, dass von der, durch die einfallende Welle stiumlierten Elektronenverteilung der Gitteratome Elementarwellen (Kugelwellen) ausgehen, die sich im Fernfeld zu einer ebenen Welle – der Reflektion – überlagern.

Die Reflektionen addieren sich entsprechend ihrer Phasenbeziehung.
Die Phasendrehung, die die einfallende Normalkomponente auf der Strecke $d_{h,k,l}$ von einer Ebene bis zur nächsten erfährt, ist

$$\vec{k}_G \; d_{h,k,l} \; \vec{e}_G = k_G \; d_{h,k,l} \; . \tag{1.92}$$

Sie wird dort reflektiert und erfährt nochmals die gleiche Phasendrehung, bis sie wieder an der davor liegenden Ebene ist. Dort überlagert sie sich mit einer von dieser Ebene reflektierten Welle. Die Überlagerung erfolgt konstruktiv (gleichphasig) unter der Bedingung

$$2 \; k_G \, d_{h,k,l} = n \cdot 2\pi \Rightarrow k_G = \frac{n \, \pi}{d_{h,k,l}} \; . \tag{1.93}$$

Mit der Konvention $\frac{d_{h,k,l}}{n} = d_{nh,nk,nl} = d_{h,k,l}$ ergibt sich

$$k_G = \frac{\pi}{d_{h,k,l}} = k \sin \phi \; . \tag{1.94}$$

Diese Reflektionen finden zwischen allen Ebenen der Ebenenschar statt und ergeben, wenn Gl. (1.94) erfüllt ist, eine konstruktive Überlagerung.
Wir bezeichnen Gl. (1.94) nach seinem Entdecker als Bragg-Bedingung und die, in diesem Fall stattfindende Reflektion als Bragg-Reflektion.
Gehen wir davon aus, dass die der Reflektion zugrunde liegende Streuung elastisch ist, so bleibt die Energie erhalten. Die Energie der einfallenden Welle

$$W_G = h f_G = h \frac{v_{\mathrm{ph}}}{\lambda_G} = h \frac{v_{\mathrm{ph}}}{2\pi} k_G = \hbar \, v_{\mathrm{ph}} \, k_G \tag{1.95}$$

ist dann gleich der Energie der reflektierten Welle. Deren Wellenvektor muss daher den gleichen Betrag, aber wegen der entgegengesetzten Ausbreitungsrichtung ein entgegengesetztes Vorzeichen haben.
Unter der Annahme, dass die Tangentialkomponente $\vec{k}_t$ der einfallenden Welle aufgrund ihrer Orientierung im Kristall keine Reflektion erfährt, überlagert sie sich unverändert mit der reflektierten Normal-Komponente. Daraus ergibt sich der Wellenvektor der reflektierten Welle

$$\vec{k}' = -\vec{k}_G + \vec{k}_t \; . \tag{1.96}$$

Abb. 1.31 veranschaulicht die Beziehung zwischen den Wellenvektoren von einfallender und reflektierter Welle.

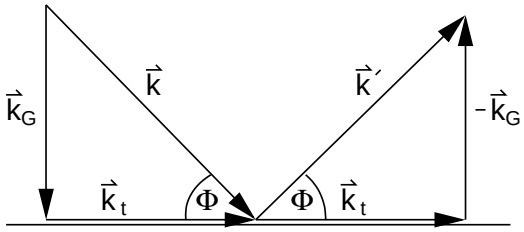

Abb. 1.31: Zusammenhang zwischen den Wellenvektoren von ein- und ausfallender Welle bei Annahme von elastischer Streuung.

Aus Abb. 1.31 kann direkt abgelesen werden

$$\vec{k} - \vec{k}' = 2\,\vec{k}_G = \frac{2\pi}{d_{h,k,l}}\vec{e}_G\ . \tag{1.97}$$

Wir definieren mit

$$\vec{G}_{h,k,l} := \frac{2\pi}{d_{h,k,l}}\vec{e}_G \tag{1.98}$$

einen Vektor des reziproken Gitters mit den Eigenschaften

1. $\vec{G}_{h,k,l}$ steht senkrecht auf der Ebenenschar (h,k,l) (1.99)

2. $$|\vec{G}_{h,k,l}| = \frac{2\pi}{d_{h,k,l}} \tag{1.100}$$

und können damit für Gl. (1.97) schreiben

$$\vec{k} - \vec{k}' = \vec{G}_{h,k,l}\ . \tag{1.101}$$

Dies ist eine sehr einfache vektorielle Beziehung, wobei berücksichtigt werden muss, dass diese Gleichung im dreidimensionalen Raum gilt. Abb. 1.32 zeigt Lösungen der Beziehung in einer zweidimensionalen Projektion.

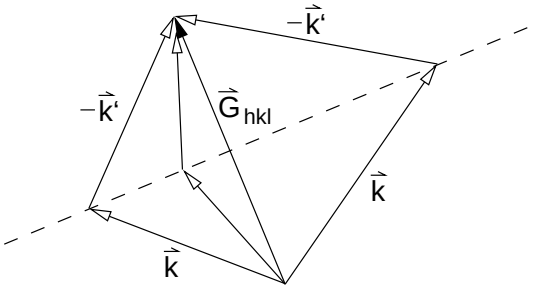

Abb. 1.32: Darstellung der Bragg-Bedingung für (Total-)Reflektion einer Elektronenwelle am Kristallgitter anhand der Wellenvektoren $\vec{k}$ und $\vec{k}'$ der einfallenden und reflektierten Welle. $\vec{G}_{h,k,l}$ wird durch den Aufbau eines Kristallgitters bestimmt. Alle $\vec{k}$, die auf der Mittelhalbierenden von $\vec{G}_{h,k,l}$ liegen, sind Lösungen von Gl. (1.101). Der reflektierte Vektor $\vec{k}'$ besitzt dann automatisch den richtigen Betrag ($|\vec{k}| = |\vec{k}'|$)und Winkel ϕ (vgl. Abb. 1.31).

Beispiel: Wir können anhand der Eigenschaften (1.99) und (1.100) für das zweidimensionale Kristallgitter in Abb. 1.33 links einige Vektoren $\vec{G}_{h,k,l}$ des reziproken Gitters zeichnen. Diese sind rechts ausgehend von einem gemeinsamen Ursprung zusammengefasst.

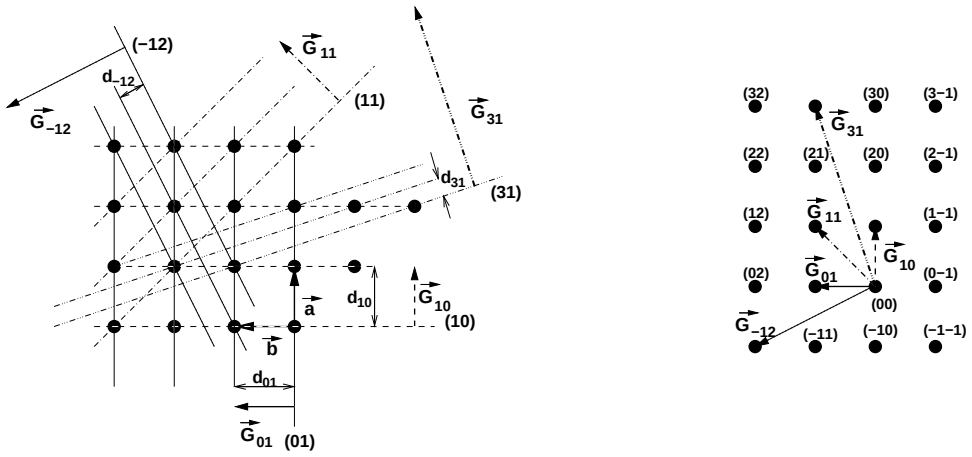

Abb. 1.33: Links: Zweidimensionales Kristallgitter mit einigen eingezeichneten Ebenen (Netzebenen) und deren reziproken Gittervektoren $\vec{G}_{h,k}$. Rechts: Jedes $\vec{G}_{h,k}$ zeigt auf einen Punkt im reziproken Gitter.

Beachten: Die Einheit der reziproken Gittervektoren ist m^{-1}. Zur besseren Darstellung ist in Abb. 1.33 daher für die $\vec{G}_{h,k}$ ein anderer Maßstab als für die Kristallgitter gewählt worden. Durch das quadratisch gewählte Gitter erscheint auch das reziproke Gitter quadratisch und hat dieselbe Struktur wie das Kristallgitter. Bei Abweichungen von der quadratischen Kristallgitterform ändert auch das reziproke Gitter seine Form. Diese ist dann aber nicht mehr identisch mit dem Kristallgitter.

Abb. 1.33 rechts zeigt eine wichtige Eigenschaft der reziproken Gittervektoren. Sie bilden ein reziprokes Gitter, das sich aus einer reziproken Elementarzelle mit dem reziproken Translationsvektoren $\vec{G}_{01}$ und $\vec{G}_{10}$ zusammensetzt. Der Index der Gitterpunkte stimmt mit den Miller-Indizees (h,k,l) der jeweiligen Ebenenschar überein.

Wir führen in Analogie zu den Translationsvektoren $(\vec{a}, \vec{b}, \vec{c})$ des Kristallgitters für das reziproke Gitter die Translationsvektoren $(\vec{A}, \vec{B}, \vec{C})$ ein.
Für das zweidimensionale Beispiel (ohne $\vec{c}$) in Abb. 1.33 rechts gilt dann z. B. eine Zuordnung $\vec{A} = \vec{G}_{01},\ \vec{B} = \vec{G}_{10}$. Allgemein im Dreidimensionalen ist jeder Gitterpunkt über einen Vektor

$$\vec{G_{h,k,l}} = (h\vec{A} + k\vec{B} + l\vec{C}) \tag{1.102}$$

erreichbar.
Im Beispiel in Abb. 1.33 wurden die Regeln (1.99) und (1.100) zur Konstruktion einiger reziproker Gittervektoren im Zweidimensionalen angewandt.
Um auf einfache Weise die Konstruktion aller reziproker Gittervektoren im Dreidimensionale zu ermöglichen, ist nach Gl. (1.102) nur die Kenntnis der Translationsvektoren des reziproken Gitters nötig. Diese ergeben sich formal aus der Berechnungsvorschrift[9]

$$\vec{A} = 2\pi\frac{\vec{b}\times\vec{c}}{V_{EZ}},\ \vec{B} = 2\pi\frac{\vec{c}\times\vec{a}}{V_{EZ}},\ \vec{C} = 2\pi\frac{\vec{a}\times\vec{b}}{V_{EZ}}\ , \tag{1.103}$$

wobei $(\vec{a}, \vec{b}, \vec{c})$ die Translationsvektoren des Kristallgitters und $V_{EZ} = (\vec{a} \times \vec{b})\vec{c} = (\vec{b} \times \vec{c})\vec{a} = (\vec{c} \times \vec{a})\vec{b}$ (Spatprodukt) das Volumen der Elementarzelle ist. Dass Gl. (1.103) unserer Berechnung für reziproke Gittervektoren genügt, zeigt eine einfache Betrachtung:

- Das Kreuzprodukt im Zähler steht immer senkrecht auf der, durch die beiden Vektoren gebildeten Ebene. Damit ist Forderung (1.99) erfüllt.
- Der Betrag des Kreuzproduktes im Zähler ist gleich der von seinen Vektoren aufgespannten Fläche. Wir nennen sie F und wissen, dass F verschiedene Werte F_{bc}, F_{ca}, F_{ab} annehmen kann. Im Nenner steht

[9] Es lässt sich aufgrund der Periodizität zeigen, dass das reziproke Gitter durch Fourier-Transformation aus dem Kristallgitter (Ortsgitter) hervorgeht.

das Volumen der Elementarzelle. Diese berechnet sich aus dem Produkt von F und der, auf F senkrecht stehenden Komponente des dritten Vektors. Die Länge dieser Komponente ist genau der Abstand ($d = d_{100},\ d_{010}\ d_{001}$) von zwei benachbarten (100), (010) oder (001)-Flächen. Für das Volumen der Elementarzelle kann man also $V_{EZ} = F \cdot d$ schreiben, wodurch sich F in Zähler und Nenner kürzt und wir die Beträge

$$|\vec{A}| = \frac{2\pi}{d_{100}},\ |\vec{B}| = \frac{2\pi}{d_{010}},\ |\vec{C}| = \frac{2\pi}{d_{001}} \tag{1.104}$$

für die reziproken Translationsvektoren erhalten. Damit ist auch Forderung (1.100) erfüllt.

Wir stellen fest, dass durch formales Anwenden von Gl. (1.103) auf ein Kristallgitter ein zweites Gitter – das reziproke Gitter – entsteht. Dieses repräsentiert den k-Raum und eignet sich daher mit Gl. (1.102) zur Darstellung von Wellenvektoren entsprechend der Reflektonsbedingung nach Gl. (1.101). Danach erfüllen nur die Kombinationen $\vec{k} - \vec{k}'$ die Reflektionsbedingung, die wie in Abb. 1.32 gezeigt einen reziproken Gittervektor ergeben. Die dort gezeigte Konstruktion der $\vec{k}$ und $\vec{k}'$ mit Hilfe der Mittelhalbierenden lässt sich direkt auf das reziproke Gitter übertragen, da jeder Vektor zwischen Punkten des Gitters ein reziproker Gittervektor ist.
Abb. 1.33 links zeigt die Konstruktion ausgehend von einem (beliebigen) Ursprung des Gitters, für die Gittervektoren, deren Mittelhalbierenden die kleinstmögliche Fläche (hier ein Quadrat) umschließen. Für den realen dreidimensionalen Kristall werden die Mittelhalbierenden zu Flächen und der einbeschriebene Körper zu einem Polyeder (bei dreidimensionaler Fortsetzung von Abb. 1.33 ergibt sich ein Kubus).
Wir nennen den kleinsten, durch die mittelhalbierenden Flächen eingefassten Polyeder die erste Brillouin-Zone. Die gleiche Bezeichnung verwenden wir für die von den mittelhalbierenden Geraden eingeschlossene Fläche in der zweidimensionalen Darstellung.
Den nächstmöglichen Körper geringsten Flächeninhalts nennen wir die zweite Brillouin-Zone (BZ), den dritten, die dritte BZ u.s.w. Jede Brillouin-Zone hat das Volumen der Elementarzelle. Abb. 1.34 rechts zeigt für das einfache zweidimensionale Beispiel die Konstruktion der zweiten BZ.
In Abb. 1.35 sind für die komplizierten kubisch-flächen- und -raumzentrierten Gitter die Körper der ersten BZ gezeigt.
Wir merken uns, dass alle Wellenvektoren $\vec{k}$, die (ausgehend vom Ursprung)

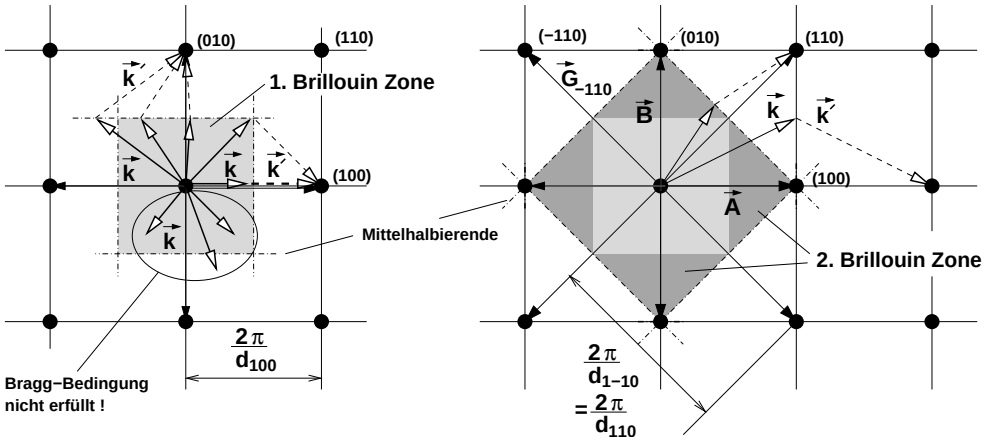

Abb. 1.34: Konstruktion der ersten (links) und zweiten (rechts) Brillouin-Zone (BZ) im Zweidimensionalen. Die Flächen ergeben sich als kleinstmögliche, durch die Mittelhalbierenden der reziproken Gittervektoren einbeschriebenen Fläche. Zur Fläche der 2. BZ zählt die der 1. BZ nicht dazu. Zu sehen ist, dass alle $\vec{k}$ die auf einer BZ enden, die Bragg-Bedingung erfüllen.

auf einer der Flächen der Brillouin-Zonen liegen, die Bedingung für Bragg-Reflektion nach Gl. (1.101) erfüllen.

1.26 Bandlücke durch Bragg-Reflektion

Je weiter ein Wellenvektor der Wellenfunktion eines Elektrons von einer BZ entfernt ist, umso weniger wird die Wellenfunktion reflektiert. Das Elektron erfährt dann keine Auswirkung der Gitterstruktur und verhält sich näherungsweise wie ein freies Elektron.
Wir betrachten die Wellenfunktion ψ eines reflektierten Elektrons. Sie besteht aus einer hinlaufenden und einer, aufgrund der Reflektion zurücklaufenden Wellenfunktion mit den Wellenvektoren $\vec{k}$ und $\vec{k}'$. Wir können allgemein für die Überlagerung dieser beiden Wellen schreiben:

$$\psi = ae^{j\vec{k}\vec{r}} \pm ae^{j\vec{k}'\vec{r}} \,. \tag{1.105}$$

Aufgrund der als elastisch angenommenen Streuung besitzen beide Wellen die gleiche Amplitude. Da wir über das Vorzeichen der reflektierten Welle

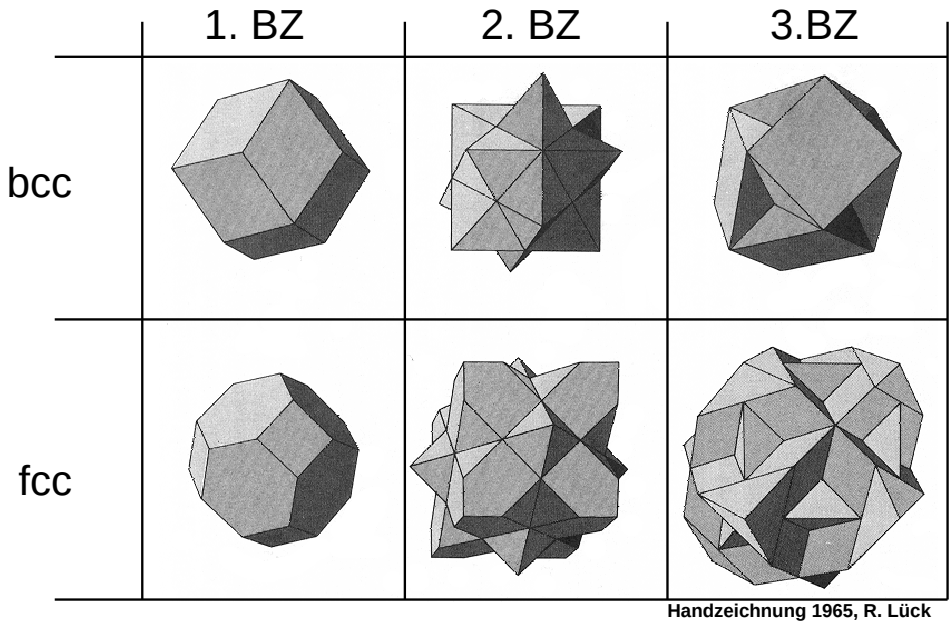

Abb. 1.35: Die ersten drei Brillouin-Zonen eines kubisch raumzentrierten (bcc) und flächenzentrierten (fcc) Gitters.

keine Aussage machen können, ist sowohl eine Überlagerung mit positivem als auch negativem Vorzeichen möglich. Um diesen Ausdruck interpretieren zu können, zerlegen wir die Wellenvektoren der Wellen wie schon zuvor die Wellenvektoren nach Abb. 1.31 in zwei Komponenten. Eine Komponente $\vec{k}_G$ zeigt in Richtung des reziproken Gittervektors $\vec{G}$ (wir schreiben zur Abkürzung $\vec{G}$ anstatt $\vec{G}_{h,k,l}$). Die andere Komponente, wir nennen sie Transversalkomponente $\vec{k}_t$, steht senkrecht auf $\vec{k}_G$. Sie ist für unsere Untersuchung der Reflektion ohne Bedeutung, da sie durch die Reflektion nicht verändert wird (vgl. Abb. 1.31). Mit $\vec{k}' = -\vec{k}_G + \vec{k}_t$ nach Gl. (1.96) gilt:

$$\psi = ae^{j(\vec{k}_G+\vec{k}_t)\vec{r}} \pm ae^{j(-\vec{k}_G+\vec{k}_t)\vec{r}} \tag{1.106}$$

$$= ae^{j\vec{k}_t\vec{r}}(e^{j\vec{k}_G\vec{r}} \pm e^{-j\vec{k}_G\vec{r}}) \; . \tag{1.107}$$

Betrachten wir mit $\vec{r} = \vec{r}_G$ die Wellenfunktion in Richtung der Reflektion, also in Richtung $\vec{k}_G$, so gilt wegen $\vec{k}_t \perp \vec{r}_G$ und $\vec{k}_G \parallel \vec{r}_G$:

$$\psi_G = a(e^{jk_Gr_G} \pm e^{-jk_Gr_G}) \; . \tag{1.108}$$

a ist ein konstanter Faktor, der sich aus der Normierungsbedingung ergibt. Für unsere Betrachtung ist er ohne Bedeutung und wir setzen im Folgenden

$a = \frac{1}{2}$.
Um zwischen den beiden Wellenfunktionen unterscheiden zu können, nennen wir

$$\psi_G^+ = \frac{1}{2}\left(e^{jk_G r_G} + e^{-jk_G r_G}\right) \tag{1.109}$$

$$\psi_G^- = \frac{1}{2}\left(e^{jk_G r_G} - e^{-jk_G r_G}\right) \ . \tag{1.110}$$

Mit diesem Ergebnis lässt sich die Wellenfunktion des reflektierten Elektrons interpretieren. Dazu ermitteln wir die Aufenthaltswahrscheinlichkeiten zu den beiden Lösungen, die sich durch Anwenden der Euler-Umformung nach kurzer Rechnung angeben läßt:

$$|\psi_G^+|^2 = \psi_G^+(\psi_G^+)^* = 4\cos^2 k_G r_G \propto \cos^2 \frac{G}{2} r_G = \cos^2 \frac{2\pi}{2d_{h,k,l}} r_G \tag{1.111}$$

$$|\psi_G^-|^2 = \psi_G^-(\psi_G^-)^* = 4\sin^2 k_G r_G \propto \sin^2 \frac{G}{2} r_G = \sin^2 \frac{2\pi}{2d_{h,k,l}} r_G \tag{1.112}$$

In den letzten beiden Umformungen in Gl. (1.111) und (1.112) wurde anstelle k_G der reziproke Gittervektor $G = \frac{2\pi}{d_{h,k,l}}$ nach Gl. (1.100) eingesetzt, durch den die Bragg-Bedingung erfüllt wird. Das war der Ausgangspunkt für den Ansatz in Gl. (1.105).

Gl. (1.111) und (1.112) beschreiben zwei mögliche Lösungen der Aufenthaltswahrscheinlichkeit der Elektronen, die vom Gitter, genauer gesagt von Gitterebenen im Abstand $d_{h,k,l}$, reflektiert werden. Wir stellen die beiden Lösungen in Abb. 1.36 grafisch dar. Zusätzlich zeichnen wir die Atome im Abstad $d_{h,k,l}$ des Gitters sowie deren periodischen Verlauf aus der Überlagerung der Potentialtöpfe.
Beide Aufenthaltswahrscheinlichkeiten sind im Gegensatz zum freien Elektronengas stehende Wellen. Sie sind so zu den Atomen (genauer: Ionenrümpfen) angeordnet, dass sich in jedem der sich periodisch wiederholenden Lösungsräume (der Raum/Abstand zwischen zwei Gitterebenen im Abstand d_{hkl}) identische und stetig fortsetzende Verläufe ergeben.
Aus der Aufenthaltswahrscheinlichkeit der beiden Lösungen ergibt sich eine überaus wichtige Schlussfolgerung:
Offensichtlich halten sich die ψ_G^+-Elektronen bevorzugt in der Nähe des Potentials der Gitterionen auf. Sie besitzen daher eine niedrigere potentielle Energie als freie Elektronen, die sich überall gleich wahrscheinlich aufhalten und daher sich auch in dem bevorzugten Aufenthaltsbereich der

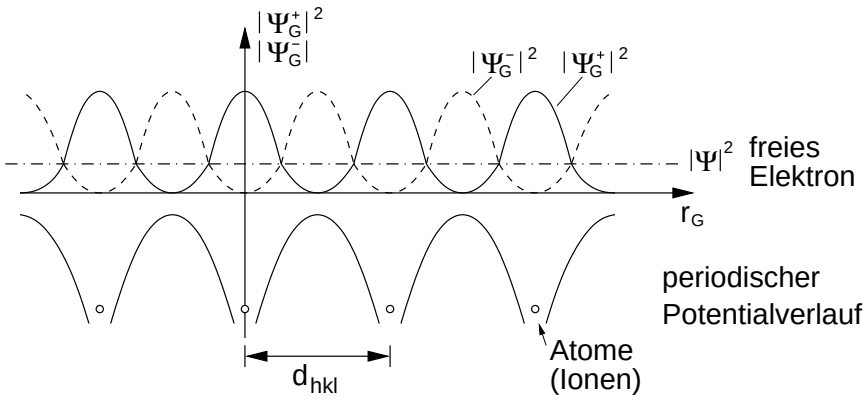

Abb. 1.36: Aufenthaltswahrscheinlichkeiten $|\psi_G^+|^2$ und $|\psi_G^-|^2$ Bragg-reflektierter Elektronen zwischen Gitterebenen mit einem Abstand $d_{h,k,l}$. Zum Vergleich ist auch die in allen Punkten gleiche Aufenthaltswahrscheinlichkeit freier Elektronen dargestellt.

ψ_G^--Elektronen befinden. Da sich diese bevorzugt zwischen den Ionenrümpfen aufhalten, besitzen sie eine höhere potentielle Energie als freie Elektronen.

Wir fassen die bisherigen Erkenntnisse zusammen:

- Freie Elektronen mit Wellenvektoren, die die Bragg-Bedingung nicht erfüllen, können sich als Elektronengas ungehindert im Kristall bewegen. Sie sind über den Kristall ausgeschmiert. Die von ihnen einnehmbaren Quantenzustände besitzen Energiewerte nach Gl. (1.76), die quantisiert, aber so eng aufeinanderfolgend sind, so dass sie als eine kontinuierliche Parabel angenommen werden können.

- Freie Elektronen, deren Wellenvektor die Bragg-Bedingung erfüllt, werden durch das Gitter des Kristalls reflektiert. Sie haben Wellenvektoren $\vec{k}_{BZ}$, die auf den Grenzflächen der Brillouin-Zonen liegen. Zu diesen Wellenvektoren existieren zwei Energiewerte, der eine kleiner, der andere größer als die Energie der freien Elektronen mit dem gleichen Wellenvektor. Wir definieren die Abweichung von der Energie der freien Elektronen mit $\pm\Delta W$, wobei wir eine symmetrische Abweichung unterstellen. Damit können wir allgemein für die Energie eines Elektrons

mit $k = k_{BZ}$ schreiben:

$$W(k_{BZ}) = \frac{\hbar^2 k_{BZ}^2}{2m_e} \pm \Delta W \ . \tag{1.113}$$

Je mehr sich ein Wellenvektor von der Bragg-Bedingung entfernt, umso mehr wird sich das Elektron wie ein freies Elektron verhalten.

Wir können mit diesen Ergebnissen das Bild der Energie der freien Elektronen aus Abb. 1.24 um die Besonderheiten aufgrund der Bragg-Reflektion an den Grenzen der Brillouin-Zonen erweitern. Es ergibt sich Abb. 1.37.

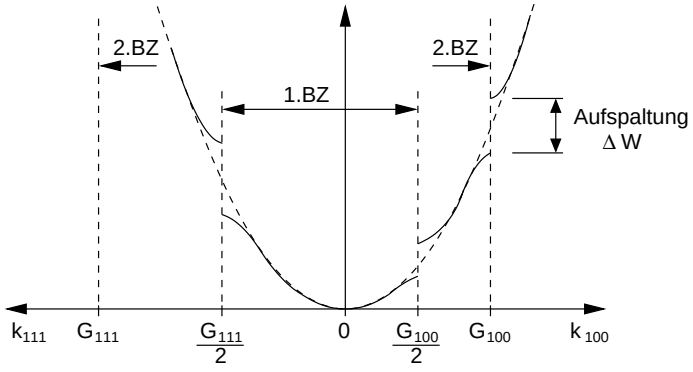

Abb. 1.37: Aufspaltung der Energieniveaus der freien Elektronen-Parabel an den Grenzen der Brillouin-Zonen (BZ). Dargestellt sind zwei Wellenvektoren, die in Richtung der Ebenen (100) und (111) (Abstand d_{100} und d_{111}) reflektiert werden.

Deutlich gezeigt ist die Aufspaltung der Energieniveaus an den Grenzen der Brillouin-Zonen. Aufgrund der räumlichen Gestalt der Brillouin-Zonen ergeben sich die Aufspaltungen abhängig von der Richtung des Wellenvektors bei unterschiedlichen Beträgen. Außerhalb der Grenzen geht der Verlauf in den parabelförmigen Verlauf der freien Elektronen über.
Durch diese Aufspaltung entsteht in der Dispersionskurve einer $\vec{k}$-Richtung eine Energielücke. Dies muss jedoch nicht zwangsläufig bedeuten, dass kein Elektron im Kristall einen Energiewert in dieser Lücke annehmen kann. Gibt es in einer anderen Richtung von $\vec{k}$ einen kontinuierlichen Verlauf der Dispersionskurve im Bereich der Energielücke, so können diese Energien

durch Elektronen mit einem $\vec{k}$ der entsprechenden Richtung besetzt werden. Dies ist z. B. in Abb. 1.37 für die unterste Energielücke für k_{100} der Fall, die durch Wellenfunktionen in k_{111}-Richtung geschlossen werden.

1.27 Reduziertes Bandschema

Aufgrund der Periodizität gibt es im reziproken Gitter keine ausgezeichneten Punkte. Die Wahl eines bestimmten Gitterpunktes erfolgt willkürlich. Die von einem gewählten Ursprung ausgehend ermittelten Ergebnisse, z. B. die Lage einer Brillouin-Zone, müssen auch (für andere Elektronen) um einen beliebigen anderen reziproken Gitterpunkt als Ursprung gelten.
Basierend auf dieser Überlegung lässt sich eine gegenüber Abb. 1.37 reduzierte Darstellung der Dispersionskurven erzielen. Dabei stellt man sich vor, dass sich ausgehend von dem nächsten benachbarten Gitterpunkt als Ursprung wieder die gleiche Dispersionskurve ergibt, ebenso von dem darauffolgenden und allen weiteren Gitterpunkten. Die Ursprünge der Dispersionskurven sind dann jeweils um einen reziproken Gittervektor der betrachteten Richtung von $\vec{k}$ entfernt. Im Beispiel in Abb. 1.37 befinden sich also die Ursprünge in der $\vec{k}_{100}$-Richtung im Abstand G_{100} und in der $\vec{k}_{111}$-Richtung im Abstand G_{111}.
Abb. 1.38 zeigt die Überlagerung der einzelnen Dispersionskurven mit den Ursprüngen im Abstand der Gittervektoren. Es ergeben sich periodisch fortgesetzte Energiebänder.
Aus der Darstellung kann man sehen, dass sich innerhalb des reduzierten Bereichs im Abstand der Länge eines halben reziproken Gittervektors der jeweiligen Richtung ($\frac{1}{2}G_{100}$, $\frac{1}{2}G_{111}$) sämtliche Verläufe der Energiebänder wiederfinden.
Man verwendet daher häufig die sogenannte reduzierte Darstellung, bei der die außerhalb liegenden Verläufe um einen reziproken Gittervektor verschoben werden, so dass sie im reduzierten Bereich liegen. Die reduzierte Darstellung für Abb. 1.37 bzw. Abb. 1.38 ist in Abb. 1.39 links nochmals zur Verdeutlichung gezeigt.
Auf der rechten Seite ist das zu den Verläufen gehörende Bändermodell angegeben. Es besitzt eine Energielücke W_g im Bereich der nicht überlappenden Verläufe. Ist das obere Band ohne zusätzliche Anregung noch unbesetzt, so stellt es das Leitungsband dar. Das darunter liegende Band ist das Valenzband.

Die realen Bandverläufe sind natürlich viel komplizierter als die hier mit der vereinfachten Annahme Bragg-reflektierter freier Elektronen hergeleiteten Kurven.
Um sie genauer zu bestimmen, müsste die Schrödingergleichung für den gesamten Kristall gelöst oder mit aufwendigen Verfahren und Annahmen Näherungslösungen erarbeitet werden.
Die für unser Verständnis der Bauelemente wichtigen Aussagen bleiben jedoch qualitativ die gleichen und können daher auf dem einfachen Stand der hier gebrachten Herleitung belassen werden.
Eine Erweiterung ist jedoch anzubringen, die auch in aufwendigeren Lösungen notwendig ist. Sie betrifft die Lage der Minima und Maxima der jeweiligen Bänder. Diese liegen sich in unserem einfachen Modell immer bei gleichen Werten der k-Vektoren auf der Grenze der Brillouin-Zone gegenüber (vgl. Abb. 1.39).
Abb. 1.40 zeigt demgegenüber die realen Verläufe der einzelnen Bänder für **Si**, **Ge** und **GaAs**. Hier ist zu erkennen, dass sich Maxima und Minima nicht immer an den Grenzen der Brillouin-Zone ergeben.
Als eine wichtige Eigenschaft ist zu sehen, dass sich außer bei **GaAs**

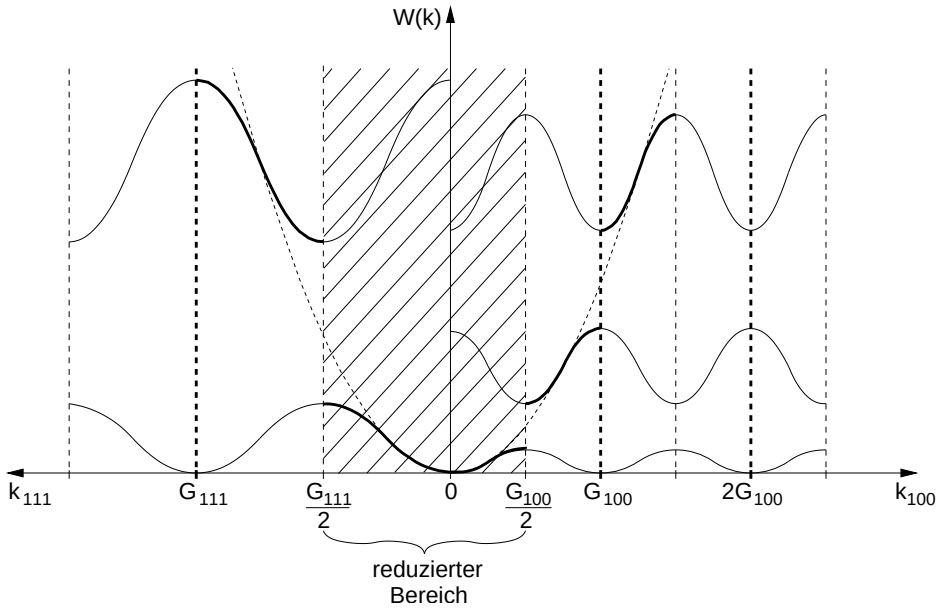

Abb. 1.38: Überlagerung der Dispersionskurven von $\vec{k}$-Vektoren mit Ursprung an den einzelnen reziproken Gitterpunkten in der jeweiligen Richtung von $\vec{k}$.

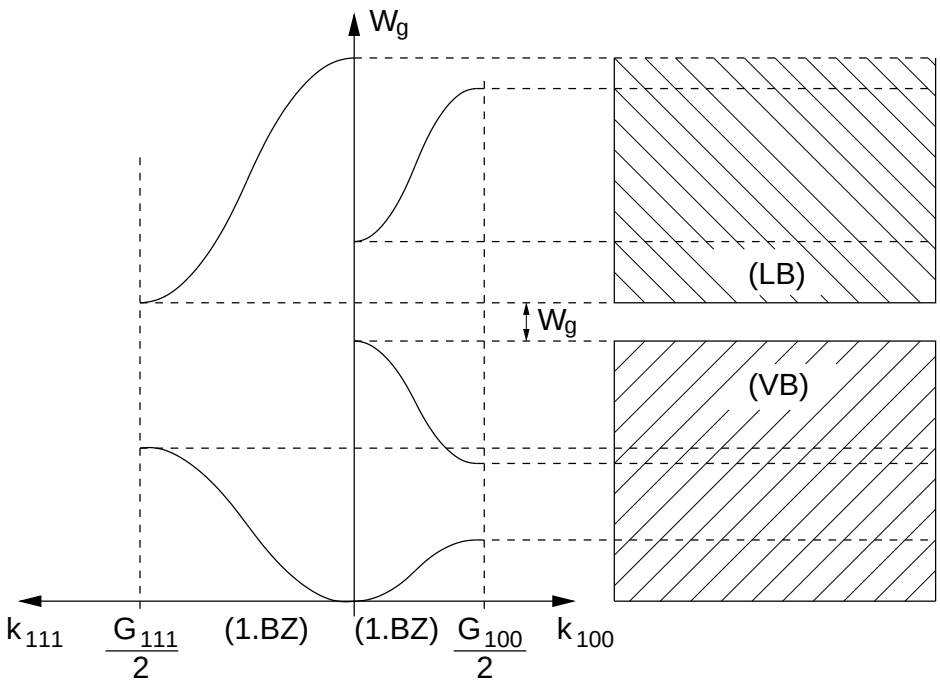

Abb. 1.39: Links: Reduziertes Dispersionsdiagramm. Rechts: Vereinfachtes Bändermodell mit einer Energielücke W_g aufgrund des mit Vektoren in $\vec{k}_{111}$- und $\vec{k}_{100}$-Richtung nicht zu schließenden Bereichs W_g.

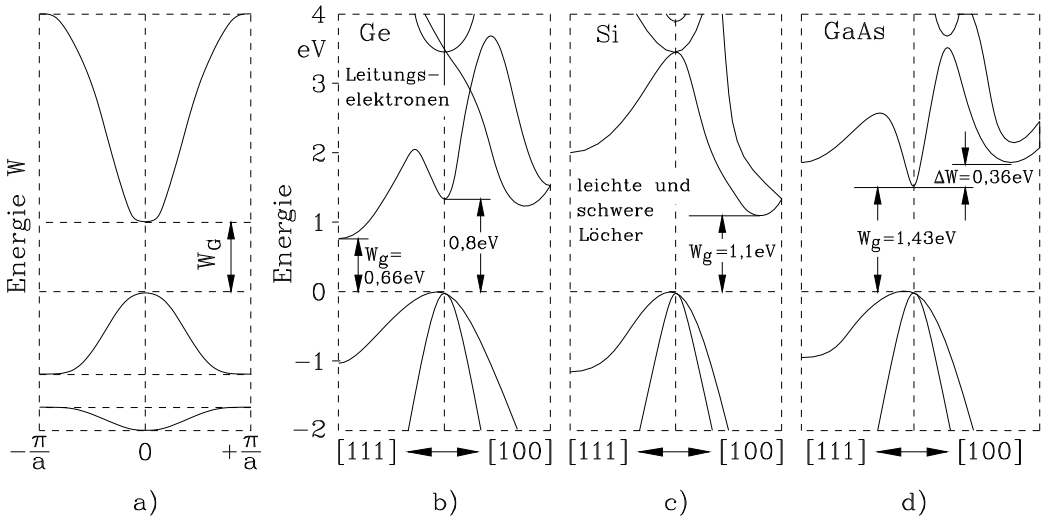

Abb. 1.40: Skizzierte reale Bandverläufe und Energielücke für **Si**, **Ge** und **GaAs**. Ganz links ist nochmals der bisher besprochene vereinfachte Verlauf dargestellt.

das Minimum des Leitungsbandes und das Maximum des Valenzbandes nicht gegenüberliegen, d. h. sie liegen nicht bei dem gleichen Wellenvektor. Diese Halbleiter nennt man indirekte Halbleiter. **GaAs** dagegen wird als direkter Halbleiter bezeichnet. Diese Unterscheidung ist wichtig, wenn es um die Fähigkeit eines Halbleitermaterials geht, Licht zu emittieren. Diese Fähigkeit besitzen nur direkte Halbleiter, weswegen **GaAs** zur Herstellung von Leuchtdioden verwendet wird. Wir kommen später im Kapitel „Direkte Rekombination“ darauf zurück.
Abb. 1.40 zeigt eine weitere Besonderheit in der Art des Maximums des Valenzbandes. Dieses wird durch die Maxima von zwei Bändern gebildet, die jedoch eine unterschiedliche Krümmung besitzen. Wie wir später sehen werden, besteht ein direkter Zusammenhang der Krümmung der $W(\vec{k})$-Kurve mit der Masse der Ladungsträger.

Wir haben in den letzten Kapiteln mit Hilfe des Modells freier Elektronen mit Streuung an der Gitterstruktur ein genaueres Verständnis für die Ursache und den Verlauf der Bandstruktur in Halbleitern erlangt. Das angewendete Modell stellt eine andere Betrachtungsweise dar als die eingangs verwendete Vorstellung sich aufspaltender Energieniveaus bei der Bildung von Kristallen. Letztendlich liefern beide Vorstellungen ausgehend von ihrer Betrachtungsweise eine Beschreibung des gleichen Festkörpers und führen zu den gleichen Resultaten. So sind z. B. die sich aus Abb. 1.18 ergebenden Energiebänder mit der Energielücke W_g identisch mit denen in Abb. 1.18. Der Vorteil des Modells freier Elektronen ist, dass es gerade die freien Ladungsträger sind, die die für uns wichtigen Eigenschaften von Halbleitern bestimmen.

1.28 Unschärferelation, Phasengeschwindigkeit, Gruppengeschwindigkeit

1.28.1 Unschärferelation

Wir betrachten das Elektron in seiner Darstellung als Wellenfunktion. Abb. 1.41 zeigt eine Wellenfunktion, die sich aus der Überlagerung mehrerer Wellenfunktionen ψ_n mit unterschiedlichen Frequenzen ergibt.
Jede der Wellenfunktionen ψ_n beschreibt eine zeitabhängige, sich im Raum ausbreitende Welle. Wir betrachten willkürlich und ohne Einschränkung der

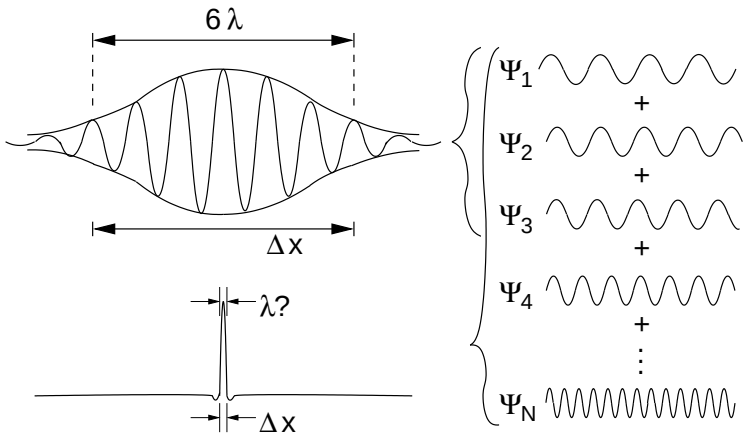

Abb. 1.41: Überlagerung von monofrequenten Wellenfunktionen zu einer Wellenfunktion ψ mit örtlich lokalisierter Aufenthaltswahrscheinlichkeit $|\psi|^2$. Je mehr Schwingungen pro Frequenzintervall bis zu $\omega \to \infty$ hinzukommen, umso kleiner wird die örtliche Lokalisation Δx.

allgemeinen Aussage eine Ausbreitung in x-Richtung. Die Wellengleichung nach Gl. (1.30) lautet allgemein

$$\psi_n = \cos(\omega_n t - k_n x) \; . \tag{1.114}$$

Durch die Überlagerung einer Anzahl N dieser Wellengleichungen entsteht die in Abb. 1.41 rechts gezeigte Wellengruppe

$$\psi = \sum_{n=1}^{N} \psi_n \; . \tag{1.115}$$

Sie besitzt je nach der Größe von N eine örtliche Ausdehnung ihres Maximums von Δx. Das ist i. e. der Bereich, in dem wir das Elektron erwarten. Er ist in dem oberen Beispiel mit $N = 3$ aufgrund der vielen (6) Schwingungszyklen in der Wellengruppe relativ groß. Jedoch ermöglichen es die vielen in Δx enthaltenen Wellenzyklen auch, die Wellenlänge genau zu bestimmen.
Wird N erhöht, werden mehr Schwingungen aus einem größer werdenden Frequenzbereich überlagert. Dadurch verringert sich die Breite Δx der Ortsunsicherheit. Im Grenzfall $N \to \infty$ geht $\Delta x \to 0$.
Demgegenüber erhöht sich die Unsicherheit bei der Bestimmung der Wellenlänge mit dem größer werdenden Frequenzbereich, da Δx abnimmt (vgl. Abb. 1.41 unten).

Über die de Broglie-Wellenlänge nach Gl. (1.3) ist der Wellenlänge ein Impuls $p = \frac{h}{\lambda}$ und damit auch eine Impulsunsicherheit (man sagt auch Unschärfe oder Unbestimmtheit) Δp zugeordnet. Mit Δp ergibt sich die Heisenbergsche Unschärferelation in ihrer bekannten Formulierung

$$\Delta p \cdot \Delta x \geq h \quad . \tag{1.116}$$

Setzt man für $\Delta p = \frac{h}{\Delta\lambda} = \hbar\Delta k$, erhält man eine Formulierung für die Unbestimmtheit der Wellenfunktion

$$\Delta k \cdot \Delta x \geq 2\pi \tag{1.117}$$

Darin ist Δk der Bereich (Bandbreite) der Wellenvektoren $k_1 \ldots k_N$, die nach Gl. (1.115) überlagert werden.

1.28.2 Phasengeschwindigkeit

Wir betrachten eine der Wellenfunktionen nach Gl. (1.114) und fragen, in welcher Zeit Δt eine bestimmte Phase φ den Weg Δx (Δx hat hier keinen Zusammenhang mit der Ortsunsicherheit des vorangegangenen Kapitels) zurückgelegt hat.
Zu Beginn sei der Einfachheit halber $t = 0$, dann ist die Phase (Argument des cos) an der Stelle x_1

$$\varphi = -kx_1 \; . \tag{1.118}$$

An der Stelle $x_1 + \Delta x$ müssen wir eine Zeit Δt warten, bis diese Phase φ erscheint. Es gilt daher

$$\varphi = -kx_1 = \omega\Delta t - k(x_1 + \Delta x) \tag{1.119}$$

und daraus folgt durch Umstellen

$$\omega\Delta t = k\Delta x \tag{1.120}$$

$$\mathrm{v}_p := \frac{\Delta x}{\Delta t} = \frac{\omega}{k} \tag{1.121}$$

die Phasengeschwindigkeit v_p.
Zu beachten ist, dass die Phasengeschwindigkeit der de Broglie-Welle wegen Gl. (1.8) immer größer als die Lichtgeschwindigkeit ist.
Für die Geschwindigkeit des Elektrons, das wegen des Maximums von $|\psi|^2$ innerhalb des Bereichs Δx der Wellengruppe lokalisiert ist (vgl. Abb. 1.41), ist die Geschwindigkeit der Wellengruppe maßgeblich. Diese Gruppengeschwindigkeit betrachten wir im nächsten Kapitel.

1.28.3 Gruppengeschwindigkeit

Wir betrachten die Überlagerung von zwei Wellenfunktionen im Sinne von Gl. (1.115) ($N = 2$). Wir nehmen an, die beiden unterscheiden sich in Frequenz bzw. Wellenvektor um $\pm\Delta\omega$ bzw. $\pm\Delta k$ von ihren Mittelwerten ω und x. Die Wellenfunktionen lauten dann:

$$\psi_1 = \cos((\omega + \Delta\omega)t - (k + \Delta k)x) \tag{1.122}$$

$$\psi_2 = \cos((\omega - \Delta\omega)t - (k - \Delta k)x) \; . \tag{1.123}$$

Die Überlagerung $\psi_1 + \psi_2$ lässt sich mittels Additionstheorem umformen zu

$$\psi_1 + \psi_2 = 2\cos(\omega t - kx)\cos(\Delta\omega t - \Delta k x) \; . \tag{1.124}$$

Der erste cos-Term beschreibt eine bekannte Wellenfunktion mit der Phasengeschwindigkeit nach Gl. (1.121). Der zweite cos-Term beschreibt eine Hüllkurve, ähnlich der gestrichelten Umhüllenden der Gruppe in Abb. 1.41 oben links. Analog zur Herleitung der Phasengeschwindigkeit lässt sich die Geschwindigkeit v_{gr} der Hüllkurve der Gruppe anhand des Arguments der zweiten cos-Funktion bestimmen.

$$v = \frac{\Delta x}{\Delta t} = \frac{\Delta\omega}{\Delta k} \; , \tag{1.125}$$

das für kleine Änderungen Δk und $\Delta\omega$ übergeht in die Gruppengeschwindigkeit

$$\mathrm{v}_{gr} = \frac{d\omega}{dk} \; . \tag{1.126}$$

Dies ist die Geschwindigkeit, mit der sich das Elektron bewegt.
Für beliebige Richtungen von $\vec{k}$ im Dreidimensionalen geht Gl. (1.126) über in

$$\vec{\mathrm{v}}_{gr} = \mathrm{grad}_{\vec{k}}\,\omega(\vec{k}) \; . \tag{1.127}$$

1.29 Konzept der effektiven Masse

Reale Bandverläufe haben eine kompliziertere Struktur, die von dem einfachen parabelförmigen Verlauf des freien Elektronengases abweicht (vgl. Abb. 1.37 und 1.40).
In den meisten Fällen[10] genügt die Betrachtung der Verläufe an den Bandkanten, die die Energielücke bilden, da diese Bereiche den Stromfluss in einem

[10] Wir betrachten hier z. B. nicht die Möglichkeit, dass zwischen den verschiedenen, in Leitungs- und Valenzband enthaltenen Bändern mit Nebenminima und -maxima Wechselwirkungen auftreten können (Interband- bzw. Intrabandstreuungen).

Halbleiter maßgeblich bestimmen (vgl. Kap. 1.15). Wir betrachten daher in Abb. 1.40 die Dispersionskurven an der Oberkante (Maximum) des Valenzbandes und an der Unterkante (Minimum) des Leitungsbandes.
Um eine Näherung der Bandverläufe im Bereich dieser Punkte zu erhalten, können wir die Kurven in den Punkten als Taylorreihe entwickeln. Wir beschränken uns hier der Einfachheit halber auf den zweidimensionalen Fall, also die Reihenentwicklung in eine bestimmte kristallographische Richtung. Mit k_0 als dem Betrag des Wellenvektors im Minimum/Maximum der Dispersionskurve des Leitungs-/Valenzbandes in der betrachteten Richtung, ergibt sich als Taylorentwicklung um k_0

$$W(k) = W_{V,C} + \Delta k \frac{dW(k)}{dk}\bigg|_{k_0} + \frac{\Delta k^2}{2}\frac{d^2W(k)}{dk^2}\bigg|_{k_0} + \dots \ . \tag{1.128}$$

Darin ist $\Delta k = k - k_0$ die Abweichung des Wellenvektors gegenüber dem k_0 im Minimum/Maximum und $W_{V,C}$ alternativ die Energie des Leitungsband-Minimums W_C oder des Valenzband-Maximums W_V.
Wir brechen die Entwicklung nach dem quadratischen Glied ab, nähern also den tatsächlichen Verlauf durch eine Parabel an. Abb. 1.42 zeigt an einem hypothetischen Bandverlauf die Näherung für Leitungs- und Valenzband.

Da wir beide Bänder um ihren Extremwert entwickeln, gilt

$$\frac{dW(k)}{dk}\bigg|_{k_0} = 0 \ , \tag{1.129}$$

und Gl. (1.128) vereinfacht sich zu

$$W(k) = W_{V,C} + \frac{\Delta k^2}{2}\frac{d^2W(k)}{dk^2}\bigg|_{k_0} \ . \tag{1.130}$$

Bevor wir anhand dieser Beziehung eine Näherung für Elektronen an den Bandkanten herleiten, betrachten wir die Aussage von Gl. (1.129). Aus ihr geht hervor, dass die Energie eines Elektrons an der Bandkante nicht vom Wellenvektor abhängt. Dies hat eine unmittelbare Auswirkung auf die Geschwindigkeit des Elektrons:
Für die Geschwindigkeit v_e eines Elektrons können wir nach Gl. (1.126) die Gruppengeschwindigkeit

$$\mathrm{v}_e = \mathrm{v}_{gr} = \frac{d\omega}{dk} \tag{1.131}$$

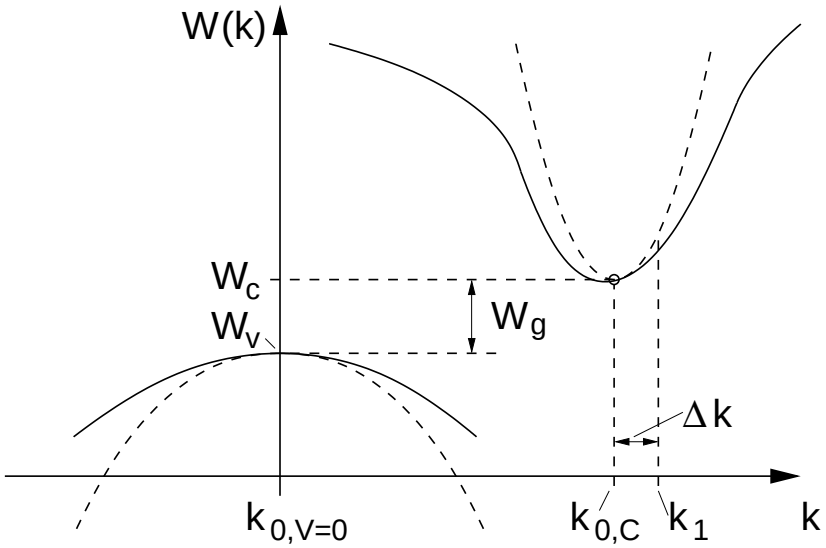

Abb. 1.42: Approximation von Maximum des Valenzbandes bzw. Minimum des Leitungsbandes durch eine Taylorreihenentwicklung zweiter Ordnung. Den Wert der Approximation an der Stelle k_1 in Gl. (1.128) erhält man z. B. für $\Delta k = k_1 - k_{0,C}$. Da für das Maximum des Valenzbandes $k_{0,V} = 0$ gilt, ist für die Reihenentwicklung des Valenzbandverlaufs in diesem Beispiel $\Delta k = k$.

setzen. Mit der Energie des Elektrons nach Gl. (1.5), $W = h \cdot f$ lässt sich die (Kreis-)Frequenz der Elektronenwelle angeben

$$W = \hbar\omega \rightarrow \omega = \frac{W}{\hbar} \,, \tag{1.132}$$

wodurch sich aus Gl. (1.131) die allgemeingültige Beziehung

$$v_e = \frac{1}{\hbar}\frac{dW}{dk} \tag{1.133}$$

ergibt.
Da an den Bandkanten ($k = k_{0,V}$, $k_{0,C}$) aber die Energie des Elektrons nicht vom Wellenvektor abhängt, ist die Geschwindigkeit der Elektronen an den Bandkanten gleich Null.
Dieses Ergebnis geht auch anschaulich aus der zuvor zur Entstehung der Bandlücken verwendeten Erklärung der Bragg-Reflektion an den Grenzen der Brillouin-Zone hervor: Wir betrachten an der Bandkante Elektronen mit einem Wellen-Vektor k_0, der der Bedingung für Bragg-Reflektion genügt.

Die Wellenfunktion für Bragg-reflektierte Wellen ist aber eine stehende Welle (vgl. Abb. 1.36). D. h. die Elektronen sind an einem Ort lokalisiert und bewegen sich nicht.[11] Daher ist $v_e = 0$.
Wir wollen im Folgenden eine Näherung herleiten, die es uns erlaubt, Elektronen im Bereich der Bandkanten von Leitungs- und Valenzband des Halbleiters genau wie Elektronen im freien Elektronengas zu behandeln. Dies gelingt uns, indem wir den Elektronen im Halbleiter eine effektive Masse m_e^* geben, durch die sie Eigenschaften wie Elektronen mit der Elektronenmasse m_e^* in einem freien Elektronengas besitzen. Wir vergleichen dazu die (mittlere) Geschwindigkeit eines Elektrons im Elektronengas mit der im Halbleiterkristall. Mit der Energie eines Elektrons in freien Elektronengas nach Gl. (1.77 b)

$$W = \frac{\hbar^2 k^2}{2m_e} \tag{1.134}$$

ergibt sich aus Gl. (1.133) die Geschwindigkeit des Elektrons im freien Elektronengas (vgl. auch Gl. (1.77 d))

$$v_{e,0} = \frac{\hbar k}{m_e} = \frac{\hbar\, 2\pi}{m_e\, \lambda} \overset{(1.3)}{=} \frac{p}{m_e} = \frac{m_e \cdot v}{m_e} = v \quad . \tag{1.135}$$

Im zweiten Teil von Gl. (1.135) wurde über die de Broglie-Wellenlänge nach Gl. (1.3) der Übergang zur Teilcheneigenschaft hergestellt. Über den Impuls $m_e \cdot v$ ergibt sich die kinetische Geschwindigkeit v identisch zu $v_{e,0}$, die über die Gruppengeschwindigkeit der Wellenfunktion des Elektrons mit Gl. (1.131) hergeleitet wurde.

Wir berechnen zum Vergleich die Geschwindigkeit des Elektrons im Halbleiter. Dazu verwenden wir die Näherung des Bandverlaufs nach

[11] Es sei hier nochmals darauf hingewiesen, dass es sich bei der Betrachtung um statistische Größen handelt. Genaugenommen ist nur die mittlere Geschwindigkeit der Elektronen mit der Energie $W(k_0)$ gleich Null.

Gl. (1.130). Durch Einsetzen in Gl. (1.133) ergibt sich

$$v_{e,krist} = \frac{1}{\hbar}\frac{dW}{dk} = \frac{1}{\hbar}\frac{d}{dk}\left(W_{V,C} + \frac{\Delta k^2}{2}\frac{d^2W(k)}{dk^2}\bigg|_{k_0}\right) \tag{1.136}$$

$$= \frac{1}{\hbar}\frac{d^2W(k)}{dk^2}\bigg|_{k_0}\frac{1}{2}\underbrace{\frac{d\Delta k^2}{dk}}_{=\frac{d}{dk}(k-k_0)^2=2\Delta k} \tag{1.137}$$

$$v_{e,krist} = \frac{1}{\hbar}\frac{d^2W(k)}{dk^2}\bigg|_{k_0}\Delta k\ . \tag{1.138}$$

Legen wir den Ursprung der Wellenvektorskala an die Stelle des jeweiligen Minimums/Maximums des Leitungs- oder Valenzbandes, dann ist $k_0 = 0$ (z. B. ist dann $k_{0,C} = 0$ bei Betrachtung des Leitungsbandes in Abb. 1.42). Damit wird aus Gl. (1.138)

$$v_{e,krist} = \frac{1}{\hbar}\frac{d^2W(k)}{dk^2}\bigg|_{k_0}k \;=\; \frac{\hbar k}{m_e^*}\ . \tag{1.139}$$

Auf der rechten Seite wird durch Vergleich mit der Geschwindigkeit eines Elektrons im freien Elektronengas (Gl. (1.135)) die effektive Masse

$$\frac{1}{m_e^*} := \frac{1}{\hbar^2}\frac{d^2W(k)}{dk^2}\bigg|_{k_0} \tag{1.140}$$

eines Elektrons im Kristall definiert. Die effektive Masse m_e^* beschreibt das Verhalten von Elektronen in den als parabelförmig angenommenen Extremwerten eines beliebigen Bandverlaufs so, als wären es freie Elektronen mit der Masse m_e^*. Die so definierte effektive Masse ist unabhängig von k, da für die Approximation des Bandverlaufs eine Parabel verwendet wurde, deren zweite Ableitung eine Konstante ergibt.
Aus Gl. (1.140) lassen sich einige wichtige Eigenschaften der effektiven Masse herleiten:

1. Da die zweite Ableitung einer Kurve ein Maß für deren Krümmung ist, besagt Gl. (1.140), dass die effektive Masse umso kleiner ist, je stärker die Dispersionskurve $W(k)$ gekrümmt ist. An der Bandkante verläuft die Kurve flach mit einer horizontalen Tangente, wodurch die effektive Masse der Elektronen unendlich wird.

2. Minima haben eine positive zweite Ableitung, Maxima einen negativen Wert. Daher gilt zwischen der effektiven Masse eines Elektrons an

der Bandkante des Leitungsbandes und des Valenzbandes bei gleichem Betrag der Krümmung

$$m^*_{e,C} = -m^*_{e,V} \quad . \tag{1.141}$$

3. Aus der Definitionsgleichung der effektiven Masse in Gl. (1.139) ergibt sich unmittelbar der Impuls

$$m^*_e \mathrm{v}_{e,krist} = \hbar k \tag{1.142}$$

eines Elektrons im Kristall an den quadratisch genäherten Verläufen von Leitungsband und Valenzband. Er wird zur Unterscheidung zu dem eines freien Elektrons mit „Quasi-Impuls“ oder Kristallimpuls bezeichnet.

4. Für ein freies Elektron mit der (effektiven) Masse m^*_e brauchen die inneren Kräfte des Kristalls nicht berücksichtigt werden, da ihre Wirkung in der effektiven Masse erfasst ist. Daher können direkt die Beziehungen der Newtonschen Mechanik angewendet werden. So gilt z. B. bei einem von außen an den Kristall angelegten elektrischen Feld $\vec{E}$ für die Kraft auf ein Elektron

$$\vec{F} = -e\vec{E} \; . \tag{1.143}$$

Dadurch erfährt das Elektron eine Beschleunigung $\frac{d\vec{\mathrm{v}}_e}{dt}$ entsprechend

$$\vec{F} = m^*_e \frac{d\vec{\mathrm{v}}_e}{dt} = -e\vec{E} \; . \tag{1.144}$$

Gl. (1.144) beschreibt eine Bewegungsgleichung für Elektronen im Kristall aufgrund eines z. B. von außen angelegtes Feldes.

Wir haben bisher eine beliebige, aber bestimmte Richtung von k im Zweidimensionalen betrachtet. Dabei sind die Richtung der Geschwindigkeit und die Richtung des Wellenvektors identisch.
Im Dreidimensionalen erfolgt die Richtungsableitung durch Bildung des Gradienten im $\vec{k}$-Raum und aus Gl. (1.133) wird

$$\vec{v_e} = \frac{1}{\hbar} \mathrm{grad}_{\vec{k}} W(\vec{k}) = \frac{1}{\hbar} \sum_{i=1}^{3} \frac{\partial W}{\partial k_i} \vec{e_i} \; . \tag{1.145}$$

Die Vektoren $\vec{e}_i$ sind die Einheitsvektoren in den drei Koordinatenrichtungen von $\vec{k}$. Die Beschleunigung, die ein Elektron erfährt, ist damit für die i-te Komponente

$$\frac{d\vec{v}_{e,i}}{dt} = \frac{1}{\hbar}\frac{d}{dt}\frac{\partial W}{\partial k_i} = \frac{1}{\hbar}\sum_{j=1}^{3}\frac{\partial^2 W}{\partial k_i \partial k_j}\frac{dk_j}{dt} \tag{1.146}$$

Mit

$$\frac{d\vec{k}}{dt} = \frac{1}{\hbar}\frac{d\vec{p}}{dt} = \frac{\vec{F}}{\hbar} \tag{1.147}$$

ergibt sich für jede Komponente

$$\frac{d\vec{v}_{e,i}}{dt} = \left(\frac{1}{\hbar}\right)^2 \sum_{j=1}^{3}\frac{\partial^2 W}{\partial k_i \partial k_j}F \; . \tag{1.148}$$

Die drei Gleichungen für die drei Komponenten von $\frac{d\vec{\mathrm{v}}_e}{dt}$ zusammengefasst ergibt

$$\frac{d\vec{\mathrm{v}}_e}{dt} = \frac{1}{[m_e^*(\vec{k})]}\vec{F} \tag{1.149}$$

mit dem Tensor

$$\frac{1}{[m_e^*(\vec{k})]} = \frac{1}{\hbar^2}\begin{pmatrix} \frac{\partial^2 W}{\partial k_x^2} & \frac{\partial^2 W}{\partial k_x \partial k_y} & \frac{\partial^2 W}{\partial k_x \partial k_z} \\ \frac{\partial^2 W}{\partial k_y \partial k_x} & \frac{\partial^2 W}{\partial k_y^2} & \frac{\partial^2 W}{\partial k_y \partial k_z} \\ \frac{\partial^2 W}{\partial k_z \partial k_x} & \frac{\partial^2 W}{\partial k_z \partial k_y} & \frac{\partial^2 W}{\partial k_z^2} \end{pmatrix} \; . \tag{1.150}$$

Die effektive Masse des Elektrons im Dreidimensionalen ist also ein Tensor. Dies hat zur Folge, dass z. B. in Gl. (1.149) die Beschleunigung eines Elektrons im Halbleiterkristall nicht in Richtung einer äußeren Kraft $\vec{F}$ erfolgen muss. Für bestimmte Werte von $\vec{k}$ entartet der effektive Masse-Tensor zum Skalar.

Da der Tensor in Gl. (1.150) symmetrisch ist, lässt sich eine Hauptachsentransformation durchführen, wodurch nur die Elemente der Hauptdiagonalen ungleich Null sind. Durch Inversion ergibt sich daraus der Tensor der effektiven Masse

$$[m_e^*] = \begin{bmatrix} m_{e,1}^* & 0 & 0 \\ 0 & m_{e,2}^* & 0 \\ 0 & 0 & m_{e,3}^* \end{bmatrix} \; . \tag{1.151}$$

Aus diesem Tensor lässt sich für makroskopische Betrachtungen ein Skalar bilden. Das Bildungsgesetz, also die Kombination und Wichtung der einzelnen Komponenten des Tensors hängt von der betrachteten Größe ab.

Für Leitfähigkeitsuntersuchungen ist wegen $\vec{J} = -en\vec{v}_e$ eine skalare Masse zu bestimmen, die zu einer guten Näherung für den Mittelwert der Geschwindigkeit $\vec{v}_e$ der Elektronen führt. Wir nennen diese Masse Leitfähigkeitsmasse m^*_{ec}. Sie ist in Tabelle 1.3 für wichtige Halbleitermaterialien angegeben.
Für die Bestimmung der Zustandsdichte ist zu berücksichtigen, dass sich anstelle der Kugel des freien Elektronengases für Flächen gleicher Energie die in Abb. 1.43 gezeigten Rotationsellipsoide bilden. Daher muss die effektive Masse zur Berechnung der Zustandsdichte so gewählt werden, dass die sich

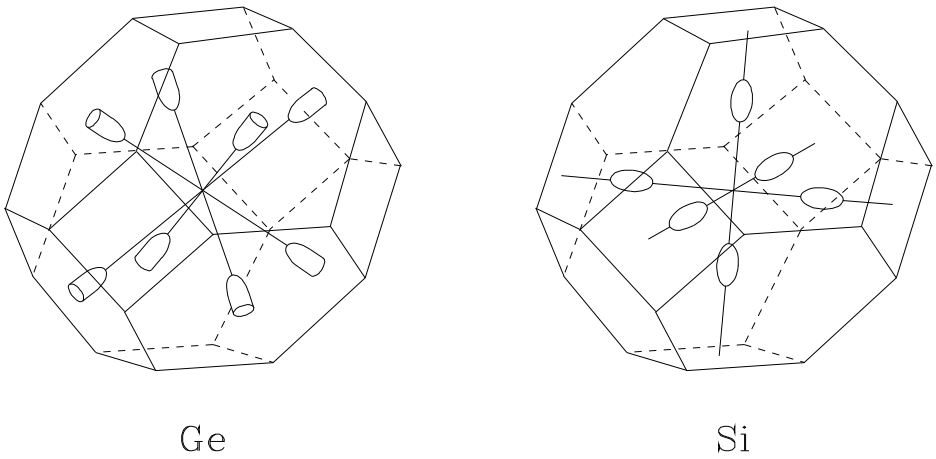

Abb. 1.43: Flächen gleicher Energie in der ersten Brillouin-Zone für **Ge** und **Si**. Es handelt sich bei den Flächen um Rotations-Ellipsoide, die bei **Ge** nur jeweils zur Hälfte in der ersten Brillouin-Zone liegen.

damit ergebende Zustandsdichte bei Berechnung als freies Elektronengas mit der der Ellipsoiden übereinstimmt. Der hieraus berechnete effektive Masse-Skalar heißt Zustandsdichte-Masse m^*_{ed}. Sie ist ebenfalls in Tab. 1.3 angegeben.

Bei den Berechnungen der effektiven Massen von **Ge**, **Si**, **GaAs** ist auch die Entartung des Valenzbandes zu berücksichtigen. In Abb. 1.40 ist zu sehen, dass die an der Bandoberkante überlappenden Bänder unterschiedliche Krümmungen und damit auch unterschiedliche Massen aufweisen.
Im nächsten Kapitel werden wir den Begriff „Löcher“ für die an der Oberkante des Valenzbandes zur Stromleitung beitragenden Ladungsträger einführen. Bei Löchern in einem Valenzband mit geringer Krümmung spricht man daher von „schweren Löchern“. Im Band mit der starken Krümmung tragen „leichte Löcher“ zur Stromleitung bei.

1.30 Elektronen und Löcher

Halbleiter zeichnen sich dadurch aus, dass für $T = 0$ das Valenzband voll besetzt und das Leitungsband leer ist. Bereits in Kap. 1.15 wurde festgestellt, dass ein voll besetztes Band keinen Strom leiten kann, da alle Energieniveaus besetzt sind. Die gleiche Aussage gewinnt man aus der Betrachtung aller $\vec{k}$-Vektoren in einer Brillouin-Zone. Sind alle Zustände besetzt, gibt es zu jedem Elektron mit einer Wellenfunktion $\psi(\vec{k})$ aufgrund der geometrischen Symmetrie auch eines mit $\psi(-\vec{k})$. D. h. zu jedem Elektron existiert ein Partner, dessen Richtung genau entgegengesetzt ist, so dass im Mittel kein Stromfluss zustande kommt.

Das Ziel bei der Realisierung von Halbleiterbauelementen besteht darin, Elektronen aus dem Valenzband zu entfernen, um so einen Stromfluss zu ermöglichen. Wie sie entfernt werden und wohin, wird später noch ausführlich diskutiert.

Durch das Fehlen von Elektronen im Valenzband sind freie Zustände vorhanden. Die restlichen Elektronen des Valenzbandes können dann an einem Stromfluss teilnehmen, indem sie durch Aufnahme von Energie durch ein von „außen“ an den Halbleiter angelegtes elektrisches Feld diese Plätze besetzen. Die Stromdichte im Halbleiter aufgrund dieser Elektronen im Valenzband beträgt

$$\vec{J}_{VB} = \frac{1}{V} \sum_{\substack{\text{besetzte} \\ \text{Zustände}}} (-e)\vec{\mathrm{v}}_e \,, \tag{1.152}$$

worin V das Volumen des Halbleiterkristalls ist, $-e$ die Ladung eines Elektrons und $\vec{\mathrm{v}}_e$ die mittlere Geschwindigkeit in dem jeweiligen Zustand ist.
Diese Formulierung ist insofern unpraktisch, da wir über alle besetzten Zustände summieren müssen, wogegen wir zur Ermöglichung der Stromleitung unbesetzte Zustände geschaffen haben. Es fällt uns daher in der Regel leichter, Aussagen über die Anzahl der von uns geschaffenen unbesetzten Zustände zu machen. Dafür kann Gl. (1.152) umformuliert werden, indem die besetzten Zustände als Differenz aller Zustände minus nicht besetzte Zustände ausgedrückt wird.

$$\vec{J}_{VB} = \frac{1}{V} \sum_{\substack{\text{alle} \\ \text{Zustände}}} (-e)\vec{\mathrm{v}}_e \; - \; \frac{1}{V} \sum_{\substack{\text{unbesetzte} \\ \text{Zustände}}} (-e)\vec{\mathrm{v}}_e \tag{1.153}$$

Der erste Term kann nicht zum Stromfluss beitragen, da ein volles Band keinen Strom leitet. Daher vereinfacht sich Gl. (1.153) zu

$$\vec{J}_{VB} = -\frac{1}{V} \sum_{\substack{\text{unbesetzte} \\ \text{Zustände}}} (-e)\vec{\mathrm{v}}_e \; = \; \frac{1}{V} \sum_{\substack{\text{unbesetzte} \\ \text{Zustände}}} (+e)\vec{\mathrm{v}}_e \; . \tag{1.154}$$

Diese Beziehung kann so interpretiert werden, dass der Stromfluss in einem teilbesetzten Band durch positiv geladene Teilchen ($+e$) entsteht, die auf den (von Elektronen) unbesetzten Zuständen sitzen. Wir nennen diese Teilchen Löcher[12] und werden sie im Folgenden immer dann verwenden, wenn wir im physikalischen Sinn (in der Realität) fehlende Elektronen meinen.
Nach Gl. (1.154) besitzt ein Loch die gleiche Geschwindigkeit wie ein Elektron, das dem zugehörigen unbesetzten Zustand zugeordnet ist. Dies bedeutet, dass sich das Loch im Ortsraum genauso bewegt, wie ein Elektron, das sich auf dem Zustand des Loches befindet.
Wir können daher für die Bewegungsgleichung eines Lochs mit Gl. (1.144) schreiben

$$m_h \cdot \frac{d\vec{\mathrm{v}}_e}{dt} = +e\vec{E} \; . \tag{1.155}$$

Darin haben wir mit m_h formal eine Löcher-Masse eingeführt. Für ein Elektron in diesem Zustand gilt entsprechend

$$m_e \cdot \frac{d\vec{\mathrm{v}}_e}{dt} = -e\vec{E} \; , \tag{1.156}$$

woraus durch Vergleich direkt folgt, dass

$$m_h = -m_e \tag{1.157}$$

für die Masse eines Lochs gelten muss.

Die effektiven Massen von **Si**, **Ge** und **GaAs** sind in Tab. 1.3 bezogen auf die Masse eines freien Elektrons angegeben.

[12] Häufig wird auch der Begriff des „Defektelektrons“ verwendet.

	Ge	**Si**	**GaAs**
Zustandsdichte-Masse			
Elektron $\frac{m^*_{ed}}{m_e}$	0,55	1,08	0,067
Loch $\frac{m^*_{hd}}{m_e}$	0,37	0,811	0,45
Leitfähigkeits-Masse			
Elektron $\frac{m^*_{ec}}{m_e}$	0,12	0,26	0,067
Loch $\frac{m^*_{hc}}{m_e}$	0,21	0,386	0,34

Tabelle 1.3: Effektive Massen für Elektronen und Löcher in **Ge**, **Si**, **GaAs**.

2 Ladungsträger in Valenz- und Leitungsband

2.1 Fermi-Dirac-Verteilungsfunktion bei Eigenleitung

Bei Halbleitern ist für $T = 0$ das Valenzband voll besetzt und das Leitungsband leer. Im Unterschied zum Isolator ist jedoch die Energielücke zwischen den beiden Bändern so klein, dass z. B. durch Zufuhr von thermischer Energie ($T > 0$) einige Elektronen genügend Energie aufnehmen, um in das Leitungsband zu gelangen. Thermische Energie führt auch dazu, dass die Ionenrümpfe der Gitteratome in Schwingung geraten. Ähnlich wie bei den Schwingungszuständen der Elektronen, sind auch die Energien der Schwingungszustände des Kristallgitters quantisiert. Je nachdem, ob benachbarte Gitterebenen in gleicher oder entgegengesetzter Richtung schwingen, spricht man von akustischer oder optischer Gitterschwingung[13]. Die verschiedenen Schwingungen sind durch ihre Kreisfrequenz gekennzeichnet und bilden daher ein Ensemble unterscheidbarer Elemente. Für die Energie der Gitterschwingung gilt daher die Boltzmann-Verteilung (Besetzungswahrscheinlichkeit für nicht interagierende Teilchen bei geringer Dichte, z. B. Gase)

$$f_{MB}(W) \sim e^{\frac{-W}{kT}} \; . \tag{2.1}$$

Sie gibt die mittlere Anzahl von Gitterschwingungen mit einer Energie W bei einer (absoluten) Temperatur T an. Darin ist $k = 1,381 \cdot 10^{-23} \frac{J}{K} = 8,617 \cdot 10^{-5} \frac{eV}{K}$ die Boltzmannkonstante. Anders als bei der Besetzung von Zuständen durch Elektronen nach dem Pauli-Prinzip, existiert für die Besetzung der Zustände der Gitterschwingung kein Exklusivitätsprinzip.
Um Konfusion zu vermeiden, sei angemerkt, dass jedem Schwingungszustand auch formal ein Teilchen mit der Energie $\hbar\,\omega_p$ (ω_p ist die Kreisfrequenz der Gitterschwingung) zugeordnet werden kann, das dann der Bose-(Einstein-)Verteilungsfunktion folgt. Dieses Teilchen nennt man Phonon. Es lässt sich zeigen, dass ein Phonon mit anderen Teilchen in Wechselwirkung tritt, als hätte es einen Impuls

$$\vec{p} = \hbar\,\vec{q}\,, \tag{2.2}$$

[13]Der Begriff akustische Schwingung rührt daher, dass die Massen benachbarter Gitterebenen, wie bei akustischen Wellen in gleicher Richtung schwingen. Optisch wurde als Begriff gewählt, da durch die entgegengesetzte Schwingung starke elektrische Dipolmomente auftreten, die sich im optischen Verhalten des Kristalls bemerkbar machen.

wobei $\vec{q}$ der Wellenvektor der zugeordneten Gitterschwingung ist.

Wir wollen im Folgenden die Wahrscheinlichkeit berechnen, dass ein Energieniveau W von einem Elektron besetzt wird. Dazu betrachten wir Stöße zwischen einem Elektron und dem Gitter. Sind W_e^a und W_e^b die Energien des Elektrons vor und nach dem Stoß, sowie W_p^a und W_p^b die Energie der Gitterschwingung vor und nach dem Stoß, so gilt nach dem Energieerhaltungssatz

$$W_e^a + W_p^a = W_e^b + W_p^b \,. \tag{2.3}$$

$f(W)$ sei die Wahrscheinlichkeit, dass ein Energieniveau mit der Energie W durch ein Elektron besetzt ist. Dann ist $1 - f(W)$ die Wahrscheinlichkeit, dass für ein Elektron das Niveau der Energie W noch frei ist.
Zur Vereinfachung nehmen wir an, dass keine Entartung vorliegt. In diesem Fall ist jedem Zustand ein Energieniveau zugeordnet, das durch zwei Elektronen entgegengesetzter Spins besetzt werden kann.
Die nachfolgenden Überlegungen gelten jedoch entsprechend auch bei Entartung und führen auf das gleiche Ergebnis, was zur Übung einmal überprüft werden sollte.
Die Wahrscheinlichkeit, dass ein Stoß mit dem Übergang der Energien $(W_e^a, W_p^a) \rightarrow (W_e^b, W_p^b)$ stattfindet, ist

$$\alpha = 2\, f(W_e^a)\, 2\,(1 - f(W_e^b))\, f_{MB}(W_p^a) \,, \tag{2.4}$$

also das Produkt aus der Wahrscheinlichkeit, dass das Elektron eine Energie mit dem Niveau W_e^a vor dem Stoß besitzt, der Wahrscheinlichkeit, dass das Energieniveau W_e^b für das Elektron nach dem Stoß noch frei ist (d. h. besetzt werden kann) und der Wahrscheinlichkeit, dass die Gitterschwingung die Energie W_p^a vor dem Stoß besitzt (vgl. Gl. (2.1)). Die Wahrscheinlichkeit, dass das Energieniveau W_p^b der Phononenschwingung noch frei ist, ist 100%, da die Energieniveaus der Boltzmannverteilung durch beliebig viele Teilchen besetzt werden können. Der Faktor 2 resultiert jeweils aus der Tatsache, dass jedes Energieniveau durch zwei Elektronen mit entgegengesetztem Spin besetzt werden kann, und sich dadurch die Wahrscheinlichkeit, ein Elektron mit der Energie W zu haben bzw. ein freies Niveau der Energie W vorzufinden, verdoppelt.
Im thermodynamischen Gleichgewicht ist ein Übergang in entgegengesetzter

Richtung $(W_e^b, W_p^b) \rightarrow (W_e^a, W_p^a)$ gleich wahrscheinlich $(= \alpha)$. Daher gilt

$$f(W_e^a)\,(1 - f(W_e^b))\,f_{MB}(W_p^a) = f(W_e^b)\,(1 - f(W_e^a))\,f_{MB}(W_p^b)\,. \qquad (2.5)$$

(Die Faktoren 2 kürzen sich dabei heraus.)
Durch Einsetzen von Gl. (2.1) für $f_{MB}(W_p^a)$ bzw. $f_{MB}(W_p^b)$ und z. B. Substitution von W_p^b durch Gl. (2.3) ergibt sich aus Gl. (2.4)

$$\frac{f(W_e^a)}{1 - f(W_e^a)}\,e^{\frac{W_e^a}{kT}} = \frac{f(W_e^b)}{1 - f(W_e^b)}\,e^{\frac{W_e^b}{kT}}\,. \qquad (2.6)$$

Da die linke Seite nur von der Energie W_e^a vor dem Stoß und die rechte Seite nur von der Energie W_e^b nach dem Stoß abhängt, müssen beide Seiten gleich derselben Konstanten sein. Wir wählen als Konstante den Ausdruck $e^{\frac{W_F}{kT}}$ und schreiben für die Energie der Elektronenzustände vor und nach dem Stoß allgemein W. Damit ergibt sich aus Gl. (2.6)

$$f(W) = \frac{1}{1 + e^{\frac{W - W_F}{kT}}}\,. \qquad (2.7)$$

Das ist die Fermi-Dirac-Verteilungsfunktion. Sie berücksichtigt gegenüber der Maxwell-Bolzmann- und Bose-Einstein-Verteilungsfunktion das Pauli-Prinzip, nach dem ein Elektron einen Elektronenzustand nur dann annehmen kann, wenn er noch nicht besetzt ist. Die von uns gewählte Konstante W_F heißt Fermi-Energie. Das zugehörige Energieniveau z. B. bei einer Darstellung im Bändermodell wird Fermi-Niveau genannt. Aus Gl. (2.7) ist unmittelbar zu sehen, dass für $W = W_F$ die Besetzungswahrscheinlichkeit gleich $\frac{1}{2}$ ist. Dies gilt unabhängig von der Temperatur.
Abb. 2.1 zeigt den Verlauf der Verteilungsfunktion für $T = 0$ und eine Temperatur $0 < T \ll \frac{W_F}{k}$.

Für $T \rightarrow 0$ sind also alle Energieniveaus unterhalb des Fermi-Niveaus besetzt und oberhalb unbesetzt. Dies entspricht genau der bereits in Kap. 1.22 für das freie Elektronengas definierten Fermi-Energie. Dort waren alle Zustände im Innern der Fermikugel besetzt, und alle außerhalb unbesetzt.
Wir kennen jetzt anhand der Fermi-Dirac-Verteilungsfunktion die Gesetzmäßigkeit, mit der bei steigender Temperatur auch Energiezustände außerhalb der Fermikugel angenommen werden.
Dabei ist zu beachten, dass wir bei der Herleitung zwar die Fermi-Energie formal eingeführt, aber noch nicht bestimmt haben. Erst nach Bestimmung der

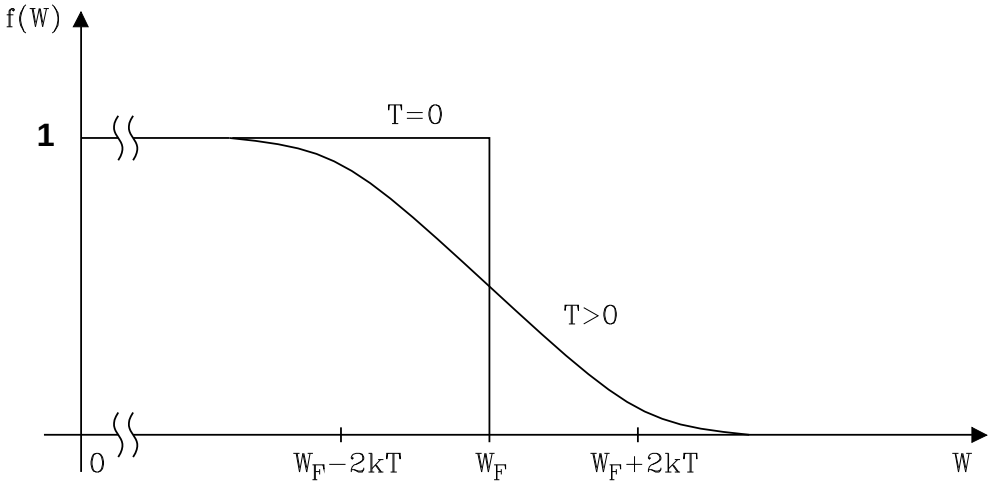

Abb. 2.1: Verlauf der Fermi-Dirac-Verteilungskunktion für $T = 0$ und $0 < T \ll \frac{W_F}{k}$.

Lage des Fermi-Niveaus lässt sich eine Ladungsträgerdichte in den Bändern angeben.
Für Abschätzungen ist es hilfreich, einen Wert jeweils an den Enden der Verteilungsfunktion zu kennen. Es gilt:

$$f(W_F - 2kT) = 0,88 \tag{2.8}$$

$$f(W_F + 2kT) = 0,12\,. \tag{2.9}$$

Für Energien weit oberhalb der Fermi-Energie $W - W_F \gg kT$ ist der Exponentialterm im Nenner der Fermi-Dirac-Verteilungsfunktion groß gegen eins und es ergibt sich die Näherung

$$f(W) \approx e^{-\frac{W-W_F}{kT}}\,, \tag{2.10}$$

die wir im Folgenden immer verwenden, wenn wir $f(W)$ als Näherung schreiben. Das ist der bekannte Ausdruck der Boltzmann-Verteilung nach Gl. (2.1). Dies lässt sich dadurch erklären, dass bei den in der Näherung beschriebenen sehr kleinen Besetzungswahrscheinlichkeiten das Pauli-Prinzip eine nur vernachlässigbare Auswirkung hat.
Wir bezeichnen Halbleiter, für die die Boltzmann-Verteilungsfunktion als Näherung der Fermi-Dirac-Verteilungsfunktion verwendet werden kann, als nicht degeneriert oder nicht entartet. Abb. 2.2 zeigt den Vergleich zwischen Fermi-Dirac-und Boltzmann-Verteilung im interessierenden Bereich für

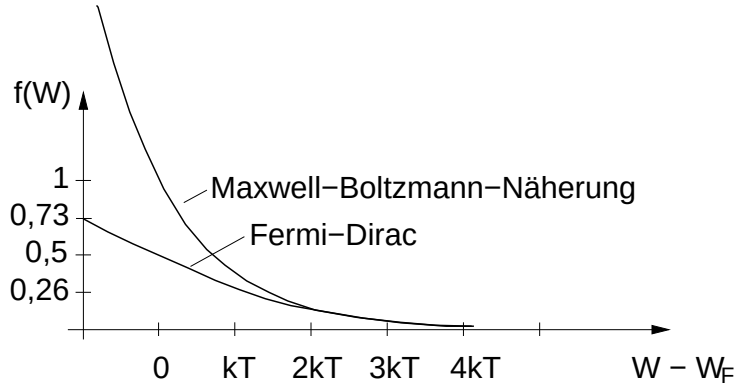

Abb. 2.2: Besetzungswahrscheinlichkeit nach der Fermi-Dirac-Verteilungsfunktion und der Maxwell-Boltzmann-Näherung.

$W - W_F > 0$.
Beachtet werden muss, dass die Näherung nur ab einem Bereich $W - W_F > 2kT$ gilt. Für kleinere Werte ergibt sich ein Verlauf der Näherung der unsinnige Ergebnisse liefert, da die Wahrscheinlichkeit niemals größer als 100 % sein kann.
Wir wollen im Sinne einer guten Näherung für nicht entartete Halbleiter fordern, dass $W - W_F \geq 3kT$ gilt. Das bedeutet, dass die Fermi-Energie mindestens um $3kT$ von den Bandkanten (W_C, W_V) entfernt in der Bandlücke liegt.

2.2 Ladungsträgerdichten

2.2.1 Allgemeine Betrachtung am Beispiel der Eigenleitung

Bisher wurde ein reiner Halbleiter betrachtet, bei dem zur Stromleitung ein Elektron von dem Valenz- in das Leitungsband gelangen muss. Dabei hinterlässt es im Valenzband ein Loch. Elektronen im Leitungsband und Löcher im Valenzband treten in einem reinen Halbleiter immer als Paar, also in gleicher Anzahl, auf. Beide tragen zur Stromleitung bei. Die Elektronen im Leitungsband, da sie dort (beliebig) viele freie Zustände besetzen können und so in einem elektrischen Feld Energie aufnehmen können. Die Löcher im Valenzband, weil sie für die Elektronen des Valenzbandes noch freie besetzbare Zustände darstellen, die die Elektronen durch Aufnahme oder Abgabe von Energie annehmen können. Bei der Besetzung eines Lochs durch ein Elektron hinterlässt das Elektron wiederum ein Loch, das von einem nach-

folgenden Elektron eingenommen werden kann usw.
Der Vorgang der Elektron-Loch-Bildung heißt Paarbildung oder Generation. Der umgekehrte Vorgang heißt Rekombination. Das durch z. B. thermische Anregung entstandene Elektron-Loch-Paar rekombiniert in einer Zeitspanne, die Lebensdauer, die vom Halbleitermaterial und dem Rekombinationsmechanismus abhängt. Hiermit werden wir uns in einem der folgenden Kapitel genauer befassen.
Die Anzahl der vorhandenen Zustände in Valenz- und Leitungsband in Abhängigkeit von der Energie W erhalten wir über die Näherung der Zustandsdichte des freien Elektronengases mit der (effektiven) Zustandsdichte-Masse der Elektronen des Halbleiters. Die Fermi-Dirac-Verteilung nach Gl. (2.7) gibt Auskunft darüber, ob bei einer Temperatur T die vorhandenen Zustände bei der Energie W besetzt sind.

2.2.2 Berechnung von Ladungsträgerdichten allgemein

Die auf das Volumen des Halbleiters bezogene spezifische Anzahl von Elektronen $dn(W)$ in den Elektronenzuständen $\frac{dN_{EZ}(W)}{V}$ in einem Energieintervall $W \ldots W + dW$ ergibt sich aus der Zustandsdichte $D_C(W)$ und der Besetzungswahrscheinlichkeit dieser Zustände $f(W)$ nach der Fermi-Dirac-Verteilungsfunktion

$$dn(W) = f(W)\frac{dN_{EZ}(W)}{V} = D_C(W)\, f(W)\, dW\ . \tag{2.11}$$

Da wir im Folgenden stets mit der spezifischen Anzahl, also der auf das Volumen bezogenen Anzahl von Elektronen (und Löchern) arbeiten, enthält Gl. (2.11) die Definition der Elektronendichte $dn(W)$. Für die Löcherdichte gilt entsprechend $dp(W)$. Wir werden im Folgenden immer Kleinbuchstaben zur Bezeichnung von auf das Volumen bezogenen Dichten verwenden.
Die gesamte Dichte der Ladungsträger in einem Band erhält man durch Integration von Gl. (2.11) über das gesamte Band. Bei der Betrachtung von Löchern im Valenzband gilt entsprechend

$$dp(W) = D_V(W)\,(1 - f(W))\, dW\ . \tag{2.12}$$

Der Faktor $1 - f(W)$ ist ersichtlich, wenn man berücksichtigt, dass Löcher nicht besetzte Zustände im Valenzband repräsentieren. Daher ist die Wahrscheinlichkeit für ein Loch bei der Energie W gleich der Wahrscheinlichkeit, dass der Zustand bei dieser Energie nicht besetzt ist.

2.2.3 Ladungsträgerdichte in Leitungs- und Valenzband

Die Dichte der Elektronen im gesamten Leitungsband ergibt sich durch Integration von Gl. (2.11) von der Unterkante bis zur Oberkante des Leitungsbandes. Für die Zustandsdichte $D_C(W)$ wird als Näherung die Zustandsdichte des freien Elektronengases nach Gl. (1.86) mit der Zustandsdichte-Masse des Halbleiters für Elektronen verwendet:

$$D_C(W) = \frac{8\pi\sqrt{2}}{h^3}\, {m^*_{ed}}^{\frac{3}{2}}\, \sqrt{W - W_C}\,,\; W \geq W_C\;. \tag{2.13}$$

W_C ist die Energie an der Unterkante des Leitungsbandes, unterhalb der es keine Zustände gibt. Der Ausdruck $W - W_C$ verschiebt daher die freie Elektronen-Parabel aus Gl. (1.71) mit der Zustandsdichte nach Gl. (1.86) auf das Energieniveau W_C an der Unterkante des Leitungsbandes.
Wir können D_C aus Gl. (2.13) und die Fermi-Dirac-Verteilungsfunktion nach Gl. (2.7) in Gl. (2.11) einsetzen und erhalten durch Integration über das gesamte Leitungsband (LB) die Dichte der darin enthaltenen Ladungsträger

$$n_0 \;=\; \int\limits_{\text{untere LB-Kante}}^{\text{obere LB-Kante}} dn(W) \;=\; \int\limits_{W_C}^{\infty} D_C(W)\, f(W)\, dW\;. \tag{2.14}$$

W_C ist die Energie an der Unterkante des Leitungsbandes. Bei der Obergrenze kann aufgrund des starken Abfalls von $f(W)$ anstelle der (unbekannten) Oberkante des Leitungsbandes in guter Näherung bis ∞ integriert werden. Der Index an n_0 zeigt uns im Folgenden immer an, dass es sich um die Ladungsträgerdichte im thermodynamischen Gleichgewicht handelt. Wir setzen die Beziehungen für Zustandsdichte und Verteilungsfunktion in Gl. (2.14) ein und erhalten

$$n_0 \;=\; \int\limits_{W_C}^{\infty} \frac{8\pi\sqrt{2}}{h^3}\, {m^*_{ed}}^{\frac{3}{2}}\, \sqrt{W - W_C}\, \frac{1}{1 + e^{\frac{W - W_F}{kT}}}\, dW\;. \tag{2.15}$$

Abb. 2.3 veranschaulicht die Zusammensetzung der Ladungsträgerdichte n_0 aus Zustandsdichte und Fermi-Dirac-Verteilungsfunktion. Ganz analog erhalten wir mit der Zustandsdichte der Löcher im Valenzband mit der Näherung des freien Elektronengases mit der Zustandsdichte-Masse der Löcher im Halbleiter

$$D_V(W) \;=\; \frac{8\pi\sqrt{2}}{h^3}\, {m^*_{hd}}^{\frac{3}{2}}\, \sqrt{W_V - W} \qquad ;\, W \;\leq\; W_V \tag{2.16}$$

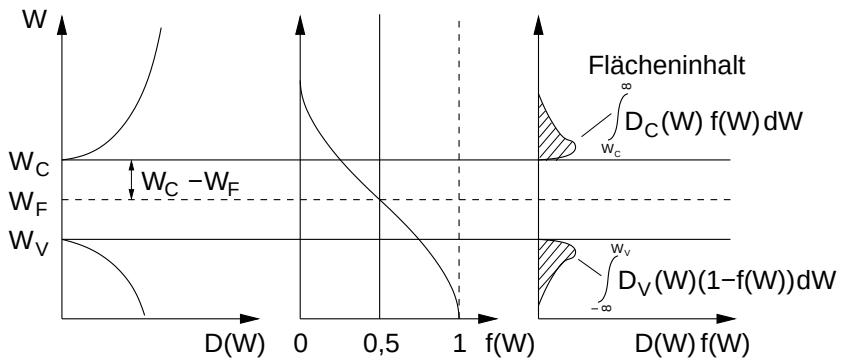

Abb. 2.3: Zustandsdichte (links), Fermi-Dirac-Verteilungsfunktion (Mitte) und Ladungsträger-Dichte (rechts) als Flächeninhalt des Produktes der beiden Verläufe.

die Dichte der Löcher im Valenzband

$$p_0 = \int_{-\infty}^{W_V} D_V(W)\,(1 - f(W))\,dW \tag{2.17}$$

$$p_0 = \int_{-\infty}^{W_V} \frac{8\pi\sqrt{2}}{h^3}\, m_{hd}^{*\,\frac{3}{2}}\, \sqrt{W_V - W}\, \frac{1}{1 + e^{\frac{W_F - W}{kT}}}\, dW\ . \tag{2.18}$$

Beide Integrale Gl. (2.15) und Gl. (2.18) sind analytisch nicht lösbar. Wir können jedoch für die darin enthaltene Fermi-Dirac-Verteilungsfunktion die Boltzmann-Näherung nach Gl. (2.10) einsetzen, wenn $|W_F - W| > 2\,kT$ ist (vgl. Abb. 2.2). Diese Näherung ist für die Eigenleitung der hier betrachteten Halbleitermaterialien erfüllt. Die Überprüfung erfolgt in einem Beispiel im Kapitel Eigenleitungsdichte.

Mit Näherung der Boltzmann-Verteilungsfunktion wird aus der Elektronendichte im Leitungsband nach Gl. (2.15)

$$n_0 = \int_{W_C}^{\infty} \frac{8\pi\sqrt{2}}{h^3} m_{ed}^{*\,\frac{3}{2}} \sqrt{W - W_C}\ e^{-\frac{W - W_F}{kT}}\, dW \tag{2.19}$$

und durch Substitution $\frac{W - W_C}{kT} = x$

$$n_0 = \frac{8\pi\sqrt{2}}{h^3}\, m_{ed}^{*\,\frac{3}{2}}\, (kT)^{\frac{3}{2}}\, e^{-\frac{W_C - W_F}{kT}} \int_0^\infty \sqrt{x}\, e^{-x} dx \tag{2.20}$$

$$= \frac{8\pi\sqrt{2}}{h^3} (m_{ed}^*\, kT)^{\frac{3}{2}}\, e^{-\frac{W_C - W_F}{kT}}\, \frac{\sqrt{\pi}}{2} \tag{2.21}$$

$$= 2\left(\frac{2\pi m_{ed}^*\, kT}{h^2}\right)^{\frac{3}{2}}\, e^{-\frac{W_C - W_F}{kT}} \tag{2.22}$$

$$n_0 = N_C\, e^{-\frac{W_C - W_F}{kT}} = N_C f(W_C)\ . \tag{2.23}$$

Darin wird

$$N_C = 2\left(\frac{2\pi\, m_{ed}^*\, kT}{h^2}\right)^{\frac{3}{2}} \tag{2.24}$$

als die effektive Zustandsdichte im Leitungsband bezeichnet. Sie ist die Dichte, die an der Leitungsbandkante vorhanden wäre, wenn alle Niveaus dort konzentriert wären. Die Löcherdichte im Valenzband ergibt sich entsprechend durch Integration über die möglichen Zustände (Zustandsdichte) im Valenzband $D_V(W)$ mal deren Besetzungswahrscheinlichkeit $1 - f(W)$ nach Gl. (2.12). In einem Rechengang analog zu Gl. (2.14) bis (2.19) ergibt sich

$$p_0 = \int_{-\infty}^{W_V} \frac{8\pi\sqrt{2}}{h^3}\, m_{hd}^{*\,\frac{3}{2}}\, \sqrt{W_V - W}\ \ e^{-\frac{W_F - W}{kT}} dW\ , \tag{2.25}$$

und nach Rechnung analog zu Gl. (2.20) bis Gl. (2.23)

$$p_0 = N_V\, e^{-\frac{W_F - W_V}{kT}} = N_V(1 - f(W_V)) \tag{2.26}$$

mit der effektiven Zustandsdichte im Valenzband

$$N_V = 2\left(\frac{2\pi\, m_{hd}^*\, kT}{h^2}\right)^{\frac{3}{2}}\ . \tag{2.27}$$

Beispiel: Bestimmung der effektiven Zustandsdichten des Leitungs- und Valenzbandes für **Ge**, **Si** und **GaAs** bei Raumtemperatur T=300 K.
Für das Leitungsband gilt Gl. (2.24)

$$N_C = 2\left(\frac{2\pi\, m_{ed}^* kT}{h^2}\right)^{\frac{3}{2}} \tag{2.28}$$

$$= 2\left(\frac{2\pi \cdot 9,1 \cdot 10^{-31} \cdot 1,38 \cdot 10^{-23} \cdot 300\,\mathrm{kg\,J}}{(6,626 \cdot 10^{-34})^2\,\mathrm{J^2 s^2}}\right)^{\frac{3}{2}} \left(\frac{m_{ed}^*}{m_e}\right)^{\frac{3}{2}} \tag{2.29}$$

$$= 2,5 \cdot 10^{25}\,\mathrm{m^{-3}} \left(\frac{m_{ed}^*}{m_e}\right)^{\frac{3}{2}} \tag{2.30}$$

$$= 2,5 \cdot 10^{19}\,\mathrm{cm^{-3}} \left(\frac{m_e^*}{m_e}\right)^{\frac{3}{2}} \quad . \tag{2.31}$$

Das Ergebnis lässt sich direkt zur Berechnung von N_V anwenden, wenn m_{ed}^* durch m_{hd}^* ausgetauscht wird. Mit den effektiven Zustandsdichte-Massen aus Tabelle (1.3) ergeben sich die effektiven Zustandsdichten in der nachfolgenden Tabelle 2.1.

	Ge	**Si**	**GaAs**
$N_C\,\mathrm{cm^3} \cdot 10^{-19}$	1,02	2,81	0,0435
$N_V\,\mathrm{cm^3} \cdot 10^{-19}$	0,564	1,83	0,757

Tab. 2.1: Effektive Zustandsdichten gebräuchlicher Halbleitermaterialien bei Raumtemperatur T=300 K.

Aus der Herleitung ergibt sich, dass die effektiven Zustandsdichten N_C, N_V Materialkonstanten sind, die jedoch von der Temperatur abhängen. Solange die Boltzmann-Näherung der Fermi-Dirac-Verteilungsfunktion gilt, hängt die Ladungsträgerdichte in Leitungs- und Valenzband (Gl. (2.23) und Gl. (2.26)) außer von der Temperatur nur von dem Abstand der Fermi-Energien von den Bandkanten ab ($W_C - W_F$ und $W_F - W_V$). Diese Aussage gilt immer, unabhängig davon, ob ein eigenleitender oder dotierter Halbleiter betrachtet wird. Dies ist eine sehr wichtige Erkenntnis.
Wir fassen also nochmals zusammen: Da W_C und W_V Materialkonstanten sind, hängt die Ladungsträgerdichte in Leitungs- und Valenzband nur (von der Temperatur und) von der Lage des Fermi-Niveaus ab. Dieses ist für

eigenleitende und dotierte Halbleiter unterschiedlich und hängt von der Dotierung und immer von der Temperatur ab.
Die Lage des Fermi-Niveaus bestimmt also die Ladungsträgerdichten im Leitungs- und Valenzband. Aufgrund dieser außerordentlich wichtigen Bedeutung beschäftigen wir uns in den folgenden Kapiteln näher mit der Bestimmung des Fermi-Niveaus.

2.3 Eigenleitungsdichte

Aufgrund der Elektron-Loch-Paarbildung müssen bei eigenleitenden (nicht dotierten) Halbleitern die Dichte der Elektronen im Leitungsband n_0 und die Dichte der Löcher im Valenzband gleich sein. Es gilt also

$$n_0 \;=\; p_0 \;=\; n_i \tag{2.32}$$

mit n_i als die Eigenleitungsdichte des Halbleitermaterials, (Index i steht für intrinsic). Wir können die Eigenleitungsdichte einfach ermitteln, indem wir für n_0 und p_0 die Bestimmungsgleichungen (2.23) und (2.26) verwenden und das Produkt bilden

$$n_0\, p_0 \;=\; n_i^2 \;=\; N_C\, N_V\, e^{-\frac{W_C - W_F}{k\,T}}\, e^{-\frac{W_F - W_V}{k\,T}} \tag{2.33}$$

$$n_i^2 \;=\; N_C\, N_V\, e^{-\frac{W_g}{k\,T}} \tag{2.34}$$

$$n_i \;=\; \sqrt{N_C\, N_V}\, e^{-\frac{W_g}{2kT}}\;. \tag{2.35}$$

Diese Beziehung gilt entsprechend der Herleitung von n_0 und p_0 im Bereich der Maxwell-Boltzmann-Näherung der Fermi-Dirac-Verteilungsfunktion. Die Eigenleitungsdichte eines Halbleitermaterials wird also zum einen bestimmt durch die effektiven Zustandsdichten N_C, N_V die als materialabhängigen Parameter die Zustandsdichte-Massen für den jeweiligen Halbleiter enthalten. Zum anderen besteht eine sehr starke exponentielle Abhängigkeit von dem Bandabstand des jeweiligen Materials. Bemerkenswert ist, dass die Eigenleitungsdichte nicht von der Fermi-Energie abhängt. Sie ist eine Materialkonstante.
Zu beachten ist die starke Temperaturabhängigkeit durch den Exponential-Term, die über die Temperaturabhängigkeit von $N_C(T)$ und $N_V(T)$ dominiert.

Zur Übung sollte einmal $n_i(T)$ für **Si**, **Ge** und **GaAs** bei T=(200,

300, 400, 500) K berechnet werden. Die Ergebnisse der Berechnung mit den vereinfachten Annahmen in diesem Skript ergeben bei Raumtemperatur die, in Tabelle 2.2 dargestellten Ergebnisse.

	Ge	**Si**	**GaAs**	
$n_i\,\mathrm{cm}^3$	$2,02 \cdot 10^{13}$	$8,72 \cdot 10^{9}$	$2,03 \cdot 10^{6}$	berechnet (Näherung)
$n_i\,\mathrm{cm}^3$	$2,8 \cdot 10^{13}$	$1,0 \cdot 10^{10}$	$2,0 \cdot 10^{6}$	Messung (genau)

Tabelle 2.2: Eigenleitungsdichte n_i für verschiedene Halbleitermaterialien bei T=300 K. Die erste Zeile gibt die, anhand der Näherung in diesem Skript berechneten Werte an. Genauere Meßwerte sind in der zweiten Zeile angegeben.

Zur Beurteilung der Temperaturabhängigkeit kann die Proportionalität

$$n_i^2(T) \;\sim\; T^3\, e^{-\frac{W_g}{kT}} \tag{2.36}$$

herangezogen werden, die sich unmittelbar durch Einsetzen von Gl. (2.24) und (2.27) in (2.34) ergibt.

2.4 Fermi-Niveau bei Eigenleitung

Wir können anhand Gl. (2.32) auch die Lage des Fermi-Niveaus W_i bei Eigenleitung bestimmen. Einsetzen der Ladungsträgerdichten aus Gl. (2.23) und (2.26) für $W_F = W_i$ liefert

$$N_C\, e^{-\frac{W_C - W_i}{kT}} \;=\; N_V\, e^{-\frac{W_i - W_V}{kT}} \;. \tag{2.37}$$

Durch Umstellen nach der Fermi-Energie ergibt sich

$$W_i \;=\; \frac{W_C \,+\, W_V}{2} \;+\; \frac{1}{2}\,kT\,\ln\left(\frac{N_V}{N_C}\right) \;. \tag{2.38}$$

Das Fermi-Niveau bei Eigenleitung liegt also mit einer Abweichung von $\frac{1}{2}\,kT\,\ln\left(\frac{N_V}{N_C}\right)$ außerhalb der Mitte zwischen Leitungs- und Valenzbandkante. Zur Übung sollten Sie diese Abweichung einmal berechnen. Sie ist sehr gering, so dass näherungsweise für den eigenleitenden Halbleiter gilt

$$W_i \;\approx\; \frac{W_C \,+\, W_V}{2} \;, \tag{2.39}$$

d. h. das Fermi-Niveau liegt wie in der Abb. 2.4 in der Mitte zwischen der Leitungs- und Valenzbandkante.

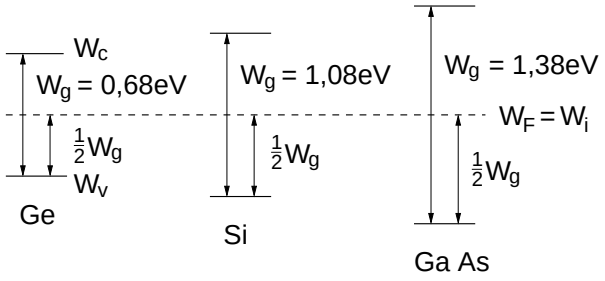

Abb. 2.4: Lage des Fermi-Niveaus W_i bei Eigenleitung in der Mitte der Bandlücke bei **Ge**-, **Si**- und **GaAs**-Halbleitern.

2.5 Dotierte Halbleiter

Eine gezielte Beeinflussung der elektrischen Eigenschaften eines Halbleiters ist durch den Einbau von Fremdatomen in das Kristallgitter möglich. Diesen Vorgang nennt man Dotierung. Ziel der Dotierung ist, dass freie Ladungsträger in Leitungs- oder Valenzband in einer definierten Menge enstehen. Dadurch lässt sich die Stromleitung entscheidend gegenüber dem undotierten Halbleiter verändern, wodurch die Realisierung der heute bekannten Halbleiterbauelemente möglich wird.
Zur Dotierung werden Fremdatome eines Elements aus der V. Gruppe (5 Valenzelektronen) wie z. B. $\mathbf{P}^{15}$, $\mathbf{As}^{33}$, $\mathbf{Sb}^{51}$ oder aus der III. Gruppe (3 Valenzelektronen) wie z. B. $\mathbf{B}^{5}$, $\mathbf{Al}^{13}$, $\mathbf{Ga}^{31}$, $\mathbf{In}^{49}$ verwendet.

Bei dem Einbau in das Kristallgitter, wie in Abb. 2.5 gezeigt, werden vier Elektronen für die kovalente Bindung mit den vier nächsten Nachbarn im Kristallgitter des Ausgangsmaterials (Wirtsgitter) benötigt. Hierdurch ergibt sich die geringstmögliche Störung im Kristall. Bei der Dotierung mit fünfwertigen Atomen besitzt das fünfte Elektron keinen Bindungspartner und besitzt daher auch nur eine sehr geringe Bindung an das ortsfeste Dotierungsatom. Bereits durch Zufuhr einer sehr kleinen Energie kann das Elektron vom Atomrumpf getrennt werden und steht als freies Elektron im Leitungsband zur Verfügung. Anders als bei der Elektron-Loch-Paarbildung ensteht hierdurch aber kein Loch im Valenzband. Durch Dotierung mit fünfwertigen Atomen werden also durch Ionisierung der Dotierungsatome „nur“ freie Elektronen gewonnen. Aus diesem Grund nennt man diese Dotierungsatome auch Donatoren(lat. dotare = ausstatten). Wichtig im Zusammenhang mit der Dotierung ist zu verstehen, dass sich an der elektri-

schen Neutralität (Summe positiver und negativer Ladungen) nichts ändert. Zu der negativen Ladung, auf Grund der freien Elekronen im Leitungsband, gehört immer die entgegengesetzte, gleichgroße Ladung der ionisierten Dotierungsatome. Im Gegensatz zu den im Halbleiter frei beweglichen

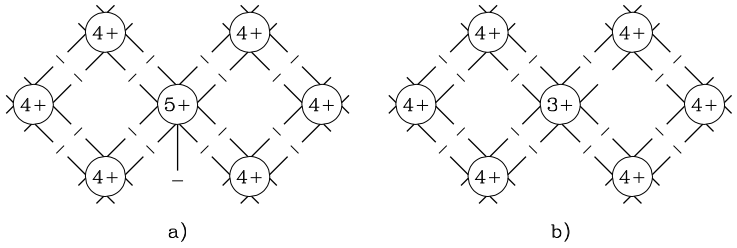

Abb. 2.5: Einbau von Fremdatomen in das Kristallgitter eines Halbleiters (Dotierung). Die Linien symbolisieren überlappende Orbitale bzw. Bindungskräfte. Links: Einbau eines fünfwertigen Donator-Atoms. Rechts: Dotierung mit dreiwertigen Akzeptor-Atom.

Elektronen sind diese jedoch ortsfest im Kristallgitter eingebunden, so dass es durch eine Verschiebung der Elektronen lokal im Halbleiter unterschiedlich geladene Bereiche gibt. Nach außen (der Kristall als Ganzes) bleibt der Halbleiter jedoch ladungsneutral.

Bei Dotierung mit dreiwertigen Atomen gilt Entsprechendes. Hier fehlt durch den Einbau in das Kristallgitter jeweils ein Elektron, das zur kovalenten Bindung benötigt wird. Da wir fehlende Elektronen als Löcher beschreiben können, werden durch den Einbau von Akzeptor-Atomen Löcher in den Halbleiter eingebaut. Die Besetzung eines Lochs in der kovalenten Bindung ist ein energetisch günstiger Zustand (abgeschlossene Schale) so dass die bei Zimmertemperatur vorhandene Energie ausreicht, um das dicht über der Valenzband-Kante liegende Loch mit einem Elektron aus dem Valenzband zu besetzen.

Abb. 2.6 zeigt die Donator- und Akzeptorniveaus W_D, W_A in einem dotierten Halbleiter. Das Donator-Energieniveau W_D ist vor der Ionisierung mit Elektronen gefüllt. Die Ionisierung bewirkt, dass das Donator-Niveau unbesetzt ist wodurch jeweils ein Elektron im Leitungsband und ein positiv geladenes Donator-Ion entstehen.
Das Akzeptor-Energieniveau W_A ist vor der Ionisierung unbesetzt. Die

Ionisierung des Akzeptors entspricht der Besetzung eines Akzeptor-Niveaus durch ein Elektron aus dem gefüllten Valenzband. Oder in der Löcher-Terminologie: Das Akzeptor-Niveau ist vor der Ionisierung mit Löchern gefüllt. Die Ionisierung bewirkt, dass das Akzeptor-Niveau unbesetzt ist, wodurch jeweils ein Loch im Valenzband und ein negativ geladenes Akzeptor-Ion entstehen.

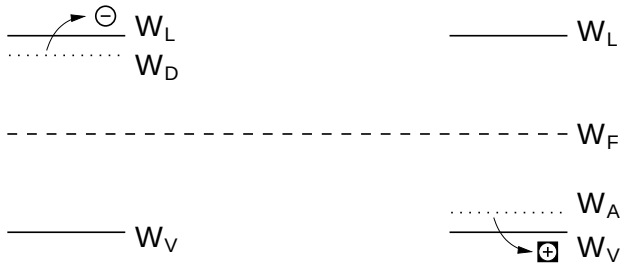

Abb. 2.6: Donator- (links) und Akzeptor- (rechts) Energieniveaus. Der Pfeil deutet die Ionisierung eines Dotierungs-Atoms an, wodurch ein Elektron in das Leitungsband bzw ein Loch, in das Valenzband gelangen.

Wir können mit den Grundlagen der vorangegangenen Kapitel bereits genauere Aussagen über die Lage der Dotierungsniveaus und die zur Ionisierung notwendige Energie machen.
Dazu betrachten wir in Abb. 2.7, links ein einfaches Modell entsprechend Abb. 2.6, für den Einbau eines Donator-Atoms in ein Kristallgitter. Die Punkte darin deuten die Elektronen in der kovalenten Bindung an.
In diesem einfachen Modell bewegt sich das ungebundene Elektron des Donator-Atoms in dem Potentialfeld seines einfach positiv geladenen Atomrumpfes, ähnlich dem Elektron eines Waserstoffatoms. Eine sehr ähnliche Modellvorstellung hatten wir bereits in Kap 1.8 für die Anwendung der Ergebnisse des Wasserstoffatoms auf allgemeine Atomaufbauten verwendet. Im Unterschied liegt hier das Atom in einem Kristall eingebettet. Die Potentiale der Kristallatome schirmen das Potentialfeld des Atomrupfes stark ab. Sie wirken als Dielektrikum. Die Coulomb-Wechselwirkung wird also durch den Halbleiterkristall abgeschwächt. Wir können diese Wirkung mit Hilfe der Dielektrizitätskonstante des Halbleitermaterials berücksichtigen. Für **Si** beträgt sie $\varepsilon_r = 11,7 \approx 12$.

Damit lässt sich ein einfaches Modell zur Abschätzung der Ionisie-

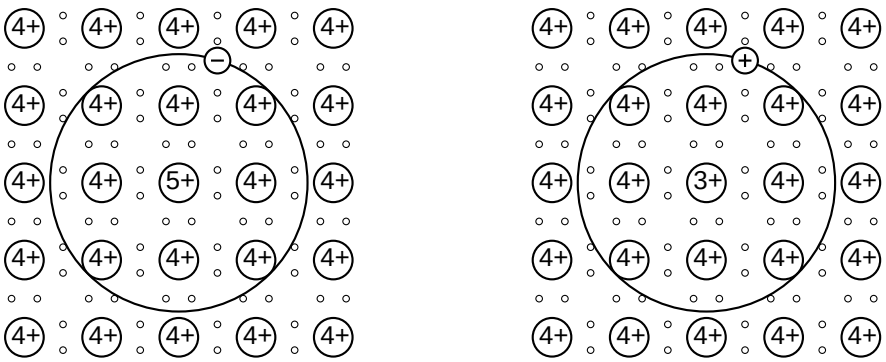

Abb. 2.7: Darstellung des ungebundenen Elektrons bzw. Lochs bei einer Dotierung mit Donatoren bzw Akzeptoren. Aufgrund der geringen Bindung bewegt sich die ungebundene Ladung ähnlich wie bei einem einzelnen Atom auf einem Radius um den Atomkern. Der dargestellte Radius ist nicht maßstabsgerecht und in der Regel viel größer.

rungsernergie W_D eines Donators aufstellen. Wir behandeln das ionisierte Donator-Atom im Halbleiterkristall wie ein einzelnes Wasserstoffatom. Der Kristall wird durch die Dielektrizitätskonstante des Halbleitermaterials und Leitfähigkeitsmasse des Elektrons berücksichtigt. Wir können mit dieser Modifikation direkt Gl. (1.19) für die Energien des Wasserstoffelektrons verwenden und erhalten für die Energien des Donator-Elektrons

$$W_n \;=\; \frac{-m^*_{el}\, e^4}{8\,(\varepsilon_0\,\varepsilon_r\,h)^2}\,\frac{1}{n^2} \qquad n = 1, 2, 3, \ldots\,. \tag{2.40}$$

n ist die Hauptquantenzahl. Für n=1 liefert Gl. (2.40) die Ionisierungsenergie des Donator-Elektrons.

Beispiel: Wir berechnen die Ionisierung eines Donator-Atoms in einem **Si**-Halbleiterkristall. Für **Si** gilt nach Tab. 1.3 $\dfrac{m^*_{ec}}{m_e} = 0,26$ und $\varepsilon_r \approx 12$. Damit ist die Ionisierungsenergie

$$W_1 \;=\; \frac{0,26 \cdot 9,1 \cdot 10^{-31}\,\mathrm{kg}\,(1\,\frac{\mathrm{eV}}{1\mathrm{V}})^4}{8(8,9 \cdot 10^{-12}\,\frac{\mathrm{As}}{\mathrm{Vm}} \cdot 12 \cdot 4,14 \cdot 10^{-15}\,\mathrm{eVs})^2} \;\approx\; 24\,\mathrm{meV}$$

Das Beispiel zeigt, dass das Donator-Niveau sehr nahe an der Leitungskante liegt. Für **Si** beträgt die Energiedifferenz für Elektronen vom Donator-Niveau in das Leitungsband nur $\dfrac{1}{45}$ des Bandabstandes.

Die Zustände für $n > 1$ nach Gl. (2.40) beschreiben die Energieniveaus in energetisch angeregten Zuständen. Sie liegen zwischen dem Niveau des Grundzustandes und der Leitungsbandkante. Sie folgen mit größer werdendem n immer dichter aufeinander und gehen für große n quasi kontinuierlich in das Leitungsband über.
Dass es sich bei dem Donator-Elektron quasi um ein freies Elektron handelt, kann auch anhand des großen Radius seiner Kreisbahn im Verhältnis zum Abstand der Kristallatome vermutet werden. Berechnen Sie doch zur Übung einmal den Radius für **Si** (Hinweis: Gl. (1.19), **Si** kristallisiert in Diamant-Struktur (fcc) mit der Gitterkonstanten a=5,43). Durch die großen Radien überlappen die Wellenfunktionen der Donatorzustände und es ergibt sich eine weitere Aufspaltung der Energieniveaus zu einem Donator-Energieband.

Die gleichen Überlegungen gelten auch für die Dotierung mit Akzeptoren nach Abb. 2.7 rechts, wenn man die Betrachtungsweise mit den Löchern Ladungsträger anwendet. Hier umkreist ein positiv geladenes Loch den negativ geladenen Atomrumpf des Akzeptors. Zur Ionisierung des Loches muss dieses in das Valenzband gesenkt werden. Dies ist gleichbedeutend mit einer Anhebung eines Valenzband-Elektrons auf das Akzeptorniveau. Die dafür nötige Ionisierungsenergie lässt sich wiederum mit Gl. (1.19) bzw. (2.40) abschätzen, wenn anstelle der Leitfähigkeitsmasse des Elektrons die eines Lochs eingesetzt wird. Ein Vergleich der Massen zeigt i.e. die gleiche Größenordnung für die Ionisierungsernergie der Akzeptorniveaus.

Die Wirkung einer Dotierung kann verringert werden, indem gleichzeitig Donatoren und Akzeptoren eingebaut werden. Dann können z. B. Elektronen von Donator-Niveaus Bindungslücken an Akzeptoratomen ausfüllen und die beiden Dotierungen kompensieren sich. Eine Störstellenleitung aufgrund der Dotierung ist nur dann möglich, wenn die Konzentration von Donatoren und Akzeptoren unterschiedlich ist.
Überwiegt die Leitung aufgrund der, durch eine hohe Donatorkonzentration eingebrachten Elektronen, so spricht man von einem n-(dotierten)Halbleiter oder von n-Leitung. Elektronen sind dann in der Überzahl und man nennt sie auch Majoritäten oder Majoritätsträger. Hingegen sind Löcher in der Minderzahl und stellen die Minoritäten oder Minoritätsträger dar.

Entsprechendes gilt bei überwiegender Akzeptor-Dotierung. Hier liegt ein p-Halbleiter mit p-Leitung vor, bei dem Löcher die Majoritätsträger und Elektronen die Minoritäten sind.

2.6 Fermi-Dirac-Verteilungsfunktion bei Dotierung

Wir wollen die Verteilungsfunktion für Ladungsträger aus Dotierungen bestimmen. Prinzipiell gelten dafür die gleichen Überlegungen wie für die Eigenleitung, jedoch mit einem Unterschied: An der Eigenleitung waren nur Leitungs- und Valenzband beteiligt. In beiden Bändern kann jeder Zustand zwei Elektronen mit entgegengesetztem Spin aufnehmen bzw. abgeben. Die Wahrscheinlichkeit, ein Elektron oder ein Loch mit einer bestimmten Energie zu finden bzw. ein freies Energie-Niveau mit einer bestimmten Energie zu finden, ist daher doppelt so groß wie für den Fall nur einfach besetzbarer Zustände. Dieser Fall liegt jedoch bei den Dotierungs-Niveaus vor.
Wir betrachten zur Erläuterung ein ionisiertes Donator-Niveau. Dieses kann prinzipiell zwei Elektronen mit antiparalleler Spinrichtung aufnehmen, bietet also zwei Besetzungsmöglichkeiten. Fragt man, ob ein Donator-Niveau W_e^a noch frei ist, so beträgt die Wahrscheinlichkeit dafür $2(1 - f(W_e^a))$. Dabei ist wie zuvor bei der Eigenleitung $f(W_e^a)$ die Wahrscheinlichkeit, dass ein Energie-Niveau der Energie W_e^a mit einem Elektron besetzt ist.
Ist hingegen das Donator-Niveau mit einem Elektron besetzt, kann kein weiteres Elektron auf diesem Niveau aufgenommen werden, da das Donator-Ion bereits durch die Aufnahme des einen Elektrons neutralisiert wurde. Es ist daher kein Zustand mehr vorhanden. Daher beträgt die Wahrscheinlichkeit, ein Elektron auf einem Donator-Niveau der Energie W_e^a zu finden, nur $f(W_e^a)$. Wie zuvor bei der Eigenleitung formulieren wir die Wahrscheinlichkeit, dass ein Stoß zwischen Elektron und Gitter mit dem Übergang der Energien $(W_e^a, W_p^a) \to (W_e^b, W_p^b)$ stattfindet. Dabei berücksichtigen wir, wie erläutert, die einfache Besetzung des Donator-Niveaus

$$\alpha = f(W_e^a) f_{MB}(W_p^a) 2(1 - f(W_e^b)) \,, \tag{2.41}$$

wobei W_e^a für ein Donator-Energie-Niveau und W_e^b für ein Leitungsband-Niveau steht.
Im thermodynamischen Gleichgewicht erfolgt der Übergang von Leitungsband auf das Donator-Niveau mit gleicher Wahrscheinlichkeit. Unter Berücksichtigung der zweifachen Besetzungsmöglichkeiten des Donator-Niveaus gilt

dann

$$\alpha = 2\, f(W_e^b) f_{MB}(W_p^b)\, 2\,(1 - f(W_e^a))\ . \tag{2.42}$$

Der Energieerhaltungssatz nach Gl. (2.3) gilt weiterhin.
Die gleiche Rechnung wie bei der Besetzungswahrscheinlichkeit bei Eigenleitung (Gleichsetzen von Gl. (2.41) und Gl. (2.42), Substitution von W_p^a oder W_p^b, Separation der Variablen W_e^a, W_e^b und Gleichsetzen mit der Konstanten $e^{\frac{W_F}{kT}}$) liefert die Besetzungswahrscheinlichkeit für ein Donator-Niveau der Energie W_D

$$f(W_D) = \frac{1}{1 + \frac{1}{2} e^{\frac{W_D - W_F}{kT}}}\ . \tag{2.43}$$

Darin ist wieder die Fermi-Energie W_F die noch zu bestimmende Größe.
Ein ionisiertes Akzeptor-Niveau W_e^a bietet dagegen nur eine Besetzungsmöglichkeit, da der zweite Zustand schon durch ein Elektron einer Spinrichtung besetzt ist. Bei der Ionisierung sind dafür jedoch zwei mögliche Elektronen auf dem gleichen Niveau vorhanden. Es herrschen also die umgekehrten Verhältnisse bei den Wahrscheinlichkeiten wie bei den Donator-Niveaus.
Eine entsprechende Rechnung liefert die Besetzungswahrscheinlichkeit für ein Akzeptor-Niveau

$$f(W_A) = \frac{1}{1 + 2 e^{\frac{W_A - W_F}{kT}}}\ . \tag{2.44}$$

Zur Übung sollte die Berechnung von Gl. (2.43) und (2.44) nachvollzogen werden.
Für praktische Überlegungen ist es einfacher, anstelle Gl. (2.43) und (2.44) weiter mit der Fermi-Dirac-Verteilungsfunktion ohne den Faktor vor dem Exponentialterm zu arbeiten. Dazu kann der Faktor in den Exponenten gezogen werden und es ergibt sich für die Besetzungswahrscheinlichkeiten des Donator- bzw. Akzeptor-Niveaus

$$f(W_D^*) = \frac{1}{1 + e^{\frac{W_D^* - W_F}{kT}}}, \qquad W_D^* = W_D - kT \ln 2 \tag{2.45}$$

$$f(W_A^*) = \frac{1}{1 + e^{\frac{W_A^* - W_F}{kT}}}, \qquad W_A^* = W_A + kT \ln 2\ . \tag{2.46}$$

Damit haben wir wieder die Standardform der Fermi-Dirac-Verteilungsfunktion nach Gl. (2.7), wie sie auch für die Berechnung bei Eigenleitung verwendet wird. Es sind lediglich die effektiven Donator-

bzw. Akzeptor-Niveaus W_D^* und W_A^* zu verwenden. W_D^* liegt im Bänderdiagramm etwas unterhalb W_D, W_A^* etwas oberhalb W_A.
Zu beachten ist, dass durch die Dotierung das Fermi-Niveau in der Nähe der effektiven Energie-Niveaus der Dotierung liegen kann. In diesem Fall ist $W_D^* - W_F$ bzw. $W_A^* - W_F$ nicht mehr groß gegen kT und es muss die Fermi-Dirac-Verteilungsfunktion nach Gl. (2.45) bzw. (2.46) verwendet werden.
Mit Hilfe der Besetzungswahrscheinlichkeiten nach Gl. (2.45) und (2.46) können die Verhältnisse in einem dotierten Halbleiter anhand des Bänderdiagramms einfach dargestellt werden. Abb. 2.8 zeigt als Beispiel einen n-dotierten Halbleiter.

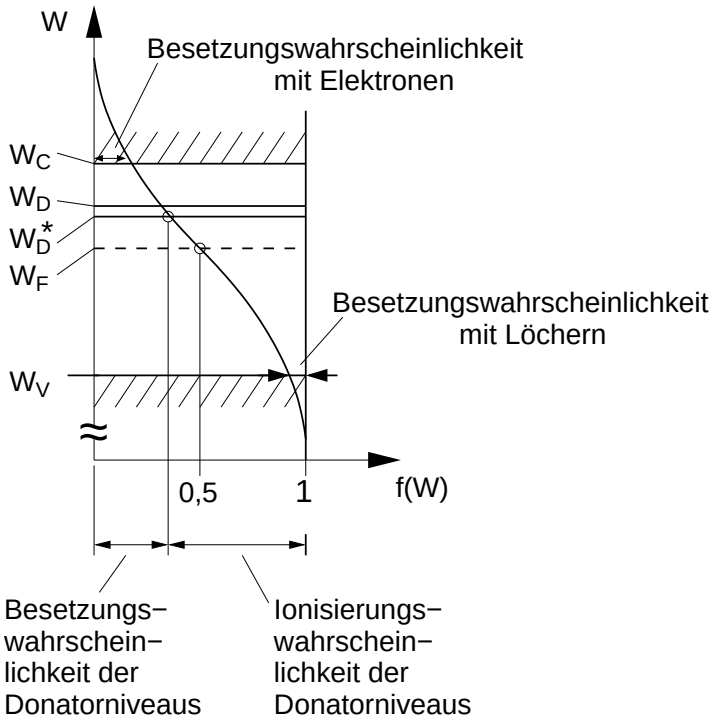

Abb. 2.8: Darstellung der Besetzungswahrscheinlichkeiten von Leitungs-, Valenz- und Donator-Niveau anhand eines Bändermodells.

Deutlich zu sehen ist die Abhängigkeit aller Wahrscheinlichkeiten von der Lage des Fermi-Niveaus.

2.7 Ladungsträgerdichten von Dotierungen

Ist der Halbleiter dotiert, so geben wir die Dichte der für die Dotierung verwandten Akzeptoren mit N_A und die Dichte der Donatoren mit N_D an. Wir hatten diese Volumendichten zur besseren Unterscheidung auch als spezifische Anzahl bezeichnet.
Zur Vereinfachung nehmen wir an, dass alle Donator-Energie-Niveaus innerhalb eines infinitesimalen Energiebereichs dW angesiedelt sind. Das gleiche nehmen wir für die Akzeptor-Niveaus an. Es existiert dann eine spezifische Anzahl (Dichte) von N_A bzw. N_D Zuständen innerhalb eines Bereichs dW.
Die Zustandsdichte auf den beiden Dotierungs-Niveaus ist also

$$D_A(W) = \frac{N_A}{dW} \quad \text{für} \quad W = W_A ... W_A + dW \tag{2.47}$$

und

$$D_D(W) = \frac{N_D}{dW} \quad \text{für} \quad W = W_D ... W_D + dW \; . \tag{2.48}$$

Wir wollen wissen, wieviele Donatoren ionisiert sind. Die Dichte der ionisierten Donatoren N_D^+ ist identisch mit der Dichte der von den Donatoren an das Leitungsband abgegebenen Elektronen. Die Wahrscheinlichkeit, dass ein Donatorzustand ionisiert, also nicht besetzt ist, ist nach Gl. (2.45) $(1 - f(W_D^*))$ $(= const.)$. Die Dichte der ionisierten Donatoren lässt sich allgemein über das Integral von Zustandsdichte und Besetzungswahrscheinlichkeit bestimmen

$$N_D^+ = \int_{W_D}^{W_D + dW} D_D(W)(1 - f(W_D^*))dW = N_D(1 - f(W_D^*)) \; . \tag{2.49}$$

Wir fragen nach der Dichte N_A^- der ionisierten Akzeptoren. Ein Akzeptor ist dann ionisiert, wenn sich ein Elektron auf dem Akzeptor-Niveau befindet, es also besetzt ist. Mit der Besetzungswahrscheinlichkeit $f(W_A^*)$ $(= const.)$ ergibt sich für die Dichte der ionisierten Akzeptoren

$$N_A^- = \int_{W_A}^{W_A + dW} D_A(W) f(W_A^*) dW = N_A f(W_A^*) \; . \tag{2.50}$$

Es ergeben sich also ganz ähnliche Ausdrücke der Form „effektive Zustandsdichte (N_A, N_D, N_C, N_V) mal Wahrscheinlichkeit“ wie schon bei der allgemeinen Berechnung der Ladungsträgerdichten in Leitungs- und Valenzband (vgl.

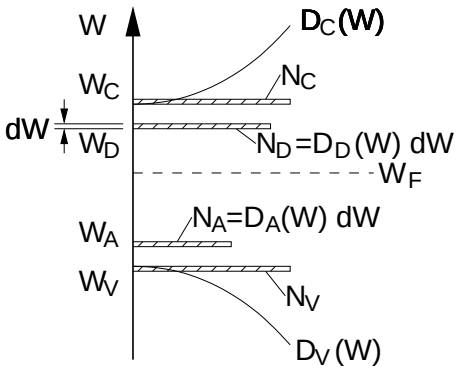

Abb. 2.9: Zustandsdichten und Ladungsträgerdichten in Leitungs- und Valenzband sowie für Donator- und Akzeptor-Niveau.

Gl. (2.23) und (2.26)). Dies war zu erwarten, da durch die Einführung der effektiven Zustandsdichten in Gl. (2.23) und (2.26) sämtliche in den Bändern vorhandene Niveaus an den Bandkanten konzentriert wurden. Dies ist in Abb. 2.9 veranschaulicht.

2.8 Das Massenwirkungsgesetz

In Kapitel 2.2.3 wurden mit Gl. (2.23) und (2.26) zwei allgemeingültige Beziehungen zur Ermittlung der Ladungsträgerdichten n_0 und p_0 in Leitungs- und Valenzband hergeleitet. Sie gelten für alle Arten von Halbleitern, unabhängig davon, ob n-, p- oder i-(eigen-)leitend. Die Unterscheidung des jeweiligen Leitungsmechanismus erfolgt einzig über die davon bestimmte Lage des Fermi-Niveaus.
Multipliziert man die beiden Gleichungen miteinander, so ergibt sich, wie bereits in Gl. (2.33) bei der Ermittlung der Eigenleitungsdichte geschehen, das Massenwirkungsgesetz

$$n_0\, p_0 = n_i^2\ . \tag{2.51}$$

Es ist entsprechend der vorangegangenen Herleitung allgemeingültig. Dabei muss jedoch der Gültigkeitsbereich der verwendeten Boltzmann-Näherung anstelle der Fermi-Dirac-Verteilungsfunktion berücksichtigt werden (vgl. Gl. (2.10)).
Das Massenwirkungsgesetz besagt, dass das Produkt aus den Ladungsträgerdichten aus Valenz- und Leitungsband konstant ($= n_i^2(T)$) ist. Es ist damit unabhängig von der Lage des Fermi-Niveaus und damit von der Dotierung und wird nur durch die Temperatur (vgl. Gl. (2.36)) bestimmt.

2.9 Neutralitätsbedingung

Eine wichtige Bedingung, die u. a. zur Berechnung der Fermi-Energie benötigt wird, ist die Neutralitätsbedingung. Sie fordert, dass der Halbleiter in jedem Ort (lokal) elektrisch neutral ist, solange er sich im thermodynamischen Gleichgewicht befindet.
Physikalisch bedeutet dies, dass die Raumladungsdichte an jedem Ort des Halbleiters gleich Null sein muss. Wäre dies nicht so, würde daraus ein elektrisches Feld resultieren, durch das die Ladungen wieder so verschoben werden, dass sich ein Gleichgewichtszustand (Neutralität) einstellt. Aus dieser Überlegung wird auch deutlich, dass die Neutralitätsbedingung nur im thermodynamischen Gleichgewicht, also ohne von außen an den Halbleiter angelegte Spannung gilt.
Die Ladungen im Halbleiter bestehen aus den frei beweglichen Elektronen im Leitungsband mit der Ladungsdichte $-e\,n_0$, frei beweglichen Löchern der Dichte $e\,p_0$ und den Dichten der ortsfest im Gitter eingebauten ionisierten Akzeptoren $-e\,N_A^-$ und Donatoren $e\,N_D^+$. Hierbei haben wir vorausgesetzt, dass die Dotierungsatome aus der dritten bzw. fünften Gruppe stammen und daher durch Abgabe bzw. Aufnahme eines Elektrons ionisieren. Zum Verständnis sei darauf hingewiesen, dass sich die Dichten der frei beweglichen Ladungsträger n_0, p_0 jeweils aus der Dichte der Ladungsträger durch Eigenleitung und den Ladungsträgerdichten der ionisierten Dotierungsatome zusammensetzt. Die Bedingung für Ladungsneutralität lautet somit

$$\rho = 0 = e\,(-n_0 - N_A^- + p_0 + N_D^+) \tag{2.52}$$

bzw. umgestellt als Bilanzgleichung von positiver und negativer Ladung

$$n_0 + N_A^- = p_0 + N_D^+ \,. \tag{2.53}$$

Das ist die wichtige Neutralitätsbedingung für Halbleiter im thermodynamischen Gleichgewicht. Die darin enthaltenen Beiträge haben wir bereits ermittelt. Wir können die Neutralitätsbedingung daher zunächst allgemein mit n_0 nach Gl. (2.23), p_0 nach Gl. (2.26), N_D^+ nach Gl. (2.49) und N_A^- nach Gl. (2.50) formulieren:

$$N_C\,f(W_C) + N_A\,f(W_A^*) = N_V\,(1 - f(W_V)) + N_D\,(1 - f(W_D^*)) \tag{2.54}$$

Darin sind N_C und N_V die effektiven Zustandsdichten von Leitungs- und Valenzband nach Gl. (2.24) und (2.27). N_A und N_D sind die Dichten der Dotierung, die bei reiner Akzeptor- oder Donator-Dotierung entsprechend auf

Null zu setzen sind. Nur bei „vergifteten“ oder „kompensierten“ Halbleitern sind gleichzeitig beide Dotierungen vorhanden. In Gl. (2.54) ist mit $f(W)$ zunächst die Fermi-Dirac-Verteilungsfunktion nach Gl. (2.7) verwendet worden, da a priori aus Gl. (2.54) nicht ersichtlich ist, ob sich ein Fermi-Niveau ergibt, das die Verwendung der Boltzmann-Näherung für einen der Terme zulässt.
Wir rufen uns nochmals die Fermi-Dirac-Verteilungsfunktion $f(W)$ in Erinnerung

$$f(W) = \frac{1}{1 + e^{\frac{W - W_F}{kT}}} \,, \tag{2.7}$$

die durch Substitution von W mit W_C, W_V, W_A^*, W_D^* direkt in die Neutralitätsbedingung nach Gl. (2.54) eingesetzt werden kann.

2.10 Ermittlung der Fermi-Energie

Alle in der Neutralitätsbedingung (Gl. (2.54)) enthaltenen Größen bis auf die Fermi-Energie W_F sind bekannt, so dass Gl. (2.54) als Bestimmungsgleichung für die Fermi-Energie im allgemeinen Fall verwendet werden kann.
Leider ist Gl. (2.54) eine transzendente Gleichung, die sich nicht geschlossen nach W_F umformen lässt. Zur Lösung bietet sich ein Computer oder eine grafische Lösung an.
Wir favorisieren hier wegen der Unabhängigkeit von elektronischen Hilfsmitteln und der Schulung der Intuition die grafische Variante. Versuchen Sie doch einmal als Kompromiss die grafische Lösung auf dem Computer zu programmieren.
Für die grafische Lösung benötigen wir eine geeignete Näherung zur Darstellung der Fermi-Dirac-Verteilungsfunktion. Wir unterteilen den Verlauf der Funktion daher in zwei Bereiche mit den Näherungen

$$f(W) \approx \begin{cases} 1 \,, & W_F \gg W \\ e^{-\frac{W - W_F}{kT}} \,, & W_F \ll W \,, \end{cases} \tag{2.55}$$

die im Punkt $W = W_F$ mit $f(W_F) = \frac{1}{2}$ zusammentreffen. Der maximale Fehler tritt bei dieser Approximation bei $W = W_F$ auf. Hier nimmt die Näherung den Wert 1 anstatt den richtigen Wert 0,5 an. Wir werden aber sehen, dass dieser Fehler in der Regel ohne Auswirkung auf die Bestimmung des Fermi-Niveaus ist. Wählen wir eine Darstellung mit logarithmischer (Basis 10) y-Achse und tragen auf der x-Achse die auf kT normierte Fermi-Energie

auf, so ergibt sich der in Abb. 2.10 gezeigte Verlauf von $f(W, W_F)$ bzw. $1 - f(W, W_F)$.

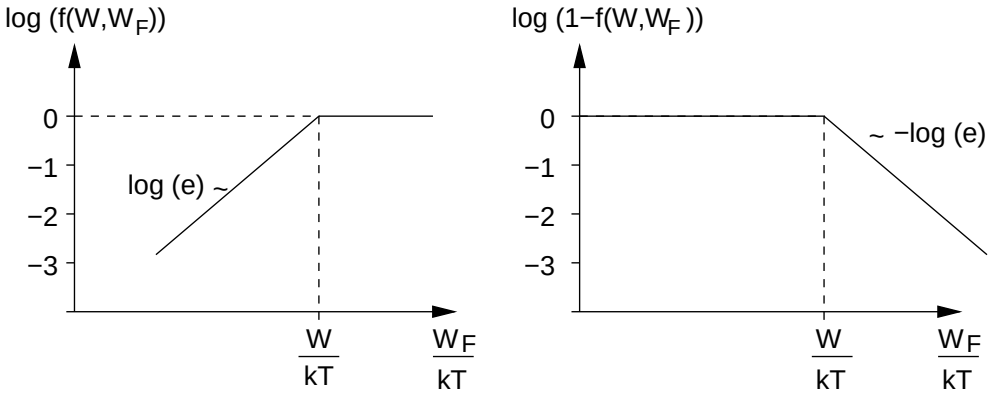

Abb. 2.10: Geradenapproximation bei logarithmischer Darstellung des Verlaufs der Fermi-Dirac-Verteilungsfunktion (links) und der Nicht-Besetzungsfunktion (rechts).

Dabei sind die Verläufe der beiden Bereiche durch Geradenabschnitte angenähert worden. Im Bereich $W_F < W$ ergibt sich durch die logarithmische y-Achse die Gerade

$$y = \log\left(e^{-\frac{W-W_F}{kT}}\right) = \left(-\frac{W}{kT} + \frac{W_F}{kT}\right) \log(e) \; . \tag{2.56}$$

Die Gerade hat also eine Steigung von $\log(e)$ und mündet im Punkt $y = \log(1) = 0$, $x = \frac{W}{kT}$ in den horizontalen Verlauf.
Durch entsprechende Überlegungen erhält man den in Abb. 2.10 rechts gezeigten Verlauf für die Wahrscheinlichkeit $1 - f(W, W_F)$ eines nicht besetzten Niveaus. Durch Multiplikation mit den effektiven Zustandsdichten in Gl. (2.54) verschieben sich die Approximationsverläufe nach oben um den Logarithmus der jeweiligen Zustandsdichte (z. B. liegt der waagerechte Verlauf von $N_C f(W_C)$ bei $\log(N_C)$).
Zur Ermittlung der Fermi-Energie aus der Neutralitätsbedingung (Gl. (2.54)) stellen wir die linke und rechte Seite der Gleichung grafisch in Abhängigkeit der Fermi-Energie dar. Die Fermi-Energie im Schnittpunkt der beiden Kurven erfüllt die Neutralitätsbedingung und ist die gesuchte Lösung.
Abb. 2.11 zeigt die grafische Lösung der Neutralitätsbedingung für einen **Si**-Halbleiter mit p- und n-Dotierung. Dargestellt sind die Verläufe ohne die

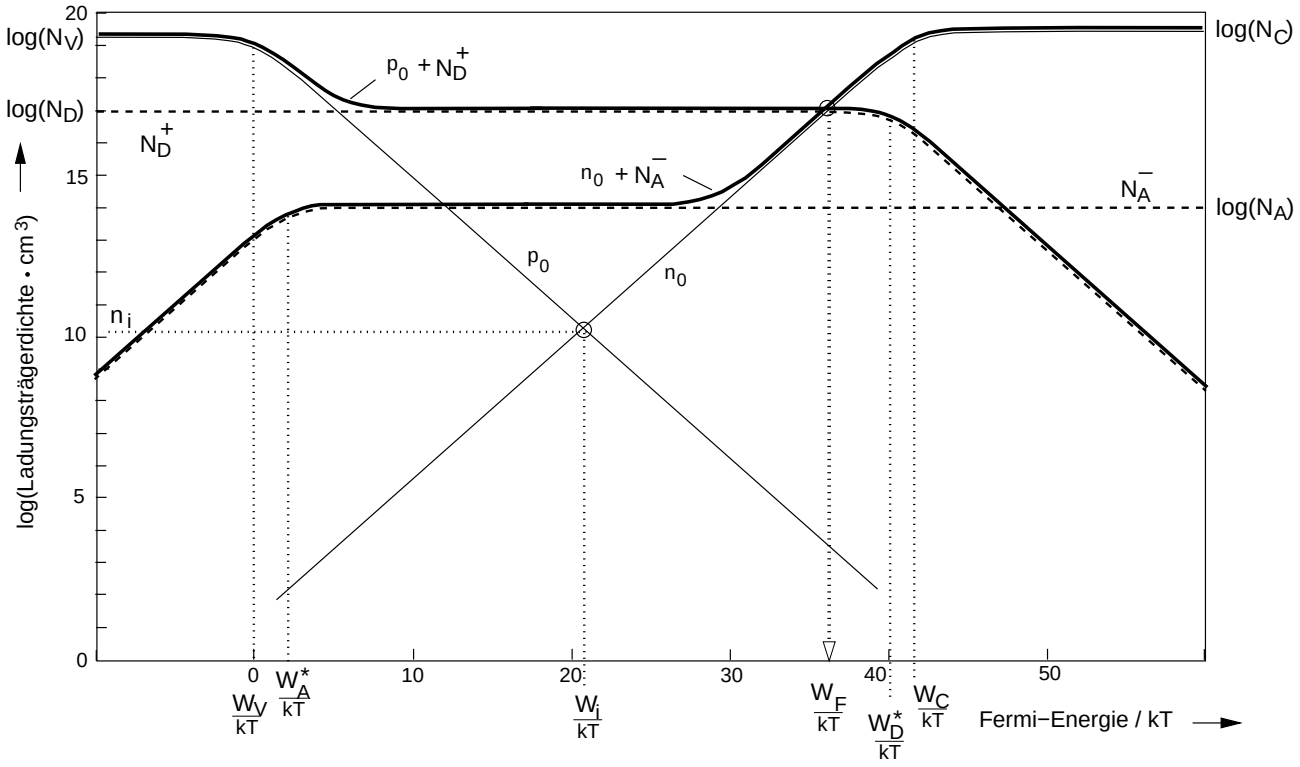

Abb. 2.11: Grafische Bestimmung der Fermi-Energie für einen Si-Halbleiter bei $T = 300\,\mathrm{K}$ mit einer Dotierung $N_A = 10^{14}\,\mathrm{cm}^{-3}$ und $N_D = 10^{17}\,\mathrm{cm}^{-3}$. Die beiden dicken Kurven stellen die Verläufe der beiden Seiten der Neutralitätsbedingung dar (ohne Geradenapproximation).

Geradenapproximation.
Der Nullpunkt der Energie wurde an die Kante des Valenzbandes gelegt.
Bei Verwendung der Geraden-Näherung ergibt sich als einzige Abweichung zu Abb. 2.11 jeweils eine Ecke an den Schnittpunkten der Geradenabschnitte anstelle des gekrümmten Verlaufs. Der maximale Fehler entsteht am Schnittpunkt der Geraden und beträgt den Faktor Zwei. Aus Abb. 2.11 erkennt man, dass dieser Fehler in dem vorliegenden Beispiel keinen Einfluss auf das Ergebnis hat, da der gesuchte Schnittpunkt davon weit entfernt liegt. Der Fehler bei einer Verwendung der Geradenapproximation ist daher in der Regel vernachlässigbar.
Der waagerechte Verlauf liegt für die vier Ladungsträgerdichten bei dem Wert der effektiven Zustandsdichten (N_C, N_V) im Fall der freien Ladungsträger

bzw. bei den Dotierungskonzentrationen (N_D, N_A) bei den Ladungsträgerdichten der ortsfesten, ionisierten Dotierungsatome. Bei der grafischen Addition von zwei Beiträgen wirkt sich der logarithmische Maßstab vorteilhaft aus. Hier geht der Summenverlauf direkt vom Verlauf eines Beitrags zum anderen über, sobald der eine Beitrag um eine Zehnerpotenz unter den anderen gefallen ist.
Die Konstruktion der Kurven zur grafischen Lösung ist aufgrund der auf kT normierten x-Achse besonders einfach. Dadurch besitzen alle Verläufe unabhängig von der Temperatur die gleiche Steigung (betragsmäßig). Leitungsband und ionisiertes Akzeptorniveau haben eine positive, Valenzband und ionisiertes Donator-Niveau eine negative Steigung (vgl. Abb. 2.10).
Durch die Überlagerung der einzelnen Kurven lassen sich anhand der grafischen Darstellung verschiedene Fälle einfach darstellen. So beschreiben z. B. die Verläufe von n_0 und p_0 alleine den nicht dotierten Halbleiter. Ihr Schnittpunkt liefert die Fermi-Energie W_i für Eigenleitung und die Eigenleitungsdichte n_i. Auch die Abhängigkeit der Fermi-Energie von der Temperatur lässt sich mit ein wenig Übung sehr einfach ermitteln. Dies betrachten wir im nächsten Kapitel.

2.11 Temperaturabhängigkeit von Fermi-Niveau und Ladungsträgerdichte

Wir vermuten, dass durch die starke Temperaturabhängigkeit der Exponentialfunktionen $(\exp(\frac{W}{kT}))$ in der Fermi-Dirac-Verteilungsfunktion die Eigenschaften eines Halbleiters ebenfalls stark von der Temperatur abhängen. Als Maß für die Abhängigkeit ermitteln wir die Lage des Fermi-Niveaus und die Dichte der Ladungsträger in den Bändern in Abhängigkeit von der Temperatur.
Zur Erhöhung der Übersichtlichkeit und ohne Einschränkung der Allgemeingültigkeit betrachten wir als Beispiel einen Halbleiter mit n-Dotierung. Die Verhältnisse bei p-dotierten Halbleitern stellen sich dann symmetrisch (bezogen auf die Bandmitte) dazu dar.
Ein immer anwendbares Lösungsverfahren zur Ermittlung des Fermi-Niveaus ist die im vorangegangenen Kapitel beschriebene grafische Lösung. Abb. 2.12 zeigt ein Beispiel für schwache n-Dotierung mit $N_D = 10^{14}$ cm^{-3} bei $T = 200$, 300 und 600 K.

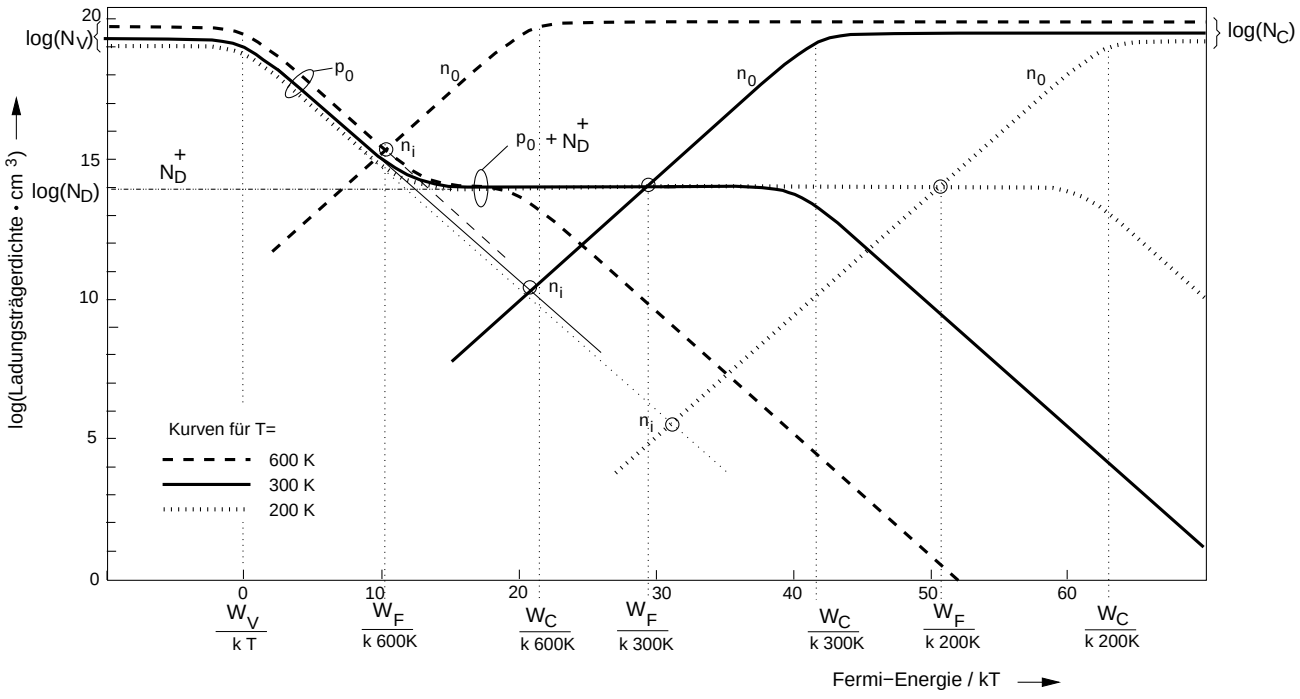

Abb. 2.12: Verläufe von p_0, n_0 und N_D^+ für $T = 200$, 300 und 600 K bei n-dotiertem Halbleiter. Der Schnittpunkt der $p_0 + N_D^+$-Kurve mit der entsprechenden n_0-Kurve ergibt die Fermi-Energie der jeweiligen Temperatur.

Die Temperaturabhängigkeit stellt sich in dem Diagramm recht einfach dar. Zu berücksichtigen sind die temperaturabhängigen effektiven Zustandsdichten N_C und N_V, die über Gl. (2.24) und (2.27) ermittelt werden können. Sie legen das Start-Plateau der Geradennäherung für die beiden Seiten der Neutralitätsgleichung fest (ganz links beginnend für die positiven Ladungen auf der p_0-Kurve, ganz rechts für die negativen Ladungen auf n_0). An den Energien der Bandkanten W_V und W_C fallen beide Verläufe ab[14]. Bedingt durch die Normierung von W_C und kT geschieht dies temperaturabhängig an unterschiedlichen Stellen. Prinzipiell gilt dies auch für W_V. Jedoch ändert sich durch die Wahl von $W_V = 0$ als Nullpunkt die Position hier nicht.

[14]Mathematisch steigt natürlich der Verlauf von n_0. In dieser, an die Konstruktion des Diagramms angelehnten Beschreibung fällt der Verlauf ab, da wir für den n_0-Verlauf von der rechten Seite des Diagramm-Randes kommen.

Sobald der Verlauf der positiven Ladungen auf der p_0-Kurve das Niveau der komplett ionisierten Donatoren ($N_D^+ = N_D$) erreicht, geht er in diesen Verlauf über. Das Niveau dieses mittleren Plateaus ist temperaturunabhängig, da sich die Anzahl (Dichte) der in den Kristall eingebrachten (dotierten) Atome nicht ändert. Bei steigender Fermi-Energie (von dem mittleren Plateau nach rechts) nimmt die Anzahl der ionisierten Donatoren ab, sobald W_F die Energie des Donator-Niveaus W_D^* erreicht. Diese Energie ist in Abb. 2.12 zur Erhöhung der Übersichtlichkeit nicht eingezeichnet, da sie sehr dicht an der Energie W_C der Leitungsbandkante liegt und in erster Näherung in der Darstellung mit dieser gleichgesetzt werden kann.

Die Fermi-Energie ergibt sich als Schnittpunkt des n_0-Verlaufs mit dem $p_0 + N_D^+$ Verlauf. Abb. 2.12 zeigt deutlich, dass dies bei 200 K und 300 K auf dem mittleren Plateau des $p_0 + N_D^+$ Verlaufs erfolgt. Wir lesen $W_F(200\,\mathrm{K}) = 0,87\,\mathrm{eV}$ und $W_F(300\,\mathrm{K}) = 0,76\,\mathrm{eV}$ ab[15]. Das Fermi-Niveau wandert also mit steigender Temperatur in Richtung Mitte der Bandlücke ($= \frac{1}{2} \cdot 1,08$ eV). Die Erklärung hierfür wird deutlich, wenn wir die Temperatur noch weiter erhöhen. Bei höherer Temperatur verschiebt sich der Schnittpunkt auf den abfallenden Ast des p_0-Anteils. Der Schnittpunkt ist dann bei einer Ladungsdichte $p_0 \gg N_D$, so dass die Fermi-Energie $W_F = W_i$ den Wert für Eigenleitung besitzt (in etwa Bandmitte). Bei hohen Temperaturen (in Abb. 2.12 ca. $T > 600$ K) liegt also trotz der Dotierung Eigenleitung vor.
Dies ist die Erklärung für das Wandern der Fermi-Energie in Richtung Bandmitte: Mit steigender Temperatur nimmt die Energie der Elektronen im Valenzband immer weiter zu und es gelangen immer mehr von dort in das Leitungsband. Die Donator-Niveaus sind bei diesen Temperaturen bereits alle ionisiert (Störstellenerschöpfung), da die hierzu benötigte Energie viel geringer als die Energie der Bandlücke ist. Der Zuwachs an Elektronen im Leitungsband erfolgt also nur durch Elektronen aus dem Valenzband. Ist die Temperatur so groß, dass deren Zahl viel (Faktor 10) größer ist als die Zahl der Dotierungs-Atome, ist die Dotierung vernachlässigbar. Dann liegen die Verhältnisse bei Eigenleitung vor und das Fermi-Niveau muss in der Bandmitte liegen.

[15] Genauer gesagt lesen wir $W_F/(k \cdot 200\,\mathrm{K}) \approx 51$ und $W_F/(k \cdot 300\,\mathrm{K}) \approx 29$ ab, woraus wir durch Umstellen und Einsetzen der Boltzmannkonstanten die oben angegebenen Werte berechnen.

Wir wissen jetzt, dass das Fermi-Niveau mit steigender Temperatur in Richtung Bandmitte wandert. Aber von wo kommt es bei tiefen Temperaturen? Hierzu machen wir (anstelle der immer möglichen grafischen Lösung) einige einfache Überlegungen, die für unsere Belange vollständig ausreichen. Wir nehmen an, die Temperatur ist auf dem absoluten Nullpunkt ($T = 0$). Dann ist das Leitungsband leer und keines der Donator-Atome ionisiert. Steigt die Temperatur auch nur ein wenig an (z. B. 1 K), dann reicht die thermische Energie aus um nach der Fermi-Dirac-Statistik Elektronen mit einer kleinen aber von Null verschiedenen Wahrscheinlichkeit in das Leitungsband zu heben. Dies ist von dem dicht darunter liegenden Donator-Niveau viel wahrscheinlicher als von dem im Vergleich viel weiter entfernten Valenzband. Wir vernachlässigen daher den Beitrag des Valenzbandes bei tiefen Temperaturen. Da das Donator-Niveau bei tiefen Temperaturen komplett besetzt ist mit Elektronen, die mit steigender Temperatur in das Leitungsband gehen, liegen hier die gleichen Verhältnisse wie bei Eigenleitung vor. Die Rolle des Valenzbandes wird hier vom Donator-Niveau übernommen.
Damit ist klar, dass die Fermi-Energie wie bei der Eigenleitung in der Mitte zwischen den Bändern liegen muss. Das Fermi-Niveau startet also in dieser einfachen Betrachtungsweise für $T = 0$ bei

$$W_F(T = 0) \approx \frac{1}{2}\,(W_C + W_D^*)\ . \tag{2.57}$$

Abb. 2.13 zeigt zur Verdeutlichung diese Lage in einem Bändermodell.

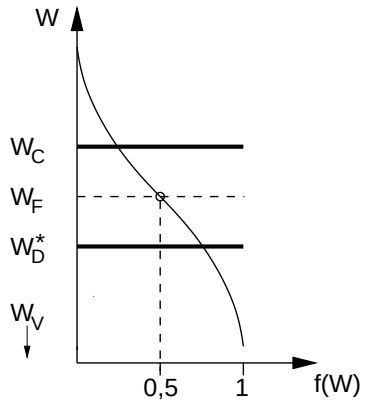

Abb. 2.13: Bei tiefen Temperaturen liegt das Fermi-Niveau zwischen effektivem Donator-Niveau und Leitungsbandkante.

Da die gleichen Verhältnisse wie bei Eigenleitung vorliegen, kann die La-

dungsträgerdichte mit Hilfe von Gl. (2.35) zur Bestimmung der Eigenleitungsdichte ermittelt werden. Dazu muss nur der Bandabstand W_g durch den Abstand zwischen effektivem Donator-Niveau und Leitungsbandkante ersetzt werden ($W_g \to W_C - W_D^*$) und anstelle der effektiven Zustandsdichte des Valenzbandes die Dichte der Donator-Dotierung eingesetzt werden ($N_V \to N_D$). Damit ergibt sich die Ladungsträgerdichte im Leitungsband bei tiefen Temperaturen zu

$$n_0 = \sqrt{N_C N_D}\, e^{-\frac{W_C - W_D^*}{2kT}} \ . \tag{2.58}$$

Sie nimmt von $n_0(T = 0) = 0$ an zu und erreicht bei einer Temperatur T_α die Größenordnung der Dotierungsdichte. Theoretisch ließe sich T_α aus Gl. (2.58) berechnen. Jedoch muss dazu auch die Temperaturabhängigkeit von N_C berücksichtigt werden, wodurch keine analytische Lösung existiert. Da bis T_α die Zahl der Ladungsträger zunimmt, spricht man auch von (Störstellen-)Reserve („Freeze-out“).
Gl. (2.58) besitzt als Näherung einen nur sehr eingeschränkten Gültigkeitsbereich. So kann z. B. n_0 in dieser Näherung auch Werte größer als N_D annehmen, was entgegen unserer Annahme für den betrachteten Bereich ist, wonach alle Ladungsträger im Leitungsband aus dem Donator-Niveau stammen. Wir leiten daher eine genauere Beziehung für die Ladungsträgerdichte in diesem Bereich her:
Bei der niedrigen Temperatur stammen alle Elektronen im Leitungsband aus dem Donator-Niveau. Die Elektronendichte im Leitungsband ist demnach gleich der Dichte der nicht mit einem Elektron besetzten Donator-Zustände. Die mathematische Formulierung dafür lautet

$$n_0 = N_D \left(1 - \frac{1}{1 + e^{\frac{W_D^* - W_F}{kT}}}\right) \ . \tag{2.59}$$

Weiterhin gilt allgemein, solange die Boltzmann-Näherung gilt:

$$n_0 = N_C\, f(W_C) = N_C\, e^{-\frac{W_C - W_F}{kT}} \ . \tag{2.60}$$

Diese Beziehung kann dazu verwendet werden, um die Fermi-Energie in der vorangegangenen Gl. (2.59) zu ersetzen. Durch Umformen und Lösen einer quadratischen Gleichung ergibt sich

$$n_0 = \frac{2\, N_D}{1 + \sqrt{1 + 4\, \frac{N_D}{N_C}\, e^{\frac{W_C - W_D^*}{kT}}}} \ . \tag{2.61}$$

Anhand der Fallunterscheidung
für $T < T_\alpha$:

$$4\,\frac{N_D}{N_C}\,e^{\frac{W_C - W_D^*}{kT}} \gg 1 \;\rightarrow\; n_0 = \sqrt{N_C N_D}\,e^{-\frac{W_C - W_D^*}{2kT}} \tag{2.62}$$

und $T_\alpha \leq T \leq T_\beta$:

$$4\,\frac{N_D}{N_C}\,e^{\frac{W_C - W_D^*}{kT}} \ll 1 \;\rightarrow\; n_0 = N_D \tag{2.63}$$

lassen sich zwei Temperaturbereiche für Näherungslösungen von Gl. (2.61) bestimmen.
Für den Bereich bis T_α nimmt Gl. (2.62) als Näherung von Gl. (2.61) die bereits vorhergesagte Form der Eigenleitungsdichte nach Gl. (2.58) für Störstellenreserve an.
Für $T > T_\alpha$ geht dieser Verlauf in das konstante Plateau $n_0 = N_D$ nach Gl. (2.63) über. Ab T_α sind sämtliche Donatoren ionisiert. Man spricht daher auch von (Störstellen-)Erschöpfung.
Ab $T = T_\beta$ wird die Anzahl der Elektronen, die durch Eigenleitung aus dem Valenzband stammen, größer werden als die der Dotierung und die Annahme die zu Gl. (2.61) führte, gilt nicht mehr. Ab dieser Temperatur geht das konstante Plateau in den Verlauf der Eigenleitungsdichte nach Gl. (2.35) über.
Abb. 2.14 fasst die vorangegangenen Aussagen über die Temperaturabhängigkeit der Elektronendichte im Leitungsband und Lage des Fermi-Niveaus zusammen. Es ist unmittelbar anhand des Schnittpunktes der Gerade für Eigenleitung mit dem Plateau $n_0 = N_D$ zu sehen, dass sich T_β bei höherer Dotierung zu höheren Temperaturen verschiebt.
Die gleichen Überlegungen gelten für die Löcherdichte im Valenzband. Daher gilt Abb. 2.14 links ebenso für Löcher. Entsprechend verläuft die Lage des Fermi-Niveaus für p-dotierte Halbleiter i. e. gespiegelt zu dem Verlauf in Abb. 2.14 rechts.

Halbleiterbauelemente werden im Bereich der Störstellen-Erschöpfung betrieben. In diesem Temperaturbereich hängt die Zahl der freien Ladungsträger und damit die Leitfähigkeit von der Dotierung ab. Die Temperaturabhängigkeit ist gering und die Eigenschaften des Halbleiters können von dem Entwickler des Bauelementes gezielt über die Stärke der Dotierung eingestellt werden. Die vorangegangenen Überlegungen

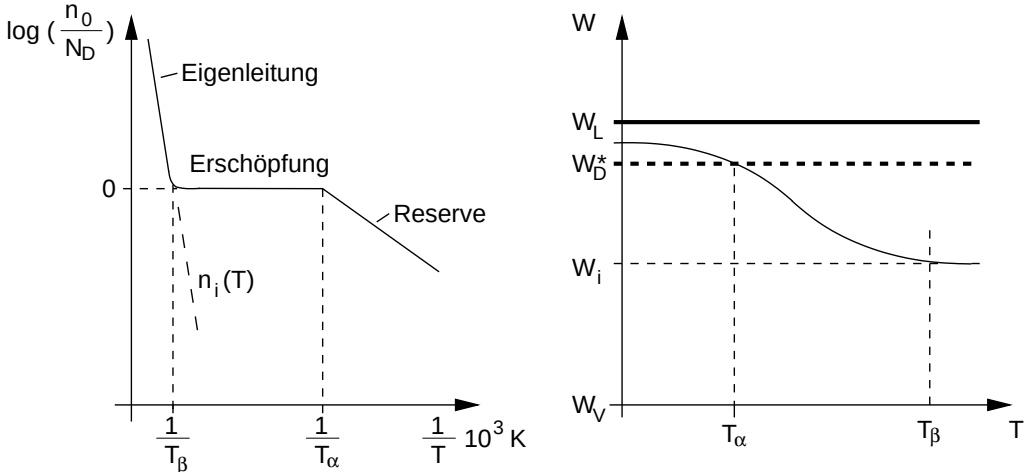

Abb. 2.14: Temperaturabhängigkeit der Elektronendichte im Leitungsband eines n-dotierten Halbleiters (links) und die entsprechende Lage des Fermi-Niveaus (rechts). Für die Kurvenabschnitte links gilt bei Eigenleitung Gl. (2.35), bei Erschöpfung Gl. (2.63) und bei Reserve Gl. (2.62).

verdeutlichen auch den Grund für die bevorzugte Verwendung von **Si** anstelle von **Ge**, wenn es darum geht, Halbleiterbauelemente für möglichst hohe Temperaturen zu entwerfen. Da aufgrund der kleineren Bandlücke bei **Ge** gilt $n_i(\mathbf{Ge}) > n_i(\mathbf{Si})$, setzt für einen **Si**-Halbleiter der Beginn der Eigenleitung (T_β) später ein.

Für die Berechnung der Halbleitereigenschaften im Bereich der Störstellenerschöpfung kann wegen der Ionisation aller Dotierungsatome immer

$$N_D^+ = N_D \ , \ N_A^- = N_A \tag{2.64}$$

gesetzt werden. Da im Bereich der Störstellenerschöpfung die Eigenleitung vernachlässigbar ist, bestehen die sich in der Majorität befindenden Ladungsträger ausschließlich aus den Ladungsträgern der ionisierten Dotierungsatome. Daher gilt für alle nachfolgenden Berechnungen der Halbleiter die Zusammenfassung in Tabelle 2.3.

Eine Beziehung für die Berechnung bei gleichzeitiger p- und n-Dotierung wird in einer Übung hergeleitet.

	Majoritätsträger	Minoritätsträger
n-Dotierung	$n_0 = N_D^+ = N_D$	$p_0 = \frac{n_i^2}{n_0} = \frac{n_i^2}{N_D}$
p-Dotierung	$p_0 = N_A^- = N_A$	$n_0 = \frac{n_i^2}{p_0} = \frac{n_i^2}{N_A}$

Tabelle 2.3: Ladungsträgerdichten bei n- und p-Dotierung.

2.12 Bewegte Ladungsträger

Wir sind nach dem vorangegangenen Kapitel in der Lage, die Zahl (genauer die Dichte) der Elektronen und Löcher in Leitungs- und Valenzband zu berechnen. Für die Funktion von Halbleiterbauelementen ist deren Bewegung entscheidend. Bei gerichteter Bewegung entsteht ein Strom von Ladungsträgern, der den bekannten elektrischen Strom bildet.
Aufgrund der Ursache der gerichteten Bewegung unterscheiden wir zwischen Feld- oder Driftstrom, der aufgrund der Kraftwirkung eines elektrischen Feldes auf die Ladungsträger entsteht und den Diffusionsstrom, der sich aufgrund eines Konzentrationsgefälles von Ladungsträgern ergibt.
Der gerichteten Bewegung überlagert ist eine zufällige ungerichtete Bewegung der (freien) Ladungsträger, die sich aus der thermischen Anregung ergibt. Der Mittelwert des Geschwindigkeitvektors dieser ungerichteten Bewegung ist Null, so dass hierdurch kein Stromfluss entsteht.
Bei ihrer Bewegung durch den Kristall stoßen die Ladungsträger gegeneinander, gegen Störstellen und Grenzflächen im Kristall und gegen das Kristallgitter. Stöße mit dem Kristallgitter sind im Temperaturbereich der Störstellenerschöpfung meist dominant. Durch die Stöße wird die Beweglichkeit der Ladungsträger eingeschränkt und die Leitfähigkeit des Stroms verringert. Beide Begriffe stellen wichtige Parameter von Halbleiterbauelementen dar, die im weiteren Verlauf ermittelt werden.

2.12.1 Der Halbleiter ohne elektrisches Feld

Ohne elektrisches Feld bewegen sich die Ladungsträger mit ihrer effektiven Masse m^* $(= m_{ec}^*, m_{hc}^*)$ aufgrund ihrer thermischen Energie „frei“ im Halbleiter.
Sie stoßen dabei gegen das Kristallgitter bzw. das Gitter, das selbst auch

in thermischer Schwingung ist, stößt gegen die Ladungsträger. Die Stoßprozesse im Einzelnen sind nicht vorhersagbar, führen aber zu einer mittleren Geschwindigkeit der Ladungsträger. Abb. 2.15 rechts zeigt eine beliebige Geschwindigkeitsverteilung für einen Ladungsträger in einer willkürlich gewählten Richtung x.

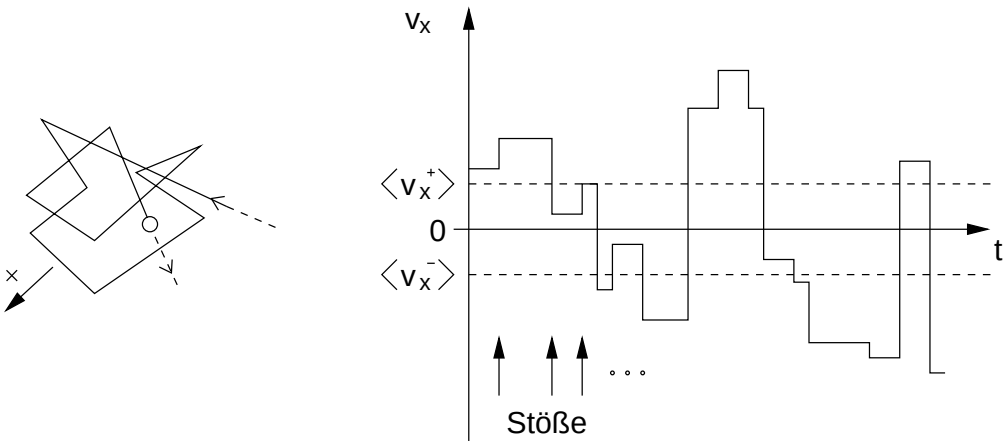

Abb. 2.15: Links: Zufällige ungerichtete Bewegung eines Ladungsträgers im Zweidimensionalen. Rechts: Geschwindigkeit des Ladungsträgers in x-Richtung.

Für die Bewegung in $+x$-Richtung stellt sich eine mittlere Geschwindigkeit $\langle \mathrm{v}_x^+ \rangle$ ein. Diese muss, aufgrund der ungerichteten, zufälligen Bewegung entgegengesetzt gleich der mittleren Geschwindigkeit in $-x$-Richtung sein, so dass im Mittel für die resultierende Geschwindigkeit in x-Richtung gilt $\langle \mathrm{v}_x \rangle = 0 = \langle \mathrm{v}_x^+ \rangle + \langle \mathrm{v}_x^- \rangle$. Die gleichen Aussagen gelten für die Mittelwerte der Geschwindigkeit in allen drei Dimensionen und Richtungen. Wir ersetzen daher im Folgenden die x-Richtung durch einen allgemeinen Ortsvektor $\vec{r}$. Wenn wir im Folgenden ein elektrisches Feld $\vec{E}$ annehmen, soll $\vec{r}$ in Richtung von $\vec{E}$ zeigen.
In einem einfachen Modell kann ein Ladungsträger als Teilchen nach der klassischen Thermodynamik behandelt werden. Danach besitzt er aufgrund seiner Bewegung in x-Richtung eine kinetische Energie von

$$W_{kin_x} = \frac{1}{2} m^* \langle \mathrm{v}_x^2 \rangle = \frac{1}{2} kT \; . \tag{2.65}$$

Für eine Bewegung in y- und z-Richtung gilt aufgrund der Unabhängigkeit der drei orthogonalen Raumvektoren die gleiche Beziehung. Die Gesamtener-

gie ist demnach

$$W_{kin} \;=\; W_{kin_x} + W_{kin_y} + W_{kin_z} \;=\; \frac{1}{2} m^* \langle \mathrm{v}^2 \rangle \;=\; \frac{3}{2} kT \;. \tag{2.66}$$

Daraus berechnet sich die mittlere (thermische) Geschwindigkeit für eine Bewegung

$$\mathrm{v}_{th} \;=\; \sqrt{\langle \mathrm{v}^2 \rangle} \;=\; \sqrt{\frac{3kT}{m^*}} \tag{2.67}$$

der Ladungsträger in drei Dimensionen ohne elektrisches Feld.

Beispiel: Für ein Elektron in **Si** gilt $m^* = m^*_{ec} = 0,26\, m_e$.
Mit $kT \approx 0{,}026\,\mathrm{eV}$ bei $300\,\mathrm{K}$ und $1\,\mathrm{eV}{=}1,6 \cdot 10^{-19}\,\mathrm{J}$ ($1\,\mathrm{J}{=}1\frac{\mathrm{kg\,m^2}}{\mathrm{s^2}}$) ergibt sich

$$\mathrm{v}_{th} \;=\; \left(\frac{3 \cdot 0,026 \cdot 1,6 \cdot 10^{-19}\,\mathrm{kg\,m^2}}{0,26 \cdot 9,1 \cdot 10^{-31}\,\mathrm{kg\,s^2}} \right)^{-\frac{1}{2}} \approx 2 \cdot 10^5 \tfrac{\mathrm{m}}{\mathrm{s}} = 2 \cdot 10^7 \tfrac{\mathrm{cm}}{\mathrm{s}} \;.$$

2.12.2 Ladungsträgerdichten außerhalb des thermodynamischen Gleichgewichts

Änderungen des Gleichgewichtszustandes in einem Halbleiter(gebiet) können durch „äußere“ Einflüsse hervorgerufen werden. Dazu gehören z. B. ein Spannungsabfall über dem Gebiet bzw. ein Stromfluss durch das Gebiet. Dazu gehört auch das Einbringen zusätzlicher, ggf. auch ortsabhängiger Ladungsträgerdichten (z. B. durch Lichteinstrahlung). Es ergeben sich dann die „Überschussladungsträgerdichten“ n und p, die um Δn bzw. Δp von ihrem Gleichgewichtsdichten n_0, p_0 abweichen.

$$n \;=\; n_0 + \Delta n \;=\; N_C\, e^{-\frac{W_C - W_{Fn}}{kT}} \tag{2.68}$$

$$p \;=\; p_0 + \Delta p \;=\; N_V\, e^{-\frac{W_{Fp} - W_V}{kT}} \tag{2.69}$$

Um weiterhin mit den uns schon vertrauten Verteilungsfunktionen für die Ladungsträgerdichten ähnlich wie in Gl. (2.23) und (2.26) arbeiten zu können, haben wir auf der rechten Seite eine Formulierung mit der gleichen Struktur gewählt. Als Unterschied zu den Berechnungen im Gleichgewichtsfall müssen wir anstelle des Fermi-Niveaus jetzt mit zwei unterschiedlichen „Quasi-Fermi-Niveaus“ W_{Fp}, W_{Fn} arbeiten. Dadurch wird sowohl der Überschussanteil

Δn, Δp als auch die Abweichung von der Eigenleitungsdichte n_i berücksichtigt. Dies zeigt sich durch Multiplikation von p und n.

$$p\,n \;=\; \underbrace{N_C\,N_\mathrm{V}\,e^{-\frac{W_g}{kT}}}_{n_i^2}\; e^{\frac{W_{Fn}-W_{Fp}}{kT}} \tag{2.70}$$

$$p\,n \;=\; \qquad n_i^2 \qquad\; e^{\frac{W_{Fn}-W_{Fp}}{kT}} \tag{2.71}$$

Der Abstand zwischen W_{Fn} und W_{Fp} bestimmt also die Abweichung von der Gleichgewichtsdichte. Für $pn > n_i^2$ liegt W_{Fn} im Bändermodell über W_{Fp}. Für $pn < n_i^2$ liegt W_{Fp} über W_{Fn}. Wir werden gegen Ende dieses zweiten Hauptkapitels genug Wissen haben, um genauere Aussagen über den Verlauf der Quasi-Fermi-Niveaus machen zu können. Wir werden dann auch anhand der Kontinuitätsgleichung sehen, dass die Erhöhung einer Ladungsträgerdichte aufgrund der Ladungsneutralität ($\Delta n = \Delta p$, wird später gezeigt) nahezu unmittelbar eine Erhöhung der anderen Ladungsträgerart bewirkt. Daher folgt für $p\,n > n_i^2 : p > p_0,\, n > n_0$ bzw. für $p\,n \,<\, n_i^2 :\; p < p_0,\, n < n_0$.

2.12.3 Der Halbleiter im elektrischen Feld

Wir betrachten den Halbleiter, in dem ein elektrisches Feld wirkt. Wir nehmen zur Vereinfachung an, dass das elektrische Feld durch eine Anordnung wie in Abb. 2.16 erzeugt wird und zwischen den beiden stirnseitig angebrachten Elektroden homogen ist.

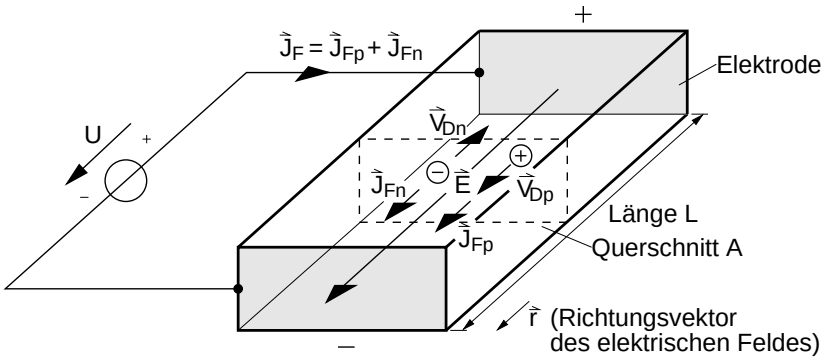

Abb. 2.16: Definition von Stromdichten für Löcher und Elektronen.

Aufgrund der Kraftwirkung des elektrischen Feldes

$$\vec{F} \;=\; q\,\vec{E} \quad , \quad q \;=\; -e,\, +e \tag{2.72}$$

erfährt ein Ladungsträger eine, zu der Kraft proportionale Beschleunigung entsprechend

$$\vec{F} = m^* \vec{a} = m^* \frac{d\vec{v}}{dt} \quad , \quad m^* = m^*_{ec},\, m^*_{hc} \,, \tag{2.73}$$

die ihn nach jedem Stoß wieder beschleunigt. Der Startwert der Geschwindigkeit nach einem Stoß ist wieder ein Zufallsprozess mit dem Mittelwert Null (keine Vorzugsrichtung, d. h. beliebige Richtung). Die Beschleunigung erfolgt jedoch gerichtet und lässt Elektronen entgegen und Löcher in der Richtung des elektrischen Feldes driften. Abb. 2.17 zeigt das an einem Beispielverlauf.

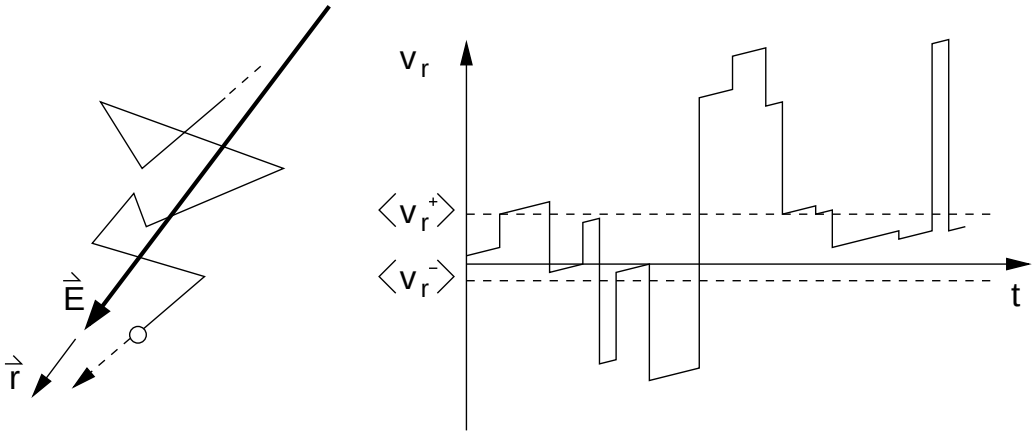

Abb. 2.17: Links: Gerichtete Driftbewegung eines Ladungsträgers aufgrund der Kraftwirkung eines elektrischen Feldes. Rechts: Beschleunigung der Ladungsträger nach jedem Stoß. Der Geschwindigkeitsvektor $\vec{v}_r$ zeigt in Richtung von $\vec{E}$.

Aufgrund der Richtung des elektrischen Feldes werden die Ladungsträger in dem Beispiel in positive $\vec{r}$-Richtung beschleunigt und in entgegengesetzter Richtung abgebremst. Anhand dieses Verhaltens ist klar, dass es sich in diesem Beispiel bei dem Ladungsträger um ein Loch handelt (vgl. Feldrichtung in Abb. 2.17). Die gleiche Überlegung gilt für Elektronen, die sich aufgrund der entgegengesetzten Ladung in die entgegengesetzte Richtung bewegen. Die gerichtete Bewegung drückt sich dadurch aus, dass eine resultierende Geschwindigkeitskomponente, hier in x-Richtung $\langle v \rangle = \langle v_r^+ \rangle + \langle v_r^- \rangle > 0$ entsteht.

Wir wollen die resultierende Geschwindigkeit (Driftgeschwindigkeit) ermitteln und mit ihrer Hilfe die Stromdichte als die Zahl der Elektronen, die pro Zeiteinheit einen Querschnitt durchströmen, bestimmen. Dazu nehmen wir an, dass für viele Stöße eines Elektrons die gleichen statistischen Aussagen wie für jeweils einen Stoß vieler Elektronen gelten (ergodisch). Da sehr viele Elektronen den Querschnitt in einem Zeitintervall von der mittleren Dauer zwischen zwei Stößen passieren, ist die Statistik in jedem dieser Zeitintervalle erfüllt. Gemäß der Statistik besitzen die Ladungsträger unmittelbar nach dem Stoß keine Vorzugsrichtung und tragen daher im Mittel nicht zum Stromfluss bei. Daher braucht nur der Geschwindigkeitszuwachs der Ladungsträger zwischen den Stößen berücksichtigt zu werden. Unmittelbar nach dem Stoß beträgt er Null und wächst linear bis zum nächsten Stoß an. Abb. 2.18 links zeigt dies anhand eines Beispielverlaufs, der aus Abb. 2.17 hervorgeht, wenn der Startwert nach jedem Stoß zu Null gesetzt wird.

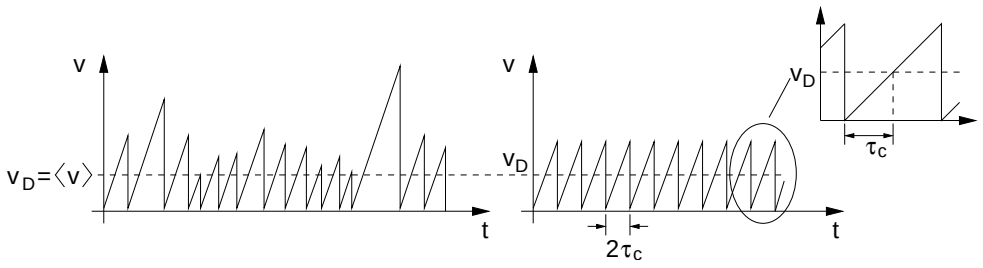

Abb. 2.18: Links: Zeitverlauf der Driftgeschwindigkeit eines Ladungsträgers. Rechts: Definition einer mittleren Stoßzeit.

Der Mittelwert

$$\mathrm{v}_D = \langle \mathrm{v} \rangle \tag{2.74}$$

dieses Geschwindigkeitsprofils ist die sog. Driftgeschwindigkeit. Sie ist identisch mit dem Mittelwert der Geschwindigkeit $\mathrm{v}_r = \langle \mathrm{v}_r^+ \rangle + \langle \mathrm{v}_r^- \rangle$ in Abb. 2.15. Die Driftgeschwindigkeit ist die Geschwindigkeit, mit der sich eine Ladungsträgerart unter Einfluss eines elektrischen Feldes im Mittel bewegt.
Auch für die Zeitachse ist es sinnvoll, mit Mittelwerten zu arbeiten. Wir definieren die mittlere Stoßzeit[16] τ_c als die Zeit, die ein Ladungsträger benötigt,

[16] Besser wäre der Begriff „mittlere Freiflug-Zeit“, da unter Stoßzeit irrtümlich die Dauer eines Stoßes verstanden werden könnte.

um von Null auf Driftgeschwindigkeit zu beschleunigen. Sie beträgt die Hälfte der mittleren Zeit zwischen zwei Stößen. Die Beschleunigung in diesem linearen Modell ist dann

$$\vec{a} = \frac{\vec{\mathrm{v}}_D}{\tau_c} . \tag{2.75}$$

Diese Beschleunigung findet in jedem der äquidistanten Zeitintervalle (vgl. Abb. 2.18 rechts) τ_c statt und lässt sich daher zu einer konstanten Beschleunigung über t fortsetzen. Wir setzen $\vec{a}$ in Gl. (2.73) ein und setzen diese mit Gl. (2.72) gleich und erhalten einen Zusammenhang zwischen elektrischer Feldstärke und Driftgeschwindigkeit

$$m^* \frac{\vec{\mathrm{v}}_D}{\tau_c} = q\vec{E} . \tag{2.76}$$

Die Driftgeschwindigkeit der Ladungsträger

$$\vec{\mathrm{v}}_D = \frac{\tau_c\, q}{m^*}\,\vec{E} = \pm\,\mu\vec{E} \quad , \quad \mu := \frac{\tau_c\, e}{m^*} \tag{2.77}$$

ist für Elektronen und Löcher unterschiedlich. Für Elektronen gilt

$$\vec{\mathrm{v}}_{Dn} = -\frac{\tau_{cn}\, e}{m^*_{ec}}\,\vec{E} = -\mu_n\vec{E} \tag{2.78}$$

mit der Beweglichkeit der Elektronen

$$\mu_n := \frac{\tau_{cn}\, e}{m^*_{ec}} . \tag{2.79}$$

Entsprechend gilt für Löcher

$$\vec{\mathrm{v}}_{Dp} = \frac{\tau_{cp}\, e}{m^*_{hc}}\,\vec{E} = \mu_p\vec{E} \tag{2.80}$$

mit der Beweglichkeit der Löcher

$$\mu_p := \frac{\tau_{cp}\, e}{m^*_{hc}} . \tag{2.81}$$

Wir können jetzt die Stromdichte $\vec{J}_F$ in der Fläche A des Halbleiters aufgrund des elektrischen Feldes berechnen. Sie setzt sich zusammen aus den Stromdichten der Elektronen und Löcher:

$$\vec{J}_F = \vec{J}_{Fn} + \vec{J}_{Fp} = \frac{1}{A}\left(I_{Fn} + I_{Fp}\right)\vec{e}_r \tag{2.82}$$

$$= \frac{1}{A}\left(\frac{-Q_n}{T_n} + \frac{Q_p}{T_p}\right)\vec{e}_r . \tag{2.83}$$

Darin ist Q_n, Q_p die gesamte Elektronen- und Löcherladung, die in der Zeit

$$T_n = \frac{L}{|\vec{\mathrm{v}}_{Dn}|} \text{ bzw. } T_p = \frac{L}{|\vec{\mathrm{v}}_{Dp}|} \tag{2.84}$$

den Halbleiter mit der Länge L von einer Elektrode zur anderen durchströmt (vgl. Abb. 2.16). Einsetzen von T_n und T_p ergibt

$$\vec{J}_F = \frac{1}{A}\left(\frac{-Q_n\,|\vec{\mathrm{v}}_{Dn}|}{L} + \frac{Q_p\,|\vec{\mathrm{v}}_{Dp}|}{L}\right)\vec{e}_r = e\,(n_0\,\vec{\mathrm{v}}_{Dn} + p_0\,\vec{\mathrm{v}}_{Dp})\,. \tag{2.85}$$

Darin sind $-en_0 = \frac{Q_n}{V}$ und $ep_0 = \frac{Q_p}{V}$ die Ladungsträgerdichten im Halbleitervolumen $V = A \cdot L$.
Einsetzen der Driftgeschwindigkeiten ergibt die wichtige Beziehung für den Feldstrom aufgrund eines elektrischen Feldes in einem Halbleiter:

$$\vec{J}_F = \underbrace{e\,n_0\,\mu_n\,\vec{E}}_{\vec{J}_{Fn}} + \underbrace{e\,p_0\,\mu_p\,\vec{E}}_{\vec{J}_{Fp}}\,. \tag{2.86}$$

Mit der Verwendung der Driftgeschwindigkeit für die Ladungsträger haben wir einen proportionalen Zusammenhang zwischen dem elektrischen Feld im Halbleiter und der, aus dessen Kraftwirkung resultierenden Geschwindigkeit angenommen. Im Folgenden werden wir sehen, dass diese Annahme nur für Feldstärken unterhalb der sog. Sättigungsfeldstärke gilt. Die Leitfähigkeit des Halbleiters ist definiert als der Quotient

$$\sigma := \left|\frac{\vec{J}_F}{\vec{E}}\right| = e\,(n_0\,\mu_n + p_0\mu_p) = \sigma_n + \sigma_p\,. \tag{2.87}$$

Damit lässt sich Gl. (2.86) verallgemeinert als das bekannte Ohmsche Gesetz

$$\vec{J}_F = \sigma\vec{E} \tag{2.88}$$

darstellen.

2.12.4 Beweglichkeit

Die Beweglichkeit wird durch die Art der Stöße, bzw. im Wellenmodell, durch die Art der Streuung, die der Ladungsträger erfährt, bestimmt. Eine Streuung führt immer zu einem Energieaustausch mit dem Streukörper sowie zu einem Richtungswechsel. Wir betrachten kurz die wichtigsten Arten der Streuung.

2.12.5 Streuung an Störstellen

Die Streuung an Störstellen ist besonders stark, wenn die Störstellen eine Ladung besitzen. Dies ist z. B. bei ionisierten Dotierungen der Fall. Die Stärke der auf elektrostatischen Kräften zwischen Ionen und Ladungsträger beruhenden Streuung nimmt mit der Dauer der Wechselwirkung und der Anzahl der beteiligten Ionen zu.
Daher führen hohe Dotierungen zu einer Abnahme der Beweglichkeit.
Die Dauer der Wechselwirkung nimmt ab mit steigender Temperatur, da die thermische Geschwindigkeit (vgl. Gl. (2.67)) der Ladungsträger mit der Temperatur zunimmt.[17]
Daher nimmt die Streuung mit der Temperatur ab und die Beweglichkeit steigt.
In erster Näherung ergibt sich aus den Abhängigkeiten eine Proportionalität der Beweglichkeit von

$$\mu \sim \frac{T^{\frac{3}{2}}}{N^{\pm}} , \tag{2.89}$$

wobei $N^{\pm}$ die Dichte der ionisierten Störstellen erfasst.

2.12.6 Gitterstreuung

Schwingungen des Gitters können durch Phononen beschrieben werden. Streuung am Kristallgitter bedeutet daher eine Wechselwirkung mit Phononen in Form einer Aufnahme oder Abgabe von Energie. Aus der Herleitung der Besetzungswahrscheinlichkeit wissen wir, dass die Dichte der Gitterschwingungen (Phononen) in einem Energieintervall mit der Temperatur zunimmt (vgl. Gl. (2.1)). Hierdurch sinkt die Streuzeit und damit die Beweglichkeit.
Theoretische Untersuchungen zeigen, dass bei **Si** und **Ge** die Beweglichkeit aufgrund der Streuung an akustischen Phononen proportional $T^{-\frac{3}{2}}$ von der Temperatur abhängt.
Bei Streuung an optischen Phononen ist eine $T^{-\frac{1}{2}}$-Abhängigkeit zu erwarten.
Experimentelle Ergebnisse der Temperaturabhängigkeit der Beweglichkeit aufgrund von Gitterstreuungen zeigt Tabelle 2.4. Im Bereich der Tem-

[17] Genauer müsste die relative Geschwindigkeit zwischen Ladungsträger und Ionen betrachtet werden. Wir nehmen in der Relation die Geschwindigkeit der im Gitter eingebauten Ionen zu Null an.

	Ge	**Si**	**GaAs**
μ_n	$\sim T^{-1,7}$	$\sim T^{-2,4}$	$\sim T^{-1,0}$
μ_p	$\sim T^{-2,3}$	$\sim T^{-2,2}$	$\sim T^{-2,1}$

Tabelle 2.4: Temperaturabhängigkeit der Beweglichkeit aufgrund von Phononenstreuung für verschiedene Halbleitermaterialien.

peraturen der Störstellenerschöpfung wird die Beweglichkeit meist von Gitterstreuung dominiert.

2.12.7 Oberflächenstreuung

Die Beweglichkeit von Ladungsträgern an Oberflächen und Grenzflächen wird durch die Beweglichkeit der Ladungsträger in dem angrenzenden Material bestimmt. Ursache hierfür ist die Unschärferelation, sowie die von Null verschiedenen Wellenfunktionen der Ladungsträger in der Umgebung des Ortes mit der höchsten Aufenthaltswahrscheinlichkeit. Daher befinden sich Ladungsträger an der Grenze eines Mediums z. T. auch in dem angrenzenden Medium. Die Wellenfunktionen können dabei z. B. im Bereich von 1-10 nm in den angrenzenden Bereich überlappen. Die resultierende Mobilität ist dann eine Kombination der Beweglichkeiten der beiden Bereiche.
Dies führt z. B. dazu, dass Ladungsträger in der Inversionschicht eines MOS-Feldeffekttransistors eine bis zu drei mal geringere Beweglichkeit wie im Halbleitermaterial selbst besitzen. Die Ursache liegt zum einen an der deutlich geringeren Beweglichkeit der Elektronen in dem darüber liegenden amorphen Silizium (Polysilizium). Zum anderen existieren geladene Oberflächenzustände, die der Beweglichkeit entsprechend der Wirkungsweise der ionisierten Störstellen reduzieren.

2.12.8 Werte für die Beweglichkeit

Die resultierende Beweglichkeit ergibt sich unter Berücksichtigung aller Streuprozesse zu

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i} \,. \tag{2.90}$$

Tabelle 2.5 gibt einige Werte der Beweglichkeit für p- und n-Dotierung in Abhängigkeit der Dotierungskonzentration bei Raumtemperatur an.

Abb. 2.19 zeigt den dazu gehörenden Verlauf in Abhängigkeit von der Temperatur.

$\frac{N}{cm^3}$	**As**	**P**	**B**
10^{15}	1359	1362	462
10^{16}	1177	1184	429
10^{17}	727	721	317
10^{18}	284	277	153
10^{19}	108	115	71

Tabelle 2.5: Beweglichkeit von Majoritäten in **Si** bei Raumtemperatur für verschiedene Dotierungen und Konzentrationen in $\frac{cm^2}{Vs}$.

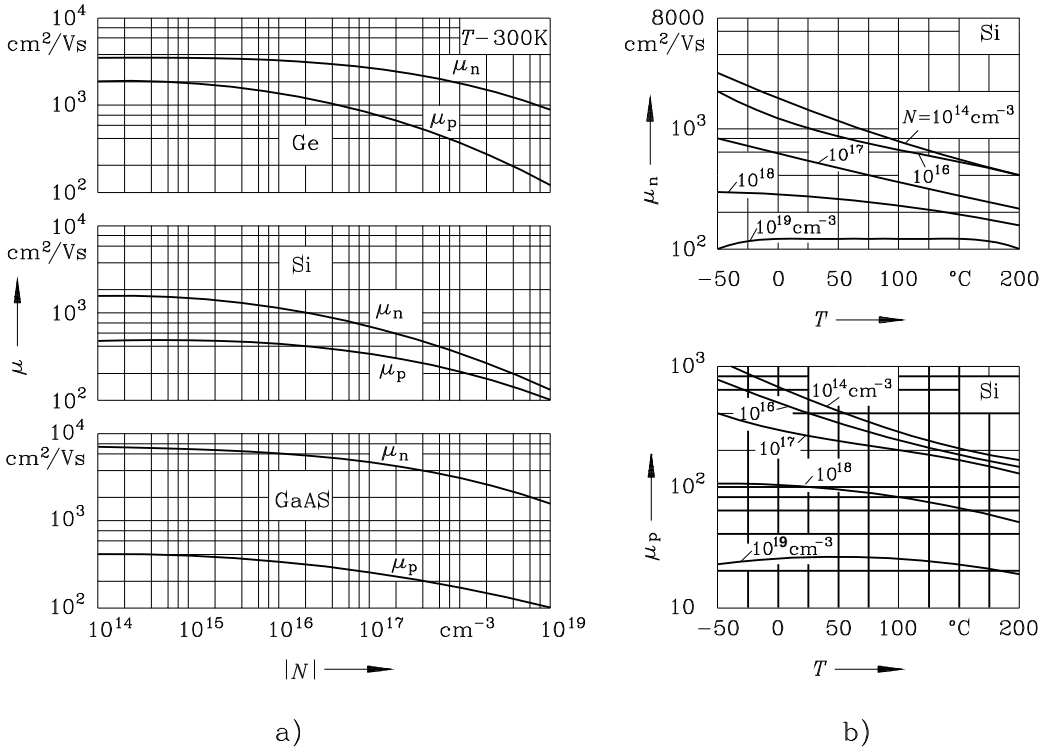

Abb. 2.19: Beweglichkeit von Elektronen und Löchern als Funktion der Dotierungskonzentration (links) und der Temperatur (rechts).

2.12.9 Heiße Ladungsträger, Sättigungsgeschwindigkeit

Bei der Ermittlung der Driftgeschwindigkeit haben wir die mittlere Stoßzeit τ_c definiert. Sie wurde bestimmt durch die Hälfte des zeitlichen Mittelwertes zwischen zwei Stößen. Die Zeit zwischen zwei Stößen wird durch die Gesamtgeschwindigkeit aus thermischer Geschwindigkeit und Driftgeschwindigkeit

$$\mathrm{v} = \mathrm{v}_{th} + \mathrm{v}_D \tag{2.91}$$

der Ladungsträger bestimmt. Solange $\mathrm{v}_{th} \gg \mathrm{v}_D$, ist die Stoßzeit unabhängig von einem angelegten Feld und damit konstant.
Bei sehr hohen Feldstärken erreicht die Driftgeschwindigkeit der Ladungsträger die Größenordnung der thermischen Geschwindigkeit und kann nicht mehr vernachlässigt werden. Dadurch steigt die Gesamtgeschwindigkeit der Ladungsträger durch das elektrische Feld an. Als Folge werden die mittlere Stoßzeit und die Beweglichkeit abhängig von der Feldstärke.
Da sich die Gesamtgeschwindigkeit mit steigender Feldstärke erhöht, nimmt die mittlere Stoßzeit τ_c und wegen Gl. (2.77) auch die Beweglichkeit ab. In diesem Bereich gilt das Ohmsche Gesetz nach Gl. (2.88) nicht mehr.
In der Beschreibung im Wellenmodell entsteht die Abnahme der Geschwindigkeit dadurch, dass die Energie der Ladungsträger durch das elektrische Feld steigt. Ist das elektrische Feld so stark, dass die Energie der Ladungsträger größer wird als die Energie der optischen Phononen, steigt die Wahrscheinlichkeit, Energie an ein optisches Phonon abzugeben, abrupt an. Dies führt dazu, dass die Geschwindigkeit der Ladungsträger abnimmt und bei höheren Feldstärken in einen Sättigungsverlauf, wie in Abb. 2.20 links gezeigt, übergeht. Der Maximalwert der Driftgeschwindigkeit ist die sog. Sättigungsgeschwindigkeit v_{sat}.
Die Geschwindigkeit von Ladungsträgern in Materialien wie **Si**, die keine erreichbaren höheren Bänder bzw. Bandstrukturen, wie z. B. bei **GaAs** besitzen, kann beschrieben werden durch

$$\mathrm{v}_D(E) = \frac{\mu\, E}{1 + \frac{\mu\, E}{\mathrm{v}_{sat}}} \,. \tag{2.92}$$

Wir merken uns den Übergang der Driftgeschwindigkeit in die konstante Sättigungsgeschwindigkeit bei hohen Feldstärken. Dieser Übergang wird uns später bei der Behandlung des Feldeffekt-Transistors im Abschnürbereich begegnen.

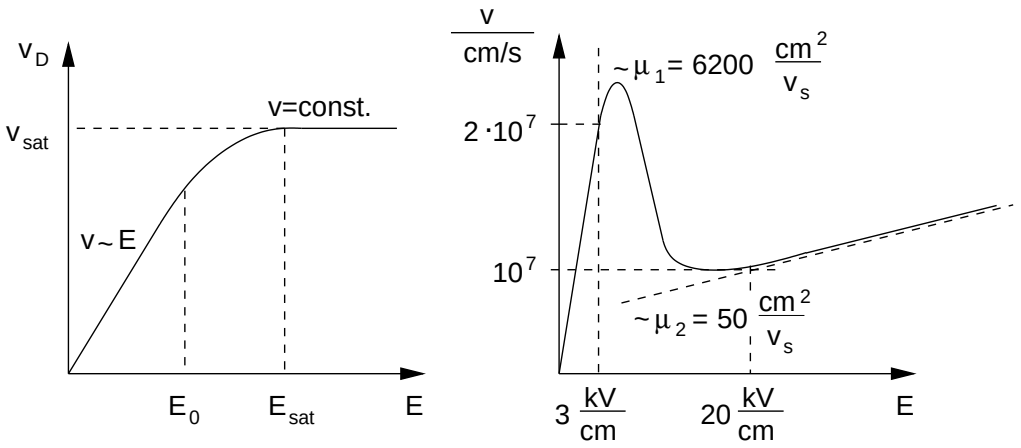

Abb. 2.20: **Links**: Abhängigkeit der Driftgeschwindigkeit v_D von der Größe des elektrischen Feldes. Bei $E = E_0$ geht die linerare Abhängigkeit in einen Sättigungsverlauf über. Bei $E = E_{sat}$ ist die Sättigungsgeschwindigkeit v_{sat} erreicht. **Rechts**: Feldabhängigkeit der Driftgeschwindigkeit bei **GaAs**.

2.12.10 Gunn-Effekt, velocity overshoot

Bei Halbleitermaterialien wie **GaAs** oder **InP** tritt ebenfalls eine Abweichung von einem linearen $v_D(E)$-Zusammenhang bei hohen Feldstärken auf. Ursache ist jedoch nicht eine verringerte mittlere Stoßzeit, sondern eine vergrößerte effektive Masse. Dies führt nach Gl. (2.78) und (2.79) ebenfalls zu einer Verringerung von Driftgeschwindigkeit und Beweglichkeit.
Die Ursache für die bei hohen Feldstärken vergrößerte effektive Masse liegt in der Struktur des Leitungsbandes dieser Halbleiter. Betrachten wir z. B. die Bandstruktur von **GaAs** in Abb. 1.40, so sehen wir, dass um $\Delta W = 0,36\,\mathrm{eV}$ über dem Hauptminimum, das den Bandabstand von 1,43 eV bildet, ein Nebenminimum liegt. Die Krümmung des Nebenminimums ist geringer als die des Hauptminimums. Nach Gl. (1.140) ist damit die Masse der Elektronen im Nebenminimum größer als die im Hauptminimum.

Bei hohen Feldstärken nehmen die Elektronen im Hauptminimum so viel Energie auf, dass sie die Energiedifferenz $\Delta W = 0,36\,\mathrm{eV}$ überwinden und in das Nebenminimum übergehen (Intrabandübergang/-streuung). Hier besitzen sie eine größere Masse und damit eine geringere Beweglichkeit und Driftgeschwindigkeit. Die Abnahme von Beweglichkeit und Driftgeschwindigkeit beim Übergang zu hohen Feldstärken zeigt Abb. 2.20

rechts. Die Spitze in der Geschwindigkeitskurve wird als sog. „velocity overshoot“ bezeichnet. Die Änderung der effektiven Masse durch Streuung der Elektronen in ein Nebenminimum mit größerer Masse wird nach seinem Entdecker als „Gunn“-Effekt bezeichnet.
Eine wesentliche Eigenschaft des Gunn-Effektes ist die negative differentielle Beweglichkeit beim Übergang in das Nebenminimum. Unter bestimmten Voraussetzungen, die über den Rahmen dieser Vorlesung hinausgehen, können durch Ausnutzen dieses Effektes Verstärker, Oszillatoren und Impulsgeneratoren für die Mikrowellentechnik realisiert werden.

2.12.11 Diffusion von Ladungsträgern

Bisher haben wir die Bewegung von Ladungsträgern verursacht durch die Kraftwirkung eines elektrischen Feldes betrachtet. Bei der Diffusion von Ladungsträgern ist die Ursache der Bewegung ausschließlich die thermische Energie der Ladungsträger von $\frac{1}{2}\,kT$ je Freiheitsgrad (x, y, z). Diese thermische Energie ist der Motor des Diffusionsprozesses. Bei $T = 0$ findet daher keine Diffusion statt.
Wir haben bereits in Kap. 2.12.1 gesehen, dass der Ladungsträger aufgrund dieser thermischen Bewegung zwischen zwei Stößen eine mittlere Geschwindigkeit v_{th} besitzt. Die Geschwindigkeit gemittelt über eine ausreichende Zahl von Stößen oder eine ausreichende Zahl von Ladungsträgern war jedoch Null, da zu jeder Geschwindigkeitskomponente $\langle \mathrm{v}_x^+ \rangle$ auch Ladungsträger mit einem Mittelwert der Geschwindigkeit $\langle \mathrm{v}_x^- \rangle$ in entgegengesetzter Richtung existieren. Die Beiträge durch die entgegengesetzen Komponenten heben sich auf, wodurch der Stromfluss gleich Null ist.
Diese Aussage gilt jedoch für den gesamten Halbleiter nur dann, wenn in diesem eine homogene Ladungsträgerverteilung herrscht. Bei inhomogener Verteilung, also bei unterschiedlicher Ladungsträgerdichte im Volumen des Halbleiters, kommt es zu einer Diffusion der Ladungsträger. Sie fließen dabei, getrieben von der thermischen Energie vom Ort der Anhäufung (höherer Dichte), in Richtung der niedrigeren Dichte. Das Bestreben des Auseinanderfließens hat eine Gleichverteilung der Ladungsträger über das gesamte Volumen zum Ziel. Im Zweidimensionalen lässt sich dieser Vorgang mit dem Auseinanderfließen eines Sandhaufens auf einer Rüttelplatte vergleichen. Der Rüttelvorgang entspricht der Anregung der Ladungsträger durch die thermische Energie.

Wir wollen eine mathematische Beschreibung für den Diffusionsvorgang herleiten. Dabei beschränken wir uns auf den wichtigsten Fall der eindimensionalen Dichteverteilung. Wir nehmen dabei an, dass sich die Dichte der Ladungsträger nur in Richtung der x-Koordinate ändert, wobei die Wahl der x-Koordinate hier willkürlich erfolgt.
Aufgrund dieser Wahl ist die Dichte in y- und z-Richtung konstant, so dass $n(x, y, z) = n(x)$ gilt. Wir führen die folgenden Überlegungen zur Vereinfachung für eine allgemeine Ladungsträgerdichte $n(x)$ durch und schließen aus den Ergebnissen im Anschluss auf die Diffusion von Elektronen und Löchern. Abb. 2.21 links zeigt beispielhaft einen Dichteverlauf in x-Richtung mit willkürlich gewähltem Nullpunkt.

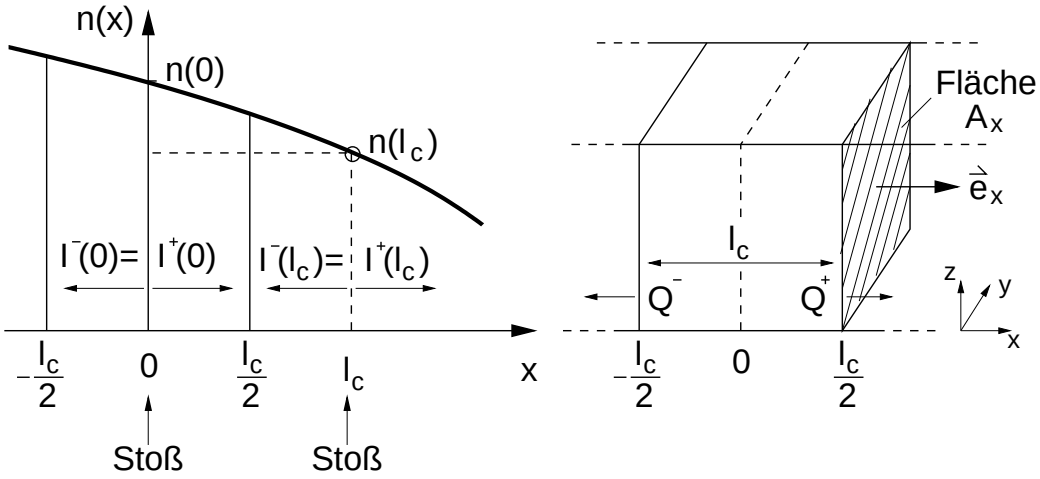

Abb. 2.21: Links: Ladungsträgerdichte $n(x)$ in Abhängigkeit vom Ort. Stöße finden am Ort $x = 0$, l_c statt. Die Ströme in $+x$- und $-x$-Richtung an jedem der beiden Orte sind gleich.
Rechts: Ein Volumenelement mit homogener Ladungsträgerdichte auf der Fläche A_x in der y-z-Ebene in Richtung $\vec{e}_x$.

Da die thermische Energie eine Zufallsgröße darstellt, erfordert eine allgemeine mathematische Behandlung der thermischen Anregung der Ladungsträger die Beschreibung durch einen Zufallsprozess. Es zeigt sich anhand der daraus erhaltenen Ergebnisse, dass das Arbeiten mit Mittelwerten für unsere Fragestellungen zu den gleichen Ergebnissen führt.
Wir verwenden daher die bereits zuvor bei der Betrachtung von Ladungsträgern im elektrischen Feld definierte mittlere Stoßzeit (Freiflugzeit) τ_c. Die mittlere Geschwindigkeit der Ladungsträger ist gleich der thermischen Ge-

schwindigkeit. Wir ermitteln über τ_c und die thermische Geschwindigkeit der Ladungsträger die mittlere (freie) Weglänge zwischen zwei Stößen

$$l_c = \mathrm{v}_{th}\,\tau_c\ . \tag{2.93}$$

In unserem auf Mittelwerten basierenden Modell bewegt sich der Ladungsträger auf einer Strecke l_c frei, d. h. unbeeinflusst von jeglicher Wechselwirkung. Wir legen den Nullpunkt der x-Achse an die Stelle eines Stoßes. Dann erfolgt der nächste Stoß (im Mittel)[18] an der Stelle $x = l_c$.
Zur Herleitung einer Bilanzgleichung (Kontinuitätsgleichung) für die Ladungsträgerdichte in einem Volumenelement betrachten wir die Stromdichte am Ort $x = \frac{l_c}{2}$ zwischen den beiden Stoßzentren.
Vom Stoßzentrum ($x = 0$) links von unserem Betrachtungsort erfolgt ein Beitrag

$$\vec{J}_x^{+}(0) = \frac{I^{+}(0)}{A_x}\,\vec{e}_x\ , \tag{2.94}$$

der aufgrund des Zufallscharakters des Stoßvorgangs auch in gleicher Größe in $-\vec{e}_x$-Richtung erfolgt. Zu einer Aussage über die y- und z-Richtung gelangen wir später sehr einfach.
Wir betrachten den Strom über die Dauer der halben mittleren Stoßzeit τ_c, in der eine Ladung $Q^{+}(0) = \frac{1}{2}\tau_c\, I^{+}(0)$ vom Stoßzentrum bei $x = 0$ nach rechts durch die Fläche A_x an die Stelle $x = \frac{l_c}{2}$ transportiert wird (vgl. Abb. 2.21 rechts) und schreiben

$$\vec{J}_x^{+}(0) = \frac{Q^{+}(0)}{A_x\,\frac{\tau_c}{2}}\,\vec{e}_x \tag{2.95}$$

und mit Gl. (2.93)

$$\vec{J}_x^{+}(0) = \frac{Q^{+}(0)\,\mathrm{v}_{th}}{A_x\,l_c\frac{1}{2}}\vec{e}_x = q\,n(0)\,\mathrm{v}_{th}\,\vec{e}_x\ . \tag{2.96}$$

Darin bezeichnet $n(0)$ die Ladungsträgerdichte in dem Volumenelement $\frac{1}{2}\,A_x\,l_c$, die wegen $Q^{+}(0) = Q^{-}(0)$ identisch mit der Ladungsträgerdichte im Volumenelement $A_x\,l_c$ ist.
Analog ergibt sich ein weiterer Beitrag zur Stromdichte bei $x = \frac{l_c}{2}$ von der aus dem Stoßzentrum bei $x = l_c$ nach links strömenden Ladung in Höhe von

$$\vec{J}_x^{-}(l_c) = q\,n(l_c)\,\mathrm{v}_{th}(-\vec{e}_x)\ . \tag{2.97}$$

[18] Wir lassen im Folgenden diesen erläuternden Einschub weg und berücksichtigen weiterhin, dass es sich bei den betrachteten Größen um Mittelwerte handelt.

Die resultierende Stromdichte bei $x = \frac{l_c}{2}$ ist dann

$$\vec{J}_x\left(\frac{l_c}{2}\right) = \vec{J}_x^{+}(0) + \vec{J}_x^{-}(l_c) \tag{2.98}$$

$$= q\,\mathrm{v}_{th}\,(n(0) - n(l_c))\,\vec{e}_x \tag{2.99}$$

$$= -q\,\mathrm{v}_{th}\,l_c\,\frac{n(l_c) - n(0)}{l_c}\,\vec{e}_x\ . \tag{2.100}$$

Da l_c die in unserem Modell verwendete kleinstmögliche Quantisierung des Ortes ist, interpretieren wir den Quotiententerm als Differentialquotienten und schreiben allgemein

$$\vec{J}_x(x) = -q\,\mathrm{v}_{th}\,l_c\,\frac{d\,n(x)}{d\,x}\,\vec{e}_x = -q\,D\,\frac{d\,n(x)}{d\,x}\,\vec{e}_x \tag{2.101}$$

$$\text{mit } D := \mathrm{v}_{th}\,l_c \quad \text{Diffusionskonstante.} \tag{2.102}$$

Dies ist die Formulierung des Diffusionsstromes aufgrund einer ortsabhängigen Ladungsträgerdichte $n(x)$.
Da wir uns auf eindimensionale Dichteprofile $n(x)$ beschränken, ist $\frac{dn}{dy} = 0$ und $\frac{dn}{dz} = 0$, wodurch $\vec{J}_y = 0$ und $\vec{J}_z = 0$ folgen.
Daher existiert nur eine Komponente der Stromdichte in Richtung der Ladungsträgerdichte-Änderung. Der Stromfluss ist dabei immer so gerichtet, dass es durch den Ladungstransport zu einem Abbau von Dichteunterschieden kommt.

Wir haben in Gl. (2.102) allgemein eine Diffusionskonstante definiert. Diese lässt sich auf die Beweglichkeit der Ladungsträger zurückführen, was wir im Folgenden zeigen.
Dazu schätzen wir die thermische Geschwindigkeit ab, indem wir berücksichtigen, dass die thermische Energie je Freiheitsgrad $\frac{1}{2}\,k\,T$ beträgt. Für eine eindimensionale Bewegung in Richtung des Stromflusses (hier x-Richtung) gilt dann

$$\frac{1}{2}\,k\,T = \frac{1}{2}\,m^*\,\mathrm{v}_{th}^2\ . \tag{2.103}$$

Darin ist m^* die effektive Masse der Ladungsträger.
Mit v_{th} aus Gl. (2.103) und l_c aus Gl. (2.93) lässt sich die Diffusionskonstante schreiben

$$D = \mathrm{v}_{th}\,l_c = \mathrm{v}_{th}^2\,\tau_c = k\,T\,\frac{\tau_c}{m^*} = \frac{k\,T}{e}\,\mu\ . \tag{2.104}$$

Für die letzte Umformung wurde die Definition der Beweglichkeit nach Gl. (2.77) verwendet.
Es ist gebräuchlich, den Ausdruck

$$\frac{k\,T}{e} := U_T \tag{2.105}$$

als Temperaturspannung U_T zu definieren.

Es ergeben sich dann die Diffusionskonstanten für Elektronen und Löcher (Nernst-, Townsend- und Einstein-Beziehungen)

$$D_n = \frac{k\,T}{e}\,\mu_n = U_T\,\mu_n \tag{2.106}$$

$$D_p = \frac{k\,T}{e}\,\mu_p = U_T\,\mu_p \tag{2.107}$$

mit den dazugehörigen Diffusionsströmen

$$\vec{J}_{Dn}(x) \;=\; e\,D_n\,\frac{d\,n}{d\,x}\,\vec{e}_x \tag{2.108}$$

$$\vec{J}_{Dp}(x) \;=\; -e\,D_p\,\frac{d\,p}{d\,x}\,\vec{e}_x\;. \tag{2.109}$$

Die Ortsangabe x lassen wir zur Übersichtlichkeit im Folgenden weg und versehen die Stromdichten mit dem Index D für „Diffusion“.

Wir erkennen, dass die Beweglichkeit der Ladungsträger einen entscheidenden Einfluss auf den Vorgang des Stromflusses hat. Sie erscheint sowohl in den Drift- als auch in den Diffusionsgleichungen als Proportionalitätskonstante der Ursache ($\vec{E}$ bei Drift, $\frac{d\,n}{d\,x}$ bei Diffusion).

Beachten: Die Herleitung der Diffusionsgleichungen nimmt als Ursache nur die örtliche Änderung der Ladungsträgerdichte. Entsteht aufgrund dieser Änderung ein elektrisches Feld, so bewirkt dies zusätzlich einen Feld-Strom (Driftstrom).

2.12.12 Strom-Transportgleichungen

Wir haben zwei voneinander unabhängige Ursachen für einen Ladungsträgerstrom ermittelt. Im vorangegangenen Kapitel war die Ursache eine ortsabhängige Ladungsträgerdichte. Sie erzeugt den Diffusionsstrom, der

unabhängig von einem elektrischen Feld ist. Ein elektrisches Feld erzeugt den Feld- oder Driftstrom. Das Feld kann dabei entweder von außen an den Halbleiter gelegt werden oder durch eine ortsabhängige Ladungsverteilung hervorgerufen werden.

Der Gesamtstrom für Elektronen und Löcher ergibt sich aus der Überlagerung von Feld- und Diffusionsstrom aus Gl. (2.86) bzw. (2.108) und (2.109):

$$\vec{J}_n = \vec{J}_{Fn} + \vec{J}_{Dn} = e\,n\,\mu_n\,\vec{E} + e\,D_n\,\frac{d\,n}{d\,x}\,\vec{e}_x \tag{2.110}$$

$$\vec{J}_p = \vec{J}_{Fp} + \vec{J}_{Dp} = e\,p\,\mu_p\,\vec{E} - e\,D_p\,\frac{d\,p}{d\,x}\,\vec{e}_x\ . \tag{2.111}$$

Wir betrachten im Folgenden nur elektrische Felder in x-Richtung. Wir verzichten daher zur Vereinfachung der Schreibweise auf die Vektordarstellung und betrachten J_n und J_p als Komponenten der Stromdichte in $+x$-Richtung.

$$J_n = e\,n\,\mu_n\,E + e\,D_n\,\frac{d\,n}{d\,x} \tag{2.112}$$

$$J_p = e\,p\,\mu_p\,E - e\,D_p\,\frac{d\,p}{d\,x} \tag{2.113}$$

E zeigt in dieser Definition in $+x$-Richtung. Für die entgegengesetzte Richtung ist ein negatives Vorzeichen zu wählen.
Der Gesamtstrom im Halbleiter ergibt sich aus der Überlagerung von Elektronen und Löcherstrom

$$J = J_n + J_p\ . \tag{2.114}$$

2.12.13 Kontinuitätsgleichung

Die Fähigkeit, das elektrische Verhalten von Halbleiterbauelementen beschreiben zu können, geht einher mit der Fähigkeit, das Verhalten der Ladungsträger im Halbleiter zu beschreiben. Für das thermodynamische Gleichgewicht haben wir bereits mit Gl. (2.52) die wichtigste Neutralitätsbedingung als Bilanzgleichung aufgestellt.
Außerhalb des thermodynamischen Gleichgewichts fließen die Ladungsträger durch den Halbleiter (Ladungsträger werden bewegt). Der Vorgang des Stromflusses ist in der zuvor abgeleiteten Transportgleichung beschrieben. Wir wissen daher, dass ein elektrisches Feld oder ein Dichtegradient die Ursache für einen Ladungsträgertransport und damit für einen Stromfluss sind.

Durch den Stromfluss befindet sich der Halbleiter außerhalb des thermodynamischen Gleichgewichts, wodurch nach Gl. (2.68) und (2.69) eine Abweichung der Ladungsträgerdichten von den Gleichgewichtsdichten und damit von der Ladungsneutralität resultiert. Die Abweichung kann dabei lokal (Anhäufung von Ladungsträgern) oder global (homogener Halbleiter in elektrischem Feld) sein.
Über die Erzeugung und Vernichtung (Generation und Rekombination) von Ladungsträgern haben wir bisher noch keine Aussage gewonnen. Auch der Aspekt der Zeitabhängigkeit von Vorgängen, die mit Ladungsträgern verknüpft sind, wurde bisher nicht beschrieben.
Das wollen wir im Folgenden nachholen, indem wir eine vereinheitlichte Beschreibung in Form einer orts- und zeitabhängigen Bilanzgleichung für Ladungsträger aufstellen. Wir verwenden dazu die erste Maxwellsche Gleichung, die in allgemeingültiger Form Auskunft über die Ursache für einen Strom von Ladungsträgern gibt:

$$\vec{J} = \text{rot } \vec{H} - \frac{\partial \vec{D}}{\partial t} . \tag{2.115}$$

Danach können die Wirbel (rot) eines magnetischen Feldes $\vec{H}$ oder ein zeitlich veränderliches elektrisches Feld $\vec{E} = \frac{1}{\varepsilon}\vec{D}$ die Ursache sein. Das Bilden der Divergenz gibt die Quellen der Stromdichte an:

$$\text{div } \vec{J} = -\text{ div } \frac{\partial \vec{D}}{\partial t} = -\frac{\partial}{\partial t} \text{div } \vec{D} . \tag{2.116}$$

Dabei wird die Zulässigkeit der Vertauschung von Orts- und Zeitdifferentiation vorausgesetzt. Der Ausdruck div $\vec{D}$ steht für die Quellen des elektrischen Feldes.

Wir wissen durch die dritte Maxwellsche Gleichung

$$\rho = \text{div } \vec{D} = \text{div } \varepsilon\vec{E} = \varepsilon \text{ div } \vec{E} \tag{2.117}$$

mit $\varepsilon = const.$, dass die Quellen des elektrischen Feldes aus einer Raumladungsdichte ρ gebildet werden. Für unsere eindimensionale Betrachtung geht die Maxwellsche Gleichung wegen div $\vec{D}(x) = \frac{d}{dx}D$ über in

$$\rho = \frac{d\,D}{d\,x} = \varepsilon \frac{d\,E}{d\,x} . \tag{2.118}$$

Dabei ist die Raumladungsdichte $\rho = \rho(x)$ auch ortsabhängig. Generell wird

die Raumladungsdichte an einem Ort immer durch die Summe aller Ladungsdichten an diesem Ort gebildet. Daher gilt immer

$$\rho = e\,(p + N_D^+ - n - N_A^-)\ . \tag{2.119}$$

Je nach Bedarf kann ρ auch alternativ mit Gl. (2.68) und (2.69) geschrieben werden als

$$\rho = e\,(p_0 + \Delta p + N_D^+ - n_0 - \Delta n - N_A^-) \tag{2.120}$$

$$\rho = e(\Delta p - \Delta n)\ , \tag{2.121}$$

wobei für die alternative Darstellung im zweiten Schritt die Neutralitätsbedingung verwendet wurde.
Für alle Überlegungen im Rahmen dieser Vorlesung wird der Halbleiter im Bereich der Störstellenerschöpfung betrieben. Das bedeutet, dass alle Dotierungsatome ionisiert sind ($N_D^+ = N_D$, $N_A^- = N_A$) und es auch bleiben. Daher gilt

$$\frac{\partial\, N_D^+}{\partial\, t} = 0\ , \qquad\qquad \frac{\partial\, N_A^-}{\partial\, t} = 0\ . \tag{2.122}$$

Dies berücksichtigen wir, wenn wir in Gl. (2.116) für die Quellen des elektrischen Feldes die Raumladung entsprechend Gl. (2.118) mit den beiden alternativen Darstellungen nach Gl. (2.119) und (2.121) einsetzen:

$$\text{div}\ \vec{J} = -\frac{\partial}{\partial\, t}\,\rho = -e\,\frac{\partial}{\partial\, t}(p - n) = -e\,\frac{\partial}{\partial\, t}\,(\Delta p - \Delta n)\ . \tag{2.123}$$

Wir werden (nur) in diesem Kapitel mit beiden alternativen Darstellungen arbeiten. Für die spätere Arbeit mit den Gleichungen kann dann die für das jeweilige Problem passende Darstellung verwendet werden.
Gleichung (2.123), die wir erhalten haben, ist die Kontinuitätsgleichung für Raumladungen. Wir werden sie im weiteren Verlauf genauer diskutieren und erläutern. Die von uns benötigte eindimensionale Darstellung lautet

$$\frac{d}{dx}\vec{J}(x,t) = -\frac{\partial}{\partial\, t}\rho(x,t) = -e\,\frac{\partial}{\partial\, t}(p(x,t) - n(x,t)) = -e\,\frac{\partial}{\partial\, t}(\Delta p(x,t) - \Delta n(x,t))\ . \tag{2.124}$$

Wir wissen, dass sich die Stromdichte auf der linken Seite der Kontinuitätsgleichung aus einem Elektronen- und einem Löcherstrom (vgl. Gl. (2.114)) zusammensetzt. Indem wir dies explizit schreiben, erhalten wir aus Gl. (2.124)

eine Bilanzgleichung für die Quellen des Elektronen- und Löcherstroms (zunächst wieder in der allgemeinen dreidimensionalen Darstellung)

$$\operatorname{div}\vec{J}_n + \operatorname{div}\vec{J}_p = e\frac{\partial n}{\partial t} - e\frac{\partial p}{\partial t} = e\frac{\partial \Delta n}{\partial t} - e\frac{\partial \Delta p}{\partial t} . \tag{2.125}$$

Die Angabe der Orts- und Zeitabhängigkeit aller darin enthaltenen Größen lassen wir der Übersichtlichkeit wegen weg.
Wir können diese Bilanzgleichung in zwei separate Bilanzgleichungen für Elektronen und Löcher aufspalten:

$$\frac{1}{e}\operatorname{div}\vec{J}_n = \frac{\partial n}{\partial t} + R = \frac{\partial \Delta n}{\partial t} + R \tag{2.126}$$

$$\frac{1}{e}\operatorname{div}\vec{J}_p = -\frac{\partial p}{\partial t} - R = -\frac{\partial \Delta p}{\partial t} - R , \tag{2.127}$$

die durch Überlagerung (Addition) wieder die Gesamtbilanz nach Gl. (2.125) ergeben. In den separaten Gleichungen ist hier zunächst rein formal eine Funktion $R = R(\vec{r}, t)$ eingeführt worden, da die Aufspaltung diesen Freiheitsgrad erlaubt, denn $R(\vec{r}, t)$ verschwindet durch die Überlagerung der beiden Ladungsträgerströme identisch.
Wir nennen $R(\vec{r}, t)$ die „Nettorekombinationsrate“ (oder auch den „Rekombinationsüberschuss“). Sie ist maßgeblich bestimmend für den Stromfluss in Halbleiterbauelementen und wird ausführlich im folgenden Kapitel behandelt. An dieser Stelle können wir schon folgende Eigenschaften von R anhand der Art der Einbettung in die Bilanzgleichungen feststellen:

1. Die Bilanzgleichungen für Elektronen und Löcher sind nicht unabhängig. Sie werden über die Nettorekombinationsrate gekoppelt. (Die direkte Kopplung über R ist das Resultat der Annahme nach Gl. (2.122).)

2. Die Einheit von R ist $\frac{1}{\mathrm{cm^3 s}}$, also Ladungsträgerdichte pro Sekunde. Daher erläutert sich der Begriff „Rate“ in der Bezeichnung Nettorekombinationsrate. Es wird demnach die Änderung einer Ladungsträgerdichte pro Zeiteinheit beschrieben. Es lässt sich daher immer für R eine formale Darstellung in der Form

$$R = \frac{\Delta n}{\tau} \tag{2.128}$$

verwenden. Darin repräsentiert Δn eine allgemeine Änderung einer Ladungsträgerdichte (Elektronen sowie Löcher). Auf die physikalische Repräsentation der darin enthaltenen Größen werden wir in Kapitel 2.21 zurückkommen.

3. Stellen wir uns die Divergenz der Stromdichten in der Integralform vor, so besagt die linke Seite von Gl. (2.126) und (2.127), dass durch die Hüllfläche des Integrals ein Strom entsprechend der darin enthaltenen Quellen oder Senken, die auf der rechten Seite der Gleichungen stehen, fließt. Die Nettorekombination R beschreibt also einen Vorgang, der innerhalb des betrachteten Halbleiterbereichs (im allgemeinen ein Volumenelement) abläuft. Wegen Punkt 2. besteht dieser Vorgang in der Erzeugung bzw. Vernichtung (Generation, Rekombination) von Ladungsträgern.

Allgemein können wir formal immer annehmen, dass eine Generation g und eine Rekombination r stattfinden, aus der die Nettorekombinationsrate hervorgeht:

$$R = r - g \; . \tag{2.129}$$

Wir schreiben damit Gl. (2.126) und (2.127) in der allgemein gebräuchlichen Form der Kontinuitätsgleichungen für beide Ladungsträgerdichten:

$$\frac{1}{e} \operatorname{div} \vec{J}_n \; = \; r - g \; + \; \frac{\partial \, n}{\partial \, t} \quad = \quad r - g \; + \; \frac{\partial \, \Delta n}{\partial \, t} \tag{2.130}$$

$$-\frac{1}{e} \operatorname{div} \vec{J}_p \; = \; r - g \; + \; \frac{\partial \, p}{\partial \, t} \quad = \quad r - g \; + \; \frac{\partial \, \Delta p}{\partial \, t} \; . \tag{2.131}$$

Bei der in dieser Vorlesung angenommenen eindimensionalen Ortsabhängigkeit geht die Bildung der Divergenz in die Ortsableitung nach x über und wir erhalten die eindimensionalen Kontinuitätsgleichungen in der Form, die wir im Folgenden verwenden werden:

$$\frac{1}{e} \frac{dJ_n}{dx} = \; R \; + \; \frac{\partial \, n}{\partial t} \quad = \quad R \; + \; \frac{\partial \, \Delta n}{\partial t} \; , \qquad R = r - g \tag{2.132}$$

$$-\frac{1}{e} \frac{dJ_p}{dx} = \; R \; + \; \frac{\partial \, p}{\partial t} \quad = \quad R \; + \; \frac{\partial \, \Delta p}{\partial t} \; , \qquad R = r - g \; . \tag{2.133}$$

Die linke Seite lässt sich, wie in Abb. 2.22 gezeigt, als die Differenz zwischen dem herausfließenden und dem hineinfließenden Strom in ein Volumenelement $A\,dx$ interpretieren. Dabei ist A die Fläche, durch die der Strom homogen fließt. Die linke Seite liefert also die Ladungsträgerdichte pro Zeiteinheit, die in dem Volumen verbleibt. Die in dem Volumen verbleibende Ladungsträgerdichte entspricht der darin netto rekombinierten Ladungsträgerdichte (Rekombinierte minus Generierte) und der zeitlichen Änderung der darin enthaltenen Ladungsträgerdichte.

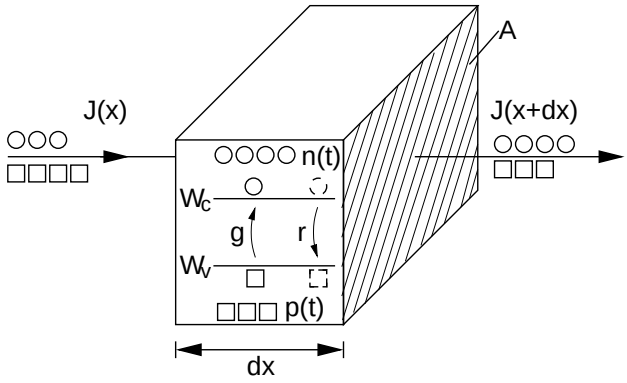

Abb. 2.22: Ladungsträger-Bilanz der Kontinuitätsgleichungen für ein Volumenelement $A\,dx$. g deutet die Generation von Ladungsträgern in Form eines Elektronen-Lochpaares an. r zeigt die Vernichtung (Rekombination) des Elektronen-Lochpaares, wobei das Elektron aus dem Leitungsband den Platz eines Lochs im Valenzband besetzt.

Die Kontinuitätsgleichung ist eine der wichtigsten Bilanzgleichungen für die Berechnung von Halbleiterbauelementen. Sie gilt an jedem Ort und zu jeder Zeit. Daher sind alle darin enthaltenen Größen $\{n, p, g, r, J_p, J_n\}$ Funktionen von Ort und Zeit. Dies erlaubt die Berechnung von zeitabhängigen Abläufen wie z. B. Ausgleichsvorgängen bei Störung der Neutralitätsbedingung.
Wir werden dazu im Kapitel 2.22 ein Beispiel rechnen.

2.13 Generation von Ladungsträgern

Wir haben uns bereits mit der Generation von Ladungsträgern bei der Herleitung der Fermi-Dirac-Verteilungsfunktion beschäftigt. Dort war die thermische Energie kT die Energie, die ein Elektron aus dem Valenzband über die Bandlücke W_g in das Leitungsband gehoben hat. Neben der Zufuhr thermischer Energie kann dem Elektron auch „optische“ Energie durch Bestrahlung

mit Licht in Form von Photonen zugeführt werden. Dieser Vorgang dominiert in Photodioden und Solarzellen.
Wir betrachten im Folgenden den Vorgang der Energiezufuhr auf Ladungsträger auf allgemeiner Ebene etwas genauer. Wir beschränken unsere Betrachtungen der Einfachheit halber auf Elektronen. Für Löcher ergeben sich die gleichen Überlegungen, wobei zu beachten ist, dass die Energieskala der Löcher nach unten steigt.
Wir nehmen im Folgenden an, dass einem Elektron eine Energie

$$\Delta W = h \cdot f \tag{2.134}$$

z. B. durch ein Photon zugeführt wird. Vor dem Treffen mit dem Photon soll das Elektron einen Energiezustand W_1 im Valenzband entsprechend Abb. 2.23 besetzt haben.

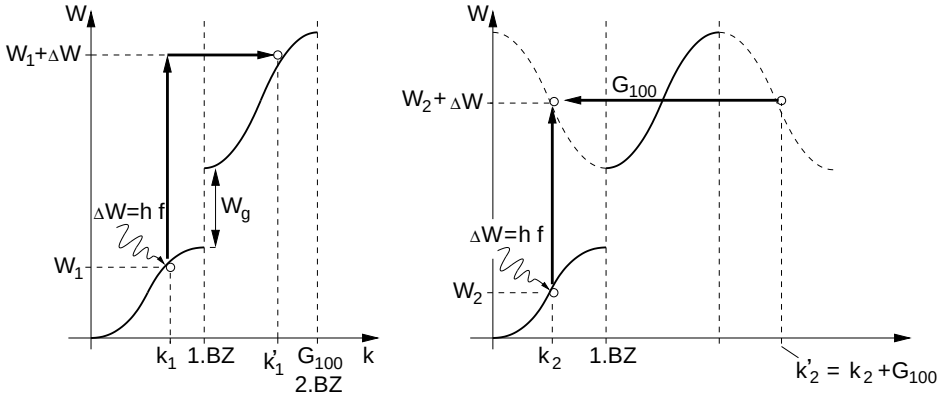

Abb. 2.23: Zufuhr einer Energie $\Delta W = h \cdot f$ für ein Elektron im VB am Beispiel einer idealisierten Bandstruktur in [100]-Richtung. **Links:** Nicht möglicher Übergang wegen Verletzung des Impulserhaltungssatzes. **Rechts:** Möglicher Übergang wegen Erfüllung von Energie- und Impulserhaltungssatz.

Für $\Delta W = h \cdot f < W_g$ existiert kein Zustand und es findet keine Absorption des Photons statt. Der Kristall ist für diese Frequenz f durchsichtig.

Wir unterstellen jetzt, dass die zugeführte Energie ausreicht, um die Energielücke zu überbrücken und eine Absorption stattfindet. Das Elektron geht dann von dem Valenzband in das Leitungsband. Ein solcher Band-Band-Übergang erfolgt immer unter Beachtung der auch in der Quantentheorie gültigen Sätze zur Energie- und Impulserhaltung. Nach dem

Wechsel hat das Elektron sowohl eine andere Energie als auch einen anderen Impuls.
Abb. 2.23 links zeigt einen Übergang, bei dem der Energieerhaltungssatz erfüllt ist, da die Energiedifferenz zwischen altem und neuem Zustand der zugeführten Energie entspricht. Der Impulserhaltungssatz ist jedoch wegen

$$\Delta p = \hbar\,(k_1 - k_1') \neq 0 \tag{2.135}$$

verletzt. Ein solcher Übergang ist für ein Elektron in dem Zustand mit der Energie W_1 daher nicht möglich. Wir erkennen weiterhin anhand der Abbildung, dass der Impulserhaltungssatz für alle Übergänge $\Delta p \neq 0$ ergibt und daher verletzt wird.
Als Konsequenz benötigt ein solcher Übergang einen dritten Partner, der die Impuls-Differenz aufnimmt. Diese Funktion erfüllt der Kristall selbst. Dabei ist zu beachten, dass der Kristall bei quantenmechanischer Betrachtung nur diskrete Impulse aufnehmen kann.
Eine entsprechende Rechnung (ohne Beweis) führt auf den Kristallimpulserhaltungssatz

$$\vec{k} - \vec{k'} = \vec{G}_{hkl}\ , \tag{2.136}$$

der gleichlautend mit der von uns in Gl. (1.101) hergeleiteten Bragg-Bedingung ist. Als Unterschied nimmt bei der hier angestellten Betrachtung der Kristall einen Impuls auf. Wir haben hier den Fall der sog. inelastischen Streuung, für die

$$|\vec{k}| \neq |\vec{k'}| \tag{2.137}$$

gilt, während bei der Bragg-Bedingung in Gl. (1.101) elastische Streuung mit $|\vec{k}| = |\vec{k'}|$ vorausgesetzt wurde.
Der Kristallimpulserhaltungssatz besagt also, dass nur der Übergang möglich ist, dessen Wellenvektor k' sich von dem ursprünglichen Wellenvektor um einen reziproken Gittervektor unterscheidet. Dieser Fall ist in Abb. 2.23 rechts eingezeichnet. Dabei wurde die Periodizität der Dispersionskurven berücksichtigt. Besonders einfach ist die Überlegung anhand des in Kap. 1.27 hergeleiteten reduzierten Bandschemas, bei dem alle Bandverläufe durch Verschiebung mit dem reziproken Gittervektor in der 1. Brillouin-Zone (1. BZ) abgebildet werden. Da die Verschiebung mit dem reziproken Gittervektor erfolgt, ist für einen senkrechten Band-Band-Übergang in der reduzierten Bandschema-Darstellung der Kristallimpulserhaltungssatz immer erfüllt. Der Übergang erfolgt dann, wie in Abb. 2.23 rechts gezeigt, für einen Wellenvektor k_2, für den die Energiedifferenz zur senkrecht darüber liegenden Energie

des höheren Bandes der zugeführten Energie ΔW entspricht. Damit ist auch der Energieerhaltungssatz erfüllt.

2.14 Thermalisierung

In Abb. 2.23 (rechts) haben wir gezeigt, wie das Elektron durch Absorption des Photons mit der Energie ΔW einen Zustand mit der Energie $W_2 + \Delta W$ im Leitungsband einnimmt. Dieser Zustand ist in Abb. 2.24 nochmals gezeigt. Dabei ist auch das Loch im Valenzband berücksichtigt, das aufgrund

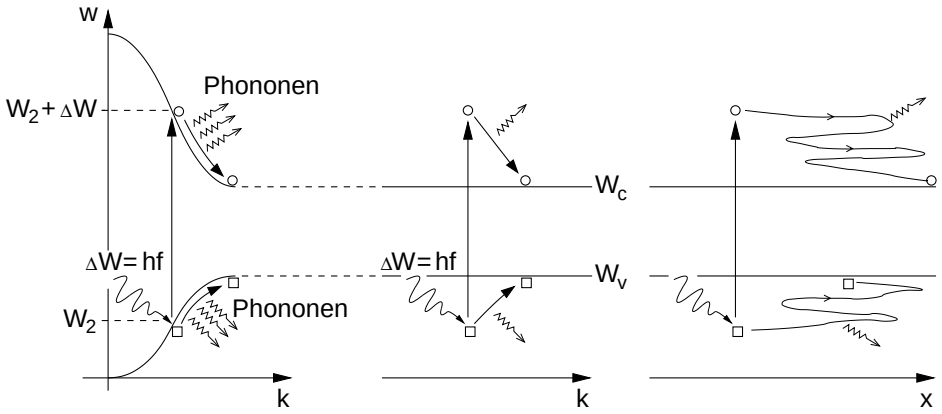

Abb. 2.24: **Links:** Dielektrische Relaxation (Thermalisierung) von Elektron und Loch nach der durch Energiezufuhr ΔW verursachten Paarbildung. **Mitte:** Darstellung des gleichen Sachverhaltes wie in der linken Abbildung, jedoch unter Verwendung des zeichnerisch einfacheren Bändermodells. **Rechts:** Darstellung des gleichen Vorgangs wie in den beiden linken Abbildungen, jedoch in eindimensionaler Orts-Darstellung.

der Paarbildung entsteht. Elektron und Loch werden nach ihrer Entstehung den energetisch niedrigsten, noch freien Zustand in ihrem Band besetzen. Elektronen „rollen“ dabei die Dispersionskurve bergab. Löcher steigen aufgrund ihrer entgegengesetzten Energieskala wie Luftblasen entlang der Dispersionskurve nach oben. Um sich auf den Dispersionskurven bewegen zu können, müssen beide Ladungsträger Energie abgeben. Sie erreichen dies, indem sie die überschüssige Energie in Quanten durch Emission von Phononen an das Kristallgitter abgeben. Das Kristallgitter nimmt diese Energie auf und erwärmt sich dabei. Dieser Vorgang wird Thermalisierung oder dielektrische Relaxation genannt. Er läuft vergleichsweise schnell, mit Zeitkonstanten zwischen $0,1 \ldots 10\,\mathrm{ps}$ ab, die dem im weiteren Verlauf besprochenen Relaxati-

onsprozess zur Neutralisierung von Raumladungen entsprechen.
Der gleiche Vorgang kann anstelle des Dispersionsdiagramms auch zur Vereinfachung im Bändermodell dargestellt werden. Dabei sind Darstellungen im k-Raum (Abb. 2.24 Mitte) und im Ortsraum (Abb. 2.24 rechts) möglich. Nach der Relaxation besitzen aufgrund der in dem Beispiel gewählten Bandstruktur sowohl Elektron als auch Loch den gleichen Wellenvektor. Dies ist eine wichtige Eigenschaft von Halbleitermaterialien für optisch emittierende Bauelemente, wie in nächsten Kapitel gezeigt wird.
Zu beachten ist, dass die Ladungsträger, auch wenn sie den gleichen Wellenvektor besitzen, nach der Paarbildung nicht (bzw. nur höchst unwahrscheinlich) am gleichen Ort verharren. Beide werden die zuvor beschriebenen Zufallsbewegungen ausführen und sich daher zu bestimmten Zeiten an verschiedenen Orten befinden. Diese Erkenntnis ist wichtig für das Verständnis des Rekombinationsvorgangs, bei dem das Elektron mit einem benachbarten Loch rekombiniert, d. h. den Zustand des Lochs besetzt. Dieses wird aufgrund der Wanderung durch den Kristall und der hohen Ladungsträgerdichte, wenn überhaupt, dann nur höchst unwahrscheinlich der selbe Partner aus der Bildung des Elektron-Loch-Paares sein.

2.15 Rekombination

Bei der Rekombination passiert mit einem Elektron und einem Loch genau das, was der Begriff beschreibt. Sie werden „wieder kombiniert“, d. h. das Elektron besetzt wieder den freien Zustand des Lochs. Wir erinnern uns an die Feststellung des letzten Kapitels und bemerken, dass es nicht genau das ursprünglich gebildete Elektron-Loch-Paar ist, das rekombiniert, sondern irgend ein Elektron-Loch-Paar, für das die Bedingungen für Rekombination erfüllt sind. Welche das sind, werden wir im Folgenden anschauen.
Dass es den Vorgang der Rekombination geben muss, macht eine einfache Überlegung deutlich. Würden durch Energiezufuhr nur Ladungsträger generiert, aber keine rekombiniert, dann würden alle Halbleiter nach genügend langer Energiezufuhr zu Leitern, da ihre Bänder mit freien Ladungsträgern gefüllt sind. Wir wissen aber zuverlässig, z. B. von Solarzellen oder Photodioden, dass nach Ausschalten des Lichts auch der Stromfluss versiegt. Es muss also einen Rekombinationsvorgang geben, der dies bewirkt.
Aus der Rekombination als Gegenspieler der Generation resultieren auch die Ladungsträgerdichten n_0, p_0 im thermodynamischen Gleichgewicht. Hier ist

die Rekombinationsrate r gleich der Generationsrate g. D. h. die pro Zeiteinheit in einem Volumen entstehenden Ladungsträger sind gleich der Zahl der in der gleichen Zeit und im gleichen Volumen verschwindenden Ladungsträger. Es gilt also im thermodynamischen Gleichgewicht

$$n = n_0, \qquad p = p_0, \qquad r = g, \qquad R = r - g = 0. \tag{2.138}$$

Bei einer Anhebung der Ladungsträgerdichte gegenüber den Gleichgewichtskonzentrationen wird die Rekombination die Generation überwiegen, da aufgrund der Neutralitätsbedingung das Bestreben besteht, die Gleichgewichtskonzentration wieder herzustellen. Es gilt bei Ladungsträgeranhebung gegenüber der Gleichgewichtskonzentration für die Nettorekombination

$$R = r - g > 0\ . \tag{2.139}$$

Im entgegengesetzten Fall, bei einer Absenkung der Ladungsträgerdichte gegenüber den Gleichgewichtskonzentrationen, gilt

$$R = r - g < 0\ . \tag{2.140}$$

Um genauere, für die Beschreibung von Halbleiterbauelementen ausreichende Aussagen über R machen zu können, benötigen wir tiefere Einblicke in die Abläufe bei der Generation und Rekombination von Ladungsträgern. Wir beschäftigen uns im Folgenden mit den Mechanismen der Rekombination. Bereits bei der Generation haben wir festgestellt, dass durch Einbeziehung eines dritten Partners Energie- und Impulserhaltungssatz bei Band-Band-Übergang erfüllt werden können. Wir betrachten daher zunächst im folgenden Kapitel die drei uns schon bekannten (Quasi-)Teilchen Elektron, Photon und Phonon, die sich als Partner für eine Wechselwirkung anbieten.

2.16 Energie und Impuls von (Quasi-)Teilchen

Energie und Impuls von Teilchen lassen sich entsprechend Gl. (1.4) und (1.27) immer über

$$W = \hbar\omega \tag{2.141}$$

$$p \ = \hbar k = \hbar \cdot \frac{2\pi}{\lambda} \tag{2.142}$$

berechnen. Die Energie wird dabei über ω und der Impuls über k (bzw. λ) bestimmt. Zwischen ω und k besteht eine Dispersionsbeziehung, die je nach

Teilchen und Vorgang unterschiedlich ist.

Für Photonen erhalten wir die Dispersionsbeziehung direkt aus der Lichtgeschwindigkeit:

$$c = f\lambda = \frac{\omega}{k} \,. \tag{2.143}$$

Für Elektronen im Kristall gilt nach Gl. (1.71) mit der effektiven Masse m_e^* die Dispersionsbeziehung

$$W_k = \hbar\omega_k = \frac{\hbar^2 k^2}{2m_e^*} \tag{2.144}$$

und für Phononen gewinnt man aus Überlegungen zu Schwingungssystemen (hier ohne Herleitung und Beweis) die Kreisfrequenz der Gitterschwingungen für den Fall einer einatomigen Basis

$$\omega_k = \sqrt{\frac{2Ya_0}{m_a}(1 - \cos(ka_0))} \tag{2.145}$$

mit Y=Elastizitätsmodul (Youngs Modulus), m_a=Atommasse, a_0=Gitterkonstante. Durch Einsetzen von Gl. (2.145) in Gl. (2.141) erhält man die Energie der Gitterschwingungen.

Während die Energie von der Frequenz des Teilchens abhängt, wird der Impuls über die Abhängigkeit von k nach Gl. (2.142) durch die Wellenlänge bestimmt. Diese ist bei Photonen über Gl. (2.143) mit der Lichtgeschwindigkeit umgekehrt proportional zur Frequenz.
Bei Elektronen im Kristall wissen wir, dass sie sich i. e. am Rand der Bandkante, also am Rand der 1. Brillouin-Zone mit $k = \frac{\pi}{a_0}$ aufhalten.
Bei Phononen hilft für die Abschätzung von k die einfache Überlegung nach Abb. 2.25, wonach die kleinste Wellenlänge dann erreicht ist, wenn benachbarte Gitteratome die entgegengesetzte Auslenkung mit der Amplitude der Schwingung besitzen (Fall der transversalen, optischen Phononen). Es gilt daher für Phononen mit maximalem Impuls $\lambda_{min} = 2a_0$.
Mit den vorangegangenen Überlegungen lässt sich für Beispielwerte (in einer Übung) die Größenordnung von Energie und Impuls der Teilchen ermitteln. Tabelle 2.6 zeigt die Ergebnisse.

Wir merken uns das daraus folgende, wichtige Ergebnis:

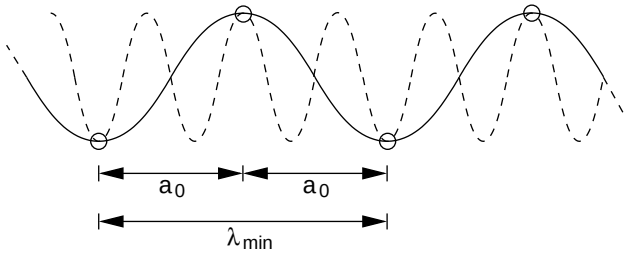

Abb. 2.25: Der Abstand a_0 der Gitteratome legt die kleinste mögliche Wellenlänge λ_{min} der Schwingung fest. Die gestrichelte Schwingung ist nicht möglich, da sie eine zu kleine Wellenlänge besitzt.

	Elektron	Photon	Phonon
$\frac{W}{W_{Elektron}}$	1	≈1	0,02
$\frac{p}{p_{Elektron}}$	1	0,001	≈1

Tabelle 2.6: Vergleich von Energie und Impuls von (Quasi-)Teilchen.

- Elektronen besitzen Energie und Impuls,
- Photonen besitzen Energie und (vernachlässigbar) kleinen Impuls.
- Phononen besitzen Impuls und (vernachlässigbar) kleine Energie.

Dabei ist die Einschätzung als vernachlässigbar immer in Relation zu dem zum Vergleich herangezogenen Teilchen zu werten.

2.17 Direkte Rekombination (Direkter Halbleiter)

Wir erinnern uns an den Vorgang der Relaxation aus Kap. 2.14. Nach Abschluss des Vorgangs befinden sich Loch und Elektron auf dem jeweils energetisch niedrigsten unbesetzten Niveau in ihrem Band. Im Beispiel in Abb. 2.24 lagen Elektron und Loch nach der Relaxation bei dem gleichen Wellenvektor, da aufgrund der Bandstruktur Minimum des Leitungsbandes und Maximum des Valenzbandes dort zusammenfielen.
Das Elektron kann unter diesen Bedingungen direkt, wie in Abb. 2.26 links gezeigt, durch Abgabe eines Photons der Energie

$$hf = W_g \tag{2.146}$$

mit einem Loch mit i. e. gleichem Wellenvektor rekombinieren. Dies ist möglich, da der Impuls von Photonen, wie im vorangegangenen Kapitel festgestellt, vernachlässigbar klein ist und daher bei einem senkrechten Übergang erhalten bleibt. Der Band-Band-Übergang kann daher im k-Raum direkt senkrecht unter Emission eines Photons erfolgen. Es wird bei dem direkten Übergang also Licht der Wellenlänge $f = \frac{W_g}{h}$ abgestrahlt.
Wir nennen Halbleiter, bei denen Minimum des Leitungsbandes und Maximum des Valenzbandes im k-Raum übereinander liegen, direkte Halbleiter. Sie erlauben die direkte Rekombination eines Elektrons aus dem Leitungsband mit einem Loch im Valenzband durch Emission von Photonen, also sichtbarem oder unsichtbarem Licht.
Direkte Halbleiter sind Materialien für optoelektronische Bauelemente. Zu den wichtigsten Materialien gehört **GaAs** mit dem Banddiagramm nach Abb. 1.40.

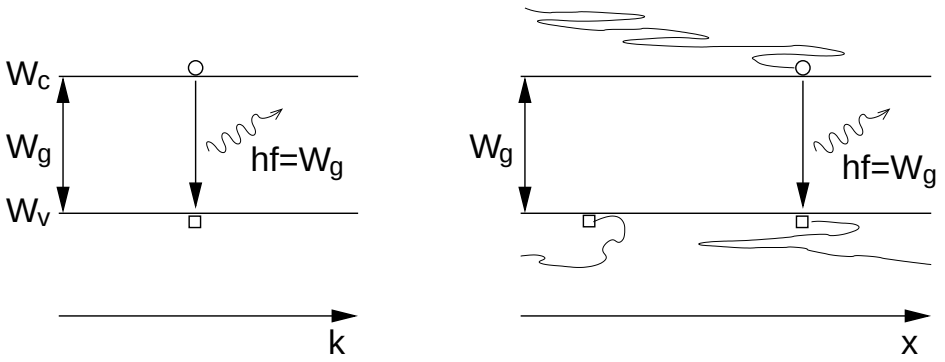

Abb. 2.26: **Links**: Direkte Rekombination unter Emission eines Photons, dargestellt im k-Raum. **Rechts**: Der gleiche Vorgang, dargestellt im Ortsraum. Die zufällige Bewegung der Ladungsträger ist symbolisch durch Zickzacklinien angedeutet.

Auch bei der Darstellung im Ortsraum in Abb. 2.26 rechts erfolgt der Band-Band-Übergang der direkten Rekombination senkrecht, da sich die beiden Rekombinationspartner am selben Ort befinden müssen.
Die Häufigkeit der direkten Rekombination eines Elektrons mit einem Loch, genauer gesagt die Rekombinationsrate r, ist proportional zur Elektronendichte n und Löcherdichte p. Sie ist auch proportional zur thermischen Geschwindigkeit v_{th}, da diese die Geschwindigkeit, mit der sich mögliche Partner an einem Ort treffen, bestimmt. Mit dem sog. Wirkungsquerschnitt der Fläche A_r führen wir eine weitere Proportionalitätskonstante ein, die die Ef-

fizienz der Rekombination charakterisiert. Für die Rekombinationsrate kann damit die Gleichung

$$r = A_r \, \mathrm{v}_{th} \, n \, p \, . \tag{2.147}$$

aufgestellt werden.
Der Rekombination wirkt ein Generationsvorgang mit der Generationsrate g entgegen. Die Nettorekombinationsrate ergibt sich nach Gl. (2.129) zu

$$R = r - g = A_r \, \mathrm{v}_{th} \, n \, p - g \, . \tag{2.148}$$

Da im thermodynamischen Gleichgewicht

$$R = 0, \; n \, p = n_0 \, p_0 = n_i^2 \tag{2.149}$$

gilt, erhalten wir aus Gl. (2.148) direkt

$$g = A_r \, \mathrm{v}_{th} \, n_0 \, p_0 \tag{2.150}$$

und damit für die Nettorekombinationsrate bei direkten Band-Band-Übergängen

$$R = A_r \mathrm{v}_{th} (np - n_i^2) = r - g \, . \tag{2.151}$$

Auch wenn hier nicht explizit geschrieben, ist zu beachten, dass die Nettorekombinationsrate aufgrund der Ortsabhängigkeit von n, p und A_r ebenfalls vom Betrachtungsort im Halbleiter abhängt.
Die Beziehung drückt direkt aus, dass die Nettorekombination zu Null wird, wenn an der betrachteten Stelle im Halbleiter thermodynamisches Gleichgewicht vorliegt.

2.18 Indirekte Rekombination (Indirekte Halbleiter)

Entsprechend den Überlegungen zu direkten Halbleitern sind **Si** und **Ge** aufgrund ihrer Bandstruktur indirekte Halbleiter (vgl. Abb. 1.40). Nach der Relaxation liegen Elektronen und Löcher in Energieminima, die sich durch einen Wellenvektor, der ungleich einem reziproken Gittervektor ist, unterscheiden. D. h. sie liegen im reduzierten Gitterschema nicht direkt übereinander. Der Kristallimpulserhaltungssatz Gl. (2.136) ist daher verletzt. Die Rekombination in indirekten Halbleitern kann dadurch nicht direkt über Emission eines („impulslosen“) Photons stattfinden.
Die Rekombination benötigt einen Partner, der einen Impuls besitzt, so dass

durch Impulsaustausch der Impulserhaltungssatz erfüllt wird. Als Partner bieten sich daher nach Kap. 2.16 Elektronen und Phononen an:
Bei der indirekten Rekombination durch Phononenprozesse erfolgt ein Energie- und Impulsaustausch mit den Gitterschwingungen des Kristalls. Aufgrund der geringen Energie von Phononen erfolgt die Rekombination schrittweise über Zwischenenergieniveaus in der Bandlücke. Die Beschreibung dieses Prozesses führt auf die bekannte „Shockley-Read-Hall"-Beziehung, die im folgenden Kapitel hergeleitet wird.
Bei der indirekten Rekombination durch Elektronen gibt ein Leitungselektron an ein anderes Leitungselektron oder an ein Defektelektron bei einem Stoß Energie ΔW und Impuls Δp ab. Dadurch erhält es die passende Energie und den Impuls eines Zustandes im Valenzband (unwahrscheinlich) oder eines Zwischenzustandes in der Bandlücke. Die Energie des anderen Elektrons erhöht sich um den entsprechenden Betrag ΔW. Sein Wellenvektor wird mit entgegengesetztem Vorzeichen $-\Delta p$ geändert, so dass der Impulserhaltungssatz erfüllt ist. Abb. 2.27 zeigt diesen als „Auger-Prozess" bezeichneten Vorgang bei einem Übergang ohne Zwischenenergieniveau. Wir werden den

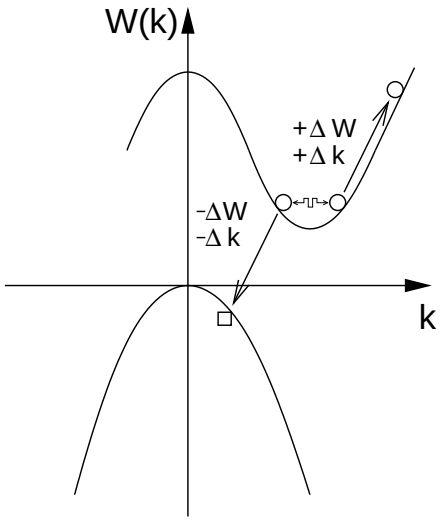

Abb. 2.27: Änderung von Energie und Impuls bei einem Auger-Prozess.

Augerprozess mit Zwischenenergieniveaus im übernächsten Kapitel 2.20 kurz betrachten.

2.19 Shockley-Read-Hall-(SRH-)Rekombination

Bei der indirekten Rekombination mit Phononen-Wechselwirkung kann das Phonon den Impuls des Elektrons aufnehmen, da die Wellenvektoren beider Teilchen die gleiche Größenordnung besitzen (vgl. Kap. 2.16). Der Impulserhaltungssatz kann also erfüllt werden.
Jedoch ist die Energie von Phononen klein im Vergleich zu der Energie von Elektronen. Um einen LB-VB-Übergang zu ermöglichen, müssten daher viele Phononen emittiert werden. Ein solcher Vorgang ist auf ein Elektron bezogen sehr unwahrscheinlich. Daher erfolgt der Band-Band-Übergang bei der Wechselwirkung mit Phononen über Zwischenniveaus der Energie W_t in etwa in der Mitte der Bandlücke (vgl. Abb. 2.28).
Zwischenniveaus werden auch als „Rekombinationszentren“ oder „Traps“ bezeichnet. Sie entstehen durch Störstellen in Form von Fremdatomen (z. B. **Au**), Verunreinigungen und Gitterdefekten. Mit zunehmender Reinheit des Kristalls nimmt die Störstellenkonzentration und damit auch die Rekombination ab. Zur Veranschaulichung haben LB-Elektronen in sehr reinem **Si** Lebensdauern im ms-Bereich und legen Strecken in mm-Größenordnung zurück, ehe sie rekombinieren. In stark verunreinigtem **Si** liegen Lebensdauer und Diffusionslänge im μs- und μm-Bereich.

Abb. 2.28 veranschaulicht den indirekten Rekombinationsvorgang anhand eines Bändermodells. Durch die Lage i. e. der Mitte der Bandlücke ist eine Besetzungswahrscheinlichkeit mit einem Elektron oder Loch näherungsweise gleich groß. In der Realität liegt nicht nur ein, sondern eine Vielzahl von Energieniveaus für Rekombinationszentren vor. Bei einem LB-VB-Übergang durchlaufen die Elektronen vielmehr eine Treppe von mehreren Zwischenniveaus. Wir begnügen uns in unserem einfachen Modell jedoch mit einem Energieniveau W_t, da die damit erzielten Ergebnisse eine für den Rahmen der Vorlesung ausreichende Übereinstimmung mit dem Experiment ergeben.

Wir bezeichnen entsprechend Abb. 2.28 mit

N_t die Dichte der Rekombinationszentren,

u_{ct} die Einfangrate von Elektronen des LB durch die Rekombinationszentren,

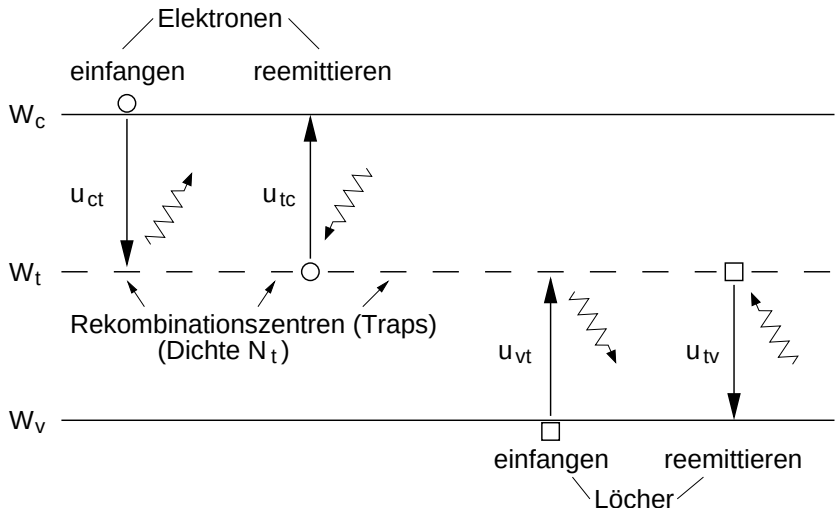

Abb. 2.28: Emissions- und Einfangvorgänge bei indirekten Rekombinationsvorgängen über Rekombinationszentren.

u_{tc} die Reemissionsrate von Elektronen aus Rekombinationszentren in das Leitungsband,

u_{vt} die Einfangrate von Löchern des Valenzbandes durch die Rekombinationszentren,

u_{tv} die Reemissionsrate von Löchern aus Rekombinationszentren in das Valenzband.

Unter der Größe „Rate“ ist immer eine Ladungsträgerdichte pro Zeiteinheit im Sinne von $\frac{dn}{dt}$ zu verstehen. Sie beschreibt die pro Volumen und Zeiteinheit eingefangenen oder emittierten Ladungsträger.

Mit f_t bezeichnen wir die Besetzungswahrscheinlichkeit eines Energieniveaus der Rekombinationszentren mit Elektronen. Wir betrachten den Halbleiter bei geringer Abweichung vom thermodynamischen Gleichgewicht, so dass für die Besetzungswahrscheinlichkeit die Fermi-Dirac-Verteilungsfunktion

$$f_t = \frac{1}{1 + e^{\frac{W_t - W_F}{kT}}} \tag{2.152}$$

verwendet werden kann.
Dann ist die Einfangrate u_{ct} der Elektronen aus dem Leitungsband wie bei

der direkten Rekombination proportional dem Produkt der Dichten der beteiligten Partner. Die Ladungsträgerdichte der Elektronen im Leitungsband ist n. Die Dichte der unbesetzten Rekombinationszentren (RZ), auf den die Elektronen übergehen könnten, ist $N_t(1 - f_t)$.
Mit der Proportionalitätskonstante c_{ct}, die den LB→RZ-Übergang charakterisiert (zu erkennen am Index *ct* für *conduction →trap*), ergibt sich

$$u_{ct} = c_{ct}\, n\, N_t\, (1 - f_t) \tag{2.153}$$

Für den entgegen laufenden Reemissionsvorgang RZ→LB gilt dementsprechend mit $(N_C - n) \approx N_C$ freien Zuständen im LB

$$u_{tc} = c_{tc}\, N_t\, f_t\, N_C\ . \tag{2.154}$$

Entsprechend ergibt sich für das Einfangen und Reemittieren der Löcher

$$u_{vt} = c_{vt}\, p\, N_t\, f_t \tag{2.155}$$

$$u_{tv} = c_{tv}\, N_t\, (1 - f_t)\, N_V\ . \tag{2.156}$$

Daraus ergeben sich die Netto-Einfangrate für Elektronen aus dem Leitungsband, also die Rate von Elektronen, die auf dem Rekombinationszentrum zur Rekombination verbleiben

$$\frac{\partial n}{\partial t} = U_{ct} = u_{ct} - u_{tc} = c_{ct}\, n\, N_t\, (1 - f_t) - c_{tc}\, N_t\, f_t\, N_C \tag{2.157}$$

und die Netto-Einfangrate für Löcher aus dem Valenzband, also die Rate von Löchern, die auf dem Rekombinationszentrum zur Rekombination mit den Elektronen verbleiben[19]

$$\frac{\partial p}{\partial t} = U_{vt} = u_{vt} - u_{tv} = c_{vt}\, p\, N_t\, f_t - c_{tv}\, N_t\, (1 - f_t)\, N_V\ . \tag{2.158}$$

Jeweils eine der Proportionalitätskonstanten in Gl. (2.157) und (2.158) kann über die Bedingung

$$\frac{\partial p}{\partial t} = U_{vt} = 0, \qquad \frac{\partial n}{\partial t} = U_{ct} = 0\ , \tag{2.159}$$

[19]Diese Betrachtungsweise ist identisch mit der Betrachtungsweise, dass eine Rate (Anzahl pro Zeit) von Elektronen von den Rekombinationszentren in das Valenzband übergeht.

die im thermodynamischen Gleichgewicht gilt, bestimmt werden. Wir erhalten mit Gl. (2.159) aus Gl. (2.157) und (2.158) mit n_0 und p_0 als Ladungsträgerdichten im thermodynamischen Gleichgewicht

$$c_{tc} = c_{ct} \frac{n_0}{N_C} \frac{1 - f_t}{f_t} = c_{ct} \cdot e^{-\frac{W_C - W_t}{kT}} \tag{2.160}$$

$$c_{tv} = c_{vt} \frac{p_0}{N_V} \frac{f_t}{1 - f_t} = c_{vt} \cdot e^{-\frac{W_t - W_V}{kT}} \tag{2.161}$$

Die beiden rechten Ausdrücke ergeben sich nach kurzer Rechnung, die zur Übung einmal nachvollzogen werden sollte (zu verwenden sind Gl. (2.152) sowie Gl. (2.23) und (2.26)).

Damit wird aus den Netto-Einfangraten aus Gl. (2.157) und (2.158)

$$U_{ct} = c_{ct}\, n\, N_t\, (1 - f_t) - c_{ct}\, N_t\, f_t\, N_C\, e^{-\frac{W_C - W_t}{kT}} \tag{2.162}$$

$$U_{ct} = c_{ct}\, N_t\, ((1 - f_t)\, n - f_t\, n_1) \tag{2.163}$$

$$\text{mit } n_1 := N_C\, e^{-\frac{W_C - W_t}{kT}} \tag{2.164}$$

und

$$U_{vt} = c_{vt}\, p\, N_t\, f_t - c_{vt}\, N_t\, (1 - f_t)\, N_V\, e^{-\frac{W_t - W_V}{kT}} \tag{2.165}$$

$$U_{vt} = c_{vt}\, N_t\, (p\, f_t - (1 - f_t)\, p_1) \tag{2.166}$$

$$\text{mit } p_1 := N_V\, e^{-\frac{W_t - W_V}{kT}} \tag{2.167}$$

Die Dichten n_1 und p_1 sind Hilfsgrößen zur Vereinfachung und beschreiben die Elektronen- und Löcherdichte für den Fall, dass das Fermi-Niveau mit dem Energieniveau W_t der Rekombinationszentren zusammenfällt. Es lässt sich leicht nachprüfen, dass entsprechend Gl. (2.33)

$$n_1 \cdot p_1 = n_i^2 \tag{2.168}$$

gelten muss, da wir den Halbleiter im thermodynamischen Gleichgewicht betrachten.
Im stationären Zustand muss die Netto-Einfangrate der Elektronen und die der Löcher gleich sein.

$$U_{ct} = U_{vt} \Leftrightarrow \frac{\partial p}{\partial t} = \frac{\partial n}{\partial t}\ . \tag{2.169}$$

Dann ist die Besetzung der Rekombinationszentren konstant (zeitunabhängig) und es ergibt sich ein kontinuierlicher Rekombinationsfluss der Elektronen von LB $\rightarrow$ RZ $\rightarrow$ VB. Diese Bedingung nutzen wir und setzen die entsprechenden Beziehungen aus Gl. (2.163) und (2.166) in die Bedingung für Stationarität ein. Umstellen führt auf die Besetzungswahrscheinlichkeit eines Energieniveaus der Rekombinationszentren mit einem Elektron

$$f_t = \frac{c_{vt}\, p_1 + c_{ct}\, n}{c_{vt}(p_1 + p) + c_{ct}(n_1 + n)} \; . \tag{2.170}$$

Einsetzen der Besetzungswahrscheinlichkeit in eine der Beziehungen für die Netto-Einfangrate Gl. (2.163) oder (2.166) (beide sind wegen Gl. (2.169) identisch) liefert nach kurzer einfacher Rechnung

$$U_{ct} = U_{vt} = \frac{pn - n_i^2}{(p_1 + p)\tau_n + (n_1 + n)\tau_p} = R = r - g \tag{2.171}$$

mit

$$\tau_n := \frac{1}{c_{ct} N_t} \; , \qquad \tau_p := \frac{1}{c_{vt} N_t} \; . \tag{2.172}$$

Das ist die Shockley-Read-Hall-(SRH-)Formel für die Netto-Rekombination $R = r - g$ bei indirekter Rekombination über Rekombinationszentren. Aus ihrem Zählerterm geht hervor, dass die Nettorekombinationsrate zu Null wird, wenn die Ladungsträgerdichten p und n die Werte im thermodynamischen Gleichgewicht annehmen.

2.20 Auger-Rekombination

Wir wollen den allgemeinen Fall der Auger-Rekombination über Rekombinationszentren betrachten. Dabei stoßen zwei Ladungsträger zusammen. Der eine Ladungsträger gibt dabei Energie ab und geht dadurch auf das Energieniveau eines Rekombinationszentrums. Der andere Ladungsträger nimmt die abgegebene Energie auf und gelangt auf ein höheres Energieniveau innerhalb seines Bandes. Dieser Mechanismus nach Abb. 2.27 und 2.29 ist vergleichbar mit dem Einfangvorgang bei der SRH-Rekombination.
Bei den Auger-Prozessen gibt es jedoch jeweils zwei Möglichkeiten, wie Elektronen vom Leitungsband ins Rekombinationszentrum bzw. Löcher vom Valenzband ins Rekombinationszentrum gelangen können (vgl. Abb. 2.29).
In Analogie zur Herleitung von Gl. (2.153) ermitteln wir die Einfangrate

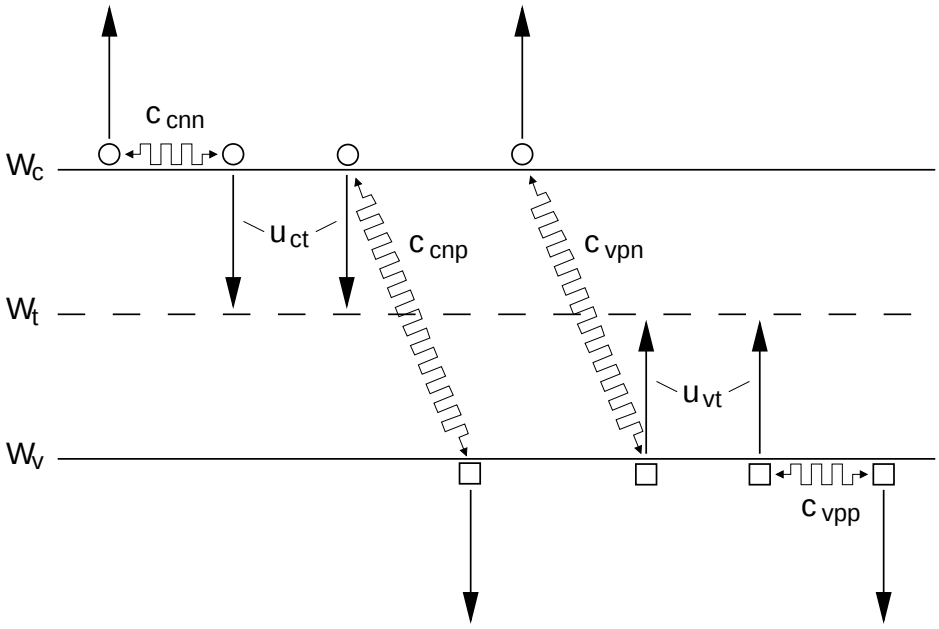

Abb. 2.29: Mögliche Einfang-Vorgänge von Ladungsträgern bei einem Auger-Prozess.

von Elektronen in den Rekombinationszentren mit einer Proportionalitätskonstante c_{cnn} für einen Elektron-Elektron-Stoß und c_{cnp} für einen Elektron-Loch-Stoß. Die Einfangraten für die beiden Stoßmechanismen müssen proportional den Dichten n, p der beiden Stoßpartner und der Dichte $(1-f_t)\cdot N_t$ der von Elektronen nicht besetzten Rekombinationszentren sein. Wir erhalten damit

$$u_{ct} = c_{cnn}\cdot n\cdot n\cdot N_t\cdot(1-f_t) + c_{cnp}\cdot n\cdot p\cdot N_t\cdot(1-f_t) \tag{2.173}$$

$$= (c_{cnn}\cdot n + c_{cnp}\cdot p)\cdot n\cdot N_t\cdot(1-f_t)\ . \tag{2.174}$$

Analog ergibt sich für die Einfangrate von Löchern in den Rekombinationszentren mit den entsprechenden Proportionalitätskonstanten der Stoßprozesse

$$u_{vt} = c_{vpp}\cdot p\cdot p\cdot N_t\cdot f_t + c_{vpn}\cdot p\cdot n\cdot N_t\cdot f_t \tag{2.175}$$

$$u_{vt} = (c_{vpp}\cdot p + c_{vpn}\cdot n)\cdot p\cdot N_t\cdot f_t\ . \tag{2.176}$$

Wie bei der SRH-Rekombination gibt es auch hier Reemissionsvorgänge aus

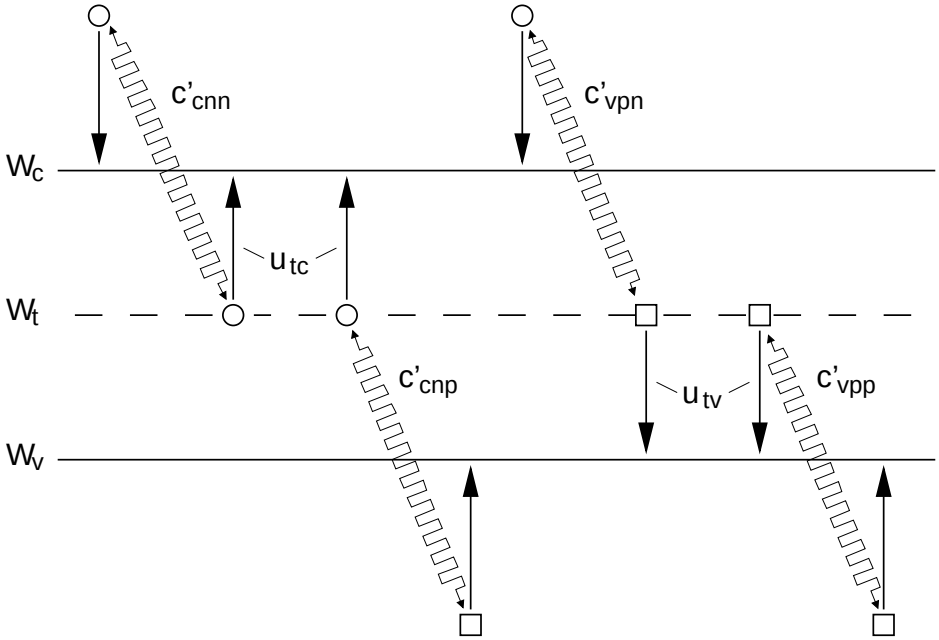

Abb. 2.30: Mögliche Reemissionsvorgänge von Ladungsträgern bei einem Auger-Prozess.

den Rekombinationszentren, die zur Übersicht in Abb. 2.30 dargestellt sind. Wir erhalten mit den gleichen Überlegungen wie zuvor für die Reemissionsrate der Elektronen von RZ $\rightarrow$ LB über Elektron-Elektron- und Elektron-Loch-Stöße mit den für den umgekehrten Stoßprozess charakteristischen Proportionalitätskonstanten

$$u_{tc} = c'_{cnn} \cdot n \cdot N_t \cdot f_t \cdot N_C + c'_{cnp} \cdot N_t \cdot f_t \cdot p \cdot N_C \tag{2.177}$$

$$= (c'_{cnn} \cdot n + c'_{cnp} \cdot p) \cdot N_t \cdot f_t \cdot N_C \; . \tag{2.178}$$

Für die Reemissionsrate der Löcher gilt entsprechend

$$u_{tv} = c'_{vpn} \cdot n \cdot N_t \cdot (1 - f_t) \cdot N_V + c'_{vpp} \cdot p \cdot N_t \cdot (1 - f_t) \cdot N_V \tag{2.179}$$

$$= (c'_{vpp} \cdot p + c'_{vpn} \cdot n) \cdot N_t \cdot (1 - f_t) \cdot N_V \; . \tag{2.180}$$

Wir haben bewusst die Umformungen mit dem vorgezogenen Klammer-Term in Gl. (2.174), (2.176), (2.178) und (2.180) gemacht. Durch Vergleich mit den entsprechenden Gleichungen für Einfangen und Reemittieren bei der SRH-

Rekombination erkennen wir, dass für den Austausch

$$\text{Gl. (2.174)}: \ (c_{cnn}\, n + c_{cnp}\, p) \rightarrow c_{ct} : \text{Gl. (2.153)} \tag{2.181}$$

$$\text{Gl. (2.176)}: \ (c_{vpp}\, p + c_{vpn}\, n) \rightarrow c_{vt} : \text{Gl. (2.155)} \tag{2.182}$$

$$\text{Gl. (2.178)}: \ (c'_{cnn}\, n + c'_{cnp}\, p) \rightarrow c_{tc} : \text{Gl. (2.154)} \tag{2.183}$$

$$\text{Gl. (2.180)}: \ (c'_{vpp}\, p + c'_{vpn}\, n) \rightarrow c_{tv} : \text{Gl. (2.156)} \tag{2.184}$$

die Auger-Rekombination durch die Beziehungen der SRH-Rekombination beschrieben werden.
Wir können uns daher viel Rechnung sparen und direkt das Ergebnis der SRH-Rekombination für die Auger-Rekombination verwenden. Es gilt also für die Nettorekombinationsrate des Auger-Prozesses die SRH-Beziehung nach Gl. (2.171)

$$R = r - g = \frac{p\, n - n_i^2}{(p_1 + p)\tau_n + (n_1 + n)\tau_p} \ . \tag{2.185}$$

Der Unterschied liegt in den Zeitkonstanten. Aus Gl. (2.172) folgt durch Substitution entsprechend Gl. (2.181) und (2.182)

$$\tau_n = \frac{1}{(c_{cnn} n + c_{cnp} p) N_t} \ , \qquad \tau_p = \frac{1}{(c_{vpp} p + c_{vpn} n) N_t} \ . \tag{2.186}$$

Daran ist zu sehen, dass die Zeitkonstanten des Auger-Prozesses abhängig von den Ladungsträgerdichten sind. Dies ist unmittelbar einsichtig, da für hohe Dichten viele Auger-Stöße erfolgen (ähnlich einer dicht gedrängten Menschenmasse), wodurch die Rekombinationsrate steigt.

2.21 Einfaches Generations-Rekombinations-Modell

In den vorangegangenen Kapiteln haben wir die Netto-Rekombinationsraten für direkte und indirekte Rekombination über Rekombinationszentren und Auger-Prozesse ermittelt.
Wir wissen daher, dass die Netto-Rekombinationsrate R nur dann ungleich Null ist, wenn die Ladungsträgerdichten (p,n) von den Gleichgewichtsdichten (p_0,n_0) abweichen. Wir sehen diesen Zusammenhang anhand des Terms $(np - n_i^2)$ in allen abgeleiteten Gleichungen für die Nettorekombinationsrate (Gl. (2.151), (2.171), (2.185)), durch den $R = 0$ für $np = n_0 p_0 = n_i^2$ folgt.
Im stationären Zustand ergeben sich wegen Gl. (2.169) für Elektronen und Löcher gleiche Rekombinationsraten

$$R = R_p = R_n \ , \tag{2.187}$$

da gleich viele Elektronen und Löcher an der Rekombination beteiligt sind ($U_{ct} = U_{vt}$, vgl. z. B. Gl. (2.171)). Dies hat eine weitreichende Konsequenz: Betrachten wir dotierte Halbleiter, so hängt die Rekombinationsrate der Majoritätsladungsträger von der Gleichgewichtsstörung Δn, Δp der Minoritätsträger ab, da diese aufgrund ihrer Minderheit die Rekombinationsrate bestimmen.
Da die Rekombinationsrate eine abgebaute bzw. generierte Ladungsträgerdichte pro Zeiteinheit angibt, liegt es nahe, für diesen Vorgang Modellgleichungen mit einer Struktur entsprechend „Gleichgewichtsstörung bezogen auf Abbaudauer“ aufzustellen.
Für die Nettorekombinationsrate von Elektronen in einem p-Halbleiter lässt sich dementsprechend eine Modellgleichung angeben:

$$R_n = R_p = R = \frac{n - n_0}{\tau_n} = \frac{\Delta n}{\tau_n} \,. \tag{2.188}$$

Entsprechend gilt für Löcher in einem n-Halbleiter

$$R_p = R_n = R = \frac{p - p_0}{\tau_p} = \frac{\Delta p}{\tau_p} \,. \tag{2.189}$$

Diese Form der Modellgleichungen berücksichtigt auch, dass für $n = n_0$ bzw. $p = p_0$ die Nettorekombinationsrate zu Null wird.
Wir verwenden diese Modellgleichungen bevorzugt dann, wenn wir formal den Vorgang der Rekombination beschreiben wollen, ohne direkt die doch recht umfangreichen Formeln der Rekombination zu verwenden.
In welchem Bereich die eingangs gemachten Überlegungen zur Struktur der einfachen Modellgleichungen richtig sind, ergibt sich aus der Bedingung, für die sich aus der Rekombinationsbeziehung die Modellgleichungen ergeben. Hier zeigt sich, dass sich unsere Gleichungen für die Nettorekombination bei kleinen Abweichungen von den Gleichgewichtsdichten (quasineutraler Halbleiter) so vereinfachen lassen, dass sie in die Modellgleichungen übergehen.
Zum Beweis verwenden wir die SRH-Gleichung Gl. (2.171) und ersetzen $n = n_0 + \Delta n$ und $p = p_0 + \Delta p = p_0 + \Delta n$. Dabei haben wir mit $\Delta n = \Delta p$ schon auf die erst später hergeleitete Neutralitätsbedingung nach erfolgter Relaxation vorgegriffen. Es ergibt sich damit

$$R = \frac{(p_0 + \Delta n)(n_0 + \Delta n) - n_i^2}{(p_1 + p_0 + \Delta n)\tau_n + (n_1 + n_0 + \Delta n)\tau_p} \,. \tag{2.190}$$

Für kleine Störungen $\Delta n \ll n_0, p_0$ ergibt sich

$$R \approx \frac{\Delta n(n_0 + p_0)}{(p_1 + p_0)\tau_n + (n_1 + n_0)\tau_p} = \frac{\Delta n}{\tau_{eff}} \tag{2.191}$$

mit der effektiven Lebensdauer-Zeitkonstanten

$$\tau_{eff} := \frac{p_1 + p_0}{n_0 + p_0}\tau_n + \frac{n_1 + n_0}{n_0 + p_0}\tau_p \; . \tag{2.192}$$

Dieses Ergebnis gilt entsprechend der Voraussetzung $\Delta n \ll n_0$, p_0 bei schwachen Abweichungen Δn von den Gleichgewichtsdichten sowohl bei dotierten als auch bei eigenleitenden Halbleitern.
Für $\frac{n_0}{n_i} > 1$ liegt ein n-Halbleiter, für $\frac{n_0}{n_i} < 1$ ein p-Halbleiter vor. Mit $n_0 p_0 = n_i^2$ gehen aus τ_{eff} durch Vereinfachung für n- bzw. p-Halbleiter die Zeitkonstanten

$$n\text{-Halbleiter: } \frac{n_0}{n_i} \gg 1 : \qquad \tau_{eff} \approx \tau_p \tag{2.193}$$

$$p\text{-Halbleiter: } \frac{n_0}{n_i} \ll 1 : \qquad \tau_{eff} \approx \tau_n \tag{2.194}$$

hervor. Damit nimmt die Näherung der SRH-Beziehung in Gl. (2.191) den Wert der einfachen Modellgleichung (2.188) an.
Die Zeitkonstanten τ_p, τ_n werden entsprechend ihrer Bedeutung als Lebensdauer für Löcher in n-Halbleitern (τ_p) bzw. als Lebensdauer für Elektronen in p-Halbleitern (τ_n) interpretiert.

2.22 Neutralisation und Abbau von Gleichgewichtsstörungen

Bisher haben wir den Halbleiter mit seinen Kenngrößen im statischen bzw. thermodynamischen ($f(t) = const.$) und stationären $\left(\frac{\partial f(t)}{\partial t} = const.\right)$ Gleichgewichtszustand betrachtet. Dabei war der Halbleiter immer homogen und elektrisch neutral.
Die Kenntnis der Halbleitereigenschaften in diesem Zustand ist wichtig, denn der Halbleiter befindet sich in der Regel bevor eine Störung eintritt und nach Ausgleich der Störung in diesem Gleichgewichtszustand.
Im Betrieb des Halbleiters in elektronischen Bauelementen sind Störungen des Gleichgewichts der Regel- oder besser der Betriebsfall. Allgemein besteht die Aufgabe des Bauelementes darin, bestimmte Änderungen in und zu

bestimmten Zeiten herbeizuführen oder auf Änderungen zu reagieren. Wir benötigen dafür eine Beschreibung der Vorgänge, die in einem Halbleiter bei Eintritt einer Gleichgewichtsstörung bis zur Wiedereinstellung eines statischen oder stationären Gleichgewichts ablaufen. Dies ist das Ziel der folgenden Überlegungen.

2.23 Ladungs-Neutralisation durch dielektrische Relaxation

Wir betrachten ganz allgemein eine plötzliche Abweichung der Ladungsträgerdichten um Δn und Δp gegenüber den Gleichgewichtswerten n_0 und p_0. Dieses plötzliche Ereignis soll zum Zeitpunkt $t = 0$ eintreten. Wir können diese Störung mit Hilfe der Neutralitätsbedingung ganz formal wie folgt formulieren:

$$0 = \rho = e(p_0 + N_D^+ - n_0 - N_A^-) \qquad , t < 0 \; , \tag{2.195}$$

$$\rho = e(p_0 + \Delta p + N_D^+ - n_0 - \Delta n - N_A^-) \qquad , t \geq 0 \; . \tag{2.196}$$

Der Halbleiter war also für $t < 0$ elektrisch neutral. Für $t \geq 0$ ist die Neutralität durch eine Raumladung gestört:

$$\rho = e(p + N_D^+ - n - N_A^-) \tag{2.197}$$

$$\text{mit } p = p_0 + \Delta p \text{ und } n = n_0 + \Delta n \quad , \tag{2.198}$$

bzw. in einer anderen Formulierung durch Einsetzen von Gl. (2.195) in Gl. (2.196)

$$\rho = e(\Delta p - \Delta n) \; . \tag{2.199}$$

Bevor wir rechnen, überlegen wir kurz, was wir erwarten:
Wir vermuten (die nachfolgende Rechnung bringt die Bestätigung), dass in dem Halbleiter ein Vorgang ablaufen wird, der die Neutralität wieder herstellt. Da dies ein Vorgang ist, der über die Zeit abläuft, muss sich ρ von seinem Anfangswert der Störung immer weiter verringern, bis er schließlich zu Null wird und zeitlich konstant bleibt. Mathematisch formuliert, bedeutet dies, dass

$$\rho = 0, \qquad \frac{\partial \rho}{\partial t} = 0 \qquad \text{für } t \to \infty \; . \tag{2.200}$$

Die dafür erforderliche zeitliche Abnahme der Raumladung ergibt sich mit der Neutralitätsbedingung Gl. (2.197) zu

$$\frac{\partial \rho}{\partial t} = e \left(\frac{\partial p}{\partial t} + \frac{\partial N_D^+}{\partial t} - \frac{\partial n}{\partial t} - \frac{\partial N_A^-}{\partial t} \right) . \tag{2.201}$$

Unter der Annahme, dass alle Dotierungsatome ionisiert sind und es auch für immer bleiben, sind deren Ladungsträgerdichten zeitlich konstant und Gl. (2.200) vereinfacht sich zu

$$\frac{\partial \rho}{\partial t} = e \left(\frac{\partial p}{\partial t} - \frac{\partial n}{\partial t} \right) . \tag{2.202}$$

Nennen wir die Zeitkonstante, mit der die Raumladung neutralisiert wird, τ_r. Dann wird für $t \gg \tau_r$ die Raumladung neutralisiert sein ($\rho = 0$) und es gilt wegen der Überlegung zu Gl. (2.199)

$$\frac{\partial \rho}{\partial t} = 0 \rightarrow \frac{\partial p}{\partial t} = \frac{\partial n}{\partial t}, \qquad t \gg \tau_r \tag{2.203}$$

$$\text{und } \rho = 0 \rightarrow \Delta p = \Delta n, \qquad t \gg \tau_r \, . \tag{2.204}$$

Wir erinnern uns an das vorangegangene Kapitel. Dort waren $\frac{\partial p}{\partial t}$ und $\frac{\partial n}{\partial t}$ als die Netto-Rekombinationsrate R (vgl. z. B. Gl. (2.157), (2.158) und (2.171)) für Löcher und Elektronen bestimmt worden. Außerhalb des thermodynamischen Gleichgewichtes ist $R \neq 0$. Wir erwarten in diesem Fall aufgrund von Gl. (2.203), dass nach Neutralisation der Raumladung ein Rekombinationsvorgang mit einer Nettorekombinationsrate ungleich Null fortbesteht.
Alternativ zur Formulierung von Gl. (2.203) kann auch mit Hilfe von Gl. (2.199) für den Endzustand nach Gl. (2.200) geschrieben werden

$$\frac{\partial \rho}{\partial t} = 0 \rightarrow \frac{\partial \Delta p}{\partial t} = \frac{\partial \Delta n}{\partial t}, \qquad t \gg \tau_r \tag{2.205}$$

$$\text{und } \rho = 0 \rightarrow \Delta p = \Delta n, \qquad t \gg \tau_r \, . \tag{2.206}$$

Wir berechnen im Folgenden den Abbau der Raumladung:
Dafür machen wir eine wichtige Annahme, die uns die Rechnung sehr vereinfacht und die in einer Vielzahl der später betrachteten Fälle erfüllt ist: Wir nehmen an, dass es sich um kleine Störungen der Ladungsträgerdichte handelt, für deren Maximalwert zum Zeitpunkt $t = 0$ gilt

$$\Delta n(x) \ll n_0, \qquad \Delta p(x) \ll p_0 \, . \tag{2.207}$$

Bezüglich der Dotierung des Halbleiters machen wir keine Annahmen. Es kann sich daher um einen n-, p- oder n_i-leitenden Halbleiter handeln. Die aus der Störung resultierende Raumladung ergibt sich gemäß Gl. (2.199) zu

$$\rho(x,t) = e(\Delta p(x,t) - \Delta n(x,t)) \; . \tag{2.208}$$

Da der Abbau der Raumladung eine Generation bzw. Rekombination von Ladungsträgern erfordert, bauen wir die Lösung sinnvollerweise auf die bereits in Kap. 2.12.13 besprochene Kontinuitätsgleichung (2.124)

$$-\frac{dJ(x)}{dx} = \frac{\partial \rho(x,t)}{\partial t} \tag{2.209}$$

auf. Der darin, in der Stromdichte enthaltene Ladungsträgerstrom übernimmt den notwendigen An- und Abtransport der Ladungsträger. Das Stichwort „Transport von Ladungsträgern“ zeigt direkt den nächsten notwendigen Schritt zur Lösung an. Die Stromdichte kann mit Hilfe der Transportgleichung (2.114) ermittelt werden. Sie lautet allgemein

$$J = e(n\mu_n + p\mu_p)E + e\left(D_n\frac{dn}{dx} - D_p\frac{dp}{dx}\right) \; . \tag{2.210}$$

Zur Wahrung der Übersichtlichkeit haben wir die Kennzeichnung der Orts- und Zeitabhängigkeit weggelassen. Es gilt aber für $J = J(x,t)$, $n = n(x,t)$, $p = p(x,t)$ und $E = E(x,t)$.

Wir haben für die Berechnung den Fall kleiner Störungen entsprechend Gl. (2.207) zugrunde gelegt. Daher ist die Abweichung der Ladungsträgerdichte von den Gleichgewichtswerten n_0, p_0 überall (für alle x) gering. Der Anteil des Diffusionsstromes in Gl. (2.114) ist daher wegen $\frac{dn}{dx} \approx 0$ und $\frac{dp}{dx} \approx 0$ vernachlässigbar. Es bleibt der Feldstrom aufgrund der Feldstärke E. Für die Ladungsträgerdichten n, p können wegen der Annahme (2.207) in guter Näherung die Gleichgewichtsdichten n_0, p_0 verwendet werden. Es ergibt sich daher mit der Definition der Leitfähigkeit σ aus Gl. (2.87) das Ohmsche Gesetz:

$$J = e(n_0\mu_n + p_0\mu_p)E = \sigma E \tag{2.211}$$

bzw. in der zum Einsetzen in Gl. (2.209) geeigneten Form

$$\frac{dJ}{dx} = \sigma\frac{dE}{dx} \; , \tag{2.212}$$

wobei wir eine homogene Dotierung annehmen, so dass σ keine Ortsabhängigkeit aufweist.
Wir müssen als nächsten Lösungsschritt die elektrische Feldstärke E bestimmen. Sie bewirkt gemäß Gl. (2.211) oder (2.212) durch die Leitfähigkeit σ einen Ladungsträgerstrom (Stromdichte J), der nach der Gesetzmäßigkeit der 3. Maxwellschen Gleichung zum Abbau der Raumladungsdichte führt.
Aus der 3. Maxwellschen Gleichung (2.118) wissen wir, dass eine Raumladungsdichte ρ die Quelle (Ursache) eines elektrischen Feldes sein muss. In unserem ansonsten neutralen Halbleiter befindet sich als einzige Raumladung die Störung nach Gl. (2.208) selbst. Wir setzen also diese Störung in die 3. Maxwellschen Gleichung ein, die damit

$$\varepsilon \frac{dE(x,t)}{dx} = \rho(x,t) \tag{2.213}$$

lautet.
Wir haben damit einen geschlossenen Regelkreis identifiziert: Die Raumladungsstörung erzeugt ein elektrisches Feld (Gl. (2.213)), das einen Stromfluss von Ladungsträgern bewirkt (Gl. (2.212)), der zu einer Neutralisierung der Raumladungsstörung (Gl. (2.209)) führt.
Wir bringen diese Gleichungen in einen Zusammenhang. Dazu setzen wir Gl. (2.213) in (2.212) und diese wiederum in Gl. (2.209) ein und erhalten die Differentialgleichung

$$\frac{\sigma}{\varepsilon}\rho + \frac{\partial \rho}{\partial t} = 0 \; . \tag{2.214}$$

Deren Lösung ist bekannt und lautet

$$\rho(x,t) = \rho(x,0) e^{\frac{-t}{\tau_r}} \; . \tag{2.215}$$

Darin ist mit

$$\tau_r := \frac{\varepsilon}{\sigma} \tag{2.216}$$

die so genannte „dielektrische Relaxations-Zeitkonstante“ definiert worden.
Die Lösung besagt, dass eine Raumladungsstörung exponentiell mit der Zeitkonstanten τ_r abgebaut wird. Als Näherung gilt für $t > 3\tau_r$:

$$t > 3\tau_r : \qquad \rho(x,t) \approx 0 \tag{2.217}$$

Das ist das Ergebnis, was wir bereits in Form von Gl. (2.200) zuvor überlegt hatten. Es ist wegen Gl. (2.199) gleichbedeutend mit der wichtigen Neutralitätsbedingung nach Abschluss der Relaxation

$$t > 3\tau_r : \qquad \Delta p(x,t) = \Delta n(x,t) \tag{2.218}$$

Das Ergebnis in dieser Formulierung besagt, dass es gar keinen Abbau der Ladungsträger gegeben hat. Vielmehr sind die Ladungsträger neutralisiert worden, indem sich für $t > 3\tau_r$ in jedem Ort und zu jeder Zeit gleich viel Ladungsträger entgegengesetzter Ladung (Löcher Δp, Elektronen Δn) befinden.
Dieser mit Relaxation bezeichnete Vorgang läuft sehr schnell ab. Er liegt im Bereich $0{,}1 \ldots 10\,\mathrm{ps}$. Außer im Bereich der Höchstfrequenztechnik oder für integrierte Hochgeschwindigkeitsschaltungen ist der Relaxationsvorgang im Vergleich zu allen anderen ablaufenden Prozessen so schnell, dass er als sofort abgeschlossen betrachtet werden kann. Alle anderen Vorgänge im Halbleiter erfolgen dann auf Basis der durch die Relaxation eingestellten Neutralitätsbedingung Gl. (2.218).
Zur Vollständigkeit sei angemerkt, dass der Relaxationsvorgang bei Minoritäten- und Majoritätenstörung unterschiedlich abläuft. Prinzipiell gibt es zwei Möglichkeiten der Neutralisation:
Zum einen kann eine Raumladung aufgrund der zwischen ihren gleichen Ladungsträgern herrschenden abstoßenden Kräfte auseinanderdriften, wodurch die Konzentration der Ladungsträger (Dichten Δn, Δp) geringer wird. Es genügen dann wenige(r) Ladungsträger der anderen Polarität zur Neutralisation.
Die andere Möglichkeit besteht darin, dass sich Ladungen im Halbleiter so verschieben (zum Ort der Raumladungsstörung oder davon weg), dass die sich noch am gleichen Ort befindende Raumladung neutralisiert wird.
Der Vorgang der zweiten Möglichkeit läuft bevorzugt bei Minoritätenstörungen ab. Hier sind die zur Neutralisierung benötigten Ladungsträger in der Überzahl (Majoritäten), so dass schon eine sehr geringe Verschiebung von ihnen zum Ausgleich führt.
Bei Majoritätsstörungen ist dies nicht möglich, da nicht genügend Ladungsträger zur Neutralisierung vorhanden sind. In diesem Fall läuft der Vorgang entsprechend der ersten geschilderten Möglichkeit ab.
Im Prinzip können beide Möglichkeiten als entgegengesetzt ablaufende Vorgänge gesehen werden. Bei Möglichkeit eins driftet eine hohe lokale Konzentration auseinander und wird im umgebenden Raum neutralisiert. Bei Möglichkeit zwei verschieben (konzentrieren) sich im umgebenden Raum die Ladungsträger so, dass sich lokal die Neutralisation einstellt.
Beachten: Beide Verschiebungen (Drift der Ladungsträger) erfolgen aufgrund der von der Raumladung ausgehenden (Coulomb-)Kraft. Es findet in dem

hier betrachteten Fall keine Diffusion statt. Diese ist durch die Ortsabhängigkeit der Ladungsträgerdichte zwar vorhanden, ist aber, wie zuvor festgestellt wurde, gegenüber dem Driftstrom aufgrund des elektrischen Feldes vernachlässigbar.

2.24 Ladungsträger-Abbau durch Rekombination

Im vorangegangenen Kapitel haben wir gesehen, wie eine Raumladung durch Relaxation neutralisiert wird. Nach Abschluss der Relaxation (schnell) liegt eine neutrale Ladungsträgeransammlung vor, für die wegen der Neutralitätsbedingung $\Delta p = \Delta n$, Gl. (2.218) gilt

$$p(x) = p_0 + \Delta n = p_0 + \Delta p \tag{2.219}$$

$$n(x) = n_0 + \Delta n = n_0 + \Delta p \; . \tag{2.220}$$

Wir haben also eine Abweichung um die Dichte $\Delta n = \Delta p$ von den Gleichgewichtsdichten. Wegen der Neutralitiät gilt mit Gl. (2.209)

$$\frac{dJ}{dx} = \frac{\partial \rho}{\partial t} = 0 \; . \tag{2.221}$$

Das bedeutet, dass die Stromdichte über dem Ort konstant ist. In diesem Fall führt die Abweichung $\Delta n = \Delta p$ aufgrund der Kontinuitätsgleichungen (2.132), (2.133) zu einer Nettorekombination der Ladungsträger

$$R = -\frac{\partial \Delta n}{\partial t} = -\frac{\partial \Delta p}{\partial t} \; . \tag{2.222}$$

Je nach vorherrschendem Rekombinationsprozess ist für R die entsprechende Beziehung einzusetzen (z. B. Gl. (2.171) für SRH-Rekombination).
Wir wollen im Folgenden den Fall untersuchen, dass die den bisherigen Berechnungen zugrunde liegenden kleinen Störungen gegenüber den Gleichgewichtswerten Störungen in der Minoritätsträgerdichte sind. Dann können wir für die Nettorekombination R in allen Fällen die einfachen Modellgleichungen (2.188) und (2.189) verwenden. Es gilt dann für eine Minoritätsträgerstörung Δn (die durch eine entsprechende Majoritätsträgerdichte Δp neutralisiert wurde) in einem p-Halbleiter

$$R_n = R_p = R = \frac{\Delta n}{\tau_n} = -\frac{\partial \Delta n}{\partial t} \;\Leftrightarrow\; \frac{\Delta n}{\tau_n} + \frac{\partial \Delta n}{\partial t} = 0 \tag{2.223}$$

bzw. bei Störung mit Δp in einem n-Halbleiter

$$R_p = R_n = R = \frac{\Delta p}{\tau_n} = -\frac{\partial \Delta p}{\partial t} \;\Leftrightarrow\; \frac{\Delta p}{\tau_p} + \frac{\partial \Delta p}{\partial t} = 0 \; . \tag{2.224}$$

Diese Differentialgleichungen sind vom gleichen Typ wie zuvor bei der Berechnung des Relaxationsvorgangs (vgl. Gl. (2.212)). Anstelle von ρ steht hier $\Delta n, \Delta p$ und anstelle τ_r stehen τ_n, τ_p. Die Lösung lautet demenstprechend für den Abbau einer Minoritätsträgerstörung Δn durch Rekombination

$$\Delta n(x,t) = \Delta n(x,0) e^{-\frac{t}{\tau_n}} \; , \tag{2.225}$$

wobei der Zeitpunkt $t = 0$ hier mit dem Abschluss der Relaxation (nach ca. $3\,\tau_r$) beginnt. Entsprechend erfolgt der Abbau einer Minoritätsträgerstörung Δp

$$\Delta p(x,t) = \Delta p(x,0) e^{-\frac{t}{\tau_p}} \; . \tag{2.226}$$

Nach Ablauf einer Zeit von ca. $t > 3\tau_n$, $3\tau_p$ gilt dann

$$t > 3\tau_n, 3\tau_p : \; \Delta n = 0, \; \Delta p = 0 \; , \tag{2.227}$$

d. h. die Minoritätsträgerstörung ist abgebaut (sie ist nicht mehr vorhanden). Der Abbau erfolgt nach Gl. (2.222), wonach sich eine Nettorekombinationsrate $R = r - g \neq 0$ einstellt. Je nachdem welche Art der Störung abgebaut werden muss, werden mehr Ladungsträger generiert als rekombiniert bzw. andersherum. Dies ist der gleiche Sachverhalt, wie er bereits in der Einführung dieses Kapitels im Zusammenhang mit Gl. (2.138)–(2.140) erläutert wurde.
Die Lebensdauer-Zeitkonstanten der Minoritäten liegen in den Zeitspannen

$$\begin{aligned} \textbf{Si} &: 10^{-10}...10^{-3}\,\mathrm{s} \\ \textbf{Ge} &: 10^{-6}...10^{-3}\,\mathrm{s} \\ \textbf{GaAs} &: 10^{-10}...10^{-8}\,\mathrm{s} \end{aligned}$$

Bei **Si** und **Ge** hängen sie sehr stark von der Art und Dichte der Rekombinationszentren ab. Bei **GaAs** (direkter Halbleiter) ist die Abhängigkeit geringer, da dort die direkte Rekombination dominiert.
Da die Lebensdauer-Zeitkonstanten in der Regel um mehrere Größenordnungen größer als die Relaxations-Zeitkonstante ist, können zur Vereinfachung diese beiden Vorgänge getrennt und nacheinander ablaufend behandelt werden.

2.25 Bändermodell außerhalb des thermodynamischen Gleichgewichts

In den vergangenen Kapiteln haben wir bereits ausführlich mit den in Kap. 2.12.2 definierten Überschussladungsträgerdichten n und p gearbeitet. Dort wurden auch bereits die beiden Quasi-Fermi-Niveaus W_{Fn} und W_{Fp} eingeführt, mit denen n und p geschrieben werden können als

$$n = n_0 + \Delta n = N_C \, e^{-\frac{W_C - W_{Fn}}{kT}} \tag{2.228}$$

$$p = p_0 + \Delta p = N_V \, e^{-\frac{W_{Fp} - W_V}{kT}} \; . \tag{2.229}$$

Wir wollen im Folgenden Lage und Verlauf der Quasi-Fermi-Niveaus im Bändermodell ermitteln. Damit lassen sich dann aufgrund der gleichen Struktur von Gl. (2.228) und (2.229) wie bei den Gleichgewichtsdichten in Gl. (2.23), (2.26) die gleichen einfachen Überlegungen für die Überschussladungsträgerdichten n, p wie für die Gleichgewichtsdichten n_0, p_0 durchführen.
Wir setzen wieder eine eindimensionale Ortsabhängigkeit im Halbleiter voraus, wodurch die Gradientenbildung zu einer Ortsableitung nach x wird. Wir wollen im Folgenden immer mit Bändermodellen arbeiten, deren x-Achse tatsächlich auch die ortsabhängige x-Koordinate des Halbleiters repräsentiert[20]. Aufgrund dieser Konvention verzichten wir in der Regel auf die explizite Darstellung und Bezeichnung der Achse.
Wir wollen im Folgenden Konstruktionsvorschriften für die Lage der Bandkanten und Quasi-Ferminiveaus im Bändermodell herleiten. Die Steigung der Energieverläufe W_V, W_C, W_{Fn}, W_{Fp} in dem Bändermodell entspricht der Ableitung nach dem Ort x. Wir leiten daher Gl. (2.228) und (2.229) nach x ab und erhalten

$$\frac{dn}{dx} = \underbrace{N_C \, e^{-\frac{W_C - W_{Fn}}{kT}}}_{n} \frac{1}{kT} \left(\frac{dW_{Fn}}{dx} - \frac{dW_C}{dx} \right) \tag{2.230}$$

$$\frac{dp}{dx} = \underbrace{N_V \, e^{-\frac{W_{Fp} - W_V}{kT}}}_{p} \frac{1}{kT} \left(\frac{dW_V}{dx} - \frac{dW_{Fp}}{dx} \right) \; . \tag{2.231}$$

Wir stellen das Ergebnis nach den Steigungen der Quasi-Fermi-Energien um und multiplizieren dieses Ergebnis noch mit der jeweiligen Beweglichkeit, um

[20] Falls anstelle der x-Koordinate ein Bändermodell mit einer Darstellung über den Wellenvektor verwendet wird, werden wir an der entsprechenden Stelle ausdrücklich darauf hinweisen.

einen Vergleich mit den Transportgleichungen zu ermöglichen:

$$\mu_n\, n\, \frac{dW_{Fn}}{dx} = \mu_n\, n\, \frac{dW_C}{dx} + \mu_n\, kT\, \frac{dn}{dx} \tag{2.232}$$

$$\mu_p\, p\, \frac{dW_{Fp}}{dx} = \mu_p\, p\, \frac{dW_V}{dx} - \mu_p\, kT\, \frac{dp}{dx}\ . \tag{2.233}$$

Nach Gl. (1.61) ist der Zusammenhang zwischen Energie W und zugehörigem Potential φ eines Elektrons

$$W = -e\,\varphi + const. \tag{2.234}$$

Die Konstante berücksichtigt den frei wählbaren Energienullpunkt.
Über den Zusammenhang zwischen Potential und elektrischem Feld

$$E = -\frac{d\varphi}{dx} \tag{2.235}$$

wird daraus[21]

$$\frac{dW}{dx} = e\,E\ . \tag{2.236}$$

Dies ist eine wichtige Konstruktionsregel für das Bändermodell. Sie besagt, dass die Steigung des Leitungsbandes $\frac{dW_C(x)}{dx}$ an einem Ort x proportional zur Feldstärke $E(x)$ an diesem Ort ist.
Da W_C und W_V immer um ein in erster Näherung konstantes W_g auseinander liegen, weisen beide Verläufe die gleiche Steigung auf, so dass mit Gl. (2.236) gilt

$$\frac{dW_C}{dx} = \frac{dW_V}{dx} = e\,E\ . \tag{2.237}$$

Mit dieser Beziehung können die Verläufe von W_C und W_V in Gegenwart eines elektrischen Feldes konstruiert werden. Zeigt z. B. E in positive x-Richtung, steigen die Verläufe mit $e\,E$ in dieser Richtung an. Beide Bandkanten verlaufen parallel zueinander im Abstand W_g.
Wir leiten noch eine Beziehung für die Bestimmung der Steigung der Quasiferminiveaus ab:
Setzen wir Gl. (2.237) in (2.232), (2.233) ein, so erhalten wir mit D_n, D_p nach

[21] Zu beachten ist, dass Gl. (2.236) aufgrund der Herleitung nur für die Energien der Leitungs- und Valenzbandkanten gilt. Für die Quasi-Fermi-Niveaus sind in jedem Fall Gl. (2.232), (2.233) auszuwerten, die nur im Fall eines reinen Feldstromes eine Steigung gemäß Gl. (2.236) aufweisen.

Gl. (2.106), (2.107) auf der rechten Seite die bekannten Stromtransportgleichungen aus Gl. (2.112), (2.113)

$$\mu_n\, n\, \frac{dW_{Fn}}{dx} = \mu_n\, n\, e\, E + e\, D_n\, \frac{dn}{dx} = J_n \tag{2.238}$$

$$\mu_p\, p\, \frac{dW_{Fp}}{dx} = \mu_p\, p\, e\, E - e\, D_p\, \frac{dp}{dx} = J_p\ . \tag{2.239}$$

Mit diesen Gleichungen können wir den Verlauf (Steigung) der Quasi-Fermi-Niveaus aus den Ladungsträgerströmen J_n, J_p ermitteln. Durch die Fallunterscheidung $E = 0$ oder $\frac{dn}{dx} = 0$, $\frac{dp}{dx} = 0$ kann dabei auch zwischen reinem Feld- bzw. Diffusionsstrom unterschieden werden.

Abb. 2.31 zeigt die unterschiedlichen Verläufe im Bändermodell, die anhand von Gl. (2.237), (2.238) und (2.239) ermittelt werden können. Dabei wird, wie bei allen folgenden Bändermodellen auch, nur Wert auf eine qualitativ richtige Darstellung gelegt. Die maßstäbliche Darstellung einer bestimmten Steigung ist in der Regel nicht erforderlich. Auf der linken Seite ist das Bändermodell für den Fall eines reinen Feldstroms dargestellt. Aus Gl. (2.237), (2.238) und (2.239) ergeben sich die gleichen Steigungen für alle Verläufe. Abb. 2.31 Mitte zeigt der Fall eines reinen Diffusionsstroms aus Löchern (Minoritäten) in einem n-Halbleiter. Wegen $E = 0$ verlaufen nach Gl. (2.237) W_C und W_V horizontal. Aus Gl. (2.238), (2.239) folgt

$$\frac{dW_{Fn}}{dx} \sim \frac{1}{n}\,\frac{dn}{dx}, \qquad\qquad \frac{dW_{Fp}}{dx} \sim \frac{-1}{p}\,\frac{dp}{dx}\ . \tag{2.240}$$

Bei homogener Dotierung gilt $\frac{dn}{dx} = \frac{dn_0}{dx} + \frac{d\Delta n}{dx} = \frac{d\Delta n}{dx}$ bzw. $\frac{dp}{dx} = \frac{d\Delta p}{dx}$. Aus der Neutralitätsbedingung ($\Delta n = \Delta p$) folgt $\frac{dp}{dx} = \frac{dn}{dx}$. Da es sich um einen n-Halbleiter handelt, gilt $n \gg p$ und daher gilt wegen Gl. (2.240) $\frac{dW_{Fn}}{dx} \ll \frac{dW_{Fp}}{dx}$, d. h. die Steigung von W_{Fn} ist vernachlässigbar, während W_{Fp} eine um $\frac{n}{p}$ größere Steigung aufweist. Zu Beginn der Diffusionsstrecke ($x = 0$) liegen W_{Fn} und W_{Fp} entsprechend der Abweichungen Δn, Δp nach Gl. (2.71) um $kT \ln \frac{pn}{n_i^2}$ auseinander. Ist nach einer ausreichenden Strecke die Abweichung abgebaut, fallen W_{Fp} und W_{Fn} zusammen. Die rechte Seite von Abb. 2.31 zeigt den analog zu behandelnden Fall des Minoritätsstroms aus Elektronen in einem p-Halbleiter.

Hilfreich: Oft muss man die Richtung der Steigung („nach rechts oder links steigend“) der Quasi-Fermi-Energien aus der Stromrichtung bestimmen

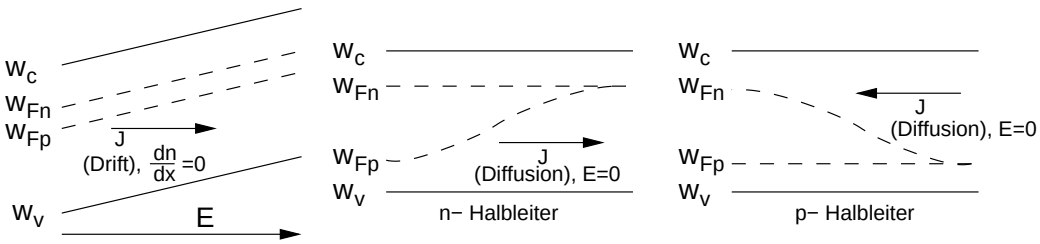

Abb. 2.31: Energieverläufe im Bändermodell außerhalb des thermodynamischen Gleichgewichts für reinen Feldstrom (Links) und Diffusionsströme in n- und p-Halbleiter (Mitte und Rechts).

(oder umgekehrt). Hilfreich ist dabei wieder die Vorstellung der Löcher als Luftblasen. Sie sind in unserem Modell eigentlich im Valenzband lokalisiert. Hier stellen wir uns als Eselsbrücke die Löcher unterhalb der Quasi-Fermi-Energien vor. Als Luftblasen steigen sie entlang eines steigenden Verlaufs empor. Ihre Richtung stimmt dabei mit der Stromflussrichtung (vgl. Abb. 2.16) überein. Dies kann z. B. anhand Abb. 2.31 überprüft werden. Natürlich kann diese Merkhilfe auch anhand von Elektronen erfolgen. Dazu betrachtet man die Elektronen wieder als Kugeln, die auf einer fallenden Bandkante (in Richtung niedrigerer Energie) herunterrollen. Setzen wir die Kugeln auf die Quasi-Fermi-Niveaus, so rollen sie entsprechend der Definition entgegen der Stromflussrichtung nach unten.

Thermodynamisches Gleichgewicht: Hier fallen die Quasi-Fermi-Energien mit der Fermi-Energie zusammen, d. h. es gilt $W_{Fn} = W_{Fp} = W_F$. Da im thermodynamischen Gleichgewicht die Ströme $J_n = 0$, $J_p = 0$ Null sind, muss gelten

$$\frac{dW_F}{dx} = 0 \,, \tag{2.241}$$

d. h. die Fermi-Energie verläuft in allen Bändermodellen im thermodynamischen Gleichgewicht waagerecht. Dies ist eine sehr wichtige Hilfe bei der Konstruktion von Bändermodellen. Da Gl. (2.241) unabhängig von der Art der Dotierung gilt, hat sie auch bei der Kontaktierung von verschiedenen Halbleitern (z. B. p-n-Diode) oder bei Metall-Halbleiter-Übergängen ihre Gültigkeit.

2.26 Drift-Diffusions-Modell (DDM)

Das Drift-Diffusions-Modell (DDM) besteht aus einem System von Annahmen und Gleichungen, die zur Beschreibung und Berechnung von Halbleiterbauelementen verwendet werden. Abhängig vom Schwerpunkt der Betrachtungen gibt es verschiedene, problemangepasste Drift-Diffusions-Modelle. Wir verwenden hier ein einfaches, aber für die Beschreibung der grundlegenden Eigenschaften der betrachteten Halbleiterelemente ausreichendes DDM. Eine wesentliche Einschränkung, die für unsere Betrachtungen jedoch ohne Auswirkung ist, besteht darin, dass der Effekt des velocity-overshoots (Ladungsträger können eine Geschwindigkeit größer als ihre Sättigungsgeschwindigkeit annehmen, vgl. Kap. 2.12.10) nicht berücksichtigt wird. Dieser Effekt schränkt die Anwendung des DDMs z. B. bei MESFETs und MOS-Transistoren mit Kanallängen im Sub-Mikrometerbereich (z. B. $0,3\,\mu$m) ein. Hierfür eignet sich ein erweitertes System von Differentialgleichungen, die die Energiebilanz und den Energiefluss berücksichtigen, das sogenannte hydrodynamische Modell (HDM).
Wir werden ein DDM mit 10 Variablen und 10 Bestimmungsgleichungen verwenden, das auf den folgenden Annahmen und Näherungen basiert:

1) Störstellenerschöpfung: Alle Dotierungsatome sind ionisiert.

2) Keine Entartung: Die Fermi-Energie liegt um mindestens $3 \cdot kT$ von den Bandkanten entfernt in der Bandlücke.

3) Stationärer Zustand: Alle Größen sind zeitunabhängig.

4) Konstante Temperatur: Die Temperatur ist im gesamten Halbleiter konstant.

5) Eindimensionales Modell: Alle Variablen haben nur eine Abhängigkeit von der x-Koordinate. In y- und z-Richtung sind die Halbleiterparameter konstant (homogen).

6) Die Energielücke W_g ist ortsunabhängig. Es gilt $\frac{dW_g}{dx} = 0$.

Die 10 Variablen des DDMs sind

$\rho(x)$,	Raumladungsdichte,
$n(x)$,	(Überschuss-)Elektronendichte ($n = n_0 + \Delta n$),
$p(x)$,	(Überschuss-)Löcherdichte ($p = p_0 + \Delta p$),
$E(x)$,	elektrische Feldstärke,
$\varphi(x)$,	elektrisches Potential,
$W_C(x)$,	Energie der Leitungsbandkante ($W_V(x) = W_C(x) + W_g$),
$W_{Fn}(x)$,	Quasi-Fermi-Energie für Elektronen,
$W_{Fp}(x)$,	Quasi-Fermi-Energie für Löcher,
$J_n(x)$,	Stromdichte des Elektronen-Stroms,
$J_p(x)$,	Stromdichte des Löcher-Stroms.

Die 10 Bestimmungsgleichungen der Variablen sind die Gleichungen für

Raumladungsdichte, Gl. (2.119)

$$\rho = e\,(p + N_D^+ - n - N_A^-)\;, \tag{2.242}$$

Elektrische Feldstärke (3. Maxw.), Gl. (2.118)

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon}\;, \tag{2.243}$$

Elektrisches Potential, Gl. (2.235)

$$\frac{d\varphi}{dx} = -E\;, \tag{2.244}$$

Feldenergie, Gl. (2.237)

$$\frac{dW_C}{dx} = \frac{dW_V}{dx} = e\,E\;, \tag{2.245}$$

Ladungsträgerdichten, Gl. (2.228), (2.229)

$$n = N_C\, e^{-\frac{W_C - W_{Fn}}{kT}}\;, \tag{2.246}$$

$$p = N_V\, e^{-\frac{W_{Fp} - W_V}{kT}}\;, \tag{2.247}$$

Ladungsträger-Transport, Gl. (2.112), (2.113)

$$J_n = e\left(n\mu_n E + D_n \frac{dn}{dx}\right) , \tag{2.248}$$

$$J_p = e\left(p\mu_p E - D_p \frac{dp}{dx}\right) , \tag{2.249}$$

mit $D_{n/p} = U_T\,\mu_{n/p}$ nach Gl. (2.106), (2.107) und Ladungsträger-Kontinuität, Gl. (2.132), (2.133)

$$\frac{dJ_n}{dx} = eR \approx e\,\frac{p - p_0}{\tau_p} , \tag{2.250}$$

$$\frac{dJ_p}{dx} = -eR \approx e\,\frac{n - n_0}{\tau_n} . \tag{2.251}$$

Die Näherungen in Gl. (2.250), (2.251) gelten für schwache Überschüsse ($\Delta p = p - p_0$ bzw. $\Delta n = n - n_0$) von Minoritätsträgern nach Gl. (2.188), (2.189). Genauer sind Gl. (2.171) bzw. (2.185) für indirekte Band-Band-Übergänge (SRH, Auger) bzw. Gl. (2.151) für direkte Band-Band-Übergänge.

3 p-n-Übergang

3.1 Struktur und Betrieb von p-n-Übergängen

Bisher haben wir Halbleiter mit homogener n- oder p-Dotierung in x-Richtung betrachtet (vgl. z. B. Abb. 2.16)[22]. Ein solcher Halbleiter weist prinzipiell immer die gleichen Eigenschaften für einen Strom in $+x$- und $-x$-Richtung auf. Dies ändert sich, wenn der Halbleiter, wie in Abb. 3.1 gezeigt, aus einem n- und einem p-dotierten Bereich besteht. Hier können wir

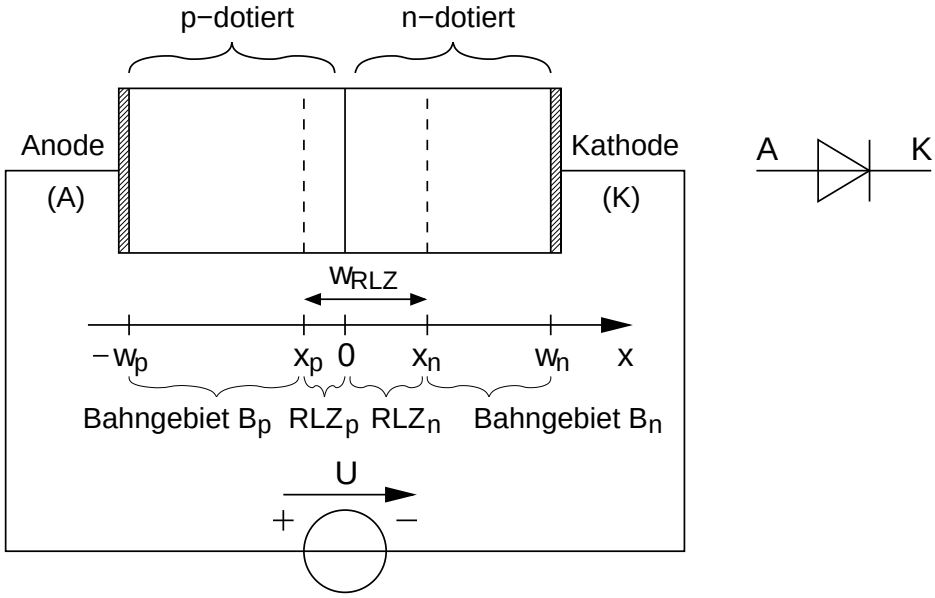

Abb. 3.1: p-n-Übergang mit Vorspannung U in Flussrichtung. Auf der rechten Seite ist das Schaltungssymbol des als Diode bezeichneten Übergangs dargestellt.

erwarten, dass sich bei unterschiedlicher Polung der Vorspannung U auch unterschiedliche Eigenschaften des Halbleiters ergeben.
Ein Halbleiter mit diesem Aufbau wird als p-n-Diode oder kürzer meist nur als Diode bezeichnet. Das Schaltungssymbol der Diode ist in Abb. 3.1 rechts gezeigt. Der Anschluss an die p-Region wird mit Anode, der an die n-Region mit Kathode bezeichnet. Anode und Kathode stellen jeweils einen Metall-Halbleiterkontakt dar, den wir zunächst als idealen Kontakt annehmen wol-

[22] Eine homogene Dotierung in y- und z-Richtung liegt in unserer vereinfachten eindimensionalen Betrachtungsweise immer vor.

len.
Im Folgenden werden wir die Eigenschaften der p-n-Diode mit Hilfe der bereits ermittelten Gleichungen des Drift-Diffusions-Modells herleiten. Wir werden sehen, dass die wesentlichen Eigenschaften der Diode durch den Bereich des Übergangs zwischen p- und n-Gebiet bestimmt werden. Diesen Bereich bezeichnen wir mit p-n-Übergang.
p-n-Übergänge bilden zusammen mit Metall-Halbleiter-Übergängen (Kontakte und Schottky-Dioden) und Metall-Oxid-Übergängen (MOS/MIS-Transistoren) die Grundlage für die Funktion aller Halbleiterbauelemente. Haben wir den p-n-Übergang verstanden, so können wir dieses Wissen z. B. direkt auf Bipolar-Transistoren (pnp-, npn-Übergänge), Feldeffekttransistoren (JFET) und Mehrschicht-Halbleiter (z. B. Thyristoren) anwenden. Dieses Kapitel bildet daher die Grundlage zum Verständnis der Funktion von Halbleiterbauelementen.

3.2 Konvention für Dichtenindizierung

Da wir im Folgenden zwischen den Ladungsträgern im p- und im n-Gebiet unterscheiden müssen, benötigen wir eine Indizierung, die Auskunft über das betrachtete Gebiet gibt. Wir verwenden daher den Index n, wenn wir eine Ladungsträgerdichte in einem n-dotierten Halbleiter angeben. p steht entsprechend als Index bei Ladungsträgerdichten bei Halbleitern mit p-Leitung. So bezeichnet z. B. n_{n0} die Gleichgewichtsdichte der Elektronen in einem n-leitenden Halbleiter. n_n ist demnach die gesamte Elektronendichte (Majoritätsträgerdichte) in diesem Halbleiter.

3.3 Modell des abrupten p-n-Übergangs

Wir verwenden für die folgenden Überlegungen ein einfaches Modell des p-n-Übergangs. Dabei nehmen wir zunächst an, dass

1. beide Halbleiterbereiche beliebig weit in $\pm x$-Richtung ausgedehnt sind (p-Halbleiter in $-x$, n-Halbleiter in $+x$ Richtung). Dies ermöglicht die Betrachtung, dass sich bei genügend großem Abstand vom p-n-Übergang (bei $x = 0$, vgl. Abb. 3.1) die bereits bekannten Eigenschaften des homogen dotierten p- oder n-dotierten Materials ergeben.

2. der Halbleiter in y- und z-Richtung homogen und unendlich ausgedehnt ist (eindimensionales Modell).

3. der Übergang zwischen p- und n-Bereich abrupt und ohne Störung des Kristallgitters erfolgt.

4. die Energielücke orts- und dotierungsunabhängig in allen Halbleiterbereichen konstant ist ($W_C(x) - W_V(x) = W_g = const.$).

Annahme 3) ist eine sehr idealisierte Vorstellung, da hierfür ein Dotierungsprofil mit einer idealen, rechteckförmigen Kante erzeugt werden muss. In der Praxis lassen sich Gauß-förmige oder lineare Übergänge realisieren. Da sich deren Eigenschaften mit der gleichen Vorgehensweise beschreiben lassen, beschränken wir uns im Folgenden auf die Darstellung des abrupten Übergangs.

3.4 Flachband-Diagramm

Wir können uns in einem Gedankenexperiment den abrupten Übergang aus zwei getrennten Teilen des gleichen Halbleitermaterials, das eine p-, das andere n-dotiert, vorstellen, die an der Stelle $x = 0$ aneinandergefügt werden. Beide Teile sollen in ihrem gesamten Bereich und an der Kontaktfläche ideal sein, so dass die Energielücke durchgehend zwischen p- und n-Gebiet gleich groß bleibt. Wir fügen zunächst die beiden Bänderdiagramme der getrennten Gebiete in Abb. 3.2 links so zusammen, dass Leitungs- und Valenzband horizontal verlaufen und erhalten die Darstellung in Abb. 3.2 rechts.

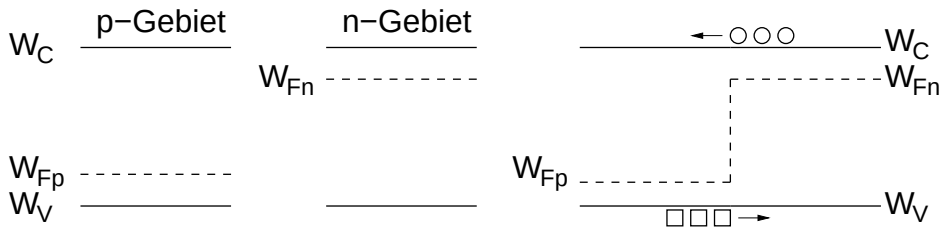

Abb. 3.2: **Links und Mitte:** Bändermodelle von p- und n-Gebiet vor dem Zusammenfügen. **Rechts:** Flachbanddiagramm als Gedankenexperiment unmittelbar nach dem Zusammenfügen der beiden Gebiete.

Aufgrund des flachen Verlaufs der Bänder wird dieses Diagramm auch Flachband-Diagramm („flatband diagram“) genannt.

Wir sehen in dem Flachband-Diagramm, dass der Verlauf der Fermi-Energie am Übergang zwischen p- und n-Bereich einen Sprung besitzt. Da

nach Gl. (2.241) im thermodynamischen Gleichgewicht die Fermi-Energie waagerecht verlaufen muss, stellt das Flachband-Diagramm in Abb. 3.2 rechts nicht das thermodynamische Gleichgewicht dar. Wir haben aus diesem Grund bereits in Abb. 3.2 die Quasi-Fermi-Energien anstelle der Fermi-Energie verwendet.
Dass der im Flachband-Diagramm gezeigte Zustand nicht das thermodynamische Gleichgewicht darstellen kann, wird auch deutlich, betrachtet man die Situation der Ladungsträger an der Stelle des Sprungs der Quasi-Fermi-Energien: Berücksichtigt man die Bedeutung der Fermi-Energie, wonach die Wahrscheinlichkeit, einen Ladungsträger oberhalb bzw. unterhalb der Fermi-Energie anzutreffen, jeweils 50% beträgt. An der Stelle des Sprungs werden daher Löcher vom p-Gebiet (sind dort Majoritäten) in das n-Gebiet (sind dort Minoritäten) „gehen“, da sie dort eine geringere Energie besitzen (Energie-Skala nimmt für Löcher nach „oben“, ab). Entsprechend werden Elektronen vom n- in das p-Gebiet „gehen“, da sie dadurch ihre Energie verringern können.
Wir erwarten daher einen Strom von Ladungsträgern über den p-n-Übergang, durch den sich das thermodynamische Gleichgewicht einstellt. Die diesem Vorgang zugrunde liegenden Ursachen und Wirkungen schauen wir uns im nächsten Kapitel an.

3.5 p-n-Übergang im thermodynamischen Gleichgewicht

Wir wissen bereits durch Gl. (2.241), dass im thermodynamischen Gleichgewicht die Fermi-Energie waagerecht im Bändermodell verlaufen muss. Wir verwenden diese Eigenschaft, um das Bändermodell des p-n-Übergangs in Abb. 3.3 zu konstruieren.
In genügend großem Abstand vom Übergang erwarten wir keine Auswirkung der Ausgleichsvorgänge am Übergang. Daher können wir in diesen Bereichen, deren Grenzen wir mit $x < -x_p$ und $x > x_n$ bezeichnen wollen, die Bändermodelle der homogenen p- und n-Gebiete aus Abb. 3.2 links und Mitte verwenden. An den zunächst noch unbekannten Stellen $-x_p$ und x_n müssen wir Valenz- und Leitungsband so verbiegen, dass sie in diesem Übergangsbereich stetig verlaufen. Die Stetigkeit folgt direkt aus

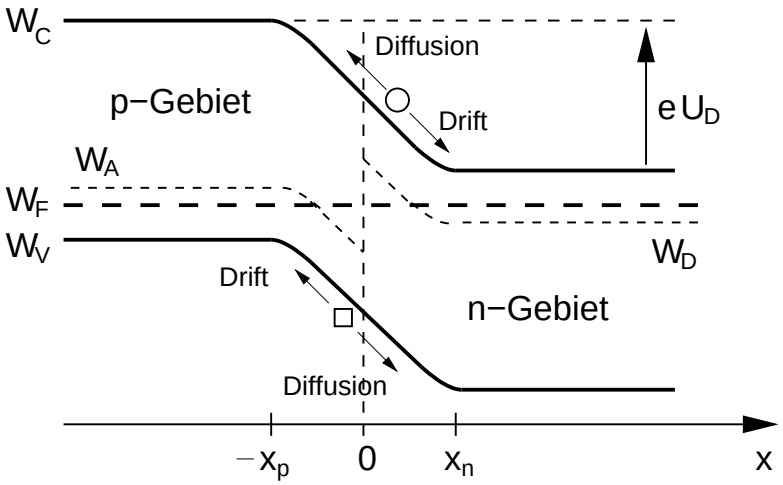

Abb. 3.3: p-n-Übergang im thermodynamischen Gleichgewicht.

Gl. (2.236), da $E(x)$ nirgendwo unendlich wird[23]. Bei der Konstruktion muss berücksichtigt werden, dass $W_C(x) - W_V(x) = W_g = const.$ im gesamten Bereich gilt.

Diese Konstruktion konnten wir rein formal basierend auf den Gesetzmäßigkeiten von Fermi-Energie und Bandverläufen herleiten. Als Resultat erhalten wir eine Ortsabhängigkeit der Bandverläufe, aus der eine Reihe von Eigenschaften resultiert, die wir im Folgenden genauer untersuchen wollen.
Wir betrachten zunächst die Ladungsträgerdichten auf beiden Seiten des p-n-Übergangs im thermodynamischen Gleichgewicht. Es gilt entsprechend der Annahme der Störstellenerschöpfung unseres DDMs:

$$\text{p-Gebiet: } p_{p0} \approx N_A^- \approx N_A \,,\; N_D = 0 \,,\; n_{p0} = \frac{n_i^2}{p_{p0}} \tag{3.1}$$

$$\text{n-Gebiet: } n_{n0} \approx N_D^+ \approx N_D \,,\; N_A = 0 \,,\; p_{n0} = \frac{n_i^2}{n_{n0}} \tag{3.2}$$

[23]Eine unendlich große Feldstärke wäre nur theoretisch im Innern von elektrischen Ladungen, z. B. bei Oberflächen- oder Grenzflächenladungen, möglich. Diese sind jedoch in dem einfachen, zugrunde gelegten Modell nicht enthalten.

Beispiel: (Abb. 3.5 a)):
Mit $N_A = 10^{17}$ und $N_D = 2 \cdot 10^{18}$ ergibt sich für **Si** ($n_i = 1,5 \cdot 10^{10}\text{cm}^{-3}$)

$$p_{p0} = 10^{17}\text{cm}^{-3},\ n_{p0} = 2,25 \cdot 10^{3}\text{cm}^{-3}$$

$$p_{n0} = 1,13 \cdot 10^{2}\text{cm}^{-3},\ n_{n0} = 2 \cdot 10^{18}\text{cm}^{-3}$$

Das Beispiel verdeutlicht, dass ein extrem starkes Konzentrationsgefälle der Ladungsträger zwischen beiden Seiten des Übergangs herrscht. Die Löcherdichte als Majoritätsträgerdichte im p-Gebiet fällt abrupt von $10^{17}\,\text{cm}^{-3}$ auf die Minoritätsträgerdichte der Löcher im n-Gebiet von $1,13 \cdot 10^{2}\,\text{cm}^{-3}$ ab. Das sind ca. 15 Zehnerpotenzen! Die gleiche Größenordnung besitzt auch das Konzentrationsgefälle der Elektronen vom n- zum p-leitenden Gebiet.
Aus den Transport-Gleichungen (2.248), (2.249) wissen wir, dass mit einem Konzentrationsgefälle von Ladungsträgern immer ein Diffusionsstrom verbunden ist, der so gerichtet ist, dass das Konzentrationsgefälle abgebaut wird (vgl. Diffusionsstrom in Abb. 3.3).
Danach werden frei bewegliche Löcher aus dem p-Gebiet über den Übergang in das n-Gebiet diffundieren. Sie hinterlassen dabei ihre ortsfest in das Gitter eingebauten negativ geladenen Akzeptor-Ionen. Der gleiche Diffusionsvorgang erfolgt auch für Elektronen von der anderen Seite des Übergangs. Sie hinterlassen ortsfest eingebaute positiv geladene Gitter-Ionen. Abb. 3.4 zeigt diesen Vorgang.
Durch das Abwandern der beweglichen Ladungsträger von ihren ortsfesten Ionen-Rümpfen sind die davon betroffenen Halbleiterbereiche elektrisch nicht mehr neutral. Die ortsfesten Ionen können ihren neutralisierenden Partnern nicht folgen und bilden auf beiden Seiten des Übergangs eine Zone mit starker Raumladung. Wir nennen sie Raumladungszone (RLZ) und markieren ihren Anfang in p- und n-Gebiet mit $-x_p$ und x_n. Die gesamte Weite nennen wir w_{RLZ}.
Die Raumladung entsteht im Wesentlichen durch die ortsfesten Ionen-Rümpfe, da die frei beweglichen Ladungsträger als Minoritätsträger im angrenzenden Gebiet rekombinieren und damit elektrisch neutralisiert sind. Ihre Konzentration ist daher vernachlässigbar gering gegenüber der Raumladungsdichte der Ionen-Rümpfe. Diese bildet im n-Gebiet eine positive, im p-Gebiet eine negative Raumladung.
Die Raumladung ist nach Gl. (2.243) des DDMs die Ursache eines elektri-

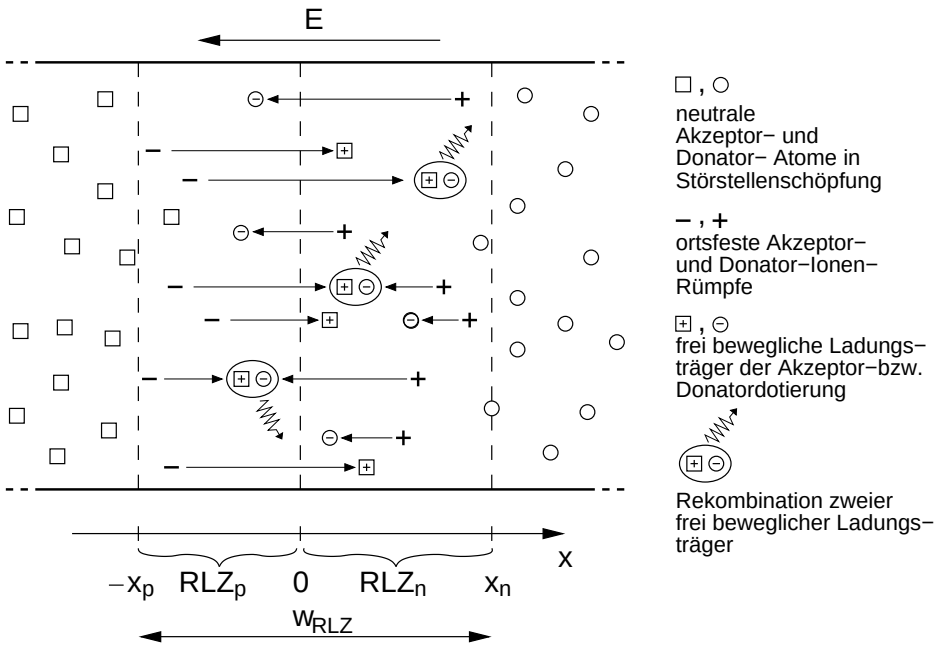

Abb. 3.4: Diffusion, Drift und Rekombination von Ladungsträgern in der Raumladungszone (RLZ) mit der Weite w_{RLZ}.

schen Feldes, dessen Wirkung darin besteht, die Raumladung abzubauen. Da die ortsfesten Ionen-Rümpfe durch das Feld nicht verschoben werden können, erfolgt die (Kraft-)Wirkung des Feldes nur auf die beweglichen, in die Raumladungszone diffundierenden Ladungsträger. Dadurch verursacht das Feld einen Driftstrom, der dem Diffusionsstrom entgegen gerichtet ist. Die Summe aus beiden Strömen bildet den Gesamtstrom an jedem Ort x entsprechend der Transportgleichung des DDMs. Da im thermodynamischen Gleichgewicht die Stromsumme gleich Null ist, muss das elektrische Feld genau so groß sein, dass dessen Driftstrom exakt den Diffusionsstrom kompensiert. Beide Ströme befinden sich dann im Gleichgewicht. Die sich dabei einstellende Weite der RLZ ist genau so groß, dass die darin enthaltene Raumladungsdichte das zur Kompensation notwendige elektrische Feld hervorruft.
Wir wollen daher im Folgenden den p-n-Übergang entsprechend Abb. 3.1 in die neutralen Bahngebiete an den beiden Enden

$$\begin{aligned} B_p &: w_p \leq x \leq x_p \\ B_n &: x_n \leq x \leq w_n \end{aligned} \tag{3.3}$$

und die Raumladungszone (RLZ)

$$x_p < x < x_n\,, \qquad w_{RLZ} = x_n - x_p$$

mit den beiden Teilen in p- und n-Gebiet

$$\begin{aligned} \mathrm{RLZ}_p &: x_p < x < 0 \\ \mathrm{RLZ}_n &: 0 < x < x_n \end{aligned} \tag{3.4}$$

unterteilen. Anstelle des Begriffs Raumladungszone wird häufig auch der Begriff Sperrschicht verwendet.

3.6 Berechnung des p-n-Übergangs im thermodynamischen Gleichgewicht

3.6.1 Rechteck-Profil-Näherung

Wir nehmen zur Vereinfachung für die folgenden Rechnungen an, dass die ortsfesten Raumladungsdichten im p- und n-Gebiet ein Rechteck-Profil wie in Abb. 3.5 b) gezeigt besitzen. Danach gilt für die Raumladungsdichte in der Diode

$$\rho(x) = \begin{cases} 0 & , \text{ in } B_p \\ -e\,N_A & , \text{ in } \mathrm{RLZ}_p \\ e\,N_D & , \text{ in } \mathrm{RLZ}_n \\ 0 & , \text{ in } B_n\,. \end{cases} \tag{3.5}$$

In dieser Störstellennäherung werden freie Ladungsträger in der Raumladungszone vernachlässigt. Man sagt zu diesem Zustand auch, die Raumladungszone ist „verarmt“ (full-depletion). Dass diese Näherung gerechtfertigt ist, erkennt man z. B. an dem starken Konzentrationsgefälle zwischen p- und n-Gebiet, wodurch im Übergangsbereich die Ladungsträgerdichte auf n_i abfällt (vgl. z. B. Bändermodell Abb. 3.3 mit W_F in Bandmitte). Außerhalb der Raumladungszone gilt aufgrund der abrupten Änderung durch das Rechteck-Profil für die Bahngebiete unmittelbar Gl. (3.1) und (3.2).
Bei bekannter Dotierung sind dadurch insbesondere die Ladungsträgerdich-

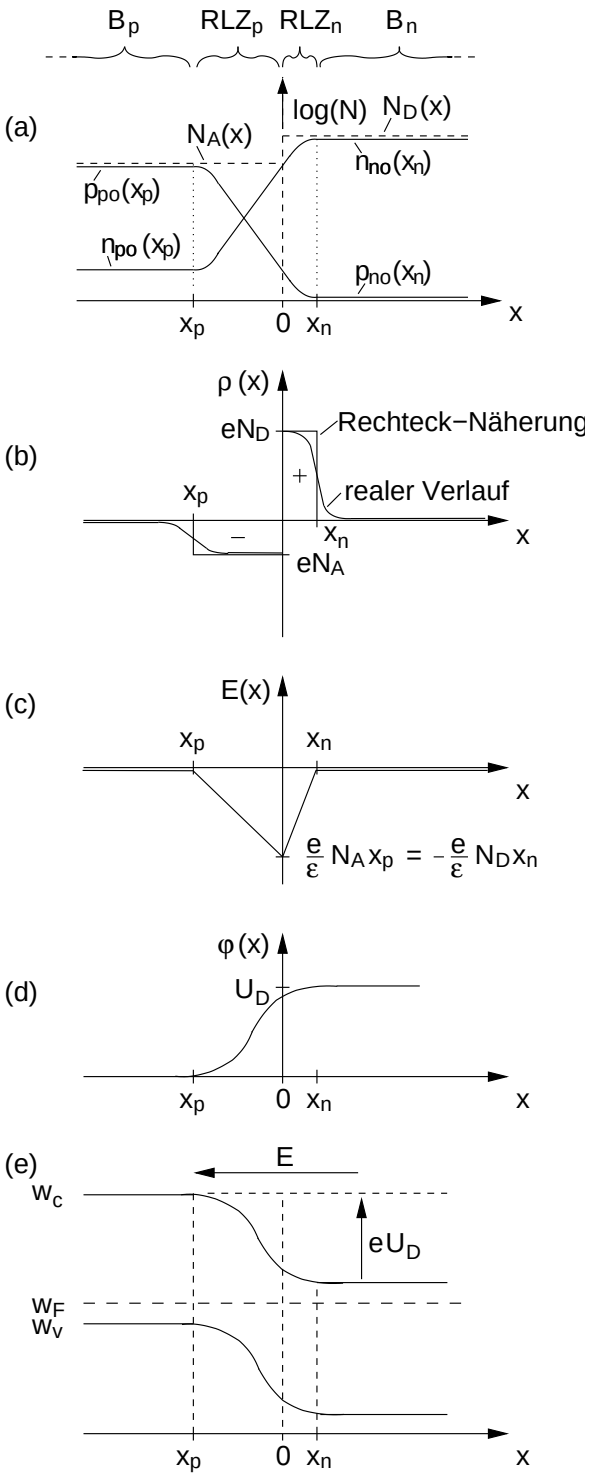

Abb. 3.5: Verläufe verschiedener Kenngrößen des p-n-Übergangs. Erläuterungen hierzu vgl. Text.

ten in den Bahngebieten bis an die Grenze zur Raumladungszone bekannt.

$$\text{Ladungsträger-dichten:} \begin{cases} p_0(x) = p_{p0} = N_A,\ n_0(x) = n_{p0} = \frac{n_i^2}{N_A} & \text{in } B_p \\ p_0(x),\ n_0(x) & \text{in RLZ}_p \text{ und RLZ}_n \\ n_0(x) = n_{n0} = N_D,\ p_0(x) = p_{n0} = \frac{n_i^2}{N_D} & \text{in } B_n \end{cases} \tag{3.6}$$

Im Folgenden werden wir die Ladungsträgerdichten insbesondere an den Grenzen der Raumladungszone bei x_n und x_p als Randbedingung bei der Bestimmung von Potentialen verwendet.

3.7 Elektrisches Feld am p-n-Übergang

Wir wissen, dass eine Raumladung die Quelle eines elektrischen Feldes ist. Zur Berechnung des Feldes verwenden wir Gl. (2.243) des DDMs, die wir auf die Raumladungsdichte in der Rechteck-Profil-Näherung nach Gl. (3.1) anwenden.

$$\frac{dE}{dx} = \frac{1}{\varepsilon} \begin{cases} 0 & , \text{ in } B_p \\ -e\,N_A & , \text{ in RLZ}_p \\ e\,N_D & , \text{ in RLZ}_n \\ 0 & , \text{ in } B_n \end{cases} \tag{3.7}$$

Wir integrieren über jedes der vier Bahngebiete getrennt, wobei wir die Feldstärke an den Rändern der Raumladungszone als Integrationskonstante im Sinne von $\int_{x_0}^{x} \frac{dE}{dx}\,dx = E(x) - E(x_0)$ (x_0 ist ausgezeichneter x-Wert mit bekannter Feldstärke) verwenden. Wir erhalten

$$\begin{aligned} E(x_p) - E(x) &= const. && \text{, in } B_p \\ E(x) - E(x_p) &= -\tfrac{e}{\varepsilon}\,N_A\,(x - x_p) && \text{, in RLZ}_p \\ E(x_n) - E(x) &= \tfrac{e}{\varepsilon}\,N_D\,(x_n - x) && \text{, in RLZ}_n \\ E(x) - E(x_n) &= const. && \text{, in } B_n \end{aligned} \tag{3.8}$$

Für die Ermittlung der Randbedingungen gelten folgende Überlegungen:

- Die Feldstärke an den Rändern x_p, x_n der Raumladungszone muss Null sein, da wir außerhalb der Raumladungszone keine Ursache für ein elektrisches Feld haben. Daraus folgt $E(x_p) = 0,\ E(x_n) = 0$.

- Da wir keinen Stromfluss in den Bahngebieten haben (thermodynamisches Gleichgewicht), muss $E(x) = 0$ sein in den beiden Bahngebieten.

Es ergibt sich dann aus Gl. (3.4)

$$E(x) = \begin{cases} 0 & , \text{ in } B_p \\ -\frac{e}{\epsilon}\, N_A\,(x - x_p) & , \text{ in } \mathrm{RLZ}_p \\ \frac{e}{\epsilon}\, N_D\,(x - x_n) & , \text{ in } \mathrm{RLZ}_n \\ 0 & , \text{ in } B_n \end{cases} . \tag{3.9}$$

Den Feldstärkeverlauf zeichnen wir in Abb. 3.5 c).

Wir haben im Übergangsbereich von p auf n bei $x = 0$ in der Definition der Raumladung eine Lücke vgl. Gl. 3.4. Unter der Annahme, dass dort keine Flächenladungsdichte vorhanden ist, ist $E(x)$ dort stetig und es gilt mit Gl. (3.9) an der Stelle $x = 0$ (beachten: x_p ist negativ)

$$-N_A\, x_p = N_D\, x_n\ . \tag{3.10}$$

Dies ist die Neutralitätsbedingung für die Raumladungen. Sie besagt, dass die Raumladungen (vgl. Raumladungsdichte in Gl. (3.1)) auf beiden Seiten des Übergangs den gleichen Betrag besitzen. Ist die Dotierung einer Seite geringer, wird dies durch eine entsprechend größere Weite der Raumladungszone ausgeglichen. Für das Zeichnen der Raumladungs-Rechtecke bei der Rechteck-Profil-Näherung bedeutet dies, dass beide Rechtecke die gleiche Fläche gemäß Gl. (3.10) besitzen müssen.

3.8 Elektrische Spannung am p-n-Übergang

Über Gl. (2.244) des DDMs ist das elektrische Feld identisch mit der negativen Ableitung des Potentials. Um das Potential zu ermitteln, integrieren wir daher das elektrische Feld in gleicher Weise wie wir zuvor die Raumladungsdichte integriert haben.

$$\frac{d\varphi(x)}{dx} = \begin{cases} 0 & , \text{ in } B_p \\ \frac{e}{\varepsilon}\, N_A\,(x - x_p) & , \text{ in } \mathrm{RLZ}_p \\ -\frac{e}{\varepsilon}\, N_D\,(x - x_n) & , \text{ in } \mathrm{RLZ}_n \\ 0 & , \text{ in } B_n \end{cases} \tag{3.11}$$

$$\begin{aligned}
&\varphi(x_p) - \varphi(x) = const. && \text{, in } B_p \\
&\varphi(x) - \varphi(x_p) = \tfrac{e}{\varepsilon}\, N_A \int_{x_p}^{x} (x - x_p)\, dx = \tfrac{e}{2\varepsilon}\, N_A\, (x - x_p)^2 && \text{, in RLZ}_p \\
&\varphi(x_n) - \varphi(x) = -\tfrac{e}{\varepsilon}\, N_D \int_{x}^{x_n} (x - x_n)\, dx = \tfrac{e}{2\varepsilon}\, N_D\, (x - x_n)^2 && \text{, in RLZ}_n \\
&\varphi(x) - \varphi(x_n) = const. && \text{, in } B_n
\end{aligned} \tag{3.12}$$

Da die Bahngebiete stromlos sind, fällt an ihnen auch keine Spannung ab. Daher gilt $\varphi(x_p) - \varphi(x) = 0$ und $\varphi(x) - \varphi(x_n) = 0$. Wir definieren die Diffusionsspannung als die Spannung zwischen dem Ende der RLZ im n-Bereich und dem Ende im p-Bereich:

$$U_D := \varphi(x_n) - \varphi(x_p)\ . \tag{3.13}$$

Eines der Potentiale ist als Bezugspotential frei wählbar. Wir wählen $\varphi(x_p) = 0$, so dass $U_D = \varphi(x_n)$ gilt. Damit vereinfacht sich der Potentialverlauf aus Gl. (3.12) zu

$$\varphi(x) = \begin{cases} 0 & \text{, in } B_p \\ \frac{e\,N_A}{2\varepsilon}\,(x - x_p)^2 & \text{, in RLZ}_p \\ -\frac{e\,N_D}{2\varepsilon}\,(x - x_n)^2 + U_D & \text{, in RLZ}_n \\ U_D & \text{, in } B_n\ . \end{cases} \tag{3.14}$$

Darin sind die Weiten x_p, x_n der Raumladungszone und die Diffusionsspannung U_D noch unbekannt und müssen im Folgenden noch berechnet werden. Formal lässt sich jedoch schon der Verlauf $\varphi(x)$ in Abb. 3.5 d) zeichnen. Er besteht aus zwei Parabelabschnitten, die für $x = 0$ stetig ineinander übergehen müssen, damit $E(0)$ einen endlichen Wert besitzt.

Berechnung der Diffusionsspannung

Wir überlegen, über welche Beziehung des DDMs wir die Diffusionsspannung berechnen können. Wir wissen, dass die Diffusionsspannung die Potentialdifferenz zwischen beiden Enden der Raumladungszone darstellt. Wir wissen auch, dass mit dieser Potentialdifferenz unmittelbar das zuvor berechnete elektrische Feld verknüpft ist. Im thermodynamischen Gleichgewicht kompensieren sich an jedem Ort x der durch dieses Feld hervorgerufene Driftstrom und der durch das Konzentrationsgefälle der

Ladungsträger hervorgerufene Diffusionsstrom. Dieser Vorgang wird durch die Transportgleichungen des DDMs beschrieben. Wir verwenden daher die Transportgleichung, um über das darin enthaltene elektrische Feld das Potential $\varphi(x)$ an der Stelle x_n und damit die Diffusionsspannung $\varphi(x_n) = U_D$ zu gewinnen.
Die Transportgleichungen (2.248), (2.249) des DDMs (Index 0, da thermodynamisches Gleichgewicht) lauten

$$J_n(x) = 0 = n_0(x)\, E(x) + U_T \frac{dn_0(x)}{dx}, \tag{3.15}$$

$$J_p(x) = 0 = p_0(x)\, E(x) - U_T \frac{dp_0(x)}{dx}. \tag{3.16}$$

Aus jeder der beiden Gleichungen lässt sich die elektrische Feldstärke bestimmen, da die Ladungsträgerdichten n_0, p_0 über das Massenwirkungsgesetz voneinander abhängen. Zur Vollständigkeit für den späteren Gebrauch rechnen wir jedoch mit beiden Gleichungen weiter. Es ergibt sich durch Umstellen:

$$-E(x) = U_T \frac{1}{n_0(x)} \frac{dn_0(x)}{dx}, \tag{3.17}$$

$$E(x) = U_T \frac{1}{p_0(x)} \frac{dp_0(x)}{dx}. \tag{3.18}$$

Wie zuvor (vgl. z. B. Gl. (3.12)) berechnet sich daraus durch Integration das Potential $\varphi(x)$. Dabei wählen wir durch die Verwendung der Integrationsgrenzen x_n, x_p die bekannten Potentiale $\varphi(x_p) = 0$ und $\varphi(x_n) = U_D$ als Integrationskonstanten.

$$-\int_x^{x_n} E(x)\, dx = \underbrace{\varphi(x_n)}_{U_D} - \varphi(x) = U_T \int_x^{x_n} \frac{dn_0(x)}{n_0(x)} = U_T \ln \frac{n_0(x_n)}{n_0(x)} \tag{3.19}$$

$$\int_{x_p}^{x} E(x)\, dx = -\varphi(x) + \underbrace{\varphi(x_p)}_{0} = U_T \int_{x_p}^{x} \frac{dp_0(x)}{p_0(x)} = U_T \ln \frac{p_0(x)}{p_0(x_p)} \tag{3.20}$$

Mit den Ladungsträgerdichten $n_0(x_n) = n_{n0}(x_n) = N_D$ und $p_0(x_p) = p_{p0}(x_p) = N_A$ entsprechend der Rechteck-Profil-Näherung aus Gl. (3.6) ergeben sich daraus unmittelbar

$$U_D - \varphi(x) = U_T \ln \frac{n_{n0}(x_n)}{n_0(x)} = U_T \ln \frac{N_D}{n_0(x)} \tag{3.21}$$

$$\varphi(x) = U_T \ln \frac{p_{p0}(x_p)}{p_0(x)} = U_T \ln \frac{N_A}{p_0(x)}\,. \tag{3.22}$$

Wir werten z. B. Gl. (3.21) für das Potential an dem anderen Ende der Raumladungszone bei $x = x_p$ aus. Dort gilt wegen unserer Wahl des Bezugspunktes $\varphi(x_p) = 0$. Die Ladungsträgerdichte $n_0(x_p)$ ist gemäß der Rechteck-Profil-Näherung $n_0(x_p) = n_{p0} = \frac{n_i^2}{N_A}$. Damit wird aus Gl. (3.21)

$$U_D = U_T \ln \frac{N_A N_D}{n_i^2} . \tag{3.23}$$

Das gleiche Ergebnis erhält man durch Auswertung von Gl. (3.22).

Beispiel:
Für $N_D = 2 \cdot 10^{18}$ cm^{-3}, $N_A = 10^{17}$ cm^{-3}, $n_i = 1,45 \cdot 10^{10}$ cm^{-3} ergibt sich bei $T = 300$ K:

$$U_T = \frac{k \cdot T}{e} = \frac{0,026 \text{ eV}}{1 \text{ eV}} \text{ V} = 26 \text{ mV}$$

$$U_D = 26 \text{ mV} \ln \frac{2 \cdot 10^{35}}{(1,45 \cdot 10^{10})^2} \approx 900 \text{ mV}$$

Beachten: Die Diffusionsspannung ist weder direkt messbar noch als Spannungsquelle zu gebrauchen. Der Grund dafür sind die Kontaktspannungen, die bei der Verbindung zweier unterschiedlich leitender Bereiche entstehen. In einem geschlossenen Stromkreis ist im thermodynamischen Gleichgewicht die Summe der Kontaktspannungen gleich Null. Beim Kontaktieren der Diode entsteht dadurch keine verwertbare Spannung.

3.9 Berechnung der Raumladungsweiten

Die Raumladungsweiten sind bereits in der Potentialgleichung (3.14) enthalten. Da wir Stetigkeit des Potentials in $x = 0$ fordern (ansonsten wäre wegen $E = -\frac{d\varphi}{dx}$ die Feldstärke unendlich), müssen die beiden Ausdrücke für $x = 0$ übereinstimmen und man erhält durch Gleichsetzen:

$$U_D = \frac{e N_A}{2 \varepsilon} x_p^2 + \frac{e N_D}{2 \varepsilon} x_n^2 \tag{3.24}$$
$$U_D = \frac{e}{2 \varepsilon} (N_A x_p^2 + N_D x_n^2) .$$

Als zweite Gleichung haben wir die Neutralisierungsbedingung (Gl. (3.10))

$$-N_A x_p = N_D x_n , \tag{3.10}$$

wodurch wir zwei Gleichungen mit zwei Unbekannten haben, durch die wir die Raumladungsweiten bestimmen können.

Einsetzen von Gl. (3.10) in (3.24) für x_n bzw. x_p liefert die gesuchten Weiten

$$x_n = \sqrt{\frac{2\,\varepsilon\,U_D}{e}\,\frac{N_A}{N_D}\,\frac{1}{N_D+N_A}} = -x_p\,\frac{N_A}{N_D} \tag{3.25}$$

$$-x_p = \sqrt{\frac{2\,\varepsilon\,U_D}{e}\,\frac{N_D}{N_A}\,\frac{1}{N_D+N_A}} = x_n\,\frac{N_D}{N_A}\,. \tag{3.26}$$

Die gesamte Raumladungsweite beträgt

$$w_{RLZ} = x_n - x_p = \sqrt{\frac{2\,\varepsilon\,U_D}{e\,(N_D+N_A)}}\left(\sqrt{\frac{N_A}{N_D}}+\sqrt{\frac{N_D}{N_A}}\right) \tag{3.27}$$

$$= \sqrt{\frac{2\,\varepsilon\,U_D}{e}\left(\frac{1}{N_A}+\frac{1}{N_D}\right)}\,. \tag{3.28}$$

Beispiel: Mit U_D=900 mV aus dem vorangegangenen Beispiel mit $N_D = 2\cdot 10^{18}\,\mathrm{cm}^{-3}$ und $N_A = 10^{17}\,\mathrm{cm}^{-3}$ beträgt $x_n \approx 5,2\cdot 10^{-9}$m, $x_p \approx -1,05\cdot 10^{-7}$m und $w_{RLZ} \approx -1,05\cdot 10^{-7}\mathrm{m} \approx -x_p$. D. h. die Weite der Raumladungszone wird in erster Näherung von der in diesem Fall 20-fach geringeren p-Dotierung bestimmt. Die Raumladungszone erstreckt sich daher näherungsweise nur in das p-Gebiet.
Als Näherung kann bei stark unterschiedlicher Dotierung wie im vorliegenden Fall die Weite der Raumladungszone für $N_D \gg N_A$ mit

$$w_{RLZ} \approx -x_p \approx \sqrt{\frac{2\,\varepsilon\,U_D}{e\,N_A}} \tag{3.29}$$

berechnet werden. Analog gilt für $N_A \gg N_D$

$$w_{RLZ} \approx x_n \approx \sqrt{\frac{2\,\varepsilon\,U_D}{e\,N_D}}\,. \tag{3.30}$$

Das Beispiel zeigt, dass die Darstellung der Verläufe in Abb. 3.5 bezüglich der Weiten x_n, x_p nicht maßstabsgerecht ist. Bei maßstabsgerechter Darstellung wäre x_n vernachlässigbar klein gegenüber x_p.

3.10 Berechnung der Ladungsträgerdichten

Aufgrund der Rechteck-Profil-Näherung gilt $n_0(x_n) = n_{n0}(x_n) = N_D$ und $p_0(x_p) = p_{p0}(x_p) = N_A$. Gl. (3.21) und (3.22) liefern nach Umstellen direkt die Ladungsträgerdichten

$$n_0(x) = n_{n0}(x_n)\, e^{\frac{\varphi(x)-U_D}{U_T}} = N_D\, e^{\frac{\varphi(x)-U_D}{U_T}} \tag{3.31}$$

$$p_0(x) = p_{p0}(x_p)\, e^{-\frac{\varphi(x)}{U_T}} = N_A\, e^{-\frac{\varphi(x)}{U_T}}\ . \tag{3.32}$$

Das dazu gehörende Potential $\varphi(x)$ haben wir bereits in Gl. (3.14) berechnet:

$$\varphi(x) = \begin{cases} 0 & , B_p \\ \frac{e\,N_A}{2\,\varepsilon}\,(x-x_p)^2 & , RLZ_p \\ -\frac{e\,N_D}{2\,\varepsilon}\,(x-x_n)^2 + U_D & , RLZ_n \\ U_D & , B_n\ . \end{cases} \tag{3.33}$$

Und auch die beiden Weiten x_n, x_p der Raumladungszone sind über Gl. (3.25) und (3.26) mit U_D nach Gl. (3.23) bekannt.
Durch Einsetzen der jeweiligen Position $x = x_p, 0, x_n$ ergeben sich z. B. die drei ausgezeichneten Ladungsträgerdichten

$$n_0(x_p) = n_{p0}(x_p) = n_{n0}(x_n)\, e^{-\frac{U_D}{U_T}} \qquad p_0(x_p) = p_{p0}(x_p) = N_A \tag{3.34}$$

$$n_0(\,0\,) = n_{n0}(x_n)\, e^{-\frac{U_D}{U_T}\left(1-\frac{N_D}{N_D+N_A}\right)} \qquad p_0(\,0\,) = p_{p0}(x_p)\, e^{-\frac{U_D}{U_T}\,\frac{N_D}{N_D+N_A}} \tag{3.35}$$

$$n_0(x_n) = n_{n0}(x_n) = N_D \qquad p_0(x_n) = p_{n0}(x_n) = p_{p0}(x_p)\, e^{-\frac{U_D}{U_T}}\ . \tag{3.36}$$

Darin haben wir mit $n_0(x_n) = n_{n0}(x_n)$ bzw. $p_0(x_p) = p_{p0}(x_p)$ formal unserer Konvention zur Dichteindizierung Genüge getan. Aufgrund der Rechteck-Profil-Näherung gilt $n_{n0}(x_n) = N_D$ und $p_{p0}(x_p) = N_A$. Wir verwenden jedoch weiterhin die ausführlichere Bezeichnungsweise, da sie mehr Information enthält (Majoritätendichte an der Stelle x_n,x_p im Gleichgewicht).

Beispiel: Durch den Exponentialterm wird $n_0(x_p)$ mit den Werten aus den vorangegangenen Beispielen um den Faktor $e^{-\frac{U_D}{U_T}} = e^{-\frac{900}{26}} \approx 10^{-15}$ gegenüber $n_0(x_n) = N_D$ reduziert.

Für $p_0(x_p) = \frac{n_i^2}{n_o(x_p)}$ ergibt sich $p_0(x_p) \approx 10^{17}\,\mathrm{cm}^{-3} = N_A$. Dieser Wert ist konsistent mit unserer Rechteckprofil-Näherung nach Gl. (3.6). An der Stelle von $x = 0$ gilt wegen $N_D \gg N_A$ $n_0(0) \approx N_D$ und $p_0(0) = \frac{n_i^2}{n_o(0)} = \frac{n_i^2}{N_D} = 1,1 \cdot 10^2\,\mathrm{cm}^{-3}$.

3.11 Berechnung der Bandverläufe

Nach Gl. (1.61) gilt $W = -e \cdot \varphi + const.$ Daher verlaufen die Bandkanten W_C und W_V mit $-e$ proportional zum Potential $\varphi(x)$ nach Gl. (3.14).
Der entsprechende Beispielverlauf ist in Abb. 3.5 e) dargestellt.

3.12 p-n-Übergang außerhalb des thermodynamischen Gleichgewichts

Formal könnte der p-n-Übergang außerhalb des thermodynamischen Gleichgewichts mit Hilfe des DDM's berechnet werden. Dies kann jedoch nur numerisch erfolgen und ergibt somit keine analytische Lösung. Wir machen daher im Folgenden eine Reihe vereinfachender Annahmen, die eine analytische Berechnung ermöglichen.

Rechteck-Profil-Näherung
Wir verwenden für alle Überlegungen den zuvor betrachteten abrupten p-n-Übergang mit der Rechteck-Profil-Näherung.

Feldfreie Bahngebiete
Die Bahngebiete, B_n, B_p (vgl. Abb. 3.1, 3.5) sind in unserer Betrachtung so niederohmig, dass bei Stromfluss nur ein vernachlässigbarer Spannungsabfall $\varphi(x)$ an ihnen entsteht. Wegen $E = -\frac{d\varphi}{dx}$ nehmen wir in unserem Modell an,

dass das elektrische Feld in den Bahngebieten vernachlässigbar ist. Es gilt also

$$E \approx 0 \,\text{in den Bahngebieten } B_n,\, B_p \,. \tag{3.37}$$

Als Folge der geringeren Feldstärke sind die Feldströme der Minoritätsträger in den Bahngebieten vernachlässigbar. Feldströme der Majoritätsträger sind aufgrund der hohen Majoritätsträgerkonzentration vorhanden.

Äußere Spannung, Boltzmann-Randbedingungen

An die Kontakte der p-n-Diode soll eine Spannung U (vgl. Abb. 3.1) angelegt werden. Die Kontaktierung des Halbleiters soll ideal sein. Da wir feldfreie Bahngebiete annehmen, liegt die äußere Spannung direkt über der RLZ. Die Wirkung der Spannung lässt sich anhand des Bändermodells in Abb. 3.6 erkennen.

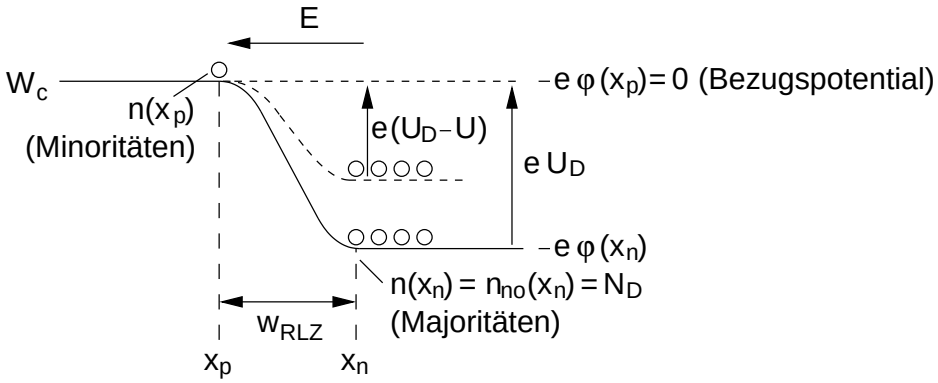

Abb. 3.6: Verlauf der Leitungsbandkante bei äußerer Spannung $U > 0$.

Wir betrachten darin nur das Leitungsband, da für Löcher im Valenzband die entsprechenden Überlegungen gelten. Der durchgezogene Verlauf zeigt die Leitungsbandkante im thermodynamischen Gleichgewicht. Entsprechend der Berechnung in Gl. (3.19) und (3.20) ist die Majoritätsträgerdichte (Elektronendichte) am Sperrschichtrand im n-Gebiet $n_{n0}(x_n) = N_D$. Dieses Ergebnis geht unmittelbar aus der Rechteck-Profil-Näherung hervor. Nach Gl. (3.34) ist die Elektronendichte auf der anderen Seite der Sperrschicht (Minoritätsträgerdichte) nur noch $n_{p0}(x_p) = N_D\, e^{\frac{-U_D}{U_T}}$ oder als Energie formuliert

$$n_{p0}(x_p) = N_D\, e^{\dfrac{-e(\varphi(x_n)-\varphi(x_p))}{kT}} = N_D\, e^{\dfrac{W_C(x_n)-W_C(x_p)}{kT}} \,. \tag{3.38}$$

(Wir hatten $\varphi(x_p) = 0$ gesetzt, daher ist $W_C(x_p) = 0$ der Nullpunkt der Energieskala in Abb. 3.6). In der physikalischen Vorstellung kann Gl. (3.38) als Boltzmann-Gleichung interpretiert werden. Darin wäre N_D die effektive Zustandsdichte und $W_C(x_n) - W_C(x_p)$ die zu überwindende Energie zwischen Fermi-Energie und Bandkante. Die Minoritätsträgerdichte hängt demnach von der Energiedifferenz der Bandkanten auf den beiden Seiten der RLZ ab. Diese beträgt im thermodynamischen Gleichgewicht $e\,U_D$. Für Elektronen an der Stelle x_n bedeutet das, dass ihnen eine Energie $e\,U_D$ zugeführt werden muss, damit sie die Potentialbarriere überwinden und auf die andere Seite der RLZ gelangen können. Die Potentialbarriere besteht ursächlich in dem elektrischen Feld E über der RLZ, gegen dessen Kraftwirkung die Elektronen unter Aufwendung der Energie $e\,U_D$ in das p-Gebiet gebracht werden müssen. Anders ist es für Elektronen auf der anderen Seite der RLZ im p-Gebiet. Gelangen sie in das elektrische Feld, so werden sie durch das elektrische Feld automatisch auf die n-Seite befördert. Dies entspricht auch der Vorstellung von Elektronen als Murmeln, die die Barriere hinunterrollen. Legen wir eine äußere Spannung an, so erzeugt diese zusätzliche Spannung über der RLZ ein zusätzliches Potentialfeld (Energie). Wir nehmen an, dass sich die beiden Potentialfelder von Diffusionsspannung und äußerer Spannung bezüglich ihrer Wirkung auf die Ladungsträger überlagern. Dann ist die Energie der Potentialbarriere

$$W_C(x_n) \,-\, W_C(x_p) \,=\, -\,e\,U_{RLZ} \,=\, -\,e(U_D - U) \tag{3.39}$$

Dabei wird U gemäß der Definition in Abb. 3.1 positiv in Richtung vom p- in das n-Gebiet (von der Anode zur Kathode) gezählt.
Bei einer positiven Spannung in dieser Richtung gelangen mehr Ladungsträger auf die andere Seite der RLZ. Die Minoritätsträgerdichte dort steigt. Wir erwarten daher einen höheren Stromfluss und bezeichnen die angelegte Spannung als Flussspannung. Bei negativem Vorzeichen von U wird die Potentialbarriere höher und es gelangen weniger Ladungsträger auf die andere Seite. Die Minoritätsträgerdichte sinkt und wir sprechen von U als Sperrspannung.

Wegen der Überlagerung der Spannungen kann die Wirkung einer äußeren Spannung z. B. auf die Raumladungsweite einfach dadurch berücksichtigt werden, dass in Gl. (3.25) bis (3.28) U_D ausgetauscht wird durch

$$U_{RLZ} \,=\, U_D \,-\, U \tag{3.40}$$

Für die Ladungsträgerdichten außerhalb des Gleichgewichts (Überschußdichten) an den Rändern der RLZ erhalten wir durch diesen Austausch aus Gl. (3.34) – (3.36)

$$n_n(x_n) = n_{n0}(x_n) = N_D \,, \qquad p_n(x_n) = p_{p0}(x_p)\, e^{-\frac{U_D - U}{U_T}} \,, \tag{3.41}$$

$$n_p(x_p) = n_{n0}(x_n)\, e^{-\frac{U_D - U}{U_T}} \,, \qquad p_p(x_p) = p_{p0}(x_p) = N_A \,. \tag{3.42}$$

Darin lassen sich die Exponentialterme durch Einsetzen der Diffusionsspannung aus Gl. (3.23) noch weiter vereinfachen. Wir erhalten damit die Boltzmann-Randbedingungen

$$n_n(x_n) = n_{n0} = N_D \,, \qquad p_n(x_n) = p_{n0}\, e^{\frac{U}{U_T}} \,, \tag{3.43}$$

$$n_p(x_p) = n_{p0}\, e^{\frac{U}{U_T}} \,, \qquad p_p(x_p) = p_{p0} = N_A \,. \tag{3.44}$$

mit $n_{p0} = \frac{n_i^2}{N_A}$ und $p_{n0} = \frac{n_i^2}{N_D}$ nach Gl. (3.6).
Wir haben darin zur Vereinfachung der Schreibweise die Ortsabhängigkeit der Gleichgewichtsdichten aufgrund der Konvention im nächsten Kapitel weggelassen.

3.13 Konvention zur Bezeichnung der Gleichgewichtsdichten

Wir haben bisher für Gleichgewichtskonzentration der Ladungsträger an den Rändern der RLZ an den Stellen

$$x_n : \qquad n_{n0}(x_n) \qquad p_{n0}(x_n) \tag{3.45}$$

$$x_p : \qquad n_{p0}(x_p) \qquad p_{p0}(x_p) \tag{3.46}$$

geschrieben. Diese Ladungsträgerdichten sind in dem Rechteck-Profil-Modell in den gesamten Bahngebieten konstant. Sie ändern sich nur in der RLZ, so dass dort die Angabe der Ortsabhängigkeit von x durchaus berechtigt ist.

Im Folgenden werden wir im überwiegenden Maße nur die Ladungsträgerdichten am RLZ-Rand und in den Bahngebieten benötigen. Um Schreibarbeit zu sparen und die Übersichtlichkeit zu erhöhen, werden wir

im Weiteren auf die Angabe der Ortsabhängigkeit verzichten, wenn wir die Dichten am RLZ-Rand bzw. deren im anschließenden Bahngebiet konstanten Wert verwenden. Beachten: die Konstanz in den Bahngebieten gilt nur im thermodynamischen Gleichgewicht. Daher bezieht sich die Konvention auch nur auf die Ladungsträgerdichten im thermodynamischen Gleichgewicht (Index 0).

3.14 Strom-Spannungskennlinie

Die Strom-Spannungskennlinie beschreibt den Zusammenhang zwischen Strom und Spannung an den äußeren Klemmen des p-n-Übergangs, also an den Anschlüssen der p-n-Diode. Formal lässt sich die Kennlinie durch numerische Lösung des DDM's bestimmen. Wir bevorzugen eine analytische Lösung, da anhand des Aufbaus der Formeln Rückschlüsse auf die Funktionsweise der Diode möglich sind. Das so gewonnene Verständnis lässt sich dann auch auf den Bipolar-Transistor übertragen.
Für eine analytische Lösung sind Vereinfachungen des DDM's basierend auf Näherungen und begründbaren Annahmen nötig. Aus der Transportgleichung des Modells geht hervor, dass der Gesamtstrom durch den Übergang aus Drift- und Diffusionsstrom von Elektronen und Löchern bestehen kann. Wir betrachten den Gesamtstrom mit seinen Komponenten in verschiedenen Abschnitten des p-n-Übergangs genauer und haben dabei das Ziel, Vereinfachungen herbeizuführen.

3.14.1 Gesamtstrom

Der Gesamtstrom durch den p-n-Übergang wird durch die stationäre eindimensionale Kontinuitätsgleichung (vgl. Gl. (2.124)) beschrieben. Sie ergibt sich durch Überlagerung (Addition) der beiden Kontinuitätsgleichungen für Elektronen und Löcher des DDM's zu

$$\frac{dI}{dx} = \frac{dI_p}{dx} + \frac{dI_n}{dx} = 0 \tag{3.47}$$

Sie bedeutet, dass der Gesamtstrom an jeder Stelle des Halbleiters konstant ist. Es geht kein Strom verloren. Dies entspricht der verallgemeinerten Kirchhoffschen Knotenregel, bei der die Summe aller Ströme, die in einen

Körper (hier ein Halbleiter) hineinfließen, gleich Null ist. Ein Strom, der auf der einen Seite in den Halbleiter (in die Diode) fließt, kommt auf der anderen Seite heraus.
Dies führt zur einfachen, aber im Folgenden wichtigen Erkenntnis, dass die Summe aus Drift- und Diffusionsstrom von Elektronen und Löchern an jedem Ort konstant und gleich dem Gesamtstrom ist.
Wichtig ist jedoch zu bemerken, dass sich der Löcher- oder Elektronen-Strom über dem Ort ändern kann. Der Strom der jeweils anderen Ladungsträgerart stellt sich dann nach Gl. (3.47) so ein, dass der Summenstrom über dem Ort konstant bleibt.

3.14.2 Ströme in der Raumladungszone

Aufgrund der eingangs getroffenen Annahme, dass die Bahngebiete niederohmig (nahezu feldfrei) sind, liegt eine von außen angelegte Spannung direkt an der RLZ an. Daraus folgt direkt, dass der Strom durch den p-n-Übergang durch die Spannung über der RLZ bestimmt wird.
Durch die, über der RLZ angelegte Spannung weichen die Ladungsträgerdichten in der RLZ von ihren Gleichgewichtswerten ab. Am Rand der RLZ stellen sich dadurch die Randkonzentrationen nach Gl. (3.42) und (3.44) ein. Bei Flusspolung der Spannung ($U>0$) wird die Potentialbarriere und damit die Feldstärke über der RLZ verringert. Im thermodynamischen Gleichgewicht hatten sich Drift- und Diffusionsstrom in der RLZ gerade kompensiert. Durch die jetzt verringerte Feldstärke nimmt der, dem Diffusionstrom entgegenwirkende Driftstrom ab. In Folge kommen durch den nun nicht mehr kompensierten Diffusionsstrom mehr Löcher aus dem p-Gebiet und Elektronen aus dem n-Gebiet (beides dort Majoritäten) in die RLZ. Die Ladungsträgerdichten in der RLZ werden also durch eine Flussspannung angehoben. Auf der jeweils anderen Seite der RLZ sind die Ladungsträger Minoritäten. Deren Erhöhung haben wir bereits als Boltzmann-Randbedingung in Gl. (3.44) berechnet. Da aufgrund der Neutralitätsbedingung bei Gleichgewichtsstörungen $\Delta n(x)=\Delta p(x)$ gilt, wird bei Flusspolung die Ladungsträgerdichte in der RLZ erhöht und es gilt $n(x)\,p(x) > n_i^2$.
Bei Sperrpolung ($U<0$) kommt es entsprechend zu einer Absenkung der Ladungsträgerkonzentration gegenüber der Gleichgewichtskonzentration in

der RLZ. Die Minoritätendichten können ebenfalls mit Gl. (3.44) berechnet werden. Es gilt allgemein bei Sperrpolung in der RLZ $n(x)\,p(x) < n_i^2$. Abb. 3.7 zeigt die Verhältnisse in der RLZ für die beiden Fälle im Vergleich zu den Gleichgewichtsdichten.

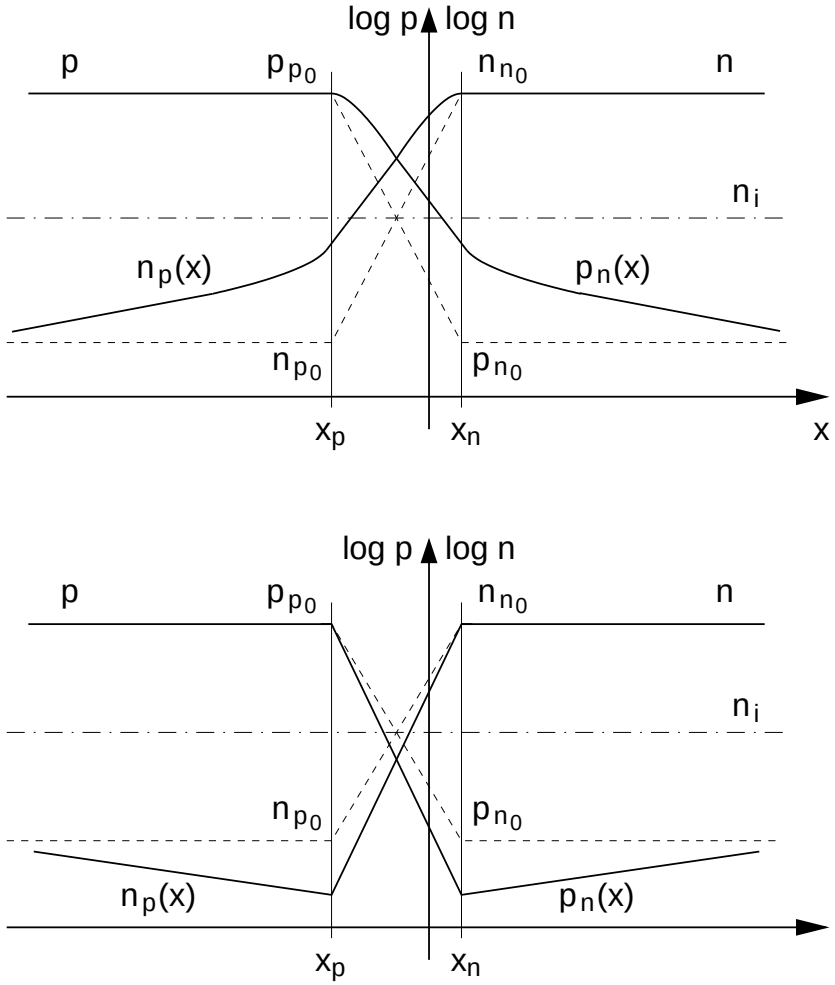

Abb. 3.7: Ladungsträgerdichten in der RLZ und den angrenzenden Bahngebieten des p-n Übergangs. Oben: Bei Flusspolung ($U > 0$). Unten: Bei Sperrpolung ($U < 0$).

Wir fassen aufgrund der Wichtigkeit zusammen:
Eine von außen an den p-n-Übergang angelegte Spannung fällt aufgrund der Niederohmigkeit der Bahngebiete über der RLZ ab. Dort steuert sie die Ladungsträgerdichten in der RLZ. Die Ladungsträgerdichten an

den RLZ-Rändern ergeben sich über die Boltzmann-Randbedingungen. Danach werden die Minoritätsträgerdichten exponentiell über die Spannung gesteuert. Aufgrund der Abweichung der Ladungsträgerdichten von der Gleichgewichtsdichte wird eine Nettorekombination einsetzen, um das Gleichgewicht wieder herzustellen. Hieraus resultiert in der RLZ ein Nettorekombinationsstrom I_{rg}, der in einem der nachfolgenden Kapitel berechnet wird.

3.14.3 Ströme in den Bahngebieten

An den Rändern der RLZ zu den Bahngebieten weichen die Minoritätsträgerdichten entsprechend den Boltzmann-Randbedingungen um Δp_n bzw. Δn_p von ihren Gleichgewichtsdichten p_{n0} bzw. n_{p0} ab. Dieser Konzentrationsunterschied verursacht jeweils einen Diffusionsstrom der Minoritätsträger auf beiden Seiten der RLZ in die jeweiligen Bahngebiete. Da diese Randkonzentrationen durch die, über der RLZ angelegte Spannung verursacht wird, spricht man auch von einer Injektion von Ladungsträgern aus der RLZ in die Bahngebiete.
Entsprechend den Überlegungen in Kap. 2.23 führt die erhöhte Minoritätsträgerkonzentration zu einer entsprechenden Erhöhung der Majoritätsträgerkonzentration um $\Delta p_p{=}\Delta n_p$ im p-Gebiet bzw. $\Delta n_n{=}\Delta p_n$ im n-Gebiet, wodurch sich Ladungsneutralität einstellt. Dies geschieht durch eine leichte Verschiebung der Majoritätsträger, die gegenüber der Gleichgewichtskonzentration p_{p0} bzw. n_{n0} vernachlässigbar (Majoritäten) ist, solange die im Folgenden vorausgesetzte niedrige Injektion vorliegt.
Durch die Abweichung Δn_p, Δp_p im p- und Δp_n, Δn_n im n-Gebiet setzt eine Nettorekombination mit dem Ziel des Abbaus der Abweichung ein. Je weiter ein Ort in den Bahngebieten von den Rändern der RLZ entfernt ist, umso geringer wird die Abweichung sein. Bei einer sog. „langen Diode“ klingt die Konzentration der injizierten Ladungsträger (Minoritäten) innerhalb der Bahngebiete noch vor Erreichen der Kontakte auf den Gleichgewichtswert ab. Dann ist kein Konzentrationsgefälle mehr vorhanden und der Diffusionsstrom wird zu Null.
Da in den Bahngebieten der Driftstrom der Minoritätsträger aufgrund der als gering ($E \approx 0$) angenommenen Feldstärke vernachlässigbar ist, muss der

Strom in der langen Diode, weit entfernt von der RLZ, ein Driftstrom der Majoritätsträger sein. Dies ist möglich, da die Feldstärke gering aber nicht Null ist, sie aber die hohe Dichte der Majoritätsträger antreibt.
Der Übergang zwischen dem Diffusionsstrom der injizierten Minoritätsträger an den RLZ-Rändern in den Driftstrom der Majoritätsträger muss gemäß der Kontinuitätsgleichung (3.47) so erfolgen, dass der Gesamtstrom über das Bahngebiet konstant bleibt. Wie wir noch sehen werden, erfolgt dies durch eine Nettorekombination in den Bahngebieten der langen Diode, an der die Minoritätsträger aus der RLZ und die Majoritätsträger aus den kontaktnahen Bahngebieten beteiligt sind.
Neben der „langen“ Diode gibt es den Sonderfall der „kurzen“ Diode. Hier ist das Bahngebiet so kurz, dass die Minoritätsträgerkonzentration an den Kontakten auf den Wert der Randbedingung, den die Kontakte vorgeben, gezwungen wird. In der Regel kann wegen der Anwesenheit eines Metalls als Kontakt eine unendlich hohe Rekombinationsgeschwindigkeit angenommen werden, wodurch sich am Ort des Kontaktes als Randbedingung ebenfalls die Gleichgewichtsdichte einstellt. Wir werden im Folgenden mit dieser Randbedingung arbeiten.

3.15 Ortsabhängigkeit der Ströme am p-n-Übergang

Wir fassen die vorangegangenen qualitativen Überlegungen in Abb. 3.8 und 3.9 zusammen. Die Wirkung der von außen angelegten Spannung U auf die Ströme in der Diode lässt sich wie folgt darstellen:

1. Durch die niederohmigen Bahngebiete fällt U über der RLZ ab.
2. Durch die Spannung über der RLZ weichen die Ladungsträgerdichten in der RLZ von ihren Gleichgewichtswerten ab.
3. Dadurch entsteht in der RLZ
   (a) der Nettorekombinationsstrom I_{rg} und
   (b) eine ortsabhängige Ladungsträgerdichte mit den Werten $n_n(x_n)$, $p_n(x_n)$ sowie $p_p(x_p)$, $n_p(x_p)$ an den Rändern x_n und x_p der RLZ.
4. Für die Minoritätsträger entsteht aufgrund des Konzentrationsgefälles in Richtung der Bahngebiete ein Diffusionsstrom ausgehend von den Rändern der RLZ.

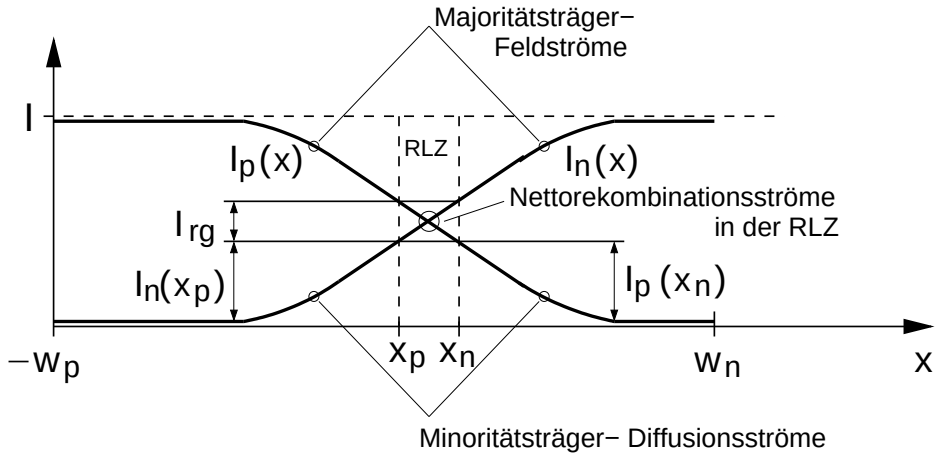

Abb. 3.8: Aufteilung des Gesamtstroms I durch den p-n Übergang in die ortsabhängigen Elektronen- und Löcherströme $I_n(x)$ und $I_p(x)$. Die Summe der beiden Ströme ist über dem Ort konstant. Dargestellt ist der Sonderfall gleicher Minoritätsträger-Diffusionsströme in beiden Bahngebieten.

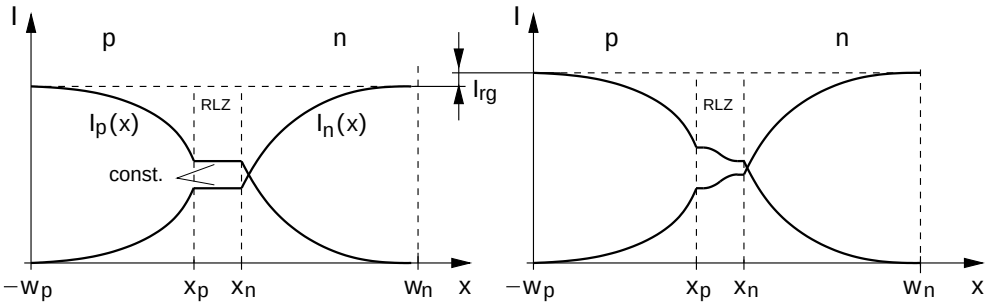

Abb. 3.9: Wie Abb. 3.8, jedoch mit unterschiedlichen Diffusionsströmen in den Bahngebieten. **Links:** Fall ohne Nettorekombination (R=0) in der RLZ. **Rechts:** Mit Nettorekombination in der RLZ.

5. Der Driftstrom der Minoritätsträger ist aufgrund der geringen Dichte der Minoritätsträger und $E \approx 0$ in den Bahngebieten vernachlässigbar.

6. Durch Rekombination im Verlauf der Bahngebiete in Richtung der Kontakte nimmt die Minoritätsträgerkonzentration und damit auch der Diffusionsstrom ab (Dies gilt für die in diesem Beispiel diskutierte „lange Diode“. Bei der „kurzen Diode“ erfolgt die Rekombination nicht in den Bahngebieten sondern nur am Kontakt).

7. Zur Rekombination der Minoritätsträger sind Majoritätsträger erfor-

derlich. Deren Strom nimmt in Richtung zu den Kontakten zu, um die zur Rekombination erforderlichen Ladungsträger zu leiten.

8. Aufgrund der hohen Konzentration der Majoritätsträger ist deren Diffusionsstrom bei niedriger Injektion vernachlässigbar klein ($p_p \approx N_A$, $n_n \approx N_D$) gegenüber deren Driftstrom der durch die Feldstärke in den Bahngebieten ($E \approx 0$) angetrieben wird.

9. Bei genügend langen Bahngebieten sind die, an den RLZ-Rändern injizierten Minoritätsträgerkonzentrationen bis auf die Gleichgewichtswerte abgeklungen. Der Minoritätsträgerstrom ist in diesem Bereich vernachlässigbar.

10. Feldstrom der Majoritäten und Diffusionsstrom der Minoritäten ergänzen sich in den Bahngebieten zum Gesamtstrom, der über die gesamte Länge des p-n-Übergangs konstant ist.

Um den Gesamtstrom durch den Übergang zu berechnen genügt es wegen 10., ihn an einer geeigneten Stelle zu bestimmen. Wir wählen dazu mit $x = x_n$ einen der beiden Ränder der RLZ. Hier gilt für den Gesamtstrom z.B. nach Abb. 3.8 oder 3.9:

$$I = I_n(x_n) + I_p(x_n) = I_n(x_p) + I_{rg} + I_p(x_n) \; . \tag{3.48}$$

Der Gesamtstrom setzt sich demnach zusammen aus den beiden Diffusionsströmen der jeweiligen Minoritätsträger an den beiden Rändern der RLZ und dem Nettorekombinationsstrom in der RLZ.
Die Bestimmung dieser drei Ströme wird Aufgabe der nächsten beiden Kapitel sein.

3.16 Nettorekombinationsstrom in der Raumladungszone

Wir wissen aus den vorangegangenen Überlegungen, dass durch eine äußere Spannung die Ladungsträgerdichten in der RLZ von ihren Gleichgewichtsdichten abweichen. Aus den Überlegungen zur Generation und Rekombination von Ladungsträgern (vgl. z. B. Kap. 2.21) wissen wir, dass sich bei einer Abweichung der Ladungsträgerdichten von den Gleichgewichtsdichten eine Nettorekombinationsrate $R = r - g$ ungleich Null einstellt. Ziel

der Nettorekombination ist es, durch einen höheren Ladungsträger-Abbau als Generation ($R > 0$) bzw. mehr Generation als Abbau ($R < 0$) wieder die Gleichgewichtsdichten herzustellen. Genau dieser Vorgang läuft auch in der RLZ ab. Abb. 3.10 zeigt die beiden Vorgänge schematisch im Bänderdiagramm.

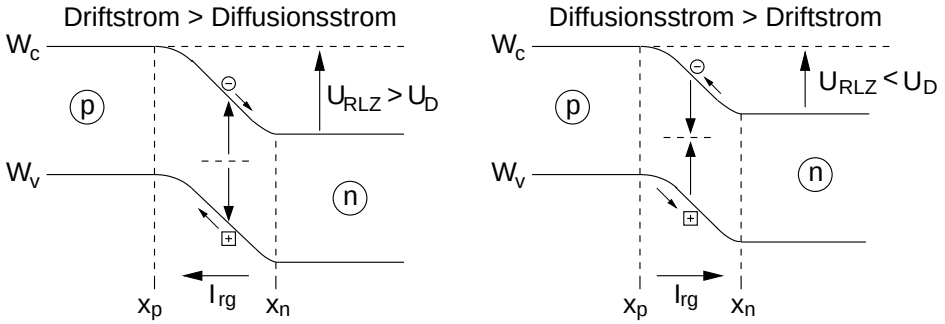

Abb. 3.10: Links: Generation ($R < 0$) von Ladungsträgern in der RLZ zur Erhöhung der Ladungsträgerdichte bei Sperrspannungen ($U < 0$ bzw. $U_{RLZ} > U_D$). Rechts: Rekombination ($R > 0$) von Ladungsträgern in der RLZ zur Verringerung der Ladungsträgerdichte bei Flusspolung ($U > 0$ bzw. $U_{RLZ} < U_D$).

Bei Sperrpolung überwiegt der Drift- den Diffusionsstrom. Durch den, gegenüber dem Gleichgewicht erhöhten Driftstrom (genauer: durch das Feld) verarmt die RLZ an Ladungsträgern. Daher wird $R < 0$, wodurch Ladungsträger in der RLZ generiert werden (Abb. 3.10 links). Aufgrund der Richtung des vorherrschenden elektrischen Feldes ergibt sich ein Driftstrom I_{rg} der generierten Ladungsträger von n- in Richtung p-Gebiet.

Bei Flusspolung überwiegt der Diffusionsstrom und R ist > 0. Durch den höheren Diffusionsstrom gelangen Elektronen von x_n und Löcher von x_p in die RLZ (vgl. Pfeile an den Ladungsträgern in Abb. 3.10) und rekombinieren dort. Der von x_p in die RLZ fließende reine Löcherstrom setzt sich daher als ein bei x_n aus der RLZ austretender reiner Elektronenstrom fort (beachten: Elektronen fließen entgegen der Stromrichtung). Abb. 3.11 verdeutlicht diesen Vorgang.
Der Gesamtstrom I_{rg} ist entsprechend den Überlegungen zu Kap. 3.14.1 über dem Ort konstant.
I_{rg} entsteht aufgrund der Nettorekombinationsrate $R \neq 0$ in der RLZ.

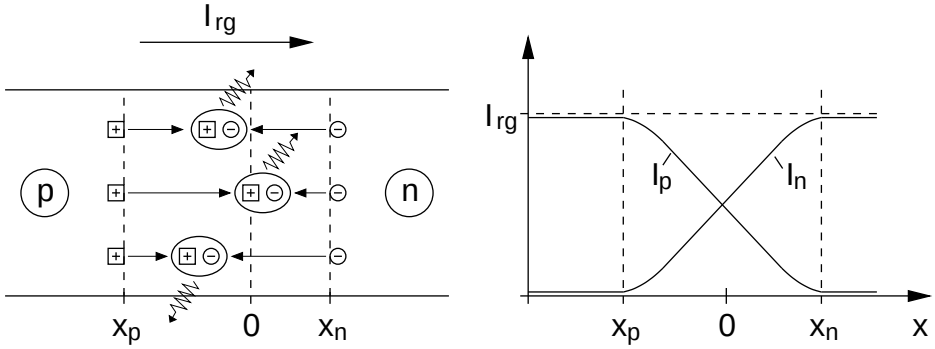

Abb. 3.11: Rekombination ($R > 0$) in der Sperrschicht aufgrund der Flusspolung des p-n-Übergangs. Links: Schematische Darstellung des Rekombinationsvorgangs. Rechts: Übergang von Löcherstrom bei x_p in einen Elektronenstrom bei x_n. Die Summe der Ströme, der Nettorekombinationsstrom, ist konstant $I_{rg} = I_p + I_n$.

Die beiden darin enthaltenen Elektronen- und Löcherströme erfüllen die Kontinuitätsgleichungen des DDMs für Elektronen und Löcher Gl. (2.250), (2.251) im stationären Fall

$$\frac{dJ_n}{dx} = \frac{1}{A}\frac{dI_n}{dx} = e\,R\,, \tag{3.49}$$

$$\frac{dJ_p}{dx} = \frac{1}{A}\frac{dI_p}{dx} = -e\,R\,. \tag{3.50}$$

Die Addition beider Gleichungen verifiziert nochmals die bereits erhaltene Aussage, dass der Summenstrom (Nettorekombinationsstrom I_{rg}) konstant über x ist. Da aufgrund des großen Konzentrationsgefälles an den Rändern der RLZ gilt

$$I_p(x_p) = I_{rg}\,,\quad I_p(x_n) = 0\,, \tag{3.51}$$

$$I_n(x_n) = I_{rg}\,,\quad I_n(x_p) = 0\,, \tag{3.52}$$

genügt eine der beiden Kontinuitätsgleichungen, um I_{rg} zu berechnen. Wir wählen willkürlich Gl. (3.50) für den Löcherstrom. Durch Integration über die RLZ erhalten wir mit dem Strom an den Integrationsgrenzen mit den Randbedingungen nach Gl. (3.51)

$$\int_{x_p}^{x_n} \frac{dI_p}{dx}\,dx = I_p(x_n) - I_p(x_p) = -e\,A\int_{x_p}^{x_n} R(x)\,dx \tag{3.53}$$

$$I_{rg} = e\,A\int_{x_p}^{x_n} R(x)\,dx\,. \tag{3.54}$$

Für $R(x)$ müssen wir die in Kap. 2.15 hergeleitete Beziehung für den vorherrschenden Rekombinationsmechanismus einsetzen. Für den wichtigsten Fall der Rekombination durch Phononenwechselwirkung (SRH) erhalten wir mit Gl. (2.171)

$$I_{rg} = e\,A \int_{x_p}^{x_n} \frac{p(x)\,n(x) - n_i^2}{(p_1 + p(x))\,\tau_n + (n_1 + n(x))\,\tau_p}\,dx \tag{3.55}$$

Die ortsabhängigen Ladungsträgerdichten haben wir für das thermodynamische Gleichgewicht bereits in Gl. (3.31), (3.32) berechnet. Setzen wir die Gültigkeit unserer Überlegungen zu Gl. (3.40) voraus, können wir die Ladungsträgerdichten bei äußerer Spannung U durch Austausch von U_D gegen $U_{RLZ} = U_D - U$ in Gl. (3.31) und (3.32) erhalten. Dazu muss der Austausch auch in der darin enthaltenen Bestimmungsgleichung (3.14) für $\varphi(x)$ erfolgen.
Das sich daraus ergebende Integral (Gl. (3.55)) ist geschlossen nicht mehr lösbar. Es gibt jedoch zahlreiche Näherungslösungen, die auf Fallunterscheidungen zwischen hoher Ladungsträgerkonzentration (Flusspolung) und niedriger Ladungsträgerkonzentration (Sperrpolung) beruhen.

Beispiel: (Übung)
Eine einfache Lösung ergibt sich für Sperrpolung, wenn $n(x) \ll n_1$ und $p(x) \ll p_1$ gilt. Das Produkt $p(x)\,n(x)$ kann in der Form des verallgemeinerten Massenwirkungsgesetzes

$$n(x)\,p(x) = n_i^2\,e^{\frac{U}{U_T}} \tag{3.56}$$

geschrieben werden (zur Übung zeigen).
Damit lässt sich zeigen (zur Übung), dass der Nettorekombinationsstrom bei Sperrpolung

$$I_{rg}(U \ll 0) = -I_{rg,S} = -\frac{e\,A\,n_i}{\tau_{eff}}\,w_{RLZ}(U) \tag{3.57}$$

$$\text{mit } \tau_{eff} := \frac{n_1\,\tau_p + p_1\,\tau_n}{n_i} \tag{3.58}$$

ist. Er steigt also proportional mit der Diodenfläche A und der Weite der RLZ an. Durch die quadratische Abhängigkeit von n_i weist der Nettorekombinationsstrom auch eine Temperaturabhängigkeit auf.
Die Spannungsabhängigkeit des Stroms wird über die Spannungsabhängigkeit der Raumladungsweite gesteuert.

Für Flusspolung ergibt sich nach einer Näherungsrechnung (vgl. Anhang A) ein Ausdruck, der mit dem Nettorekombinationsstrom für Sperrpolung nach Gl. (3.57) zusammengefasst werden kann zu

$$I_{rg} = I_{rg,S}\left(e^{\frac{U}{2U_T}} - 1\right) \tag{3.59}$$

$$\text{mit } I_{rg,S} = \frac{e\,A\,n_i}{\tau_{eff}}\,w_{RLZ}(U) \text{ nach Gl. (3.57)}\ .$$

Der Nettorekombinationsstrom besitzt demnach bei Flusspolung eine starke (exponentielle) Spannungsabhängigkeit.

3.17 Minoritätsträgerströme an den Rändern der Raumladungszone

3.17.1 Differentialgleichung der Minoritätsträgerverteilung in den Bahngebieten

Der Strom durch den p-n-Übergang setzt sich nach Gl. (3.48) aus dem im vorangegangenen Kapitel berechneten Nettorekombinationsstrom I_{rg} und den im Folgenden berechneten Diffusionsströmen der Minoritätsträger an den beiden Rändern der RLZ zusammen.
Zur Berechnung verwenden wir die Transportgleichungen (2.248) und (2.249) des DDM's. Bei vernachlässigbarem Driftstrom der Minoritätsträger (wegen $E \approx 0$ in den Bahngebieten) vereinfachen sich diese für einen Querschnitt A durch den der Strom fließt zu

$$I_n(x) = e\,A\,D_n\,\frac{d\,n_p(x)}{dx} \qquad \text{(Elektronenstrom in } B_p\text{)}, \tag{3.60}$$

$$I_p(x) = -e\,A\,D_p\,\frac{d\,p_n(x)}{dx} \qquad \text{(Löcherstrom in } B_n\text{)}. \tag{3.61}$$

Die Änderung der Ströme über den Ort wird durch die Kontinuitätsgleichungen des DMM's (2.250), (2.251) bestimmt

$$\frac{dI_n(x)}{dx} = e\,A\,R(x) = e\,A\,\frac{n_p(x) - n_{p_0}}{\tau_n} \quad \text{, in } B_p \tag{3.62}$$

$$\frac{dI_p(x)}{dx} = -e\,A\,R(x) = -e\,A\,\frac{p_n(x) - p_{n_0}}{\tau_p} \quad \text{, in } B_n \tag{3.63}$$

Für die Gleichgewichtsdichten darin wird aufgrund der Rechteckprofil-Näherung keine Ortsabhängigkeit angenommen. Die Nettorekombinationsraten $R(x)$ (unterschiedlich für die beiden Ströme in B_n und B_p) werden darin durch das einfache Rekombinationsmodell für schwache Abweichungen von der Gleichgewichtskonzentration beschrieben. Die Diffusionsströme nach Gl. (3.60), (3.61) müssen der Kontinuitätsgleichung gehorchen. D.h. deren Ableitung muss Gl. (3.62), (3.63) erfüllen. Daraus folgt durch Gleichsetzen

$$e\,A\,D_n\,\frac{d^2 n_p(x)}{dx^2} = e\,A\,\frac{n_p(x) - n_{p_0}}{\tau_n} \quad \text{, in } B_p \tag{3.64}$$

$$-e\,A\,D_p\,\frac{d^2 p_n(x)}{dx^2} = -e\,A\,\frac{p_n(x) - p_{n_0}}{\tau_p} \quad \text{in } B_n. \tag{3.65}$$

Mit der Definition der Diffusionslängen für Elektronen und Löcher

$$L_n := \sqrt{\tau_n \, D_n} \quad , \tag{3.66}$$

$$L_p := \sqrt{\tau_p \, D_p} \quad , \tag{3.67}$$

deren Bedeutung uns anhand der nachfolgenden Lösung klar wird, ergeben sich aus Gl. (3.64), (3.65) die Differentialgleichungen

$$\frac{d^2 n_p(x)}{dx^2} = \frac{n_p(x) - n_{p_0}}{L_n^2} \quad , \text{in } B_p \tag{3.68}$$

$$\frac{d^2 p_n(x)}{dx^2} = \frac{p_n(x) - p_{n_0}}{L_p^2} \quad , \text{in } B_n. \tag{3.69}$$

Mit ihnen lässt sich die Verteilung der Minoritätsträger in den Bahngebieten berechnen. Dazu benötigen wir noch die im nächsten Kapitel definierten Randbedingungen.

3.18 Annahmen und Randbedingungen

Die Minoritätsträger werden an den Rändern der RLZ in die Bahngebiete in Abhängigkeit der von außen angelegten Spannung injiziert. Die Minoritätsträgerdichte an den Rändern hatten wir bereits in Form der Boltzmann-Randbedingung in Gl. (3.43), (3.44) berechnet

$$n_p(x_p) = n_{p_0} \, e^{\frac{U}{U_T}} \tag{3.70}$$

$$p_n(x_n) = p_{n_0} \, e^{\frac{U}{U_T}} \tag{3.71}$$

Darin sind die Gleichgewichtsdichten n_{p_0}, p_{n_0} über die gesamten Bahngebiete wegen der Rechteckprofil-Näherung konstant.

Auf der anderen Seite der Bahngebiete, an den Kontakten, nehmen wir eine idealisierte, unendlich hohe Rekombinationsgeschwindigkeit an. Dies ist gerechtfertigt, da die Rekombinationsgeschwindigkeit durch die Anwesenheit eines Kontaktmaterials stark erhöht ist. Dies wird zusätzlich dadurch unterstützt, dass die Rekombination an Oberflächen (Kontakt) und Grenzflächen in der Regel höher ist.
Durch diese Annahme existiert an den Kontakten keine Abweichung von der Gleichgewichtsdichte, da diese unmittelbar durch Nettorekombination

verschwinden. An den Kontakten muss daher die Minoritätsträgerdichte auf den Wert der Gleichgewichtsdichten abgeklungen sein

$$n_p(-w_p) = n_{p0} \tag{3.72}$$

$$p_n(w_n) = p_{n_0} \tag{3.73}$$

Dies gilt sowohl für die kurze Diode als auch für die lange Diode, bei der sich diese Randbedingung definitionsgemäß ohnehin aufgrund der Rekombination über die Länge der Bahngebiete eingestellt hat. Anhand dieser Randbedingung lässt sich ein prinzipieller Verlauf der Minoritätsträgerverteilung in den Bahngebieten wie z.B. in Abb 3.12 zeichnen.

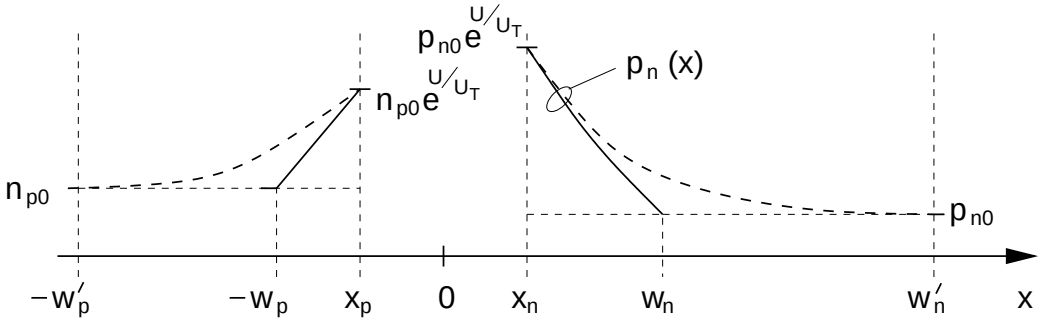

Abb. 3.12: Konzentration der Minoritätsträger in den Bahngebieten des p-n-Übergangs für den Fall einer langen Diode (gestrichelt, Kontakte bei w'_p, w'_n) und einer kurzen Diode (durchgezogen, Kontakte bei w_p, w_n).

3.19 Berechnung der Minoritätsträger-Diffusionsströme

Um die Minoritätsträger-Diffusionsströme zu erhalten sind zuerst die Minoritätsträgerverteilungen in den Bahngebieten zu ermitteln. Dazu sind die Differentialgleichungen zweiter Ordnung aus Gl. (3.68), (3.69)

$$\frac{d^2 n_p(x)}{dx^2} = \frac{n_p(x) - n_{p0}}{L_n^2} \quad \text{, in } B_p \tag{3.74}$$

$$\frac{d^2 p_n(x)}{dx^2} = \frac{p_n(x) - p_{n_0}}{L_p^2} \quad \text{, in } B_n \tag{3.75}$$

für das p- bzw. n-Bahngebiet mit den Randbedingungen aus Gl. (3.70)-(3.73)

$$n_p(x_p) = n_{p0}\, e^{\frac{U}{U_T}} \tag{3.76}$$

$$n_p(w_p) = n_{p0} \tag{3.77}$$

$$p_n(x_n) = p_{n0}\, e^{\frac{U}{U_T}} \tag{3.78}$$

$$p_n(-w_n) = p_{n0} \tag{3.79}$$

zu lösen.
Die allgemeine Lösung der Differentialgleichung ist bekannt und hat z.B. für die Elektronendichte im p-Bahngebiet die Form

$$n_p(x) = A + B\, e^{bx} \tag{3.80}$$

Die Löcherdichte im n-Bahngebiet besitzt die gleiche Form. Anstatt beide Bahngebiete hier parallel zu behandeln wird im Folgenden nur das p-Bahngebiet betrachtet. Die Lösung für das n-Bahngebiet ergibt sich dann einfach durch Substitution der Größen des p-Gebiets durch die des n-Gebiets. Einsetzen der allgemeinen Lösung nach Gl. (3.80) in die Differentialgleichung (3.74) ergibt wegen der beiden Lösungen $b = \pm L_n^{-1}$ die Lösungsfunktion

$$n_p(x) = n_{p0} + B\, e^{\frac{x}{L_n}} + C\, e^{\frac{-x}{L_n}} \quad . \tag{3.81}$$

Hieraus wird die Bedeutung der Diffusionslängen L_n, L_p nach Gl. (3.66) und (3.67) klar, die den Abfall der Minoritätsträgerdichte über dem Ort bestimmen (ähnlich einer Zeitkonstante). Die Konstanten B und C werden über die Randbedingungen (3.76), (3.77) durch Einsetzen in Gl. (3.81) bestimmt. Nach einfacher und kurzer Rechnung (gut zum Üben) ergibt sich die Minoritätsträgerdichte im p-Bahngebiet

$$n_p(x) = n_{p0} + n_{p0}(e^{\frac{U}{U_T}} - 1)\, \frac{\sinh\left(\frac{w_p+x}{L_n}\right)}{\sinh\left(\frac{w_p+x_p}{L_n}\right)} \tag{3.82}$$

Durch Differentiation ergibt sich über Gl. (3.60) der Elektronenstrom im p-Bahngebiet

$$I_n(x) = e\, A\, D_n\, \frac{dn_p(x)}{dx} = \frac{1}{L_n}\, e\, A\, D_n\, n_{p0}\, (e^{\frac{U}{U_T}} - 1)\, \frac{\cosh\left(\frac{w_p+x}{L_n}\right)}{\sinh\left(\frac{w_p+x_p}{L_n}\right)} \tag{3.83}$$

und an der Stelle $x = x_p$

$$I_n(x_p) = e\,A\,\frac{D_n}{L_n}\,n_{p_0}\,(e^{\frac{U}{U_T}} - 1)\,\coth\left(\frac{w_p + x_p}{L_n}\right) \tag{3.84}$$

Durch Substitution $n_{p_0} \to p_{n_0}$, $-w_p \to w_n$, $x_p \to x_n$, $-e \to e$, $L_n \to L_p$, $D_n \to D_p$ ergibt sich aus Gl. (3.82) die Löcherdichte im n-Bahngebiet

$$p_n(x) = p_{n_0} + p_{n_0}\,(e^{\frac{U}{U_T}} - 1)\,\frac{\sinh\frac{(w_n - x)}{L_p}}{\sinh\left(\frac{w_n - x_n}{L_p}\right)} \quad , \tag{3.85}$$

sowie aus Gl. (3.84) der Löcherstrom im n-Bahngebiet an der Stelle $x = x_n$, wobei zusätzlich das negative Vorzeichen des Löcher-Gradienten aus dem DDM berücksichtigt werden muss

$$I_p(x_n) = e\,A\,\frac{D_p}{L_p}\,p_{n_0}\,(e^{\frac{U}{U_T}} - 1)\,\coth\left(\frac{w_n - x_n}{L_p}\right) \ . \tag{3.86}$$

Zur Abkürzung führen wir die Länge der Bahngebiete

$$d_n := w_n - x_n \tag{3.87}$$

$$d_p := -(-w_p - x_p) \tag{3.88}$$

ein, die immer positive Werte besitzen.

3.20 Gesamtstrom des p-n-Übergangs

Mit Gl. (3.84), (3.86) und den Längen der Bahngebiete d_n, d_p nach Gl. (3.87) und (3.88) lassen sich, unter Berücksichtigung, dass der coth eine ungerade Funktion ist, Löcher- und Elektronenminoritätsträgerstrom zum Gesamtstrom des p-n-Übergangs entsprechend Gl. (3.45) zusammenfassen.

$$\begin{aligned} I &= I_n(x_p) + I_p(x_n) + I_{rg} \\ &= \underbrace{e\,A\,\left(n_{p_0}\,\frac{D_n}{L_n}\,\coth\frac{d_p}{L_n} + p_{n_0}\,\frac{D_p}{L_p}\,\coth\frac{d_n}{L_p}\right)}_{I_S}\,(e^{\frac{U}{U_T}} - 1) + I_{rg} \end{aligned} \tag{3.89}$$

Wir definieren den Sättigungsstrom I_S des p-n-Übergangs

$$I_S := e\,A\,\left(n_{p_0}\frac{D_n}{L_n}\coth\left(\frac{d_p}{L_n}\right) + p_{n_0}\frac{D_p}{L_p}\coth\left(\frac{d_n}{L_p}\right)\right) \tag{3.90}$$

mit dem sich der Gesamtstrom schreiben lässt

$$I = I_S(e^{\frac{U}{U_T}} - 1) + I_{rg} \quad . \tag{3.91}$$

Für den Nettorekombinationsstrom lässt sich die Näherung nach Gl. (3.59) einsetzen

$$I = I_S\,(e^{\frac{U}{U_T}} - 1) + I_{rg,s}\,(e^{\frac{U}{2U_T}} - 1) \tag{3.92}$$

$$\text{mit} \quad I_{rg,s} = \frac{eA}{\tau_{eff}} n_i\, w_{RLZ}(U)\ . \tag{3.93}$$

Aufgrund der unterschiedlichen Exponenten der e-Funkion lassen sich die beiden Summanden nur in Form eines interpolierten Ausdrucks der Form

$$I = I_{s_0}(U)\,(e^{\frac{U}{mU_T}} - 1) \tag{3.94}$$

zusammenfassen. Darin sind $1{<}m(U){<}2$ und $I_{s_0}(U)$ so zu wählen, dass sich eine möglichst gute Anpassung an die Originalfunktion (3.92) ergibt.
Abb. 3.13 zeigt den Verlauf des Gesamtstroms des p-n-Übergangs nach Gl. (3.92) bei Fluss- und Sperrpolung. Zusätzlich ist auch der idealisierte Verlauf ohne Berücksichtigung des Nettorekombinationsstroms in der RLZ eingezeichnet ($I_{rg,s} = 0$). Die Spannung in Flussrichtung ist in die verschiedenen Bereiche (a) bis (d) unterteilt, in denen der Interpolationsfaktor m entsprechend der Bildunterschrift gewählt werden muss, wenn mit der Näherungsgleichung (3.94) gearbeitet wird. Da $I_{rg,s}{\sim}n_i$ aber $I_s{\sim}n_i^2$ ist der Beitrag des Rekombinationsstroms in Gl. (3.92) bei Materialien mit großem n_i (Ge) vernachlässigbar. Für Materialien mit kleinem n_i wie **Si** gilt $I_{rg,S} = (10^2...10^3)I_S$, so dass für $U{<}0$ der Gesamtstrom durch den Nettorekombinationsstrom in Form des zweiten Terms in Gl. (3.92) bestimmt wird. Dieser Term ist auch noch im Bereich $0{<}U{<}7\,U_T$ dominant.
Für $U{>}10\,U_T$ ist die der vorangegangenen Rechnung zugrunde liegenden Annahme der geringen Abweichung der Ladungsträgerdichten von den Gleichgewichtswerten nicht mehr erfüllt. Die Minoritätsträgerdichte in den Bahngebieten ist dann nicht mehr vernachlässigbar gegenüber der Dichte der Majoritätsträger, wodurch die verwendeten Näherungen $n_n = N_D$, $p_p = N_A$ nicht mehr gelten. Eine analytische Berechnung der Kennlinie ist dann nicht mehr möglich.
Für den Gesamtstrom bei großen Abweichungen von den Gleichgewichtswerten gilt dann (ohne Beweis) die Proportionalität

$$I \sim e^{\frac{U}{2U_T}}, \tag{3.95}$$

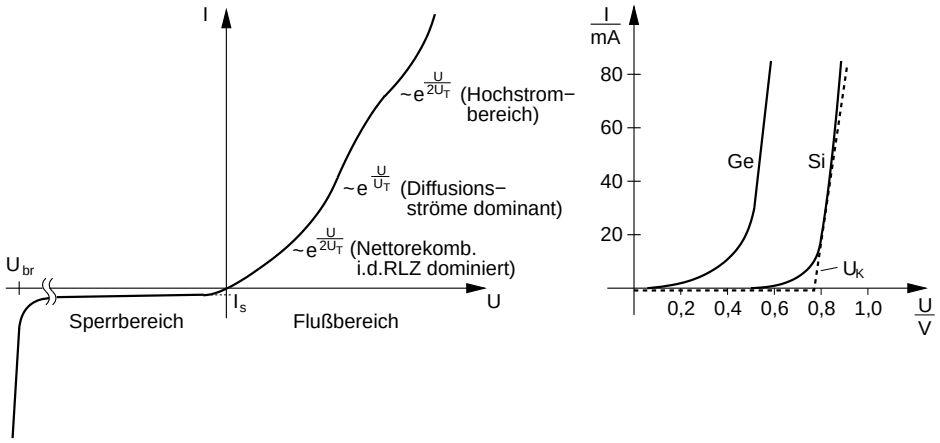

Abb. 3.13: Gleichstromkennlinien von p-n-Übergängen. **Links**: Kennlinie, wie sie sich aus den Berechnungen aus dem Skript ergibt. Tatsächlich ist I_S im Sperrbereich nicht konstant, sondern wächst aufgrund der Spannungsabhängigkeit der RLZ-Weite in Richtung negativer Ströme an. Der Hochstrombereich geht aufgrund von Spannungsabfällen in den Bahngebieten in einen $\sim U^2$ Verlauf über. **Rechts**: Kennlinien realer **Si**- und **Ge**-Dioden. Für den praktischen Gebrauch kann der exponentielle Verlauf näherungsweise mit Geradenabschnitten und einer Knickspannung $U_k (< U_D)$ angenähert werden.

so dass Gl. (3.94) mit $m = 2$ angenommen werden kann. Zusätzlich ist die Wirkung der Bahnwiderstände zu berücksichtigen durch die bei hohen Strömen ein Spannungsabfall in den Bahngebieten entsteht.

Aufgrund der Ähnlichkeiten der Kennlinienverläufe wird in den meisten einfachen Dioden- und Transistormodellen (z.B. Gummel-Poon), die auch in Schaltungs-Simulationsprogrammen wie z.B. SPICE eingesetzt werden, ein idealisierter Kennlinienverlauf entsprechend Gl. (3.94) angenommen. In der in SPICE verwendeten Nomenklatur lautet Gl. (3.94)

$$I = I_S \, (e^{\frac{V}{nV_T}} - 1) \tag{3.96}$$

I_S und n werden darin so gewählt, dass sie die bestmögliche Anpassung an die realen Kennlinienverläufe erzielen. Diese Anpassung der Parameter wird auch als „fitting“ bezeichnet. Beachten: I_S in Gl. (3.96) ist in der Regel nicht identisch mit dem in Gl. (3.90) definierten Sättigungsstrom sondern besitzt nur aufgrund der im SPICE Gummel-Poon Modell verwendeten

Nomenklatur die gleiche Bezeichnung. n liegt dabei in der Regel zwischen 1...2. Zur Vereinfachung von Überlegungen und Rechnungen werden wir in Zukunft meist mit $n = 1$ arbeiten.

3.21 Näherung für kurze und lange Diode

Für den Sättigungsstrom I_S nach Gl. (3.90) können, je nachdem ob die Diffusionslängen der Minoritätsträger sehr viel größer oder kleiner als die Längen d_n und d_p der Bahngebiete sind, die folgenden Näherungen gemacht werden:

- **Kurze Diode:** $d_p \ll L_p,\ d_n \ll L_n$ (3.97)

  Dann gilt wegen $\coth x \approx \frac{1}{x}$ für $x \ll 1$

$$I_S = e\,A\left(n_{p_0}\frac{D_n}{L_n}\coth\frac{d_p}{L_n} + p_{n_0}\frac{D_p}{L_p}\coth\frac{d_n}{L_p}\right),$$
$$I_S \approx e\,A\left(n_{p_0}\frac{D_n}{d_p} + p_{n_0}\frac{D_p}{d_n}\right) \approx e\,A\left(n_{p_0}\frac{D_n}{w_p} + p_{n_0}\frac{D_p}{w_n}\right). \qquad (3.98)$$

  Meist ist, wie im zweiten Schritt Gl. (3.98) berücksichtigt, die Sperrschichtweite vernachlässigbar gegenüber der Weite der Bahngebiete und es gilt dann wegen $d_p \approx w_p$, $d_n \approx w_n$ und mit $n_{p_0} = \frac{n_i^2}{N_A}$, $\quad p_{n_0} = \frac{n_i^2}{N_D}$.

$$I_S \approx e\,A\,n_i^2\left(\frac{D_n}{w_p N_A} + \frac{D_p}{w_n N_D}\right). \qquad (3.99)$$

- **Lange Diode:** $d_p \gg L_p, \quad d_n \gg L_n\,.$ (3.100)

  Dann gilt wegen $\coth x \approx 1$ für $x \gg 1$

$$I_S \approx e\,A\left(n_{p_0}\frac{D_n}{L_n} + p_{n_0}\frac{D_p}{L_p}\right) \qquad (3.101)$$

  und wegen $L_n^2 = \tau_n D_n$ und $L_p^2 = \tau_p D_p$ aus Gl. (3.66) und (3.67) auch

$$I_S \approx e\,A\left(n_{p_0}\frac{L_n}{\tau_n} + p_{n_0}\frac{L_p}{\tau_p}\right) \qquad (3.102)$$
$$\approx e\,A\,n_i^2\left(\frac{L_n}{N_A\tau_n} + \frac{L_p}{N_D\tau_p}\right) \qquad (3.103)$$

3.22 Verlauf der Quasi-Ferminiveaus

Mit Hilfe der Ergebnisse der vorangegangenen Kapitel können wir direkt den Verlauf der Quasi-Ferminiveaus zeichnen. Wir betrachten hier den Fall der Flusspolung und unterscheiden dabei die drei Gebiete mit unterschiedlicher Teilchendichte und Transportvorgängen.

1. **Neutrale Bahngebiete**
   Bei einer langen Diode sind in ausreichend weitem Abstand von den Diffusionszonen die Bahngebiete bis zu den Kontakten bei w_p und w_n neutral. Den Spannungsabfall, und damit auch die Feldstärke über die Bahngebiete, haben wir als vernachlässigbar angenommen. Daher fallen wie in Abb. 3.14 in diesen äußeren Bereichen die beiden Quasi-Ferminiveaus mit dem Ferminiveau bei thermodynamischen Gleichgewicht zusammen.

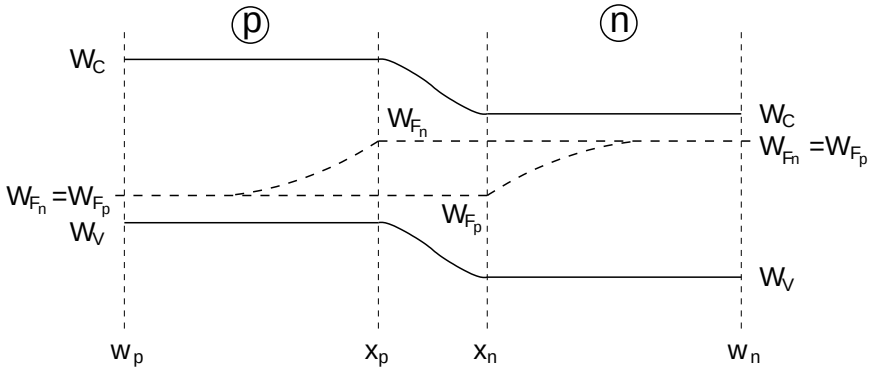

Abb. 3.14: Verlauf der Quasi-Ferminiveaus bei einer (langen) Diode mit Flusspolung ($W_{F_n} > W_{F_p}$).

   Bei einer kurzen Diode erzwingt die Randbedingung $n_p(w_p) = n_{p_0}$ und $p_n(w_n) = p_{n_0}$, aufgrund der Annahme einer beliebig schnellen Rekombination an den Kontakten, das Zusammenfallen der Ferminiveaus zumindest an den Kontaktstellen.

2. **Diffusionszonen in den Bahngebieten**
   In den Diffusionszonen ergeben sich die bereits in Kap. 2.25 ermittelten Verläufe. Danach verläuft das Quasi-Ferminiveau der jeweiligen Majoritätsträger aufgrund ihrer hohen Dichte konstant (Steigung $\sim \frac{1}{\text{Dichte}}$).

Die Steigung des Quasi-Ferminiveaus der Minoritätsträger ist propotional $\frac{1}{n}\frac{dn}{dx}$ für Elektronen, mit $n(x)$ nach Gl. (3.82). Für Löcher ist die entsprechenden Gleichung (3.85) zu verwenden. Es ergeben sich die Verläufe in Abb. 3.14. [24]

3. **Raumladungszone (Sperrschicht)**
In der Raumladungszone können wir den Verlauf der Quasi-Ferminiveaus über Gleichsetzen der beiden identischen Darstellungen der Ladungsträgerdichten nach Gl. (3.31), (3.32) mit Gl. (2.228), (2.229) ermitteln, dabei ist zu berücksichtigen, dass aufgrund der von außen an die Diode gelegten Spannung, U_D durch die über der Raumladungszone liegende Spannung U_{RLZ} ausgetauscht werden muss.
Nach kurzer Rechnung (zur Übung) ergibt sich

$$\frac{d\,W_{F_n}}{d\,x} = 0, \qquad \frac{d\,W_{F_p}}{d\,x} = 0\,, \tag{3.104}$$

d.h. die Quasi-Ferminiveaus verlaufen wie in Abb. 3.14 gezeigt horizontal in der Raumladungsszone. Aufgrund der Bandverbiegung von Leitungs- und Valenzband spiegelt dies die uns schon bekannte Tatsache wieder, dass die Ladungsträgerdichten mit wachsender Entfernung zu dem jeweiligen Bahngebiet in dem die Ladungsträger Majoritäten sind abnehmen.

3.23 Temperaturabhängigkeit des Diodenstroms

Wir verwenden zur Bestimmung des Temperaturkoeffizienten des Diodenstroms die aufwendige Beziehung ohne Interpolation nach (3.92).
Dieser erlaubt es uns, zwischen dem bei Sperrpolung dominanten Beitrag des Rekombinationsstroms I_{rg} und dem Beitrag des Diffusionsstroms in Form des Sättigungsstroms I_S zu unterscheiden. Der Diodenstrom ergibt sich demnach aus

$$I = I_S\,(e^{\frac{U}{U_T}} - 1) + I_{rg,s}\,(e^{\frac{U}{2U_T}} - 1)\,, \tag{3.105}$$

mit

$$I_S = e\,A\,n_i^2 \cdot C \tag{3.106}$$

[24] Streng genommen gilt der so ermittelte Verlauf der Quasi-Ferminiveaus aufgrund der Gültigkeit von Gl. (3.82), (3.85) auch in dem Bereich der neutralen Bahngebiete nach 1.

nach Gl. (3.99) bzw. (3.103), wobei C für den jeweiligen Klammerausdruck in den bei Gleichungen (3.99) bzw. (3.103) für die kurze bzw. lange Diode steht.
Für $I_{rg,s}$ verwenden wir aus Gl. (3.59)

$$I_{rg,s} = \frac{e\,A\,w_{RLZ}(U)}{\tau_{eff}} n_i\,. \tag{3.107}$$

Der Diodenstrom besitzt außer der direkten Temperaturabhängigkeit über $U_T = \frac{k \cdot T}{e}$ in den Ausdrücken $I_S(T)$ und $I_{rg,s}(T)$ noch weitere implizite Temperaturabhängigeiten in Form von $n_i(T)$, $C(T)$, $\tau_{eff}(T)$ und $w_{RLZ}(T)$.
Zur Vereinfachung wollen wir aufgrund der starken Temperaturabhängigkeit von n_i nach Gl. (2.36)

$$n_i(T) = \alpha T^{\frac{3}{2}}\, e^{\frac{-W_g}{2kT}} \tag{3.108}$$

die Temperaturabhängigkeiten von $C(T)$, $\tau_{eff}(T)$ und $w_{RLZ}(T)$ vernachlässigen. Eine ausführliche Rechnung, die den Gültigkeitsbereich dieser Annahme zeigt, findet sich z.B. in [Möschwitzer, Lunze].
Aufgrund der impliziten Temperaturabhängigkeit wenden wir die Kettenregel auf Gl. (3.105) an und erhalten

$$\begin{aligned}
\frac{d\,I}{d\,T} =& \frac{\partial I}{\partial U_T}\frac{d\,U_T}{d\,T} + \frac{\partial I}{\partial I_S}\frac{d\,I_S}{d\,T} + \frac{\partial I}{\partial I_{rg,s}}\frac{d\,I_{rg,s}}{d\,T} && (3.109)\\
=& \left[I_S\left(-\frac{U}{U_T^2}\right)e^{\frac{U}{U_T}} + I_{rg,s}\left(-\frac{U}{2U_T^2}\right)e^{\frac{U}{2U_T}}\right]\cdot\frac{U_T}{T}\\
&+ \left[e^{\frac{U}{U_T}} - 1\right]\frac{d\,I_S}{d\,T} + \left[e^{\frac{U}{2U_T}} - 1\right]\frac{d\,I_{rg,s}}{d\,T}\\
=& -\left[I_S\frac{U}{T\,U_T}e^{\frac{U}{U_T}} + I_{rg,s}\frac{U}{2T\,U_T}e^{\frac{U}{2U_T}}\right] && (3.110)\\
&+ \left[e^{\frac{U}{U_T}} - 1\right]\frac{d\,I_S}{d\,n_i}\,\frac{d\,n_i}{d\,T} + \left[e^{\frac{U}{2U_T}} - 1\right]\frac{d\,I_{rg,s}}{d\,n_i}\cdot\frac{d\,n_i}{d\,T}
\end{aligned}$$

Wir bestimmen die einzelnen darin enthaltenen Terme.
Aus Gl. (3.106) folgt

$$\frac{d\,I_S}{d\,n_i} = 2\,e\,A\,n_i\,C = 2\frac{I_S}{n_i} \tag{3.111}$$

und aus Gl. (3.107) folgt

$$\frac{d\,I_{rg,s}}{d\,n_i} = \frac{I_{rg,s}}{n_i} \tag{3.112}$$

Für den Temperaturkoeffizienten von n_i folgt nach kurzer einfacher Rechnung aus Gl. (3.108)

$$\frac{d\,n_i}{d\,T} = \frac{n_i}{T}\left(\frac{3}{2} + \frac{W_g}{2kT}\right) = \frac{n_i}{2T}\left(3 + \frac{U_g}{U_T}\right) \tag{3.113}$$

mit der Bandabstandsspannung

$$U_g := \frac{W_g}{e} \tag{3.114}$$

Einsetzen der zuvor bestimmten Terme nach Gl. (3.111), (3.112), (3.113) in Gl. (3.110) liefert den gesuchten Temperaturkoeffizienten des Diodenstroms

$$\begin{aligned}\frac{d\,I}{d\,T} &= -\frac{U}{T\cdot U_T}\left(I_S e^{\frac{U}{U_T}} + \frac{I_{rg,s}}{2}\,e^{\frac{U}{2U_T}}\right)\\ &\quad +\left(2I_S(e^{\frac{U}{U_T}} - 1) + I_{rg,s}(e^{\frac{U}{2U_T}} - 1)\right)\cdot\frac{1}{2T}\left(3 + \frac{U_g}{U_T}\right).\end{aligned} \tag{3.115}$$

Bei Flusspolung der Diode ($I := I_F$) mit $U_F \gg U_T$ gilt $e^{\frac{U_F}{2U_T}} \gg 1$ und Gl. (3.115) kann zusammengefasst werden zu

$$\frac{d\,I_F}{d\,T} = \left(I_S\,e^{\frac{U}{U_T}} + \frac{I_{rg,s}}{2}\,e^{\frac{U}{2U_T}}\right)\left(-\frac{U}{T\cdot U_T} + \frac{1}{T}\left(3 + \frac{U_g}{U_T}\right)\right) \tag{3.116}$$

$$= \left(I_S\,e^{\frac{U}{U_T}} + \frac{I_{rg,s}}{2}\,e^{\frac{U}{2U_T}}\right)\frac{1}{T}\left(3 + \frac{U_g - U_F}{U_T}\right). \tag{3.117}$$

Bei Sperrpolung der Diode ($I := I_R$) mit $U \ll U_T$ gilt $e^{\frac{U}{U_T}} \ll 1$ und Gl. (3.115) kann vereinfacht werden zu

$$\frac{d\,I_R}{d\,T} = -\left(I_S + \frac{I_{rg,s}}{2}\right)\frac{1}{T}\left(3 + \frac{U_g}{U_T}\right) \quad . \tag{3.118}$$

Wird mit dem vereinfachten Ausdruck nach Gl. (3.96) (z.B. in SPICE) gearbeitet ergibt sich im Flussbereich mit einem geeignet gewählten I_S (ungleich dem in Gl. (3.117) und (3.118))

$$\frac{d\,I_F}{d\,T} = I_S\cdot e^{\frac{U}{nU_T}}\cdot\frac{1}{T}\left(3 + \frac{U_g - U_F}{U_T}\right) \tag{3.119}$$

und im Sperrbereich

$$\frac{d\,I_R}{d\,T} = -I_S\,\frac{1}{T}\left(3 + \frac{U_g}{U_T}\right). \tag{3.120}$$

3.24 Stationäre Ladungssteuerung

3.24.1 Gespeicherte Minoritätsträger

Wir haben bereits ermittelt, dass ein Stromfluss der Diode in Flussrichtung dadurch entsteht, dass die Minoritätsträger-Randkonzentration gegenüber den Gleichgewichtsdichten um $e^{\frac{U}{U_T}}$ erhöht wird. In den Bahngebieten klingt die Minoritätsträgerdichte gemäß Gl. (3.82) ab. Wir wollen im Folgenden eine einfache Beziehung herleiten, die den Stromfluss in Abhängigkeit der, in den Bahngebieten vorhandenen (man sagt auch gespeicherten) überschüssigen Minoritätsträgerladungen ΔQ beschreibt.
Wir betrachten dazu exemplarisch die gespeicherten Minoritätsträger (Elektronen) im p-Bahngebiet entsprechend Abb. 3.15 und schließen von dem Ergebnis auf die entsprechende Beziehung der Minoritätsträger (Löcher) im n-Bahngebiet. Die Elektronenüberschussladung ΔQ_n im p-Bahngebiet ergibt

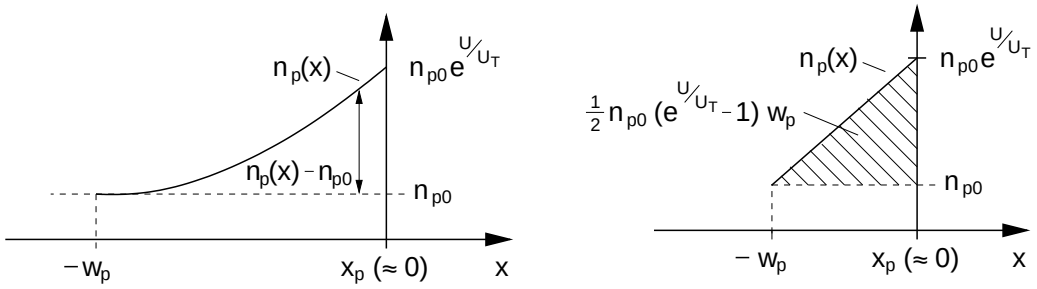

Abb. 3.15: Verlauf der Elektronendichte (Minoritätsträger) im p-Bahngebiet. Links: Lange Diode mit $w_p \gg L_p$. Rechts: kurze Diode ($w_p \ll L_p$) mit Diffusionsdreieck.

sich durch Integration von $n_p(x) - n_{p_0}$ aus Gl.(3.82) vom Rand der Sperrschicht bei x_p zum Kontakt bei $-w_p$

$$\Delta Q_n = -e\,A \int_{-w_p}^{x_p} (n_p(x) - n_{p_0})\,dx \tag{3.121}$$

$$= -e\,A\,n_{p_0}\,(e^{\frac{U}{U_T}} - 1) \int_{-w_p}^{x_p} \frac{\sinh(\frac{w_p+x}{L_n})}{\sinh(\frac{w_p+x_p}{L_n})}\,dx \tag{3.122}$$

$$= e\,A\,n_{p_0}\,(e^{\frac{U}{U_T}} - 1)\,L_n\,\frac{1-\cosh(\frac{w_p+x_p}{L_n})}{\sinh(\frac{w_p+x_p}{L_n})}\,. \tag{3.123}$$

Für $w_p \gg -x_p$, d. h. $w_p + x_p \approx w_p$ lässt sich wieder die Fallunterscheidung zwischen kurzer und langer Diode durchführen:

1. kurze Diode ($w_p \ll L_n$)

   Hier ist mit $\cosh(\varepsilon) \approx 1 + \frac{\varepsilon^2}{2}$ und $\sinh(\varepsilon) \approx \varepsilon$ für $|\varepsilon| \ll 1$

   $$\Delta Q_n \;\approx\; -e\,A\,n_{p0}\,(e^{\frac{U}{U_T}} - 1)\,\frac{w_p}{2}\;. \tag{3.124}$$

   Aufgrund der kurzen Länge kann hier der exponentielle Verlauf durch eine Gerade angenähert werden, wodurch sich das in Abb. 3.15 rechts gezeigte Diffusionsdreieck ergibt. Sein Flächeninhalt multipliziert mit der Elektronenladung und dem Diodenquerschnitt A (y-z-Ebene) ergibt die in dem Bahngebiet der kurzen Diode gespeicherte Minoritätsträgerladung.

2. lange Diode ($w_p \gg L_n$)

   Hier ist wegen $1 - \cosh(\frac{w_p}{L_n}) \;\approx\; -\cosh(\frac{w_p}{L_n})$ und $\tanh(\frac{w_p}{L_n}) \;\approx\; 1$

   $$\Delta Q_n \;\approx\; -e\,A\,n_{p0}\,(e^{\frac{U}{U_T}} - 1)\,L_n\;. \tag{3.125}$$

3.24.2 Berechnung des Diodenstroms aus den gespeicherten Ladungen

Der Diodenstrom setzt sich nach Gl. (3.89) zusammen aus

$$I \;=\; I_n(x_p) \;+\; I_p(x_n) \;+\; I_{rg}\;. \tag{3.126}$$

Wir betrachten den Beitrag der Elektronen nach Gl. (3.84) mit den Näherungen für die kurze und lange Diode. Der Beitrag der Löcher ergibt sich durch Substitution der entsprechenden Größen.

1. kurze Diode ($w_p \ll L_n$)

   Mit der Näherung $\tanh(\frac{w_p}{L_n}) \;\approx\; \frac{w_p}{L_n}$ folgt für $w_p + x_p \;\approx\; w_p$ aus Gl. (3.84)

   $$I_n(x_p) \;\approx\; e\,A\,D_n\,n_{p0}\,(e^{\frac{U}{U_T}} - 1)\,\frac{1}{w_p} \tag{3.127}$$

Wir ersetzen darin die entsprechenden Terme durch die in der Diode gespeicherte Minoritätsträgerladung nach Gl. (3.124) und erhalten

$$I_n(x_p) = \frac{2\,\Delta Q_n\,D_n}{w_p^2} = \frac{\Delta Q_n}{\tau_{Bn}} \tag{3.128}$$

$$\text{mit } \tau_{Bn} := \frac{w_p^2}{2\,D_n} \qquad \text{(Transitzeit)} \tag{3.129}$$

Da in der kurzen Diode alle Minoritätsträger ohne Rekombination durch das Bahngebiet fließen, lässt sich der Strom in diesem Fall interpretieren als die Überschussladung ΔQ_n, die in der mittleren Zeit τ_{Bn} das Bahngebiet durchquert. τ_{Bn} wird aus diesem Grund auch Transitzeit der Elektronen genannt.

2. lange Diode ($w_p \gg L_n$)

Mit der Näherung $\tanh(\frac{w_p}{L_n}) \approx 1$ folgt aus Gl. (3.84)

$$I_n(x_p) \approx e\,A\,D_n\,n_{p0}\,(e^{\frac{U}{U_T}} - 1)\,\frac{1}{L_n} \tag{3.130}$$

und durch Substitution mit Gl. (3.125)

$$I_n(x_p) = \frac{\Delta Q_n\,D_n}{L_n^2} = \frac{\Delta Q_n}{\tau_n} \tag{3.131}$$

$$\text{mit } \tau_n := \frac{L_n^2}{D_n} \qquad \text{nach Gl. (3.66)} \tag{3.132}$$

Da bei der langen Diode alle Überschussladungen rekombinieren, entspricht der Strom der Ladung der in der (mittleren) Lebensdauer rekombinierten Minoritätsträger.

Vernachlässigen wir bei Flusspolung den Rekombinationsstrom in der Sperrschicht ($I_{rg} = 0$) so ergibt sich mit Gl. (3.128) und (3.131) der Gesamtstrom der Diode zu

$$I_F = I_n(x_p) + I_p(x_n) \approx \frac{\Delta Q_n}{\tau_n^*} + \frac{\Delta Q_p}{\tau_p^*} \tag{3.133}$$

Darin ist für τ_n^* und τ_p^* der jeweilige Ausdruck für die lange bzw. kurze Diode einzusetzen (z.B. für τ_n^* Gl. (3.128) bzw. (3.132)). Der Strom durch die Diode in Flussrichtung wird also durch die Überschussladungen der Minoritäten in den Bahngebieten gesteuert. Der Elektronenstrom ist dabei direkt mit

der entsprechenden Zeitkonstanten der Elektronenüberschussladung im p-Bahngebiet proportional. Entsprechendes gilt für den Löcherstrom.
Mit anderen Worten bedeutet dies, dass solange eine Überschussladung existiert auch ein Strom durch die Diode fließt.
Betrachtet man dynamische Schaltvorgänge der Diode besteht die Forderung für ein schnelles Ausschalten der Diode (Übergang von Fluss- in Sperrrichtung) darin, die in den Bahngebieten gespeicherten Überschussladungen in möglichst kurzer Zeit auszuräumen. Um diese Vorgänge richtig zu beschreiben, reicht die vereinfachte Kontinuitätsgleichung unseres DDM's nicht mehr aus, da sie nur den stationären Fall erfasst. Für eine korrekte Beschreibung muss daher die vollständige Kontinuitätsgleichung mit dem zeitabhängigen Anteilen $\frac{\partial p}{\partial t}$ und $\frac{\partial n}{\partial t}$ verwendet werden.

Wir beschränken uns im Rahmen dieser Vorlesung jedoch auf die stationäre Betrachtung mit deren Hilfe auch die im Folgenden Kapitel beschriebene Diffusionskapazität ermittelt werden kann.

3.25 Diffusionskapazität

Durch die Flusspolung der Diode werden hohe Überschusskonzentrationen ($\sim e^{\frac{U}{U_T}}$) von Minoritätsträgern in die Bahngebiete injiziert. Aufgrund der Neutralisierung steigt die Majoritätsträgerdichte durch Umordnung der Majoritätsträger im gleichen Maße, so dass (quasi) Neutralität herrscht. Minoritäts- und Majoritätsträger ändern sich daher bei Variation des Diodenstroms gleichzeitig.
Die Variation des Diodenstroms wird durch Änderung der von außen an die Diode angelegten Spannung bewirkt. Für den Strom-Spannungszusammenhang gilt nach Gl. (3.91) bei vernachlässigbarem Nettorekombinationsstrom I_{rg} (Index F für Flusspolung)

$$I_F \;=\; I_S\,(e^{\frac{U_F}{U_T}} - 1)\,. \tag{3.134}$$

Aus der Theorie der stationären Ladungssteuerung wissen wir, dass eine Änderung des Diodenstroms gleichbedeutend ist, mit einer Änderung der Minoritätsträgerladungen in den Bahngebieten. Nach Gl. (3.133) gilt für den Zusammenhang zwischen Ladungen und Strom

$$I_F \;=\; \begin{cases} \frac{\Delta Q_n}{\tau_{Bn}} \;+\; \frac{\Delta Q_p}{\tau_{Bp}} & ,\text{kurze Diode} \\ \frac{\Delta Q_n}{\tau_n} \;+\; \frac{\Delta Q_p}{\tau_p} & ,\text{lange Diode} \end{cases} \tag{3.135}$$

Wir fassen die gespeicherten Löcher- und Elektronenladungen zusammen. Für die gesamte Minoritätsträgerspeicherladung schreiben wir

$$\Delta Q \;=\; \Delta Q_n + \Delta Q_p \tag{3.136}$$

und führen formal über eine Definitionsgleichung die gewichtete Zeitkonstante τ_T ein

$$\frac{\Delta Q}{\tau_T} := \frac{\Delta Q_n}{\tau_n^*} + \frac{\Delta Q_p}{\tau_p^*}\,. \tag{3.137}$$

Für die kurze Diode wird τ_T auch als Transitzeit bezeichnet:

$$\frac{1}{\tau_T} = \frac{1}{\tau_n^*}\frac{\Delta Q_n}{\Delta Q} + \frac{1}{\tau_p^*}\frac{\Delta Q_p}{\Delta Q}\,. \tag{3.138}$$

Bei Dioden und besonders bei $p-n$ Übergängen in Transistoren liegen meist stark unterschiedliche Dotierungen vor. Damit vereinfacht sich τ_T z.B. bei

$N_D \gg N_A$ wegen $n_{p0} = \frac{n_i^2}{N_A} \gg p_{n0} = \frac{n_i^2}{N_D}$, woraus wegen Gl. (3.125) $\Delta Q_p \ll \Delta Q_n \approx \Delta Q$ folgt, zu

$$\frac{1}{\tau_T} \approx \frac{1}{\tau_n^*} = \begin{cases} \tau_{Bn} = \frac{w_p^2}{2D_n} & \text{(Transitzeit), kurze Diode} \\ \tau_n = \frac{L_n^2}{D_n} & \text{, lange Diode} \end{cases} \tag{3.139}$$

Für den Strom in Vorwärtsrichtung erhalten wir in diesem Fall mit der gewichteten Zeitkonstanten anstelle von Gl. (3.135) den sehr einfachen Ausdruck

$$I_F \approx \frac{\Delta Q}{\tau_T} \tag{3.140}$$

der einen linearen Zusammenhang zwischen Flussstrom und gespeicherter Minoritätsträgerladung in den Bahngebieten beschreibt.
Durch Gleichsetzen von Gl. (3.140) und (3.134) erhalten wir einen Zusammenhang zwischen der gespeicherten Ladung in den Bahngebieten und der angelegten Flussspannung

$$I_S(e^{\frac{U_F}{U_T}} - 1) = \frac{\Delta Q}{\tau_T}\,. \tag{3.141}$$

Die Änderung der Ladung dQ eines Zweipols bezogen auf die zur Ladungsänderung gehörende Spannungsänderung an den Klemmen des Zweipols wird als Kapazität C bezeichnet

$$C := \frac{dQ}{dU}\,. \tag{3.142}$$

Durch Ableiten von Gl. (3.141) nach U_F erhalten wir die in der Diode zu einer Spannungsänderung dU_F gehörende Ladungsänderung

$$\frac{I_S}{U_T}\, e^{\frac{U_F}{U_T}} = \frac{1}{\tau_T}\, \frac{d\Delta Q}{dU_F} \tag{3.143}$$

Darin definieren wir den Differentialquotienten gemäß Gl. (3.142) als die Diffusionskapazität

$$C_d := \frac{d\Delta Q}{dU_F} = \tau_T\, \frac{I_S}{U_T}\, e^{\frac{U_F}{U_T}} \tag{3.144}$$

Für die in den meisten Fällen erfüllte Näherung $U_F \gg U_T$ gilt

$$I_F = I_S\,(e^{\frac{U_F}{U_T}} - 1) \approx I_S\, e^{\frac{U_F}{U_T}} \tag{3.145}$$

wodurch für C_d auch geschrieben werden kann

$$C_d = \tau_T \frac{I_F}{U_T} . \tag{3.146}$$

Das ist ein sehr wichtiger Zusammenhang, der auch beim Bipolar-Transistor auftritt. Wir wollen mit Hilfe von Abb. 3.16 am Beispiel der kurzen Diode mit $N_D \gg N_A$ die Bedeutung verdeutlichen.

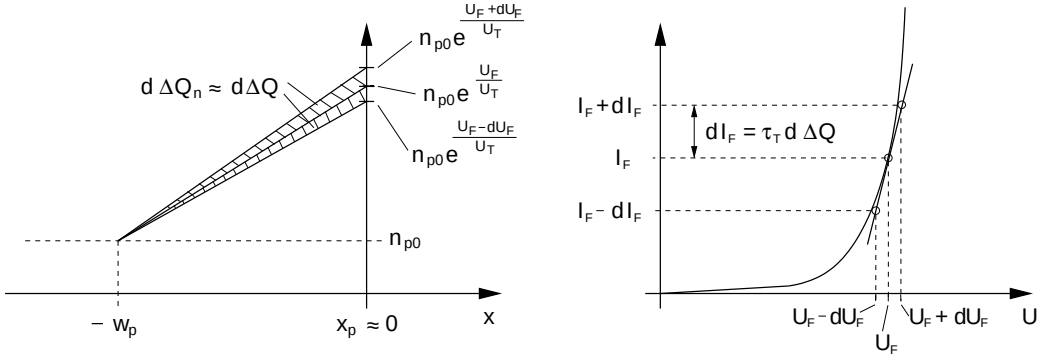

Abb. 3.16: Links: Änderung der Minoritätsträger-Überschussladung $d\Delta Q_n$ bei Änderung der Flussspannung um dU_F. Wegen $N_D \gg N_A$ gilt $\Delta Q_p \approx 0$ wodurch $\Delta Q \approx Q_n$. Rechts: Zugehörige quasistatische Diodenkennlinie.

Bei einer Flussspannung U_F soll durch die Diode ein Flussstrom I_F fließen. Die Diffusionskapazität besitzt dann den Wert C_d nach Gl. (3.146). Bei kleinen Änderungen $dU_F \ll U_F$ der Flussspannung, für die weiterhin stationärer Zustand gilt, ist dann die Diffusionskapazität C_d wirksam. Die in ihr gespeicherten Ladungen müssen durch den Flussstrom auf- bzw. abgebaut werden. Je kleiner die Diffusionskapazität umso kleiner die gespeicherte Minoritätsträgerladung und umso schneller erfolgt die Umladung.
Das Spannungs-Strompaar U_F, I_F um das die kleine Auslenkung um dU_F bzw. dI_F erfolgt, wird als der Arbeitspunkt der Diode bezeichnet. Wegen der kleinen Aussteuerungen um den Arbeitspunkt wird das aussteuernde Signal dU_F, dI_F auch als Kleinsignalaussteuerung bezeichnet. Die Kleinsignalaussteuerung ist solange gültig, wie die Abweichungen von dem Arbeitspunkt durch die im Arbeitspunkt ermittelte Diffusionskapazität C_d als Mittelwert richtig beschrieben werden.
Die Grenzen der stationären Näherung liegen dort, wo die Ladungsänderung nicht mehr als Differenz zweier Überschussladungen im stationären Fall wie

in Abb. 3.16 links ermittelt werden können. Bei schnellen Änderungen ist zu berücksichtigen, dass die Änderung der Überschussladung an der Grenze der Raumladungszone beginnt und sich von dort in die Bahngebiete fortsetzt. Dadurch ergeben sich z.B. die in Abb. 3.17 gezeigten Verläufe der Elektronenladung, wie sie als Lösung der vollständigen Kontinuitätsgleichung ermittelt werden können. Der dargestellte Fall zeigt

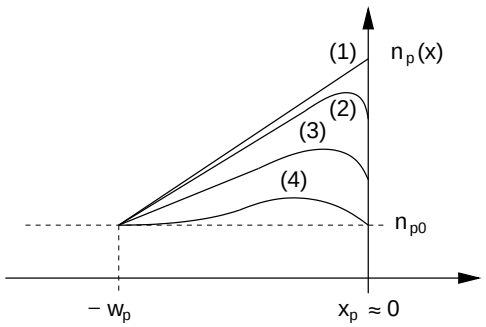

Abb. 3.17: Verlauf der Minoritätsträgerdichte bei nicht stationärem Abbau des Ladungsüberschusses in (1). Kurve (2) kann als Kleinsignalaussteuerung bei Annahme des Abeitspunktes in (1) interpretiert werden (vgl. Abb. 3.16). Großsignalaussteuerung liegt bei Kurven (3) und (4) vor, wenn der Ausgangspunkt der Aussteuerung in (1) liegt.

mit den Verläufen (3) und (4) auch den Fall der Großsignalaussteuerung, für den eine bei einem Arbeitspunkt in (1) ermittelte Diffusionskapazität nicht mehr gültig ist.

3.26 Sperrschichtkapazität

Wir haben bereits in Kapitel 3.9 gesehen, dass in der Raumladungszone der p-n Diode zwei gleich große Raumladungen entgegengesetzter Polarität vorhanden sind, die sich aufgrund der Neutralitätsbedingung nach Gl. (3.10)

$$-N_A\, x_p = N_D\, x_n \tag{3.147}$$

kompensieren (Beachten: Der metallurgische Übergang zwischen p- und n-Gebiet liegt bei dieser Definition bei $x = 0$. Aufgrund der Wahl des Nullpunktes ist die x-Koordinate des links von $x = 0$ liegenden Gebietes negativ). Unter Verwendung der Rechteck-Profil Näherung nach Gl. (3.5) ergaben sich

daraus die Ränder der Raumladungszone nach Gl. (3.25) bei

$$x_n = \sqrt{\frac{2\,\varepsilon\,U_{RLZ}}{e}\,\frac{N_A}{N_D}\,\frac{1}{N_A+N_D}} = -x_p\,\frac{N_A}{N_D}\,. \tag{3.148}$$

Darin wird die extern an die Diode angelegte Spannung U entsprechend Gl. (3.40) in

$$U_{RLZ} = U_D - U \tag{3.149}$$

berücksichtigt. Danach verringert sich die Weite der Raumladungszone bei angelegten Spannungen in Flussrichtung ($U = U_F > 0$). Entsprechend vergrößert sich die Weite bei Polung in Sperrrichtung.
Durch die Verschiebung der Ränder der Raumladungszone verändert sich auch die in der Raumladungszone befindende Ladung um den von der Verschiebung betroffenen Anteil. Die Ladungsänderung erfolgt also nur an den Rändern der Raumladungszone. Da wir eine homogene Diode in y- und z-Richtung voraussetzen, beinhaltet diese Änderung z.B. im n-Gebiet, wie in Abb. 3.18 gezeigt, die in der Raumladungsweiten-Änderung dx_n enthaltene Ladung

$$dQ_p = -e\,A\,dx_n\,N_D\,. \tag{3.150}$$

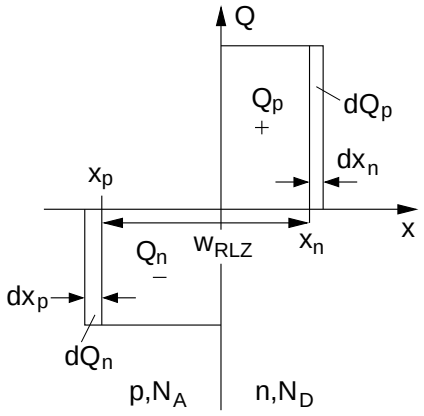

Abb. 3.18: Änderung der Weite der Raumladungszone und damit verbundene Ladungsänderung.

Mit Q_p bzw dQ_p wird die positive Ladung aufgrund der ionisierten Donatoren bezeichnet. Entsprechend stehen Q_n und dQ_n für negative Ladungen

aufgrund ionisierter Akzeptoren.
Aufgrund der Neutralitätsbedingung Gl. (3.147) erfolgt die gleiche Ladungsänderung auch auf der p-Seite der Diode, wozu auf der Seite eine Weitenänderung $dx_p = -\frac{N_D}{N_A}\, dx_n$ notwendig ist. Wir verschieben also durch eine Änderung der externen Spannung U um dU, zwei im Abstand $w_{RLZ} = x_n - x_p$ befindende Platten mit der Fläche A um dx_p bzw. dx_n und ändern dabei die sich auf[25] diesen Platten befindende Ladung um dQ_p $(= -dQ_n)$. Wie bereits bei der Diffusionskapazität ordnen wir der mit der Spannungsänderung dU verbundenen Ladungsänderung dQ_p eine Kapazität

$$C_j = \frac{dQ_p}{dU} \tag{3.151}$$

zu, die wir aufgrund ihrer Ursache mit Sperrschichtkapazität bezeichnen. Aufgrund der vorangegangenen Überlegungen erwarten wir als Ergebnis für C_j einen Plattenkondensator mit einem Plattenabstand w_{RLZ}.
Um C_j zu berechnen leiten wir Gl. (3.148) nach U ab und erhalten

$$\frac{dx_n}{dU} = -\sqrt{\frac{\varepsilon}{2eU_{RLZ}}\,\frac{N_A}{N_D}\,\frac{1}{N_A + N_D}} \tag{3.152}$$

und damit über Gl. (3.150) die mit dU verdundene Ladungsänderung

$$dQ_p = e\,A\,N_D\,dU\,\sqrt{\frac{\varepsilon}{2eU_{RLZ}}\,\frac{N_A}{N_D}\,\frac{1}{N_A + N_D}}\quad , \tag{3.153}$$

woraus sich mit der Definition nach Gl. (3.151) die Sperrschichtkapazität

$$C_j = \frac{dQ_p}{dU} = A\,\sqrt{\frac{\varepsilon\, e}{2(U_D - U)}\,\frac{N_A\,N_D}{N_A + N_D}} \tag{3.154}$$

ergibt. Zwischen der Sperrschichtkapazität und der externen Spannung besteht also ein nichtlinearer Zusammenhang. Zur Übung sollten Sie einmal versuchen zu zeigen, wie Gl. (3.154) in die Formel eines Plattenkondensators mit $C = \frac{\varepsilon \cdot A}{w_{RLZ}}$ überführt werden kann.
Aufgrund unserer einfachen Annahmen bei der Herleitung gilt Gl. (3.154) nicht mehr, wenn U sich U_D annähert, wodurch w_{RLZ} sehr klein und die Injektion von Ladungsträgern sehr groß wird. Aufwendigere Berechnungen und Messungen zeigen das in Abb. 3.19 gezeigte Verhalten der Sperrschichtkapazität in Abhängigkeit von der angelegten Spannung. Danach nimmt die

[25] Genauer müsste es „in diesen Platten“ heißen, wenn den Platten die infinitesimalen Dicken dx_p, dx_n zugeordnet sind. Da aber die Dicken sehr viel kleiner als x_n bzw. x_p sind ist die Vorstellung einer dünnen Platte, auf der Ladungen sitzen, ebenfalls richtig.

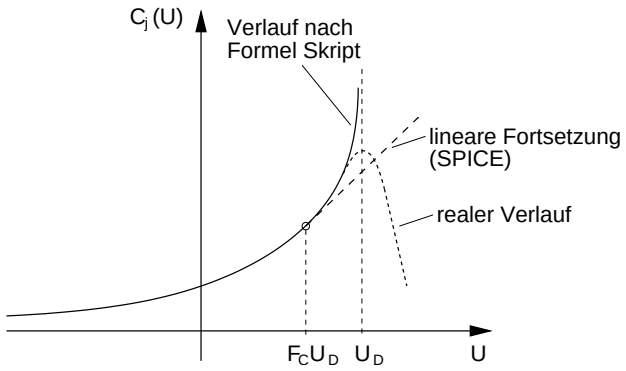

Abb. 3.19: Abhängigkeit der Sperrschichtkapazität von der angelegten Spannung.

Sperrschichtkapazität bei Werten um U_D zwar einen Maximalwert an, fällt jedoch bei größeren Werten wieder ab.
Im SPICE Gummel-Poon Modell für die Schaltungssimulation wird der Pol bei U_D dadurch umgangen und der reale Verlauf besser angenähert, indem der Verlauf nach Gl. (3.154) ab einer Spannung $F_C \cdot U_D$ linear fortgesetzt wird. F_C ist daher ein Parameter des Dioden- und des Transistormodells der je nach Charakteristik des $C(U)$ Verlaufs so gewählt (fitting) wird, dass eine bestmögliche Anpassung an den realen Verlauf in dem interessierenden Bereich um den Arbeitspunkt des Übergangs stattfindet.
Gl. (3.154) kann auch durch Messung des $C(U)$ Verlaufs dazu genutzt werden um Dotierungskonzentration und Diffusionsspannung zu ermitteln.

3.27 Großsignalersatzschaltbild

Wir haben zuvor die Diffusions- und Sperrschichtkapazität als zwei stark vom jeweiligen Arbeitspunkt der Diode abhängige Größen kennengelernt. Formal können wir für den jeweiligen Arbeitspunkt beide Kapazitäten mit Gl. (3.146) und (3.154) berechnen.
Für ein Großsignalersatzschaltbild der Diode in Abb. 3.20 berücksichtigen wir die beiden Kapazitäten daher formal mit $C_d(U)$ und $C_j(U)$. Dabei ist U die an der Diode anliegende Spannung. Aufgrund der Leitfähigkeit der Bahngebiete sowie durch Kontakt - und Zuleitungswiderstände ist je nach Erheblichkeit des Einflusses in der Anwendung noch ein zusätzlicher Serienwiderstand R zu berücksichtigen. Eine extern an der Diode angelegte Spannung U_x ist dann um $I \cdot R$ größer als die eigentliche Spannung U an der

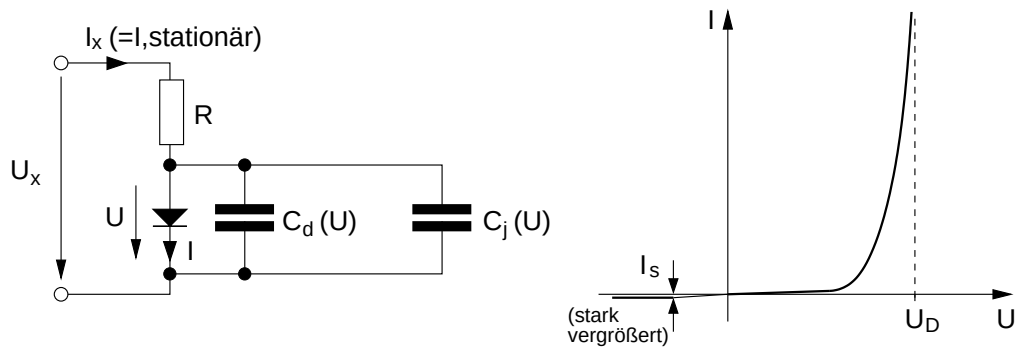

Abb. 3.20: Großsignalersatzschaltbild der p-n-Diode (links) und quasistatische Kennlinie der eigentlichen Diode (rechts).

Diode. Die Diode selbst wird in diesem Großsignalersatzschaltbild durch ihre Strom-Spannungsbeziehung nach Gl. (3.92) bzw. einfacher nach Gl. (3.94) oder (3.96) beschrieben, wodurch sich die in Abb. (3.20) rechts gezeigte Kennlinie ergibt.

3.28 Kleinsignalleitwert

Wir haben mit der Diffusions- und der Sperrschichtkapazität schon zwei Bauelemente kennengelernt, die die Eigenschaften des p-n-Übergangs in einem gewählten Arbeitspunkt (Wertepaar $\{U_0,\ I_0\}$) beschreiben. Bei kleinen Aussteuerungen um diesen Arbeitspunkt können diese Elemente als konstant angenommen werden.
Wir wollen auch die eigentliche Diode in Abb. 3.20 durch ein lineares Bauelement beschreiben, das bei kleinen Aussteuerungen um den Arbeitspunkt konstant ist.

Dazu entwickeln wir die Kennlinie der Diode wie in Abb. 3.21 gezeigt, im Arbeitspunkt ($\{U_0,\ I_0\}$) in eine Taylorreihe, die wir nach dem linearen Glied abbrechen. Wir nähern damit also den Verlauf der Kennlinie um den Arbeitspunkt durch eine Gerade an, die die gleiche Steigung hat wie die Originalfunktion im Arbeitspunkt.
Für die Kennlinie der Diode wählen wir Gl. (3.92).

$$I(U) = I_S(e^{\frac{U}{U_T}} - 1) + I_{r_g,s}(e^{\frac{U}{2U_T}} - 1) \tag{3.155}$$

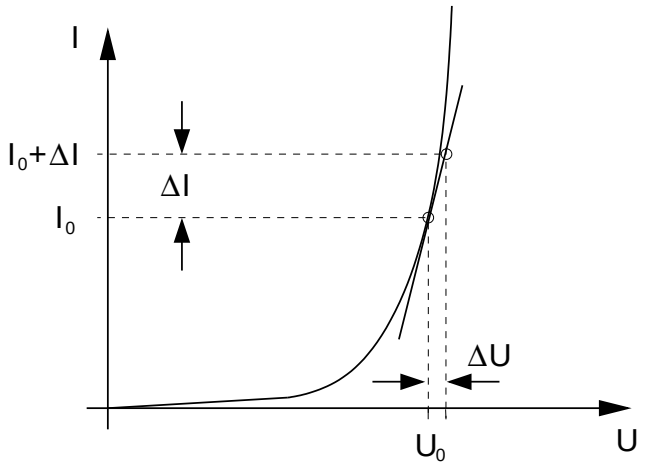

Abb. 3.21: Linearisierung der Diodenkennlinie um einen Arbeitspunkt $\{U_0,\ I_0\}$.

Die Taylorreihe dieser Kennlinie bei einer Abweichung ΔU vom Arbeitspunkt U_0 lautet

$$I(U_0+\Delta U) = I(U_0)+ \left.\frac{dI(U)}{dU}\right|_{U=U_0} \cdot \Delta U + \frac{1}{2} \left.\frac{d^2I(U)}{dU^2}\right|_{U=U_0} \cdot \Delta U^2 + \dots \quad (3.156)$$

Bei Abbruch nach dem ersten Glied ergibt sich als lieare Approximation der Tangente im Arbeitspunkt

$$I(U_0 + \Delta U) = I(U_0) + \left(\frac{I_S}{U_T}\, e^{\frac{U_0}{U_T}} + \frac{I_{rg,s}}{2\,U_T}\, e^{\frac{U_0}{2U_T}} \right) \Delta U \ . \quad (3.157)$$

Die Differenz $I(U_0 + \Delta U) - I(U_0) = \Delta I$ (vgl. Abb. (3.21) ist die sich bei linearer Approximation im Arbeitspunkt bei U_0 einstellende Stromänderung bei Änderung der Spannung um ΔU.
Lassen wir die Änderung infinitesimal klein werden, so erhalten wir den Differentialquotienten

$$g_d := \lim_{\Delta U \to 0} \frac{I(U_0 + \Delta U) - I(U_0)}{\Delta U} = \lim_{\Delta U \to 0} \frac{\Delta I}{\Delta U} = \frac{dI}{dU} = \frac{I_S}{U_T} e^{\frac{U_0}{U_T}} + \frac{I_{rg,s}}{2U_T} e^{\frac{U_0}{2U_T}} \quad (3.158)$$

den wir als Kleinsignalleitwert g_d bezeichnen.

Bei Flusspolung mit U_{0F} ist der Nettorekombinationsstrom vernachlässigbar ($I_{rg,s}$ in Gl. (3.158)) und es gilt mit

$$I_F(U_{0F}) = I_S(e^{\frac{U_{0F}}{U_T}} - 1) \approx I_S\, e^{\frac{U_{0F}}{U_T}} . \quad (3.159)$$

und für den Kleinsignalleitwert kann vereinfacht geschrieben werden

$$g_d \approx \frac{I_F(U_{0F})}{U_T}. \tag{3.160}$$

Der Kleinsignalleitwert ist also proportional dem Strom im Arbeitspunkt. Aufgrund $U_T \approx 25\text{mV}$ ergeben sich für Ströme im mA-Bereich Leitwerte im $\frac{1}{\Omega}$-Bereich.

Beispiel: Eine Diode wird im Arbeitspunkt mit einem Strom von 27 mA in Flussrichtung betrieben. Die Temperatur ist gerade so groß, dass sich ein $U_T = 27$ mV einstellt. Daraus ergibt sich ein Kleinsignalleitwert von

$$g_d \approx \frac{27\text{mA}}{27\text{mV}} = 1\text{S} = \frac{1}{1\Omega}\,.$$

3.29 Kleinsignalersatzschaltbild

Mit den bisher ermittelten Größen lässt sich das in Abb. 3.22 dargestellte Kleinsignalersatzschaltbild der Diode zeichnen. Die darin enthaltenen Größen

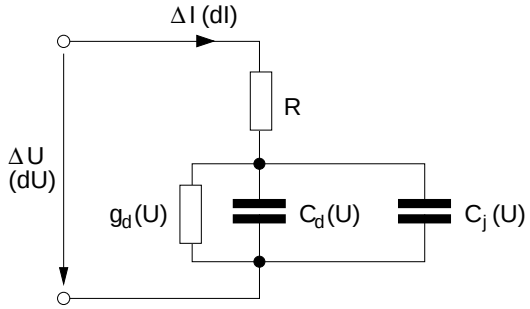

Abb. 3.22: Kleinsignalersatzschaltbild der p-n-Diode.

mit Ausnahme des Kontakt- und Bahnwiderstandes R hängen von dem gewählten Arbeitspunkt (U_0) ab. Ströme und Spannungen in diesem Ersatzschaltbild entsprechen den, sich auf der linearen Approximation um den Arbeitspunkt der spannungsabhängigen Verläufe der Großsignalelemente in Abb. 3.20 ergebenden Änderungen.
Spannungen und Ströme des Arbeitspunktes sind daher in der Kleinsignalbetrachtung in Abb. 3.22 nicht mehr enthalten. Sie bestimmen jedoch indirekt

die Werte der darin enthaltenen Bauelemente. Für die Sperrschichtkapazität gilt nach Gl. (3.154)

$$C_j(U) = A \sqrt{\frac{\varepsilon\, e}{2\,(U_D - U)} \frac{N_A\, N_D}{N_A + N_D}} \tag{3.161}$$

Für die Diffusionskapazität bei Flusspolung mit $U \gg U_T$ gilt nach Gl. (3.146)

$$C_d(U_F) = \tau_T \cdot \frac{I_F(U_F)}{U_T} \tag{3.162}$$

mit

$$I_F(U_F) \approx I_S\, e^{\frac{U_F}{U_T}} \tag{3.163}$$

Und für den Kleinsignalleitwert gilt mit Gl. (3.160) bei Flusspolung mit $U_F \gg U_T$ und mit Gl. (3.162)

$$g_d(U_F) = \frac{I_F(U_F)}{U_T} = \frac{C_d(U_F)}{\tau_T} \tag{3.164}$$

3.30 Stoßionisation, Lawineneffekt

Die Ladungsträger in der Diode sind in der Raumladungszone einer hohen Feldstärke ausgesetzt. Zum Beispiel ergibt sich bei einer Sperrspannung von 10V und $w_{RLZ} = 1\mu$m eine Feldstärke von 100kV/cm. Bei solch hohen Feldstärken nehmen die Ladungsträger in dem elektrischen Feld soviel Energie auf, dass sie bei einem Zusammenstoß mit dem Gitter ein Elektron aus dem Gitter lösen. Dabei wird dem herausgelösten Elektron die zur Ionisierung notwendige Energie zugeführt. Für den Vorgang der Stromleitung steht dann ein zusätzliches Elektron und das durch die Ionisierung vorhandene Loch zur Verfügung.
Abb. 3.23 zeigt schematisch diesen Vorgang, der sowohl durch ein Elektron wie auch durch ein Loch verursacht werden kann.

Nach dem Stoß beschleunigen die freigewordenen Ladungsträger in dem elektrischen Feld. Löcher beschleunigen in Richtung des Feldes, Elektronen in entgegengesetzter Richtung.
Beide erzeugten Ladungsträger nehmen durch die Beschleunigung im Feld Energie auf. Reicht diese Energie aus um bei einem erneuten Stoß mit dem Gitter wieder ein (oder zwei) Elektron-Loch Paar(e) zu erzeugen, so können

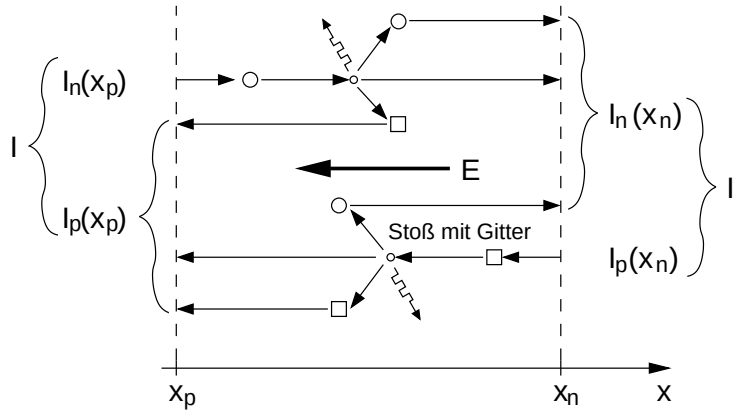

Abb. 3.23: Schematische Darstellung der Stoßionisation in der Raumladungszone.

auch diese neuen Ladungsträger an dem Paarbildungsprozess teilnehmen. Es entsteht eine Ladungsträgerlawine, die dem Lawineneffekt den Namen gegeben hat. Durch den starken Anstieg freier Ladungsträger steigt die Leitfähigkeit im selben Maße und es kommt trotz der Sperrpolung zu einem plötzlich einsetzenden Stromfluss.
Wir wollen diesen Vorgang mathematisch beschreiben. Dafür benötigen wir ein Modell für die Ladungsträgergeneration in der Sperrschicht. Hierzu werden die Ionisationskoeffizienten α_n und α_p eingeführt, die die mittlere Anzahl von Stoßionisationen angeben, die ein Ladungsträger längs eines Wegelementes dx erzielt. Aufgrund der vorangegangenen Erläuterungen ist verständlich, dass α_n und α_p stark feldabhängig sind. Abb. 3.24 zeigt diese Abhängigkeit für Si, Ge und GaAs.
Zur Beschreibung der Verläufe eignet sich z.B. folgendes Modell

$$\alpha = \alpha_0 \left| \frac{E}{E_0} \right|^m = K \cdot |E|^m \tag{3.165}$$

Mit Hilfe der Ionisationskoeffizienten kann für die Generationsrate g_s der durch Stoßionisation erzeugten Elektron-Loch Paare ein einfaches mathematisches Modell aufgestellt werden:

$$g_s = g_{s_n} = g_{s_p} = \alpha_n \, n \, v_{D_n} + \alpha_p \, p \, v_{D_p} \quad . \tag{3.166}$$

Sämtliche darin enthaltenen Größen sind aufgrund der ortsabhängigen Feldstärke auch ortsabhängig. Das Modell besagt, dass die an einem Ort pro Zeiteinheit generierten Elektron-Loch Paare proportional der

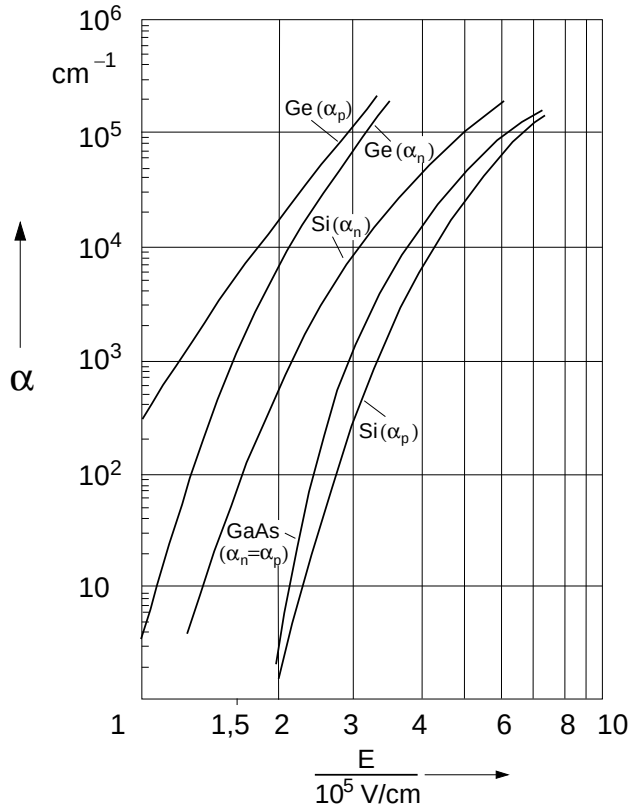

Abb. 3.24: Ionisationskoeffizienten α_n, α_p für Si, Ge und GaAs.

dort herrschenden Ladungsträgerdichte (n, p), der Geschwindigkeit[26] der stoßenden Ladungsträger und den Ionisationskoeffizienten sind. Da es sich um Elektron-Loch Paare handelt sind die Generationsraten für beide Ladungsträgerarten gleich.

Zur Vereinfachung der Rechnung nehmen wir für die Ionisationskoeffizienten einen geeigneten Mittelwert α an, wodurch aus Gl. (3.166)

$$g_s = (n\, v_{D_n} + p\, v_{D_p})\, \alpha \tag{3.167}$$

wird. Für den Mittelwert α kann z.B. für Si $K = 10^{-35}\,\mathrm{V}^{-7}\mathrm{cm}^6$ und $m = 7$ verwendet werden.

[26]Die Zahl der Paare, die ein Ladungsträger auf dem Weg dx erzeugt ist dx proportional. Betrachtet man diese Anzahl pro Zeitintervall dt (Generationsrate) so stellt $\frac{dx}{dt} = v_D$ die Driftgeschwindigkeit dar.

Mit der allgemeinen Formulierung für die (Feld)Stromdichte (vgl. Gl. (2.85))

$$J = e\,(n\,v_{D_n} + p\,v_{D_p}) \tag{3.168}$$

lässt sich Gl. (3.167) umformen, wobei zu beachten ist, dass die Stromflussrichtung aufgrund der Sperrpolung in negativer x-Richtung verläuft

$$g_s = -\frac{1}{e}\,\alpha\,J. \tag{3.169}$$

Wir setzen diese Generationsrate in die Kontinuitätsgleichung für Elektronen (für Löcher ergibt sich die gleiche Vorgehensweise) ein und berücksichtigen dabei noch den Anteil der SRH-Rekombination in der Sperrschicht mit R_{SRH}.

$$\frac{dJ_n}{dx} = e \cdot R = e\,(r - g) = e\,(R_{SRH} - g_s) = e\,R_{SRH} + \alpha\,J \tag{3.170}$$

Darin wird der Elektronenstrom mit dem Gesamtstrom (Stromdichte J), der durch die Raumladungszone fließt, verknüpft. Da dieser ortsunabhängig konstant sein muss (Kontinuitätsgleichung) ergibt eine Integration von Gl. (3.170) über die Raumladungszone

$$J_n(x_n) - J_n(x_p) = e \int_{x_p}^{x_n} R_{SRH}\,dx + J \int_{x_p}^{x_n} \alpha\,dx. \tag{3.171}$$

Wir verwenden Ströme anstelle der Stromdichten (Division durch A) und ersetzen dabei das Integral der SRH-Rekombination durch den entsprechenden Rekombinationsstrom I_{rg} aus Gl. (3.54)

$$I_n(x_n) - (I_n(x_p) + I_{rg}) = +I \int_{x_p}^{x_n} \alpha\,dx \tag{3.172}$$

Die linke Seite lässt sich umformen. Dazu verwenden wir Gl. (3.48), die uns den Gesamtstrom durch die Diode angibt

$$I_{R0} = I_n(x_p) + I_{rg} + I_p(x_n) \tag{3.173}$$

Dies ist der (Sperr)Strom ohne die Stoßionisation, denn wie man sich leicht anhand von Abb. 3.23 verdeutlichen kann, sind $I_n(x_p)$ und $I_p(x_n)$ die in die RLZ eintretenden Ströme, die noch nicht durch die Stoßionisation vergrößert sind.

Einsetzen der nach $(I_n(x_p)+I_{rg})$ umgestellten Gl. (3.173) in Gl. (3.172) liefert unter der Berücksichtigung, dass $I_n(x_n) + I_p(x_p) = I$ der Gesamtstrom der Diode mit Ladungsträgermultiplikation (vgl. Abb.3.23) ist

$$I - I_{R0} = I \int_{x_p}^{x_n} \alpha \, dx \tag{3.174}$$

und umgestellt

$$I = \frac{I_{R0}}{1 - \int_{x_p}^{x_n} \alpha \, dx} = I_{R0} \cdot M \tag{3.175}$$

mit dem Multiplikationsfaktor

$$M := \frac{1}{1 - \int_{x_p}^{x_n} \alpha \, dx} \tag{3.176}$$

Er beschreibt um wieviel der Diodenstrom aufgrund des Lawineneffektes größer als der Sperrstrom I_{R0} ohne Ladungsträgermultiplikation ist.
Für I_{R0} gilt wegen $U \gg U_T$ nach Gl. (3.92)

$$I_{R0} = I_S + I_{rg,s} \tag{3.177}$$

Für

$$\int_{x_p}^{x_n} \alpha \, dx \; = 1 \;\; \Rightarrow \;\; M \to \infty \tag{3.178}$$

geht M und damit der Diodenstrom gegen ∞. In der Praxis wird der Strom schon allein durch vorhandene Serienwiderstände begrenzt. Da α definiert ist als die mittlere Anzahl von Stoßionisationsvorgängen die ein Ladungsträger längs eines Weges dx erfährt, ist die Forderung nach Gl. (3.178) identisch mit der Aussage, dass im Mittel jeder Ladungsträger in der RLZ ein Elektron-Loch Paar durch Stoßionisation erzeugt. Dies ist der Einsatz des Lawineneffektes.

Ist der Wert der Spannung bei dem der Lawineneffekt einsetzt bekannt, kann der Multiplikationsfaktor alternativ zu Gl. (3.176) auch beschrieben werden durch

$$M = \frac{1}{1 - \left|\frac{U_R}{U_{br}}\right|^n} \tag{3.179}$$

wobei n eine empirisch zu ermittelnde Größe im Bereich $2 < n < 6$ ist. Die Größe U_{br} wird als Durchbruchspannung bezeichnet, das Einsetzen des Lawineneffektes mit Durchbruch. U_R ist die an die Diode angelegte Sperrspannung.

Temperaturabhängigkeit:

Die Wahrscheinlichkeit, dass ein Ladungsträger im elektrischen Feld genügend Energie für eine Stoßionisation aufnehmen kann, steigt mit der mittleren freien Weglänge. Bei kleinen freien Weglängen geben die Ladungsträger durch Streuung an Phononen bereits Energie ab, so dass höhere Feldstärken notwendig sind um die notwendige kinetische Energie aufnehmen zu können. Da die mittlere freie Weglänge mit steigender Temperatur abnimmt, sind höhere Feldstärken für den Durchbruch notwendig. Für den Einsatz der Diode als Bauelement bedeutet dies, dass bei höheren Temperaturen der Durchbruch erst bei höheren Durchbruchspannungen U_{br} erfolgt. Der Lawinendurchbruch hat also einen positiven Temperaturkoeffizienten (TK).

3.31 Tunnel-Effekt (Zener-Effekt)

Der Stromfluss über den p-n-Übergang infolge des Tunneleffekts wird auch Zener-Effekt genannt. Zum Verständnis dieses Effekts betrachten wir das Bändermodell des p-n-Übergangs. Das elektrische Feld in der Raumladungszone mit $E(x) = -\frac{d\varphi}{dx}$ bewirkt dort eine Bandverbiegung um $-e \cdot \varphi(x)$ mit dem Maximalwert $-e\,(U_D - U)$ zwischen den Rändern der Diffusionszonen zur Raumladungszone. Eine Sperrspannung $U < 0$ vergrößert also die Bandverbiegung und ab einer bestimmten Sperrspannung liegen Niveaus des Valenzbandes und des Leitungsbandes auf gleicher energetischer Höhe. In Abb. 3.25 sind die Bandkanten und die Quasi-Fermi-Niveaus für diesen Fall dargestellt. Die Ausdehnung der Raumladungszone ist dabei stark übertrieben.

Es stehen viele mit Elektronen besetzte Niveaus im Valenzband auf der p-Seite vielen leeren Niveaus im Leitungsband auf der n-Seite gegenüber. Nach den Gesetzen der klassischen Physik müssen die Elektronen des Valenzbandes jedoch zunächst ins Leitungsband gelangen, um durch die Raumladungszone

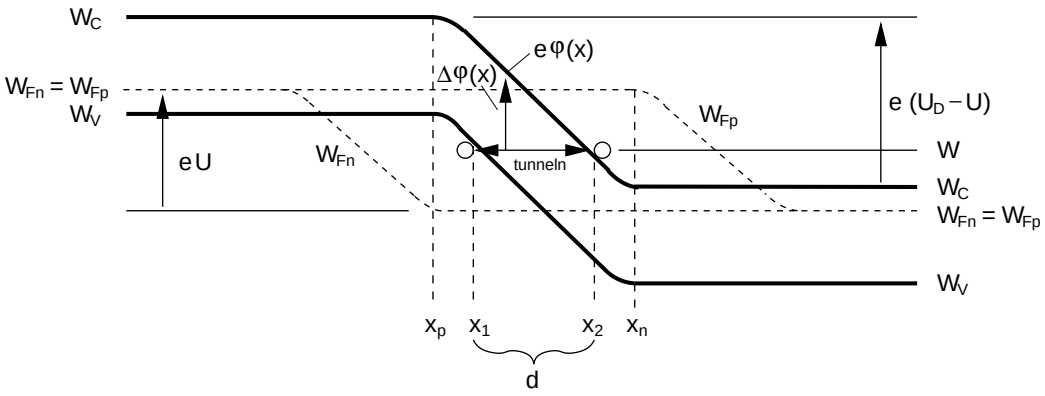

Abb. 3.25: Bandverläufe und Quasi-Fermi-Niveaus bei einem p-n-Übergang mit starker Sperrpolung ($U < 0$).

hindurchzukommen. Dann kann die aufgenommene Energie durch Thermalisierung wieder abgegeben werden, wobei ein Platz mit gleicher oder geringerer Energie als zunächst (vorher) im Valenzband eingenommen werden kann. Aufgrund der Wellennatur können Elektronen jedoch einen Potentialwall der Höhe $\Delta\varphi$ und der Längsausdehnung d mit der Wahrscheinlichkeit

$$f_T(W) = \exp\left(-\frac{4\pi}{h}\left(2m_e^* e\Delta\varphi\right)^{\frac{1}{2}} d\right) \tag{3.180}$$

durchqueren ohne vorher Energie aufzunehmen, um den Potentialwall zu überwinden. In Analogie an das Durchqueren eines Berges in einem Tunnel anstelle des Aufnehmens und Abgebens potentieller Energie beim Überfahren des Passes nennt man diesen Vorgang auch tunneln. Die Wahrscheinlichkeit des Tunnelns ist umso größer, je niedriger und je dünner der Potentialwall ist. Bei ortsabhängigem Potential ist die Tunnelwahrscheinlichkeit gegeben durch

$$f_T(W) = \exp\left(-\frac{4\pi}{h}\int\limits_{x_1}^{x_2}\left(2m_e^*\, e\Delta\varphi(x)\right)^{\frac{1}{2}} dx\right). \tag{3.181}$$

Mit $\Delta\varphi(x) = e\,\varphi(x) - W$, wobei W das Energieniveau ist bei dem der äquienergetische Tunnelübergang stattfindet (vgl. Abb. 3.25). Die Inegrationsgrenzen ergeben sich aus der Lage der Bandkanten bei der Energie W. Durch das Tunneln fließt ein Strom vom Leitungsband zum Valenzband (I_{CV}) und vom Valenzband zum Leitungsband (I_{VC}). Diese Ströme sind jeweils proportional der Zahl der besetzten Zustände auf der Ausgangsseite und der Zahl

der unbesetzten Zustände auf der Zielseite. Mit $f_C(W)$ ist die Besetzungswahrscheinlichkeit eines Zustandes im Leitungsband und mit $f_V(W)$ die eines Zustandes im Valenzband bezeichnet, wobei zu beachten ist, dass in $f_C(W)$ und $f_V(W)$ unterschiedliche Quasi-Fermi-Niveaus einzusetzen sind, nämlich W_{Fn} in $f_C(W)$ und W_{Fp} in $f_V(W)$. $D_C(W)$ und $D_V(W)$ sind die Zustandsdichten des Leitungsbandes und des Valenzbandes. A_T ist eine kompliziert zusammengesetzte Konstante, die eA als Faktor enthält. Die Stromanteile ergeben sich durch Integration von der Leitungsbandkante auf der n-Seite bis zur Valenzbandkante auf der p-Seite, da innerhalb dieses Energiebereichs äquienergetische Durchquerungen der Energiebarriere möglich sind.

$$I_{CV} = A_T \int_{W_C(x_n)}^{W_V(x_p)} f_T(W)\, D_C(W)\, f_C(W)\, D_V(W)\, (1 - f_V(W))\, dW, \quad (3.182)$$

$$I_{VC} = A_T \int_{W_C(x_n)}^{W_V(x_p)} f_T(W)\, D_V(W)\, f_V(W)\, D_C(W)\, (1 - f_C(W))\, dW. \quad (3.183)$$

Der Tunnelstrom ist die Differenz beider Ströme:

$$I_T = A_T \int_{W_C(x_n)}^{W_V(x_p)} f_T(W)\, D_C(W)\, D_V(W)\, (f_C(W) - f_V(W))\, dW. \quad (3.184)$$

Wegen $W - W_{Fn} > 0$ und $W - W_{Fp} < 0$ gilt $f_C(W) - f_V(W) < 0$, so dass $I_T < 0$ ist.

Unterscheidung zwischen Lawinen- und Zenerdurchbruch:

Ob der Einsatz des Durchbruchs auf Stoßionisation oder den Beginn der Bandüberlappung, also auf den Zener-Effekt zurückzuführen ist, lässt sich aus dem unterschiedlichen Temperaturverhalten entscheiden. Die Stoßionisation wird durch Temperaturerhöhung behindert, da dabei die mittlere freie Weglänge der Ladungsträger abnimmt. Der Durchbruch setzt also erst bei höherer Sperrspannung ein.
Dagegen wird die Tunnelwahrscheinlichkeit durch Temperaturerhöhung vergrößert, da der Bandabstand mit der Temperatur abnimmt. Genauer gilt nämlich:

- für **Si**: $E_G(T) = (1.21 - 3.60 \cdot 10^{-4} \cdot T)\,\mathrm{eV}$,
- für **Ge**: $E_G(T) = (0.785 - 2.23 \cdot 10^{-4} \cdot T)\,\mathrm{eV}$.

4 Bipolar-Transistor

4.1 Prinzipieller Aufbau und Definition

Der Bipolar-Transistor wurde 1948 von John Bardeen und Walter Brattain entdeckt und zusammen mit William Shockley weiterentwickelt. Sie erhielten 1956 fr ihre Forschungsarbeiten und die Entdeckung des Transistor-Effekts den Nobel-Preis in Physik. Als Halbleitermaterial diente damals *Ge*, da mit den damals realisierbaren großen Basisweiten nur *Ge* eine ausreichend große Diffusionslänge aufwies um den bipolaren Transistoreffekt zu ermöglichen. Warum die nötig ist, werden wir im Folgenden verstehen.
Je nach Anwendung gibt es zahlreiche Arten Bipolar-Transistoren die sich über Strukturfolgen und Technologieschritte unterscheiden. Allen gemeinsam ist, dass der eigentliche Transistor sich auf eine einfach *npn*- bzw. *pnp*-Folge *n*- bzw. *p*-dotierter Halbleiterschichten reduzieren läßt. Die Wirkung der Umgebung (peripherer Transistor) kann durch zusätzliche Elemente im Ersatzschaltbild oder durch Korrekturfaktoren zu den Parametern des eigentlichen Transistors berücksichtigt werden.
Wir beschäftigen uns im Rahmen dieser Vorlesung nur mit dem eigentlichen Transistor und meinen, wenn wir im Folgenden von Transistor sprechen immer den eigentlichen Transistor. Um das Wesentliche in den Transistoreigenschaften herauszustellen, werden wir den (eigentlichen) Transistor als nahezu ideales Bauelement mit entsprechenden Vereinfachungen für die Standardanwendung beschreiben. Aufgrund seiner, speziell in integrierten Schaltungen, höheren Verbreitung, die aus seiner im Vergleich zum *pnp*-Transistor höheren Geschwindigkeit resultiert, werden wir uns ausschließlich mit dem *npn*-Bipolar-Transistor beschäftigen. Alle Ergebnisse sind jedoch uneingeschränkt auch für den *pnp*-Typ gültig, wenn die Polarität, der an den Transistor angelegten Spannungen im Arbeitspunkt umgekehrt wird.
Abb. 4.1 zeigt die Schaltungssymbole mit den, für den normal-aktiven Betriebsfall notwendigen Vorspannungen (Arbeitspunkt).
Zur Eindeutigkeit der Spannungsrichtungen wird üblicherweise (bitte nicht darauf verlassen, da Abweichungen möglich) die Reihenfolge der Indizees in Pfeilrichtung genannt. Danach ist U_{BE} eine positive Spannung von der Basis zum Emitter gepolt. $|U_{BE}|$ liegt in der Größenordnung der Diffusionsspannung U_D, ist aber in jedem Fall kleiner als U_D. Die Transistoranschlüsse werden mit

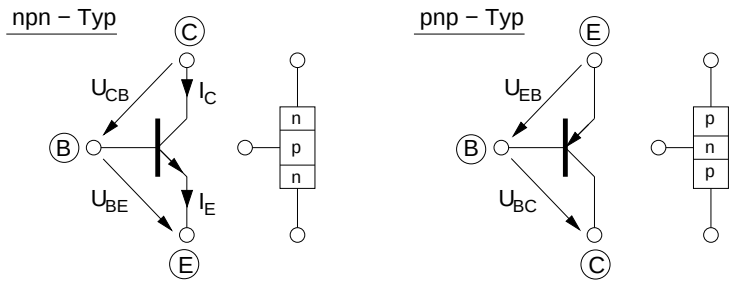

Abb. 4.1: Schaltungssymbole des Bipolar-Transistors mit Schichtenfolge für *npn*-Typ (links) und *pnp*-Typ (rechts).

E: Emitter (engl: to emit = aussenden)
B: Basis (engl: base = Basis, da dieser Anschluß bei den ersten Prototypen die Grundplatte (Basis) war)
C: Kollektor (engl: to collect = einsammeln)

bezeichnet. Die Richtung des Emitterpfeils lässt sich leicht merken, berücksichtigt man die Schichtfolge. Beim *npn*-Typ zeigt der Emitterpfeil von der Basis zum Emitter, also vom *p*- in das *n*-Gebiet. Das ist die gleiche Richtung, den der Pfeil einer *p*-*n*-Diode hatte. Man stellt sich daher immer für die Schichtfolge die entsprechende Richtung des Diodenpfeils vor. Neben dem normal-aktiven Betriebsbereich gibt es noch eine Reihe anderer Betriebszustände des Transistors, die in Tab. 4.1 zusammengefasst sind. Sie ergeben sich durch eine entsprechende Einstellung der Spannungen und/oder Ströme am Transistor, die der Schaltungsentwickler vornehmen muss. Die Spannungen und Ströme, die den Transistor in den jeweiligen Betriebsbereich bringen, bezeichnen wir in ihrer Gesamtheit als Arbeitspunkt.

Wir konzentrieren uns im Folgenden ausschließlich auf den *npn*-Typ, im normal-aktiven Zustand. Abb. 4.2 zeigt schematisch die entsprechende Schichtfolge mit einer äußeren Arbeitspunkt-Beschaltung für den normal-aktiven Betrieb.
Die Dicke der Schichten bezeichnen wir näherungsweise mit w_e, w_b und w_c wobei bei genauerer Betrachtung von den Weiten noch die Raumladungsweiten abgezogen werden müssen (vgl. z.B. Abb. 4.4). Es sei vorweggenommen, dass die Darstellung der Weite in Abb. 4.2 nicht maßstäblich ist. Insbesondere die Kollektorweite w_c ist in der Regel viel größer als w_b. Die Emitterweite

Betriebszustand	BE-Diode	BC-Diode	Anwendung
normal aktiv	leitend	gesperrt	Verstärker, Schalter, Logik
invers aktiv	gesperrt	leitend	-
normal gesättigt	leitend	schwach leitend (Flusspolung)	Low Power Elektronik, Gesättigte Logik
invers gesättigt	schwach leitend (Flusspolung)	leitend	-
gesperrt			Kapazitätsdioden, $C(U)$

Tabelle 4.1: Einteilung der verschiedenen Betriebszustände des Bipolar-Transistors.

ist bei integrierten Bipolar-Transistoren meist in der gleichen Größenordnung wie w_b, dies ist jedoch für den Transistoreffekt nicht unbedingt erforderlich.

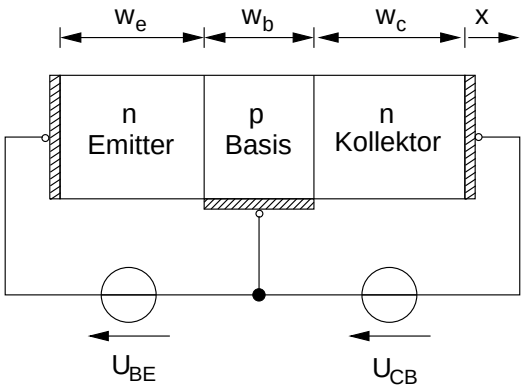

Abb. 4.2: Prinzipieller Aufbau eines *npn*-Transistors mit einer Arbeitspunkt-Beschaltung für den normal-aktiven Bereich ($U_{BE} \approx U_D$, $U_{CB} \geq 0$).

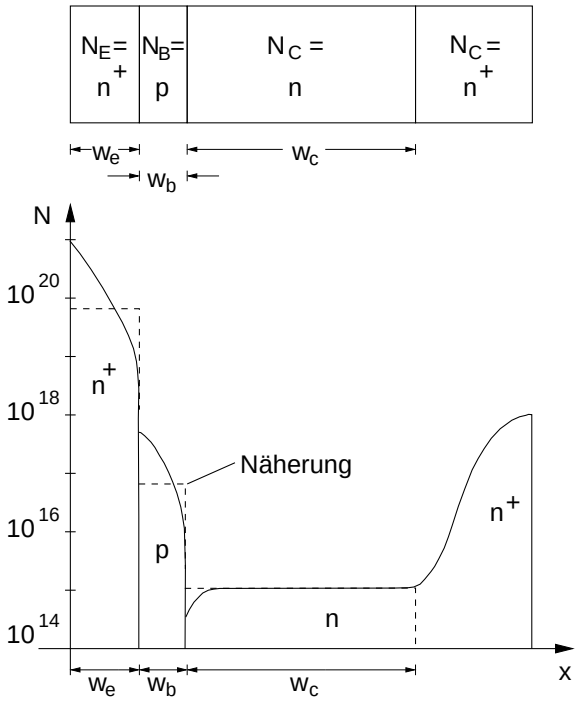

Abb. 4.3: Querschnitt durch den eigentlichen Transistor mit entsprechendem Dotierungsprofil.

Abb. 4.3 zeigt ein reales Dotierungsprofil der drei Bereiche für das Beispiel eines Hochfrequenztransistors für den GHz-Bereich. Wir nehmen zur Vereinfachung eine homogene Dotierung der drei Bereiche mit

$$N_E(x), \text{Donatordotierung} = \text{const. in } w_e, \text{ sonst} = 0$$
$$N_B(x), \text{Akzeptordotierung} = \text{const. in } w_b, \text{ sonst} = 0$$
$$N_C(x), \text{Donatordotierung} = \text{const. in } w_c, \text{ sonst} = 0$$

an. Weiterhin nehmen wir, wie bei der p-n-Diode, Störstellenerschöpfung und Rechteckprofil-Näherung für die ortsfeste Raumladung an den beiden p-n-Übergängen an.

4.2 Ladungsträgerdichten in einer npn-Schichtfolge

Um das prinzipielle Verhalten der npn-Schichtfolge zu verstehen, analysieren wir in einzelnen Schritten die Verhältnisse in den einzelnen Bahngebieten und Raumladungszonen. Wir benötigen dazu nur unser bereits für den p-n-

Übergang hergeleitetes Wissen. Die einzelnen Schritte werden anhand von Abb. 4.4 illustriert.

1. Durch die Flusspolung der Basis-Emitter-Diode wird das Gleichgewicht zwischen Diffusions- und Driftstrom in der Basis-Emitter-Raumladungszone zugunsten des Diffusionsstroms verändert. Als Resultat diffundieren Löcher aus der Basis-Emitter-Raumladungszone in den Emitter und Elektronen in die Basis. Die Randkonzentrationen dieser Minoritätsträger sind entsprechend den Boltzmann-Randbedingungen (Gl. (3.41) und (3.42)) um den Faktor $e^{\frac{U_{BE}}{U_T}}$ gegenüber den Gleichgewichtsdichten erhöht. Abb. 4.4b zeigt diese Verhältnisse.

2. Aufgrund der Flusspolung sind die Randkonzentrationen gegenüber der jeweiligen Gleichgewichtsdichte in den Bahngebieten erhöht. Hierdurch stellt sich ein Diffusionsstrom der Minoritätsträger in die Bahngebiete ein (vgl. Richtung der eingezeichneten Ladungsträger in Abb. 4.4b).

3. Die Minoritätsträgerkonzentration in den Bahngebieten können wir wie bei der p-n-Diode ermitteln. Als Randbedingung hatten wir bei der p-n-Diode angenommen, dass an den Enden der Diffusionsstrecke (in Abb. 4.4b bei $x = 0$ für die Diffusion im Emitter-Bahngebiet) die Ladungsträgerdichten die Gleichgewichtsdichte annehmen.
   Dies ist aufgrund einer beliebig hohen Rekombinationsrate für den Emitterkontakt bei $x = 0$ gewährleistet. Für das Ende der Diffusionsstrecke in der Basis (bei x_c) ist dies für $U_{BC} = 0$ erfüllt. Diese Randbedingung ist mathematisch äquivalent mit einem idealen Kontakt bei x_c. Jedoch besteht physikalisch der Unterschied, dass die bei x_c am Basis-seitigen Rand der BC-RLZ ankommenden Minoritätsträger nicht rekombinieren sondern durch die BC-RLZ driften und auf dem Kollektor-seitigen Rand als Majoritätsträger austreten.

4. Wir betrachten exemplarisch die Minoritätsträgerdichte in der Basis. Für den Verlauf der Minoritätsträgerdichte im Emitter, sowie die im weiteren Verlauf über eine Spannung $U_{CB} \neq 0$ in Kollektor und Basis verursachten zusätzlichen Minoritätsträgerströme, gelten die gleichen Überlegungen.
   Nach Gl. (3.82) gilt für die Minoritätsträgerdichte in der Basis mit

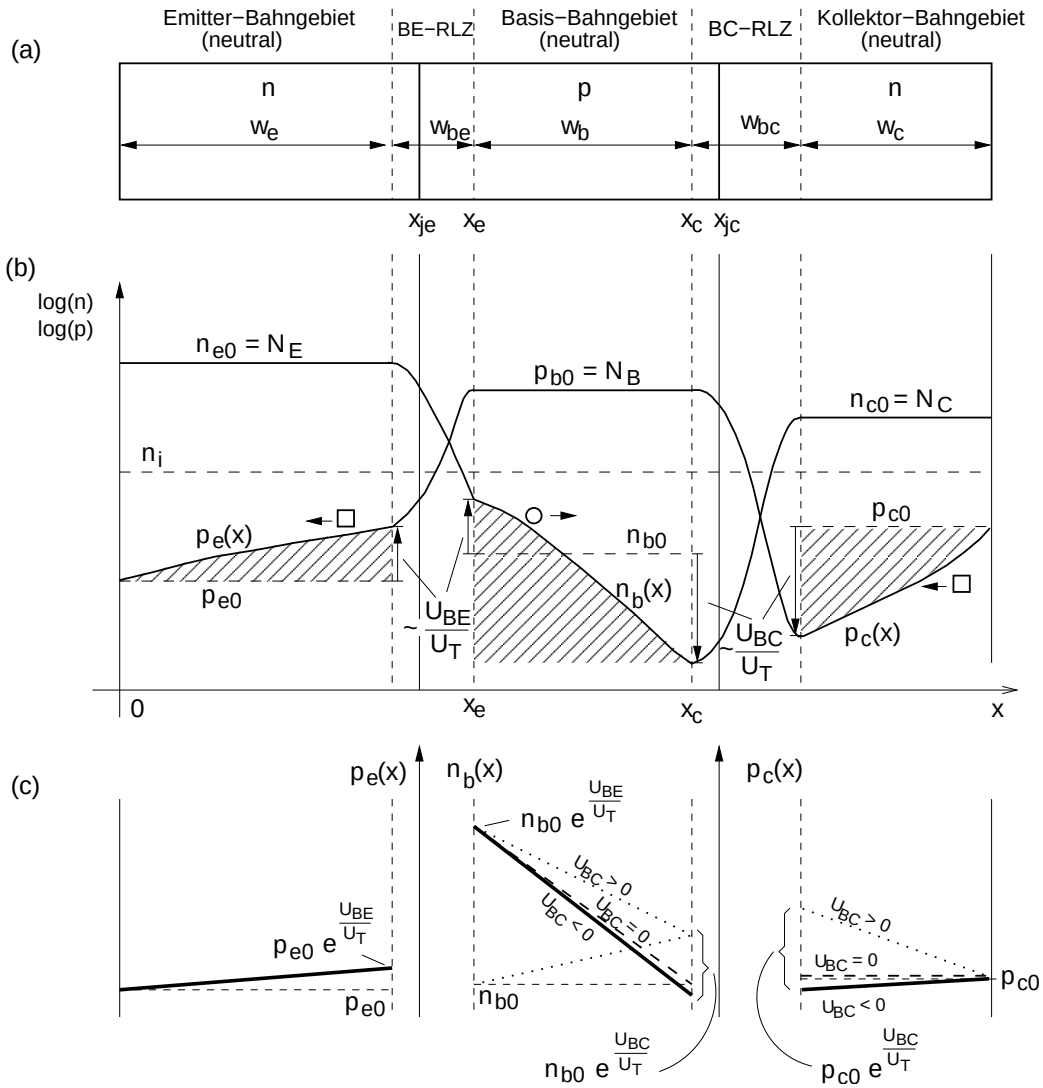

Abb. 4.4: Ladungsträgerdichten in einem *npn*-Bipolartransistor

a) Definition der Raumladungszonen (RLZ) und Bahngebiete. Zur Beachtung: Die Spannungsabhängigkeit der BC-RLZ ist in der Darstellung nicht berücksichtigt.

b) Ladungsträgerdichten bei normal-aktivem Betrieb (BC-Diode in Sperrpolung)

c) Minoritätsträgerverteilung in den Bahngebieten bei Geradennäherung für kurze Bahngebiete.

$x_e \leq x \leq x_c$ bei $U_{BC} = 0$

$$n_b(x) \;=\; n_{b0} \;+\; n_{b0} \left(e^{\frac{U_{BE}}{U_T}} - 1 \right) \frac{\sinh\left(\frac{x_c - x}{L_{nb}}\right)}{\sinh\left(\frac{x_c - x_e}{L_{nb}}\right)} \tag{4.1}$$

Darin kann x maximal x_c annehmen (Ende der Diffusionsstrecke). Das Argument des *sinh* kann damit maximal

$$\frac{x_c \;-\; x_e}{L_{nb}} \;=\; \frac{w_b}{L_{nb}} \tag{4.2}$$

betragen. Darin ist $w_b \;=\; x_c - x_e$ die Weite der Basis zwischen den Raumladungszonen-Rändern zu Emitter und Kollektor (vgl. Abb. 4.4a). Für eine dünne Basis mit $\frac{w_b}{L_n} \ll 1$ (entspricht Näherung für kurze Diode) gilt dann mit der Näherung $\sinh(x) \approx x$ für $x \ll 1$

$$n_b(x) \;=\; n_{b0} + n_{b0}\,(e^{\frac{U_{BE}}{U_T}} - 1)\,\frac{x_c - x}{w_b} \tag{4.3}$$

D.h. die Ladungsträgerdichte fällt linear von dem, über die Boltzmann-Randbedingung eingestellten Maximalwert von dem Rand der in Flussrichtung gepolten Raumladungszone ab, auf die Gleichgewichtsdichte am Ende der Diffusionszone. Abb. 4.4c zeigt diese Geradennäherung in Form der dicken, gestrichelten, fallenden Kurve.
Ebenfalls dargestellt sind ein Fall für Vorspannung der Basis-Kollektordiode in Sperrpolung $U_{BC} < 0$ sowie für den normal gesättigten Betriebsfall mit $U_{BC} > 0$.

5. Für lange Bahngebiete (gilt niemals für die Basis) muss anstelle der Geradennäherung Gl. (4.1) oder die Näherung für die lange Diode verwendet werden, wodurch $\sinh(x) = \frac{1}{2}\,e^x$ gesetzt werden kann. Bei einem langen Bahngebiet (gilt unter Umständen für Emitter und Kollektor) ist das Bahngebiet viel länger als die Diffusionslänge der Minoritäten. D.h. die Minoritäten sind länger als ihre mittlere Lebensdauer in den Bahngebieten unterwegs. Sie rekombinieren daher mit den Majoritäten. Durch die Rekombination geht der Minoritätsträgerstrom in einen Majoritätsträgerstrom über. Die Summe beider Ströme über dem Ort bleibt wegen der Aussage der Kontinuitätsgleichung konstant.

6. In einem kurzen Bahngebiet (gilt immer für die Basis) verweilen die Minoritäten nur kurze Zeit, verglichen mit ihrer (mittleren) Lebensdauer. Es kommt daher nur zu einer geringen (fr die Basis im Idealfall

vernachlässigbaren) Rekombination, wodurch am Ende der Diffusionsstrecke (Unterschied zu langem Bahngebiet) noch immer ein Strom aus Minoritätsträgern vorliegt. Eine Netto-Rekombination findet in diesem Fall erst an dem Kontakt statt. Fr die ber die Basis-Emitter-RLZ in die Basis injizierten Minoritätsträger ist das der Kollektor-Kontakt an der äußersten rechten Seite.

7. Wir nehmen im Folgenden zur Vereinfachung kurze Bahngebiete in Emitter, Basis und Kollektor an, so dass die Minoritätsträgerkonzentrationen näherungsweise durch die in Abb. 4.4c gezeigten linearen Verläufe angenähert werden können. Es kann dann anstelle von Abb. 4.4b mit der linearen Approximation der Minoritätsträgerverläufe in 4.4c gearbeitet werden, wodurch die Ermittlung der Steigung der Minoritätsträgerverläufe sehr einfach wird. Aufgrund ihrer Form wird der Verlauf der Minoritätsträgerkonzentration als Diffusionsdreieck bezeichnet.

4.3 Ströme in der *npn*-Schichtfolge ohne Rekombination

Wie bei der *p*-*n*-Diode ermitteln wir die Ströme in der *npn*-Schichtfolge über die Diffusionsströme der Minoritätsträger an den Grenzen der Bahngebiete zur jeweiligen Raumladungszone. Wir verwenden dabei die beiden schon bei der *p*-*n*-Diode angewandten Annahmen.

1. Die Feldstärke in den Bahngebieten ist so gering, dass die Minoritätsträger-Feldströme wegen der geringen Minoritätsträger-Konzentration vernachlässigbar sind. Es existieren daher nur Minoritätsträger-Diffusionsströme.

2. Der Netto-Rekombinationsstrom in der Raumladungszone überlagert sich den anderen Stromkomponenten. Er kann und wird daher zur Vereinfachung für die folgenden Betrachtungen zu Null gesetzt werden. Seine Wirkung kann im Nachhinein durch das Hinzufügen eines Netto-Rekombinationsstrom-Terms in den Gleichungen bzw. eines zusätzlichen Bauelements in den Ersatzschaltbildern berücksichtigt werden.

Wegen der zweiten Annahme ermitteln wir die Ströme in den Schichten für den Fall, dass keine Netto-Rekombination der beiden Raumladungs-

zonen stattfindet. Ein Elektronen- oder Löcherstrom bleibt daher in den Raumladungszonen konstant. Er hat daher an der Grenze zur Raumladungszone, wo er ein Majoritätsträgerstrom ist, den gleichen Wert, wie an der anderen Grenze der Raumladungszone, wo er wegen Annahme 1) ein Minoritätsträger-Diffusionsstrom ist.
Diese Überlegungen gelten unabhängig von der Länge der Bahngebiete. Die Minoritätsträger-Diffusionsströme ergeben sich allgemein aus der Transportgleichung.

An den Grenzen der Basis-Emitter-Raumladungszone ergeben sich die Diffusionsströme zu:

$$I_{ep}(x_e - w_{be}) \;=\; -e\,A\,D_p\;\left.\frac{d\,p_e(x)}{d\,x}\right|_{x_e - w_{be}} \tag{4.4}$$

$$I_{bn}(x_e) \;=\; e\,A\,D_n\;\left.\frac{d\,n_b(x)}{d\,x}\right|_{x_e} . \tag{4.5}$$

An den Grenzen der Basis-Kollektor-Raumladungszone gilt:

$$I_{bn}(x_c) \;=\; e\,A\,D_n\;\left.\frac{d\,n_b(x)}{d\,x}\right|_{x_c} \tag{4.6}$$

$$I_{cp}(x_c + w_{bc}) \;=\; -e\,A\,D_p\;\left.\frac{d\,p_c(x)}{d\,x}\right|_{x_c + w_{bc}} . \tag{4.7}$$

Im allgemeinen Fall muss für die Ableitung der Ladungsträgerverteilung die entsprechende Beziehung wie z.B. in Gl. (4.1) für $n_b(x)$ bei $U_{BC} = 0$ eingesetzt werden. Für $I_{bn}(x) = e\,A\,D_n \frac{d\,n_b}{d\,x}$ ergibt sich dann direkt der bereits für die p-n-Diode bei der Randbedingung $n_b(x_c) = n_{b0}$ also für $U_{BC} = 0$ ermittelte Stromverlauf

$$I_{bn}(x) \;=\; -e\,A\,D_n\,n_{b0}\,(e^{\frac{U_{BE}}{U_T}} - 1)\frac{1}{L_n}\frac{\cosh\left(\frac{x_c - x}{L_{nb}}\right)}{\sinh\left(\frac{w_b}{L_{nb}}\right)} \tag{4.8}$$

wodurch sich für $x = x_e$ und $x = x_c$ die gesuchten Werte nach Gl. (4.5) und Gl. (4.6) ergeben. Für unsere vereinfachende Annahme kurzer Bahngebiete können wir die Steigung einfach an den Diffusionsdreiecken in Abb. 4.4 ablesen. Es ergibt sich unter Berücksichtigung der Boltzmann-Randbedingung für den Kollektor-seitigen Rand der Basis. An den Grenzen der Basis-Emitter-

Raumladungszone:

$$I_{ep}(x_e - w_{be}) \;=\; -e\,A\,D_p\,\frac{p_{e0}(e^{\frac{U_{BE}}{U_T}} - 1)}{w_e} \tag{4.9}$$

$$I_{bn}(x_e) \;=\; -\,e\,A\,D_n\,\frac{n_{b0}(e^{\frac{U_{BE}}{U_T}} - e^{\frac{U_{BC}}{U_T}})}{w_b}\;. \tag{4.10}$$

An den Grenzen der Basis-Kollektor-Raumladungszone:

$$I_{bn}(x_c) \;=\; I_{bn}(x_e) \tag{4.11}$$

$$I_{cp}(x_c + w_{bc}) \;=\; e\,A\,D_p\,\frac{p_{c0}(e^{\frac{U_{BC}}{U_T}} - 1)}{w_c}\;. \tag{4.12}$$

Dies sind die gleichen Ergebnisse, wie sie zuvor auch für die kurze p-n-Diode erhalten wurden. Wir definieren zur Abkürzung der Schreibweise die Sättigungsströme

$$I_{es} \;:=\; e\,A\,D_p\,\frac{p_{e0}}{w_e} \tag{4.13}$$

$$I_{bs} \;:=\; e\,A\,D_n\,\frac{n_{b0}}{w_b} \tag{4.14}$$

$$I_{cs} \;:=\; e\,A\,D_p\,\frac{p_{c0}}{w_c}\;. \tag{4.15}$$

Aufgrund des linearen Verlaufs der Minoritätsträgerkonzentration in den Diffusionsdreiecken sind die Diffusionsströme in ihrem gesamten Bahngebiet konstant. Wir benötigen sie für die weitere Rechnung zwar nur an den Rändern der Raumladungszonen, lassen aber aufgrund der Konstanz die Angabe der Ortsabhängigkeit im Folgenden weg. Mit den zuvor definierten Sättigungsströmen lassen sich Gl. (4.9) bis (4.12) schreiben

$$I_{ep} \;=\; -I_{es}(e^{\frac{U_{BE}}{U_T}} - 1) \tag{4.16}$$

$$I_{bn} \;=\; -\underbrace{I_{bs}(e^{\frac{U_{BE}}{U_T}} - 1)}_{I_{ben}} + \underbrace{I_{bs}(e^{\frac{U_{BC}}{U_T}} - 1)}_{I_{bcn}} \;=\; I_{bcn} \,-\, I_{ben} \tag{4.17}$$

$$I_{cp} \;=\; I_{cs}(e^{\frac{U_{BC}}{U_T}} - 1)\;. \tag{4.18}$$

Der Minoritätsträgerstrom in der Basis wurde in Gl. (4.17) so umgeformt (die -1 in den beiden Klammern hebt sich weg), dass er sich aus einem von U_{BE} bei $U_{BC} = 0$ gesteuerten Anteil I_{ben} und einem von U_{BC} bei $U_{BE} = 0$

I_{bcn} gesteuerten Anteil zusammensetzt. Jeder der beiden Terme steht für ein eigenes Diffusionsdreieck in der Basis, an dessen Ende die Gleichgewichtskonzentration n_{b0} erreicht wird. Abb. 4.4c zeigt dies fr den Fall, dass das sich fr $U_{BC} > 0$ ergebende Viereck in das gepunktete Dreieck mit der kollektorseitigen Konzentration $n_{b0}e^{\frac{U_{BC}}{U_T}}$ und das gestrichelte Dreieck mit der emitterseitigen Konzentration $n_{b0}e^{\frac{U_{BE}}{U_T}}$ zerlegt werden kann.
Entsprechend der mathematischen Vorzeichendefinition im DDM sind Ströme mit positivem Vorzeichen in Richtung der positiven x-Achse gerichtet. Abb. 4.5 links zeigt die dementsprechenden Ströme aus Gl. (4.16) bis (4.18) an den Rändern der sie steuernden Raumladungszone. Im Folgenden

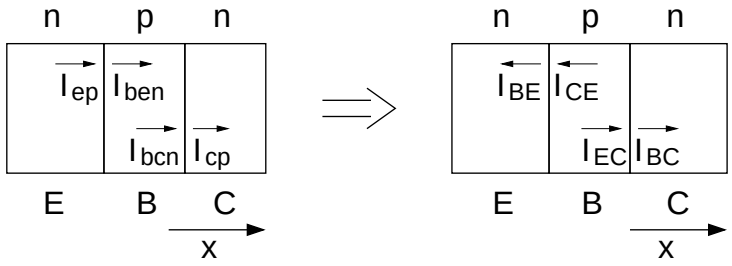

Abb. 4.5: Links: Über die Minoritätsträger-Injektion von Basis-Emitter- und Basis-Kollektor- Raumladungszone gesteuerte Minoritätsträgerströme in den Bahngebieten von Emitter, Basis und Kollektor. Rechts: Die selben Ströme wie links, jedoch mit physikalisch orientierte Stromflussrichtung.

werden wir anstelle der mathematischen Richtung $(+x)$ die Ströme physikalisch positiv orientieren. Das bedeutet, dass die von der Spannung über eine der beiden Raumladungszonen gesteuerten Ströme positiv gezählt werden, wenn diese Raumladungszone in Flussrichtung gepolt ist. Wir definieren daher die auch in der Literatur üblichen Ströme

$$I_{BE} := -I_{ep} = I_{es}\,(e^{\frac{U_{BE}}{U_T}} - 1) \tag{4.19}$$

$$I_{CE} := -I_{ben} = I_{bs}\,(e^{\frac{U_{BE}}{U_T}} - 1) \tag{4.20}$$

$$I_{EC} := I_{bcn} = I_{bs}\,(e^{\frac{U_{BC}}{U_T}} - 1) \tag{4.21}$$

$$I_{BC} := I_{cp} = I_{cs}\,(e^{\frac{U_{BC}}{U_T}} - 1) \tag{4.22}$$

die in Abb. 4.5 rechts dargestellt sind. Der Stromfluss (Transfer) vom Kollektor zum Emitter erfolgt über die Basis mit den beiden Strömen I_{CE} und

I_{EC}. Es ist daher üblich diesen Strom als Transferstrom I_T mit einem Transfersättigungsstrom I_S zu definieren

$$I_T = I_{CE} - I_{EC} = I_S\,(e^{\frac{U_{BE}}{U_T}} - e^{\frac{U_{BC}}{U_T}}) \approx I_S\,e^{\frac{U_{BE}}{U_T}} \approx I_{CE}\ . \tag{4.23}$$

Mit

$$I_S := I_{bs} = e\,A\,D_n\,\frac{n_{b0}}{w_b} \tag{4.24}$$

aus Gl. (4.14). Im normal-aktiven Betrieb ist die Basis-Emitter-Strecke leitend und die Basis-Kollektor-Strecke gesperrt. Dann gilt die Näherung in Gl. (4.23).

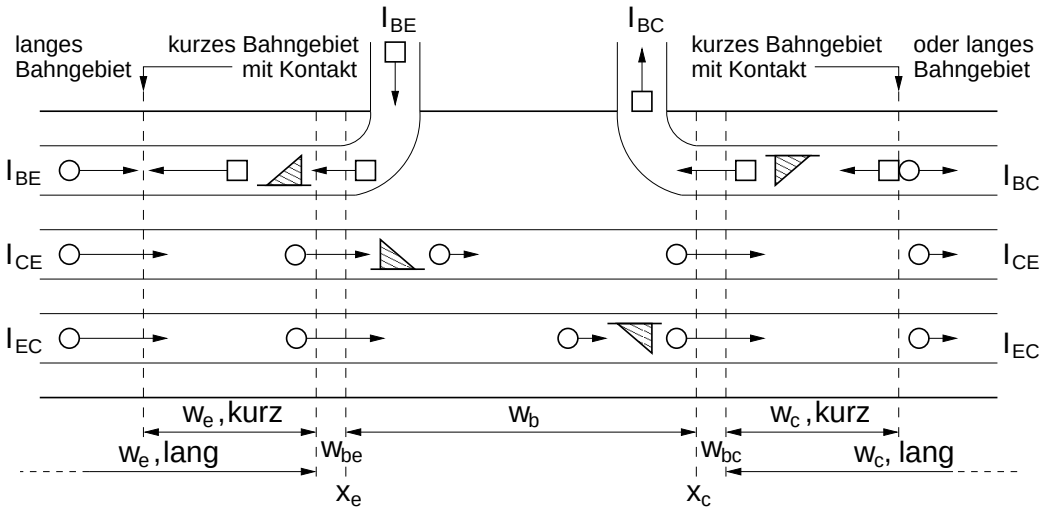

Abb. 4.6: Ladungsträgerflüsse und Ströme in einem *npn*-Transistor im normal-aktiven Betrieb ohne Rekombination in der Basis und ohne Netto-Rekombination in den Raumladungszone.

Abb. 4.6 zeigt den zu den Minoritätsträgerströmen gehörenden Ladungsträgerfluss für einen Transistor im normal-aktiven Betrieb. Die darin eingezeichneten Dreiecke deuten die zuvor besprochenen Diffusionsdreiecke an. Der verlängerte Strich an den Diffusionsdreiecken steht für die Gleichgewichtskonzentration. Entsprechend der Polung der beiden p-n-Übergänge zeigen die beiden Basis-Emitter-Dreiecke einen Ladungsträgerüberschuss, die beiden Basis-Kollektor-Diffusionsdreiecke zeigen einen Ladungsträgermangel gegenüber der Gleichgewichtskonzentration. Das Bild unterscheidet zwischen kurzen Emitter- und Kollektor-Bahngebieten, in denen die Minoritäten bis

zum Kontakt gelangen ehe sie dort rekombinieren, und langen Bahngebieten, in denen über eine Netto-Rekombination der Minoritätsträgerfluss in einen Majoritätsträgerfluss übergeht. Die Netto-Rekombination im Emitter-Bahngebiet ist wegen des Ladungsträgerüberschusses eine Rekombination, die im Kollektor-Bahngebiet aufgrund des Ladungsträgermangels (Sperrpolung) eine Generation.

4.4 Einfaches Großsignalmodell des Transistors

Über die physikalische Zuordnung der einzelnen Ladungsträgerströme in Abb. 4.6 zu einzelnen Bauelementen kann das in Abb. 4.7 gezeigte elektrische Ersatzschaltbild des *npn*-Transistors gezeichnet werden. Das Ersatzschaltbild

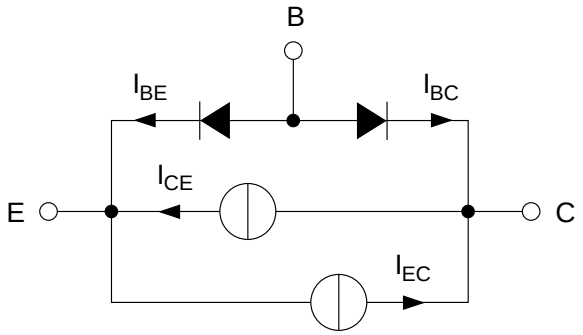

Abb. 4.7: Elektrisches Ersatzschaltbild (Großsignalmodell) eines *npn*-Transistors mit den Modellgleichungen (4.19) bis (4.22).

gilt für alle Betriebsbereiche des Transistors, da hinsichtlich der darin enthaltenen Bauelemente noch keine Näherungen eingeführt wurden. Aufgrund seiner Gültigkeit über große Spannungsbereiche (und damit auch Strombereiche) wird es auch als Großsignalmodell bezeichnet.
I_{BE} und I_{CE} sind in Abb. 4.7 über die Spannung U_{BE} voneinander abhängig. Entsprechend sind I_{BC} und I_{EC} über U_{BC} voneinander abhängig.
Die Basis soll der Eingang des Transistors sein, der mit wenig Strom einen größeren Strom zwischen Emitter und Kollektor steuert. Wir definieren daher für das Verhältnis der beiden jeweils von einer Spannung abhängigen Ströme die Vorwärtsstromverstärkung (Gl. (4.19), (4.20), (4.16), (4.17))

$$B_F := \frac{I_{CE}}{I_{BE}} = \frac{I_{bs}}{I_{es}} = \frac{D_n}{D_p}\frac{n_{b0}\,w_e}{p_{e0}\,w_b} \tag{4.25}$$

und die Rückwärtsstromverstärkung (Gl. (4.21), (4.22), (4.17), (4.18))

$$B_R := \frac{I_{EC}}{I_{BC}} = \frac{I_{bs}}{I_{cs}} = \frac{D_n}{D_p}\frac{n_{b0}\,w_c}{p_{c0}\,w_b} \quad . \tag{4.26}$$

Ihre Namen erklären sich anschaulich dadurch, dass im normal-aktiven Bereich die Basis-Emitter-Diode in Flussrichtung (vorwärts, „forward“) gepolt ist und die Basis-Kollektor-Diode gesperrt ist. Dann ist B_F das Verhältnis von dem in die Basis fließenden Steuerstrom und dem dadurch gesteuerten Strom, der über den Transistorausgang (Kollektor-Emitter-Strecke) fließt. Entsprechendes gilt für die Rückwärtsstromverstärkung bei gesperrter Basis-Emitter- und leitender Basis-Kollektor-Diode.
Damit ergeben sich die in Abb. 4.8 gezeigten Großsignalmodelle bei denen zusätzlich die beiden Stromquellen I_{EC}, I_{CE} entsprechend Gl. (4.23) durch eine Transferstromquelle I_T ersetzt worden sind. Beide Modell sind identisch,

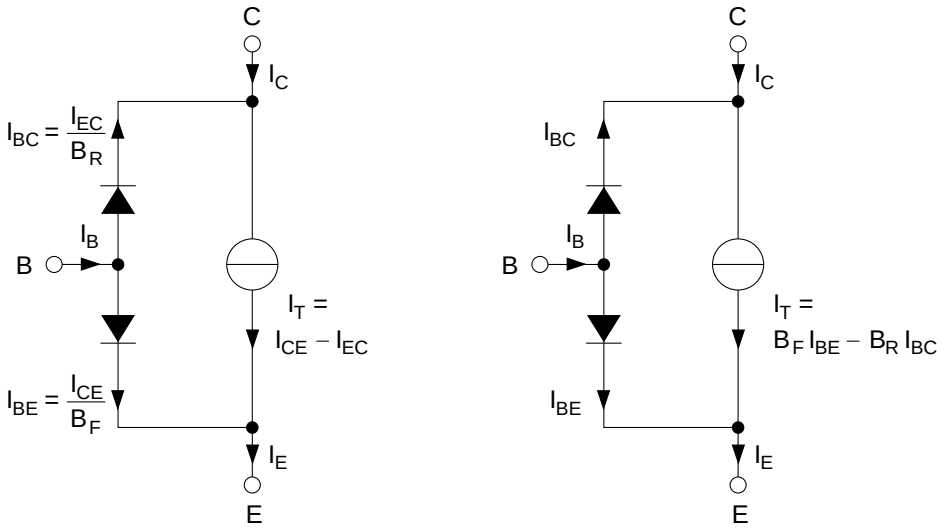

Abb. 4.8: Transferstrom-Großsignalmodelle des *npn*-Transistors. Links: Modell nach Gummel-Poon, das in SPICE Verwendung findet. Rechts: Identische Variante, bei der der Transferstrom über die steuernden Ströme ausgedrückt wird.

besitzen jedoch aufgrund ihrer über die Stromverstärkung umgerechneten Ströme je nach Aufgabenstellung Vorteile. Die Variante links findet z.B. im Schaltungssimulator SPICE zusammen mit vielen Erweiterungen als Grundlage des Gummel-Poon-Modells Einsatz. Die rechte Variante ist von Vorteil bei der Schaltungsentwicklung, denn hier wird in der Regel die Ausgangsgröße in Abhängigkeit von der steuernden Eingangsgröße benötigt.

4.5 Stromverstärkung (quasistatisch)

Für den Einsatz in elektrischen Schaltungen sind nicht die einzelnen Ladungsträgerströme, sondern nur der in dem jeweiligen Transistoranschluss fließende Gesamtstrom von Bedeutung. Für die Verstärkung, die ein in die Basis fließender Strom zum Kollektor erfährt, wird für diesen Zweck die quasistatische Stromverstärkung

$$B := \frac{I_C}{I_B} \tag{4.27}$$

definiert.
Mit den aus Abb. 4.8 ablesbaren Zusammengängen kann B ausgedrückt werden als

$$B = \frac{I_C}{I_E - I_C} = \frac{I_T - I_{BC}}{I_T + I_{BE} - I_T + I_{BC}} = \frac{I_T - I_{BC}}{I_{BE} + I_{BC}} \tag{4.28}$$

und mit:

$$I_T = B_F\, I_{BE} - B_R\, I_{BC} \tag{4.29}$$

$$B = \frac{B_F\, I_{BE} - (1 + B_R) I_{BC}}{I_{BE} + I_{BC}}\,. \tag{4.30}$$

Aus dieser Beziehung lassen sich einige Dimensionierungsvorschriften für den Transistor ableiten, wenn eine hohe Stromverstärkung erwünscht ist. Im normal-aktiven Bereich gilt $I_{BE} \gg I_{BC}$ mit mehreren Zehnerpotenzen Unterschied. Für eine große Stromverstärkung ist daher in erster Linie eine große Vorwärtsstromverstärkung B_F notwendig. Dafür gilt in Kurzschreibweise die folgende Überlegung (vgl. Gl. (4.25))

$$B_F \uparrow \Rightarrow n_{b0} \uparrow,\, p_{e0} \downarrow,\, w_e \uparrow,\, w_b \downarrow \tag{4.31}$$

$$\Rightarrow \frac{ni^2}{p_{b0}} \uparrow,\, \frac{ni^2}{n_{e0}} \downarrow,\, w_e \uparrow,\, w_b \downarrow \tag{4.32}$$

$$\Rightarrow p_{b0} \downarrow,\, n_{e0} \uparrow,\, w_e \uparrow,\, w_b \downarrow \tag{4.33}$$

$$\Rightarrow N_B \downarrow,\, N_E \uparrow,\, w_e \uparrow,\, w_b \downarrow \tag{4.34}$$

Für eine hohe Vorwärtsstromverstärkung ist daher eine niedrige Basisdotierung und eine kurze Basisweite, sowie eine hohe Emitterdotierung und ein langer Emitter notwendig. Die Forderung des langen Emitters würde konsequenterweise ein Emitterbahngebiet mit $w_e \gg L_{pe}$ (Diffusionslänge der Löcher im Emitter) verlangen. In einigen Büchern wird daher der Bipolartransistor mit langem Emitter behandelt, wodurch w_e durch L_{pe} in Gl. (4.25) und (4.30) ersetzt werden muss.

Für die Rückwärtsstromverstärkung B_R ergeben sich aufgrund $I_{BC} \ll I_{BE}$ keine Forderungen um eine hohe Stromverstärkung B zu erhalten (vgl. (4.30)). Prinzipiell gelten jedoch für kleines B_R die entgegengesetzten Forderungen aus Gl. (4.31)-(4.34) wobei $E \to C$ zu ersetzen ist. Danach kann, ohne die Dimensionierung für $B_F \uparrow$ zu beeinträchtigen, $w_C \downarrow$ und $N_C \downarrow$ gefordert werden. Für einen so dimensionierten Transistor gilt die Näherung

$$B \approx B_F \; . \tag{4.35}$$

Typische Werte für B bei *npn*-**Si**-Bipolar-Transistoren liegen zwischen 100-300.

4.6 Ebers-Moll-Ersatzschaltbild

Neben dem Transferstrom-Großsignalmodell aus Abb. 4.8 wird häufig auch das Ebers-Moll-Ersatzschaltbild des Bipolar-Transistors verwendet. Es ist gleichwertig mit dem Transferstrom-Modell, wobei jedoch die darin enthaltenen Elemente sich nicht direkt den jeweiligen physikalischen Ursachen zuordnen lassen.
Das Ebers-Moll-Ersatzschaltbild kann durch die einfache, in Abb. 4.9 gezeigte Umformung aus dem Transferstrommodell gewonnen werden. Ziel der Umformung ist es, die eine, von I_{EC} und I_{CE} abhängige Transferstromquelle, durch zwei einzelne, jeweils nur von einem der beiden Ströme abhängige Stromquelle, umzuwandeln.

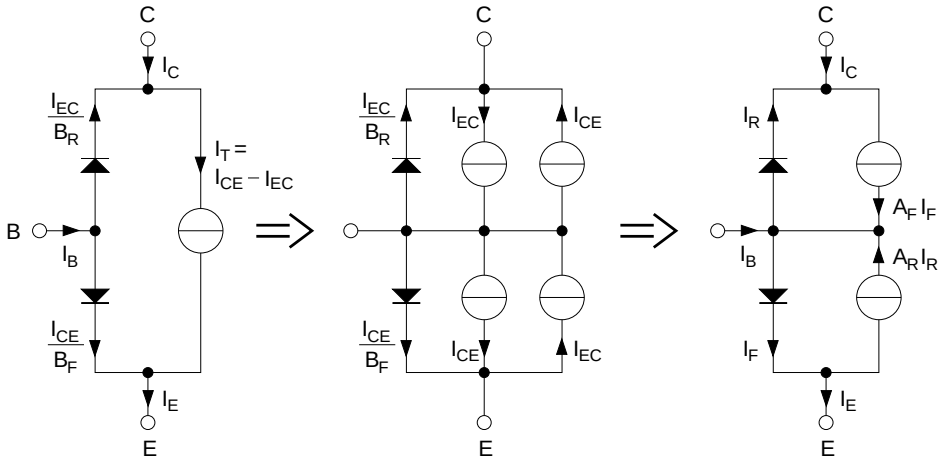

Abb. 4.9: Umwandlung des Transfermodells in das Ebers-Moll-Modell.

Die Umwandlung von dem linken zum mittleren Bild in Abb. 4.9 erfolgt rein auf zeichnerischer Basis ohne Rechnung. Dabei ist nur die Kirchhoffsche Knotenregel für den Basisanschluss zu berücksichtigen. Das Ebers-Moll-Ersatzschaltbild auf der rechten Seite geht durch Definition der Vorwärts-(Forward) und Rückwärts-(Reverse) Ströme

$$I_F := I_{CE} + \frac{I_{CE}}{B_F} = I_{CE}(1 + \frac{1}{B_F}) = \frac{I_{CE}}{A_F} \tag{4.36}$$

$$I_R := I_{EC} + \frac{I_{EC}}{B_R} = I_{EC}(1 + \frac{1}{B_R}) = \frac{I_{EC}}{A_R} \tag{4.37}$$

$$\text{mit} \quad A_F := \frac{B_F}{1 + B_F} \quad \text{und} \quad A_R := \frac{B_R}{1 + B_R} \tag{4.38}$$

aus der mittleren Abbildung hervor.
Der Strom-Spannungszusammenhang ergibt sich damit zu

$$I_F := I_{ES} \left(e^{\frac{U_{BE}}{U_T}} - 1 \right) \quad \text{mit} \quad I_{ES} = \frac{I_S}{A_F} \tag{4.39}$$

und

$$I_R := I_{CS} \left(e^{\frac{U_{BC}}{U_T}} - 1 \right) \quad \text{mit} \quad I_{CS} = \frac{I_S}{A_R} \tag{4.40}$$

worin I_S aus Gl. (4.25) stammt.

Für den normal-aktiven Betrieb ist $I_R \approx 0$ und die betreffende Diode und Stromquelle können aus dem Modell entfernt werden. Es ergibt sich dann das einfache Großsignalmodell in Abb. 4.10.

4.7 Rekombination in der Basis

Wir haben im vorangegangenen Kapitel die Forderung nach einer geringen Basisweite w_b aufgestellt, um eine hohe Stromverstärkung zu erzielen. Dies rechtfertigt im Nachhinein auch die eingangs aufgestellte Forderung, dass die Basisweite klein gegen die Diffusionslänge L_{nb} der Minoritäten in der Basis sein soll. Diese Forderung sorgt neben einer hohen Stromverstärkung aufgrund der kurzen durch Diffusion in der Basis zurückgelegten Strecke, auch für eine geringe Netto-Rekombination in der Basis.
Da die Diffusionslänge nur ein Mittelwert ist, gibt es immer auch Ladungsträger die selbst bei kürzester Basis rekombinieren. Wir wollen im Folgenden

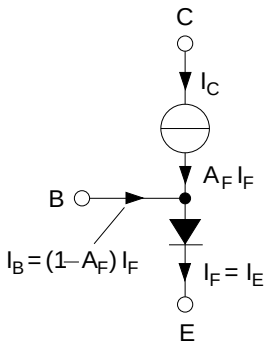

Abb. 4.10: Vereinfachtes Ebers-Moll-Ersatzschaltbild für den normal-aktiven Betrieb eines *npn*-Bipolar-Transistors.

die Verhältnisse bei Netto-Rekombination in der Basis betrachten.
Abb. 4.11 zeigt den, um die Netto-Rekombination in der Basis erweiterten Ladungsträgerfluss im Transistor aus Abb. 4.6.

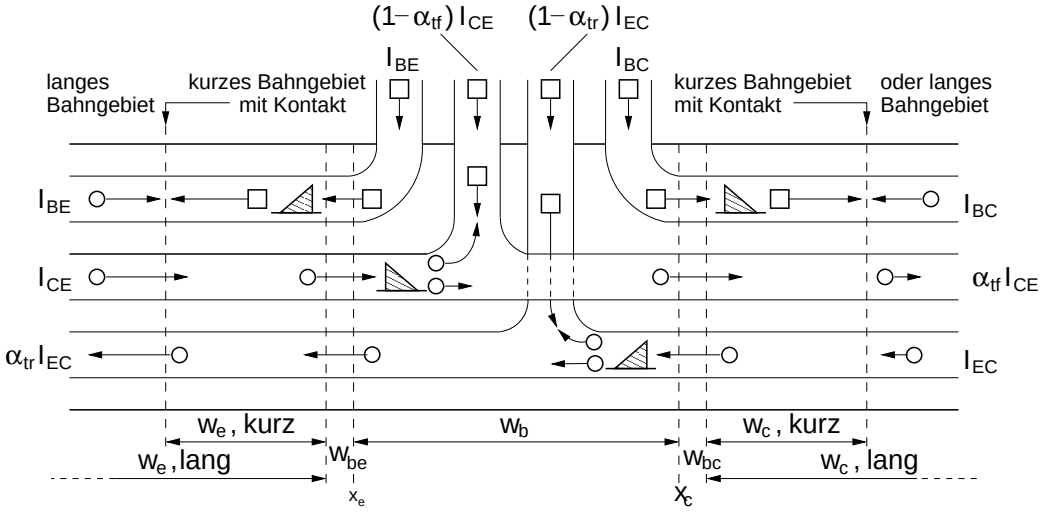

Abb. 4.11: Ladungsträgerflüsse im normal gesättigten *npn*-Transistor mit Netto-Rekombination in der Basis.

Für den Fall, dass der *npn*-Transistor im normal gesättigten Bereich betrieben wird (Basis-Kollektordiode leicht in Flusspolung). Für die nachfolgende Herleitung ist die Richtung der Polung jedoch ohne Bedeutung.

Durch die Netto-Rekombination in der Basis werden die bei x_e bzw. x_c in die Basis eintretenden Ströme I_{CE} bzw. I_{EC} beim Austritt am

jeweils anderen Ende der Basis verkleinert (Rekombination aufgrund Ladungsträger-Überschuss) oder vergrößert (Generation aufgrund Ladungsträger-Mangel).
Wir betrachten zunächst das Verhältnis des in Vorwärtsrichtung fließenden Stroms I_{CE} an den Stellen x_c und x_e und schreiben für das Verhältnis mit Hilfe der von der Diode bekannten Beziehung Gl. (3.83)

$$\alpha_{tf} = \frac{I_{CE}(x_c)}{I_{CE}(x_e)} = \frac{\cosh\left(\frac{x_c - x_c}{L_{nb}}\right)}{\cosh\left(\frac{x_c - x_e}{L_{nb}}\right)} = \frac{1}{\cosh\left(\frac{w_b}{L_{nb}}\right)} \quad . \tag{4.41}$$

α_{tf} wird als (Basis-)Transportfaktor in Vorwärtsrichtung bezeichnet. Wir können wieder den $\cosh(x)$ für kleine x annähern durch die beiden ersten Glieder der Reihenentwicklung, wodurch

$$\alpha_{tf} \approx \frac{1}{1 + \frac{1}{2}\left(\frac{w_b}{L_{nb}}\right)^2} \approx 1 - \frac{1}{2}\left(\frac{w_b}{L_{nb}}\right)^2 \tag{4.42}$$

folgt. Im zweiten Schritt darin haben wir von der Näherung $(1+\varepsilon)^{-1} \approx 1 - \varepsilon + \varepsilon^2 - \ldots \approx 1 - \varepsilon$ Gebrauch gemacht. Es gilt also für $w_b \ll L_{nb}$

$$I_{CE}(x_c) = I_{CE}(x_e) \cdot \alpha_{tf} = I_{CE}(x_e) \cdot \left(1 - \frac{1}{2}\left(\frac{w_b}{L_{nb}}\right)^2\right) \tag{4.43}$$

Da $I_{CE}(x_b)$ der in das Basis-Bahngebiet eintretende Strom noch ohne Rekombination ist, gilt $I_{CE} = I_{CE}(w_b)$, wobei I_{CE} die Stromkomponente der über den Emitter zufließenden Elektronen ist. Das Ergebnis in Gl. (4.43) bestätigt die intuitive Aussage, dass der durch Rekombination in der Basis entstehende Basisstrom

$$I_{CE}(x_e) - I_{CE}(x_c) = (1 - \alpha_{tf}) I_{CE} = \frac{1}{2}\left(\frac{w_b}{L_{nb}}\right)^2 I_{CE} \tag{4.44}$$

umso geringer wird, je geringer die Basisweite im Verhältnis zur Diffusionslänge der Minoritätsträger in der Basis ist.

Für eine Injektion von Minoritätsträgern über die Basis-Kollektor-Raumladungszone kann für das Verhältnis von $I_{EC}(x_c)$ zu $I_{EC}(x_b)$ aufgrund der Symmetrie der Anordnung die gleiche Beziehung wie für α_{tf} in

Gl. (4.41) verwendet werden. Wir erhalten für den (Basis) Transportfaktor in Rückwärtsrichtung

$$\alpha_{tr} = \frac{I_{EC}(x_e)}{I_{EC}(x_c)} = \frac{\cosh\left(\frac{x_c - x_e}{L_{nb}}\right)}{\cosh\left(\frac{x_e - x_c}{L_{nb}}\right)} = \frac{1}{\cosh\left(\frac{w_b}{L_{nb}}\right)} = \alpha_{tf} \quad . \tag{4.45}$$

Es gilt also wegen der Symmetrie der Anordnung, die man aufgrund der in beiden Richtungen identischen elektrischen Eigenschaften auch als Reziprozität bezeichnet

$$\alpha_{tr} = \alpha_{tf} =: \alpha_T = 1 - \frac{1}{2}\left(\frac{w_b}{L_{nb}}\right)^2 \quad . \tag{4.46}$$

Aufgrund der Reziprozität haben wir in Gl. (4.46) α_T allgemein für Vorwärts- und Rückwärtsrichtung als den Transportfaktor definiert.
Mit bekanntem α_T kann das aus den Stromflüssen in Abb. 4.11 abgeleitete Großsignal-Ersatzschaltbild für die Schaltungstechnik verwendet werden.

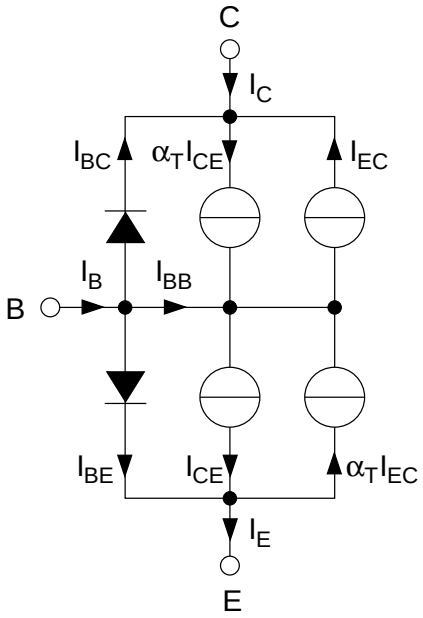

Abb. 4.12: Großsignal-Ersatzschaltbild eines *npn*-Transistors mit einer über α_T berücksichtigten Netto-Rekombination in der Basis.

Aufgrund von $\alpha_T \neq 1$ ergibt sich ein zusätzlicher Basisstrom (über Knotenregel)

$$I_{BB} = (1 - \alpha_T)(I_{CE} + I_{EC}) \tag{4.47}$$

dessen Wirkung über eine modifizierte Stromverstärkung

$$B^* = \frac{I_C}{I_{BE} + I_{BC} + I_{BB}} = \frac{1}{\frac{1}{B} + \frac{I_{BB}}{I_C}} \tag{4.48}$$

berücksichtigt werden kann. Aufgrund der Struktur von Gl. (4.48) wird B^* von der jeweils kleineren Stromverstärkung (B bzw. $\frac{I_C}{I_{BB}}$) bestimmt. Wir werden im Folgenden von einer entsprechend modifizierten Stromverstärkung ausgehen und weiterhin mit dem Transistor-Modell nach Abb. 4.8 arbeiten.

4.8 Netto-Rekombination in Raumladungszonen

Wie bei der Diode ermittelt, findet in den Raumladungszonen von Basis-Emitter und Basis-Kollektor eine Netto-Rekombination statt, die jeweils einen Netto-Rekombinationsstrom I_{rg} nach Gl. (3.59) bewirkt:

$$I_{rg,BE} = I_{rg,bes} \cdot (e^{\frac{U_{BE}}{2U_T}} - 1) = I_{SE}\,(e^{\frac{U_{BE}}{n_E U_T}} - 1) = I_{LE} \tag{4.49}$$

$$I_{rg,BC} = I_{rg,bes} \cdot (e^{\frac{U_{BC}}{2U_T}} - 1) = I_{SC}\,(e^{\frac{U_{BC}}{n_C U_T}} - 1) = I_{LC}\,. \tag{4.50}$$

Diese Ströme können als Diodenströme mit dem Nichtidealitätsfaktoren $n_E = 2, n_C = 2$ aufgefasst werden.
Im SPICE Transistor-Großsignalmodell werden sie als sogenannte Leckstromdioden mit den Strömen I_{LE} und I_{LC} und den Modellparametern I_{SE}, I_{SC} und n_E, n_C berücksichtigt. Abb. 4.13 zeigt das entsprechend modifizierte SPICE-Modell.

4.9 Sättigung

Gelangt im normal-aktiven Betrieb die Basis-Kollektorspannung in den Flussbereich, bzw. ist die Schaltung bereits für den normal gesättigten Betrieb dimensioniert, werden Minoritätsträger aus der Basis-Kollektor-Raumladungszone in das Basis- und Kollektor-Bahngebiet injiziert. Durch die Flusspolung der Basis-Kollektor-Diode steigt der Basisstrom an.

Von der Ladungssteuerungstheorie wissen wir, dass der Stromfluss durch eine Diode über die sich in ihren Bahngebieten befindenden Minoritätsträger gesteuert wird. Zu der Ladung Basis-Emitter-Diffusionskapazität kommt daher bei Sättigung noch die Basis-Kollektor-Diffusionskapazität, deren

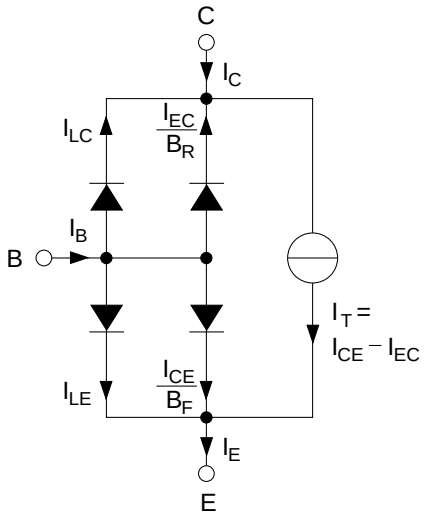

Abb. 4.13: SPICE Transistormodell mit Leckstromdioden I_{LC}, I_{LE} zur Berücksichtigung der Netto-Rekombination in den Sperrschichten.

Speicherladung umgeladen werden muss. Da dies zu einer Verlangsamung der Schaltzeiten bei schnellen Schaltvorgängen führt, ist in diesen Fällen der Sättigungsbereich unbedingt zu meiden. Die Verlangsamung betrifft insbesondere das Ausschalten des Transistors, nachdem er sich im gesättigten Betrieb befunden hat. Damit der Kollektorstrom sinken kann, muss zunächst durch einen negativen Basisstrom und durch Rekombination die Überschussladung in der Basis abgebaut werden.
Diese Vorgänge lassen sich quantitativ in einer einfachen Näherungsrechnung sehr gut mit Hilfe der Diffusionsdreiecke und der Ladungssteuerungstheorie beschreiben. Das Vorgehen entspricht dem anhand der Diode gezeigten Überlegungen. Um den Rahmen der Vorlesung nicht zu sprengen, wird jedoch an dieser Stelle auf eine Darstellung verzichtet.
Eine einfache Abgrenzung zwischen normal-aktivem Bereich und Sättigungsbereich ergibt sich, wenn als Grenze zwischen den beiden Bereichen $U_{BC} = 0$ definiert wird. Dann gilt

$$U_{BC} = 0 \Rightarrow \; U_{BE} = U_{CE} \tag{4.51}$$

und damit durch Einsetzen in Gl. (4.23)

$$I_T \mid_{U_{BC=0}} = I_S \, e^{\frac{U_{CE}}{U_T}} \approx I_C \mid_{U_{BC=0}} \tag{4.52}$$

d.h. im $I_C(U_{CE})$ Ausgangskennlinienfeld ergibt sich als Grenze für $U_{BC} = 0$ eine Exponentialfunktion, die den Bereich zwischen normal-aktiv und Satti-

gungsbereich markiert.

4.10 Modulation der Basisweite, Early-Effekt

Wir haben die, über die Raumladungszonen gesteuerten Diffusionsströme im Transistor über die Minoritätsträgerrandkonzentration in den jeweiligen Bahngebieten bestimmt (vgl. Gl. (4.16)-(4.18)). Für den Fall kurzer Bahngebiete können wir auf die Näherung mit Hilfe des Diffusionsdreiecks zurückgreifen. Für den normal-aktiven Bereich wird der Ausgangsstrom (Transferstrom I_T) über das durch Injektion vom Emitter stammende Diffusionsdreieck in der Basis bestimmt. Für dessen Weite haben wir allgemein w_b gesetzt, wobei w_b der Abstand zwischen den basisseitigen Grenzen der Basis-Emitter-Raumladungszone und der Basis-Kollektor-Raumladungszone ist, die wir in Abb. 4.3 mit x_e und x_c bezeichnet haben. Die metallurgischen Übergänge in den Raumladungszonen hatten wir mit x_{je} und x_{jc} bezeichnet. Sie sind physikalisch durch die Breite der dotierten Bereiche gegeben und sind daher nicht ortsveränderlich. Dagegen ist die Weite der Raumladungszone und damit die Basisweite w_b von der Größe der Spannung an den beiden Raumladungszonen abhängig.

Wir untersuchen im Folgenden den Einfluss der spannungsabhängigen Basisweite auf die Eigenschaften des Transistors. Hierzu machen wir wieder einige vereinfachende Annahmen und Näherungen:

1. Wir betrachten den Transistor im normal-aktiven Bereich.

2. Die Basis-Emitter-Spannung U_{BE} soll wegen 1. näherungsweise konstant sein (erläutern Sie zur Übung, warum das angenommen werden kann).

3. Wegen 2. ist auch x_e näherungsweise konstant und die Basisweite wird nur über die Spannungsabhängigkeit von x_c verändert.

Zur Vereinfachung der Rechnung legen wir in Abb. 4.14 den Nullpunkt an die Stelle x_e und bezeichnen die basisseitige Weite der Basis-Kollektor-Raumladungszone mit Δw_{bc}.
Die Weite Δw_{bc} können wir direkt mit dem Ergebnis der p-n-Diode aus Gl. (3.26) bestimmen, wobei $U_D \rightarrow (U_{D,BC} - U_{BC})$ ausgetauscht werden

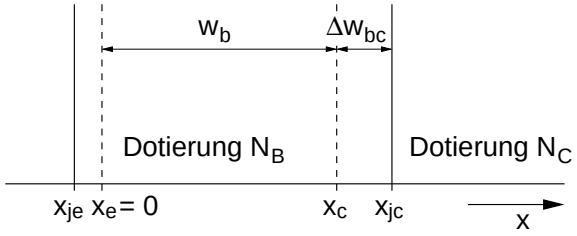

Abb. 4.14: Zur Definition der spannungsabhängigen Basisweite

muss

$$\Delta w_{bc} = \sqrt{\frac{2\varepsilon(U_{D,BC} - U_{BC})}{e} \frac{N_C}{N_B} \frac{1}{N_C + N_B}} \,. \tag{4.53}$$

Darin ist $U_{D,BC}$ die Diffusionsspannung der Kollektor-Basis-Diode und U_{BC} die über der Basis-Kollektor-Raumladungszone abfallende Spannung. Wir können diesen Ausdruck umformen in

$$\Delta w_{bc} = \underbrace{\sqrt{\frac{2\varepsilon(U_{D,BC})}{e} \frac{N_C}{N_B} \frac{1}{N_C + N_B}}}_{\Delta w_{bc(0)}} \sqrt{1 - \frac{U_{BC}}{U_{D,BC}}} \tag{4.54}$$

$$\Delta w_{bc} = \Delta w_{bc}(0) \sqrt{1 - \frac{U_{BC}}{U_{D,BC}}} \,. \tag{4.55}$$

Hiernach ergibt sich die spannungsabhängige Weite der Basis-Kollektor-Raumladungszone aus der konstanten Weite $\Delta w_{bc}(0)$ bei $U_{BC} = 0$ multipliziert mit dem spannungsabhängigen Wurzel-Term.
Im normal-aktiven Bereich ist $U_{BC} \leq 0$ und der Wurzel-Term kann in erster, grober Näherung an der Stelle $U_{BC} = 0$ durch eine Taylor-Reihe bis zum linearen Glied approximiert werden, so dass gilt

$$\Delta w_{bc} \approx \Delta w_{bc}(0) + \left. \frac{d\Delta w_{bc}}{dU_{BC}} \right|_{U_{BC}=0} \cdot U_{BC} \,, \tag{4.56}$$

$$\Delta w_{bc} \approx \Delta w_{bc}(0) - \frac{1}{2} \Delta w_{bc}(0) \frac{U_{BC}}{U_{D,BC}} \,. \tag{4.57}$$

Die Basisweite lässt sich damit schreiben als (Nullpunkt bei $x_e = 0$)

$$w_b = x_{jc} - \Delta w_{bc} \tag{4.58}$$

$$= \underbrace{x_{jc} - \Delta w_{bc}(0)}_{w_b(0)} + \frac{1}{2} \Delta w_{bc}(0) \frac{U_{BC}}{U_{D,BC}} \,. \tag{4.59}$$

Darin ist $w_b(0)$ die Basisweite bei $U_{BC} = 0$. Wir formen noch etwas um:

$$w_b = w_b(0) \left(1 + \frac{U_{BC}}{\frac{2\,w_b(0)\,U_{D,BC}}{\Delta w_{bc}(0)}} \right) \tag{4.60}$$

und erhalten mit der Definition der Early-Spannung

$$U_A := \frac{-w_b(0)}{\left.\frac{d\Delta w_{bc}}{dU_{BC}}\right|_{U_{BC}=0}} = \frac{2\,w_b(0)\,U_{D,BC}}{\Delta w_{bc}(0)} \tag{4.61}$$

das Ergebnis

$$w_b = w_b(0) \left(1 + \frac{U_{BC}}{U_A} \right) . \tag{4.62}$$

In der Definition der Early-Spannung in Gl. (4.61) ist der erste Ausdruck die formale Definition, die sich ergeben hätte, wäre die Taylor-Reihenentwicklung nicht direkt in Gl. (4.57) berechnet, sondern bis in Gl. (4.60) beibehalten worden.
Durch die spannungsabhängige Basisweite wird auch der Transferstrom I_T durch seinen Sättigungsstrom $I_S(w_b) = I_S(U_{BC})$ moduliert.
Es gilt nach Gl. (4.14)

$$I_S(U_{BC}) = e\,A\,D_N\,\frac{n_{b0}}{w_b} = e\,A\,D_N\,\frac{n_{b0}}{w_b(0)\,\left(1 + \frac{U_{BC}}{U_A}\right)} = \frac{I_S(0)}{1 + \frac{U_{BC}}{U_A}} . \tag{4.63}$$

Für $|U_{BC}| \ll U_A$ können wir mit der Näherung $\frac{1}{1+\varepsilon} \approx 1 - \varepsilon$ schreiben

$$I_S(U_{BC}) \approx I_S(0) \left(1 - \frac{U_{BC}}{U_A} \right) = I_S(0) \left(1 + \frac{U_{CB}}{U_A} \right) \tag{4.64}$$

wenn im normal-aktiven Bereich $U_{CB} \gg U_{BE}$ erfüllt ist, kann $U_{CB} \approx U_{CE}$ gesetzt werden, wodurch sich

$$I_S(U_{CE}) \approx I_S(0) \left(1 + \frac{U_{CE}}{U_A} \right) \tag{4.65}$$

ergibt. Den Transferstrom unter Berücksichtigung des Early-Effekts können wir damit im normal-aktiven Bereich schreiben als

$$I_T \approx I_S(U_{CE})\, e^{\frac{U_{BE}}{U_T}} \approx I_S(0) \left(1 + \frac{U_{CE}}{U_A} \right) e^{\frac{U_{BE}}{U_T}} . \tag{4.66}$$

Für den Schaltungsentwickler folgt aus diesem Ergebnis das wichtige Fazit, dass obwohl die Basis-Kollektor-Diode im normal-aktiven Bereich gesperrt und ihre Ströme dadurch vernachlässigbar sind, die Wirkung der Kollektor-Basis- bzw. Kollektor-Emitter-Spannung den Transferstrom (Kollektorstrom) mitbestimmt.
Auch erfolgt über diese Abhängigkeit ein Einfluss auf die Stromverstärkungen B_F und B_R nach Gl. (4.25) und (4.26) wobei im normal-aktiven Bereich nur B_F von Bedeutung ist.

4.11 Lawinen-Durchbruch, 1. Durchbruch

Für den Lawinen-Durchbruch beim Bipolar-Transistor gelten die gleichen Überlegungen wie für die p-n-Diode. Im normal-aktiven Bereich ist der Kollektor-Basis-Übergang bei großer Sperrspannung besonders gefährdet.

Definieren wir mit

$$I_{CBR} := \text{Reststrom } I_C \text{ für } I_E = 0 \text{ und } U_{CB} \gg U_T$$

und

$$U_{CB0} := \text{Durchbruchspannung der } CB\text{-Diode bei } I_E = 0$$

so können wir direkt mit $I_{CBR} = I_{R0}$ und $U_{CB0} = U_{br}$ das Ergebnis der p-n-Diode aus Gl. (3.175) mit (3.179) für den Kollektor-Basis p-n-Übergang anwenden:

$$I_C = \frac{I_{CBR}}{1 - \left(\frac{U_{CB}}{U_{CB0}}\right)^n} = M \cdot I_{CBR} \tag{4.67}$$

mit

$$M = \frac{1}{1 - \left(\frac{U_{CB}}{U_{CB0}}\right)^n} \,. \tag{4.68}$$

Gegenüber der Diode wird durch Beschalten des Emitters der Strom, der in der CB-Raumladungszone in den Bereich der Stoßionisation eintritt, um $A_F \cdot I_E$ vergrößert. Daher ist I_{R0} (hier $= I_{CBR}$) in Gl. (3.173) bzw. (3.175) um I_E zu ergänzen. Aus Gl. (4.67) wird entsprechend

$$I_C = M \cdot (I_{CBR} + A_F \cdot I_E) \,. \tag{4.69}$$

Gl. (4.69) gibt den Kollektorstrom in Abhängigkeit des Emitterstroms an. Die Einsatzspannung des Durchbruchs hängt darin nicht vom Emitterstrom

ab.

Wir leiten eine entsprechende Beziehung in Abhängigkeit des Basisstroms her. Dazu ersetzen wir in Gl. (4.69) $I_E = I_C + I_B$ und erhalten

$$I_C = \frac{M \cdot (I_B \cdot A_F + I_{CBR})}{1 - M \cdot A_F} \tag{4.70}$$

als den Kollektorstrom bei Durchbruch und endlichem Basisstrom.
Für $I_B = 0$ ergibt sich der Sonderfall

$$I_C = \frac{M}{1 - MA_F} I_{CBR} \tag{4.71}$$

darin ist

$$\frac{M}{1 - MA_F} \tag{4.72}$$

der Multiplikationsfaktor bei $I_E \neq 0$ und offener Basis ($I_B = 0$).
Der Durchbruch erfolgt in diesem Fall, wenn

$$M \cdot A_F = 1 \tag{4.73}$$

gilt. Die Spannung U_{CB}, bei der dies erfüllt ist, bezeichnen wir mit $U_{CB,CEO}$.
Durch Einsetzen von Gl. (4.68) ergibt sich

$$\frac{A_F}{1 - \frac{U_{CB,CEO}}{U_{CB0}}} = 1 \;\Rightarrow\; U_{CB,CEO} = U_{CB0}(1 - A_F)^{\frac{1}{n}} \; . \tag{4.74}$$

Bei $U_{CB,CEO}$ handelt es sich um eine Kollektor-Basis Spannung. Es wird jedoch wegen, der bei Sperrpolung der BC-Diode im Durchbruch gut erfüllten Näherung $U_{CB} \gg U_{BE} \Rightarrow U_{CB} \approx U_{CE}$, so dass wir anstelle der Kollektor-Basis Spannung die Kollektor-Emitter Spannung verwenden können. Wir setzen daher

$$U_{CE0} := U_{CB,CE0} \tag{4.75}$$

Aus Gl. (4.74) sehen wir wegen $A_F < 1$ (z.B. $A_F = 0,99$) dass immer gilt

$$U_{CE0} \;<\; U_{CB0} \; . \tag{4.76}$$

Die CB-Durchbruchspannung bei offener Basis ist also immer geringer als die bei offenem Emitter. Für die Schaltungsentwicklung ist dies von großer Bedeutung bei der Wahl der Grundschaltung in der ein Transistor betrieben wird. Die Forderung $I_E = 0$ bzw $I_B = 0$ muss in der Schaltungstechnik

nicht gleichbedeutend mit einem offenen Emitter- bzw Basisanschluss sein. Es genügt auch eine Ansteuerung durch eine hochohmige (Leerlauf) Stromquelle, deren Strom unabhängig von den Strömen im Durchbruchsfall ist. Durch eine Ansteuerung der Basis mit einer Quelle mit endlichem Widerstand ergibt sich eine Durchbruchspannung U_{CER} für die entsprechend

$$U_{CE0} \;<\; U_{CER} \;<\; U_{CB0} \tag{4.77}$$

gilt.

4.12 Zweiter Durchbruch

Setzen wir in Gl. (4.69) für $I_C = I_E - I_B$ ein und stellen die Gleichung nach I_B um, so ergibt sich

$$I_B = (1 - MA_F)\, I_E - MI_{CBR}\,, \tag{4.78}$$

$$I_B \approx (1 - MA_F)\, I_E\,. \tag{4.79}$$

Im Fall eines Lawinendurchbruchs kehrt nach Gl. (4.79) ab $MA_F > 1$ der Basisstrom sein Vorzeichen um. Dies ist die Folge des sehr großen, durch die Lawinenvervielfachung in der CB-RLZ entstehenden Stroms. Er fließt als Majoritätsträgerstrom zum großen Teil über die Basis ab und verursacht dort einen Spannungsabfall im Basis-Bahngebiet. Das Bauelement im Ersatzschaltbild, an dem dieser Spannungsabfall auftritt ist der Basisbahnwiderstand r_B, den wir in dem Transistor-Ersatzschaltbild in Abb. 4.15 berücksichtigt haben.

Aufgrund der Transistorgeometrie ist die durch den negativen Basisstrom verursachte Basis-Emitter Spannung U_{bE} ortsabhängig. Diese Ortsabhängigkeit ist in unserem eindimensionalen Modell nicht berücksichtigt, da sie aufgrund der Anordnung vertikal auftritt, so dass sich Unterschiede zwischen oberem und unterem BE-Rand ergeben. Als Folge der Ortsabhängigkeit konzentriert sich der Strom in den Bereichen der größten Basis-Emitter Spannung. Durch die daraus folgende höhere Injektion steigt I_E und damit auch die Stromdichte in diesem Bereich an. Durch den Anstieg von I_E kommt es wiederum nach Gl. (4.79) zu einem größeren Basisstrom, wodurch eine Mitkopplung einsetzt. Diese Mitkopplung begründet den Einsatz des zweiten Durchbruchs.

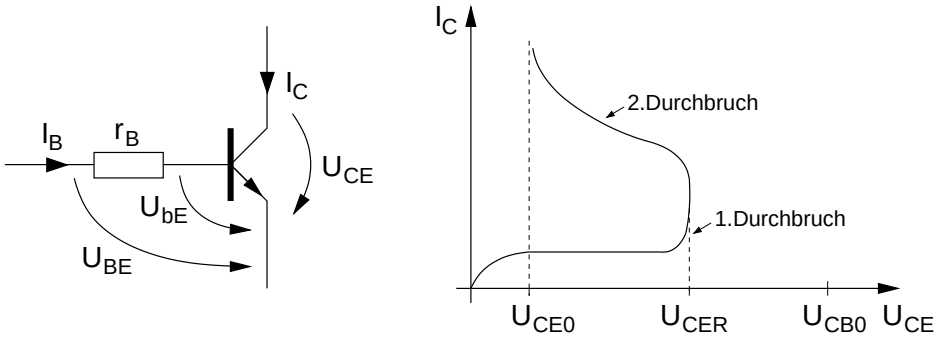

Abb. 4.15: Links: Vergrößerung der inneren Basis-Emitter Spannung U_{bE} bei $U_{BE} = const.$ durch negativen Basisstrom ($I_B < 0$). Rechts: Rückläufige Kennlinie durch negativen Basisstrom bei Einsetzen des Zweiten Durchbruchs.

Durch die hohe Stromdichte, aufgrund der lokalen Konzentration des Stroms wird, wenn keine ausreichende Strombegrenzung z.B. durch Serienwiderstände erfolgt, der Transistor durch den zweiten Durchbruch zerstört.

Dieser Effekt erklärt die Rückläufigkeit der Kennlinie in Abb. 4.15 rechts bei Einsetzen des zweiten Durchbruchs. Aufgrund der größeren inneren Basis-Emitter Spannung U_{bE} steigt bei außen konstanter Spannung U_{BE} der Kollektorstrom, wodurch die Kollektor-Emitter Spannung sinkt. Die Kennlinie in Abb. 4.15 gilt auch für $I_E = const.$ Hier ist zwar der Strom konstant, verteilt sich aber aufgrund der vertikalen Inhomogenität so, dass er sich lokal an den Stellen einer größeren inneren Basis-Emitter Spannung konzentriert, so dass es wiederum zu der damit verbundenen Mitkopplung über I_E und dem damit verbundenen Durchbruchmechanismus kommt.

4.13 Physikalisches Großsignalmodell

Wir können unter Verwendung der bisher abgeleiteten Ergebnisse für den Transistor, unter Zuhilfenahme der Ähnlichkeit zur p-n-Diode, das in Abb. 4.16 gezeigte Großsignalmodell erstellen.

Darin sind r_B, r_C und r_E die Bahn- bzw Zuleitungswiderstände des Transistors. C_{CB} und C_{BE} sind die Sperrschichtkapazitäten des BC- bzw BE-Über-

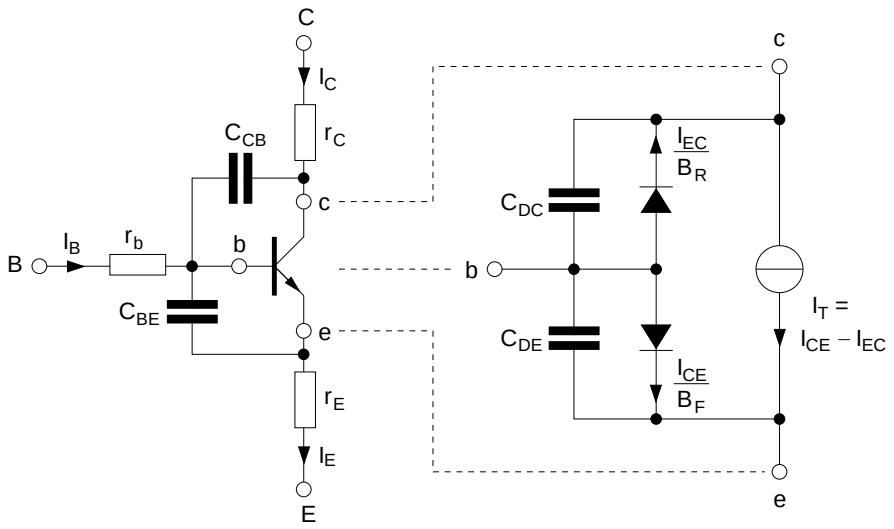

Abb. 4.16: Großsignalmodell des *npn*-Bipolartransistors. Links: Äußere Anschlüsse mit parasitären Elementen und eigentlichem (innerem) Transistor. Rechts: Großsignalersatzschaltbild des eigentlichen Transistors

gangs. Sie liegen parallel zu den Diffusionskapazitäten C_{DC} und C_{DE} der BC- und BE-Diode. Ihre Werte bestimmen sich analog zu der Vorgehensweise bei der p-n-Diode.
Im Fall des normal-aktiven Transistors ist die BC-Diode in Sperrrichtung gepolt. Daher ist $C_{DC} \ll C_{CB}$ und kann vernachlässigt werden.
Für die Diffusionskapazität der Basis-Emitter-Diode können wir mit Gl. (3.144) schreiben

$$C_{DE} := \frac{d\Delta Q}{dU_{BE}} = \tau_T \frac{I_S}{U_T} e^{\frac{U_{BE}}{U_T}} \,, \tag{4.80}$$

wobei ΔQ die in den BE-Gebieten gespeicherten Minoritätsträgerladungen sind. Mit

$$I_C \approx I_S \, e^{\frac{U_{BE}}{U_T}} \tag{4.81}$$

im normal-aktiven Bereich wird aus Gl. (4.80)

$$C_{DE} = \tau_T \frac{I_C}{U_T} \,. \tag{4.82}$$

Für τ_T können wir wegen Gl. (4.34) die Näherung aus Gl. (3.139) verwenden

$$\tau_T \approx \tau_{B_n} = \frac{w_B^2}{2\, D_n} \,. \tag{4.83}$$

4.14 Kleinsignalmodell

Wir wollen zunächst die zwei für die Schaltungstechnik wichtigen Begriffe Arbeitspunkt und Kleinsignal anhand des kompletten Kennlinienfeldes des Bipolar-Transistors in Abb. 4.17 definieren.

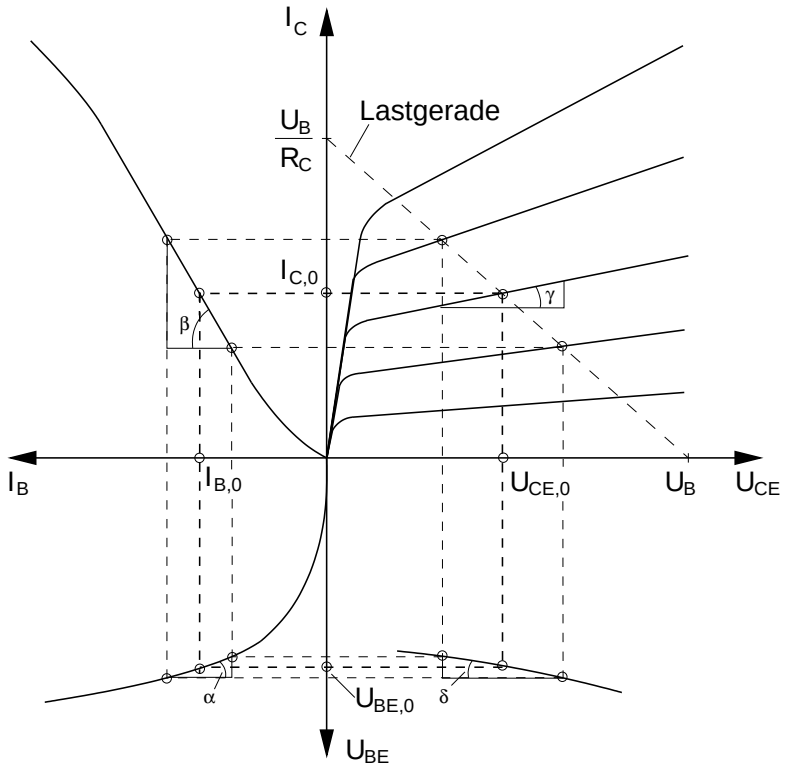

Abb. 4.17: Ausgangskennlinienfeld eines *npn*-Bipolar-Transistors mit Arbeitspunkt in $\{U_{BE,0}, I_{B,0}, I_{C,0}, U_{CE,0}\}$. Die Kleinsignalparameter ergeben sich aus den Steigungen $\{\alpha, \beta, \gamma, \delta\}$ der jeweiligen Kurven im Arbeitspunkt.

Wir nehmen dazu an, dass der Transistor in einer, vom Schaltungsentwickler vorgegebenen, Schaltung betrieben wird. Wir wollen an dieser Stelle keine speziellen Annahmen zu der Schaltung machen. Wir gehen jedoch davon aus, dass sich, durch diese Schaltung bewirkt, an den Anschlüssen des Transistors bestimmte Spannungen und Ströme einstellen.
Als Arbeitspunkt bezeichnen wir dann die Ströme und Spannungen an den Anschlüssen des Transistors, die im Ruhezustand der Schaltung vorliegen.
Ein vollständiger Satz von Spannungen und Strömen im Arbeitspunkt sind die in Abb. 4.17 eingezeichneten Größen $\{U_{CE,0}, I_{C,0}, I_{B,0}, U_{BE,0}\}$. Den mit

einem Komma getrennten Index 0 haben wir zur Kennzeichnung des Ruhezustandes („quicent“) verwendet. Wenn keine Gefahr zur Verwechslung mit der Durchbruchspannung U_{CE0} besteht, kann und wird das Komma im Folgenden auch weggelassen. Ein Arbeitspunkt kann (muss aber keinesfalls) zum Beispiel durch eine Beschaltung nach Abb. 4.18 in der folgenden Weise eingestellt werden:

1. Zwischen Basis und Emitter wird eine Spannung U_{BE0} angelegt.
2. Über die Steuerkennlinie $I_B(U_{BE})$ im dritten Quadranten folgt daraus ein Basisstrom I_{B0}.
3. Durch die Stromverstärkung $B = \frac{I_C}{I_B}$ (zweiter Quadrant) fließt ein um $B \cdot I_{B0} = I_{C0}$ größerer Kollektorstrom im Arbeitspunkt.
4. Dieser Kollektorstrom verursacht an R_C einen Spannungsabfall, so dass sich bei einem Spannungsumlauf $I_{C0}\,R_C + U_{CE0} = U_B$ die Kollektor-Emitter-Spannung U_{CE0} im Arbeitspunkt ergibt.

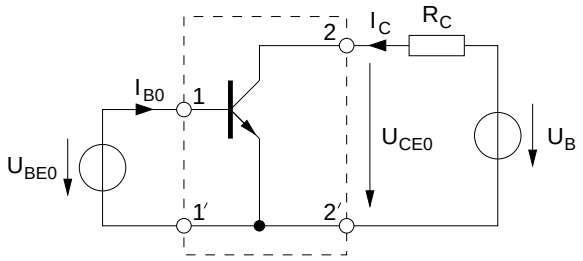

Abb. 4.18: Einfache Beschaltung zur Einstellung eines Arbeitspunktes für einen Transistor in Emittergrundschaltung.

Es lässt sich einfach zeigen (Elektronik II), dass I_C und U_{CE} in der Schaltung nach Abb. 4.18 ausschließlich Werte auf der im ersten Quadranten eingezeichneten fallenden Lastgeraden annehmen können.
Durch Einspeisen eines Signals in die Schaltung ändern sich Ströme und Spannungen am Transistor gegenüber dem Arbeitspunkt. In der Schaltung in Abb. 4.18 könnte ein Signal z.B. durch eine in Reihe zur Spannungsquelle U_{BE0} geschaltete Signal-Spannungsquelle eingespeist werden.
Abhängig von der Größe (Amplitude) des eingespeisten Signals stellen sich

die Werte $\{U_{CE}, I_C, I_B, U_{BE}\}$ am Transistor ein. Die Änderung

$$\begin{aligned}&\{\Delta U_{CE}, \Delta I_C, \Delta I_B, \Delta U_{BE}\} = \\ &\{U_{CE} - U_{CE0}, I_C - I_{C0}, I_B - I_{B0}, U_{BE} - U_{BE0}\}\end{aligned} \tag{4.84}$$

bezeichnen wir als Aussteuerung (um den Arbeitspunkt). Sie entspricht den Signalspannungen und -strömen.
Ist die Aussteuerung so klein, dass der Verlauf der Kennlinien in Abb. 4.17 durch eine Geradennäherung mit ausreichender Genauigkeit beschrieben werden kann, so sprechen wir von einer Kleinsignalaussteuerung oder von Kleinsignalbetrieb. Der Arbeitspunkt entspricht also dem ersten Glied einer Taylor-Reihenentwicklung (vgl. z.B. $I(U_0)$ in Gl. (3.156)), das den Funktionswert, um den entwickelt wird, angibt. Der lineare Teil der Kennlinie wird durch das zweite Glied der Taylorreihe beschrieben ($\frac{dI(U)}{dU}\Big|_{U=U_0} \cdot \Delta U$ in Gl. (3.156)).
Eine Kennlinie in einem Quadranten von Abb. 4.17 wird jeweils durch Wertepaare auf den Achsen dieses Quadranten gebildet. Zum Beispiel sind dies für die Steuerkennlinie $I_B(U_{BE})$ die Größen I_B und U_{BE}. Offen bleibt, welche Werte die jeweils beiden anderen Größen des aus vier Parametern bestehenden Datensatzes haben. Im Fall der Steuerkennlinie fehlt die Information darüber, welche Werte U_{CE} und I_C haben. Für den Fall der Ausgangskennlinien im ersten Quadranten ist offen, welche Bedingung für die Eingangsgrößen I_B, U_{BE} der Eingangskennlinie gelten. Diese Festlegung wird im folgenden Kapitel am Beispiel der Definition der Hybridparameter nachgeholt.

4.15 Hybridparameter

Für den Fall der Kleinsignalaussteuerung liegen lineare Abhängigkeiten zwischen den Signalspannungen und -strömen vor. Es kann daher der Überlagerungssatz angewendet werden. Dieser erlaubt die Ermittlung der Gesamtwirkung durch Überlagerung der Wirkungen der einzelnen Ursachen. Bei der Ermittlung der Wirkung einer einzelnen Ursache werden die anderen Ursachen zu Null gesetzt. Beachten: Das Null-Setzen einer Größe bezieht sich auf die Signalaussteuerung und entspricht einem konstanten Wert im Arbeitspunkt.

Bei der Ermittlung der Kleinsignalparameter eines Transistors werden je nach verwendeten Parametern unterschiedliche Größen zu Null gesetzt. Wir betrachten hier beispielhaft die Hybridparameter. Hier gilt vereinbarungsgemäß für einen Transistor in Emittergrundschaltung nach Abb. 4.18 $\Delta U_{CE} = 0$ bzw. $U_{CE0} = const.$ für den zweiten und dritten Quadranten (linke Seite von Abb. 4.17), und $\Delta I_B = 0$ bzw. $I_B = const.$ für den ersten und vierten Quadranten (rechte Seite von Abb. 4.17). Aus Abb. 4.17 lassen sich mit diesen Randbedingungen die folgenden vier Hybridparameter ablesen

$$h_{11} = \left.\frac{\partial U_{BE}}{\partial I_B}\right|_{U_{CE}=const.=U_{CE0}} = r_{be} = \frac{1}{g_{be}} \quad \text{, Eingangswiderstand} \tag{4.85}$$

$$h_{21} = \left.\frac{\partial I_C}{\partial I_B}\right|_{U_{CE}=const.=U_{CE0}} = \beta \quad \text{, Stromverstärkung} \tag{4.86}$$

$$h_{12} = \left.\frac{\partial U_{BE}}{\partial U_{CE}}\right|_{I_B=const.=I_{B0}} \quad \text{, Spannungsrückwirkung} \tag{4.87}$$

$$h_{22} = \left.\frac{\partial I_C}{\partial U_{CE}}\right|_{I_B=const.=I_{B0}} = g_0 \quad \text{, Ausgangsleitwert .} \tag{4.88}$$

Beachten: Die Hybridparameter sind Kleinsignalparameter und daher abhängig vom Arbeitspunkt $\{U_{CE0}, I_{B0}\}$ der Schaltung.
Die Forderung $U_{CE} = const.$ bzw. $I_B = const.$ entspricht für die Signalsteuerung wegen $\Delta U_{CE} = 0$ einem Kurzschluss des Signals am Ausgang bzw. wegen $\Delta I_B = 0$ einem Leerlauf des Signals am Eingang.
Zur Ermittlung der Hybridparameter kann das Großsignalmodell aus Abb. 4.16 verwendet werden. Wir betrachten zunächst den eigentlichen Transistor mit dem Ersatzschaltbild auf der rechten Seite. Für die Basis-Emitter-Diffusionskapazität kann als Kleinsignalwert der Wert im Arbeitspunkt nach Gl. (4.82) verwendet werden.
Für den invers-aktiven Betrieb kann der Beitrag des eigentlichen Transistors in Form der Kollektor-Basis-Diode und deren Diffusionskapazität vernachlässigt werden. Sie können für diesen Betriebsfall aus dem Modell entfernt werden. Für die restliche Schaltung gelten mit Gl. (4.66) für den Early-

Effekt die Beziehungen:

$$I_C \approx I_{CE} \left(1 + \frac{U_{CE}}{U_A}\right) \tag{4.89}$$

$$I_B = \frac{I_{CE}}{B_F} \tag{4.90}$$

$$B_N := B_F \left(1 + \frac{U_{CE}}{U_A}\right) = \frac{I_C}{I_B} \tag{4.91}$$

$$I_{CE} \approx I_S \, e^{\frac{U_{BE}}{U_T}} \; . \tag{4.92}$$

Mit B_N wird dabei die Vorwärtsstromverstärkung eingeführt, die sich durch Einsetzen der spannungsabhängigen Basisweite nach Gl. (4.62) in die Stromverstärkung nach Gl. (4.25) ergibt.
Die Basis-Emitter-Diode des Großsignalmodells kann wie in Gl. (3.160) bei der p-n-Diode gezeigt, durch den Kleinsignalleitwert nach Gl. (4.85)

$$g_{be} = \left.\frac{\partial I_B}{\partial U_{BE}}\right|_{U_{CE0}} = \frac{1}{B_F} \frac{I_{C0}}{U_T} = \frac{I_{B0}}{U_T} \tag{4.93}$$

ersetzt werden.
Für die Stromverstärkung wird die quasistatische Kleinsignal-Stromverstärkung nach Gl. (4.86)

$$\beta_0 = \left.\frac{\partial I_C}{\partial I_B}\right|_{U_{CE0}} = \left.\frac{\partial B_N \, I_B}{\partial I_B}\right|_{U_{CE0}} = B_N + I_B \left.\frac{\partial B_N}{\partial I_B}\right|_{U_{CE0}} \tag{4.94}$$

eingesetzt. Bei arbeitspunktunabhängiger Stromverstärkung gilt

$$\beta_0 \approx B_N \approx B_F \tag{4.95}$$

wobei im zweiten Schritt der Näherung $B_N \approx B_F$ auch der Early-Effekt vernachlässigt wurde. Es lässt sich einfach zeigen (zur Übung), dass für β_0 auch geschrieben werden kann

$$\beta_0 = \frac{g_m}{g_{be}} \tag{4.96}$$

mit

$$g_m := \left.\frac{\partial I_C}{\partial U_{BE}}\right|_{U_{CE0}} = \frac{I_{C0}}{U_T} \quad , \quad I_C \approx I_S \cdot e^{\frac{U_{BE}}{U_T}} \tag{4.97}$$

als die Steilheit des Transistors. Sie gibt an, wie stark sich der Kollektorstrom des Transistors bei kleinen Änderungen der Eingangsspannung U_{BE} ändert.

Die Spannungsrückwirkung h_{12} ist, aufgrund des exponentiellen Strom-Spannungszusammenhangs, in der Regel vernachlässigbar, so dass

$$\left.\frac{\partial U_{BE}}{\partial U_{CE}}\right|_{I_{B0}} \approx 0 \tag{4.98}$$

angenommen werden kann.
Der Ausgangsleitwert unter Berücksichtigung des Early-Effektes ergibt sich durch Ableiten von Gl. (4.89) zu

$$g_0 = \left.\frac{\partial I_C}{\partial U_{CE}}\right|_{I_{B0}} = \frac{I_C}{U_{CE} + U_A} \; . \tag{4.99}$$

Damit kann das in Abb. 4.19 gezeigte Ersatzschaltbild des eigentlichen Transistors gezeichnet werden.
Die darin enthaltenen Größen entsprechend den in Gl. (4.85) bis (4.88) defi-

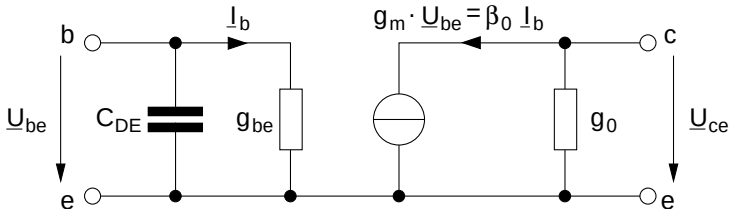

Abb. 4.19: Kleinsignalmodell des eigentlichen Transistors.

nierten Hybridparametern. Zur Kennzeichnung, dass es sich bei den Strömen und Spannungen des Ersatzschaltbildes um Signalamplituden handelt, wurde die Phasorenschreibweise gewählt.
Abb. 4.20 zeigt das vollständige Kleinsignalmodell des Transistors, das gegenüber Abb. 4.19 um die Bahnwiderstände und die Sperrschichtkapazitäten entsprechend Abb. 4.16 links erweitert wurde. Es wird auch als Giacoletto-Ersatzschaltbild bezeichnet.

Ströme und Spannungen in den Ersatzschaltbildern sind, entsprechend ihrer Herleitung, Signalamplituden, was durch die Phasorenschreibweise ausgedrückt wird.
Der Arbeitspunkt des Transistors ist in den Werten der Elemente des Ersatzschaltbildes berücksichtigt. Auch für das vollständige Ersatzschaltbild können die Hybridparameter bestimmt werden. Da sämtliche darin enthaltenen Größen bereits linear sind, können die Bestimmungsgleichungen (4.85)

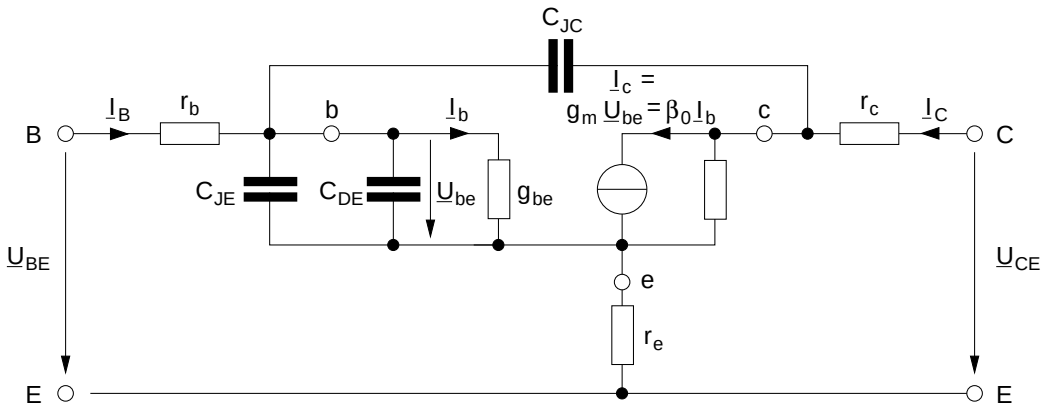

Abb. 4.20: Vollständiges Kleinsignalersatzschaltbild eines Bipolar-Transistors (Giacoletto-Ersatzschaltbild).

bis (4.88) linear und in Phasorenschreibweise ausgedrückt werden

$$\underline{h}_{11} = \left.\frac{\underline{U}_{BE}}{\underline{I}_B}\right|_{\underline{U}_{CE}=0} , \tag{4.100}$$

$$\underline{h}_{21} = \left.\frac{\underline{I}_C}{\underline{I}_B}\right|_{\underline{U}_{CE}=0} , \tag{4.101}$$

$$\underline{h}_{12} = \left.\frac{\underline{U}_{BE}}{\underline{U}_{CE}}\right|_{\underline{I}_B=0} , \tag{4.102}$$

$$\underline{h}_{22} = \left.\frac{\underline{I}_C}{\underline{U}_{CE}}\right|_{\underline{I}_B=0} . \tag{4.103}$$

Die Hybridparameter lassen sich demnach durch einfache Wechselstromrechnung mit Phasoren aus der Schaltung in Abb. 4.20 berechnen. Dies wird eine der Aufgaben der Vorlesung Elektronik II sein.

4.16 Transitfrequenz

Durch das Berücksichtigen der Kapazitäten im Giacoletto-Ersatzschaltbild in Abb. 4.20 weicht der innere Basisstrom $\underline{I}_b$ mit zunehmender Frequenz von dem äußerem Basisstrom $\underline{I}_B$ ab. Es gilt mit Gl. (4.82)

$$C_{BE} = C_{JE} + C_{DE} = C_{JE} + \tau_T \, g_m \tag{4.104}$$

für das Verhältnis der Basisströme bei vernachlässigbarem Strom durch C_{JC}

$$\frac{\underline{I}_b}{\underline{I}_B} \approx \frac{g_{be}}{g_{be} + j\omega C_{BE}} = \frac{1}{1 + j\omega \frac{C_{BE}}{g_{be}}} = \frac{1}{1 + \frac{j\omega}{\omega_\beta}} \tag{4.105}$$

$$\text{mit der } \underline{\beta \text{ -Grenzfrequenz}} \quad \omega_\beta := \frac{g_{be}}{C_{BE}} \, . \tag{4.106}$$

Der innere Kollektorstrom ist über die reellwertige Stromverstärkung β_0 nach Gl. (4.95) bzw. (4.96) vom inneren Basisstrom abhängig. Definiert man die Wechselstromverstärkung bezüglich der äußeren Transistorklemmen, so ergibt sich für $\underline{U}_{CE} = 0$ bei vernachlässigbarem r_C mit Gl. (4.105)

$$\underline{\beta} := \left. \frac{\underline{I}_C}{\underline{I}_B} \right|_{\underline{U}_{CE}=0} = \frac{\beta_0 \, \underline{I}_b}{\underline{I}_B} = \frac{\beta_0}{1 + \frac{j\omega}{\omega_\beta}} \, . \tag{4.107}$$

Die Stromverstärkung fällt also oberhalb der β-Grenzfrequenz ab, so dass für $\omega \gg \omega_\beta$ näherungsweise

$$\underline{\beta} \approx \frac{\beta_0 \, \omega_\beta}{j\omega} = \frac{\beta_0 \, f_\beta}{j \, f} = \frac{f_T}{j \, f} \tag{4.108}$$

gilt. Darin ist die Transitfrequenz

$$f_T := \beta_0 \, f_\beta \tag{4.109}$$

die Frequenz, bei der $|\underline{\beta} \; (f = f_T)| = 1$ gilt. Wir können mit Gl. (4.96) und (4.106) für den Kehrwert der Transitfrequenz auch schreiben

$$\frac{1}{f_T} = \frac{g_{be}}{g_m} \cdot \frac{2\pi C_{BE}}{g_{be}} = 2\pi \, \frac{C_{JE} + C_{DE}}{g_m} \tag{4.110}$$

woraus mit Gl. (4.82)

$$\frac{1}{2\pi f_T} = C_{JE} \, \frac{U_T}{I_C} + \tau_T \tag{4.111}$$

wird. Die Transitfrequenz ist also in erster Näherung abhängig vom Arbeitspunkt (I_{C0}), der Basis-Emitter-Sperrschichtkapazität und von der Transitzeit τ_T. Bei vernachlässigbaren Minoritätsträgerladungen in Emitter und Kollektor können wir für $\tau_T \approx \tau_{B_n}$ nach Gl. (4.83) setzen.
Bei genauerer Berechnug unter Berücksichtigung von r_e, r_c und C_{JC} (g_0 kann in der Regel weiterhin vernachlässigt werden) ergibt sich die Transitfrequenz aus

$$\frac{1}{2\pi f_T} = (C_{JE} + C_{JC}) \, \frac{U_T}{I_C} + \tau_T + (r_e + r_c) \, C_{JC} \, . \tag{4.112}$$

A Näherungsrechnung für Flusspolung

Für $U > 0$ kann $n(x) \gg n_1$ und $p(x) \gg p_1$ für $x_p \leq x \leq x_n$ angenommen werden. Für $p(x) \cdot n(x)$ im Zähler von Gl. (3.54) kann Gl. (3.55) eingesetzt werden, wodurch sich Gl. (A.1) ergibt.

$$I_{rg} = qAn_i^2(e^{\frac{U}{U_T}} - 1) \int_{x_p}^{x_n} \frac{dx}{(n(x) + n_1)\tau_p + (p(x) + p_1)\tau_n} \ . \tag{A.1}$$

Der Hauptbeitrag zum Integral in Gl. (A.1) kommt von dem Bereich, in dem $n\tau_p = p\tau_n$ ist, da dort der Nenner minimal wird. Diese Bedingung ist erfüllt für

$$e^{\frac{U}{U_T}} = \left(\frac{\tau_p N_D}{\tau_n N_A}\right)^{-\frac{1}{2}} e^{\frac{(U_D - U)}{2U_T}} \ . \tag{A.2}$$

Hiermit erhalten wir

$$n\tau_p + p\tau_n = 2\sqrt{\tau_p\tau_n}\sqrt{N_D N_A}\, e^{-\frac{U_D}{2U_T}} e^{\frac{U}{2U_T}} \ , \tag{A.3}$$

und mit Gl. (3.23)

$$n\tau_p + p\tau_n = 2\sqrt{\tau_p\tau_n}\, n_i\, e^{\frac{U}{2U_T}} \ . \tag{A.4}$$

Hiermit liefert Gl. (A.1)

$$I_{rg} = \frac{qAn_i}{\sqrt{\tau_p\tau_n}} w(U) \frac{e^{\frac{U}{2U_T}} - e^{-\frac{U}{2U_T}}}{2} \ . \tag{A.5}$$

Mit

$$\tau_{eff} := 2\sqrt{\tau_p\tau_n} \tag{A.6}$$

für $U > 0$, also einem anderen Wert als für $U < 0$ lassen sich die Gleichungen (3.56) und (A.5) zusammenfassen zu

$$I_{rg} = I_{rg,s}(U)(e^{\frac{U}{2U_T}} - 1) \tag{A.7}$$

mit

$$I_{rg,s}(U) = \frac{qAn_i}{\tau_{eff}} w(U) \ . \tag{A.8}$$

B Genauere Berechnungen zu Unipolaren Bauelementen

B.1 Berechnung der Kennlinien des Sperrschichtfeldeffekttransistors

Anmerkung: Berechnung derzeit noch nicht vorhanden.

B.2 Berechnung der Influenz von Ladungen auf einem MOS-Kondensator

Anmerkung: Berechnung derzeit noch nicht vorhanden.

B.3 Berechnung der Kennlinie eines MOS-FET

Anmerkung: Berechnung derzeit noch nicht vorhanden.