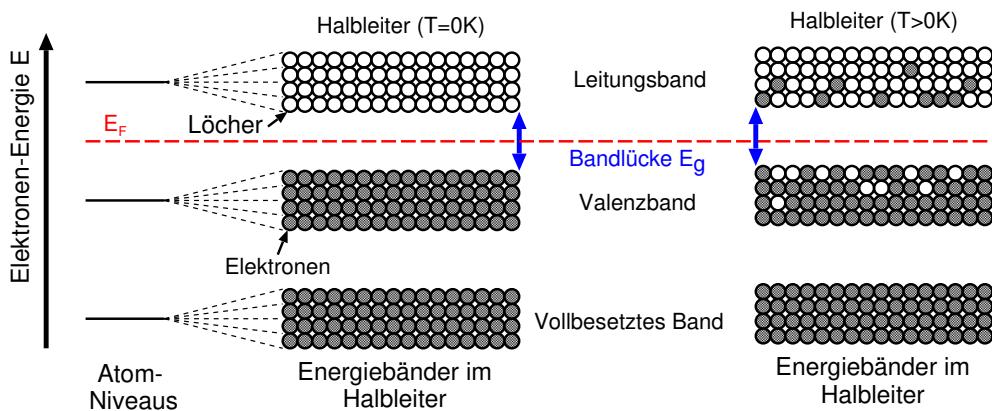


Halbleitertechnik

Peter Widerin, Stefan Stark
peter.widerin@fhv.at, stefan.stark@fhv.at



Prüfungsbedingung:

- 100 Punkte insgesamt
- 20 Punkte können während des Semesters durch Abgabe von Übungsbeispielen erworben werden (10 Punkte bei P. Widerin und 10 Punkte bei S. Stark)
- 10 Punkte können durch Abgabe des Laborprotokolls (letzte Übungsstunde) erworben werden
- 70 Punkte für Theorie und Beispiele bei der zweistündigen schriftlichen Prüfung am 30.06.2025
- Erlaubte Hilfsmittel: Taschenrechner

Literatur:

- S.M. Sze and K.K. Ng, *Physics of Semiconductor Devices*, (Wiley, New York 2006)
- N.W. Ashcroft und N.D. Mermin, *Solid State Physics*, (Cengage, Cornell 2016)
- B. Razavi, *Fundamentals of Microelectronics*, (John Wiley, Hoboken 2008)
- Y. Tsividis, *The MOS Transistor*, (McGraw-Hill, Boston 1999)
- D. C. Giancoli, *Physik, Lehr- und Übungsbuch* (Pearson, München 2019)
- R. P. Feynman, *The Feynman lectures on physics, Vol.1 and Vol 2* (Basic Books, New York 2011)
- R. Müller und W. Heywang, *Halbleiter-Elektronik, Band 1 und Band 2* (Springer, Berlin 1975)

0 – Inhaltsverzeichnis**Inhalt**

0 – Inhaltsverzeichnis	2
1 – Einleitung	3
1.1 – Wann spricht man von einem Halbleiter?	4
2 – Physikalische Grundlagen	6
2.1 – Mechanik	6
2.2 – Elektromagnetismus	8
2.3 – Kontinuitätsgleichung	11
3 – Freie Ladungen	12
3.1 – Drude-Modell	13
3.2 – Mechanische Wellen	18
3.3 – Elektronen als Welle	26
4 – Gebundene Ladungen	33
4.1 – Elektronen in einem Potentialtopf	33
4.2 – Elektronen im periodischen Potential	34
5 – Ladungskonzentration im Halbleiter	41
5.1 – Thermisches Gleichgewicht	41
5.2 – Transportphänomene	48
6 – Grundgleichungen der Halbleiterphysik	52
6.1 – Physikalische Parameter wichtiger Halbleiter bei Raumtemperatur (300K)	53
6.2 – Beispiele	54
7 – Halbleiterbauelemente	55
7.1 – Die Diode	56
7.2 – Der Bipolar-Transistor	63
7.3 – Der Feldeffekt-Transistor	70
8 – Halbleiterfertigung	78

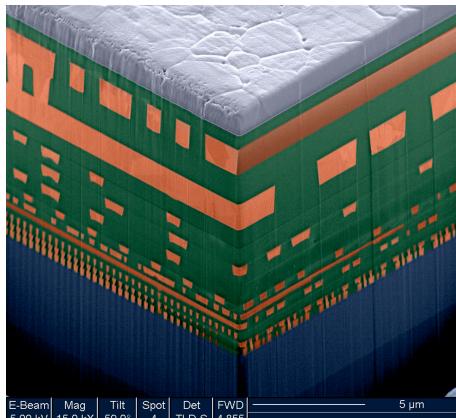
1 – Einleitung

Halbleitertechnik bestimmt einen Großteil moderner (elektronischer) Maschinen und Geräte. Basierend auf einem guten Verständnis der physikalischen Eigenschaften und Jahrzehnten technologischer Forschung und Entwicklung werden Halbleiter-Bauelemente heute auf vielfältige Weise eingesetzt:

- elektronische Leistungsschalter (Eisenbahn, Elektromotor, Gleich-Wechselrichter,...)
- Datenspeicherung durch Halbleiter (EEPROM, Flash-Speicher,...)
- Datenverarbeitung durch elektronische Schaltungen (analog und digital)
- Halbleiter als Sensoren (Bildsensoren, Temperatursensoren,.....)
- Halbleiter als Lichtquellen (LEDs,)
- Stromerzeugung (Photovoltaikzellen,....)

Die Halbleiterfertigung (Lithographie) erlaubt es, kompakt mehrere Bauelemente auf einem einzigen Halbleiter anzurichten. Die Verbindungen zwischen den Bauelementen werden in höherliegenden Metall-Lagen, getrennt durch isolierende Schichten, gefertigt.

Der Schnitt durch einen 64-bit Microprocessor Chip (IBM 90nm Server-Class CMOS Technologie), siehe Abbildung



Quelle: IBM 90nm Technologie

- Bauelemente mit minimal möglicher Abmessung (Kanal-Länge) von 90nm sind direkt im Halbleiter (blau) angeordnet
- erste Metalllage mit minimaler Dimension von 120nm für die lokalen Verbindungen zwischen den Bauelementen
- fünf Kupferlagen (Cu) mit SiCOH Isolator dazwischen für elektrische Verbindungen zwischen den Schaltungsböcken
- drei breitere Lagen mit Cu/SiCOH für grössere Ströme
- zwei breite Cu Lagen in FTEOS/SiO₂
- eine Lage Al(Cu) für Power-Verdrahtung (Versorgungsleitungen) und Lötstellen (Pads) zum Chipgehäuse

Was ist das Ziel dieser Vorlesung:

- 1 Fachliteratur zur Halbleitertechnik wird verständlich(er) (z.B. "Fundamentals of Microelectronics", B. Razavi, "Physics of Semiconductor Devices", S.M. Sze)
- 2 Physikalische Prinzipien von Halbleitern werden auf einem fundamentalen Niveau verstanden
 - Welche **Grundlegenden Konzepte** beschreiben die Festkörper?
 - Woher kommt die **elektrische Leitfähigkeit**?
 - Wie kann das **Verhalten von Festkörpern** durch zugrundeliegende Modelle verstanden werden?
 - Welche Physik beschreibt einen **pn-Übergang**?
 - Welche Physik beschreibt einen **Bipolar Transistor**?
 - Welche Physik beschreibt einen **MOS Transistor**?

Was wird vorausgesetzt:

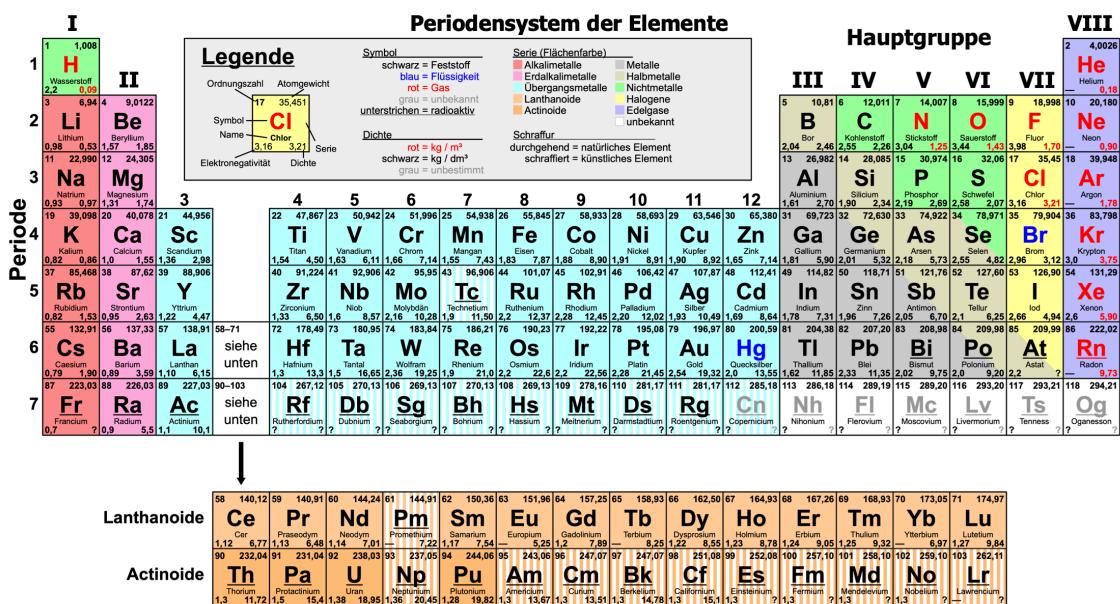
- Grundwissen in Chemie
- Grundwissen in Mechanik und Thermodynamik
- Mathematische Kenntnisse bis einschliesslich Differentialgleichungen
- Analyse und Simulation von elektrischen Schaltungen im Zeit- und Frequenz-Bereich.

1.1– Wann spricht man von einem Halbleiter?

Festkörper können in Bezug auf ihre elektrischen Eigenschaften in Leiter (hohe elektrische Leitfähigkeit, Metalle) oder Isolatoren (keine elektrische Leitfähigkeit, z.B. Glas) eingeteilt werden. Zwischen diesen beiden Extremen liegt eine Gruppe von Materialien mit einer geringen Leitfähigkeit bereits bei **Raumtemperatur**. Diese Materialien werden als **Halbleiter** bezeichnet. Eine genauere Untersuchung der physikalischen Eigenschaften von Metallen, Halbleitern und Isolatoren zeigt weitere qualitative Eigenschaften, die für die jeweilige Gruppe charakteristisch sind:

Eigenschaft	Isolator	Halbleiter	Metall
Elektrischer Widerstand:	sehr hoch	mittel	sehr klein
Temperatur Koeffizient	schwach	stark und negativ	schwach und positiv
Dotierungsabhängig	nein	sehr stark	schwach
Beleuchtungs-abhängig	ja, typisch bei UV-Licht falls $\hbar\omega > \hbar\omega_{cr}$	ja, typisch bei IR-Licht falls $\hbar\omega > \hbar\omega_{cr}$	nein
Optisch	transparent	grau, schwarz	reflektierend
Mechanisch	hart, spröde	nicht so hart, spröde	elastisch formbar
Beispiel	Diamant, Glas, Quarzglas	Si, Ge, GaAs, InP GaN, SiC	Ag, Au, Cu, Fe, Li, Na

Betrachten wir das Periodensystem der Elemente, siehe Abbildung, so sehen wir, dass die Halbleiter entweder aus der IV. Hauptgruppe (Si, Ge, SiC) oder zusammengesetzt aus jeweils einem Element der III.Hauptgruppe und der V.Hauptgruppe sind. (Es gibt auch II-VI Halbleiter.)



Quelle: Wikipedia, Hauptgruppe ergänzt

Spezifischer elektrischer Widerstand

Ein elektrisches Feld \vec{E} im Inneren eines Materials bewirkt einen Strom von beweglichen elektrischen Ladungen q . Die pro Zeit und Querschnitt transportierte Ladung wird **elektrische Stromdichte** \vec{j} genannt. Der elektrische **Leitwert** σ verknüpft die Stromdichte \vec{j} mit dem angelegten elektrischen Feld \vec{E} ,

$$\vec{j} = \sigma \vec{E}. \quad (1)$$

Der resultierende Strom ist folglich direkt proportional zum angelegten Feld. Bei isotropen Medien ist der elektrische Leitwert die zugehörige Proportionalitätskonstante. (N.B. bei anisotropen Medien ist der Leitwert in jede Richtung unterschiedlich und wird durch eine 3×3 -Matrix, dem sogenannten Leitwert-Tensor repräsentiert). Der **spezifische elektrische Widerstand** ist der Kehrwert des Leitwertes, σ^{-1} , und wird in Ωm gemessen. (Der Ohm'schen Widerstand R einer Leitung mit Querschnittsfläche A und Länge l steigt proportional der Länge und sinkt umgekehrt proportional zur Fläche, $R = \sigma^{-1}l/A$.)

Der spezifische Widerstand ist charakteristisch für jedes Material, mit einer Temperaturabhängigkeit,

$$\sigma(T)^{-1} = \sigma_0^{-1} [1 + \alpha(T - T_0)]. \quad (2)$$

α wird **Temperaturkoeffizient** genannt und gibt an, wie stark sich der Widerstand der Temperatur T vom spezifischen Widerstand σ_0^{-1} bei der Temperatur T_0 unterscheidet. In der folgenden Tabelle sind die gemessenen spezifischen elektrischen Widerstände σ_0^{-1} bei Raumtemperatur $T_0 = 20^\circ\text{C}$ mit den zugehörigen Temperaturkoeffizienten für exemplarische Leiter, Halbleiter und Isolatoren angeführt.

Material		spez. Widerstand $\sigma_0^{-1} [\Omega m]$	Temp. Koeff. $\alpha [K^{-1}]$
Metalle	Ag	1.59×10^{-8}	3.8×10^{-3}
	Cu	1.68×10^{-8}	4.0×10^{-3}
	Au	2.44×10^{-8}	3.4×10^{-3}
	Al	2.65×10^{-8}	3.9×10^{-3}
Halbleiter	Si (rein)	2.3×10^3	-75×10^{-3}
	Si (n-Typ P $10^{23} m^{-3}$)	8.7×10^{-4}	
	Si (p-Typ B $10^{23} m^{-3}$)	2.1×10^{-3}	
	GaAs (rein)	10^8	
Isolat.	Glas	10^{10} to 10^{14}	
	Quarzglas	10^{16}	

Der elektrische Widerstand σ^{-1} der reinen Halbleiter liegt bei Raumtemperatur 11-16 Größenordnungen über dem Widerstand typischer Metalle, aber immer noch einige Größenordnungen unter dem Widerstand typischer Isolatoren. Werden dem Si-Halbleiter 'Verunreinigungen' in Form von Phosphor-Fremdatomen (10^{23} pro m^3) oder Bor-Fremdatomen (10^{23} pro m^3) beigemengt, sinkt der Widerstand bei Raumtemperatur auf ein Millionstel. Die elektrischen Eigenschaften können also durch Zugabe von Fremdatome (genannt **Dotierung**) deutlich beeinflusst und daher durch entsprechende Fertigungsschritte definiert werden.

Die Temperaturabhängigkeit von Halbleitern ist ca. 20-mal stärker als bei Metallen und weist einen **negativen Temperaturkoeffizienten** auf. Das bedeutet, dass Halbleiter, im Gegensatz zu metallischen Leitern, bei steigenden Temperaturen den Strom immer besser leiten.

Ziel der Vorlesung wird es sein, die Mechanismen hinter diesen offensichtlichen Unterscheidungsmerkmalen von Halbleitern auf einem fundamentalen Niveau zu verstehen. Basierend auf diesem Wissen wird das Verhalten von **Dioden**, **Bipolar-Transistoren** und **MOS-Transistoren** hergeleitet.

2 – Physikalische Grundlagen

2.1 – Mechanik

Im Rahmen der klassischen **Newton'schen Mechanik** ist der **Zustand** eines Systems von N Teilchen durch die Angabe der Position \vec{q} und des Impulses \vec{p} jedes Teilchens vollständig bestimmt. Der **Phasenraum** wird durch alle Positionsvektoren und Impulsvektoren aufgespannt, und daher ein $6N$ -dimensionaler Raum. Jeder mögliche Zustand entspricht in eindeutiger Weise einem Punkt im Phasenraum.

Die zeitliche Entwicklung eines Zustandes beschreibt eine **Trajektorie** in diesem Phasenraum und wird durch die **Drei Newton'schen Axiome** beschrieben:

- 1 Ein Körper verharrt im Zustand der Ruhe oder gleichförmigen Translation solange keine Kraft einwirkt
- 2 Die Summe der einwirkenden Kräfte \vec{F} bewirkt eine zeitliche Änderung des Impulses \vec{p} gemäß

$$\frac{d}{dt} \vec{p} = \vec{F} \quad (3)$$

- 3 Jede Kraft von Körper A auf Körper B bewirkt eine gleichgroße entgegengerichtete Kraft von B auf A

Konservative Kräfte (z.B. die Gravitation oder elektrostatische Kraft) können als räumliche Ableitung (**Gradient**) einer skalaren Funktion der Koordinaten $V(\vec{x})$, der **Potentiellen Energie**, geschrieben werden,

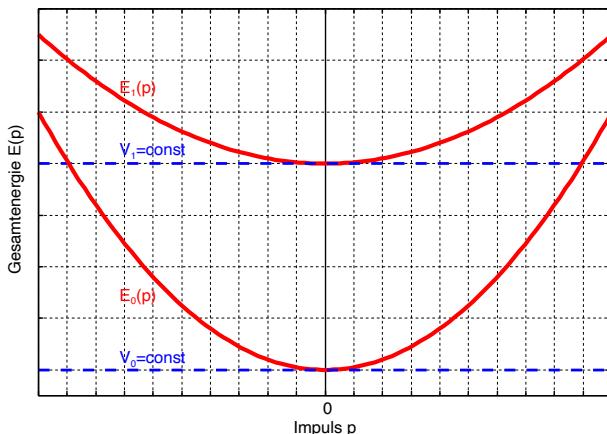
$$\vec{F}(\vec{x}) = -\vec{\nabla}V(\vec{x}). \quad (4)$$

N.B.: der Nabla-Operator $\vec{\nabla}$ ist die Verallgemeinerung der räumlichen Ableitung in alle Raumrichtungen und kann formal als Vektor geschrieben werden, dessen x-Komponente der partiellen Ableitung nach x , die y-Komponente nach y und die z-Komponente nach z entspricht, also $\vec{\nabla} := (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$.

Die Integration des zweiten Newton'schen Axioms, Gleichung (3) entlang der Trajektorie eines Teilchens in einem konservativen Kraftfeld führt zur Erhaltungsgröße der **Energie** des Teilchens im Zustand (\vec{q}, \vec{p}) :

$$E(\vec{p}, \vec{q}) = \frac{p^2}{2m} + V(\vec{q}) \quad (5)$$

$p^2 = \vec{p} \cdot \vec{p}$, m ist die Masse des Teilchens, der Impuls für nicht-relativistische Teilchen lautet $\vec{p} := m \frac{d}{dt} \vec{q} \equiv m \vec{v}$.



Ein (Kräfte-)**freies Teilchen**, d.h. ein Teilchen ohne äußere Krafteinwirkung bewegt sich in einem konstanten Potential, d.h. konstanter potentieller Energie. Die **kinetische Energie** entspricht einer Parabel in p ,

$$E_{kin} = \frac{p^2}{2m} \quad (6)$$

Die Gesamtenergie zweier freier Teilchen mit unterschiedlicher Masse und potentieller Energie ist im **Energie-Impuls Diagramm** (siehe Abbildung) dargestellt.

Die potentielle Energie entspricht der Gesamtenergie bei Impuls $p \equiv 0$, die **Krümmung der Energie-Parabel** entspricht dem Kehrwert der **Teilchenmasse**,

$$\frac{1}{m} = \frac{\partial^2}{\partial p^2} E(p, q) \quad 1 - \dim \quad (7)$$

Bei 3-dim Bewegungen wird nach jeder Impulskomponente abgeleitet, $\frac{1}{m} = \left[\frac{\partial^2}{\partial p_x^2} + \frac{\partial^2}{\partial p_y^2} + \frac{\partial^2}{\partial p_z^2} \right] E(\vec{p}, \vec{q})$.

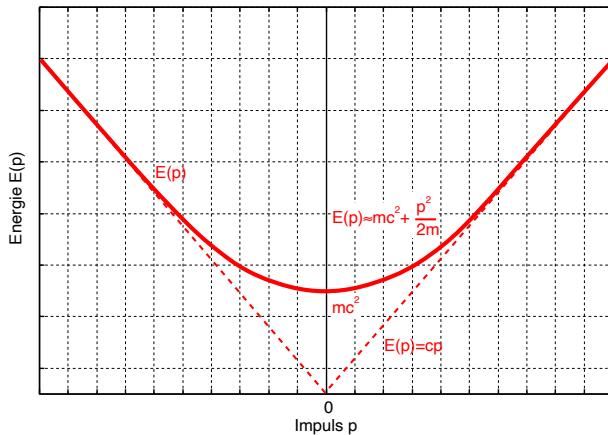
Aufgabe 1: Welche der oben dargestellten Energiefunktionen entspricht dem schwereren Teilchen?

Relativistische Mechanik

Im Rahmen der **Speziellen Relativitätstheorie** (A. Einstein 1905) wird die Newton'sche Mechanik für Geschwindigkeiten bis zur Lichtgeschwindigkeit erweitert. Die Details können hier nicht^a behandelt werden. Die **relativistische Energie-Impuls-Beziehung** für ein freies Teilchen ($V(\vec{q}) \equiv 0$) mit Masse m lautet

$$E^2 = m^2 c^4 + c^2 p^2 \quad (8)$$

$c = 2.9979 \times 10^8 \text{ m/s}$ ist die Vakuumlichtgeschwindigkeit.



Für kleine Impulse (langsame Geschwindigkeiten) reduziert sich das **Energie-Impuls-Diagramm** auf die klassische Newton'sche Energie, mit einer **Ruhe-Energie** von mc^2 , siehe Abbildung,

$$E \approx mc^2 + \frac{p^2}{2m}. \quad (9)$$

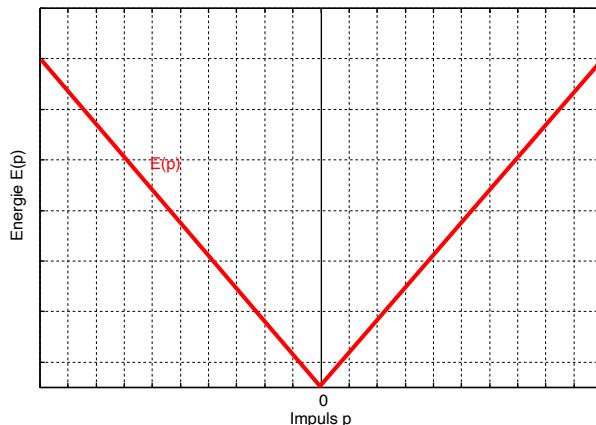
Bei großen (positiven oder negativen) Impulsen ergibt sich eine lineare Beziehung zwischen Energie und Impuls, im krassen Gegensatz zur Newton'schen Mechanik.

Im Falle eines massiven Teilchens lauten die relativistischen Verallgemeinerungen für Energie und Impuls

$$E = \frac{mc^2}{\sqrt{1 - (\frac{v}{c})^2}}, \quad \vec{p} = \frac{mv}{\sqrt{1 - (\frac{v}{c})^2}}. \quad (10)$$

Im Falle von **masselosen Teilchen**, wie z.B. bei Lichtteilchen (genannt **Photonon**) muss die relativistische Energie-Impulsbeziehung, Gleichung (8), verwendet werden und es folgt

$$E^2 = c^2 p^2 \quad (11)$$



Masselose Teilchen bewegen sich mit Lichtgeschwindigkeit c . Die Energie E ist, bis auf eine Konstante, gleich dem Betrag des Impulses $p = |\vec{p}|$,

$$E = c|\vec{p}|, \quad (12)$$

siehe Abbildung.

Energie-Impuls-Beziehungen werden auch **Dispersionsrelationen** genannt, wie später noch genauer erläutert wird.

Aus den vorhergegangenen Beispielen sollte klar sein, dass an der Form der Dispersionsrelation vieles über das betrachtete Teilchen ausgesagt werden kann. (Masse, Geschwindigkeit, potentielle Energie,...)

Aufgabe 2: Reduzieren sich die relativistischen Energien und Impulse, Gleichung (10) bei Geschwindigkeiten $v \ll c$ auf die Definitionen der Newton'schen Physik? Welche 'effektive Masse' folgt für relativistische Teilchen der Ruhemasse m aus der Energie-Impulsbeziehung (Dispersionsrelation)?

^aSiehe z.B. L. Susskind und A. Friedman: Elektrodynamik und Relativität: Das theoretische Minimum

2.2– Elektromagnetismus

Elektrostatik, Elektrisches Feld

Ein Beispiel für eine konservative Kraft ist die elektrische Anziehungskraft (**Coulomb-Kraft**) zwischen zwei elektrischen Ladungen, q und q_0 . Nehmen wir an, die Ladung q_0 sei im Koordinatenursprung fixiert. Die Kraft auf die Ladung q an der Position \vec{x} lautet in diesem Fall (ϵ ist die Dielektrizitätskonstante):

$$\vec{F}(\vec{x}) = q \frac{q_0}{4\pi\epsilon|\vec{x}|^2} \frac{\vec{x}}{|\vec{x}|} \quad (13)$$

Die Kraft wirkt demnach entlang der Verbindungsgeraden der beiden Ladungen, bei gleichen Vorzeichen der Ladungen abstoßend bei unterschiedlichen Vorzeichen anziehend, und nimmt mit dem Quadrat des Abstandes ab. Die Kraft auf jede Ladung q an der Position \vec{x} relativ zu q_0 ist folglich gegeben durch einen ortsabhängigen Vektor, genannt **Vektorfeld** $\vec{E}(\vec{x})$, multipliziert mit der Ladung q ,

$$\vec{F}(\vec{x}) = q\vec{E}(\vec{x}). \quad (14)$$

Bei einer Konstellation aus vielen statischen Ladungen q_i im Raum summieren sich die Kräfte aller Ladungen auf q vektoriell und wir können die resultierende Kraftwirkung wieder analog zu Gleichung (14) schreiben. Das Vektorfeld, welche die Summe der Kräfte pro Ladung beschreibt, nennt man **Elektrisches Feld** $\vec{E}(\vec{x})$.

Im Grenzfall sehr vieler Ladungen im Raum kann die Ladungsverteilung durch eine skalare Funktion, die **Ladungsdichte** $\rho(\vec{x})$, dargestellt werden, welche jedem betrachteten kleinen Volumen ΔV die darin enthaltenen Gesamtladung $\Delta q = \rho\Delta V$ zuordnet. Wie im Fall der einzelnen Ladung, ist auch im Falle einer Ladungsverteilung $\rho(\vec{x})$ das resultierende elektrische Feld in eindeutiger Weise bestimmt. Die Ladungsdichte $\rho(\vec{x})$ definiert die Quellstärke des elektrischen Feldes, gemäß des Gauß'schen Gesetzes (für $\epsilon \equiv \text{konst}$)

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon}. \quad (15)$$

N.B.: Die Anwendung des Nabla-Operators $\vec{\nabla}$ auf ein Vektorfeld \vec{E} wird **Divergenz** genannt und entspricht formal dem Skalarprodukt, d.h. $\vec{\nabla} \cdot \vec{E} \equiv \frac{\partial}{\partial x} E_x + \frac{\partial}{\partial y} E_y + \frac{\partial}{\partial z} E_z$.

Jede konservative Kraft kann als Gradient einer skalaren Funktion ausgedrückt werden. Daher ist auch das elektrostatische Feld (also Kraft pro Ladung) der Gradient einer skalaren Funktion (welche analog die potentielle Energie pro Ladung darstellt). Diese Funktion heißt **Elektrisches Potential** $\Phi(\vec{x})$ und es gilt

$$\vec{E}(\vec{x}) =: -\vec{\nabla}\Phi(\vec{x}). \quad (16)$$

Bei gegebenem elektrischen Feld errechnet sich das Potential $\Phi(\vec{x})$ analog zur potentiellen Energie

$$\Phi(\vec{x}) := - \int_{\vec{x}_0}^{\vec{x}} \vec{E}(\vec{x}') \cdot d\vec{x}' \quad (17)$$

Elektrische Potentiale sind nur bis auf ein beliebige Konstante bestimmt, messbar ist nur die Ableitung, das elektrische Feld. Die Potentialdifferenz zwischen zwei Punkten wird **Elektrische Spannung** U_{12} genannt:

$$U_{12} := \Phi(\vec{x}_1) - \Phi(\vec{x}_2) = \int_{\vec{x}_1}^{\vec{x}_2} \vec{E}(\vec{x}') \cdot d\vec{x}' \quad (18)$$

Die Bestimmungsgleichung für das Potential bei gegebener Ladungsdichte $\rho(\vec{x})$, die **Poisson Gleichung**, folgt durch Einsetzen der Gleichung (16) in das Gauß'sche Gesetz (15) (für räumlich konstantes ϵ):

$$\Delta\Phi(\vec{x}) = -\frac{\rho(\vec{x})}{\epsilon}. \quad (19)$$

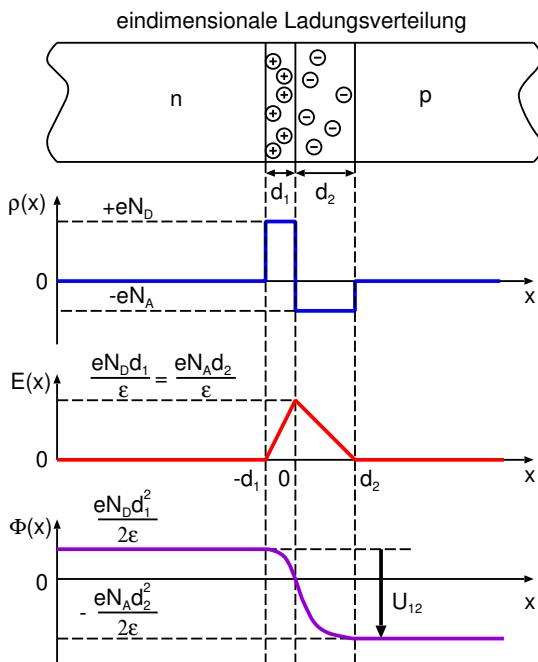
N.B.: Das Quadrat des Nabla-Operators $\vec{\nabla} \cdot \vec{\nabla}$ stellt eine räumliche Verallgemeinerung der zweiten Ableitung dar, genannt **Laplace-Operator**, und mit $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ abgekürzt.

Aufgabe 3: Eine 1-dimensionale Ladungsdichte $\rho(x)$ sei konstant ρ_0 für $x \in [-a, a]$ und 0 für $|x| > a$. Welches Potential und welches elektrische Feld resultiert?

Beispiel: 1-dimensionaler PN-Übergang

Als Beispiel betrachten wir eine räumliche Ladungsverteilung, welche sich nur in x -Richtung ändert, eine sogenannte 1-dimensionale Ladungsdichte $\rho(x)$. Die Gesamtanordnung sei elektrisch neutral, besteht aber aus einem Bereich homogener positiver Teilchendichte N_D der Breite d_1 , angrenzenden an einen Bereich homogener negativer Teilchendichte N_A der Breite d_2 , siehe Abbildung. Die Ladungsdichte im positiven Bereich ist folglich eN_D , im negativen Bereich $-eN_A$, mit der **Elementarladung** $e = 1.602 \times 10^{-19} C$ der geladenen Teilchen. Die beiden Ladungen heben sich gerade auf und es gilt

$$d_1 N_D - d_2 N_A = 0. \quad (20)$$



In den neutralen Bereichen wirkt keine elektrische Kraft auf freie Ladungen, das elektrische Feld muss also verschwinden,

$$\vec{E}(-d_1) = \vec{E}(d_2) \equiv 0. \quad (21)$$

Die Poisson-Gleichung, (19), lautet daher:

$$\frac{d^2\Phi(x)}{dx^2} = \begin{cases} -\frac{eN_D}{\epsilon}, & -d_1 \leq x \leq 0 \\ \frac{eN_A}{\epsilon}, & 0 \leq x \leq d_2 \end{cases} \quad (22)$$

Durch zweimalige Integration erhalten wir $\Phi(x)$ als x^2 Parabeln mit zwei Integrationskonstanten. Das Potential bei $x = 0$ muss stetig sein, und kann ohne Beschränkung der Allgemeinheit (beliebige Konstante des Potentials) auf $\Phi(0) \equiv 0$ gesetzt werden.

$$\Phi(x) = \begin{cases} -\frac{eN_D}{2\epsilon}x^2 + a_1x, & -d_1 \leq x \leq 0 \\ \frac{eN_A}{2\epsilon}x^2 + a_2x, & 0 \leq x \leq d_2. \end{cases} \quad (23)$$

Die verbleibenden Integrationskonstanten a_1 und a_2 folgen aus den Randbedingungen (21),

$$0 \equiv \vec{E}(-d_1) = -\frac{d}{dx}\Phi(x)|_{-d_1} = -\frac{eN_D}{\epsilon}d_1 - a_1 \quad (24)$$

$$0 \equiv \vec{E}(d_2) = -\frac{d}{dx}\Phi(x)|_{d_2} = -\frac{eN_A}{\epsilon}d_2 - a_2. \quad (25)$$

Der **Potentialverlauf** eines abrupten PN-Überganges ist in der Abbildung dargestellt und lautet

$$\Phi(x) = \begin{cases} -\frac{eN_D}{\epsilon} \left(d_1x + \frac{x^2}{2} \right), & \text{falls } -d_1 \leq x \leq 0 \\ -\frac{eN_A}{\epsilon} \left(d_2x - \frac{x^2}{2} \right), & \text{falls } 0 \leq x \leq d_2 \end{cases} \quad (26)$$

Die Annahme der fixen Ladungsdichten wird später erläutert. Am PN-Übergang bildet sich eine fixe Spannung, genannt **Diffusionsspannung** aus, welche von den Teilchendichten N_D und N_A abhängt.

$$U_D = \frac{e}{2\epsilon} (N_D d_1^2 + N_A d_2^2) = \frac{e N_D d_1^2}{2\epsilon} \left(1 + \frac{d_2}{d_1} \right). \quad (27)$$

In der letzten Beziehung wurde die Ladungsneutralität, Gleichung (20) verwendet. Das Diffusionspotential ist wesentlich für die Funktion von Halbleiter-Dioden und definiert die Durchlassrichtung.

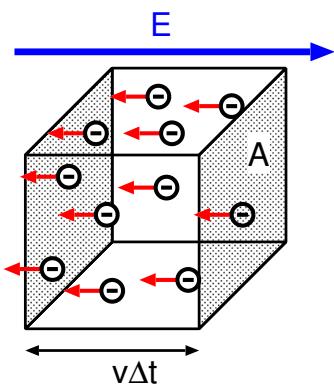
Aufgabe 4: In welche Richtung zeigt das elektrische Feld? Warum darf das Potential keine Sprünge haben? Was passiert mit freien positiven oder negativen Ladungen, die in die (Raum)-Ladungszone des PN-Überganges kommen?

Elektromagnetische Kräfte

Die elektrische Kraft auf ein Teilchen der Ladung q im homogenen Feld \vec{E} ist gegeben durch Gleichung (14). Betrachten wir nun ein Elektron der Masse m und der Ladung $q = -e$, wobei die **Elementarladung** $e := 1.602 \times 10^{-19} C$ eine Naturkonstante ist. Aus dem zweiten Newton'schen Axiom (3) folgt eine gleichmäßig beschleunigte Bewegung,

$$\vec{p}(t) = -e\vec{E}t + \vec{p}(0), \quad (28)$$

also einen zeitlich linear anwachsenden Impuls, relativ zum Anfangsimpuls $\vec{p}(0)$, entgegengesetzt der Richtung des elektrischen Feldes. Entsprechend wächst auch die Geschwindigkeit gemäß $\vec{v}(t) = -\frac{e}{m}\vec{E}t + \vec{v}(0)$.



Betrachten wir ein Quader eines Materials mit der Elektrondichte n . Die Ladung jedes Elektrons beträgt $q = -e$, und somit die **Ladungsdichte**

$$\rho := qn \quad (29)$$

Durch das Anlegen eines homogenen elektrischen Feldes in Richtung der Quaderkanten bewegen sich alle Elektronen mit Geschwindigkeit \vec{v} entgegen der Feldrichtung (negative Ladung), siehe Abbildung. Pro Zeit Δt wird folglich ein Elektronenvolumen $Av\Delta t$ entgegen der Feldrichtung bewegt. Die Gesamtladung in diesem Volumen ist $\Delta Q = \rho Av\Delta t$.

Die pro Querschnittsfläche A und Zeit Δt strömende Gesamtladung ΔQ wird, wie bereits erwähnt, **Stromdichte** \vec{j} genannt.

Für eine Menge Teilchen mit Ladungsdichte ρ , die sich jeweils mit Ladung q im Mittel mit Geschwindigkeit $\langle \vec{v} \rangle$ bewegen lautet daher die **elektrische Stromdichte**

$$\vec{j} = \rho \langle \vec{v} \rangle. \quad (30)$$

Die Annahme völlig wechselwirkungsfreier Elektronen ergibt eine zeitlich beliebig anwachsende Stromdichte - offensichtlich im Widerspruch zur Leitung in Metallen, bei der laut Ohm'schem Gesetz, Gleichung (1) die Stromdichte bei konstantem Feld \vec{E} ebenfalls konstant bleibt. Wir werden später im Drude-Modell sehen, dass die Wechselwirkung der Elektronen mit den Atomen des Materials gerade zu dieser, im Mittel zeitlich konstanten, Geschwindigkeit und daher Stromdichte führt.

Wirkt zusätzlich zu einem elektrischen Feld \vec{E} auch ein magnetisches Feld \vec{B} , werden bewegte Teilchen durch das Magnetfeld orthogonal zu den magnetischen Feldlinien abgelenkt, gemäß der **Lorentz-Kraft**:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}). \quad (31)$$

Betrachten wir den Fall eines Teilchens der Ladung q und Masse m , welches sich mit Geschwindigkeit \vec{v} orthogonal zu einem homogenen Magnetfeld $\vec{B} \equiv \text{konst}$ bewegt (das elektrische Feld sei null, $\vec{E} \equiv 0$). Die Lorentzkraft wirkt, aufgrund des Vektorproduktes, in eine Richtung orthogonal zum Magnetfeld \vec{B} und orthogonal zur Flugrichtung \vec{v} . Während die Richtung orthogonal zum Magnetfeld zeitlich konstant bleibt, ändert die Geschwindigkeit aufgrund der Lorentzkraft ständig ihre Richtung - die Lorentzkraft wirkt folglich als reine Zentripetalkraft $\vec{F}_{ZP} = m\vec{\omega} \times \vec{v}$ mit Betrag qvB und das Teilchen wird auf eine Kreisbahn gebracht, mit Kreisfrequenz

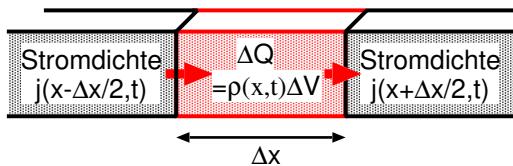
$$\omega = \frac{qB}{m}. \quad (32)$$

Der Drehsinn dieser Kreisbahn hängt von Vorzeichen der Ladung q ab. (Warum?)

Aufgabe 5: Betrachte ein homogenes Magnetfeld in $z-$ -Richtung, in das ein Elektron mit Geschwindigkeit v in negativer x -Richtung eingeschossen wird. In welche Richtung wird das Elektron abgelenkt?

2.3– Kontinuitätsgleichung

Elektrische Ströme transportieren Ladungen entsprechend der Stromdichte \vec{j} und der Querschnittsfläche A orthogonal zur Stromrichtung. Diese Ströme können sowohl zeitlich, als auch räumlich variieren und führen folglich zu einer räumlich und zeitlich veränderlichen Ladungsdichte $\rho(\vec{x}, t)$. Falls in ein betrachtetes Raumvolumen mehr Ströme hinein- als hinausfließen, wird die Ladungsdichte in diesem Volumen steigen und umgekehrt bei einem Netto-Abfluss von Ladungen. Die Stromdichte im Raum $\vec{j}(\vec{x}, t)$ ist daher unmittelbar mit der Ladungsdichte im Raum $\rho(\vec{x}, t)$. Der Zusammenhang ist die sogenannte **Kontinuitätsgleichung** welche im Folgenden hergeleitet wird.



Betrachten wir ein Volumen $\Delta V = A\Delta x$ mit der Ladungsdichte $\rho(\vec{x}, t)$, siehe Abbildung. Eine räumlich und zeitlich variable Stromdichte $\vec{j}(\vec{x}, t)$ führt zu einem Anstieg der Ladung im Volumen durch den von links durch die Fläche A in das Volumen laufende Strom und wird reduziert durch den entsprechend rechts aus dem Volumen laufenden Strom.

Unter der Annahme, dass im Inneren des Volumen weder Ladungen erzeugt noch vernichtet werden, muss folgende Bilanzgleichung für den Anstieg der Ladung ΔQ im Volumen gelten:

$$\frac{d}{dt}\Delta Q \equiv \Delta V \frac{\partial}{\partial t} \rho(\vec{x}, t) \stackrel{!}{=} A \left[j_x(x - \frac{\Delta x}{2}, y, z, t) - (j_x(x + \frac{\Delta x}{2}, y, z, t)) \right] \quad (33)$$

Im dargestellten Fall ist nur die Stromdichte in x -Richtung relevant, also nur die x -Komponente j_x . In das Volumen fliesst dieser Strom entsprechend der Stromdichte eine halbe Volumsbreite $\Delta x/2$ zu negativen x -Werten verschoben, aus dem Volumen fliesst die x -Komponente des Stromes an der x -Koordinate $x + \Delta x/2$. Im Grenzwert eines sehr schmalen betrachteten Volumens, also $\Delta x \rightarrow 0$ können die Ströme durch die niedersten Terme eine Taylor-Reihe approximiert werden,

$$j_x(x \pm \frac{\Delta x}{2}, y, z, t) \approx j_x(x, y, z, t) \pm \frac{\Delta x}{2} \frac{\partial}{\partial x} j_x(x, y, z, t). \quad (34)$$

Eingesetzt in die Bilanzgleichung (33) kann das Volumen gekürzt werden und wir erhalten den gesuchte Zusammenhang zwischen der lokalen Ladungsdichte und der Stromdichte,

$$\frac{\partial}{\partial t} \rho(\vec{x}, t) = - \frac{\partial}{\partial x} j_x(x, y, z, t). \quad (35)$$

Die Komponenten der Stromdichte in y - und z -Richtung müssen sich aus Symmetriegründen analog verhalten und wir erhalten die 3-dimensionale **Kontinuitätsgleichung**

$$\frac{\partial}{\partial t} \rho(\vec{x}, t) = - \vec{\nabla} \cdot \vec{j}(\vec{x}, t), \quad (36)$$

für den Fall einer konstant gehaltenen Anzahl an Ladungsträgern.

N.B.: Auch hier tritt wieder die 3-dimensionale räumliche Ableitung als **Divergenz** auf. Entsprechend der Intuition, nimmt die Ladungsdichte an einem Punkt zeitliche ab, wenn an dieser Stelle die Ströme 'in verschiedene Richtungen auseinandergehen'. Das Lateinische Wort für 'auseinandergehen' lautet *divergere*.

Die Kontinuitätsgleichung beschreibt die Dynamik der freien positiven und der negativen Ladungsverteilungen im Halbleiter. Die Ströme können entweder durch elektrische Felder bewirkt werden, sogenannte **Driftströme**, oder durch Konzentrationsunterschiede und Diffusion, sogenannte **Diffusionsströme**. Zusätzlich können in Halbleitern Ladungen durch einfallendes Licht freigesetzt und dadurch freie Ladungen erzeugt werden (z.B. PV-Zelle) oder durch Bindung an die Atome eingefangen und dadurch freie Ladungen reduziert werden (unter Freisetzung z.B. von Licht in der LED), siehe später.

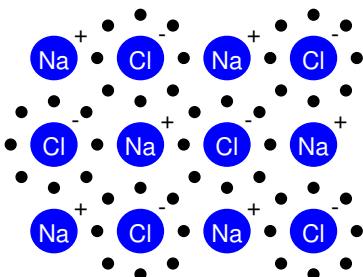
Aufgabe 6: Die Kirchhoff'sche Knotenregel ist gerade die Anwendung der Kontinuitätsgleichung auf elektrische Ströme. Warum? Wie folgt die Knotenregel aus der Kontinuitätsgleichung?

3 – Freie Ladungen

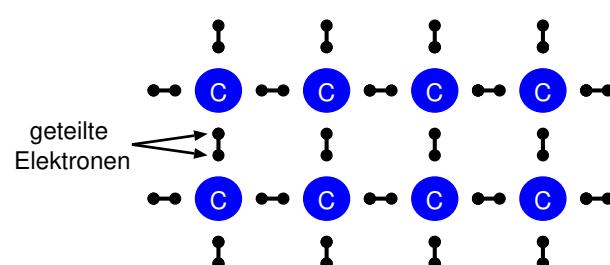
Elektrische Leitung erfolgt über frei bewegliche Ladungsträger im Medium. Je nach Art der chemischen Bindung im Festkörper sind diese Ladungsträger (bei Raumtemperatur) verfügbar. Es werden verschiedene Arten der chemischen Bindungen unterschieden, siehe Abbildung, die grob in 4 Arteneingeteilt werden:

- **Ionenbindung** Ein chemisches Element (typischerweise aus der I. Hauptgruppe, den Alkalimetallen) mit einem nur schwach gebundenen Elektron in der äußersten Hülle gibt dieses Elektron einem chemischen Element (typischerweise aus der VII Hauptgruppe, den Halogenen) mit einem Elektron weniger als der Edelgaskonstellation (Hauptgruppe VIII). Beide Atom besitzen nun eine Art Edelgashülle, sind dafür aber elektrisch geladen (Ionen). Die elektrostatischen Anziehungskräfte zwischen den Ionen verursachen die Bindung in einer regelmäßigen Kristallstruktur (z.B. Kochsalz).
- **Kovalente Bindungen** Atome mit unvollständig gefüllten äusseren Schalen teilen sich jeweils ein Elektron (oder auch mehrere) mit den unmittelbaren Nachbarn. Diese Elektronen können nun nicht mehr eindeutig einem der Atome zugeordnet werden und gehören in gewisser Weise zu beiden Atomen, die dadurch jeweils eine Art Edelgaskonfiguration erhalten (z.B. Diamantgitter, Si, GaAs,...).
- **Metallische Bindung** Die Elektronen der äussersten Schale können keinem Atom mehr zugeordnet werden und bewegen sich als quasi freie Elektronen zwischen den ionisierten Atomrumpfen (Atomkern und innere Schalen), welche typischerweise eine regelmäßige Struktur bilden (Metalle).
- **Molekulare Bindung** Unter diesem Begriff werden schwache elektrostatische Anziehungen zwischen elektrisch neutralen Molekülen zusammengefasst. Liegt der Ladungsschwerpunkt der Elektronenhülle eines Moleküls nicht an der gleichen Stelle wie der Ladungsschwerpunkt der beteiligten Atomkerne, spricht man von einem elektrischen Dipolmoment. Das Molekül hat folglich eine elektrisch positive und eine negative Seite. Die Seiten unterschiedlicher Polarität benachbarter Moleküle ziehen sich elektrostatisch an und bilden eine schwache Bindung (z.B. Van-der-Waals Kraft zwischen Wassermolekülen).

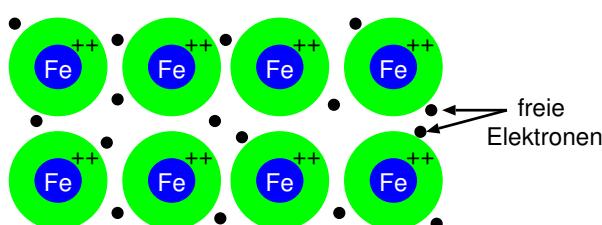
Ionenbindung
elektrische Anziehung zwischen ionisierten Atomen



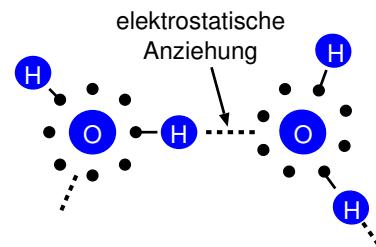
Kovalente Bindung
Atome teilen sich Elektronen zu gleichen Teilen



Metallische Bindung
Ionen umgeben von freien Elektronen



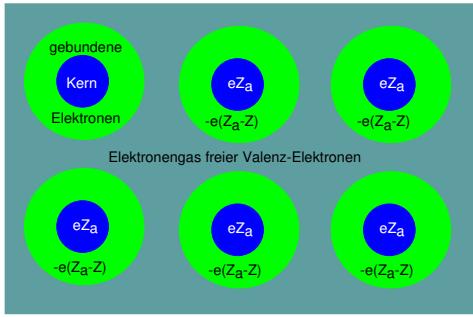
Molekulare Bindung
schwache elektrische Anziehung zwischen Molekülen



Aufgabe 7: Welche Bindung liegt bei Halbleitern vor? Warum gibt es III-V und II-VI Halbleiter?

3.1– Drude-Modell

Im Jahre 1900 veröffentlichte P. Drude eine Modell für Metalle, indem die äussersten Elektronen, genannt **Valenz-Elektronen**, ein **Elektronengas** freier Elektronen bilden. Die Atome eines Metalls mit Ordnungszahl Z_a und Z Valenz-Elektronen bilden ein starres Gitter. Die $Z_a - Z$ inneren Elektronen sind fest an das Atom gebunden und bilden zusammen mit dem Atomkern ionisierte Atomrümpfe. Die Z Valenz-Elektronen bilden ein Elektronengas, welches sich als ideales Gas zwischen den Ionen bewegt, siehe Abbildung



Folgende Annahmen werden getroffen:

- **Unabhängige Elektronen:** keine Elektron-Elektron-Wechselwirkung zwischen den Stößen
- **Freie Elektronen:** keine Elektron-Ionen-Wechselwirkungen zwischen den Stößen
- **Kinetische Gastheorie:** Das Elektronengas ist im thermischen Gleichgewicht durch Kollisionen mit der Stoßzeit τ_c .

Drude Model: Elektrische Leitfähigkeit

Die Atomrümpfe sind in Ruhe und definieren den mittleren Impuls $\langle \vec{p} \rangle \equiv 0$. Elektronen des Impulses $\vec{p}(t)$ stehen durch Stöße gemäß der kinetischen Gastheorie ein thermisches Gleichgewicht (Boltzmann-Verteilung). Die Stoßzeit τ_c definiert die Zeitskala in der sich das thermische Gleichgewicht entsprechend der **Relaxationsgleichung** einstellt (siehe Physik I).

In Anwesenheit einer äußeren Kraft \vec{F} lautet die Relaxationsgleichung:

$$\frac{d}{dt} \vec{p}(t) = -\frac{\vec{p}(t)}{\tau_c} + \mathbf{F} \quad (37)$$

Im Fall der **elektrischen Leitung** bewirkt ein elektrisches Feld im Material eine elektrische Stromdichte, d.h. auf die jedes Elektronen wirkt entsprechend der Ladung $q = -e$ eine elektrische Kraft $\mathbf{F} = -e\mathbf{E}$. Aus der Relaxationsgleichung (37) ergibt sich nun das Verhalten des Elektronimpulses $\vec{p}(t)$

$$\vec{p}(t) = \Delta \vec{p} e^{-\frac{t}{\tau_c}} - e\vec{E}\tau_c. \quad (38)$$

Wie erwartet, führen die Kollisionen (mit der typischen Zeitskala τ_c) zu einer exponentiellen Angleichung des mittleren Impulses an einen asymptotischen Wert, der durch das elektrische Feld bestimmt wird. Die mittlere Elektronengeschwindigkeit, genannt **Drift-Geschwindigkeit**, nach mehreren Kollisionen mit den Atomrumpfen, $\vec{v}_{\text{drift}} := \vec{p}(t \rightarrow \infty)/m$, lautet (m ist die Masse des Elektrons)

$$\vec{v}_{\text{drift}} = -\frac{e\vec{E}\tau_c}{m}, \quad (39)$$

und ist proportional zum elektrischen Feld. Die Proportionalitätskonstante wird auch **Mobilität** μ genannt,

$$\vec{v}_{\text{drift}} =: -\mu \vec{E}. \quad (40)$$

N.B.: Im Gegensatz zum freien Elektron (28) steigt der Impulse nicht beliebig an.

Die **elektrische Strom** im Drude Modell passt zum gemessenen Verhalten einer konstanten Stromdichte

$$\vec{j} = -ne\vec{v}_{\text{drift}} = \left(\frac{ne^2\tau_c}{m} \right) \vec{E} \quad (41)$$

Das Drude Modell liefert einen analytischen Ausdruck für die **Leitfähigkeit** $\sigma = \left(\frac{ne^2\tau_c}{m} \right)$. Der elektrische Strom aufgrund eines elektrischen Feldes ist der bereits erwähnte **Driftstrom**.

Aufgabe 8: Hängt die Leitfähigkeit vom Vorzeichen der elektrischen Ladung der Ladungsträger ab? Warum?

Drude-Modell: Elektronendichten und Stoßzeiten

Betrachten wir einige typische Metalle, mit der entsprechenden Anzahl Valenz-Elektronen, kann die **Dichte der freien Elektronen** bestimmt werden, siehe Tabelle:

Metall	Valenzen	$n[10^{28}/m^3]$	$r_s[nm]$
Li (78K)	1	4.7	0.172
Na (5K)	1	2.65	0.208
K (5K)	1	1.4	0.257
Rb (5K)	1	1.15	0.275
Cs (5K)	1	0.91	0.298
Cu	1	8.47	0.141
Ag	1	5.86	0.160
Au	1	5.90	0.159
Be	2	24.7	0.099
Fe	2	17.0	0.112
Ga	3	15.4	0.116
Al	3	18.1	0.110

Quelle: N. W. Ashcroft and D. D. Mermin, Solid State Physics

Aus den gemessenen Leitwerten können die, dem Drude-Modell entsprechenden, **Kollisionszeiten** berechnet werden, siehe Tabelle:

Metall	$\tau_c[10^{-14}s]$		
	$T = 77K$	$T = 273K$	$T = 373K$
Li	7.3	0.88	0.61
Na	17	3.2	
K	18	4.1	
Rb	14	2.8	
Cs	8.6	2.1	
Cu	21	2.7	1.9
Ag	20	4.0	2.8
Au	12	3.0	2.1
Be		0.51	0.27
Fe	3.2	0.24	0.14
Ga	0.84	0.17	
Al	6.5	0.8	0.55

Quelle: N. W. Ashcroft and D. D. Mermin, Solid State Physics

Die **Driftbewegung** ist folglich relativ zur thermischen Bewegung um Größenordnungen untergeordnet und tatsächlich nur eine langsame Drift.

Aufgabe 9: Passt die gemessene Temperaturabhängigkeit mit den Erwartungen aus der kinetischen Gastheorie zusammen?

Die **Elektronenmasse** beträgt

$$m = 9.1 \times 10^{-31} kg. \quad (42)$$

Um eine Vorstellung des typischen Abstandes zwischen zwei Leitungselektronen zu bekommen wir der Begriff des **Sphärischen Abstandes** r_s eingeführt. Hierzu wird jedem Leitungselektron eine Kugel mit Radius r_s zugeordnet, die sich berührenden Kugeln ergeben ein Volumen mit der gemessenen Elektronendichte,

$$r_s := \sqrt[3]{\frac{3}{4\pi n}}. \quad (43)$$

Der typische Abstand zwischen Elektronen beträgt folglich $0.2 - 0.6 nm$. Verglichen mit der Dichte von Luft bei $20^\circ C$ ist das Elektronengas ca. 1000 mal dichter!

Das Drude Modell liefert den Zusammenhang des Leitwertes zur Stoßzeit

$$\sigma = \left(\frac{ne^2 \tau_c}{m} \right), \quad (44)$$

und daher die typischen **Driftgeschwindigkeiten**,

$$\vec{v}_{\text{drift}} = -\frac{e\tau_c}{m} \vec{E}. \quad (45)$$

N.B.: Die Elektronen im Drude Modell sind ein Gas freie Teilchen, die Boltzmann-Verteilung gegebener Temperatur liefert die mittlere thermische Geschwindigkeit v_{thermal} (siehe Physik I). Während die thermische Bewegung zufällig und ungerichtet ist, bewirkt das elektrische Feld eine gerichtete Drift v_{drift} die der thermischen Bewegung überlagert ist. Bei Raumtemperatur und typischen Spannungen von einigen V ergibt sich in Kupfer:

- $v_{\text{thermal}} \approx 10^6 m/s$
- $v_{\text{drift}} \approx 10^{-7} m/s$

Drude-Modell: Frequenzabhängigkeit des Leitwertes und Plasmaoszillationen

Wird ein Metall mit einem zeitlich periodischen elektrischen Feld einer gegebenen Frequenz ω angeregt, also $\vec{E}(t) = \vec{E}_\omega \exp(-i\omega t)$, reagiert der Strom mit der gleichen Frequenz (siehe Resonanz, Physik II),

$$\vec{j}(t) = \vec{j}_\omega \exp(-i\omega t). \quad (46)$$

N.B. Wir rechnen hier mit komplexen Größen, die komplexen Amplituden zur betrachteten Kreisfrequenz werden mit \vec{E}_ω bzw. \vec{j}_ω bezeichnet. Physikalisch relevant ist jeweils der Realteil der komplexen Größe, also eine reine Cosinus-Schwingung mit Phasenverschiebung.

Die Stromdichte der Elektronendrift, Gleichung (30), steht in direktem Zusammenhang zum Elektronenmpuls $\vec{p} = -\frac{m}{en}\vec{j}$ welcher im elektrischen Wechselfeld der Relaxationsgleichung (37) genügen muss. Wir erhalten eine Gleichung, welche die komplexe Stromdichte mit dem komplexen Feld verknüpft, $\vec{j}_\omega =: \sigma(\omega)\vec{E}_\omega$ mit dem komplexen Leitwert

$$\sigma(\omega) = \left(\frac{ne^2\tau_c}{m} \right) \frac{1}{1 - i\omega\tau_c} \quad (47)$$

Für kleine Frequenzen ist der Leitwert gleich dem statischen Leitwert $\frac{ne^2\tau_c}{m}$. Sobald die Anregungsfrequenz ω schneller als die typische Frequenz zwischen Kollisionen ist, $\omega\tau_c \gg 0$, kann der Strom der Anregung nicht mehr folgen und hinkt um eine Viertelperiode hinterher,

$$\sigma(\omega) \approx i \left(\frac{ne^2}{m\omega} \right) \quad (48)$$

Diese Vorhersage des Drude-Modells kann in jedem Metall beobachtet werden. Wir betrachten hierzu die Wellengleichung des elektrischen Feldes (siehe Physik II) in Anwesenheit von elektrischen Strömen ^a

$$\Delta \vec{E} = \mu\epsilon \frac{\partial^2}{\partial t^2} \vec{E} + \mu \frac{\partial}{\partial t} \vec{j}. \quad (49)$$

Drückt man die frequenzabhängige Stromdichte durch den komplexen Leitwert und das komplexe Feld $\vec{E}(t)$, erhalten wir die Wellengleichung für die Amplitude des elektrischen Feldes,

$$\left[\Delta + \left(\frac{\omega}{c} \right)^2 \left(1 + i \frac{\sigma}{\epsilon\omega} \right) \right] \vec{E}(\omega) = 0. \quad (50)$$

Die erhaltene Gleichung der **Plasma-Oszillationen** ist gerade eine Wellengleichung für das elektrische Feld, mit einer frequenzabhängigen relativen Dielektrizitätskonstanten ϵ_r

$$\epsilon_r(\omega) = 1 + i \frac{\sigma}{\epsilon\omega} =: 1 - \left(\frac{\omega_p}{\omega} \right)^2, \quad (51)$$

wobei im rechten Ausdruck der hochfrequente Leitwert (48) eingesetzt wurde. Für einfallende Ebene Wellen mit Wellenvektor $\propto \exp(i\vec{k} \cdot \vec{x})$ erhalten wir die Dispersionsrelation für den Wellenvektor im Metall

$$k \equiv \frac{\omega}{c} \sqrt{1 - \left(\frac{\omega_p}{\omega} \right)^2} \quad (52)$$

Im Fall von Frequenzen unterhalb der sogenannten **Plasma Frequenz** $\omega_p := e\sqrt{n/\epsilon m}$ werden die Wellenvektoren rein imaginär und folglich die Welle zu einer räumlich exponentiell abklingenden Welle. Daher kann sich die elektromagnetische Welle nicht in das Metall ausbreiten. Die Elektronen an der Oberfläche folgen der Anregung und erzeugen durch ihre Bewegung eine elektromagnetische Welle - das Metall reflektiert die einfallende Welle vollständig. Bei Frequenzen oberhalb der Plasmafrequenz sind reelle Wellenvektoren möglich und elektromagnetische Wellen breiten sich durch das Metall aus - das Metall wird transparent.

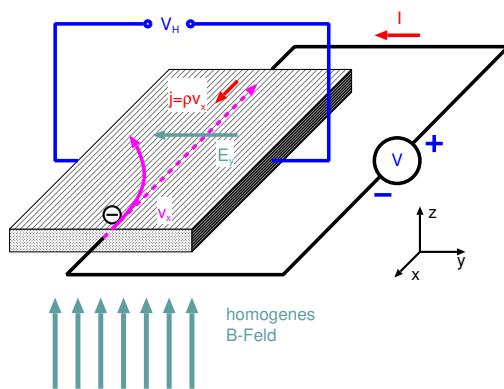
Aufgabe 10: Die Anschlüsse der Pixel eines LCD-Bildschirmes sollten transparente Metalle sein. Welche Bedingung ergibt das für die Elektrondichte?

^ad.h. Herleitung der Wellengleichung mit Stromdichte im Ampère-Gesetz $\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{j}$.

Drude-Modell: Der Hall-Effekt

Bisher wurden die erfolgreichen Vorhersagen des Drude-Modell behandelt. Beim Hall-Effekt zeigt sich erstmals, dass die Beschreibung der Leitungsmechanismen mit dem Drude-Modell unzureichend ist.

Betrachten wir ein stromdurchflossenes Metall in einem homogenen Magnetfeld orthogonal zum Strom:



Diese Ladungen werden vergrößert bis das resultierende **Feld** E_y gerade die Lorentz-Kraft kompensiert

$$E_y = v_x B_z. \quad (54)$$

Das Feld in y -Richtung (bzw. die Spannung) wird gemessen und daraus der **Hall-Koeffizient** R_H ermittelt

$$R_H := \frac{E_y}{j_x B_z} = -\frac{1}{ne}. \quad (55)$$

Metal	Valenzen	$-1/R_H ne$
Li	1	0.8
Na	1	1.2
K	1	1.1
Rb	1	1.0
Cs	1	0.9
Cu	1	1.5
Ag	1	1.3
Au	1	1.5
Be	2	-0.2
Mg	2	-0.4
In	3	-0.3
Al	3	-0.3

Quelle: N. W. Ashcroft and D. D. Mermin, Solid State Physics

Basierend auf diesen Resultaten ist klar, dass das Drude-Modell die elektrischen Eigenschaften von Leitern unzureichend erklärt und wir eine detailliertere Theorie benötigen.

Unter den gemachten Annahmen hängt der Hall-Koeffizient einzig von der Elektronendichte n ab, die elektrische Ladung der Ladungsträger $q = -e$ ist durch die Annahmen des Drude-Modells (Elektronen) festgelegt.

Messungen des Hall-Koeffizienten R_H für verschiedene Metalle ergeben die Resultate in der Tabelle. In Abhängigkeit der Valenzen, also der laut Drude-Modell für die Leitung zur Verfügung stehenden freien Elektronen, wird der theoretische Wert des Hall-Koeffizienten $-1/ne$ relativ zur Messung R_H dargestellt.

Folgende Widersprüche sind offensichtlich:

- Der gemessene Hall-Koeffizient wird durch das Drude-Modell nur sehr schlecht vorhergesagt.
- Offensichtlich gibt es Metalle deren elektrische Leitfähigkeit von 'positiven Elektronen' stammen.

Aufgabe 11: In welche Richtung fliest der elektrische Strom \vec{j} (bei gleichem Versuchsaufbau) wenn positive Ladungen für den Strom verantwortlich sind? In welche Richtung fliessen diese 'positiven Elektronen', in welche Richtung werden sie abgelenkt und wie lautet die entstehende Hall-Spannung?

Erfolge und Unzulänglichkeiten des Drude Modells

Die Annahme von **freien Elektronen** im Metall, welche sich nach den Gesetzen der **kinetischen Gastheorie** als **freies Elektronengas** bewegen (Drude Modell), kann für einige physikalische Beobachtungen als **brauchbares Modell** hergenommen werden:

- Erklärung des Ohm'schen Gesetzes (elektrische Leitfähigkeit)
- Erklärung der Frequenzabhängigkeit der Leitfähigkeit
- Modell für Plasma-Oszillationen
- Erklärung für die glänzende Oberfläche von Metallen
- Erklärung für die Transparenz von Metallen bei hochfrequenter elektromagnetischer Strahlung
-

Für gewisse physikalische Beobachtungen liefert das Drude-Modell hingegen keine oder nur **unzureichende Erklärungen**:

- Keine Erklärung für die 'positiven Elektronen' beim Hall-Effekt
- Die Elektronengeschwindigkeit aus der kinetischen Gastheorie ist ca. eine Größenordnung unter der gemessenen Elektronengeschwindigkeit bei Raumtemperatur
- Die effektive Elektronen-Masse, gemessen aus der Dispersionsrelation der Elektronen im Leiter, Gleichung (104), unterscheidet sich von der Masse freie Elektronen
- Die Richtungsabhängigkeit des Leitwertes gewisser Materialien kann nicht erklärt werden
- Die Temperaturabhängigkeit des Leitwertes kann nicht richtig erklärt werden
-

Basierend auf diesen Resultaten muss die Näherung **unabhängiger und freier Elektronen**, beschrieben als ideales Gas entsprechend der **Maxwell-Boltzmann Statistik** aufgegeben werden.

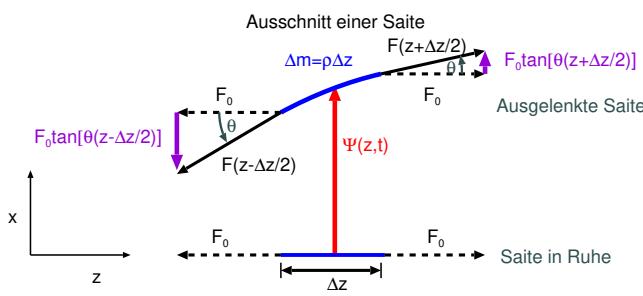
In den folgenden Kapiteln werden die physikalischen Gesetzmäßigkeiten von freien Ladungsträgern im Festkörper genauer untersucht. Es wird dafür nötig sein, die Ladungsträger als Wellen zu beschreiben (d.h. mithilfe der **Wellenmechanik** oder auch **Quantenmechanik** genannt). Im Rahmen der Vorlesung ist keine umfassende und systematische Darstellung der Quantentheorie möglich, für viele Phänomene ist aber ausreichend, die Welleneigenschaft der Teilchen als Erklärung herzunehmen. Um den scheinbaren Widerspruch zwischen Wellen-und Teilchenbeschreibung zu relativieren, werden im Folgenden die Gemeinsamkeiten von Wellen und Teilchen aufgezeigt.

Aufgabe 12: Warum wird eine 'falsche Theorie' überhaupt unterrichtet?

3.2– Mechanische Wellen

Im Rahmen der Punktmechanik wurde die Energie-Impulsrelation (die **Dispersionsrelation**) für klassische Punktteilchen hergeleitet, ebenso wurde die Dispersionsrelation eines massiven und eines masselosen relativistischen Teilchens behandelt. Aus der Form der Dispersionsrelation konnte die (effektive) Masse des entsprechenden Teilchens abgelesen werden. Wie sehen nun die entsprechenden Beziehungen bei Wellen aus?

Ein einfaches **mechanisches Modell eines Wellenleiter** ist die transversale Schwingung einer eindimensionalen massiven Saite, mit einer Liniendichten ρ (kg pro Meter), siehe Abbildung.



Die Saite sei mit Zugkraft F_0 entlang der z -Achse eingespannt. Die schwingende Saite wird transversal in x -Richtung ausgelenkt, beschrieben durch die entsprechende orts- und zeitabhängige **Elongation** $\Psi(z, t)$, die **Wellenfunktion**.

Betrachten wir die Auslenkung eines kleinen Saitenausschnitts der Länge Δz , mit Masse $\Delta m = \rho\Delta z$ und Elongation $\Psi(z, t)$.

Das 2. Newton'sche Axiom, Gleichung (3), für diesen Abschnitt bestimmt die Dynamik der Saite:

$$\Delta m \frac{\partial^2}{\partial t^2} \Psi(z, t) = F_0 \tan[\theta(z + \frac{\Delta z}{2})] - F_0 \tan[\theta(z - \frac{\Delta z}{2})] \quad (56)$$

Die Kräfte links und rechts des Abschnittes wirken jeweils in Seilrichtung, also im Winkel $\theta(z \pm \Delta z/2, t)$. Wird nun der Tangens des Winkel links und rechts des Saitenausschnittes durch die Steigung der Saite an der entsprechenden Stelle ausgedrückt, folgt für kleine Δz (Taylor-Entwicklung)

$$\tan[\theta(z \pm \frac{\Delta z}{2}, t)] \equiv \frac{\partial}{\partial z} \Psi(z \pm \frac{\Delta z}{2}, t) \approx \frac{\partial}{\partial z} \Psi(z, t) \pm \frac{\Delta z}{2} \frac{\partial^2}{\partial z^2} \Psi(z, t). \quad (57)$$

Eingesetzt in die Bewegungsgleichung (56) folgt die **Wellengleichung in 1-dim** (siehe Physik II),

$$\frac{\partial^2}{\partial t^2} \Psi(z, t) = v^2 \frac{\partial^2}{\partial z^2} \Psi(z, t). \quad (58)$$

mit der **Wellengeschwindigkeit** $v := \sqrt{F_0/\rho}$.

Die Lösung der Wellengleichung erfolgt über die Erfahrungstatsache, dass Wellen sich sowohl als Funktion der Zeit t als auch als Funktion der Position z analog verhalten. Dieser Sachverhalt ist auch in der Wellengleichung klar ersichtlich, da die linke und rechte Seite mathematisch völlig analog sind. Es macht daher Sinn, eine Lösung zu versuchen, mit Zeitanteil und Raumanteil als gleichberechtigte Faktoren,

$$\Psi(z, t) =: Z(z)T(t). \quad (59)$$

Die Funktion $Z(z)$ ist nur eine Funktion der Position und drückt die räumliche Welle aus, $T(t)$ ist die entsprechende Zeitabhängigkeit. Eingesetzt in die Wellengleichung (58) folgt

$$\frac{\ddot{T}(t)}{T(t)} \stackrel{!}{=} v^2 \frac{Z''(z)}{Z(z)} \quad (60)$$

Auf der linken Seite ist die Zeit die einzige Variable, auf der rechten die Position. Die Bedingung kann daher nur erfüllt werden, wenn beide Seiten konstant sind. Aus der Mechanik wissen wir, dass eine Schwingung der **Kreisfrequenz** ω mit dem Euler'schen Ansatz $T(t) = \exp(\pm i\omega t)$ dargestellt wird. Ein analoger Ansatz für den Raumanteil, $Z(z) = \exp(\pm ikz)$, mit der 'räumlichen Frequenz', genannt **Wellenzahl**, k ergibt,

$$\omega^2 \stackrel{!}{=} v^2 k^2. \quad (61)$$

Aufgabe 13: Wie lautet die allgemeine Lösung der Wellengleichung? Entspricht es der Vorstellung einer Welle mit Geschwindigkeit v ?

Dispersionsrelation eines kontinuierlichen Wellenleiters

Die Wellengleichung (58) für 1-dim Wellen kann in analoger Weise auf mehrdimensionale Wellen verallgemeinert werden. Die Wellenfunktion wird zu einem Feld der drei Ortsrichtungen und der Zeit $\Psi(\vec{x}, t)$. Die linke, zeitabhängige Seite der Wellengleichung bleibt unverändert, während die rechte Seite nun die drei Koordinatenrichtungen in gleichberechtigter Weise behandelt (**Laplace Operator**):

$$\frac{\partial^2}{\partial t^2} \Psi(\vec{x}, t) = v^2 \Delta \Psi(\vec{x}, t), \quad (62)$$

N.B.: Im Falle elektromagnetischer Wellen, also Licht (siehe Physik II), folgt genau diese dreidimensionale Wellengleichung für das elektrische Feld $\vec{E}(\vec{x}, t)$ und das magnetische Feld $\vec{B}(\vec{x}, t)$, mit $v \equiv c$.

Die allgemeine Lösung der 3-dim Wellengleichung bei gegebener Kreisfrequenz ω kann als **ebene Welle** (siehe Physik II) in Richtung \vec{n} mit Wellenzahl k ausgedrückt werden, bzw. durch den **Wellenvektor** \vec{k} ,

$$\vec{k} := \frac{2\pi}{\lambda} \vec{n}. \quad (63)$$

Jede Welle der Frequenz ω kann nun als räumliche Überlagerung (Interferenzmuster) ebener Wellen

$$\Psi(\vec{x}, t) = A e^{i(\vec{k} \cdot \vec{x} - \omega t)} \quad (64)$$

unterschiedlicher Wellenvektoren \vec{k} dargestellt werden. Eine 3-dim ebene Welle wird also durch die Frequenz ω (für das zeitliche Verhalten) und den Wellenvektor \vec{k} (für das räumliche Verhalten) vollständig charakterisiert. Einsetzen in die Wellengleichung (62) zeigt, dass auch in 3-dim eine zusätzliche Bedingung zwischen Frequenz und Wellenvektor erfüllt werden muss, vgl. Gleichung (61)

$$\omega^2 \stackrel{!}{=} v^2 k^2. \quad (65)$$

Drückt man die Frequenz durch die zeitliche Periodenlänge T und die Wellenzahl durch die 'räumliche Periodenlänge', die **Wellenlänge** λ aus, erhalten wir wie erwartet

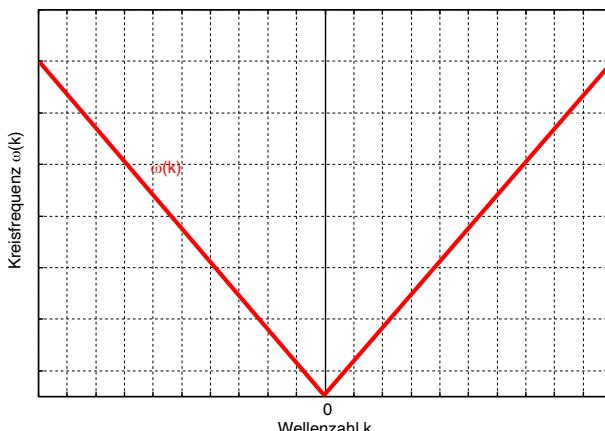
$$\lambda \stackrel{!}{=} vT, \quad (66)$$

die Wellenlänge ist also gerade die Strecke, die die Welle (der Geschwindigkeit v) in einer Periode zurücklegt.

Die notwendige Bedingung für zwischen Kreisfrequenz und Wellenvektor ist in der Abbildung dargestellt,

$$\omega = v|\vec{k}|. \quad (67)$$

Das ist die **Dispersion Relation** einer 3-dim Welle auf einem kontinuierlichen Wellenleiter. Die beiden Vorzeichen der k -Werte entsprechen der rechts-bzw. linkslaufenden Welle.



Bereits hier ist die Ähnlichkeit zur Energie-Impulsbeziehung (ebenfalls Dispersionsrelation genannt) eines masselosen Teilchens mit Lichtgeschwindigkeit $v = c$, Gleichung (12)

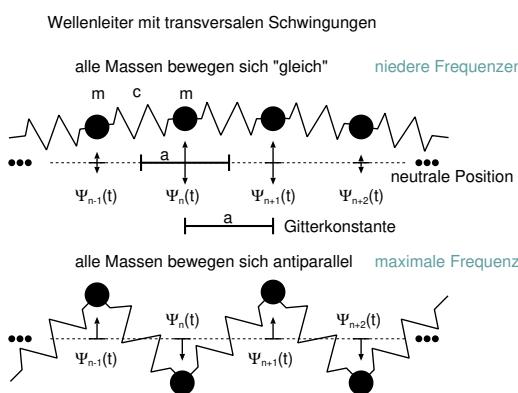
$$E = c|\vec{p}|, \quad (68)$$

offensichtlich. Die Kreisfrequenz ω einer Welle scheint analog zur Energie E eines Punktteilchens, die Wellenvektor \vec{k} entspricht dem Teilchenimpuls \vec{p}

Aufgabe 14: Ein Punktteilchen mit gegebenen Impuls werde durch eine ebene Welle mit entsprechender Wellenzahl dargestellt. Wo ist die Position dieses Teilchens? (Grüße von W. Heisenberg!)

Dispersionsrelation eines 1-dim Kristalls

Die Atome in einem Festkörper und daher die Elektronen bilden keinen kontinuierlichen Wellenleiter. Der Grund liegt in der Periodizität des Kristallgitters. Wir betrachten nun das mechanische Modell einer 'periodisch aus diskreten Massen aufgebauten Saite', einer **Perlenkette**. Identische Massen m auf einem elastischen Band sind auf einen Gitterabstand a auseinandergezogen, siehe Abbildung.



Für **diskret aufgebaute Wellenleiter** erwarten wir:

- Bei großen Wellenlängen, d.h. $k \rightarrow 0$, darf die Granularität nicht ins Gewicht fallen. Die Dispersionsrelation einer Perlenkette ist daher für **kleine Wellenzahlen identisch zur kontinuierlichen Saite**.
- Die kürzest mögliche Wellenlänge ergibt sich bei antiparallel schwingenden Perlen, d.h. $\lambda_{\min} = 2a$. Die Dispersionsrelation reicht folglich nur über einen **endlichen Bereich von Wellenzahlen** $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$
- Der minimalen Wellenlänge $\lambda_{\min} = 2a$ entspricht eine Maximalfrequenz ω_{\max} für ausbreitende Wellen.

Die Wellenfunktion der Perlenkette wird durch die jeweilige Elongation der n -ten Perle beschrieben, $\Psi_n(t)$, d.h. die räumlich kontinuierliche Wellenfunktion zerfällt in diskrete Werte an den Positionen $z = na$. Die horizontale Zugkraft zwischen benachbarten Perlen ist $F_0 = ca$. Die **transversale Zugkraft** F_t ergibt sich aus der horizontalen Zugkraft und dem Winkel θ zwischen benachbarten Auslenkungen,

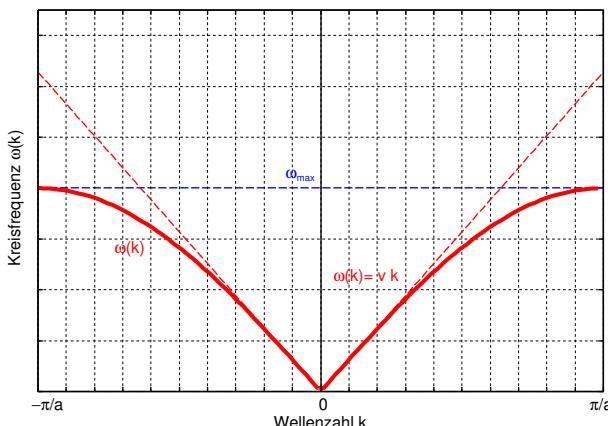
$$F_t = F_0 \tan \theta. \quad (69)$$

Die transversale Rückstellkraft auf die n -te Perle, ausgeübt von der $n+1$ -ten Perle lautet daher exemplarisch $F_0 \frac{\Psi_n - \Psi_{n+1}}{a} \equiv c(\Psi_{n+1} - \Psi_n)$ und die Bewegungsgleichung für $\Psi_n(t)$ resultiert:

$$m \ddot{\Psi}_n(t) = c(\Psi_{n+1}(t) - \Psi_n(t)) + c(\Psi_{n-1}(t) - \Psi_n(t)) = c[\Psi_{n+1}(t) + \Psi_{n-1}(t) - 2\Psi_n(t)] \quad (70)$$

Wir erwarten für laufende Wellen mit Frequenz ω und Wellenzahl k wieder 1-dim ebene Wellen, Gleichung 64, an diskreten Positionen bei $z = na$, d.h. $\Psi_n(t) = Ae^{i(kna - \omega t)}$. Eingesetzt in die Bewegungsgleichung (70) folgt die **Dispersionrelation für periodische Wellenleiter**, siehe Abbildung

$$\omega = |2\sqrt{\frac{c}{m}} \sin \frac{ka}{2}|. \quad (71)$$



Im Fall kleiner Wellenzahlen entspricht die Dispersionrelation einer kontinuierlichen Saite mit Massendichte m/a , Zugkraft ca und daher Wellengeschwindigkeit

$$v = a \sqrt{\frac{c}{m}}. \quad (72)$$

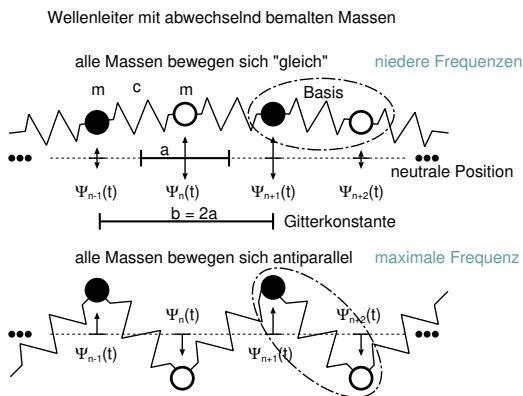
Die Maximalfrequenz für antiparallel schwingende Massen ist

$$\omega_{\max} = 2\sqrt{\frac{c}{m}}. \quad (73)$$

Aufgabe15: Wie lautet die Herleitung der Dispersionsrelation? Wie hängt die Maximalfrequenz mit der Gleichung des harmonischen Oszillators zusammen?

Dispersionsrelationen eines Kristalls mit nichttrivialen Elementarzellen

Die bisher betrachtete Dispersionsrelation (1-dim) bestand aus identischen Massen im fixen Gitterabstand a . Als Folge bestand die Dispersionsrelation aus nur einem Band, d.h. es gab zu jeder Wellenzahl nur eine mögliche Frequenz ebener Wellen. **Dispersionsrelationen mit mehreren Frequenzbändern** entstehen, sobald das Gitter aus nichttrivialen Einheiten, genannt **Basis**, aufgebaut ist. Betrachten wir hierzu die Perlenkette mit 'alternierend angemalten Perlen' siehe Abbildung. Die Basis besteht in diesem Fall aus einem schwarz-weißen Perlenpaar, die Gitterkonstante beträgt $b = 2a$.



Die Physik darf nicht von der Farbe abhängen, aber:

- Die kürzest mögliche Wellenlänge folgt wieder aus der Periode d.h. $\lambda_{\min} = 2b = 4a$. Das Fenster der Dispersionsrelation ist folglich halb so groß, $k \in [-\frac{\pi}{b}, \frac{\pi}{b}]$.
- Bei Periodenlänge $\lambda_{\min} = 2b$, d.h. $k = \pm \frac{\pi}{b}$ ruht jede zweite Perle (Schwingungsknoten). Aufgrund der Farbe gibt es nun aber zwei unterscheidbare Situationen.
 - Weiß schwingt, Schwarz ruht
 - Schwarz schwingt, Weiß ruht

Es folgen also **zwei mögliche Fälle** für $k = \pm \frac{\pi}{b}$.

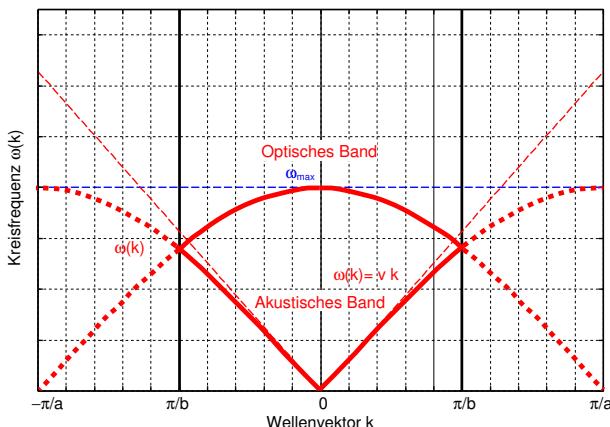
- Die Maximalfrequenz ergibt sich wie im vorigen Fall bei antiparallel schwingenden Perlen. Ausgedrückt durch unsere neuen Gitterelemente, bestehend aus einer Basis mit 2 gefärbten Perlen, schwingen jedoch bei ω_{\max} alle Gitterelemente gleich, siehe Abbildung, folglich in Phase und daher mit Wellenlänge unendlich. Bei $k = 0$ befindet sich demnach sowohl die minimale als auch die maximale Frequenz.

Die Dispersionsrelation mit nichtrivialer Basis besteht aus mehreren Frequenzbändern!

Aus der vorhin abgeleiteten Dispersion für die Perlenkette folgt mit der Vorschrift $k \in [-\frac{\pi}{b}, \frac{\pi}{b}]$ direkt die **Dispersionsrelation** der betrachteten nichtrivialen Elementarzelle, siehe Abbildung,

$$\omega = |2\sqrt{\frac{c}{m}} \sin \frac{kb \pm 2n\pi}{4}|. \quad (74)$$

N.B. Die Halbierung des Wellenzahlfensters 'faltet' formal die Dispersionsrelation bei $k = \pm \pi/b$.

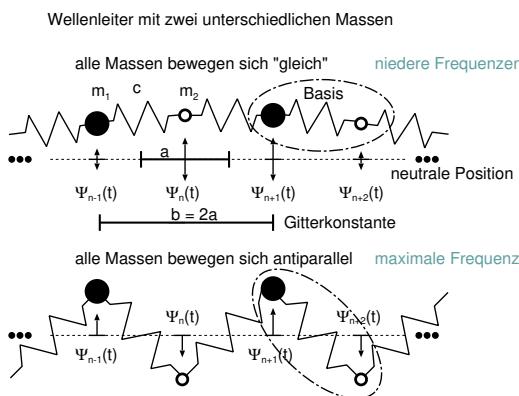


Die betrachtete Dispersionsrelation weist zu jeder Wellenzahl zwei unterscheidbare Frequenzbänder auf. Beim unteren Band schwingen Nachbarn parallel zu einander, wie bei akustischen Gitterschwingungen (**Akustisches Band**). Im oberen Band schwingen jeweils die beiden Massen der Basis gegengleich. In einem Festkörper mit Ionenbindung bedeutet das eine jeweils antiparallele Schwingung der Ionenpaare, die durch elektromagnetischen Wellen angeregt werden können (**Optisches Band**).

Aufgabe16: Wie lautet die Herleitung der Dispersionsrelation? Wie sehen die beiden Schwingungsmuster bei $k = \pm \frac{\pi}{b}$ aus?

Dispersionsrelationen mit Bandabstand

Die bisher betrachtete Dispersionsrelation (1-dim) bestand aus identisch schweren Massen im fixen Gitterabstand a . Wenn nun die Basis der Perlenkette aus tatsächlich unterschiedlichen Massen (genannt m_1 und m_2 , mit $m_1 > m_2$) besteht, siehe Abbildung, tritt ein sehr interessanter und für das Verständnis von Halbleitern essentieller Wellen-Effekt auf: Es bildet sich zwischen den Bändern eine **Bandlücke**, auch genannt **Bandabstand**, aus, d.h. es gibt nicht nur eine Maximalfrequenz sich ausbreitender Wellen, es gibt auch 'verbotene Frequenzen' auf denen **keine ebenen Wellen erlaubt** sind.



Die Physik des Bandabstandes lässt sich leicht verstehen:

- Wellenzahlfenster wie gehabt $k \in [-\frac{\pi}{b}, \frac{\pi}{b}]$.
- Bei $k = 0$ befindet sich die maximale Frequenz (optische Mode) mit antiparallen Schwingungen
- Bei maximaler Wellenzahl $k = \pm \frac{\pi}{b}$ ruht wieder jede zweite Perle (Schwingungsknoten). Aufgrund der unterschiedlichen Massen gilt nun aber
 - Schwere Masse (m_1) schwingt, leichte Masse ruht
 - Leichte Masse (m_2) schwingt, schwere Masse ruht

Die entsprechende **Dispersionsrelation** wird analog zu vorhin berechnet, sei hier aber nur angeführt,

$$\omega = \sqrt{\frac{c}{M} \left(1 \pm \sqrt{1 - \frac{4M^2 \sin^2 \frac{kb}{2}}{m_1 m_2}} \right)}, \quad (75)$$

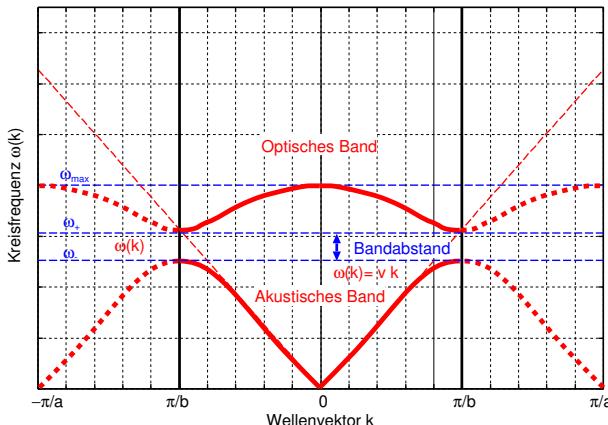
mit der Abkürzung (reduzierter Masse M)

$$\frac{1}{M} := \frac{1}{m_1} + \frac{1}{m_2}. \quad (76)$$

Im Fall $m_1 \equiv m_2$ folgt daraus das vorherige Resultat.

Die Bandlücke (der Frequenzbereich in dem keine ebene Wellen möglich sind) liegt zwischen der maximalen Frequenz des akustischen, ω_- , und der minimalen Frequenz des optischen Bandes, ω_+ , siehe Abbildung.

$$\Delta\omega_{BG} = \omega_+ - \omega_-. \quad (77)$$



Die Maximalfrequenz ist

$$\omega_{max} = \sqrt{\frac{2c(m_1 + m_2)}{m_1 m_2}}. \quad (78)$$

Bei $k = \pm \pi/b$ lauten die Frequenzen

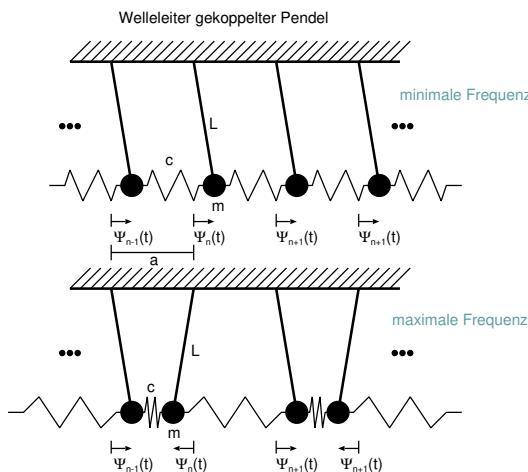
$$\omega_- := \sqrt{\frac{2c}{m_1}}, \quad (79)$$

$$\omega_+ := \sqrt{\frac{2c}{m_2}}.$$

Aufgabe17: Leite die Grenzfrequenzen der Bandlücke aus der Oszillatorgleichung des Masse-Feder-Systems her. Was passiert mit dem Wellenvektor falls die Frequenz in der Bandlücke liegt?

Dispersionsrelationen mit unterer Grenzfrequenz

Die bisher betrachteten Dispersionsrelationen (1-dim) waren im Grenzwert großer Wellenlängen im akustischen Band analog zur Dispersionsrelation relativistischer Teilchen. Ein mechanischer Wellenleiter mit der langwelligen Dispersionsrelation eines **freien Teilchens** ist die **Pendelkette**, siehe Abbildung. Die Pendellänge sei L , die Federkonstante c koppelt die Pendelmassen m mit einer Gitterkonstanten a .



Die Physik der Schwingungen im erlaubten Frequenzbandes ist leicht zu verstehen:

- Die minimale Frequenz entspricht der Pendelfrequenz

$$\omega_0^2 := \frac{g}{L}. \quad (80)$$

- Die maximal mögliche Frequenz entspricht dem Fall antiparallel schwingender Pendel, d.h. es addiert sich die Pendelkraft zur Federkraft

$$\omega_{\max}^2 = \frac{g}{L} + \frac{4c}{m}. \quad (81)$$

Die entsprechende **Dispersionsrelation** wird analog zu vorhin berechnet, sei hier aber nur angeführt,

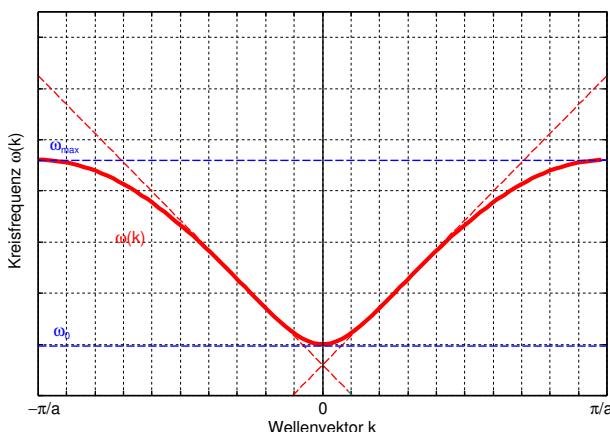
$$\omega = \sqrt{\frac{g}{L} + \frac{4c}{m} \sin^2 \frac{ka}{2}}. \quad (82)$$

Im Grenzfall kleiner Wellenvektoren $k \rightarrow 0$ kann der Sinus durch den ersten Term der Taylorreihe ersetzt werden. Zusätzlich wird der zweite Term in der Wurzel viel kleiner als der erste und es folgt

$$\omega(k) \xrightarrow{k \rightarrow 0} \omega_0 + \frac{ca^2}{\omega_0} \frac{k^2}{2m}. \quad (83)$$

Vergleicht man die Dispersionsrelation bei kleinen Wellenzahlen mit der Energie-Impulsbeziehung des massiven nichtrelativistischen Teilchens in einem konstanten Potential, Gleichung (5) sehen wir wieder die klare Analogie (wie schon bei der Perlenkette):

- Die **Energie** eines Punktteilchens entspricht der **Kreisfrequenz** einer Welle
- Der **Impuls** eines Punktteilchens entspricht dem **Wellenvektor** einer Welle



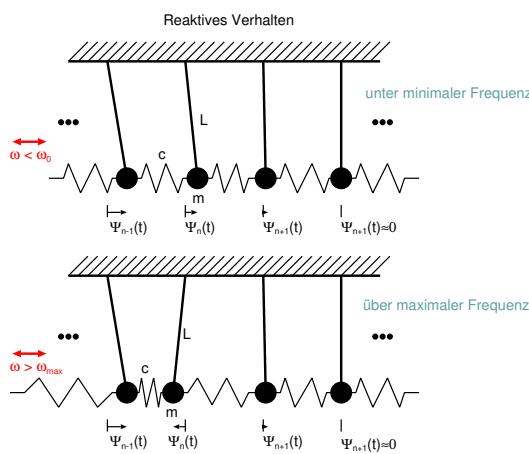
Dispersionsrelationen eines 3-dim Kristalls zeigen entsprechend die Kreisfrequenzen der möglichen ebenen Wellen als Funktion des dreidimensionalen Wellenvektors, $\omega(\vec{k})$. Dispersionsrelationen sind daher Hyperflächen in einem 4-dim Raum. Zumeist werden daher die Schnitte dieser Hyperflächen entlang wichtiger Symmetrierichtungen des Kristalls gezeichnet. Die Eigenschaften und Erklärungen bleiben aber analog zum 1-dim Kristall.

Aufgabe18: Skizziere die Dispersionsrelation einer Pendelkette mit alternierend gefärbten gleichschweren Massen. Was passiert bei alternierend unterschiedlichen Massen?

Dispersives und reaktives Verhalten

In der Herleitung der Dispersionsrelationen wurde jeweils nach den möglichen Frequenzen für einen gegebenen Wellenvektor \vec{k} einer **ebenen Welle** mit konstanter Amplitude gefragt. Diese Wellen können den Wellenleiter dissipationsfrei durchqueren, d.h. sie verlassen das gegenüberliegende Ende mit der Eingangsamplitude. Der Wellenleiter ist für diese Frequenzen 'transparent', er verhält sich **dispersiv**.

Was passiert nun mit einem Wellenleiter, der mit einer Frequenz ausserhalb des **Dispersiven Bandes** angeregt wird? Diese Anregung ist weder 'verboten' noch sind die entstehenden Wellen 'verboten' - sie sind nur eben keine dispersiven Wellen.



Anregung ausserhalb des **Dispersiven Bereiches**:

- Bei Frequenzen unterhalb der minimalen Frequenz, $\omega < \omega_0$, folgt jedes Pendel der Auslenkung des Vorgängers, erreicht aber nur eine kleinere Amplitude. Die Pendel schwingen **in Phase** mit **exponentiell abklingender Amplitude**, siehe Abbildung
- Bei Frequenzen oberhalb der maximalen Frequenz, $\omega > \omega_{\max}$, versucht jedes Pendel der Auslenkung des Vorgängers gegengleich zu folgen, erreicht die volle Amplitude aber nicht rechtzeitig. Die Pendel schwingen **gegenphasig** mit **exponentiell abklingender Amplitude**, siehe Abbildung.

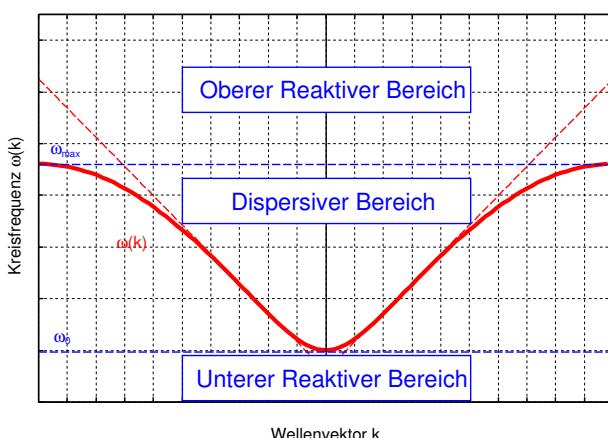
Man bezeichnet räumlich exponentiell abklingende Wellen als **reakтив**. Das Frequenzband wird entsprechend **Reaktiver Bereich** genannt.

Wie sehen nun die Wellenvektoren in den reaktiven Bereichen aus? Im **unteren reaktiven Bereich** der Pendelkette gilt $\omega^2 < \frac{g}{L}$, im **oberen reaktiven Bereich** der Pendelkette gilt $\omega^2 > \frac{g}{L} + \frac{4c}{m}$. Eingesetzt in die Dispersionsrelation (82) folgen die Bedingungen

$$\cos ka > 1 \quad \text{Unterer reaktiver Bereich} \quad (84)$$

$$\cos ka < -1 \quad \text{Oberer reaktiver Bereich} \quad (85)$$

Für reelle Wellenzahlen ist die Bedingung nicht zu erfüllen, erlaubt man hingegen komplexe Wellenzahlen kann mit der Euler'schen Darstellung $\cos \alpha = (e^{i\alpha} + e^{-i\alpha})/2$ leicht eine Lösung gefunden werden.



Oberer reaktiver Bereich

Exponentiell abklingende alternierende Wellen mit gemischt komplexer Wellenzahl $k(\omega) =: \frac{\pi}{a} + i\kappa(\omega)$:

$$\Psi_n(t) \propto e^{-\kappa an} (-1)^n e^{-i\omega t} \quad (86)$$

Unterer reaktiver Bereich

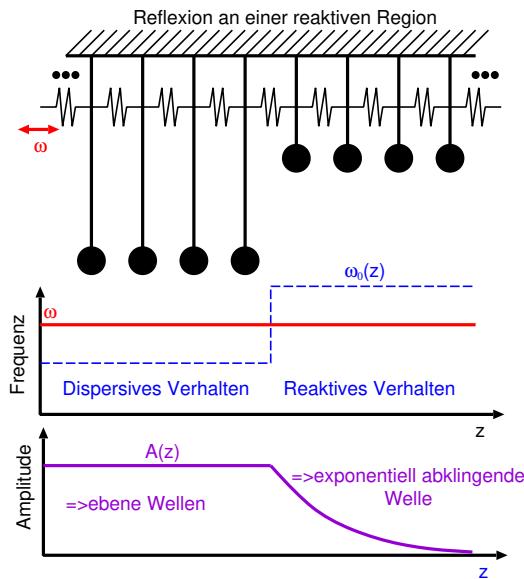
Exponentiell abklingende Welle mit rein komplexer Wellenzahl $k(\omega) =: i\kappa(\omega)$:

$$\Psi_n(t) \propto e^{-\kappa an} e^{-i\omega t} \quad (87)$$

Aufgabe19: Welche Relation ergibt sich für den Abklingexponenten κ als Funktion der Frequenz ω in den reaktiven Bereichen?

Reflexion und Tunneleffekt

Was passiert, wenn eine ebene Welle aus einem dispersiven Wellenleiter auf eine reaktive Region trifft?

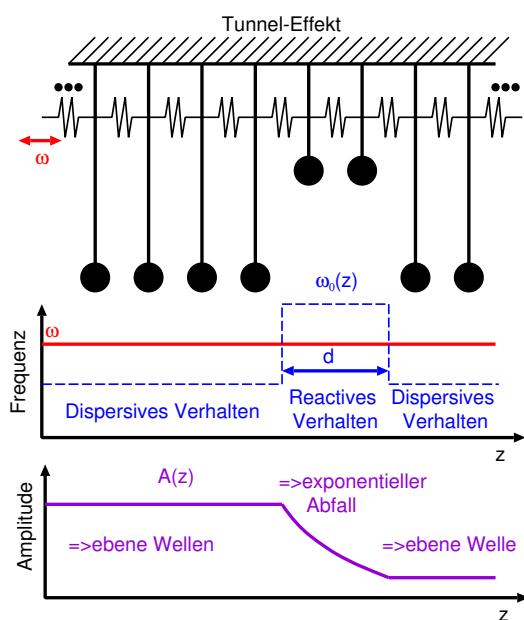


Betrachten wir dazu eine Pendelkette, bei der die Pendellänge an einer Stelle im Raum plötzlich kürzer wird, siehe Abbildung. Die untere Grenzfrequenz ω_0 der Pendelkette ist folglich ortsabhängig und springt an dieser Stelle von $\omega_0(-\infty)$ auf einen höheren Wert $\omega_0(+\infty)$.

Eine von links einlaufende ebene Welle der Frequenz ω mit $\omega > \omega_0(-\infty)$ (dispersiv) trifft auf eine reaktive Region mit $\omega < \omega_0(+\infty)$. Es wird im reaktiven Bereich eine exponentiell abklingende Welle angeregt, die nach einigen Pendeln quasi nicht mehr vorhanden ist. Gleichzeitig wirkt die reaktive Welle zurück auf den dispersiven Bereich und regt eine rücklaufende ebene Welle an.

Nachdem im betrachteten Aufbau keine dissipativen Elemente sind, muss die Energie in den Wellen erhalten bleiben. Die auslaufende Welle trägt die gleiche Leistung wie die einlaufende Welle, sie wird also **vollständig reflektiert**. (Das gleiche Verhalten tritt mikroskopisch bei jeder Totalreflexion der Optik auf.)

Basierend auf diesem Verhalten kann auch der **Wellenmechanische Tunneleffekt** anschaulich verstanden werden. Wir betrachten dazu wieder einen räumlich schmalen (Breite d) reaktiven Bereich, der auf beiden Seiten an einen dispersiven Bereich grenzt, siehe Abbildung. Im klassischen Bild einer Totalreflexion, sollte jede einlaufende Welle am 'verbotenen Bereich' (d.h. am reaktiven Bereich) vollständig reflektiert werden. Wellenmechanisch zeigt sich allerdings ein deutlich anderes Bild.



Eine von links einlaufende ebene Welle der Frequenz ω mit $\omega > \omega_0(-\infty)$ (dispersiv) trifft auf einen kurzen reaktiven Bereich. Es wird im reaktiven Bereich eine exponentiell abklingende Welle angeregt und ein Teil der Welle an der ersten Grenzfläche reflektiert.

Die reaktive Welle klingt proportional $e^{-\kappa z}$ ab und erreicht die gegenüberliegende Grenzfläche mit einer um $e^{-\kappa d}$ reduzierten Amplitude. Entsprechend wird an dieser zweiten Grenzfläche eine nach rechts auslaufende ebene Welle mit reduzierter Amplitude angeregt.

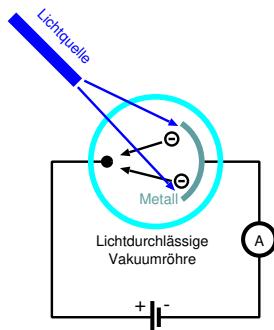
Aufgrund der Energieerhaltung muss die Leistungsdifferenz zwischen einlaufender und reflektierter Welle gerade die Leistung der durch die reaktive Region 'getunnelten' auslaufenden Welle sein.

Die Welle tunnelt durch den 'verbotenen Bereich'!

Aufgabe 20: Beobachte den Tunneleffekt bei einer optischen Totalreflexion. (Blick von oben in ein gefülltes Wasserglas auf die Rillen der Finger der Hand, die das Glas hält.)

3.3– Elektronen als Welle

Licht verhält sich oft wie eine Welle (elektromagnetische Welle), beim **Photoelektrischen Effekt** treten hingegen die Teilchenaspekte in Erscheinung, siehe Abbildung. Licht der (Kreis-)Frequenz ω fällt in einer sogenannten **Photozelle** auf eine metallische Oberfläche, welche elektrisch vorgespannt ist. Je nach Metall werden ab einer bestimmten minimalen Frequenz Elektronen freigesetzt und es wird ein Strom gemessen.



Die Intensität des einfallenden Lichtes ist für das Freisetzen der Elektronen irrelevant, jedoch steigt bei Frequenzen oberhalb dieser Minimalfrequenz die kinetische Energie der freigesetzten Elektronen. Die korrekte Interpretation ist der Energieerhaltungssatz (Einstein 1905):

$$E_{\text{kin}} = \hbar\omega - E_0 \quad (88)$$

Hier ist E_0 die metallabhängige Bindungsenergie der Elektronen, die beim Stoß mit einem Lichtteilchen der Energie $E = \hbar\omega$ mindestens überwunden werden muss. Somit wurde der Plank'sche 'Rechentrick' einer Lichtenergie in **Quanten** von $\hbar\omega$ (siehe auch Physik I) zur tatsächlichen Energie der Lichtteilchen (**Photonen**) uminterpretiert.

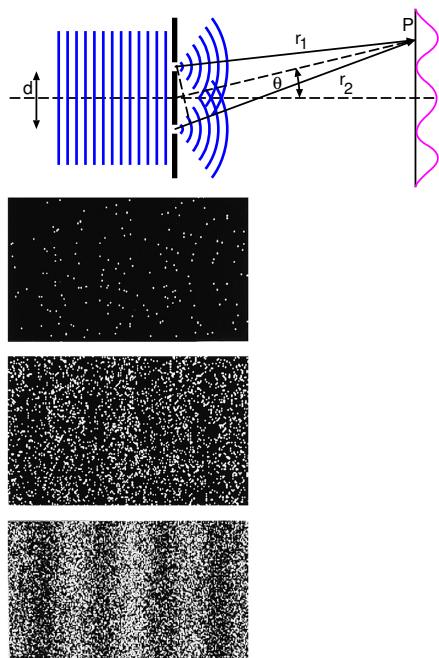
Licht als ebene Welle, $\Psi(\vec{x}, t) \propto e^{i(\vec{k} \cdot \vec{x} - \omega t)}$, wird durch die beiden Parametern Kreisfrequenz ω und Wellenvektor \vec{k} charakterisiert, Photonen hingegen durch Energie E und Impuls \vec{p} . Die Verknüpfung beider Größen ist jeweils die Dispersionsrelation mit c als Proportionalitätskonstante. Es ist also naheliegend nicht nur die Energie mit der Frequenz, sondern auch den Impuls mit dem Wellenvektor in Beziehung zu bringen:

$$E \stackrel{!}{=} \hbar\omega \quad (89)$$

$$\vec{p} \stackrel{!}{=} \hbar\vec{k} \quad (90)$$

N.B. Die Wellenlänge zum Impuls \vec{p} wird auch **de-Broglie-Wellenlänge** genannt.

Wie sieht es nun mit dem **Wellencharakter der Elektronen** aus, die sich oft wie Teilchen verhalten?



Wird, analog zum Young'schen Doppelspaltversuch in der Optik (siehe Physik II), ein Elektronenstrahl auf einen Doppelspalt geschossen, siehe Abbildung, erscheinen die Elektronen auf dem Schirm zuerst zufällig verteilt. Mit steigender Dauer des Experiments zeigt sich aber, dass an gewissen Positionen deutlich mehr und an anderen Positionen niemals Elektronen zu finden sind. Die Wahrscheinlichkeitsverteilung ein Elektron an einer bestimmten Position auf dem Schirm zu finden entspricht gerade dem Intensitätsbild des Young'schen Experimentes.

Elektronen passieren den Schirm als Wellen, wobei die Intensität (also das Quadrat der Amplitude) dieser Welle gerade der **Wahrscheinlichkeitsdichte** auf dem Schirm entspricht. In gewissem Sinn werden Elektronen als Teilchen erzeugt, breiten sich als Welle aus und werden auf dem Schirm wieder als Teilchen gemessen. Diese Interpretation wird **Welle-Teilchen-Dualismus** genannt.

Quelle: D. C. Giancoli, Physik, Pearson (2010)

Aufgabe 21: Fliegt das Elektron nun durch das obere oder das untere Loch? Wie sieht das Muster auf dem Schirm aus, wenn an jedem Loch die durchfliegenden Elektronen registriert werden?

Wellenmechanik

Nimmt man die Analogie für die **Wellengleichung des Lichtes** als Richtschnur, lässt sich folgendes Vorgehen für die Brücke zwischen der Teilchenbeschreibung und der Wellenbeschreibung definieren:

- Nimm die Energie-Impulsbeziehung der Teilchenbeschreibung
- Ersetze Energie und Impuls durch Ableitungen einer, zu bestimmenden, Wellenfunktion $\Psi(\vec{x}, t)$

$$E\Psi(\vec{x}, t) \equiv i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}, t) \quad (91)$$

$$\mathbf{p}\Psi(\vec{x}, t) \equiv \frac{\hbar}{i} \vec{\nabla} \Psi(\vec{x}, t) \quad (92)$$

- Die so erhaltene Differentialgleichung ist die entsprechende Wellengleichung
- Die Wellenfunktion definiert die Aufenthaltswahrscheinlichkeit des Teilchen. Das Quadrat des Absolutbetrags der Wellenfunktion ist gerade die lokale Wahrscheinlichkeitsdichte:

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} |\Psi(x, t)|^2 dx \quad \text{1-dim} \quad (93)$$

Mit diesem 'Rezept' kann einerseits die **einem Teilchen entsprechende Wellenfunktion** gefunden werden, aber auch die Aufenthaltswahrscheinlichkeit eines **der Wellenfunktion entsprechenden Teilchens** angegeben werden. Formal bedeutet der so definierte Formalismus, dass die klassischen Größen (z.B. Energie und Impuls) durch Differentialoperatoren ersetzt werden (z.B. \hat{E} und \hat{p}):

$$\hat{E} := i\hbar \frac{\partial}{\partial t} \quad (94)$$

$$\hat{p} := \frac{\hbar}{i} \vec{\nabla} \quad (95)$$

Die **Wellengleichung eines Elektrons** in einem Potential $V(\vec{x})$ erhält man folglich aus der Energie-Impuls Beziehung, $E = \frac{\hat{p}^2}{2m} + V$, indem der Impuls durch den entsprechenden Differentialoperator (95) ersetzt wird. Der resultierende Differentialoperator der Energie wird **Hamiltonoperator** \hat{H} genannt:

$$\hat{H} := -\frac{\hbar^2}{2m} \Delta + V(\vec{x}) \quad (96)$$

N.B.: Die Ersetzung von \hat{p}^2 durch das Quadrat des Impulsoperators \hat{p}^2 beinhaltet ein Skalarprodukt zwischen Gradienten, und entspricht daher dem Laplace-Operator, $\vec{\nabla} \cdot \vec{\nabla} \equiv \Delta$.

Die Zeitentwicklung der Wellenfunktion im Raum wird nun beschrieben durch eine Wellengleichung, die **zeitabhängige Schrödingergleichung**:

$$\hat{H}\Psi(\vec{x}, t) = i\hbar \frac{\partial}{\partial t} \Psi(\vec{x}, t) \quad (97)$$

Bei stationären Problemen gegebener Energie ist die Zeitabhängigkeit bekannt und kann abgespalten werden, d.h. $\Psi(\vec{x}, t) =: e^{-i\omega t} \Psi(\vec{x})$ und es folgt die Wellengleichung für die räumliche Wellenfunktion $\Psi(\vec{x})$, die **zeitunabhängige Schrödingergleichung**:

$$\hat{H}\Psi(\vec{x}) = E\Psi(\vec{x}) \quad (98)$$

Die Wellenmechanik von Erwin Schrödinger begründete im Jahre 1925 die moderne **Quantenmechanik**. Die Wellenfunktion beschreibt den quantenmechanischen Zustand vollständig und liefert Wahrscheinlichkeitsaussagen für das entsprechende Teilchen. Die Einschränkung auf Wahrscheinlichkeiten ist nicht, wie in der Thermodynamik, ein Hilfsmittel, um Aussagen für große Gesamtheiten zu erhalten und auch kein Hinweis auf 'versteckte Variable'. Das Verhalten des einzelnen Teilchens ist bereits probabilistisch. Beim Messprozess folgt der Teilchenzustand gemäß der Wahrscheinlichkeitsverteilung ('Kollaps der Wellenfunktion').

Aufgabe 22: Warum kann der Wellenfunktion keine Wahrscheinlichkeitsinterpretation gegeben werden?

Wellengleichung des freien Elektrons

Betrachten wir die Bewegung eines freien Elektrons in 1-dim, d.h. eines massiven nichtrelativistischen Teilchens der Energie E mit $V \equiv 0$. Die zeitunabhängige Schrödinger-Gleichung (98) lautet in diesem Fall:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Psi(x) = E \Psi(x) \quad (99)$$

N.B. Im 1-dim Fall reduziert sich die Ortsabhängigkeit auf eine Variable, x .

Die Wellengleichung ist identisch zur Gleichung des harmonischen Oszillators (siehe Physik II), wenn Zeit t gegen die Ortskoordinate x ausgetauscht wird. Die Lösung ist eine ebene Welle in positiver oder negativer Richtung zwischen $-\infty < x < \infty$,

$$\Psi(x) = A e^{\pm ikx}, \quad (100)$$

Die Amplitude A ist konstant und bis auf Weiteres unbestimmt. Eingesetzt in die Schrödinger-Gleichung (99) folgt die **Dispersionsrelation des freien Elektrons**

$$E \equiv \frac{(\hbar k)^2}{2m}, \quad (101)$$

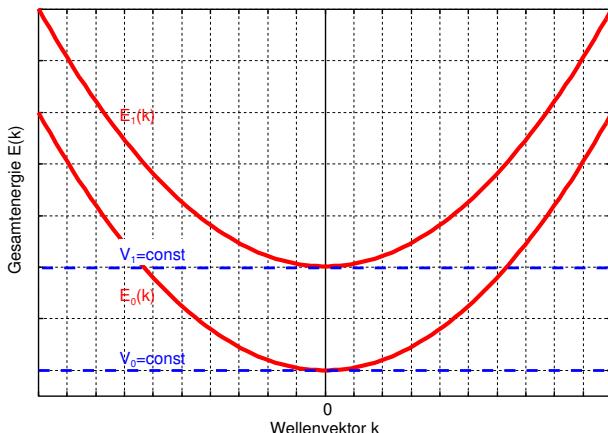
in Übereinstimmung mit der Energie eines freien nichtrelativistischen Teilchens mit Impuls $p \equiv \hbar k$.

Wir haben nun die Wellenfunktion eines freien Elektrons mit Impuls p erhalten. Wo befindet sich dieses Teilchen? Die Aufenthaltswahrscheinlichkeit wird durch die Wahrscheinlichkeitsdichte $|\Psi(x)|^2$ berechnet,

$$|\Psi(x)|^2 = |A|^2. \quad (102)$$

Die Wahrscheinlichkeit das Elektron an einer bestimmten Position x ist folglich für alle Positionen auf dem Intervall $x \in [-\infty, \infty]$ gleich. Die Position des Teilchens mit bestimmtem Impuls ist demnach völlig unbestimmt. Umgekehrt kann man zeigen, dass eine Wellenfunktion mit klar definierter Position im Raum eine Überlagerung unendlich vieler Wellen unterschiedlicher Wellenvektoren ist. Der Impuls eines Teilchens mit bestimmter Position ist demnach völlig unbestimmt. Diese Komplementarität wird **Heisenberg'sche Unschärferelation** genannt. Eine genaue quantenmechanische Rechnung ergibt eine untere Schranke für die Standardabweichung der Impuls- und Ortswerte, die hier, vollständigkeitshalber, nur angeführt wird:

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad (103)$$



Die **Dispersionsrelation eines Elektrons** in einem konstanten Potential ist in der Abbildung für zwei unterschiedliche Werte der potentiellen Energie gezeigt.

N.B. Bei Dispersionsrelationen wird oft auch die Energie als Funktion des Wellenvektors ausgedrückt, also $E(k)$.

Die effektive Masse des Elektrons folgt direkt aus der Dispersionsrelation:

$$\frac{1}{m_{\text{eff}}} := \frac{1}{\hbar^2} \frac{d^2 E(k)}{dk^2} \quad (104)$$

Aufgabe 23: Man nehme die mittlere kinetische Energie eines Leitungs-Elektrons bei Raumtemperatur laut Drudemodell als Maß für die zu erwartende Geschwindigkeits- und daher Impulsunschärfe Δp (m siehe Gleichung (42)). Macht es wellenmechanisch Sinn das Leitungselektron einem Atom zuzuordnen?

Unabhängige freie Elektronen im Raum

Die Wellenfunktion eines freien Elektrons mit Energie E in 3-dim ist gegeben durch eine räumliche Welle mit Wellenvektor \vec{k} ,

$$\Psi(\vec{x}) = Ae^{\pm i\vec{k}\cdot\vec{x}}. \quad (105)$$

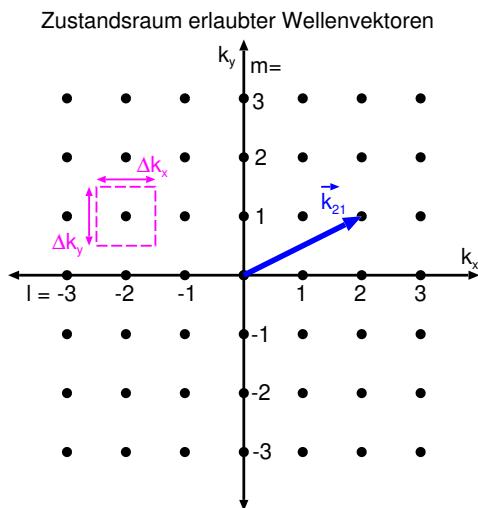
Die Anzahl möglicher Elektronenzustände pro Volumen, die sogenannte **Zustandsdichte**, wird berechnet, indem wir den Raum in hypothetische, identische kubische Volumina der Kantenlänge $L = V^{1/3}$ zerlegen. (Diese ad-hoc Annahme kürzt sich am Schluss wieder weg.) Die Annahme identischer Volumina bedeutet, dass nur Wellenvektoren erlaubt sind mit ganzzahliger Anzahl Wellenlängen im Volumen. Diese periodische Randbedingung (genannt **Born-von Karman Randbedingung**) bedeutet:

$$\Psi(x + L, y + L, z + L) \stackrel{!}{=} \Psi(x, y, z). \quad (106)$$

Als Konsequenz sind nur diskrete Wellenvektoren erlaubt, welche durch die ganzzahlige Anzahl Perioden pro Richtung charakterisiert werden können,

$$\vec{k}_{lmn} := \frac{2\pi}{L} \begin{pmatrix} l \\ m \\ n \end{pmatrix}, \quad \text{mit } l, m, n \in \{0, \pm 1, \pm 2, \pm 3, \dots\}. \quad (107)$$

Der Raum aller erlaubten Wellenvektoren sieht (für 2-dim) wie folgt aus:



Die x -Komponente jedes Wellenvektors, k_x wird durch die ganze Zahl l , die y -Komponente durch m und (nicht dargestellt) die z -Komponente durch n . Die möglichen erlaubten Wellenvektoren bilden folglich ein regelmässiges Gitter im 3-dim Raum der Wellenvektoren \vec{k} . Exemplarisch ist der erlaubte Vektor mit $l = 2$ und $m = 1$ dargestellt (blau).

Jedem erlaubten Wellenvektor kann nun ein Volumen Ω im Raum der Wellenvektoren (genannt **k -Raum**) zugesprochen werden (violett). Diese kubische Volumen mit Kantenlänge $\Delta k = \frac{2\pi}{L}$ lautet

$$\Omega := \left(\frac{2\pi}{L} \right)^3 = \frac{(2\pi)^3}{V}. \quad (108)$$

In einem Gas freier unabhängiger Elektronen werden nun diese möglichen Zustände, beginnend beim Zustand niedriger Energie, besetzt. Die Energie der möglichen Zustände wird ebenfalls wieder durch die ganzen Zahlen $l, m, n \in \{0, \pm 1, \pm 2, \pm 3, \dots\}$ ausgedrückt und ergibt sich aus der Dispersionsrelation,

$$E_{lmn} \equiv \frac{\hbar^2 \vec{k}_{lmn}^2}{2m}. \quad (109)$$

Die Energie E als Funktion des Wellenvektors \vec{k} beschreibt ein 3-dim Paraboloid über dem Raum der Wellenvektoren, d.h. über dem k -Raum. Die Schnitte dieser Hyperfläche entlang der Koordinatenebenen im k -Raum, (also z.B. der Ebene mit $k_y \equiv 0$ und $k_z \equiv 0$) ergibt die 1-dim Dispersionsrelation entlang dieser Wellenrichtung (siehe auch Phasenraum in Physik II).

Aufgabe 24: Angenommen die maximale Energie der Elektronen sei gegeben, E_{\max} , und daher er maximale Länge des Wellenvektors. Wie nimmt die Anzahl erlaubter Wellenvektoren mit der Maximalenergie E_{\max} zu?

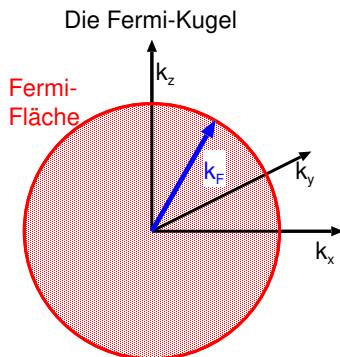
Die Fermi-Kugel

Elektronen gehören zur Familie der **Fermionen**, siehe Physik I, also zu den Teilchen mit halbzahligem Spin, $\frac{\hbar}{2}$. Fermionen unterliegen dem **Pauli-Prinzip** und folgen daher der **Fermi-Dirac Statistik**. Laut Pauli-Prinzip darf **jeder mögliche Zustand von maximal einem Elektron besetzt** sein. Aufgrund der zwei möglichen Ausrichtungen des Spins, also $+\frac{\hbar}{2}$ und $-\frac{\hbar}{2}$ kann jede mögliche Wellenzahl von maximal 2 Elektronen besetzt sein. Wir haben nun die gesuchte Vorschrift um die erlaubten Zustände zu besetzen.

Beim **absoluten Nullpunkt** $T = 0K$ verschwindet die thermische Bewegung der Teilchen (siehe Physik I) und alle Elektronen befinden sich auf den tiefsten möglichen Zuständen. Für ein Elektronengas mit vielen Teilchen sind daher alle verfügbaren Zustände bis zu einer maximalen Energie besetzt. Diese Energie ist die **Fermi-Energie** E_F (siehe Physik I) und entspricht einer maximalen Wellenzahl k_F ,

$$E_F := \frac{(\hbar k_F)^2}{2m}. \quad (110)$$

Im k -Raum sind daher alle Zustände mit $|\vec{k}_{lmn}| \leq k_F$ besetzt, alle anderen Zustände sind (beim absoluten Nullpunkt) unbesetzt siehe Abbildung.



Die besetzten Zustände beim absoluten Nullpunkt entsprechen einer Kugel im k -Raum mit 'Radius' k_F , die sogenannte **Fermi-Kugel**. Alle erlaubten Wellenvektoren innerhalb der Kugel entsprechen besetzten Zuständen, alle Wellenvektoren ausserhalb der Kugel sind unbesetzte und daher freie Zustände. Die Grenzfläche wird **Fermi-Fläche** genannt und beinhaltet die Elektronenzustände welche bereits bei mit geringer Energie freigesetzt werden können (und daher für die elektrische Leitfähigkeit verantwortlich sind, siehe später).

Betrachten wir nun ein System mit N Elektronen in einem Volumen V . Die Fermi-Energie E_F definiert das Volumen Ω_N der Fermi-Kugel im Raum der Wellenvektoren:

$$\Omega_N = \frac{4}{3}\pi k_F^3 \quad (111)$$

Die möglichen Wellenvektoren im k -Raum belegen jeweils ein Volumen Ω , Gleichung (108). Laut Pauli-Prinzip kann jedes dieser Volumina von zwei Elektronen (mit unterschiedlichem Spin) belegt sein. Wir können also das Volumen der Fermikugel mit den verfügbaren Zuständen in Verbindung setzen.

$$\Omega_N \stackrel{!}{=} \frac{N}{2} \Omega = 4\pi^3 \frac{N}{V}. \quad (112)$$

N.B.: Das willkürlich gewählte kubische Volumen V kommt nur noch als Teilchendichte $n := N/V$ vor. Unser Ergebnis gilt folglich für beliebig Volumina der Teilchendichte n .

Die maximale Wellenzahl k_F folgt nun aus Gleichung (111) und hängt nur von der Teilchendichte n ab.

$$k_F = \sqrt[3]{3\pi^2 n} \quad (113)$$

Wir erhalten für die **Fermi-Energie** freie Elektronen der Elektronendichte n ,

$$E_F = \frac{\hbar^2 (3\pi^2 n)^{2/3}}{2m}. \quad (114)$$

Die Fermi-Energie der Elektronen (in 3-dim) hängt nur von der Teilchendichte n ab und wächst $\propto n^{2/3}$.

Aufgabe 25: Berechne die Fermienergie E_F der Valenzelektronen für Kupfer am absoluten Nullpunkt (Tabelle Seite 14). Vergleiche diese Energie mit der mittleren Energie der Elektronen bei Raumtemperatur im Drudemodell. Passt das zusammen?

Zustandsdichte

In der Herleitung der Fermi-Energie, Gleichung (114), wurde angenommen, dass sehr viele Teilchen im betrachteten Volumen existieren, und daher die diskrete Natur der erlaubten Wellenvektoren nicht relevant ist. (Es wurde eine Kugel mit Würfeln aufgefüllt.) Im Falle elektrischer Leitung in Festkörpern ist zumeist von hohen Elektronendichten (z.B. $n > 10^{16} m^{-3}$) und daher sehr eng aneinanderliegenden diskreten Energien auszugehen. Aus diesem Grund ist es oft einfacher, die Energie wieder als kontinuierliche Größe zu betrachten, ohne dabei die quantenmechanische Natur der Zustände zu ignorieren.

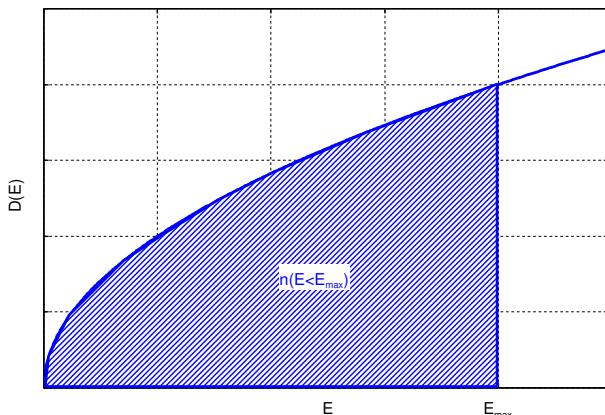
Die **Zustandsdichte** $D(E)$ ist genau diese Funktion, die die Elektronendichte als Funktion der Elektronenenergie ausdrückt,

$$D(E) := \frac{dn}{dE}. \quad (115)$$

Für das Fermigas freier Elektronen folgt die Zustandsdichte $D(E)$ aus der Fermienergie, Gleichung (114),

$$D(E) = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \sqrt{E}. \quad (116)$$

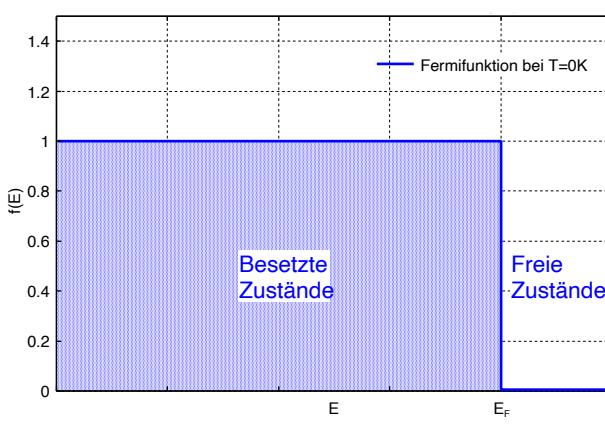
Die Zustandsdichte gibt an, wieviele neue Zustände mit steigender Energie hinzukommen, siehe Abbildung. Im Falle eines 3-dim Elektronengases in steigt die Anzahl möglicher Zustände mit der Energie $\propto \sqrt{E}$.



Die Fläche unter der Zustandsdichte gibt die Dichte der Elektronenzustände an. Die Dichte der Elektronenzustände in einem bestimmten Intervall $E \in [E_{\min}, E_{\max}]$ ist folglich gegeben durch

$$n(E_{\min} \leq E \leq E_{\max}) = \int_{E_{\min}}^{E_{\max}} D(E) dE. \quad (117)$$

Die Zustandsdichte ist lediglich eine Aussage über die verfügbaren Zustände. Ob ein Zustand tatsächlich besetzt ist wird durch die Besetzungszahl, die **Fermi-Funktion** $f(E)$, siehe Physik I, ausgedrückt.



Im betrachteten Fall freier Elektronen am absoluten Nullpunkt, sind alle Zustände bis zur Fermi-Energie E_F besetzt und alle Zustände über der Fermi-Energie unbesetzt, die Fermi-Funktion $f(E)$ ist eine Stufenfunktion, siehe Abbildung. Die Dichte **aller freier Elektronen** ist nun das Integral über die Dichte der verfügbaren Zustände (Zustandsdichte) gewichtet mit der tatsächlichen Besetzungszahl,

$$n = \int_0^{\infty} D(E) f(E) dE. \quad (118)$$

Aufgabe 26: Nehmen wir an, dass die freien Elektronen auf eine dünne Schicht beschränkt sind, das Elektronengas kann sich folglich nur in 2-dim bewegen. Zeige, dass die Zustandsdichte in 2-dim eine Konstante ist?

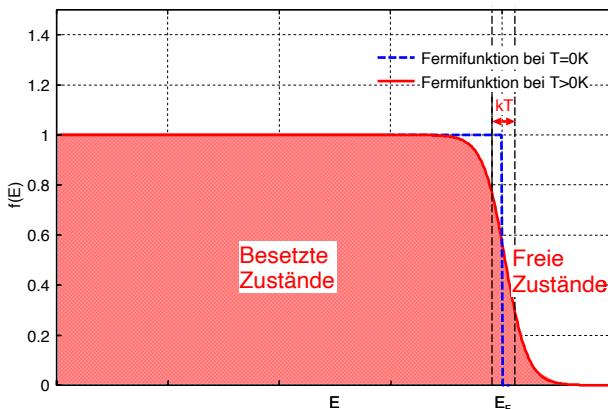
Fermigas bei T>0K

Ein Gas freier Fermionen, auch **Fermigas** genannt, verhält sich thermisch im Sinne der **Fermi-Dirac Statistik**, siehe Physik I. Das thermische Gleichgewicht entspricht dem Zustand maximaler Entropie bei der Verteilung der Teilchen, unter Einhaltung des Pauli-Prinzips und des Makrozustandes (z.B. gegebene Teilchenzahl N und innere Energie U) auf die möglichen Zustände. Während bei $T = 0K$ nur die niedersten Zustände bis zur Fermi-Energie besetzt sind, werden bei $T > 0K$ von einem gewissen Anteil Teilchen nahe der Fermienergie (d.h. Teilchen auf der Fermi-Fläche) auch Zustände oberhalb der Fermienergie besetzt und entsprechend Zustände unter der Fermi-Energie nicht besetzt, siehe Abbildung.

Bei eng aneinanderliegenden Energiezuständen kann die Besetzungszahl durch ein kontinuierliche Funktion der Energie, die **Fermi-Funktion** ausgedrückt werden, siehe Physik I

$$f(E) := \frac{1}{e^{(E-\mu)/kT} + 1}. \quad (119)$$

Das chemische Potential μ entspricht bei $T = 0K$ gerade der Fermi-Energie E_F .



Am absoluten Nullpunkt ist die Fermifunktion eine reine Stufenfunktion, bei endlichen Temperaturen weicht sich die Stufe in einen graduellen Übergang zwischen den besetzten Zuständen niederer Energie und den unbesetzten Zuständen hoher Energie auf, siehe Abbildung. Die Breite dieses Überganges wird durch die thermische Energie kT bestimmt.

Die Besetzungswahrscheinlichkeit der Zustände deutlich über der Fermi-Energie nimmt entsprechend des **Boltzmann-Faktors** ab.

$$f(E > E_F) \propto e^{-E/kT}. \quad (120)$$

Die Dichte der Ladungsträger bei Temperatur T in einem gewissen Energieintervall wird nun wieder durch die Fläche unter der Zustandsdichte, multipliziert mit der Fermi-Funktion berechnet.

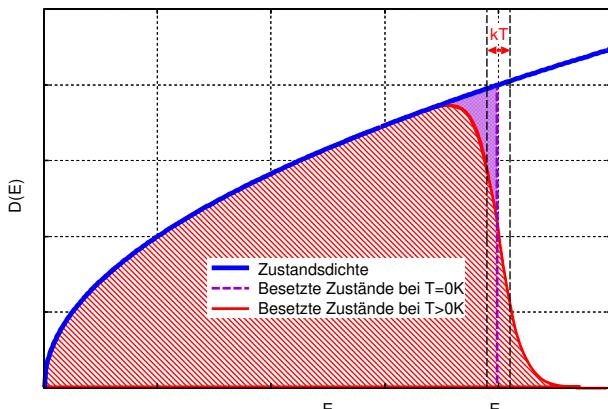
Die **Elektrondichte** n oberhalb der Fermi-Energie bei $T > 0K$ lautet z.B.

$$n(E > E_F) = \int_{E_F}^{\infty} D(E)f(E)dE. \quad (121)$$

Analog können die nichtbesetzten freien Zustände unterhalb der Fermi-Energie als fehlende negative Ladungen und daher als 'positive Elektronen' oder **Löcher** interpretiert werden. Die Löcherdichte p unterhalb der Fermi-Energie bei $T > 0K$ lautet z.B.

$$p(E < E_F) = \int_0^{E_F} D(E)(1 - f(E))dE. \quad (122)$$

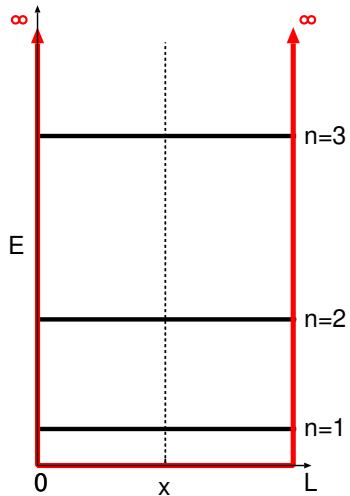
Aufgabe 27: Was folgt aufgrund der Fermi-Funktion für die Ladungsträgerdichte (Elektronen und Löcher) in Abhängigkeit der Temperatur? Welcher Vorzeichen ergibt sich für den Temperaturkoeffizienten des spezifischen Widerstandes?



4 – Gebundene Ladungen

4.1 – Elektronen in einem Potentialtopf

Bisher wurden freie Elektronen behandelt, d.h. Elektronen in einem konstanten Potential $V(\vec{x})$, die Elektronen in einem Atom sind hingegen gebundene Elektronen. Auf die Behandlung der Schrödinger-Gleichung in einem 3-dim elektrischen Potential einer Punktladung, d.h. unter der Coulombkraft (13), wird hier verzichtet. Die wesentlichen Effekte sind bereits in einem deutlich einfacheren Beispiel, dem **1-dim Potentialtopf**, ersichtlich, siehe Abbildung.



Ein Elektron sei in einem unendlich tiefen 1-dim Potentialtopf der Breite L gebunden. Die potentielle Energie im Topf sei auf $V(x) = 0$ festgelegt. Am Rand des Topfes steige das Potential jeweils auf ∞ .

$$V(x) = \begin{cases} 0, & \text{für } 0 < x < L \\ \infty, & \text{sonst.} \end{cases} \quad (123)$$

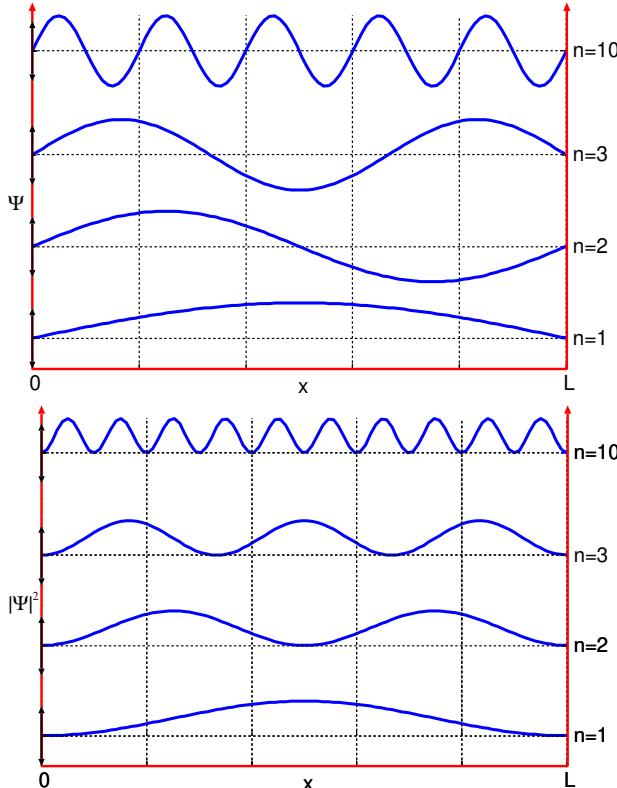
Die stationäre Schrödinger-Gleichung lautet

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \Psi(x) = E \Psi(x). \quad (124)$$

Ausserhalb des Potentialtopfes kann die Gleichung nur von einer verschwindenden Wellenfunktion erfüllt werden,

$$\Psi(x \leq 0) \equiv \Psi(x \geq L) \equiv 0, \quad (125)$$

Im Inneren des Potentialtopfes erhalten wir die Schrödinger-Gleichung des freien Elektrons, Gleichung (99).



Die Randbedingung, Gleichung (125), erlaubt nur ganzzahlige Vielfache der halben Wellenlängen in der Schachtelebene L , siehe Abbildung,

$$k_n := \frac{\pi}{L} n, \text{ for } n \in \{1, 2, 3, \dots\} \quad (126)$$

und daher **diskrete Energiewerte**:

$$E_n := \frac{\hbar^2}{2m} \left(\frac{\pi}{L} \right)^2 n^2 \quad (127)$$

Die möglichen **Energiezustände sind quantisiert**. Die Wellenfunktion eines einzelnen Elektrons mit der n -ten Energie im 1-dim Potentialtopf lautet

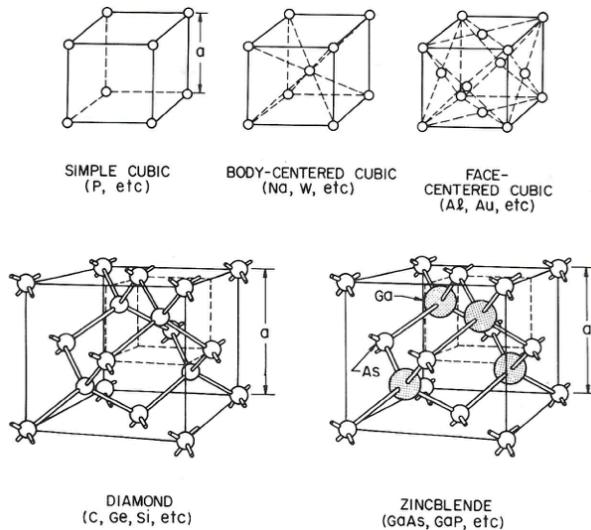
$$\Psi_n(x) := \sqrt{\frac{2}{L}} \sin \frac{\pi}{L} nx. \quad (128)$$

Die entsprechende Wahrscheinlichkeitsdichte $|\Psi_n|^2$ entspricht einer Gesamtwahrscheinlichkeit $P(-\infty < x < \infty) \equiv 1$.

Aufgabe 28: Zeige, dass die Wellenfunktion $\Psi_n(x)$ eines Elektrons der Energie E_n zu einer korrekten Aufenthaltswahrscheinlichkeit führt (d.h. Wahrscheinlichkeit 1 im Topf und Wahrscheinlichkeit 0 außerhalb).

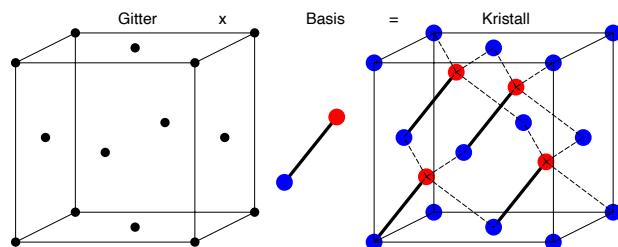
4.2– Elektronen im periodischen Potential

Gebundene Elektronenzustände (auch im 3-dim Fall) einzelner Elektronen in einem definierten Potentiaaltopf weisen ein diskretes Spektrum erlaubter Energien auf. In einem Festkörper sind jedoch immer viele Elektronen gleichzeitig vorhanden. Zusätzlich befinden sich diese Elektronen in einer periodischen Anordnung positiv geladener Atomkerne, d.h. ist einem räumlich periodischen Potential, einer **Kristallstruktur**.



Quelle: S. M. Sze, *Physics of Semiconductor Devices*

Silizium und Germanium bilden eine Diamant-Struktur mit jeweils 4 nächsten Nachbarn, III-V Halbleiter wie Galliumarsenid oder Galliumphosphid kristallisieren in der Zinkblende-Struktur, siehe Abbildung. Die Diamant-Struktur und die Zinkblende-Struktur kann aus einem kubisch-flächenzentrierten Gitter bilden.



Die meisten Halbleiter weisen ein kubisches Gitter auf. Hier wird zwischen dem einfachen kubischen Gitter (oben links), dem kubisch-raumzentrierten Gitter (oben mittte) und dem kubisch flächenzentrierten Gitter (oben rechts) unterschieden. Das kubisch-raumzentrierte Gitter erhält man, indem zwei einfache kubische Gitter um eine halbe Raumdiagonale gegeneinander verschoben werden. Das kubisch-flächenzentrierte Gitter erhält man, wenn vier einfache kubische Gitter jeweils um eine halbe Flächendiagonale jeder Raumrichtung gegeneinander verschoben werden.

Betrachten wir ein kubisch-flächenzentriertes Gitter, Abbildung links. An jedem Gitterplatz wird nun eine nichttriviale Basis aus zwei Atomen eingefügt (rot-blauer Keule) und es entsteht die Diamant- bzw. Zinkblende-Struktur. Die jeweils 4 nächsten Nachbarn sind mit strichlierten Linien verbunden.

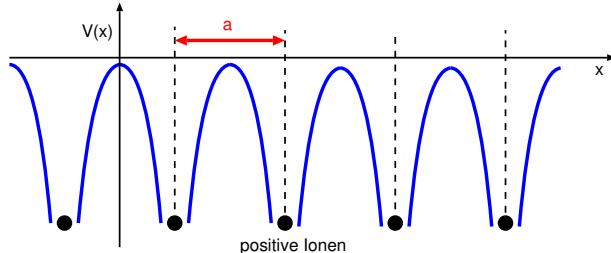
Die Lösung der Schrödinger-Gleichung für ein periodisches Gitter entspricht der Wellengleichung eines periodischen Wellenleiters (einer Perlenkette). Nachdem die Kristallstruktur aus nichttrivialen Basiselementen aufgebaut ist, erwarten wir eine Dispersionsrelation mit mehreren Frequenzbändern, welche in jede Raumrichtung unterschiedlich ausfallen.

Auf den folgenden Seiten lösen wir die Schrödinger-Gleichung eines Elektrons für den einfachsten Fall eines 1-dim Kristalls mit trivialer Basis. Die Behandlung eines 3-dim Kristallgitters mit nichttrivialer Basis verläuft analog, wird aber typischerweise numerisch gelöst.

Aufgabe 29: Wie kann die Diamantstruktur aus zwei kubisch flächenzentrierten Gittern gewonnen werden? Wieviele Atome befinden sich in einem Würfel der Kantenlänge a ?

Periodisches Potential in 1-dim: Das Bloch-Floquet Theorem

Eine 1-dimensionale **Kette positiver Ionen** bildet das periodische Potential für die Schrödinger-Gleichung eines 1-dim Kristalls, siehe Abbildung. Die potentielle Energie (blaue Linie) erreicht bei den Ionen (Abstand a) ihr Minimum, zwischen zwei Ionen ist das Potential jeweils maximal, aber endlich.



Die statische Schrödinger-Gleichung (98),

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \Psi(x) = E \Psi(x), \quad (129)$$

kann nun geschrieben werden als

$$\frac{d^2}{dx^2} \Psi(x) + \kappa^2(x) \Psi(x) = 0. \quad (130)$$

Die Abkürzung $\kappa(x) := \sqrt{(E - V(x)) \frac{2m}{\hbar^2}}$ beinhaltet die Periodizität des Problems, $\kappa(x+a) \stackrel{!}{=} \kappa(x)$.

Jede Differentialgleichung zweiter Ordnung hat zwei linear unabhängige Lösungen, d.h. **zwei linear unabhängige Eigenfunktionen** $\Psi_1(x)$ und $\Psi_2(x)$. Die Periodizität der Problemstellung muss sich in den Eigenfunktionen niederschlagen, darf aber die Physik nicht beeinflussen. Wir erwarten daher eine Aufenthaltswahrscheinlichkeit $|\Psi_i(x)|^2$ periodisch in a und eine Wellenfunktion die maximal durch einen komplexen Faktor μ_i mit Betrag 1 geändert wird. Für jede der linear unabhängigen Eigenfunktionen gilt folglich

$$\Psi_i(x+a) = \mu_i \Psi_i(x). \quad (131)$$

Die Eigenfunktionen bestehen nun aus einem periodischen Anteil $u_i(x) \equiv u_i(x+a)$ und einem Phasenfaktor,

$$\Psi_i(x) = \mu_i^{x/a} u_i(x). \quad (132)$$

Die Wellengleichung, (130), für jede Eigenmode, ergibt nun folgende Gleichungen,

$$\frac{d^2}{dx^2} \Psi_1 = -\kappa^2 \Psi_1 \quad (133)$$

$$\frac{d^2}{dx^2} \Psi_2 = -\kappa^2 \Psi_2 \quad (134)$$

welche geeignet kombiniert folgende Relation ergeben:

$$0 = \Psi_2 \frac{d^2}{dx^2} \Psi_1 - \Psi_1 \frac{d^2}{dx^2} \Psi_2 = \frac{d}{dx} \left(\Psi_2 \frac{d}{dx} \Psi_1 - \Psi_1 \frac{d}{dx} \Psi_2 \right) \quad (135)$$

Wir haben also durch Linearkombination der Moden eine **Konstante der Bewegung** erhalten

$$\left(\Psi_2 \frac{d}{dx} \Psi_1 - \Psi_1 \frac{d}{dx} \Psi_2 \right) = \text{const.} \quad (136)$$

Setzt man den Ansatz für die Eigenfunktionen (132) in Gleichung (136) ein, ergibt sich die Bedingung,

$$\mu_1 \mu_2 \stackrel{!}{=} 1. \quad (137)$$

Die Phasenfaktoren μ_i sind folglich komplexe konjugierte Wurzeln aus 1 und können geschrieben werden als

$$\mu_1 =: e^{ika} \text{ und } \mu_2 =: e^{-ika}. \quad (138)$$

Wir haben das **Bloch-Floquet Theorem** erhalten:

Die Lösung der Wellenfunktion in einem periodischen Potential hat die Form

$$\Psi(x) = e^{ikx} u_k(x) \quad (139)$$

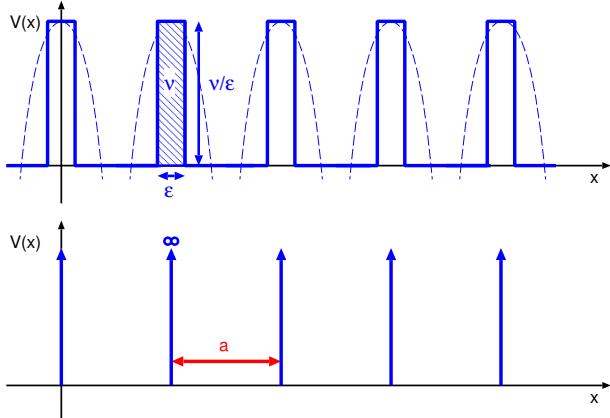
mit der periodischen Funktion $u_k(x+a) = u_k(x)$ und Eigenwerten $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$.

Aufgabe 30: Die Lösung entspricht einer periodisch amplituden-modulierten ebenen Welle der Wellenzahl k . Wie reduziert sich das Ergebnis auf die Wellengleichung der Perlenkette?

Periodisches Potential in 1-dim: Das Kronig Penney Modell

Das einfachste analytische Modell für ein 1-dim periodisches Potential ist das **Kronig Penney Modell**. Hierzu wird das Coulomb-Potential durch endliche Topfpotentiale approximiert, siehe Abbildung. Der Nullpunkt der Energie wird, ohne Beschränkung der Allgemeinheit, im Inneren der Topfpotentiale festgelegt.

Das Potential verschwindet bis auf Berge endlicher Höhe ν/ε und Breite ε im Abstand a ,



Die mittlere potentielle Energie $\langle V \rangle$ definiert die Pulshöhe $\nu = a \langle V \rangle$.

Im Grenzfall unendlich schmaler Pulse $\varepsilon \rightarrow 0$ gleicher mittlerer Energie $\langle V \rangle$ wird das Potential durch **Dirac-Pulse** ^a $\delta(x)$ dargestellt

$$V(x) = \sum_{n=-\infty}^{n=\infty} \nu \delta(x - na). \quad (141)$$

^a $\delta(x) \equiv 0$ für $x \neq 0$ und $\int_{-\infty}^{\infty} \delta(x) dx = 1$

Im Grenzfall großer Wellenlängen ($\lambda \gg a$) darf die Lösung der Schrödinger-Gleichung, Gleichung (130), nicht von den Details der Potentialfunktion abhängen und wir können die periodische Funktion $\kappa(x)$ durch ihren räumlichen Mittelwert κ ersetzen,

$$\kappa^2 \approx \frac{2m}{\hbar^2} (E - \langle V \rangle). \quad (142)$$

Die Schrödinger-Gleichung wird zu einer Wellengleichung ebener Wellen mit Wellenzahl $k \equiv \pm \kappa$ mit der **Dispersionsrelation** eines freien Elektrons in einem konstanten Potential $\langle V \rangle$, siehe Kapitel 2.1.

$$E \approx \frac{\hbar^2 k^2}{2m} + \langle V \rangle. \quad (143)$$

Die Lösung der Schrödinger-Gleichung für allgemeine Wellenlängen lässt sich am einfachsten im Grenzfall unendlich schmaler Pulse (also Dirac-Pulse) berechnen. In diesem Fall lautet die Schrödinger-Gleichung zwischen dem n -ten und dem $(n+1)$ -ten Puls, also für $na < x < (n+1)a$

$$\left[\frac{d^2}{dx^2} + \frac{2m}{\hbar^2} E \right] \Psi(x) = 0. \quad (144)$$

Mithilfe der Abkürzung $q := \sqrt{\frac{2m}{\hbar^2} E}$ lautet die allgemeine Lösung in diesem Intervall

$$\Psi(x) = A \cos qx + B \sin qx, \quad (145)$$

mit den Integrationskonstanten A und B .

Für die Lösung über alle Intervalle, müssen nun die Lösungen freier Elektronen in jedem Intervall, Gleichung (145), unter Verwendung des Bloch-Floquet Theorems (139) zu einer Gesamtlösung zusammengeklebt werden. Die notwendigen Anschlussbedingungen an den Gitterstellen $x = na$ werden im Folgenden hergeleitet.

Aufgabe 31: Zeige, dass sich die Potentialfunktion (140) im Grenzwert $\varepsilon \rightarrow 0$ auf die Serie von Dirac-Pulsen, Gleichung (141) reduziert, bei gleicher mittlerer potentiellen Energie $\langle V \rangle$.

Anschlussbedingungen: Das Kronig Penney Modell

Die geeigneten 'Klebebedingungen' an einem unendlich schmalen Potentialpuls, z.B. bei $x = 0$, ergibt sich direkt aus der Schrödinger-Gleichung. Wie integrieren hierzu die Schrödinger-Gleichung

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \sum_{n=-\infty}^{n=\infty} \nu \delta(x - na) - E \right] \Psi(x) = 0, \quad (146)$$

über ein schmales Intervall um den Dirac-Puls bei $x = 0$, also $x \in [-\epsilon, \epsilon]$:

$$-\frac{\hbar^2}{2m} \int_{-\epsilon}^{\epsilon} \frac{d^2}{dx^2} \Psi(x) dx + \nu \int_{-\epsilon}^{\epsilon} \delta(x) \Psi(x) dx - E \int_{-\epsilon}^{\epsilon} \Psi(x) dx = 0. \quad (147)$$

Das erste Integral lässt sich leicht lösen und ergibt die Differenz der Ableitungen an $x + \epsilon$ und $x - \epsilon$,

$$\int_{-\epsilon}^{\epsilon} \frac{d^2}{dx^2} \Psi(x) dx = \Psi'(x + \epsilon) - \Psi'(x - \epsilon). \quad (148)$$

Das letzte Integral kann durch den Mittelwertsatz der Integralrechnung ^a dargestellt werden,

$$\int_{-\epsilon}^{\epsilon} \Psi(x) dx = 2\epsilon \Psi(\xi), \text{ für } \xi \in [-\epsilon, \epsilon]. \quad (149)$$

Für das mittleren Integral stellen wir die Wellenfunktion im Intervall $x \in [-\epsilon, \epsilon]$ durch eine Taylor-Reihe dar, $\Psi(x) \approx \Psi(0) + x\Psi'(0) + \dots$. Laut Definition verschwindet der Dirac-Puls und daher das Integral an allen Stellen außer $x = 0$. Für die Integration ist daher nur die Wellenfunktion an der Stelle $x = 0$ relevant

$$\int_{-\epsilon}^{\epsilon} \delta(x) \Psi(x) dx = \Psi(0) \int_{-\epsilon}^{\epsilon} \delta(x) dx = \Psi(0). \quad (150)$$

Im Grenzwert $\epsilon \rightarrow 0$ wird das Argument $x - \epsilon$ bzw. $x + \epsilon$ durch $0-$ bzw. $0+$ abgekürzt,

$$\Psi(0-) := \lim_{\epsilon \rightarrow 0} \Psi(-\epsilon) \quad (151)$$

$$\Psi(0+) := \lim_{\epsilon \rightarrow 0} \Psi(\epsilon), \quad (152)$$

und das letzte Integral verschwindet. Die Resultat um den Impuls bei $x = 0$, Gleichung (147), lautet daher

$$-\frac{\hbar^2}{2m} [\Psi'(0+) - \Psi'(0-)] + \nu \Psi(0) = 0. \quad (153)$$

Wir haben nun die **Anschlussbedingungen** für die Wellenfunktionen links und rechts eines Dirac-Pulses erhalten. Die Wellenfunktion muss stetig sein, die erste Ableitung knickt abhängig von der Potentialhöhe ν :

$$\Psi(0+) \stackrel{!}{=} \Psi(0-) \quad (154)$$

$$\Psi'(0+) \stackrel{!}{=} \Psi'(0-) + \frac{2m\nu}{\hbar^2} \Psi(0). \quad (155)$$

Für die 'Klebebedingungen' an den anderen Dirac-Pulsen (d.h. bei $x = na$) kann das Bloch-Floquet Theorem verwendet werden, bzw. die Wellenfunktion durch eine sogenannte **Bloch Funktion** dargestellt werden,

$$\Psi(x + a) = e^{ika} \Psi(x). \quad (156)$$

Aufgabe 32: Warum muss die Wellenfunktion stetig sein? Was würde folgen, wenn die Wellenfunktion an der Anschlussstelle springt?

^a $\int_c^d f(x) dx = f(\xi)(d - c)$ mit $\xi \in [c, d]$

Dispersionsrelation: Das Kronig Penney Modell

Für die gesamte Wellenfunktion muss die Lösung des freien Elektrons im Intervall $x \in (0, a)$, Gleichung 145, an die Anschlussbedingungen (155) angepasst werden. Während $\Psi(0+)$ und $\Psi'(0+)$ im Intervall liegen und daher sofort angegeben werden können, muss die Wellenfunktion an der Stelle $0-$ mithilfe der Blochfunktion (156) auf einen äquivalenten Punkt im Intervall bezogen werden,

$$\Psi(0-) := \lim_{\epsilon \rightarrow 0} \Psi(\epsilon) = \lim_{\epsilon \rightarrow 0} e^{-ika} \Psi(a - \epsilon) = e^{-ika} \Psi(a-). \quad (157)$$

Die Wellenfunktion im Intervall $0 < x < a$ für die Anschlussbedingungen lauten nun

$$\Psi(0+) = A \quad (158)$$

$$\Psi(0-) = e^{-ika}(A \cos qa + B \sin qa) \quad (159)$$

$$\Psi'(0+) = qB \quad (160)$$

$$\Psi'(0-) = e^{-ika}(-qA \sin qa + qB \cos qa). \quad (161)$$

Eingesetzt in die Anschlussbedingungen (155) folgt ein Gleichungssystem für die Amplituden A und B .

$$A - e^{-ika}(A \cos qa + B \sin qa) = 0 \quad (162)$$

$$B - e^{-ika}(-A \sin qa + B \cos qa) = \frac{2m}{q\hbar^2} \nu A. \quad (163)$$

Dieses homogenes Gleichungssystem kann als Matrizengleichung geschrieben werden:

$$\begin{pmatrix} 1 - e^{-ika} \cos qa & -e^{-ika} \sin qa \\ e^{-ika} \sin qa - \frac{2m}{q\hbar^2} \nu & 1 - e^{-ika} \cos qa \end{pmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix} = 0 \quad (164)$$

Eine nichttriviale Lösung ist nur bei verschwindender Determinante möglich und ergibt folgende notwendige Bedingung zwischen die Wellenzahl k und der Größe q ,

$$\cos ka \stackrel{!}{=} \cos qa + \frac{m\nu}{\hbar^2 q} \sin qa. \quad (165)$$

Nachdem die Energie nur durch q gegeben ist, $E = \frac{\hbar^2}{2m} q^2$ handelt es sich hier um eine implizite Energie-Impulsbeziehung, d.h. um die **Dispersionsrelation** der Elektronen in einem periodischen 1-dim Potential.

Im Grenzfall verschwindender Potentiale, $\nu \equiv 0$, lässt sich die Dispersionsrelation exakt lösen und wir erhalten $k = \pm q$ und somit die Dispersionsrelation eines freien Teilchens ohne Einschränkung. Im Falle großer Wellenlängen, $\lambda \gg a$, also für $ka \ll 1$ kann die Dispersionsrelation in eine Taylorreihe entwickelt werden. Kleine Werte für ka implizieren auch kleine Werte für den Energieparameter q , es werden daher die Taylorreihen für $qa \ll 1$ benötigt. Mit Hilfe von $\cos x \approx 1 - \frac{1}{2}x^2$ und $\sin x \approx x$ folgt die Beziehung

$$E = \frac{\hbar^2 q^2}{2m} \approx \frac{\hbar^2 k^2}{2m} + \frac{\nu}{a}, \quad (166)$$

in Übereinstimmung mit Gleichung (143). Für allgemeine Wellenvektoren kann die Dispersionsrelation nur numerisch gelöst werden.

Aufgabe 33: Wie lautet die Herleitung der Dispersionsrelation aus der Determinante?

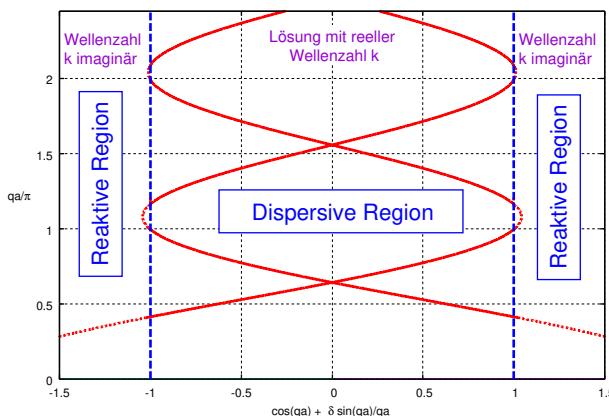
Bändermodell: Das Kronig Penney Modell

Die Dispersionsrelation des Kronig Penney Modells ist eine transzendentale Gleichung und besteht aus trigonometrischen Funktionen der Energie, ausgedrückt durch den Energieparameter $q := \sqrt{\frac{2m}{\hbar^2} E}$, und der Wellenzahl,

$$\cos ka = \cos qa + \delta \frac{\sin qa}{qa}. \quad (167)$$

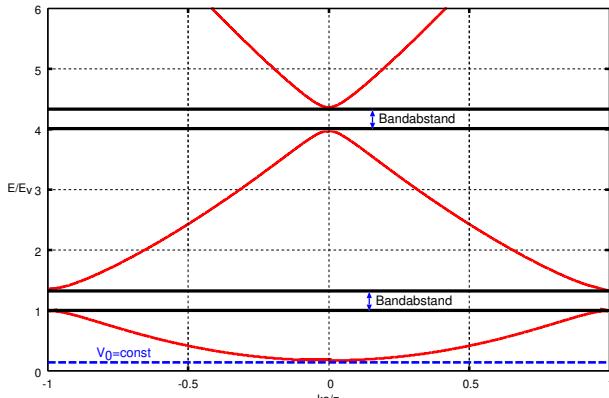
Die Potentialhöhe wurde hier durch den Parameter $\delta := \frac{mv_0}{\hbar^2}$ abgekürzt.

Die linke Seite der Dispersionsrelation kann für reelle Wellenvektoren nur Werte im Intervall $[-1, 1]$ annehmen, siehe strichlierten Bereich in der Abbildung.



Die rechte Seite kann, als Summe zweier trigonometrische Funktionen, auch Werte außerhalb dieses Intervalls annehmen. In der Abbildung links ist die rechte Seite der Dispersionsrelation als Abszisse (mit beiden möglichen Vorzeichen) und der Energieparameter q als Ordinate aufgetragen. Für Energien mit einer rechten Seite im strichlierten Intervall ergeben sich dispersive Wellen mit reellen Wellenzahlen, in den Bereichen außerhalb des Intervalls kann die Dispersionsrelation nur mit komplexen Wellenzahlen, d.h. nur von reaktiven Wellen erfüllt werden.

Die **Dispersionsrelation** besteht folglich aus mehreren Bändern, mit verbotenen Energie-Bereichen für ebenen Wellen. In der Abbildung ist das **Bänderschema** des 1-dim periodischen Potentials dargestellt.



Die Wellenvektoren können nur Werte aus dem Intervall $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$ annehmen, wobei die Bänder an den Intervallgrenzen immer horizontal verlaufen. In niedrigsten Band erhalten wir wie erwartet die Dispersionsrelation eines freien Elektrons der Masse m im konstanten Potential $< V >$. Das nächste Band entspricht der Energie eines negativen Elektrons mit Masse entsprechend der Krümmung bei $k = 0$. Das folgende Band entspricht wieder einem freien Elektron, allerdings mit geringerer effektiver Masse.

Der Abstand zwischen den Bändern wird **Bandabstand** genannt und entspricht den Energiebereichen in denen keine dispersive Wellenfunktionen möglich sind (d.h. die Wellenfunktionen dieser Energien klingen exponentiell ab). Die Energiefülle des Bandabstandes, genannt **Bandlücke** E_g , ist im Kronig Penney Modell für alle Bänder gleich und beträgt

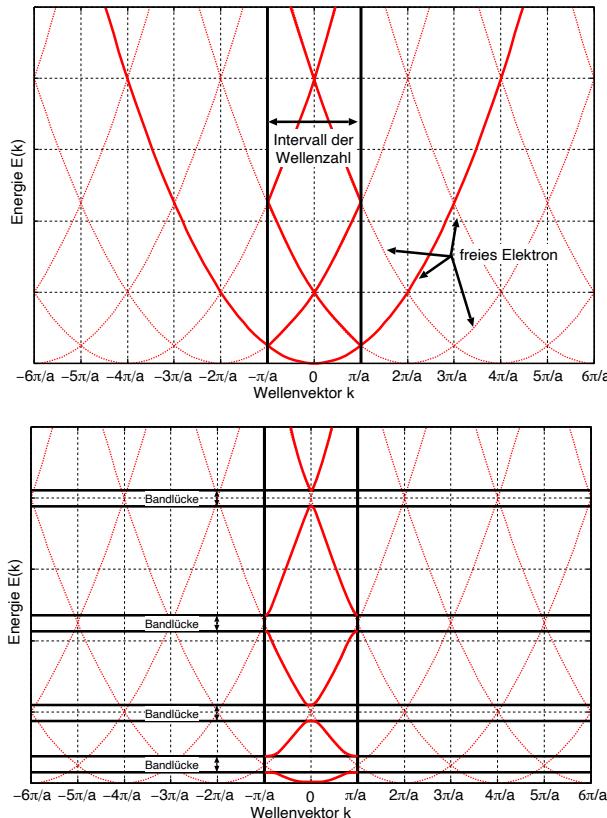
$$E_g = 2 < V >. \quad (168)$$

Im Fall eines 3-dim Kristalles kann das Bänderschema in jede Richtung des Wellenvektors unterschiedlich aussehen und die Bandlücken, also das 'verbotene' Energieintervall zwischen der höchsten Energie eines Bandes und der tiefsten Energie des folgenden Bandes, für jeden Bandabstand unterschiedlich sein.

Aufgabe 34: Berechne die Bandlücke im Kronig Penney Modell aus der Dispersionsrelation für kleine Potentialhöhen.

Bänderschema

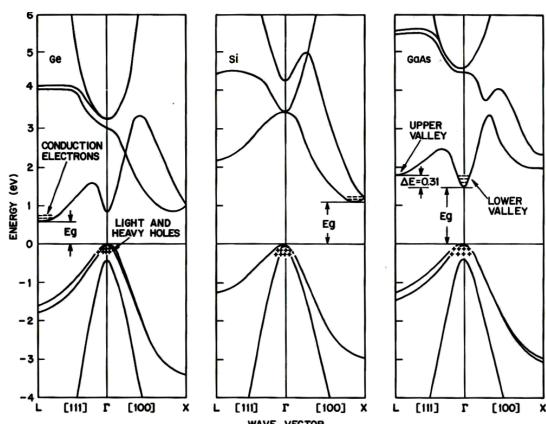
Das Bänderschema eines Kristalls kann qualitativ wie folgt konstruiert werden:



- Betrachte die Periode a eines Kristalls in eine gewisse Richtung.
- Die Wellenvektoren im Abstand $\frac{2\pi}{a}$ sind äquivalent.
- Zeichen die Parabeln freier Elektronen für alle äquivalenten Wellenvektoren.
- Aufgrund der Periodizität ist nur der Bereich $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$ relevant.
- Die Äste der freien Dispersionsrelationen im relevanten Wellenvektorbereich ergeben die einzelnen Bänder (Abbildung oben).
- An den Kreuzungspunkten der Bänder werden die freien Parabeln zu horizontalen Kurven, ebenso an den Grenzen $k = \pm\pi/a$.
- Der Abstand dieser horizontalen Bereiche ist jeweils der **Bandabstand**, z.B. im Falle des Kronig-Penney Models $E_g = 2 < V >$.

Mit dieser Konstruktion können Bänderschemata in eine bestimmte Richtung des Wellenvektors qualitativ erzeugt werden.

Im Falle eines realen 3-dim Kristalles wird das Bänderschema numerisch ermittelt. Da nun unendlich viele Richtungen von Wellenvektoren möglich sind, wird zumeist die Dispersionrelation nur entlang der wichtigsten Symmetrierichtungen des Kristalls dargestellt und die k -Werte der verschiedenen Richtungen horizontal aneinander gereiht. Die Dispersionsrelationen von drei wichtigen Halbleitern (Ge , Si , $GaAs$) lauten:



Quelle: S. M. Sze, Physics of Semiconductor Devices

In einem Halbleiter am absoluten Nullpunkt sind alle Bänder unterhalb einer gewissen Energie (Fermienergie) besetzt. Diese höchste besetzte Energie definiert die Energie des Valenzbandes, E_v . Das nächsthöhere Band entspricht den niedrigsten freien Elektronenzustände und definiert die Energie des Leitungsbandes E_c . Der Bandabstand E_g ist per Definitionem der Abstand zwischen dem Leitungsband und dem Valenzband. Die Krümmung des Leitungsbandes und des Valenzbandes entspricht der effektiven Masse der entsprechenden freien Elektronen bzw. 'Löcher'.

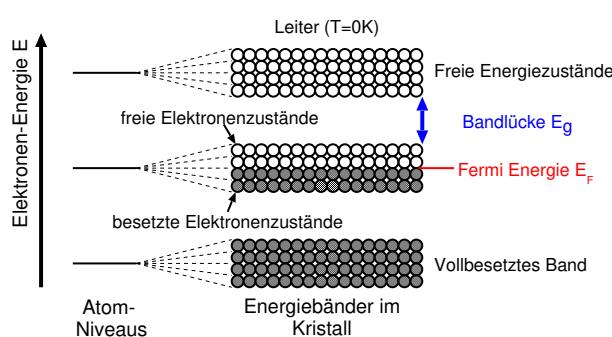
Bei Si und Ge ist die Richtung der freien Elektronen- und Lochzuständen unterschiedlich (d.h. Minimum des Leitungsbandes liegt nicht über dem Maximum des Valenzbandes). In diesem Fall spricht man von einem **indirekten Halbleiter**. Im Gegensatz dazu ist $GaAs$ ein **direkter Halbleiter**, mit freien Elektronen und Löchern bei gleichem Wellenvektor, d.h. gleichem Impulsvektor.

Aufgabe 35: Warum eignen sich direkte Halbleiter für Leuchtdioden, indirekte Halbleiter aber nicht?

5 – Ladungskonzentration im Halbleiter

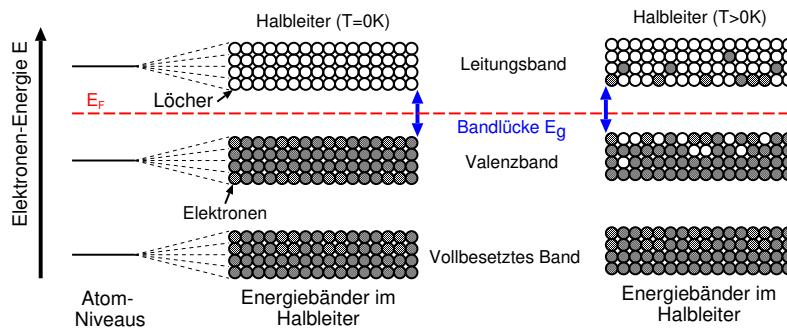
5.1 – Thermisches Gleichgewicht

Die möglichen Elektronenzustände eines einzelnen Potentialtopfes (also eines einzelnen Atoms) sind diskrete Werte. Durch die periodische Anordnung mehrere Atome ergibt sich eine kontinuierliche Anzahl an möglichen Elektronenzuständen, wobei Energie und Impuls der möglichen Zustände nicht beliebig wählbar sind. Vielmehr definiert die **Dispersionsrelation** Bereiche erlaubter und 'verbotener' Energien für die Wellenfunktionen der Elektronen. Für viele Betrachtungen ist es ausreichend, lediglich die Bänder möglicher Energien zu betrachten (d.h. nur die y-Achse der Dispersionsrelation). Die diskreten Energiezustände einzelner Atome weiten sich folglich zu Energiebändern, in denen Elektronenzustände mit (gemäß Dispersionsrelation) entsprechenden Wellenvektoren existieren. Diese Energiebänder können sich entweder überlappen oder aber durch Bandlücken getrennt sein.



Die **Fermi-Dirac Statistik** der Elektronen erlaubt nur eine einfache Besetzung jedes Elektronenzustandes (**Pauli-Prinzip**). Am absoluten Nullpunkt $T = 0K$ sind daher alle Elektronenzustände unter einer gewissen Energie, der **Fermi-Energie** E_F , besetzt, alle möglichen darüberliegenden Zustände bleiben frei, siehe Abbildung. Bei einem metallischen Leiter liegt die Fermienergie innerhalb eines Energiebandes. Die thermische Energie bei $T > 0$ genügt bereits um Elektronen auf die noch freien Zustände zu bringen und es gibt bereits bei Raumtemperatur viele frei bewegliche Elektronen.

Bei einem Halbleiter (Hauptgruppe IV, bzw. gleiche Mengen von Atomen aus den Hauptgruppen III/V bzw. II/VI) ist jedes Atom kovalent an die 4 nächsten Nachbarn gebunden, d.h. jedes Atom teilt mit jedem der Nachbarn im Schnitt ein Valenzelektron. Als Folge wird die äußerste Schale von 8 (jeweils zwischen 2 Nachbarn geteilten) Elektronen besetzt und bildet daher eine Art Edelgaskonfiguration. Am absoluten Nullpunkt $T = 0K$ füllen die Valenzelektronen ein Energieband vollständig, das sogenannte **Valenzband**, während freie Elektronenzustände erst im nächsten Energieband, dem sogenannten **Leitungsband** möglich sind. Die **Fermi-Energie** der Elektronen liegt bei $T = 0K$ daher genau in der Mitte der Bandlücke.



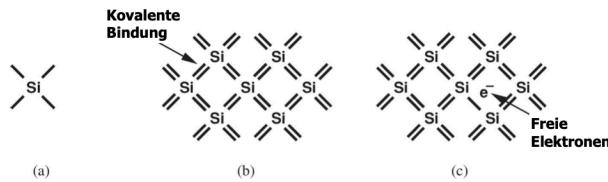
Bei $T > 0$ ist die Besetzungszahl der Elektronen gemäß **Fermi-Funktion** $f(E)$ verteilt und es sind auch (mit dem Boltzmann-Faktor gewichtet) Elektronen mit ausreichend Energie für das Leitungsband vorhanden, siehe Abbildung.

Im Folgenden wird die Dichte der freien Ladungsträger des reinen Halbleiters, d.h. die **Elektrondichte** n und die **Löcherdichte** p , berechnet.

Aufgabe 36: Die Elektronen im Bereich der Fermi Energie, $E_F \approx 1 - 10\text{ eV}$, verhalten sich wie freie Elektronen mit m_{eff} . Welche Elektronenenergie haben im Vergleich die Leitungselektronen im Drude-Modell bei Raumtemperatur? ($1\text{ eV} = 1.6 \times 10^{-19}\text{ J}$)

Intrinsischer Halbleiter: Leitfähigkeit

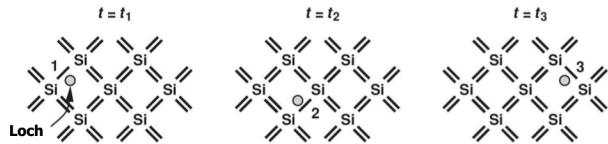
Die Valenzelektronen der Halbleiter bilden kovalente Bindungen zu den 4 nächsten Nachbarn. Am Beispiel von Silizium, siehe Abbildung (a), bilden die 4 Valenzelektronen jedes Si-Atoms mit einem der Valenzelektronen der 4 nächsten Nachbarn eine kovalente Bindung, dargestellt durch 2 Linien in Abbildung (b).



Quelle: B. Razavi, Fundamentals of Microelectronics

Bei endlichen Temperaturen können sich Valenzelektronen von ihren kovalenten Bindungen lösen und frei im Gitter bewegen. Die Eigen-Leitfähigkeit, oder auch **intrinsische Leitfähigkeit**, steigt entsprechend an.

Jedes frei-werdende Elektron im reinen Halbleiter hinterlässt einen nicht besetzten Zustand im Valenzband, ein sogenanntes **Loch**, welches wieder durch ein freies Elektronen besetzt werden kann. Durch das Besetzen des Loch-Zustandes durch ein anderes Elektron, entsteht an der ursprünglichen Stelle dieses anderen Elektrons ein neues Loch, siehe Abbildung.



Quelle: B. Razavi, Fundamentals of Microelectronics

Am absoluten Nullpunkt sind nur die niedrigen Zustände besetzt und alle Valenzelektronen in kovalenten Bindungen zu den Nachbarn, es gibt daher keine frei verfügbaren Ladungsträger und die elektrische Leitfähigkeit ist verschwindend klein.

Das Loch 'wandert' folglich durch den Halbleiter, entsprechend eines positiven Elektrons. Die frei-werdenen Löcher tragen als positive Ladungsträger ebenfalls zur **intrinsischen Leitfähigkeit** des Halbleiters bei.

In einem reinen Halbleiter bei endlichen Temperaturen $T > 0K$ kann die **thermische Energie** eines Elektronen ausreichen, um die Energiedifferenz des Bandabstandes E_g zu überwinden und daher zur Leitfähigkeit der freien Ladungsträger beizutragen (also zu freien Elektronen im Leitungsband bzw. freien Löchern im Valenzband). Die Wahrscheinlichkeit für Elektronen ausreichender Energie (d.h. mit Energie ausreichend für das Leitungsband) ist gegeben durch den Boltzmann-Faktor zur Überwindung des Bandabstandes. Die **intrinsische Leitfähigkeit** ist direkt proportional zur Anzahl der freien Ladungsträger und steigt daher ebenfalls mit dem Boltzmann-Faktor,

$$\sigma \propto e^{-\frac{E_g}{kT}}. \quad (169)$$

Man beachte die starke Temperaturabhängigkeit der **Eigenleitfähigkeit** mit positivem Temperaturkoeffizienten, eine typische Eigenschaft von Halbleitern.

Die **Dichte der Leitungselektronen** oder kurz **Elektronendichte** $n(T)$ im Leitungsband ist ebenfalls durch den Boltzmannfaktor zur Überwindung der Bandlücke gegeben,

$$n(T) \propto e^{-\frac{E_c - E_F}{kT}}. \quad (170)$$

Die **Dichte der freien Löcher** oder kurz **Löcherdichte** $p(T)$ im Valenzband folgt analog aus den Boltzmannfaktor zur Überwindung der Bandlücke,

$$p(T) \propto e^{-\frac{E_F - E_v}{kT}}. \quad (171)$$

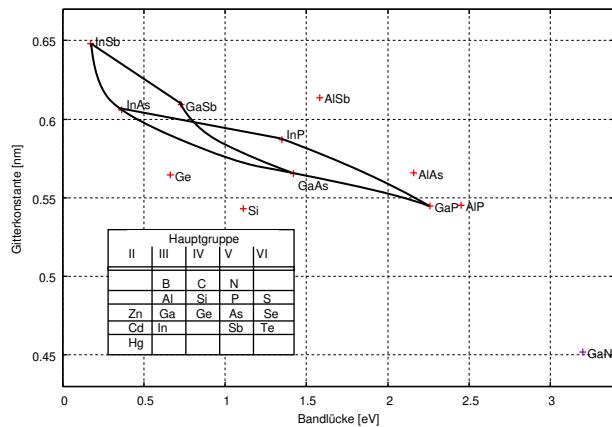
In einem **intrinsischen** (d.h. reinen) **Halbleiter** entstehen freie Elektronen und Löcher immer gemeinsam und bilden eine temperaturabhängige Größe, genannt **intrinsische (Ladungs-)Konzentration** $n_i(T)$:

$$n(T) \equiv p(T) =: n_i(T) \quad (172)$$

Aufgabe 37: Der Bandabstand von Quarzglas (SiO_2) beträgt $E_g = 9eV$, warum wird es in Siliziumtechnologien ($E_g = 1.12eV$) als Isolator verwendet? Wie verhält sich die Leitfähigkeit der beiden Materialien bei Raumtemperatur? ($1eV = 1.6 \times 10^{-19} J$)

Intrinsischer Halbleiter: Ladungskonzentration

Die meisten bekannten Halbleiter bilden ein Diamant, bzw. Zinkblende-Gitter, welches zur Familie der kubischen Kristallgitter gehört, siehe Abbildung. Eine Ausnahme bildet Siliziumkarbid *SiC*, welches in vielen unterschiedlichen Gittern kristallisiert kann!

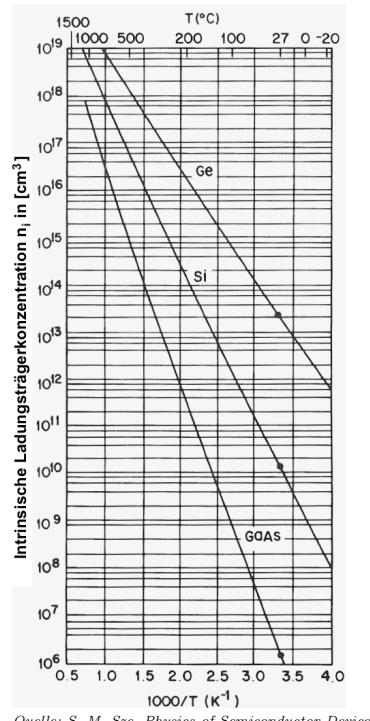


Typische Halbleiter sind hier in Bezug auf Gitterkonstante und Bandlückenenergie E_g dargestellt. Bei den zusammengesetzten Halbleitern können durch Mischung der Atome der III-ten Hauptgruppe bzw. der V-ten Hauptgruppe alle Bandlücken auf den durchgezogenen Linien erzeugt werden, z.B. alle Mischverhältnisse im *InGaAsP*-Komplex. Damit werden z.B. Leuchtdioden unterschiedlicher Farbe hergestellt. Die unterschiedlichen Gitterkonstanten der Mischverhältnisse führen allerdings zu mechanischen Verspannungen und Versetzungen im Gitter.

Die **intrinsische Konzentration** der Ladungsträger bezeichnet die bei endlicher Temperatur immer vorhandene Dichte an freien Ladungsträgern. Aus den Überlegungen zur intrinsischen Leitfähigkeit folgt

$$n_i^2(T) = p(T)n(T) \propto e^{-\frac{E_g}{kT}}. \quad (173)$$

Die intrinsische Konzentration ist also unabhängig von der konkreten Energie des Valenzbandes bzw. des Leitungsbandes und hängt nur vom Bandabstand E_g ab, mit einer sehr starken Temperaturabhängigkeit.



Die temperaturabhängige Eigenleitungskonzentration als Funktion der inversen Temperatur wird oft in einem halblogarithmischen Graphik, einem sogenannten **Arrhenius-Plot** dargestellt. In der Abbildung links sind die Kurven für *Si*, *Ge* und *GaAs* gezeichnet.

Die intrinsische Konzentration wichtiger Materialien bei Raumtemperatur 300K

Material	$E_g [eV]$	$n_i [cm^{-3}]$
<i>Si</i>	1.12	1.5×10^{10}
<i>GaAs</i>	1.42	1.8×10^6
<i>Ge</i>	0.66	2.5×10^{13}
<i>SiC</i>	3.26	$\approx 10^{-8}$
<i>SiO₂</i>	9.0	$\approx 10^{-66}$

N.B.: Man beachte die geringe Eigenleitung von Siliziumkarbid.

Die intrinsische Leitfähigkeit von Halbleitern beschränkt den Temperaturbereich in dem halbleiterbasierte Schaltungen verwendet werden können.

Aufgabe 38: Der Bandabstand eines Halbleiters ist schwach temperaturabhängig. Im Falle von *Si* gilt $E_g(T) = 1.17eV - \frac{\alpha T^2}{T+\beta}$ mit $\alpha = 4.72 \times 10^{-4}eV/K$ und $\beta = 636K$. Wie stark variiert der Bandabstand von *Si* im typischen Temperaturbereich für automotive Anwendungen $T \in [-40^\circ C, 175^\circ C]$?

Berechnung der intrinsische Ladungskonzentration

Die Dispersionsrelation der Elektronen im Leitungsband entspricht freien Elektronen mit effektiver Masse m_e^{eff} . Die Energie der freiwerdenden Löcher ist ebenfalls durch die Energiefunktion freier positiver Teilchen (Löcher) mit Masse m_h^{eff} gegeben. Die Minimalenergie der freien Elektronen ist gerade die Bandkante des Leitungsbandes E_c und entspricht hier einer konstanten potentiellen Energie. Im Falle der Löcher ist die Bandkante des Valenzbandes E_v die konstante potentielle Energie:

$$E_e = E_c + \frac{p^2}{2m_e^{\text{eff}}} \quad (174)$$

$$E_h = E_v - \frac{p^2}{2m_h^{\text{eff}}} \quad (175)$$

Die Zustandsdichte freier Teilchen wurde bereits in Gleichung (116) berechnet. Die Zustandsdichten der freien Elektronen und Löcher im Halbleiter sind lediglich um die potentielle Energie verschoben und lauten

$$D_e(E) = D(E - E_c) \quad (176)$$

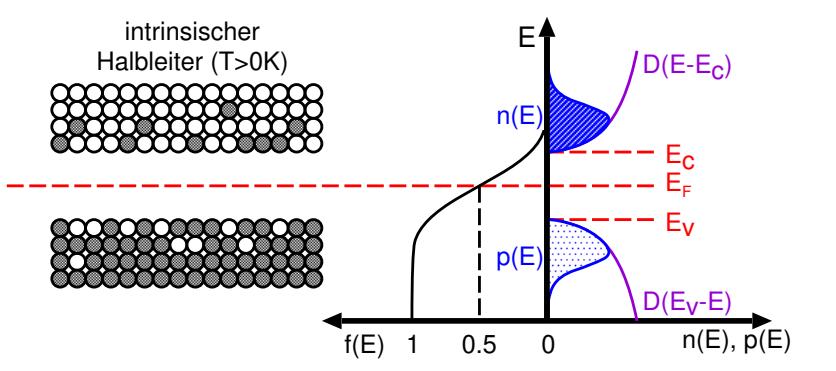
$$D_h(E) = D(E_v - E). \quad (177)$$

N.B.: Die Dichte der Elektronenzustände unter der Bandkante des Leitungsbandes und die Löcherdichte oberhalb der Bandkante des Valenzbandes ist gleich null.

Zur Berechnung der freien Elektronendichte n und Löcherdichte p wird nun die Fermi-Funktion $f(E)$ (d.h. die temperaturabhängige Besetzungszahl der verschiedenen Energien), über die Zustandsdichte $D(E)$ (d.h. die Anzahl der möglichen Zustände pro Energie) integriert, siehe Abbildung

$$n = \int_0^\infty D_e(E)f(E)dE \quad (178)$$

$$p = \int_0^\infty D_h(E)(1 - f(E))dE. \quad (179)$$



Die **Elektrondichte** n (blau schraffiert) und die **Löcherdichte** p (blau gepunktet) für $|E - E_F| \gg kT$ lassen sich leicht berechnen

$$n = \rho_e e^{-\frac{E_c - E_F}{kT}} \quad (180)$$

$$p = \rho_h e^{-\frac{E_F - E_v}{kT}}, \quad (181)$$

wobei $\rho_{e,h} := 2 \left(\frac{m_{e,h}^{\text{eff}} k T}{\hbar^2 2\pi} \right)^{3/2}$.

Die **intrinsische Konzentration** n_i ist, wie bereits vorhin vermutet, **unabhängig von der Fermi-Energie**, siehe auch Gleichung (173)

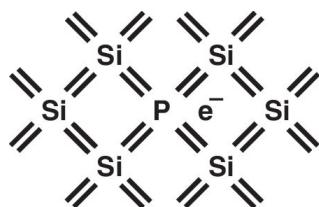
$$n_i = \sqrt{\rho_e \rho_h} e^{-\frac{E_g}{2kT}}. \quad (182)$$

N.B.: Bei Halbleitern mit gleicher effektiver Masse für Elektronen und Löcher liegt das Fermi-Niveau genau in der Mitte der Bandlücke.

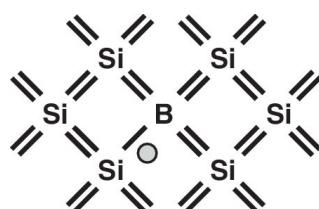
Aufgabe 39: Leite die Gleichungen (180) für einen Bandabstand deutlich über der thermischen Energie, d.h. $|E - E_F| \gg kT$, her.

Extrinsische Halbleiter: Dotierung

Werden einem reinen (d.h. intrinsischen) Halbleiter Fremdatome beigemischt, kann dadurch die Ladungsträgerdichte beeinflusst werden. Einen Halbleiter mit Fremdatomen nennt man **extrinsischen Halbleiter**, das Beifügen von Fremdatomen wird **Dotieren** genannt. Im Falle von *Si* gibt es die Möglichkeit Atome aus der V-ten Hauptgruppe, z.B. Phosphor *P* oder Arsen *As*, oder Atome aus der III-ten Hauptgruppe, z.B. Bor *B* beizumengen, siehe Abbildung.



Quelle: B. Razavi, Fundamentals of Microelectronics

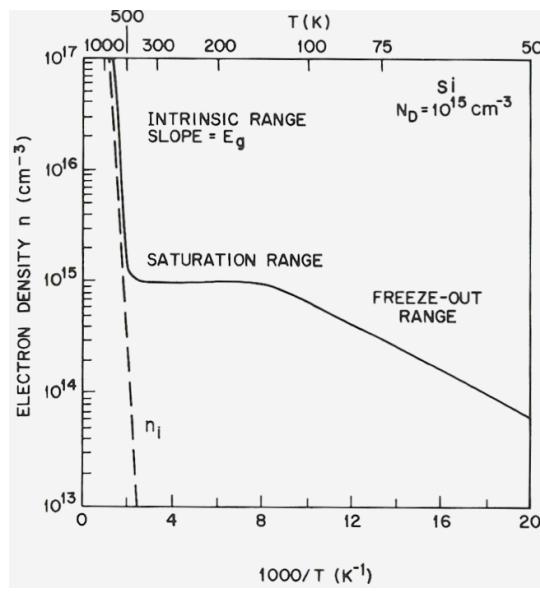


Quelle: B. Razavi, Fundamentals of Microelectronics

Wird ein gewisser Anteil der *Si*-Atome z.B. durch Phosphor ersetzt, bilden 4 der 5 Valenzelektronen des Phosphors die kovalenten Bindungen zu den nächsten *Si*-Atomen. Das verbleibende Valenzelektron ist nur schwach (Bindungsenergie $\Delta E \approx 0.01\text{eV}$) an das *P*-Atom gebunden und ist bereits für thermische Energien bei Raumtemperatur, $kT \approx 25\text{mV}$, als freies Elektronen verfügbar. Phosphor wirkt demzufolge als 'Elektronenspender', **Donor** genannt, und erhöht die Elektrondichte im resultierenden **extrinsischen Halbleiter**, genannt **n-dotiertes Silizium** oder n-Typ *Si*.

Wird ein gewisser Anteil der *Si*-Atome z.B. durch Bor ersetzt, bilden 3 Valenzelektronen des Bor die kovalenten Bindungen zu den nächsten *Si*-Atomen, eine mögliche kovalente Bindung kann nicht besetzt werden und bildet folglich ein Loch mit großer Anziehungs Kraft für Elektronen. Das Loch ist folglich nur schwach (Bindungsenergie $\Delta E \approx 0.01\text{eV}$) an das *B*-Atom gebunden und bereits für thermische Energie bei Raumtemperatur, $kT \approx 25\text{mV}$, als freies Loch verfügbar. Bor wirkt demzufolge als 'Elektronensammler', **Akzeptor** genannt und erhöht die Lochdichte im resultierenden **extrinsischen Halbleiter**, genannt **p-dotiertes Silizium** oder p-Typ *Si*.

Die Dichte der Donoren in einem n-Typ Halbleiter wird N_D , die Dichte der Akzeptoren in einem p-Typ Halbleiter wird mit N_A bezeichnet. Im Falle einer n-dotierten Siliziums mit $N_D = 10^{15}\text{cm}^{-3}$ sieht der Arrhenius-Plot nun wie folgt aus



Quelle: S. M. Sze, Physics of Semiconductor Devices

Bei sehr kalten Temperaturen (rechts in der Abbildung) sind die zusätzlichen Valenzelektronen des Phosphors z.T. noch an das *P*-Atom gebunden (**Freeze-Out**).

Mit steigender Temperatur werden alle zusätzlichen Elektronen abgelöst und die Elektronendichte ist nur durch die Donordichte bestimmt, $n \approx N_D$. Man spricht hier von **Sättigung bzw. Saturation**.

Die Ladungsträger der Sättigung werden Mehrheitsladungsträger bzw. **Majoritätsladungsträger** genannt (Elektronen bei n-Typ und Löcher bei p-Typ).

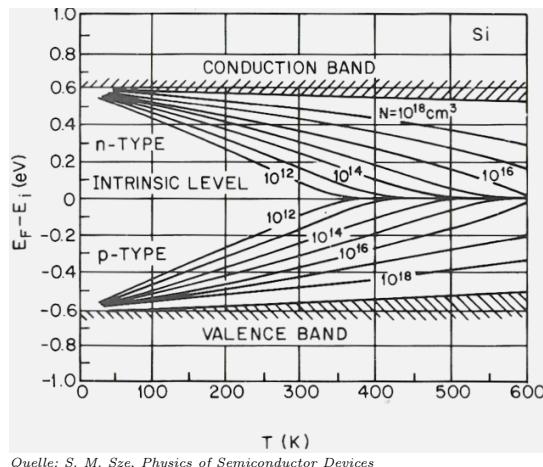
Mit steigender Temperatur (links in der Abbildung) steigt auch die intrinsische Elektronendichte und bestimmt ab einer gewissen Temperatur die Elektronendichte.

Dotierte Halbleiter werden typischerweise im Sättigungsbereich betrieben, da hier die Elektronendichte konstant und definiert ist, bei hohen Temperaturen führt der starke Anstieg der Ladungsträger zu einer thermischen Zerstörung des Halbleiters.

Aufgabe 40: Wie ändert sich der Arrhenius Plot für einen extrinsischen Halbleiter mit großem Bandabstand, wie z.B. *SiC*? Was bedeutet das für die möglichen Betriebstemperaturen?

Berechnung der extrinsische Ladungskonzentration

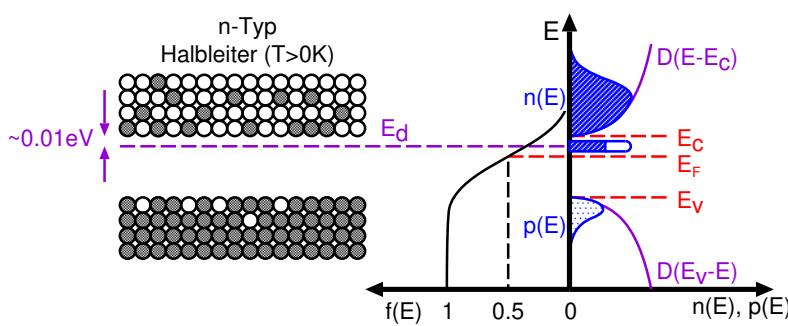
Die Fermi-Energie bei sehr hohen Temperaturen wird von den intrinsischen Eigenschaften des Halbleiters definiert und darf nicht von der Dotierung abhängen, bei tieferen Temperaturen wird hingegen die Fermi-Energie, also die Energiegrenze zwischen freien und besetzten Zuständen vom Energieniveau der Dotierungsatome bestimmt, siehe Abbildung.



Quelle: S. M. Sze, Physics of Semiconductor Devices

Die **Donoren** im n-Typ Halbleiter erhöhen die Elektronendichte im Leitungsband n relativ zur Löcherdichte p . Die **Akzeptoren** im p-Typ Halbleiter erhöhen die Löcherdichte im Valenzband p relativ zur Elektronendichte n . Die Berechnung der Ladungskonzentrationen im extrinsischen Halbleiter erfolgt analog zum intrinsischen Halbleiter, Gleichung (178), wobei die Fermi-Energie je nach Dotierung in Richtung Leitungsband oder Valenzband verschoben wird, siehe Abbildung.

Im Falle eines n-dotierten Halbleiters bei tiefen Temperaturen wird die Fermi-Energie von den schwach gebundenen Phosphor-Atomen bestimmt und liegt knapp unter der Energie des Leitungsbandes E_c . Im Falle des p-dotierten Halbleiters bei tiefen Temperaturen bestimmen die schwach gebundenen Löcher der Bor-Atome das Fermi-Niveau, welches folglich knapp über der Energie des Valenzbandes E_v liegt. Wie bereits im Arrhenius-Plot ersichtlich, wird die Dichte der Majoritätsladungsträger bei hohen Temperaturen durch die intrinsische Dichte abgelöst, und die Fermienergie liegt in der Mitte der Bandlücke, siehe Abbildung.



Die Zustandsdichte mit Dotierung erhält zusätzliche Zustände in der Bandlücke und die Ladungsträgerdichten eines extrinsischen Halbleiters, siehe Abbildung (n-Typ), lauten wie folgt:

$$n \gg p, \text{n-Typ} \quad (183)$$

$$p \gg n, \text{p-Typ.} \quad (184)$$

N.B.: Die Abbildung für p-Silizium ist analog mit E_F nahe bei E_v .

Im Sättigungsbereich ist, wie bereits vorhin gezeigt, die Ladungsträgerdichte durch die Dichte der Fremdatome bestimmt und wir erhalten

$$n \approx N_D, \text{n-Typ} \quad (185)$$

$$p \approx N_A, \text{p-Typ.} \quad (186)$$

Durch Dotierung kann folglich der Typ der dominierenden Ladungsträger und die Dichte bestimmt werden, d.h. die **Dichte der Majoritätsladungsträger**.

Aufgabe 41: Warum hängt die Temperaturabhängigkeit der Fermi-Energie von der Dotierdichte ab (siehe Abbildung oben)?

Extrinsische Halbleiter: Das Massenwirkungsgesetz

Bei endlichen Temperaturen werden in einen Halbleiter pro Zeit und Volumen, in Abhängigkeit von der verfügbaren thermischen Energie (also Temperatur), der Bindungsenergie der Valenzelektronen und der Anzahl verfügbarer Atome ständig neue Elektron-Loch-Paare erzeugt. Bei einem gegebenen Halbleiter sind Bindungsenergie und Anzahl der verfügbaren Atome konstant. Die Erzeugungsrate pro Volumen, genannt **Generationsrate** G hängt daher nur von der Temperatur ab. Dem gegenüber steht der Prozess der Rekombination, bei dem freie Elektronen und Löcher wieder einen Bindungszustand eingehen. Die **Rekombinationsrate** R , also die pro Zeit und Volumen rekombinierenden Elektron-Loch Paare, hängt ebenfalls von der (nun freiwerdenden) Bindungsenergie, der Temperatur aber auch von der vorliegenden Elektronendichte und Löcherdichte. Es gilt daher

$$R(T, n, p) \approx r(T)np \quad (187)$$

wobei die materialabhängige Funktion $r(T)$ nur noch eine Funktion der Temperatur ist.

Material	$r[cm^3/s]$
Si	1.79×10^{-15}
Ge	5.25×10^{-14}
GaP	5.37×10^{-14}
GaAs	7.21×10^{-10}
InP	1.26×10^{-9}

Im thermischen Gleichgewicht muss die Anzahl der erzeugten Elektron-Loch-Paare gerade gleich der Anzahl der rekombinierenden Elektron-Loch Paare sein, also

$$G(T) = R(T, n_0, p_0) = r(T)n_0p_0, \quad (188)$$

n_0 und p_0 bezeichnen hier die Ladungsträgerdichten im thermischen Gleichgewicht.

Die beiden Unbekannten $G(T)$ und $r(T)$ sind unabhängig von den Ladungsträgerdichten. Diese Gleichung drückt das Prinzip des **detaillierten Gleichgewichtes** zwischen Generation und Rekombination aus und ist eine direkte Folge des zweiten Hauptsatzes der Thermodynamik.

Im Falle des intrinsischen Halbleiters im thermischen Gleichgewicht gilt $n_i^2 = n_0p_0$, siehe Gleichung (173) und wir können die materialabhängige Größen durch die intrinsische Ladungsdichte ausdrücken, $G(T)/r(t) = n_i^2$. Daher ist das Produkt der Elektronendichte und der Löcherdichte auch im extrinsischen Halbleiter gleich der intrinsischen Ladungsdichte und wir erhalten das sogenannte **Massenwirkungsgesetz**:

$$n_0p_0 = n_i^2 \quad (189)$$

Im thermischen Gleichgewicht ist die Dichte der **Majoritätsladungsträger** eines extrinsischen Halbleiters in Sättigung durch die Dotierung geben. Mithilfe des Massenwirkungsgesetzes (189) ist dadurch aber auch die Dichte der komplementären Ladungsträger, der sogenannten **Minoritätsladungsträger** gegeben:

- **n-Typ Halbleiter** mit Donorendichte N_D in Sättigung:

$$n \approx N_D \text{ Majoritätsladungsträger} \quad (190)$$

$$p \approx \frac{n_i^2}{N_D} \text{ Minoritätsladungsträger} \quad (191)$$

- **p-Typ Halbleiter** mit Akzeptorendichte N_A in Sättigung:

$$p \approx N_A \text{ Majoritätsladungsträger} \quad (192)$$

$$n \approx \frac{n_i^2}{N_A} \text{ Minoritätsladungsträger} \quad (193)$$

Typische schwache Dotierungen, genannt n^-, p^- liegen im Bereich $10^{12} \dots 10^{14} cm^{-3}$ (Leistungsbauten), typische starke Dotierungen, genannt n^+, p^+ liegen im Bereich $10^{18} \dots 10^{21} cm^{-3}$ (elektrische Kontakte).

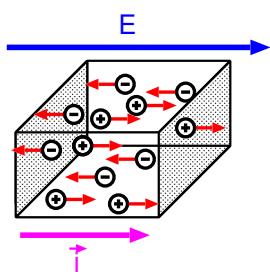
Aufgabe 42: Gegeben sei schwach n-dotiertes Si mit $N_D = 10^{14} cm^{-3}$. Wie groß ist die Majoritäts- und die Minoritätsladungsdichte bei $T = 300K$ und bei $T = 425K$. Ab welcher Temperatur wird die Minoritätsdichte von der intrinsischen Ladungsdichte dominiert?

5.2– Transportphänomene

Bisher wurden nur statische Systeme im thermischen Gleichgewicht betrachtet. Die Ladungsdichten waren deshalb im Mittel konstant und es erfolgte netto kein Ladungstransport. Die Störung dieses Gleichgewichtes durch externe Einflüsse wie externe Spannungen oder injizierte Ladungen führen zu bewegten Ladungen, d.h. elektrischen **Strömen** \vec{j} , welche entsprechend der **Koninuitätsgleichung** (36) zu **lokal geänderten Ladungsdichten** ρ führen. Geänderte Ladungsgradienten bewirken jedoch **Diffusionsströme**, geänderte Ladungen führen über das Gauss'sche Gesetz (15) zu elektrischen Feldern und in weiterer Folge zu **Driftströmen**, welche wiederum über die Kontinuitätsgleichung (36) ein neues Gleichgewicht anstreben. Im Folgenden werden die Transportgleichungen zur Beschreibung dieser Nicht-Gleichgewichtsprozesse, und daher zur Beschreibung des gestörten Halbleiters, behandelt.

Driftstrom

Ein elektrisches Feld \vec{E} im Halbleiter bewirkt eine Bewegung der freien Ladungsträger und daher eine Stromdichte, genannt **Driftstrom**. In einem Halbleiter gibt es prinzipiell freie Elektronen und freie Löcher, und daher auch einen Driftstrom für jeden der beiden Ladungsträger



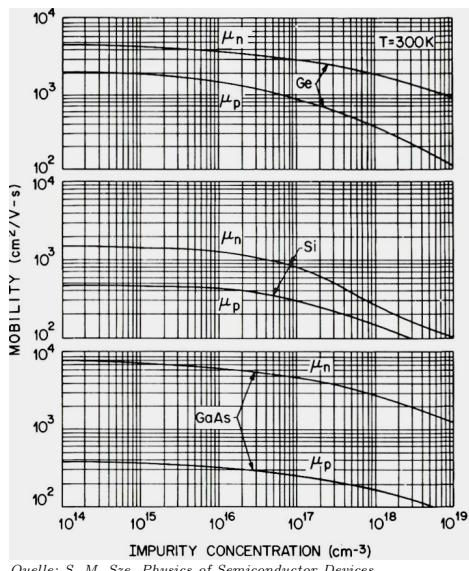
Die Ladungsdichte ρ_n der Elektronen ist gegeben durch die Elektronendichte n und die Elementarladung $-e$. Zusammen mit der Elektronenmobilität μ_n , Gleichung (40), folgt der **Driftstrom der negativen Ladungsträger**:

$$\vec{j}_n^{\text{drift}} := en\mu_n \vec{E}. \quad (194)$$

Die Ladungsdichte ρ_p der Löcher ist gegeben durch die Löcherdichte p und die Elementarladung e . Zusammen mit der Löchermobilität μ_p , Gleichung (40), folgt der **Driftstrom der positiven Ladungsträger**:

$$\vec{j}_p^{\text{drift}} := ep\mu_p \vec{E}. \quad (195)$$

N.B.: Beide Ladungsträger ergeben einen Strom in Richtung des Feldes.



Quelle: S. M. Sze, Physics of Semiconductor Devices

Die **Driftgeschwindigkeit der Elektronen** ist proportional zum negativen elektrischen Feld \vec{E} und definiert die **Elektronen-Mobilität** μ_n :

$$\vec{v}_e =: -\mu_n \vec{E} \quad (196)$$

Die **Driftgeschwindigkeit der Löcher** ist proportional zum positiven elektrischen Feld \vec{E} und definiert die **Löchermobilität** μ_p :

$$\vec{v}_h =: \mu_p \vec{E} \quad (197)$$

Im Fall von Si, siehe Abbildung, gilt $\mu_n/\mu_p \approx 3$,

$$\mu_n = 1350 \frac{\text{cm}^2}{\text{Vs}} \quad (198)$$

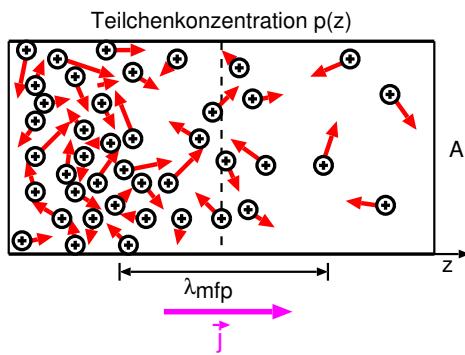
$$\mu_p = 480 \frac{\text{cm}^2}{\text{Vs}}. \quad (199)$$

Der Faktor 3 in der Mobilität der Elektronen relativ zu den Löchern in Si ist der Grund, warum in integrierten CMOS Schaltungen die Transistoren mit Löcherleitung (PMOS-Transistor) typischerweise um einen Faktor 3 größer dimensioniert sind, wie die Transistoren mit Elektronleitung (NMOS-Transistor). Außerdem werden bei integrierten niederohmigen Leistungsschaltern zumeist NMOS-Transistoren verwendet.

Aufgabe 43: Die Elektronenladung ist negativ, trotzdem tragen auch die Elektronen zu einem elektrischen Strom in Richtung des elektrischen Feldes bei. Wie kann das sein?

Diffusionsstrom

Die zufällige thermische Bewegung von freien Teilchen führt bei ortabhängigen Konzentrationsunterschieden zu einem Netto-Teilchenstrom von hoher Konzentration zu niederer Konzentration, genannt **Diffusion**, siehe Physik I. Der resultierende Teilchenstrom ist proportional zum Konzentrationsgradienten mit einer physikalischen Proportionalitäts-Konstanten, der Diffusionskonstanten D . Im Falle eines Halbleiters kann der **Diffusionsstrom** sowohl aus **freien Elektronen** als auch aus **freien Löchern** bestehen.



Bei einem Konzentrationsgefälle der Löcherdichte $p(\vec{x})$ entsteht ein Teilchenstrom in Richtung des negativen Gradienten, d.h. $-\vec{\nabla}p(\vec{x})$, und folglich, multipliziert mit der Löcherladung e ein **elektrischer Diffusionsstrom der Löcher**

$$\mathbf{j}_p^{\text{diff}} := -eD_p \nabla p. \quad (200)$$

Im Falle eines Konzentrationsgefälles der Elektrondichte $n(\vec{x})$, entsteht ein Elektronenstrom in Richtung des Konzentrationsgefälles, $-\vec{\nabla}n(\vec{x})$, und multipliziert mit der Elektronenladung $-e$ ein **elektrischer Diffusionsstrom der Elektronen**

$$\mathbf{j}_n^{\text{diff}} := eD_n \nabla n, \quad (201)$$

D_n bzw. D_p sind die jeweiligen **Diffusionskonstanten**.

Die potentielle Energie eines freien Teilchens im Halbleiter ist gegeben durch das elektrische Potential multipliziert mit der elektrischen Ladung des Teilchens, $E_{\text{pot}} := q\Phi$. Das elektrische Feld ist gerade der negative Gradient des Potentials, Gleichung (16). Eine räumlich veränderliche potentielle Energie erzeugt folglich ein lokales elektrisches Feld,

$$\vec{E} = -\frac{1}{q} \vec{\nabla} E_{\text{pot}}. \quad (202)$$

Die potentielle Energie eines freien Elektrons im Halbleiter ist gegeben durch die Energie des Leitungsbandes, E_c . Die potentielle Energie eines Loches im Valenzband des Halbleiters ist entsprechend die Energie des Valenzbandes E_v . Bei einer lokal variierenden Energie des Leitungsbandes $E_c(\vec{x})$ bzw. des Valenzbandes $E_v(\vec{x})$ entstehen folglich elektrische Felder und daher **Driftströme**, welche zu einer Ladungsverschiebung und dadurch zu einem **Konzentrationsgefälle** der freien Ladungsträger führen. Durch das Konzentrationsgefälle bildet sich ein **Diffusionsstrom** aus, der dem Driftstrom entgegenwirkt und schliesslich kompensiert.

Im Gleichgewicht ergibt sich nun für den Löcherstrom und den Elektronenstrom an einer lokalen Stufe der Bandkante folgende Relation:

$$\vec{j}_p = ep\mu_p \vec{E}(E_v) - eD_p \vec{\nabla} p(E_v) \stackrel{!}{=} 0 \quad (203)$$

$$\vec{j}_n = en\mu_n \vec{E}(E_c) + eD_n \vec{\nabla} n(E_c) \stackrel{!}{=} 0. \quad (204)$$

Das elektrische Feld der Löcher folgt via Gleichung (202) aus dem Gradienten der Valenzband-Energie E_v , das elektrische Feld der Elektronen aus dem Gradienten der Leitungsband-Energie E_c . Setzt man die Löcherdichte als Funktion der Valenzbandkante $p(E_v)$, Gleichung (181) und die Elektronendichte als Funktion der Leitungsbandkante $n(E_c)$, Gleichung (180), folgt die **Einstein Relation** für die Diffusionskonstanten.

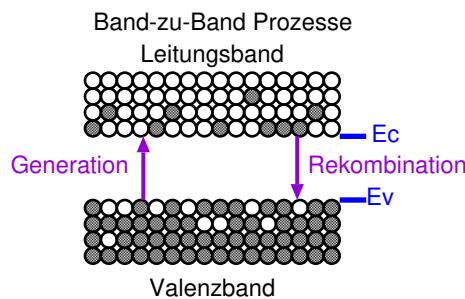
$$D_{n,p} = \mu_{n,p} \frac{kT}{e}. \quad (205)$$

Die Einstein-Relation beschreibt das Gleichgewicht zwischen Driftstrom und Diffusionsstrom an einem lokalen Übergang von einem p-dotierten zu einer n-dotierten Halbleiter, einem sogenannten **PN-Übergang**.

Aufgabe 44: Wie lauten die Zwischenschritte für die Herleitung der Einstein-Relation?

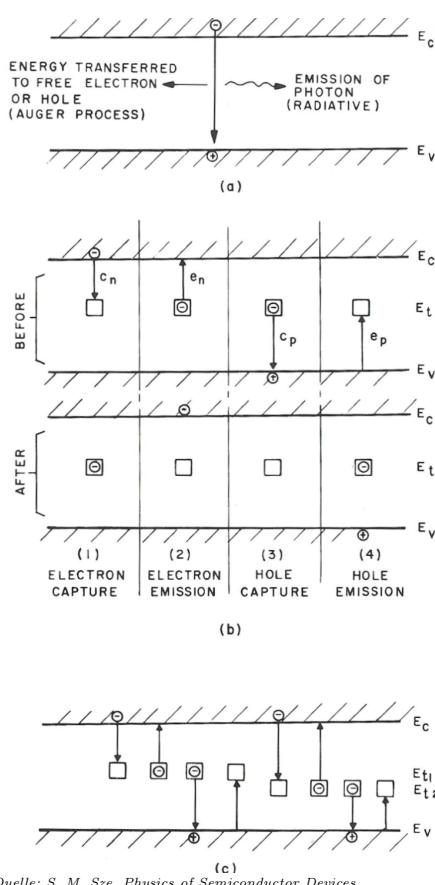
Ladungsgeneration und Rekombination

In der Herleitung des Massenwirkungsgesetzes wurde bereits verwendet, dass in einem Halbleiter bei endlichen Temperaturen ständig freie Elektron-Loch-Paare erzeugt werden (**Generation**) und freie Elektron-Loch-Paare zu gebundenen Zuständen übergehen, d.h. rekombinieren (**Rekombination**).



- Die **Generationsrate** G gibt an, wieviele freie Elektron-Loch-Paare pro Zeit und pro Volumen neu entstehen und daher als Ladungsträger dazukommen.
- Die **Rekombinationsrate** R gibt an, wieviele freie Elektron-Loch-Paare pro Zeit und Volumen in einen gebundenen Zustand übergehen und daher nicht mehr als Ladungsträger verfügbar sind.

Die **Energieerhaltung** verlangt, dass die benötigte Energie für die Erzeugung oder die freigesetzte Energie der Rekombination aus anderen Quellen stammt bzw. über andere Mechanismen abgeführt wird.



Quelle: S. M. Sze, Physics of Semiconductor Devices

Generation: Die Anregung eines Elektrons in das Leitungsband **erzeugt** ein neues Elektron-Loch-Paar und erfolgt:

- thermisch (für $T > 0K$) - z.B. Eigenleitung
- optisch (Absorption eines Photons) - z.B. PV-Zelle
- elektrisch (Ladungsinjektion) - z.B. Diode

Rekombination: Die Bindung eines Leitungselektrons an ein Loch des Valenzbandes **reduziert** die freien Elektron-Loch-Paare und gibt die Energie ab an:

- emittierte Photonen (strahlende Rekombination bei direkten Halbleitern) - z.B. Leuchtdioden
- thermische Bewegung des Kristalls (Nichtstrahlende Rekombination, bei indirekten Halbleitern)
- ein anders freigesetztes Elektron (Auger Elektron)

Neben den bisher besprochenen Band-zu-Band Übergängen, sind auch noch andere Mechanismen möglich. Die Abbildung zeigt die wichtigsten Rekombinationsmechanismen:

- Band-zu-Band Rekombination (a), welche strahlend oder nicht strahlend erfolgen kann (siehe oben)
- Rekombinationen mit einzelnen Elektron-oder Loch-Zuständen (b)
- Multi-Level Rekombinationen (c), bei denen mehrere Zustände zugleich besetzt werden.

Im Folgenden werden die Ladungsträgerinjektion und die resultierenden Prozesse im Halbleiter genauer beschrieben.

Aufgabe 45: Bei welchen Generations-und Rekombinationsprozessen ändert sich die elektrische Ladung des Halbleiters, bei welchen nicht?

Ladungsträgerinjektion

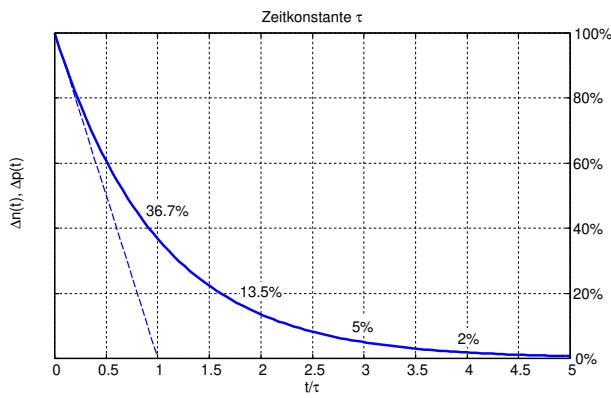
Werden einem Halbleiter im thermischen Gleichgewicht, d.h. $R_0 = G_0 = rn_0p_0$, Gleichung 188, zusätzlich zur Gleichgewichts-Generationsrate G_0 konstant Elektron-Loch-Paare zugeführt, stört diese **zusätzliche Generationsrate** G_i das thermische Gleichgewicht. Als Folge ändern sich die Ladungsträgerdichten zu $n = n_0 + \Delta n$ und $p = p_0 + \Delta p$ bis sich ein neues Gleichgewicht einstellt

$$R = G_0 + G_i = rnp. \quad (206)$$

Da Elektronen und Löcher (laut Annahme) immer als Paaren entstehen gilt für die Zunahme $\Delta n = \Delta p$. Im Bereich kleiner Injektionsraten gilt $\Delta n \ll n_0$ und $\Delta p \ll p_0$ kann der Anstieg der Ladungsträger in erster Näherung Taylor-Entwicklung direkt mit der Injektionsrate G_i verknüpft werden

$$G_i = R - G_0 \approx \frac{\Delta n}{\tau}, \quad (207)$$

Die Zeitkonstante $\tau := 1/[r(n_0 + p_0)]$ bezeichnet die Lebensdauer der injizierten Ladungsträger.

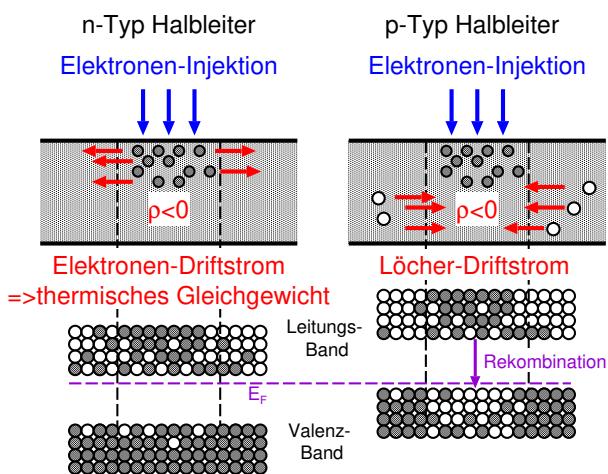


Rekombination ist gerade ein Beispiel einer Relaxationsgleichung, siehe Physik I. Der Überschuss an negativen Δn (oder analog positiven Δp) Ladungsträgern wird durch die Injektionsrate bestimmt und kehrt, sobald keine Ladungen mehr injiziert werden, durch Rekombination auf den Gleichgewichtswert zurück, siehe Abbildung:

$$\frac{d}{dt} \Delta n = G_i - \frac{\Delta n}{\tau} \quad (208)$$

$$\frac{d}{dt} \Delta p = G_i - \frac{\Delta p}{\tau} \quad (209)$$

Bisher wurden nur Elektron-Loch-Paare injiziert. Bei der Injektion nur eines Ladungstyps im extrinsischen Halbleiter gibt es eine wichtige Unterscheidung, je nachdem ob die injizierten Ladungsträger zu den **Majoritätsladungen** oder den **Minoritätsladungen** zählen, siehe Abbildung für Elektronen-Injektion.



- **Injektion von Majoritätsladungsträgern:** Der Ladungsüberschuss erzeugt ein elektrisches Feld und Driftströme stellen das Gleichgewicht sofort wieder her, in typischerweise 10^{-12}s .

- **Injektion von Minoritätsladungsträgern:** Der Ladungsüberschuss erzeugt ein elektrisches Feld, welches die **Majoritätsladungen** anzieht. Dieser **Majoritätsladungs-Überschuss** wird durch **Rekombination** (strahlend oder nichtstrahlend) in der Zeitkonstante τ , geben durch die **Rekombinationsrate**, abgebaut.

Der Fall einer Löcher-Injektion ist völlig analog.

Wir haben nun das Gleichungssystem zur Beschreibung eines Halbleiters vollständig.

Aufgabe 46: Zeige, dass die Relaxationsgleichung bei konstanter Injektionsrate G_i den Überschuss τG_i erzeugt und nach Abschalten der Injektion exponentiell auf den Gleichgewichtswert $\Delta n = 0$ zurückgeht.

6 – Grundgleichungen der Halbleiterphysik

Im vorherigen Kapitel wurden die einzelnen Transportmechanismen im Halbleiter abgeleitet:

- Externe Ladungsinjektion G_i erzeugt (lokalen) Ladungüberschüsse, Δn or Δp
- Die Zeitkonstante für den Abbau von Überschussladungen wird von der **Relaxationsgleichung** der Minoritätsladungsträger bestimmt, τ_n or τ_p
- Geänderte räumliche Ladungsdichten erzeugen **elektrische Felder** gemäß Gauss'schem Gesetz
- Elektrische Felder bewirken **Driftströme**, Dichtegradienten bewirken **Diffusionsströme** für Minoritäts- und Majoritätsladungen
- Die Änderung der Ladungsdichten n und p durch Ströme, Injektion und Rekombination muss die **Koninuitätsgleichung** erfüllen

Stromdichten:

Gesamte Stromdichte

$$\vec{j} = \vec{j}_n + \vec{j}_p \quad (210)$$

Negative Ladungsträger (Elektronen)

$$\vec{j}_n = en\mu_n \vec{E} + eD_n \vec{\nabla}n \quad (211)$$

Positive Ladungsträger (Löcher)

$$\vec{j}_p = ep\mu_p \mathbf{E} - eD_p \vec{\nabla}p \quad (212)$$

Koninuitätsgleichung:

n-Typ Halbleiter mit Injektionsrate G_i

$$-\frac{\partial}{\partial t}p = \frac{p - p_0}{\tau_p} - G_i + \frac{1}{e}\vec{\nabla} \cdot \vec{j}_p \quad (213)$$

$$-\frac{\partial}{\partial t}n = \frac{p - p_0}{\tau_p} - G_i - \frac{1}{e}\vec{\nabla} \cdot \vec{j}_n \quad (214)$$

p-Typ Halbleiter mit Injektionsrate G_i

$$-\frac{\partial}{\partial t}p = \frac{n - n_0}{\tau_n} - G_i + \frac{1}{e}\vec{\nabla} \cdot \vec{j}_p \quad (215)$$

$$-\frac{\partial}{\partial t}n = \frac{n - n_0}{\tau_n} - G_i - \frac{1}{e}\vec{\nabla} \cdot \vec{j}_n \quad (216)$$

Elektrostatische Gleichungen

Gesamte Ladungsdichte (Donoren hinterlassen positive, Akzeptoren negative Atomröhpfle)

$$\rho = e(p - n + N_D - N_A) \quad (217)$$

Gauss'sches Gesetz (für konstantes ε)

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon} \quad (218)$$

6.1 – Physikalische Parameter wichtiger Halbleiter bei Raumtemperatur (300K)

	<i>Ge</i>	<i>Si</i>	<i>GaAs</i>
Gitterkonstante [nm]	0.56	0.54	0.56
Atomdichte [cm^{-3}]	4.4×10^{22}	5.0×10^{22}	4.4×10^{22}
Bandabstand E_g [eV]	0.66	1.12	1.42
Art des Bandabstandes	indirekt	indirekt	direkt
Bindungsenergie für Fremdatome [eV]	≈ 0.01	≈ 0.05	≈ 0.01
Intrinsische Ladungsdichte n_i [cm^{-3}]	2.5×10^{13}	1.5×10^{10}	1.8×10^6
Relative Effektive Elektronen-Masse m_n^{eff}/m	0.56	1.08	0.067
Relative Effektive Löcher-Masse m_p^{eff}/m	0.29	0.55	0.47
Relative dielektrische Konstante ϵ_r	16	12	11
Elektronen-Mobilität μ_n [cm^2/Vs]	3900	1350	8500
Löcher-Mobilität μ_p [cm^2/Vs]	1900	480	450
Diffusionskonstante Elektronen D_n [cm^2/s]	101	35	221
Diffusionskonstante Löcher D_p [cm^2/V]	49	12.5	12

6.2– Beispiele

Abbau eines lokalen Majoritätsladungs-Überschusses:

Ein **lokale Ladungsüberschuss**, $\Delta n(\vec{x})$, von Majoritätsladungsträgern erzeugt ein elektrisches Feld, Gleichung (218),

$$\vec{\nabla} \cdot \vec{E} = -\frac{e}{\varepsilon} \Delta n, \quad (219)$$

und dadurch einen **Driftstrom** \vec{j}_n , Gleichung(211), (Driftstrom dominiert hier über Diffusionsstrom)

$$\vec{j}_n = \sigma \vec{E}. \quad (220)$$

Die **Kontinuitätsgleichung**, (214), wird zur **Relaxationsgleichung** der Überschussladungen,

$$\frac{\partial}{\partial t} n = \frac{1}{e} \vec{\nabla} \cdot \vec{j}_n = \frac{1}{e} \vec{\nabla} \cdot (\sigma \vec{E}) = -\frac{\sigma}{\varepsilon} \Delta n. \quad (221)$$

Die **Dielektrische Relaxationszeit**, $\tau := \varepsilon/\sigma$, bewegt sich zwischen $10^{-14}s$ für Metalle, $10^{-12}s$ für *Si* und erreicht $1s$ für Isolatoren. Der Abbau der Ladungsträger folgt einer Zerfallskurve,

$$\Delta n(t) = \Delta n(0) e^{-\frac{t}{\tau}}. \quad (222)$$

Abbau eines lokalen Minoritätsladungs-Überschusses:

Eine **konstante Injektion** von Minoritätsladungen G_i in einen p-Typ Halbleiter (z.B. durch eine Lichtquelle) erzeugt im statischen Fall eine Überschuss von Minoritätsladungsträgern Δp , Gleichung (213),

$$\Delta p := p - p_0 = \tau_p G_i. \quad (223)$$

Sobald die Injektion abgedreht wird (z.B. das Licht ausgeschaltet wird) **rekombinieren die Minoritätsladungsträger** entsprechend Gleichung (213):

$$\frac{\partial}{\partial t} p = -\frac{1}{\tau_p} (p - p_0). \quad (224)$$

Die **Rekombinationszeit der Minoritätsladungsträger**, τ_p , ist typischerweise sehr kurz für direkte Halbleiter wie *GaAs*, $10^{-9}s$, indirekte Halbleiter wie sehr reines *Si* erreichen $10^{-3}s$.

Der Abbau der Ladungsträger folgt wieder einer Zerfallskurve,

$$\Delta p(t) = \Delta p(0) e^{-\frac{t}{\tau_p}}. \quad (225)$$

Diffusionlänge der Minoritätsladungsträger

Die **Dichte der Minoritätsladungsträger** wird an einer gewissen Stelle künstlich hochgehalten. Es wird z.B. die Elektronendichte $n(x)$ in einem p-Typ Halbleiter mit einer asymptotischen Minoritätsdichte $n(\infty) = n_0$ an der Stelle $x = 0$ auf $n(0) > n_0$ gehalten. **Diffusionsströme**, Gleichung (211), bewirken ein räumliches Ladungsprofil, Gleichung (216),

$$-\frac{\partial}{\partial t} n \equiv 0 = \frac{n(x) - n_0}{\tau_n} - D_n \frac{d^2}{dx^2} n(x). \quad (226)$$

Die statische Lösung ist ein **exponentieller räumlicher Zerfall** gemäß

$$\Delta n(x) = n_0 + (n(0) - n_0) e^{-\frac{x}{L_n}}, \quad (227)$$

mit der sogenannten **Diffusionslänge der Elektronen** L_n und analog für Löcher:

$$L_n = \sqrt{D_n \tau_n} \quad \text{p-Typ} \quad (228)$$

$$L_p = \sqrt{D_p \tau_p} \quad \text{n-Typ} \quad (229)$$

Diffusion führt zu einer Überschuss an Minoritätsladungsträgern bis zu einer typischen Tiefe $L_{n,p} \approx 1\text{cm}$ (*Si*).

7 – Halbleiterbauelemente

Die elektrische Leitfähigkeit eines Halbleiters erfolgt prinzipiell über Elektronenleitung und Löcherleitung. Die vorherrschende Art der Ladungsträger kann über Dotierung bestimmt werden. Während in n-dotierten Halbleitern die Elektronenleitung dominiert (**Majoritätsladungsträger**) und die Löcherleitung untergeordnet ist (**Minoritätsladungsträger**), sind die Rollen von Majoritäts- und Minoritätsladungsträgern im p-dotierten Halbleitern gerade umgekehrt. Durch Aneinanderfügen unterschiedlich dotierter extrinsischer aber auch intrinsischer Halbleiterbereiche werden Halbleiterbauelemente mit typischen Eigenschaften erzeugt, d.h. 'elektrisch gesteuerte Bauteile' welche schnell und möglichst verlustfrei zwischen einem Durchlaßverhalten mit geringem Widerstand und einem Sperrverhalten mit hoher Spannungsfestigkeit 'geschaltet' werden können. Bauelemente, bei denen beide Leitungsmechanismen (also Elektronen und Löcher) das elektrische Verhalten bestimmen, werden **Bipolare Bauelemente** genannt, Halbleiterbauelemente, bei denen nur ein Ladungstyp die Leitfähigkeit dominiert, nennt man **Unipolare Bauelemente**.

Zur Familie der bipolaren Bauelemente gehört der Bipolartransistor (auch bipolar junction transistor oder BJT), die Diode, der Tyristor oder der IGBT (isolated gate bipolar transistor - der in gewissem Sinn eine Mischung aus unipolar und bipolarem Bauelement darstellt). Zur Familie der unipolaren Bauelemente gehören der Feldeffekttransistor (auch metal oxide semiconductor field effect transistor oder MOSFET) und die Schottky-Diode. (Der IGBT ist im Wesentlichen ein BJT, welcher mithilfe eines MOSFETs eingeschaltet wird und verhält sich daher eingangsseitig wie ein MOSFET, ausgangsseitig wie ein BJT). Die Vor- und Nachteile der beiden Arten von Halbleitern kann grob wie folgt zusammengefasst werden.

- **Bipolare Bauelemente: BJT, Diode, Tyristor, IGBT**

- + Geringer Durchlaßwiderstand (on-resistance) R_{on}
- + Hohe Sperrspannung
- Hohe Schaltverluste
- Niedere Schaltgeschwindigkeit

- **Unipolare Bauelemente: Schottky-Diode, MOSFET**

- + Geringe Schaltverluste
- + Hohe Schaltgeschwindigkeit
- Kompromiss zwischen Durchlaßwiderstand und Sperrspannung

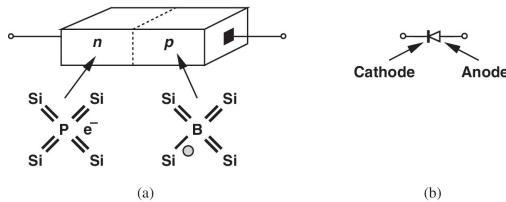
Basierend auf diesen Eigenschaften werden, je nach Anwendung, unterschiedliche Halbleiterbauelemente verwendet

- elektronische Leistungsschalter vorwiegend Bipolar Bauelemente (aber auch immer häufiger Unipolare Bauelemente, z.B. MOSFET mit großem Bandabstand)
- Datenspeicherung durch Halbleiter vorwiegend MOSFET
- Datenverarbeitung durch elektronische Schaltungen vorwiegend MOSFET
- Halbleiter als Sensoren sowohl unipolare als auch bipolare Bauelemente
- Halbleiter als Lichtquellen vorwiegend Dioden
- Stromerzeugung durch Dioden

Im Folgenden werden die drei grundlegenden Bauelemente, **Dioden**, **Bipolar-Transistoren** und **Feldeffekt-Transistoren** im Detail behandelt.

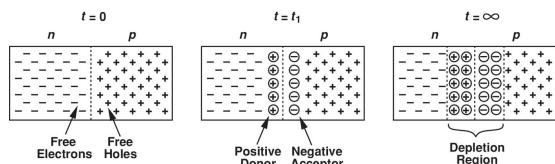
7.1– Die Diode

Der Übergang zwischen einem extrinsischen Halbleiter vom n-Typ zu einem extrinsischen Halbleiter vom p-Typ bildet den sogenannten **PN-Übergang** und somit das einfachste Halbleiter-Bauelement, die **Diode**. Der Name Diode stammt von W.H. Eccles (1919, Electrician) und bedeutet auf griechisch 'zwei-Weg'.

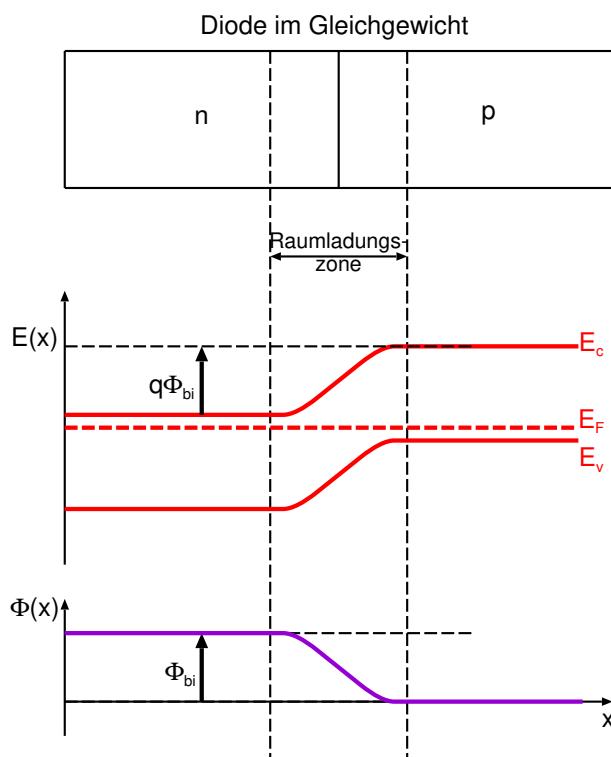


Quelle: B. Razavi, Fundamentals of Microelectronics

Am Übergang der unterschiedlichen Dotierungen kommt es zu Ladungsverschiebungen, siehe Abbildung. Elektronen bilden im n-dotierten Teil (links) die Majoritätsladungsträger, im p-dotierten Teil (rechts) die Minoritätsladungsträger (vice versa bei den Löchern). Beim Aneinanderfügen (siehe Abbildung t=0) führt der Dichtegradient der Elektronen- und Löcherdichten zu Diffusionsströmen der Elektronen und Löcher über den PN-Übergang. Die zurückgelassenen ionisierten Donatoren und Akzeptoren (z.B. positive P-Rümpfe und negative B-Rümpfe) erzeugen eine elektrische Feld, welches den Diffusionsströmen entgegenwirkt.



Quelle: B. Razavi, Fundamentals of Microelectronics



Aufgabe 47: Zeige, dass Elektronen- und Löcherdiffusion trotz unterschiedlichem Vorzeichen auf gleiche Weise zur Bildung der Raumladungszone beitragen.

Der n-dotierte Halbleiter (z.B. ein mit P dotiertes Si , siehe Abbildung) bildet einen Anschluss, genannt **Kathode**, der p-dotierte Halbleiter (z.B. ein mit B dotiertes Si , siehe Abbildung) bildet den anderen Anschluss, genannt **Anode**. Das elektrische Symbol symbolisiert die Durchlassrichtung des elektrischen Stroms.

Im PN-Übergang entsteht ein Bereich mit reduzierten freien Ladungsdichten, die **Verarmungs- oder Raumladungszone** (wie bereits im Kapitel 2.2. ohne Beweis angenommen), mit einem elektrischen Feld und folglich einer elektrischen Spannung, abhängig von den Dotierungen.

- Der n-Typ Halbleiter mit konstantem Dopingprofil $n_n = N_D$ (links) mit **Fermi-Energie** knapp unter dem Leitungsband E_c .
- Der p-Typ Halbleiter mit konstantem Dopingprofil $p_p = N_A$ (rechts) mit **Fermi-Energie** knapp über dem Valenzband E_v

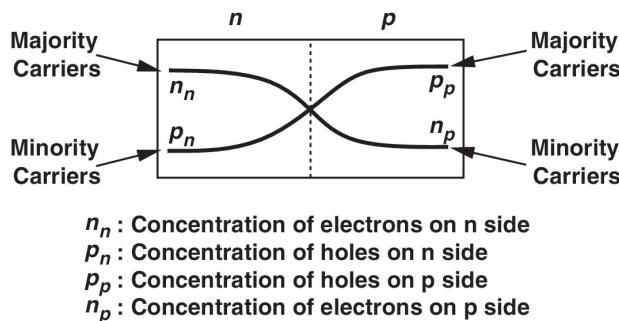
Im thermischen Gleichgewicht muss das Fermi-Niveau im gesamten Halbleiter gleich sein, die potentielle Energie des p-Typs wird folglich relativ zum n-Typ angehoben und es bildet sich eine **Potentialstufe** zwischen Kathode und Anode aus. Diese Stufe wird **Built-In Potential** Φ_{bi} oder **Diffusionspotential** genannt und entspricht der Diffusionsspannung U_D , siehe (27)

$$U_D = -\Phi_{bi}. \quad (230)$$

Die Wahrscheinlichkeit freie Elektronen auf der n-Seite mit genügend Energie für das Leitungsband der p-Seite zu haben (analog für Löcher) ist um den **Boltzmann-Faktor** reduziert.

Die Diode im Gleichgewicht

Im thermischen Gleichgewicht ohne Ladungsinjektion und externe Spannungen ist die Diffusionsspannung nur durch die Dotierung des Halbleiters in den n-Typ und p-Typ Halbleitern gegeben. Die Ladungsträgerdichten weit weg von PN-Übergang entsprechen asymptotischen den Werten des extrinsischen Halbleiters. Im Übergangsbereich ändern sich die Ladungsdichten typischerweise in einem Bereich gegeben durch die **Diffusionslänge**, Gleichung (227).

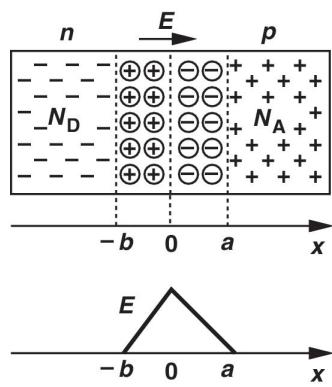


Quelle: B. Razavi, Fundamentals of Microelectronics

Der ionisierte Kristall der Raumladungszone erzeugt ein **elektrisches Feld**, welches den Diffusionsströmen entgegenwirkt. Im Falle eines 1-dimensionalen PN-Überganges gilt für das Feld in x -Richtung, Gleichung (16)

$$E(x) = -d\Phi(x)/dx. \quad (232)$$

Im Gleichgewicht kompensieren sich die Driftstöme veruracht vom elektrischen Feld $E(x)$ und die Diffusionsströme verursacht vom Konzentrationsverlauf der Ladungsträger.



Quelle: B. Razavi, Fundamentals of Microelectronics

Der Löcherstrom, Gleichung (212), im Gleichgewicht verschwindet an allen x -Positionen,

$$j_p \equiv 0 = e p \mu_p E(x) - e D_p \frac{d}{dx} p(x). \quad (233)$$

Mithilfe von Gleichung (232) wird das elektrische Feld durch den Potentialverlauf $\Phi(x)$ ausgedrückt. Durch Separation der Variablen kann der Potentialverlauf über die gesamte Diode integriert werden, siehe auch Gleichung (27),

$$\Phi_{bi} = - \int_{-\infty}^{\infty} d\Phi = \frac{D_p}{\mu_p} \int_{p_n}^{p_p} \frac{dp}{p} = \frac{D_p}{\mu_p} \ln \frac{p_p}{p_n}. \quad (234)$$

Das Diffusionspotential ist folglich durch die asymptotischen Löcherdichten vollständig bestimmt.

Im thermischen Gleichgewicht kann die Diffusionskonstante durch die **Einstein Relation**, Gleichung (205), ausgedrückt werden. Unter Verwendung der asymptotischen Löcherdichten, Gleichung (231) folgt schliesslich das **Diffusionspotential**

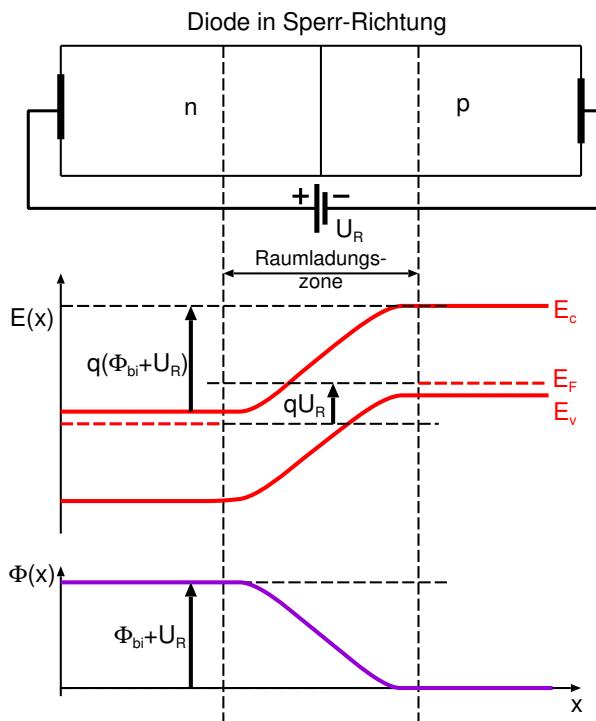
$$\Phi_{bi} = \frac{kT}{e} \ln \frac{N_A N_D}{n_i^2}. \quad (235)$$

N.B.: Das Diffusionspotential hängt von den Dotierdichten N_D und N_A in symmetrischer Weise ab. Daraus folgt unmittelbar, dass die Herleitung über den Elektronenstrom zum gleichen Ergebnis führt.

Aufgabe 48: Zeige am Beispiel eines mit $N_D = 10^{15} \text{ cm}^{-3}$ und $N_A = 10^{15} \text{ cm}^{-3}$ dotierten PN-Übergangs in Si, dass das Diffusionspotential im Wesentlichen durch den Bandabstand E_g gegeben ist. Nimmt das Diffusionspotential mit steigender Temperatur zu oder ab?

Die Diode in Sperr-Richtung

Wird zwischen Anode und Kathode eine negative Spannung angelegt, eine sogenannte **Sperrspannung** oder **Reverse Bias Voltage** U_R , vergrössert sich die Stufe der potentiellen Energie zwischen p-Typ und n-Typ, siehe Abbildung.



- Die zusätzliche potentielle Energie beträgt

$$\Delta E_{\text{pot}} = qU_R, \quad (236)$$

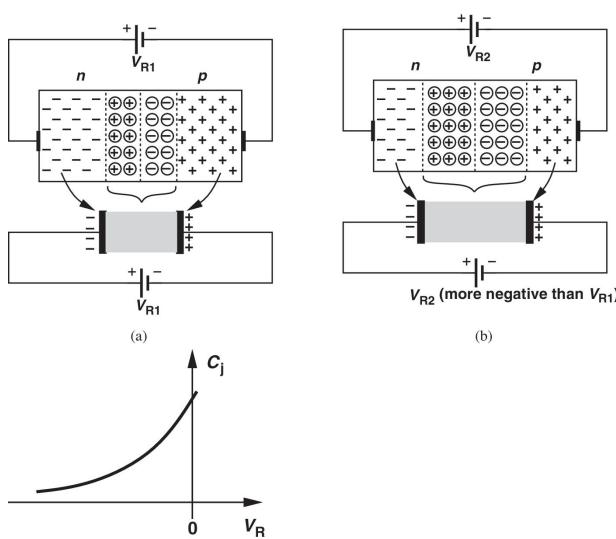
und definiert die Stufe der Fermi-Energie der Diode mit Spannung in Sperr-Richtung.

- Die Sperrspannung zieht zusätzliche Elektronen aus der Kathode (n-Typ) und zusätzliche Löcher aus der Anode (p-Typ).
- Als Folge verbreitert sich die Verarmungszone am PN-Übergang und die Raumladungszone verbreitert sich mit steigender Sperrspannung.

Für freie Ladungsträger wird es mit steigender Sperrspannung immer unwahrscheinlicher (Boltzmann-Faktor) die Verarmungszone zu durchqueren.

Die Diode blockiert Ströme in Sperr-Richtung.

Eine Diode in Sperr-Richtung kann als spannungsabhängiger Kondensator (**Varaktor**) verwendet werden. Kathode und Anode bilden mit ihren jeweiligen Majoritätsladungsträgern die geladenen Platten des Kondensators, die Verarmungszone bildet das (ladungsträgerfreie) Dielektrikum, siehe Abbildung.



Mit steigend negativer Sperrspannung, vergrößert sich die Raumladungszone und der Kapazitätswert sinkt. Die **Sperrsicht-Kapazität** C_j pro Fläche als Funktion der negativen Sperrspannung $V_R < 0$ kann wie folgt geschrieben werden (ohne Herleitung),

$$C_j = \sqrt{\frac{\varepsilon e}{2} \frac{N_a N_D}{N_A + N_D} \frac{1}{\Phi_{bi}}} \frac{1}{\sqrt{1 - V_R/\Phi_{bi}}}. \quad (237)$$

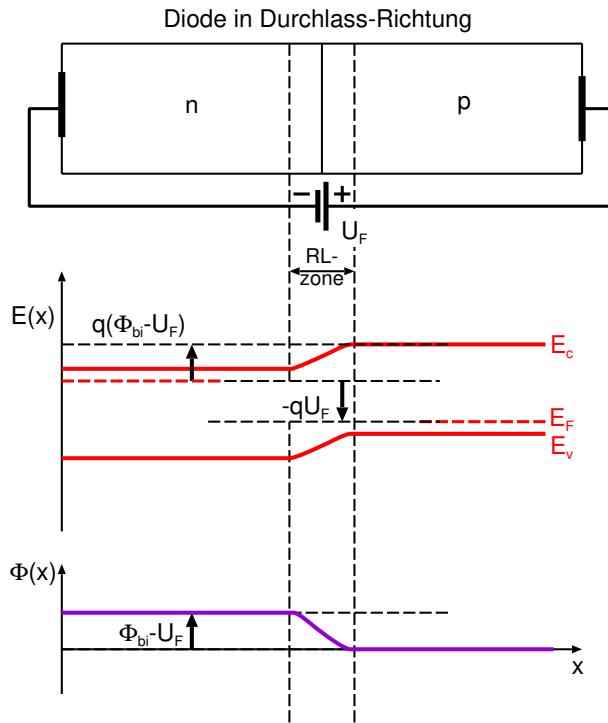
Varaktoren werden auch **Kapazitätsdioden** genannt und finden Anwendung in abstimmbaren LC-Schwingkreisen, z.B. bei spannungsgesteuerten Oszillatoren (Voltage-Controlled Oscillator oder VCO) mit Schwingkreisen hoher Güte.

Quelle: B. Razavi, Fundamentals of Microelectronics

Aufgabe 49: Wie kann ein abstimmbarer LC-Schwingkreis mit Kapazitätsdiode aufgebaut werden, ohne, im Fall einer Schwingung, die Vorspannung in Sperr-Richtung zu verlieren?

Die Diode in Durchlaß-Richtung

Wird zwischen Anode und Kathode eine positive Spannung angelegt, die sogenannte **Diodenspannung** oder **Forward Bias Voltage** U_F , verkleinert sich die Stufe der potentiellen Energie zwischen p-Typ und n-Typ, siehe Abbildung.



- Die Reduktion der potentielle Energie beträgt

$$\Delta E_{\text{pot}} = -qU_F, \quad (238)$$

und definiert die Stufe der Fermi-Energie der Diode mit Spannung in Durchlaß-Richtung.

- Die Diodenspannung zieht zusätzliche Elektronen in die Kathode (n-Typ) und zusätzliche Löcher in die Anode (p-Typ).
- Als Folge verkleinert sich die Verarmungszone am PN-Übergang die Raumladungszone schrumpft mit steigender Diodenspannung.

Für freie Ladungsträger wird es mit steigender Diodenspannung immer wahrscheinlicher (Boltzmann-Faktor) die Verarmungszone zu durchqueren.

Die Diode leitet Ströme in Durchlaß-Richtung.

Die positive Diodenspannung in Durchlaß-Richtung reduziert die Diffusionsspannung auf $\Phi_{bi} - U_F$, wodurch sich die **Minoritätsdichte vergrößert** (relativ zum Gleichgewicht mit $U_F = 0V$), siehe Abbildung.

Die asymptotische Majoritätsdichte ist durch die Dotierung gegeben und daher unverändert,

$$p_{p,f} \approx p_{p,e} \approx N_A \quad (239)$$

$$n_{n,f} \approx n_{n,e} \approx N_D. \quad (240)$$

die Minoritätsdichte wird angehoben,

$$\Delta p_n = p_{n,f} - p_{n,e} \quad (241)$$

$$\Delta n_p = n_{n,f} - n_{n,e}. \quad (242)$$

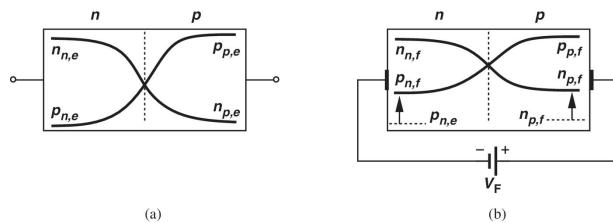
f und e bezeichnen die Größen in Durchlaßrichtung ('forward bias') und Gleichgewicht ('equilibrium').

In Gleichung (234) wurde bereits das Verhältnis der asymptotischen Löcherdichten mit dem Potentialverlauf ermittelt. Für den Fall einer Diode in Durchlaß-Richtung gilt die gleiche Beziehung mit der entsprechend reduzierten Diffusionsspannung $\Phi_{bi} - U_F$, analog die Beziehungen für die Elektronendichten:

$$\frac{p_{n,e}}{p_{p,e}} = \frac{n_{p,e}}{n_{n,e}} = e^{-\frac{e\Phi_{bi}}{kT}} \quad (243)$$

$$\frac{p_{n,f}}{p_{p,f}} = \frac{n_{p,f}}{n_{n,f}} = e^{-\frac{e(\Phi_{bi}-U_F)}{kT}} \quad (244)$$

Aufgabe 50: Wie können bei einer Diode in Durchlaß-Richtung die Anzahl der Elektronen im n-Typ und der Löcher im p-Typ steigen obwohl die asymptotischen Majoritätsdichten gleichbleiben?



Quelle: B. Razavi, Fundamentals of Microelectronics

Die Dioden-Gleichung (Shockley Gleichung)

Die Minoritätsladungen im PN-Übergang bestimmt das elektrische Verhalten einer Diode in Durchlass-Richtung. Die Diodenkennlinie (Strom durch die Diode I_D als Funktion der Diodenspannung U_D) wurde erstmals von W. Shockley^a hergeleitet und wird daher **Shockley Gleichung** genannt.

Die Anhebung der Minoritätsladungsträger in Abhängigkeit der Diodenspannung wurde bereits für die Löcherdichten (und analog für die Elektronendichten) hergeleitet, siehe Gleichung (243) und (244):

$$\Delta p_n(U_D) = p_{p,f} e^{-\frac{e(\Phi_{bi}-U_D)}{kT}} - p_{p,e} e^{-\frac{e\Phi_{bi}}{kT}} \approx \frac{N_A}{e^{\frac{e\Phi_{bi}}{kT}}} \left(e^{\frac{eU_D}{kT}} - 1 \right) \quad (245)$$

$$\Delta n_p(U_F) = n_{n,f} e^{-\frac{e(\Phi_{bi}-U_D)}{kT}} - n_{n,e} e^{-\frac{e\Phi_{bi}}{kT}} \approx \frac{N_D}{e^{\frac{e\Phi_{bi}}{kT}}} \left(e^{\frac{eU_D}{kT}} - 1 \right). \quad (246)$$

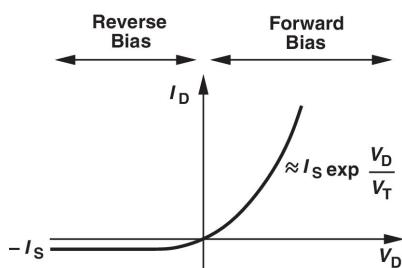
Die **Stromdichte**, Gleichung (210), als Funktion der Diodenspannung folgt aus dem Diffusionsstrom,

$$j = eD_n \frac{d}{dx} \Delta n_p(U_D) - eD_p \frac{d}{dx} \Delta p_n(U_D). \quad (247)$$

Die räumlichen Ableitung können mit der **Diffusionslänge**, Gleichung (227), approximiert werden, d.h. für Elektronen $d/dx \approx 1/L_n$ und Löcher $d/dx \approx -1/L_p$, und es folgt die **Shockley Gleichung** für Dioden

$$j = j_S \left(e^{\frac{eU_D}{kT}} - 1 \right). \quad (248)$$

Als Abkürzung wurde der **Sättigungsstromdichte** in Sperrrichtung $j_S = en_i^2 \left(\frac{D_n}{N_A L_n} + \frac{D_p}{N_D L_p} \right)$ eingeführt.



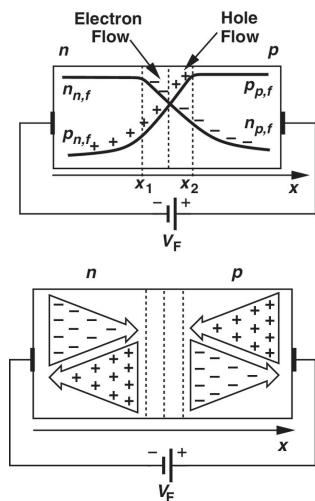
Quelle: B. Razavi, Fundamentals of Microelectronics

Der **Diodenstrom** I_D als Funktion der **Diodenspannung** U_D zwischen Anode (p-Typ) und Kathode (n-Typ) ist gerade die Stromdichte, Gleichung (248), integriert über die Querschnittsfläche des PN-Überganges (und daher von der Geometrie der Diode abhängig). Die **Diodenkennlinie** lautet:

$$I_D(U_D) = I_S \left(e^{\frac{U_D}{U_T}} - 1 \right), \quad (249)$$

mit der thermischen Spannung $U_T := kT/e$. Der Sättigungsstrom I_S gibt den thermisch erzeugten Leckstrom in Sperrrichtung an.

Die physikalischen Mechanismen der leitenden Diode sind in der folgenden Abbildung zusammengefasst:



Quelle: B. Razavi, Fundamentals of

Microelectronics

Aufgabe 51: Warum leuchten leitende Si-Dioden nicht? In welcher Farbe leuchten GaAsP Dioden?

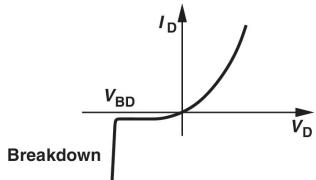
^aF. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, and W. Shockley, Phys. Rev. **81**, 4 (1951)

Effekte höherer Ordnung:

Bisher wurde das Verhalten idealer Dioden behandelt, reale Dioden können von diesem Verhalten abweichen. Im Folgenden werden zwei relevante Effekte höherer Ordnung (d.h. jenseits der idealen Diode) behandelt, der Diodendurchbruch und die Sperrverzögerungsladung (reverse recovery charge).

Diodendurchbruch

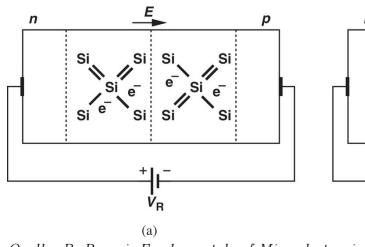
Bei hoher Sperr-Spannung kommt es zu einem sprunghaften Anstieg des Diodenstroms, siehe Abbildung.



Quelle: B. Razavi, Fundamentals of Microelectronics

Durchbruch tritt bei allen Dioden auf. Je nach Bauart der Diode, kann der Durchbruch zur (thermischen) **Zerstörung** der Diode führen oder ein verwendbarer **Betriebsmodus** sein.

- Dioden in Sperr-Richtung blockieren (bis auf den Sättigungsstrom I_S) den Stromfluss
- Wird die Sperr-Spannung zu groß, kommt es zu einem sprunghaften Anstieg des Stromes, dem sogenannten **Diodendurchbruch**
- Die Spannung des Durchbruchs heißt **Durchbruch-Spannung** V_{BD}



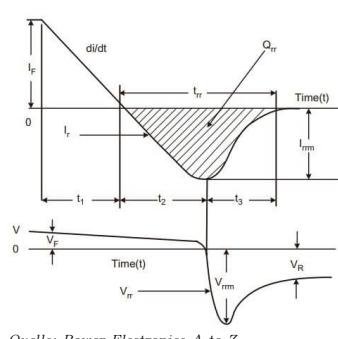
Quelle: B. Razavi, Fundamentals of Microelectronics

Beim **Zener-Effekt** erreicht das elektrische Feld in der Raumladungszone Werte in der Größenordnung $1V/\mu m$, und daher ausreichend um Elektronen aus ihren kovalenten Bindungen zu reissen. Diese zusätzlichen freien Elektronen bewirken den sprunghaften Anstieg des Stromes. Der **Zener Effekt**, also der sprunghafte Stromanstieg bei einer definierten Spannung V_{BD} , wird in speziell gebauten Dioden (**Zener Dioden**), z.B. zur Spannungsstabilisierung oder für den Überspannungsschutz verwendet. Zener Dioden zeichnen sich durch hohe Dotierungen aus, welche zu schmalen Raumladungszonen und daher großen Feldstärken führen.

Der **Lawinen-Effekt** dominiert den Diodendurchbruch bei PN-Übergängen mit moderaten Dotierungen, $< 10^{15}/cm^3$. Freie Ladungsträger (z.b. thermisch generiert) in der Raumladungszone werden durch das starke elektrische Feld beschleunigt. Sobald die so erlangte kinetische Energie ausreicht, um gebundene Elektronen loszureißen werden durch Söße zusätzliche freie Elektronen erzeugt (**Stoß-Ionisation**).

Sperrverzögerungsladung (reverse recovery charge)

Werden Dioden dynamisch von Durchlass-auf Sperr-Betrieb geschaltet, muss sich zuerst die Verarmungszone bilden, d.h. die Ladungsträger aus der Raumladungszone entfernt werden. Als Folge fliesst kurzzeitig (einige ns) ein Strom entgegen der Diodenrichtung (**reverse current**), siehe Abbildung.



Quelle: Power Electronics A to Z

Der Diodendurchbruch bei starkem E -Feld in der Raumladungszone erfolgt über zwei Mechanismen:

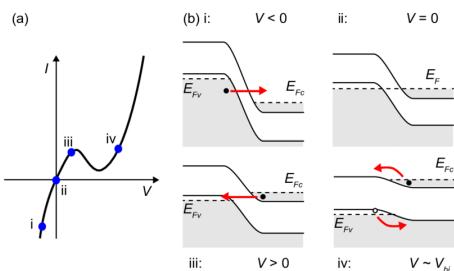
- a Zener-Effekt
- b Lawinen-Effekt (Avalanche Breakdown)

- Dieser verzögerte Start des Sperrbetriebes wird **Sperrverzögerung** genannt (**reverse recovery time**), t_{rr}
- Das zeitliche Integral des Reverse-Stromes ist die **Sperrverzögerungsladung** oder **Reverse Recovery Charge** Q_{RR} , und ein charakteristischer Parameter einer gegebenen Diode
- Die maximale Reverse-Strom I_{rrm} hängt vom zeitlichen Abfall di/dt des Diodenstroms beim Umschalten ab
- Die **Sperrverzögerung** limitiert die maximale Umschaltzeit von Dioden und erzeugt Schaltverluste

Aufgabe 52: Wie hängt die Sperrverzögerung t_{rr} vom Stromabfall di/dt ab (setze $t_3 \approx 0$)?

Diodentypen:

Bereits bei Dioden basierend auf klassischen PN-Übergängen gibt es, wie schon erwähnt, die Möglichkeit durch hohe Dotierung Dioden mit dominatem Zener-Durchbruch bei einer definierten **Zener-Spannung** herzustellen. Dioden dieser Eigenschaft werden **Zener Dioden** genannt. Wird die Dotierung weiter erhöht, tritt ein interessanter wellenmechanischer Effekt ein, der **Tunneleffekt**, welcher Dioden mit einer vollkommenen neuen Kennlinie erzeugt, die sogenannten **Tunneldioden**^a, siehe Abbildung. Tunneldioden bestehen oft aus Germanium, mit Dotierdichten im Bereich $10^{19} - 10^{21} \text{ cm}^{-3}$. Durch die sehr hohe Dotierung resultiert eine sehr schmale Raumladungszone, im Bereich von typischerweise 10 nm , welche von den quantenmechanischen Wellenfunktionen der Elektronen und Löchern durchtunnelt werden können.



a Kennlinie einer typischen Tunneldiode

b Schematisches Energiediagramm für

i : Rückwärts Spannung

ii : Thermisches Gleichgewicht

iii : Geringe Vorwärtsspannung

iv : Vorwärtsspannung im Bereich von Φ_{bi} .

Quelle: M. Malinverni, EPFL thesis (2015)

Tunneldioden sperren bei negativer Spannung nicht, da der Tunnelstrom den Teilchen ein Durchqueren der Raumladungszone ermöglicht. Der Tunnelstrom dominiert auch für geringe Spannungen in Vorwärtsrichtung. Bei großen Vorwärtsspannungen geht die Kennlinie in eine normale PN-Kennlinie über. Der Übergangsbereich zwischen diesen beiden Domänen in Vorwärstrichtung ergibt einen Bereich mit fallender Kennlinie, d.h. negativem differentiellen Widerstand. Tunneldioden in diesem Bereich werden als aktive Elemente in Verstärkern und Oszillatoren verwendet.

Wichtige **Diodenarten** sind in der folgenden Abbildung, mit Schaltungssymbol, zusammengefasst

Diode	
Zenerdiode	
Schottkydiode	
Kapazitätsdiode	
Leuchtdiode	
Photodiode	
Tunneldiode	

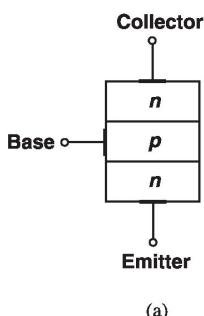
- **Zenerdioden:** hochdotierter PN-Übergang mit zerstörungsfreiem Zener-Durchbruch als Spannungsreferenz oder Überspannungsschutz
- **Schottkydioden:** eine Seite durch Metall ersetzt, Halbleiter-Metall Übergang, Majoritätsladungsstrom in Durchlass-Richtung, **unipolares** Bauteil, höhere Leckströme, geringe Kniespannung, geringe **Sperrverzögerung** und daher als Hochfrequenz-Dioden geeignet
- **Kapazitätsdioden:** in Sperr-Richtung als spannungsgesteuerte Kapazitäten verwendet (Varaktoren)
- **Leuchtdioden:** direkte Halbleiter, mit vorwiegend strahlender Rekombination, Bandabstand bestimmt die Frequenz der emittierten Photonen
- **Photodioden:** Lichteinfall auf Raumladungszone erzeugt Elektron-Loch-Paare, Diffusionsspannung bringt diese zusätzlichen Elektronen zur Kathode und Löcher zur Anode, dienen als Lichtsensoren oder PV-Zellen. Wird die Raumladungszone durch einen intrinsischen Halbleiter zwischen den extrinsischen Gebieten verbreitert, spricht man von PIN-Dioden.
- **Tunneldioden:** Sehr hoch dotierte PN-Übergänge führen zu Dioden mit einem Bereich negativer differentieller Spannung. Tunneldioden werden als Verstärker bei hohen Frequenzen eingesetzt

Aufgabe 53: Wie wirkt sich der intrinsische Bereich der PIN-Dioden auf die Sperrverzögerung aus?

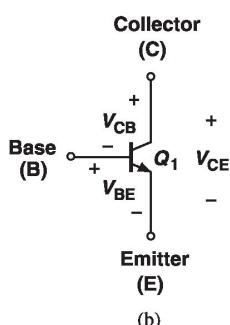
^aL. Esaki, Phys.Rev. **109** (2) (1958)

7.2– Der Bipolar-Transistor

Wird der PN-Übergang durch einen weiteren PN-Übergang ergänzt, wobei die beiden Übergänge einen extrinsischen Halbleiter teilen, siehe Abbildung, entsteht ein weiteres Halbleiterbauelement, der **Bipolar Junction Transistor** (BJT) ^a, oft auch **Bipolartransistor** oder nur **Transistor** genannt. Der Name Transistor wurde von J. Pierce (1949) geprägt und leitet sich von 'trans-resistor' ab. Ein Transistor kann folglich als schaltbarer Widerstand betrieben werden, d.h. ein elektrisch gesteuerter Schalter, der einen Strompfad zwischen leitend (niederohmig) und sperrend (hochohmig) schalten kann.



(a)

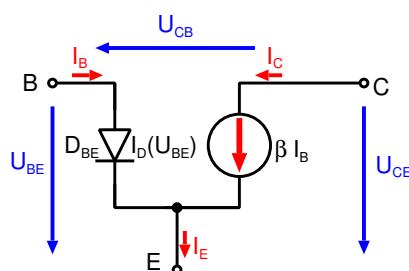


(b)

Quelle: B. Razavi, Fundamentals of Microelectronics

Der Bipolartransistor, bei dem sich die beiden PN-Übergänge den p-Typ Halbleiter teilen wird NPN Transistor genannt. Der elektrische Anschluss am p-Typ wird **Basis** (B) genannt und bildet mit dem sogenannten **Emitter** (E) eine leitende Diode bei positiver Diodenspannung U_{BE} . Die Diodenrichtung wird durch einen Pfeil von der Basis in Richtung Emitter angezeigt. Der zweite PN-Übergang führt zum sogenannten **Kollektor** mit der Sperrspannung U_{CB} . Die Kollektor-Emitter Spannung wird mit U_{CE} bezeichnet.

Der Bipolartransistor verhält sich in Durchlassrichtung wie eine **spannungsgesteuerte Strom Quelle**, wobei die Basis-Emitter Spannung U_{BE} eine Stromquelle zwischen Kollektor und Emitter auf und zu steuert, siehe Abbildung.



Die Basis (B) bildet zusammen mit dem Emittor (E) die Basis-Emitter Diode D_{BE} in Durchlassrichtung. Abhängig von der Diodenspannung U_{BE} fließt, entsprechend der Diodenkennlinie $I(U_{BE})$, Gleichung 249, ein Strom von der Basis zum Emittor, I_B . Dieser (geringe) Basis-Strom steuert einen direkt proportionalen Strom I_C , genannt Kollektorstrom, zwischen Kollektor (C) und Emittor (E), welcher folglich ebenfalls eine Diodencharakteristik in Bezug auf die Diodenspannung aufweist.

Der Kollektorstrom I_C ist typischerweise deutlich größer (z.B. Faktor 50) als der Basisstrom I_B mit der **differentiellen Stromverstärkung** $\beta := \partial I_C / \partial I_B \approx I_C / I_B \gg 1$. Der Emittorstrom I_E setzt sich aus dem Kollektorstrom I_C und dem Basisstrom I_B zusammen, $I_E = I_C(1 + 1/\beta) \approx I_C$.

Im Folgenden wird das Transistorverhalten anhand des NPN-Transistors erklärt. Analog kann ein Transistor auch aus zwei p-Typ Halbleitern mit einem n-Typ dazwischen gebildet werden. Bei diesem sogenannte PNP-Transistor wir der Anschluss in der Mitte wieder als Basis bezeichnet, mit den beiden PN-Übergängen zu Kollektor und Emittor. Das Verhalten ist analog zum NPN-Transistor und wird anschliessend erklärt.

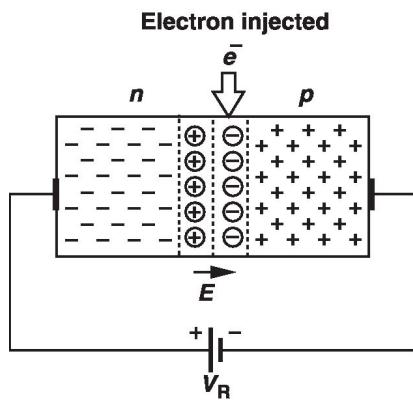
Während der Basisstrom einfach nur einer leitenden Diode entspricht ist die Entstehung des Kollektorstromes durch die sperrende Kollektor-Basis Raumladungszone nicht offensichtlich. Wir betrachten dazu zuerst die Ladungsinjektion in eine Raumladungszone.

Aufgabe 54: In welche Richtung werden injizierte Elektronen in einer Raumladungszone beschleunigt?
In welche Richtung werden injizierte Löcher beschleunigt?

^aJ. Bardeen, W. Brattain and W. Shockley (1947)

Der Bipolar-Transistor: Ladungsinjektion

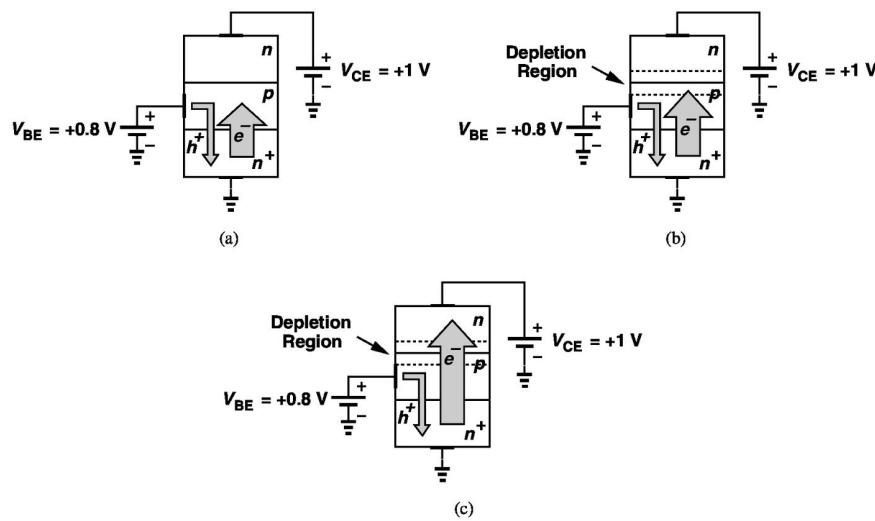
Betrachten wir eine Diode in Sperr-Richtung, also z.B. die Kollektor-Basis Diode bei positiver Spannung U_{CB} . In der Raumladungszone sind die freien Ladungsträger deutlich reduziert und der Driftstrom, bewirkt durch die Raumladung, wirkt dem Diffusionsstrom entgegen, siehe Abbildung.



Quelle: B. Razavi, Fundamentals of Microelectronics

Im p-Typ Halbleiter bilden die Elektronen die Minoritätsladungsträger. Werden nun in die Raumladungszone beim p-Typ Minoritätskladungsträger (also Elektronen) injiziert, führt das Feld der Raumladungen zu einer starken Beschleunigung der injizierten Elektronen über die Raumladungszone in den n-Typ. Die injizierten Elektronen fließen also sofort als **Dirftstrom** durch die Raumladungszone und werden vollständig vom n-Typ 'aufgesammelt' - daher der Name **Kollektor**.

Betrachten wir nun wieder den ganzen Bipolartransistor, mit leitender Basis-Emitter-Diode und positiver Kollektor-Emitter-Spannung U_{CE} , siehe Abbildung. Die positive Diodenspannung U_{BE} ergibt einen Löcherstrom (h^+) von der Basis zum Emittor, Abbildung (a), und eine Elektronenstrom (e^-) vom Emittor in die Basis. Der sperrende Übergang zwischen Kollektor und Basis bildet eine Raumladungszone, mit den Elektronen der Basis als Minoritätsladungsträgern, Abbildung (b). Alle Elektronen, welche auf dem Weg durch die Basis noch nicht rekombiniert sind und den Rand dieser Raumladungszone erreichen, werden zum Kollektor abgesaugt. Es fliesst folglich ein großer Elektronenstrom vom **Emitter** (daher der Name) zum **Kollektor** sobald aufgrund der leitenden Basis-Emitter Diode freie Elektronen in der Basis vorhanden sind. Dieser Strom ist von der Kollektorspannung quasi unabhängig, solange durch einen sperrenden Kollektor-Basis-Übergang eine Raumladungszone vorhanden ist.



Quelle: B. Razavi, Fundamentals of Microelectronics

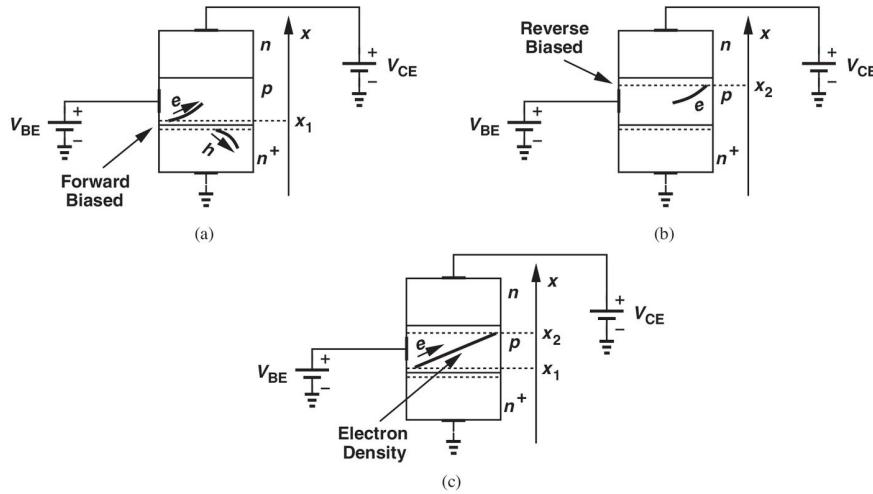
Wie erreichen die Elektronen den Rand der sperrenden Raumladungszone?

Aufgabe 55: Durch welchen physikalischen Prozess werden Elektronen in die Verarmungszone injiziert?

Der Bipolar-Transistor: Diffusionsstrom

Um den Leitungsmechanismus über den Basisbereich besser zu verstehen, ist es hilfreich sich die Majoritäts- und Minoritätsladungsträger in den einzelnen Regionen vor Augen zu führen.

- Emitter (n-Typ): Elektronen sind Majoritätsladungsträger mit Teilchendichte N_D , siehe Abbildung (a).
- Basis (p-Typ): Elektronen sind Minoritätsladungsträger, d.h. bei breiter Basis (d.h. deutlich über der Diffusionslänge L_n , Gleichung (228)) baut sich die Elektronendichte auf die Minoritätsdichte n_i^2/N_A ab.
- Kollektor (n-Typ): Elektronen sind Majoritätsladungsträger, aber von der Basis durch eine sperrende Raumladungszone mit sehr niedriger Elektronendichte getrennt, siehe Abbildung (b).



Quelle: B. Razavi, Fundamentals of Microelectronics

Der schmale p-dotierte Bereich (Basis) ist folglich ein Bereich hoher Elektronendichte an der leitenden Basis-Emitter Diode und geringer Elektronendichte an der Grenze zur Raumladungszone des Kollektors. Der große Gradient der Minoritätsdichten in der Basis, siehe Abbildung (c), bewirkt einen starken **Diffusionsstrom der Minoritätsladungen** über den Basisbereich und daher zu einem permanenten Elektronenfluss in den Randbereich der sperrenden Raumladungszone - also zu einer permanenten **Minoritätsladungs-Injektion**. Diese durch den Basis-Diffusionsstrom injizierten Elektronen werden vom Kollektor permanent als Driftstrom in der Raumladungszone abgesaugt. Der Leitungsmechanismus des **Kollektorstroms** eines Bipolartransistors setzt sich also aus dem Diffusionsstrom der Minoritätsladungen über die Basis, gefolgt von einem Driftstrom durch den sperrenden Kollektor-Basis Diode zusammen.

Der Elektronenüberschuss am **Basis-Emitter Übergang**, $x = x_1$, entspricht Gleichung, (246):

$$\Delta n(x_1) = \frac{n_i^2}{N_B} \left(e^{\frac{U_{BE}}{U_T}} - 1 \right). \quad (250)$$

Mit N_B wir hier die Dotierdichte der Basis bezeichnet. Der Dichtegradient der Elektronen über die Breite W_B der Basis bewirkt einen **Diffusionsstrom der Minoritätsladungen**,

$$j_n = eD_n \frac{d}{dx} n \underset{L_n \gg W_B}{\approx} eD_n \frac{\Delta n(x_1)}{W_B}. \quad (251)$$

Hier wurde die Ableitung durch den Abfall über die Weite W_B approximiert.

Der **Kollektorstrom** über den Emitterquerschnitt A_E des leitenden Bipolartransistors lautet nun:

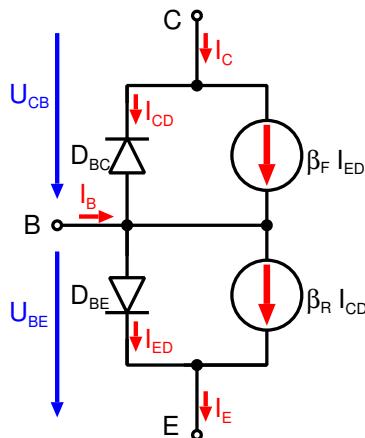
$$I_C = \frac{eA_E D_n n_i^2}{N_B W_B} \left(e^{\frac{U_{BE}}{U_T}} - 1 \right) \underset{U_{BE} \gg U_T}{\approx} I_S e^{\frac{U_{BE}}{U_T}} \quad (252)$$

Der Vorfaktor wird wieder in einen Sättigungssperrstrom $I_S := \frac{A_E e D_n n_i^2}{N_B W_B}$ zusammengefasst.

Aufgabe 56: Wie baut man einen Transistor mit großem Kollektorstrom (A_E , W_B , Dotierung Emitter)?

Der Bipolar-Transistor: Die Ebers-Moll Gleichungen

Der Aufbau eines Bipolartransistors ist, abgesehen von Dotierungsstärken und geometrischen Abmessungen symmetrisch bezüglich Emitter und Kollektor. Wird demnach in einem Bipolartransistor Emitter und Kollektor vertauscht, treten die gleichen physikalischen Effekte in Erscheinung. Das einfachste vollständige Modell eines Bipolartransistors ist das Ebers-Moll Modell ^a. Betrachten wir dazu folgende Abbildung.



Die Basis-Emitter Diode D_{BE} wird durch die Shockley-Gleichung (249) beschrieben:

$$I_{ED}(U_{BE}) = I_{SE} \left(e^{\frac{U_{BE}}{V_T}} - 1 \right) \quad (253)$$

Die Basis-Kollektor Diode D_{BE} wird durch die Shockley-Gleichung beschrieben,

$$I_{CD}(U_{CB}) = -I_{SC} \left(e^{\frac{-U_{CB}}{V_T}} - 1 \right), \quad (254)$$

wobei die Vorzeichenkonvention der Abbildung die Vorzeichenwechsel bewirkt. Mit I_{SE} und I_{SC} werden die Sperrströme der Basis-Emitter bzw. Basis-Kollektor Diode bezeichnet.

Der Strom durch die Basis-Emitter Diode, I_{ED} bewirkt einen Driftstrom zum Kollektor proportional zur Stromverstärkung β_F in Vorwärtsrichtung, also $I_{ED}\beta_F$. Analog bewirkt der Strom durch die leitende Basis-Kollektor Diode, I_{CD} eine Driftstrom zum Emitter, proportional zur Stromverstärkung β_R in Rückwärtsrichtung, also $I_{CD}\beta_R$.

In Summe setzt sich nun der Kollektorstrom I_C aus der Driftstrom des leitenden Basis-Emitter Überganges, Gleichung (252), und dem Strom der Basis-Kollektor Diode zusammen:

$$I_C = \beta_F I_{ED}(U_{BE}) + I_{CD}(U_{CB}) \quad (255)$$

Der Emitterstrom I_E ergibt sich analog aus dem Strom der Basis-Emitter Diode und dem Driftstrom:

$$I_E = I_{ED}(U_{BE}) + \beta_R I_{CD}(U_{CB}) \quad (256)$$

Der Basisstrom ergibt sich aus der Differenz zwischen dem Emitterstrom und dem Kollektorstrom:

$$I_B = I_E - I_C \quad (257)$$

Aus der Definition der Stromverstärkung und des Basistromes folgt nun, dass der Sperrstrom aus Gleichung (252), der verstärkte Sperrstrom der Emitterdiode bzw. der verstärkte Sperrstrom der Kollektordiode ist,

$$I_S \equiv \frac{\beta_F}{1 + \beta_F} I_{SE} \equiv \frac{\beta_R}{1 + \beta_R} I_{SC}. \quad (258)$$

Die vollständigen **Ebers-Moll Gleichungen** können nun als Funktion von nur drei Modellparametern I_S , β_F und β_R geschrieben werden:

$$I_C = I_S \left(e^{\frac{U_{BE}}{V_T}} - 1 \right) - I_S \left[1 + \frac{1}{\beta_R} \right] \left(e^{\frac{-U_{CB}}{V_T}} - 1 \right) \quad (259)$$

$$I_E = I_S \left[1 + \frac{1}{\beta_F} \right] \left(e^{\frac{U_{BE}}{V_T}} - 1 \right) - I_S \left(e^{\frac{-U_{CB}}{V_T}} - 1 \right) \quad (260)$$

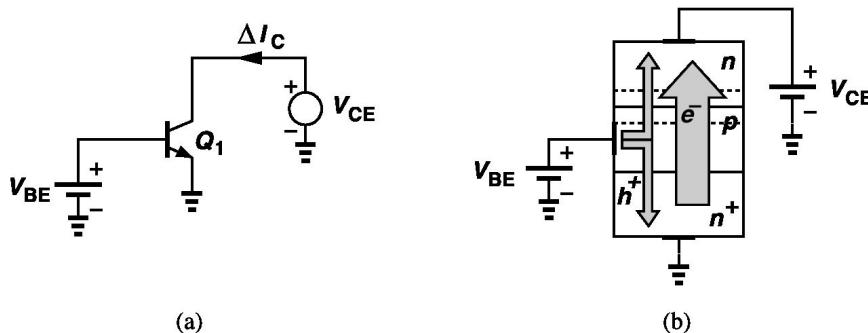
$$I_B = I_E - I_C \quad (261)$$

Aufgabe 57: Reduziert sich die Eber-Moll Gleichung bei einem leitenden Transistor in Vorwärtsrichtung auf Gleichung (252)? Was gilt in dieser Näherung für I_E und I_B ?

^aJ. L. Moll and J. J. Ebers (1954)

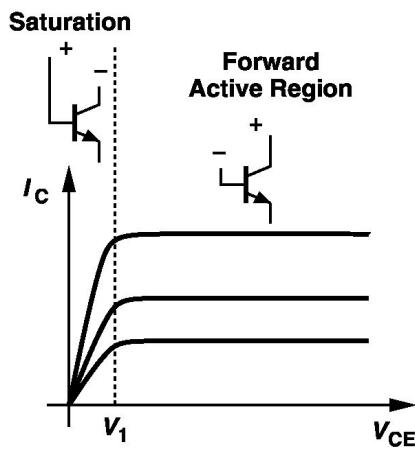
Der Bipolar-Transistor: Die Transistorkennlinie

Sobald die Kollektorspannung unter der Basisspannung liegt, $U_{CB} < 0V$, beginnt auch die Basis-Kollektor Diode zu leiten. In diesem Fall subtrahieren sich die Driftströme der beiden PN-Übergänge und der resultierende Kollektorstrom wird reduziert, siehe Abbildung. Man spricht von einem Bipolartransistor in **Sättigung**.



Quelle: B. Razavi, Fundamentals of Microelectronics

Die **Transistorkennlinie** in Sättigung ist wieder durch die Ebers-Moll Gleichungen, (259) -(261), gegeben. Wie erwartet, geht der Kollektorstrom für verschwindende Kollektor-Emitter Spannung, $U_{CE} \equiv 0V$, unabhängig von der Basisspannung gegen 0. Die Kennlinie des Kollektorstromes I_C als Funktion der Kollektor-Emitter Spannung U_{CE} für verschiedene Basis-Emitter Spannungen ist nun bekannt, siehe Abbildung:



Quelle: B. Razavi, Fundamentals of Microelectronics

- Bei verschwindender Kollektor-Emitter Spannung, $U_{CE} \equiv 0V$, verschwindet der Kollektorstrom
- Bei positiver Kollektor-Basis Spannung, $U_{CB} \gg 0V$ arbeitet der Bipolartransistor in **Durchlassrichtung** und entspricht einer **Konstantstromquelle**, deren Strom von der Basis-Emitter Spannung abhängt, $I_C(U_{BE})$
- Bei negativer Kollektor-Basis Spannung, $U_{CB} \lesssim 0V$, arbeitet der Transistor in **Sättigung** und erzeugt eine annähernd lineare Strom-Spannungskennlinie, kann also als **spannungs-gesteuerter Widerstand** betrieben werden.

Bisher wurden nur NPN-Transistoren behandelt, d.h. Transistoren mit einer n-dotierten Basis, angrenzend an Emitter und Kollektor vom p-Typ. NPN-Transistoren werden z.B. verwendet, um Konstantstromquellen nach Masse zu erzeugen. Sie sind auch die bevorzugte Wahl, wenn Basisspannungen größer als die Kniestromspannung der Basis-Emitter Diode verstärkt werden müssen. Für Konstantstromquellen zur Versorgungsspannung, bzw. für Spannungen nahe der Masse wird der komplementäre PNP-Transistor verwendet, welcher im Folgenden kurz vorgestellt wird.

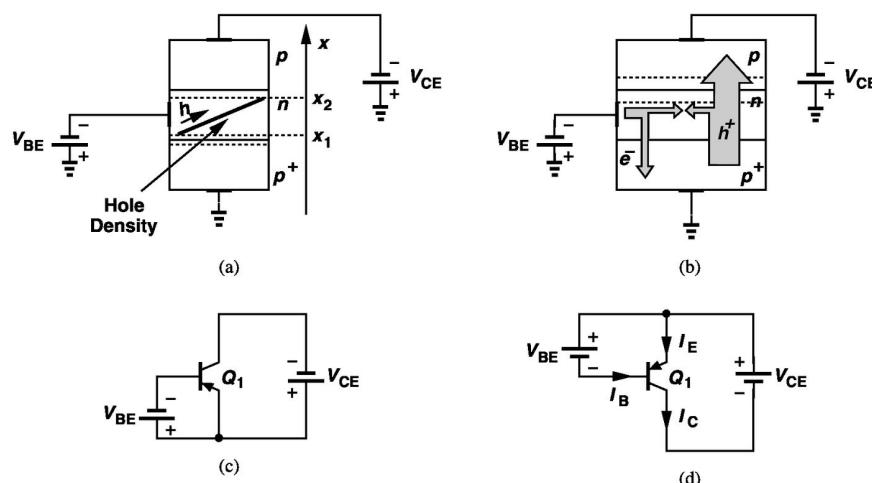
Aufgabe 58: Leite die lineare Kennlinie im Grenzfall kleiner Kollektor-Emitter Spannung, $U_{CE} \approx 0V$, her.

Der Bipolar-Transistor: Der PNP-Transistor

Werden bei einem Transistor Dotierungen zwischen n-Typ und p-Typ gerade vertauscht, entsteht ein **komplementärer Transistor** zum NPN-Transistor, genannt **PNP Transistor**. Entsprechend der umgekehrten Dotierung, vertauschen sich auch die weiteren Rollen.

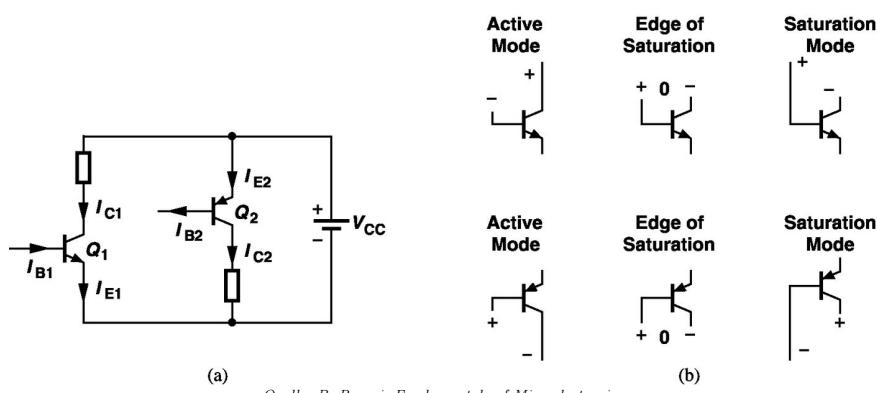
- Eine negative Basis-Emitter Spannung, Abbildung (a), ergibt eine leitenden Basis-Emitter Diode
- Elektronen sind die Majoritätsladungen in der Basis und fliessen durch die leitende Diode zum Emitter
- Der starke Gradient der Löcherdichte über die Basis-Breite ergibt einen Löcher-Diffusionsstrom.
- Bei negativer Kollektor Emitterspannung besteht eine Raumladungszone zwischen Basis und Kollektor
- Der Löcher-Diffusionsstrom injiziert Löcher in die Raumladungszone, welche als Driftstrom zum Kollektor fließen

Das Symbol des PNP-Transistors zeigt die Flussrichtung der Emitter-Diode, siehe Abbildung (c). Zumeist wird der PNP-Transistor mit dem Emitter in Richtung Versorgungsspannung betrieben, Abbildung (d).



Quelle: B. Razavi, Fundamentals of Microelectronics

Die Betriebsarten des NPN-und PNP-Transistors ist hier nochmals zusammengefasst:



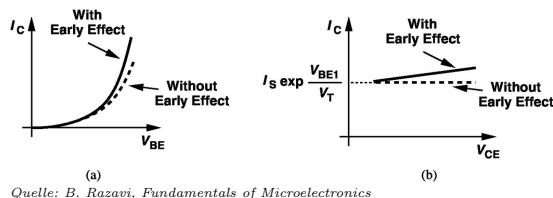
Quelle: B. Razavi, Fundamentals of Microelectronics

Bisher wurde nur das ideale Verhalten von Bipolartransistoren behandelt. Reale Transistoren sind keine idealen Stromquellen in Vorwärtsrichtung, wie in Folgenden kurz diskutiert wird.

Aufgabe 59: Warum leiten PNP-Transistoren typischerweise deutlich schlechter als NPN-Transistoren?

Effekte höherer Ordnung: Der Early Effekt

Die Kennlinien eines realen Bipolartransistors in Vorwärtsrichtung zeigt eine Abhängigkeit des Kollektorstromes I_C von der Kollektor-Emitter Spannung U_{CE} . In der bisherigen (idealen) Kennlinie hängt der Kollektorstrom eines leitenden Transistors lediglich von der Basisspannung ab, der Transistor verhält sich also wie eine ideale Stromquelle (d.h. mit unendlichem Innenwiderstand). Bei realen Transistoren zeigt sich jedoch, dass der Kollektorstrom mit steigender Kollektor-Emitter Spannung U_{CE} ebenfalls linear leicht ansteigt und daher über dem erwarteten Wert des idealen Modells liegt, siehe Abbildung.

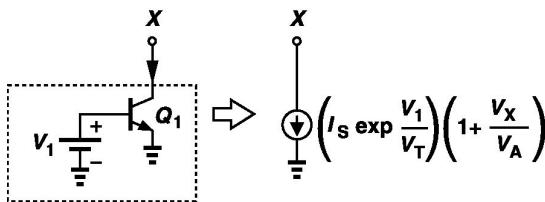


Im Modell kann dieser Anstieg

Im Modell kann dieser Anstieg durch einen Zusatzterm im Kollektorstrom berücksichtigt werden:

Die linear ansteigende Kennlinie des leitenden Transistors wird nach ihrem Entdecker **Early Effekt** genannt. Eine lineare Extrapolation dieses Anstieges ergibt einen Schnittpunkt mit der x-Achse bei negativen Kollektor-Emitter Spannungen, $U_{CE} = -U_A$.

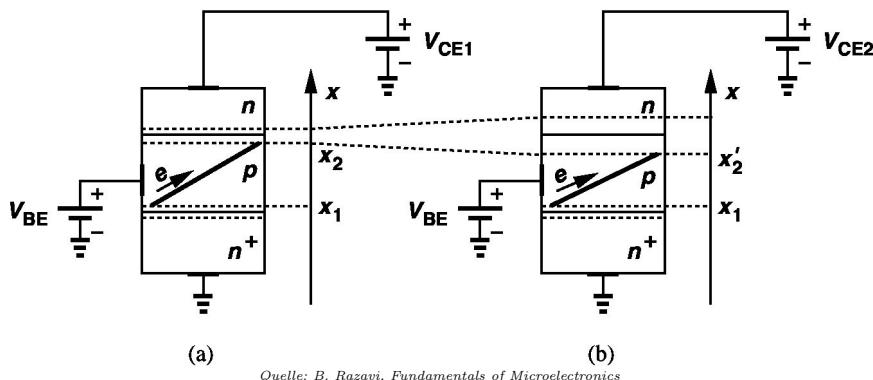
$$I_C \approx I_{SE} e^{\frac{U_{RE}}{U_T}} \left(1 + \frac{U_{CE}}{U_A} \right) \quad (262)$$



Quelle: B. Razavi, Fundamentals of Microelectronics

Der Parameter U_A mit der Dimension einer Spannung wird **Early Spannung** genannt und ist umgekehrt proportional zum Anstieg der realen Kennlinie. Die Early Spannung hängt von der Geometrie und Dotierung des Transistors ab und liegt bei realen Transistoren im Bereich von einigen Volt.

Die Ursache des **Early Effektes** liegt in der sogenannten **Basisweiten-Modulation**, siehe Abbildung:



- Steigende Kollektor-Spannung **vergrößert die Raumladungszone** der sperrenden Kollektor-Basis Diode
- Die verbreiterte Verarmungszone **reduziert die effektive Weite** der Basis, W_B
- Der Abbau der Minoritätsdichte in der Basis erfolgt daher über einen schmaleren Bereich, d.h. mit **vergrößertem Gradienten**
- Vergrößerter Gradient der Minoritätsladungen **erhöht den Diffusions Strom** über die Basis und daher den Kollektorstrom

Reale Transistoren verhalten sich im Durchlassbereich wie reale Stromquellen mit endlichem (wenn auch großem) Innenwiderstand.

Aufgabe 60: Wie lauten das elektrische Ersatzschaltbild eines realen Transistors als Stromquelle? Wie hängt der Nominalstrom der Innenwiderstand mit der Early Spannung zusammen?

7.3– Der Feldeffekt-Transistor

Der Feldeffekt-Transistor wurde deutlich vor dem Bipolartransistor von J. E. Lilienfeld erfunden (Patenteinreichung 1928), siehe Abbildung, die erste Realisierung erfolgte hingegen erst in den 1960-er Jahren.

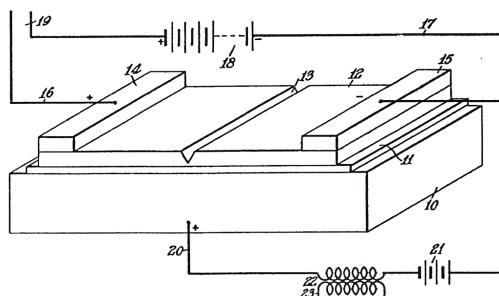
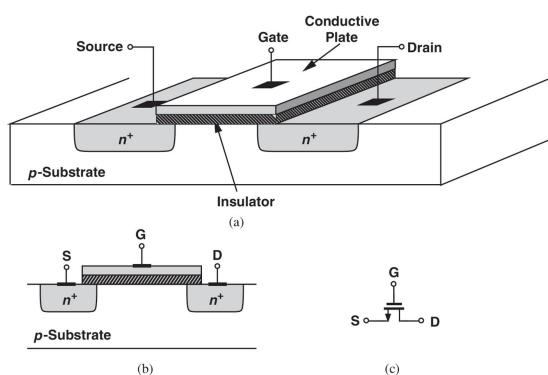


FIGURE 1.19
A figure from J. E. Lilienfeld's U.S. patent 1,900,018 (filed March 28, 1928; granted 1933). The structure should be turned upside down to correspond to the orientation used in this book (e.g., Fig. 1.20).

Quelle: Y. Tsividis, *The MOS Transistor*

Basierend auf den 3 Materialien wurde der Transistor **Metal-Oxid-Semiconductor-Field-Effect-Transistor** oder kurz **MOSFET** genannt. Es sind auch die Bezeichnungen **MOS** oder **FET** üblich. In modernen MOS-Technologien wird der Anschluss über dem Kanal aus fertigungstechnischen Gründen ebenfalls aus dotiertem Halbleitermaterial gefertigt (z.B. aus polykristallinem dotierten Silizium, genannt **Poly-silizium** oder kurz **Poly**). Analog zum Bipolartransistor kann ein MOSFET als p-dotierter Kanal zwischen zwei n-Typ Anschlüssen (ein sogenannter **N-MOS**) oder als n-dotierter Kanal zwischen zwei Anschlüssen vom p-Typ gefertigt werden (genannt **P-MOS**).

In Folgenden wird der N-MOS im Detail behandelt, der P-MOS folgt in völliger Analogie.



Quelle: B. Razavi, *Fundamentals of Microelectronics*

Moderne Mikrochips integrieren viele 100 Millionen MOSFETS auf einem einzelnen Chip.

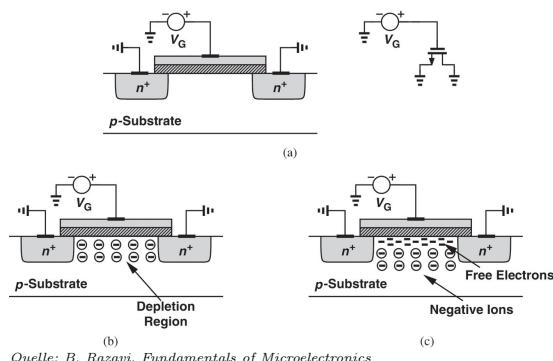
Analog zum Bipolar-Transistor ist ein MOSFET ebenfalls ein spannungsgesteuerter Schalter. Bei einer anliegenden Spannung zwischen Drain und Source (U_{DS}) verhält sich der MOSFET, in Abhängigkeit der Gate-Source Spannung (U_{GS}), entweder als hochohmige Verbindung (d.h. quasi als sperrender Schalter) oder als niederohmige Verbindung (d.h. als leitender Schalter). Nachdem das Gate elektrisch vom Rest des MOSFET getrennt ist, fließt kein Gleichstrom in das Gate, es ist lediglich die Spannung am Gate relevant. Speziell für hochintegrierte Schaltungen ist das ein wesentlicher Vorteil gegenüber Bipolar-Transistoren, weil dadurch die Bewahrung eines Schaltzustandes (also z.B. die elektrische Realisierung eines Bits) keinen Stromfluss erfordert. Lediglich das Umschalten zwischen leitend und sperrend erfordert einen Strom für die Umladung des Gates.

Aufgabe 61: Ein hochintegrierter Digitalchip mit 100 Millionen Transistoren sei in Bipolarechnologie realisiert. Wie hoch ist der Stromverbrauch des Chips, wenn die Hälfte der Bipolartransistoren zur Bewahrung ihres Bit-Zustandes mit einem Basisstrom von jeweils $1\mu A$ bestromt werden müssen? Der Chip sei mit einer Versorgungsspannung von $1.5V$ realisiert. Wie groß ist die dissipierte Leistung?

Der Feldeffekt-Transistor besteht, wie auch der Bipolartransistor, aus 3 Anschlüssen, wobei die beiden äusseren Anschlüsse extrinsische Halbleiter des gleichen Typs sind, die mit einem extrinsischen Halbleiter vom komplementären Typ verbunden sind (genannt **Kanal**). Der dritte Anschluss war ursprünglich ein Metallanschluss, der getrennt durch eine isolierende Schicht (z.B. aus **Siliziumdioxid** SiO_2) über dem Kanal angebracht ist.

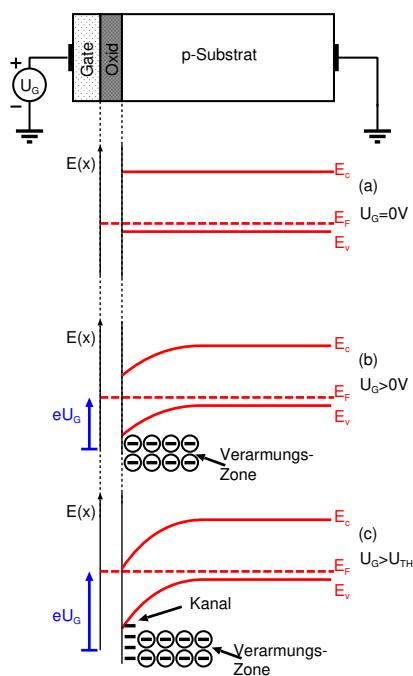
Der Feldeffekt-Transistor: Der Kanal

Das Schalten des MOSFETs erfolgt über die Gate-Source-Spannung U_{GS} . Liegt keine Spannung am Gate an, ist die elektrische Verbindung zwischen Drain und Source (beide n-Typ) durch den p-dotierten Halbleiter dazwischen unterbrochen, ab einer gewissen Gate-Source Spannung, genannt Schwellspannung oder **Thresholdspannung** U_{TH} bildet sich ein elektrisch **leitender Kanal** zwischen Drain und Source und der N-MOS wird leitend. Die Entstehung des Kanals wird im Folgenden genauer untersucht.



Drain, Source und Substrat eines N-MOS Transistors liegen auf Masse, siehe Abbildung (a). Wird nun das Gate relativ zur Masse positiv aufgeladen, also $U_{GS} > 0V$, befindet sich mit dem Gate eine positiv geladene Platte oberhalb des p-dotierten Kanals. Die freien Löcher des Kanals weichen dieser positiven Ladung und entfernen sich vom Oxid. Die zurückgelassenen Akzeptoren formen eine negativ geladene Verarmungszone, Abbildung (b). Wird die Gate-Sourcespannung über die Threshold-Spannung erhöht tritt die sogenannte **Inversion** auf, und es befinden sich freie Elektronen direkt unter dem Oxid, Abbildung (c). Diese Elektronen formen einen **elektrisch leitenden Kanal** zwischen Drain und Source, der MOSFET wird leitend.

Warum verlassen die freien Löcher den Kanal und wie kommen die freien Elektronen in den Kanal?



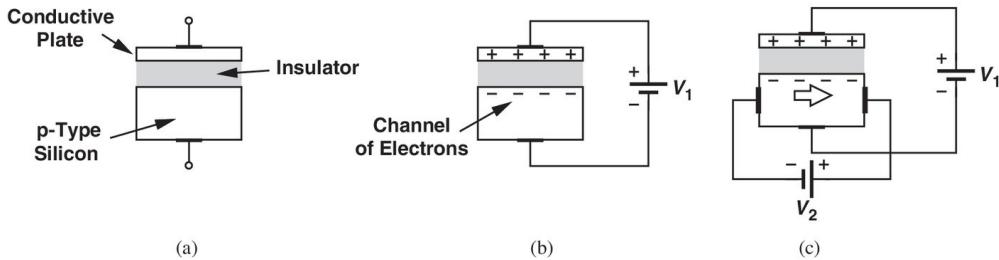
- Ohne Vorspannung, Abbildung (a), verteilen sich die freien Löcher im Valenzband mit Energie E_v gleichmäßig auf das p-dotierte Substrat.
- Durch Anbringen einer positiven Gate-Spannung $U_{GS} > 0$, Abbildung (b), wird die Energie des Valenzbandes (vergleiche PN-Übergang) nach unten gezogen, die Energie der freien Löcher bleibt aber E_v des p-Substrates im Gleichgewicht. Folglich bewegen sich die Löcher in die energetisch begünstigte Region und hinterlassen eine Verarmungszone in der Nähe des Gates. Die Fermienergie bleibt überall gleich, wodurch auch die thermische Erzeugung freier Löcher nahe des Oxids sehr unwahrscheinlich wird.
- Eine weiterer Vergrößerung des Gate-Potentials, Abbildung (c), bringt schliesslich auch das Leitungsband unter das Fermi-Niveau des Halbleiters. Diese freien Elektronen unter dem Oxid können thermisch freigesetzt werden, beim leitenden N-MOS stammen sie aber vor allem aus den benachbarten n-dotierten Anschlüssen, also Drain und Source.

Das Anbringen einer Spannung zwischen Drain und Source, U_{DS} führt nun zu einem Stromfluss durch diesen leitenden Kanal. Die Leitfähigkeit des Kanals lässt sich mit steigender Gate-Source-Spannung erhöhen. Der MOS-Transistor verhält sich also unter gewissen Bedingungen wie eine spannungsgesteuerte ohm'sche Verbindung zwischen Drain und Source.

Aufgabe 62: Was bedeutet eine positive Drain-Source Spannung für die Spannungsabfall zwischen Gate und Kanal in der Nähe der Drain?

Der Feldeffekt-Transistor: Der Pinch Off

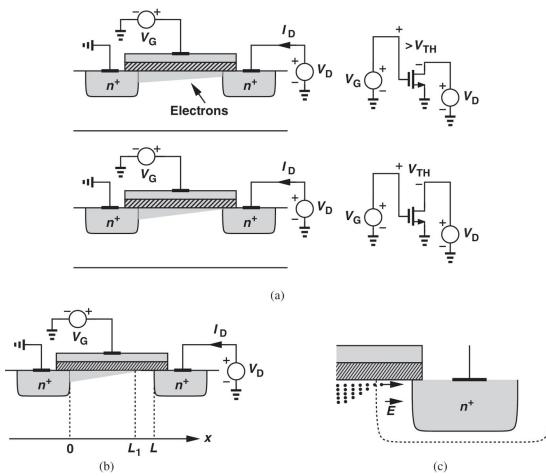
Das Ein- und Ausschalten MOSFET erfolgt über die Ausbildung des leitenden Kanals zwischen Drain und Source. Durch Anlegen einer positiven Drain-Source Spannung erwarten wir einen Driftstrom der Elektronen und daher einen linearen Zusammenhang zwischen U_{DS} und den Strom in die Drain, I_D , siehe Abbildung.



Der lineare Anstieg des Drain-Stromes I_D bei steigender Drain-Source-Spannung U_{DS} geht ab $U_{DS} \geq U_{DS} - U_{TH}$ in einen konstanten Drain-Strom über. Man spricht in diesem Fall von einem MOSFET in **Sättigung**, der Transistor verhält sich demnach wie eine reine Stromquelle. Der Grund für diesen spannungsunabhängigen Strom liegt im sogenannten **Pinch Off**. Betrachten wir dazu den Spannungsverlauf unter dem Oxid bei endlicher Drain-Source Spannung. Die Source liegt auf Masse, der Kanal in der Nähe der Masse entspricht folglich der vorhin besprochenen Situation, wobei die **Overdrive Spannung** U_{OD} ,

$$U_{OD} := U_{GS} - U_{TH}, \quad (263)$$

die Breite und daher Leitfähigkeit des Kanals bestimmt. Die Drain liegt über der Masse bei $U_{DS} > 0V$, der Kanal in der Nähe der Drain wird folglich von der Overdrive-Spannung relativ zur Drain-Spannung bestimmt, also $U_{GD} - U_{TH} = U_{GS} - U_{TH} - U_{DS}$, siehe Abbildung (a):



Sobald die Drain-Source Spannung den Wert $U_{DS} = U_{GD} - U_{TH}$ überschreitet, bildet sich kein Kanal mehr nahe der Drain, der Kanal ist folglich abgeschnürt (engl. to pinch-off) und die Drain grenzt mit einer schmalen Raumladungszone an den Kanal, siehe Abbildung (b). Der Strom durch den Kanal entspricht folglich dem Strom bei $U_{DS} = U_{GD} - U_{TH}$. Die Elektronen des Driftstromes, die die Raumladungszone erreichen, werden, analog zum Bipolar-Transistor, vom elektrischen Feld in Richtung Drain beschleunigt, Abbildung (c).

Die Kennlinie eines MOSFET lässt sich nun qualitativ wie folgt zusammenfassen:

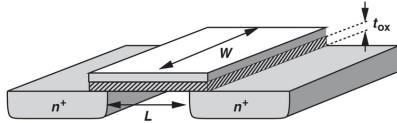
- $U_{GS} < U_{TH}$: **Transistor sperrt**, da kein leitender Kanal ausgebildet ist
- $U_{GS} > U_{TH}$:
 - * $U_{DS} \ll U_{GD} - U_{TH}$: **Linearer Bereich**, MOSFET arbeitet als Widerstand zwischen Drain und Source, wobei der Widerstandswert von U_{GS} bestimmt wird
 - * $U_{DS} > U_{GD} - U_{TH}$: **Sättigung**, MOSFET arbeitet als Stromquelle, unabhängig von U_{DS} , wobei der Stromwert von U_{GS} bestimmt wird
 - * $U_{DS} \lesssim U_{GD} - U_{TH}$: **Trioden Bereich**, weicher Übergang vom linearen Verhalten in die Sättigung

Aufgabe 63: Wodurch wird das elektrische Feld in der Raumladungszone bei Pinch-off bestimmt?

Der Feldeffekt-Transistor: Der Drain-Strom

Das Verhalten eines MOSFET ist nun qualitativ soweit verstanden, dass wir die Gleichungen für den **Drain-Strom** quantitativ ableiten können. Die **Kanalänge** L sowie die **Kanalweite** W formen zusammen mit dem Gateoxid der Dicke t_{ox} eine Kapazität zwischen Kanal und Gate der Fläche WL , siehe Abbildung.

Die Dicke des Gateoxides ergibt eine **spezifische Flächen-Kapazität** C'_{ox} . Für $U_{GS} > U_{TH}$ leitet der Kanal und die Over-drive Spannung definiert die negative Ladung im Kanal



Quelle: B. Razavi, Fundamentals of Microelectronics

$$Q = -WLC'_{ox} [U_{GS} - U_{TH}]. \quad (264)$$

Im Falle eines ortsabhängigen Potentials entlang des Kanals $U(x)$, variiert die Ladung mit der Position x , und wir erhalten die Ladungszunahme per Länge Q'

$$Q'(x) = -WC'_{ox} [U_{GS} - U_{TH} - U(x)]. \quad (265)$$

Wird nun eine positive Drain-Source Spannung angelegt, steigt die elektrische Spannung über den Kanal von $U(0) = 0V$ mit steigendem x linear bis über die Länge des Kanals bis $U(L) = U_{DS}$, siehe Abbildung.

Der **Drift Strom** der Elektronen im Kanal folgt aus der Ladung pro Länge $Q'(x)$ und der Driftgeschwindigkeit, Gleichung (41),

$$I_D = -Q'v = -Q'\mu_n \frac{d}{dx} U(x). \quad (266)$$

Hier wurde für das elektrische Feld bereits der Spannungsgradient eingesetzt, Gleichungen (16) und (18). Unter Verwendung der Elektronen-Mobilität μ_n , Gleichung (40) folgt nun eine Differentialgleichung für $U(x)$:

$$I_D = WC'_{ox} [U_{GS} - U_{TH} - U(x)] \mu_n \frac{d}{dx} U(x) \quad (267)$$

Quelle: B. Razavi, Fundamentals of Microelectronics

Durch Separation der Variablen kann der Strom über die Kanallänge integriert werden (linearer Bereich),

$$\int_0^L I_D dx = \mu_n C'_{ox} \int_0^{U_{DS}} [U_{GS} - U_{TH} - U] dU. \quad (268)$$

Der Drain-Strom bleibt über den leitenden Kanal konstant. Daher ist die linke Seite von Gleichung (268) $I_D L$ und wir erhalten den **Drain-Strom im linearen/Trioden Bereich**, d.h. $U_{DS} \leq U_{GS} - U_{TH}$

$$I_D^{\text{lin}} = \mu_n C'_{ox} \frac{W}{L} \left[(U_{GS} - U_{TH}) U_{DS} - \frac{1}{2} U_{DS}^2 \right]. \quad (269)$$

Ab einer Drain-Source Spannung von $U_{DS} = U_{GS} - U_{TH}$ wird der Kanal abgeschnürt (Pinch-Off) und der **Drain-Strom in der Sättigung**, d.h. $U_{DS} \geq U_{GS} - U_{TH}$ wird unabhängig von U_{DS} . Der Drain-Strom in der Sättigung entspricht gerade dem Maximalstrom des linearen Bereiches, also I_D^{lin} bei $U_{DS} = U_{GS} - U_{TH}$

$$I_D^{\text{sat}} = \frac{1}{2} \mu_n C'_{ox} \frac{W}{L} (U_{GS} - U_{TH})^2. \quad (270)$$

Die Gleichungen des MOSFETs im Trioden Bereich, Gleichung 269, und in der Sättigung, Gleichung 270, entsprechen dem einfachsten Spice-Modell in numerischen Simulationsprogrammen. Aufgrund der Einfachheit werden diese Gleichungen oft zur analytischen Berechnung und Dimensionierung von MOS-Schaltungen verwendet. Die Feinjustierung erfolgt dann numerisch (z.B. mit detaillierten Spice Modellen).

Aufgabe 64: Für eine Bandgap-Referenz werden zwei identische Stromquellen benötigt, welche möglichst prozessunabhängig und einstellbar den identischen Strom produzieren. Wie kann das mit 2 MOSFETs realisiert werden. Wie kann man proportionale Stromquellen mit fixem Stromverhältnis realisieren?

Der Feldeffekt-Transistor: Die MOSFET-Kennlinie

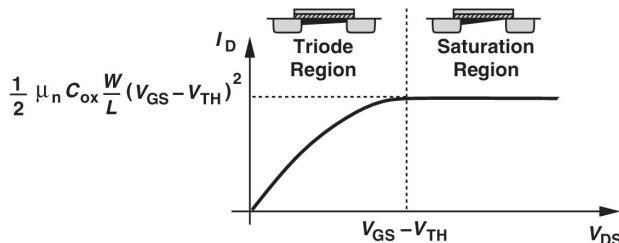
Der **Drain-Strom** I_D des MOSFET als Funktion der **Gate-Spannung** U_{GS} und der **Drain-Source-Spannung** U_{DS} kann nun wie folgt zusammengefasst werden:

$$I_D = \frac{1}{2} \mu_n C'_ox \frac{W}{L} \begin{cases} 0 & U_{GS} < U_{TH} \\ [2(U_{GS} - U_{TH})U_{DS} - U_{DS}^2] & U_{GS} > U_{TH} \& U_{DS} \leq U_{GS} - U_{TH} \\ (U_{GS} - U_{TH})^2 & U_{GS} > U_{TH} \& U_{DS} \geq U_{GS} - U_{TH} \end{cases}$$

Off-Zustand
Linear/Triode
Sättigung

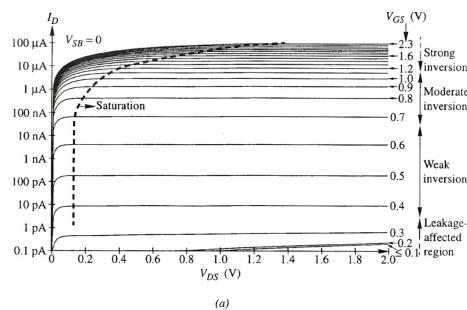
Die **MOSFET-Kennlinie** ist in der folgenden Abbildung dargestellt. Ein leitender MOSFET ($U_{GS} > U_{TH}$) verhält sich für kleine Drain-Source-Spannungen wie ein (Gate-)spannungsgesteuerter Widerstand. Diese Bereich der Kennlinie wird linearer Bereich genannt und der Widerstand $R_{on} = 1/(\partial I_D / \partial U_{DS})$ bei $U_{DS} \ll U_{GS} - U_{TH}$ lautet

$$R_{on} = \frac{1}{\mu_n C'_ox \frac{W}{L} (V_{GS} - V_{TH})}. \quad (271)$$

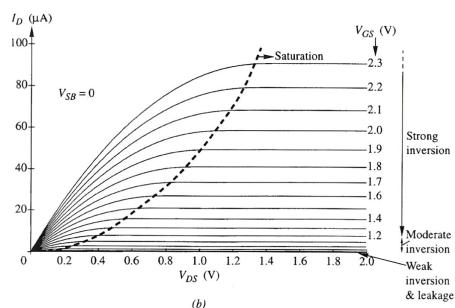


Quelle: B. Razavi, Fundamentals of Microelectronics

Das **Kennlinie-Feld** einer NMOS-Transistors für unterschiedliche Gate-Spannungen als Funktion der Drain-Source Spannung ist hier nochmals in logarithmischer und linearer Darstellung angeführt:



- Bei $U_{DS} \leq U_{GS} - U_{TH}$ arbeitet der MOS im **Trioden** oder **Linearen Bereich** mit näherungsweise konstantem on-Widerstand, siehe linearer Anstieg in Abbildung (b).



- Mit steigender Gate-Spannung vergrößert sich die Inversion, d.h. die Dicke des leitenden Kanals, und reduziert sich folglich der on-Widerstand
- Bei $V_{DS} > V_{GS} - V_{TH}$ sättigt der Drain-Strom und der Transistor verhält sich in **Saturation** wie eine Konstantstromquelle, mit sehr hohem Innenwiderstand
- Mit steigender Gate-Source-Spannung vergrößert sich der Sättigungs-Strom

Quelle: Y. Tsividis, The MOS Transistor

Reale MOSFETs zeigen leicht abweichendes Verhalten, welches im Folgenden erklärt wird.

Aufgabe 65: Die Grenze zwischen linearem Bereich und Sättigung ist abhängig von der Gate-Source-Spannung (strichlierte Linie in der obigen Abbildung). Wie lautet die Grenzspannung als Funktion des Sättigungs-Stroms?

Effekte höherer Ordnung: Die Kanallängen-Modulation

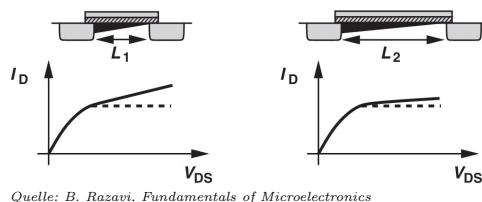
In der Herleitung der Drain-Spannung in der Sättigung wurde der Maximal-Strom des Triodenbereiches als Wert für einen konstanten Strom in der Sättigung genommen. Diese Annahme steht eigentlich im Widerspruch zum Pinch Off. Mit steigender Drain-Source-Spannung U_{DS} verbreitert sich die Verarmungszone vor der Drain. Folglich reduziert sich die Breite der effektiven Kanales auf $L_{\text{eff}} < L$. Diese effektive Kanallänge ist eine schwach monoton fallende Funktion der Drain-Source-Spannung,

$$L_{\text{eff}}(U_{DS}) = L - \Delta L(U_{DS}) \approx L(1 - \lambda U_{DS}), \quad (272)$$

wobei der Modell-Parameter λ die sogenannte **Kanallängen-Modulation** beschreibt. Wird nun im Sättigungsstrom, Gleichung (270), die Kanallänge durch L_{eff} ersetzt, folgt in der ersten Näherung Taylor-Reihe:

$$I_D = \frac{1}{2} \mu_n C'_{ox} \frac{W}{L} (U_{GS} - U_{TH})^2 (1 + \lambda U_{DS}) \quad (273)$$

Der Parameter λ entspricht der inversen Early-Spannung des Bipolartransistors. Als Folge der **Kanallängen-Modulation** steigt der Sättigungsstrom mit steigender Drain-Source Spannung linear an, siehe Abbildung.



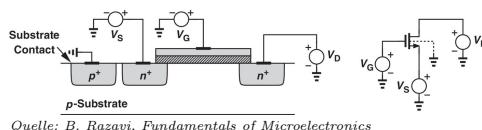
Quelle: B. Razavi, Fundamentals of Microelectronics

Die Breite der Raumladungszone hängt nur von der Spannung ab, $\Delta L(U_{DS})$, der prozentuale Einfluß auf die effektive Kanallänge ist bei Transistoren mit kleiner Länge aber dramatischer. Folglich zeigt die Kennline von Transistoren kleiner Länge, L_1 in der Abbildung, eine stärkere Abhängigkeit von U_{DS} als die Kennlinie großer Länge, L_2 .

Die Kanallängen-Modulation ist bei MOSFETs, die als Schalter verwendet werden, z.B. in **Digitalenschaltungen** zumeist **irrelevant** und es können minimale Kanallängen (z.B. $L = 90\text{nm}$) verwendet werden. Bei MOSFETs als Stromquellen, z.B. in **Anologschaltungen**, beeinträchtigt die Kanallängen-Modulation das Verhalten und es werden Kanallängen von typischerweise $L > 1\mu\text{m}$ gewählt.

Der Body Effekt

Bisher waren Source und Substrat des NMOS immer auf gleichem Potential. Das Substrat liegt immer auf Masse, die Source eines NMOS kann auch auf einen höheren Potential $U_S > 0\text{V}$ liegen, siehe Abbildung



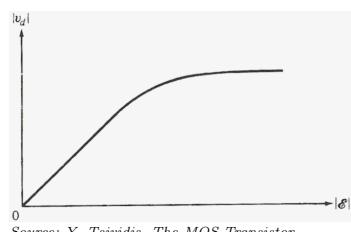
Quelle: B. Razavi, Fundamentals of Microelectronics

Für die Ausbildung des Kanals wirkt nun die Potentialdifferenz zwischen Source und Substrat (oder **Body**), U_S , wie ein **zweites Gate**, welches von unten die Wirkung des Gate-Source Potentials reduziert. Der Transistor mit **Body Effekt** wird erst bei höherem U_{GS} leitend, entsprechend einer vergrößerten Thresholdspannung,

$$U_{TH}(U_S) > U_{TH}(0\text{V}). \quad (274)$$

Die Geschwindigkeits-Sättigung

Bisher wurde ein linearer Zusammenhang zwischen Elektrischem Feld und Driftgeschwindigkeit, Gleichung (196), angenommen. Bei sehr hohen elektrischen Feldern, d.h. Kleinkanaltransistoren ($L \lesssim 100\text{nm}$), erreicht die Driftgeschwindigkeit einen Maximalwert v_{sat} , man spricht von **Geschwindigkeits-Sättigung**:



Source: Y. Tsividis, The MOS Transistor

Als Folge sättigt der **Drain Strom** und wird unabhängig von U_{DS} :

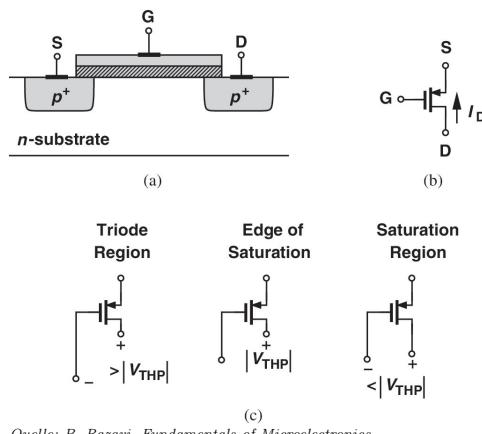
$$I_D = v_{\text{sat}} Q' = v_{\text{sat}} W C'_{ox} (V_{GS} - V_{TH}). \quad (275)$$

Bei Kleinkanaltransistoren dominiert die Geschwindigkeits-Sättigung das Verhalten in der Sättigung. Der **Sättigungsstrom** I_D steigt daher nur noch **linear** mit der Overdrive-Spannung, d.h. $I_D \propto (V_{GS} - V_{TH})$.

Aufgabe 66: Mit steigender Source-Spannung U_S kann die Thresholdspannung über den Body Effekt vergrößert werden. Warum ist eine Reduktion der Thresholdspannung mit $U_S < 0\text{V}$ nicht möglich?

Der Feldeffekt-Transistor: Der P-MOS Transistor

Analog zum Bipolar-Transistor kann auch bei den MOSFETs ein komplementärer FET zum NMOS gebildet werden. Bei diesem sogenannten **P-MOS**-Transistor werden auf einem Substrat vom n-Typ zwei stark p-dotierte Regionen für die Drain und Source aufgebracht (p^+ in der Abbildung). Der Kanal ist wieder über eine Oxidschicht vom Gate getrennt, siehe Abbildung a. Die Source wird, analog zum PNP-Transistor, durch einen Pfeil in Stromrichtung markiert:



Quelle: B. Razavi, Fundamentals of Microelectronics

In Analogie zum NMOS drücken wir den Drainstrom I_D nun als Funktion der Gate-Source Spannung U_{GS} und der Drain-Source Spannung U_{DS} aus. Die **Source** ist (im einfachsten Fall) auf dem Potential des Substrates, und daher auf den **höchsten Potential**, Drain und Gate können also nur gleich oder unter dem Potential der Source liegen. Als Folge fliesst ein negativer Drainstrom, siehe Abbildung (b). Für die Ausbildung eines Kanals muß nun das Potential des Gates eine Thresholdspannung unter das Potential der Source (und des Substrates) gebracht werden und es entsteht ein leitender Kanal von Löchern, der bei entsprechend niedrigem Drain-Potential in die Sättigung geht, Abbildung (c).

Die Gleichungen des PMOS-Transistors sind in Analogie zum NMOS gegeben, wobei lediglich die Mobilität der Löcher μ_p und die nun negative Thresholdspannung $U_{THp} < 0V$ verwendet werden.

$$I_D = -\frac{1}{2} \mu_p C'_{ox} \frac{W}{L} \begin{cases} 0 & U_{GS} > U_{THp} \\ \frac{1}{2} [2(U_{GS} - U_{THp})U_{DS} - U_{DS}^2] & U_{GS} < U_{THp} \& U_{DS} \geq U_{GS} - U_{THp} \\ \frac{1}{2} (U_{GS} - U_{THp})^2 & U_{GS} < U_{THp} \& U_{DS} \leq U_{GS} - U_{THp} \end{cases}$$

Man beachte, dass alle Spannungen und auch der Drainstrom nun negativ sind, was leicht zu Verwirrung führen kann. In der Praxis ist es daher oft einfacher, nur die Absolutbeträge der Spannungen und Ströme zu betrachten, wodurch die PMOS-Gleichungen folgende Form annehmen:

$$|I_D| = \frac{1}{2} \mu_p C'_{ox} \frac{W}{L} \begin{cases} 0 & |U_{GS}| < |U_{THp}| \\ \frac{1}{2} [2(|U_{GS}| - |U_{THp}|)|U_{DS}| - U_{DS}^2] & |U_{GS}| > |U_{THp}| \& |U_{DS}| \leq |U_{GS}| - |U_{THp}| \\ \frac{1}{2} (|U_{GS}| - |U_{THp}|)^2 & |U_{GS}| > |U_{THp}| \& |U_{DS}| \geq |U_{GS}| - |U_{THp}| \end{cases}$$

Der dominante Leitungsmechanismus des PMOS ist der **Driftstrom** der **Löcher** im Kanal, ausgedrückt durch die **Löchermobilität** μ_p . In den meisten Technologien ist die Mobilität der Löcher deutlich geringer als die Elektronenmobilität, wodurch PMOS Transistoren bei gleicher Größe deutlich schlechter leiten. Im Falle von Si-Prozessen ist der Unterschied in etwa ein Faktor 3. Daher werden PMOS-Transistoren typischerweise mit ca. 3-facher Kanalweite W eines analogen NMOS gewählt.

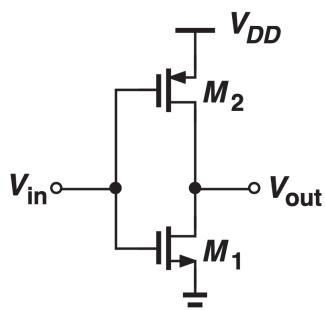
Liegt die Source Spannung unter dem n-Substrat, führt analog zum NMOS, der Bodyeffekt zu einem Anstieg der Thresholdspannung $|U_{THp}|$. Bei Kleinkanaltransistoren sind wieder die Kanallängenmodulation und die Geschwindigkeits-Sättigung zu beachten.

- **Body Effekt:** Der Absolutbetrag der Thersholtspannung wird vergrössert
- **Kanallängen-Modulation:** Der Sättigungsstrom bekommt einen zusätzlichen Faktor $(1 + \lambda|U_{DS}|)$
- **Geschwindigkeits-Sättigung:** Der Sättigungsstrom reduziert sich auf eine lineare Abhängigkeit

Aufgabe 67: Bei welchen Anwendung ist der PMOS-Transistor einem NMOS vorzuziehen?

CMOS Technologie

Werden auf einem Siliziumscheibe, einem sogenannten **Wafer**, gleichzeitig sowohl NMOS wie PMOS Transistoren gefertigt spricht man von einer **komplementären MOS Technologie**, oder einfach von einem **CMOS Prozess**. Ein CMOS Prozess verbindet die Vorteile der NMOS und PMOS Transistoren, da Signale mit NMOS Transistoren einfach gegen Masse gezogen werden können, Signale mit PMOS Transistoren hingegen einfach gegen die Versorgungsspannung. Betrachten wir als einfachste digitale Schaltung den sogenannten **Inverter**, siehe Abbildung, bei dem der logische Zustand des Ausgangs gerade der negierte logische Zustand des Einganges ist, $V_{out} = \text{NOT}(V_{in})$.

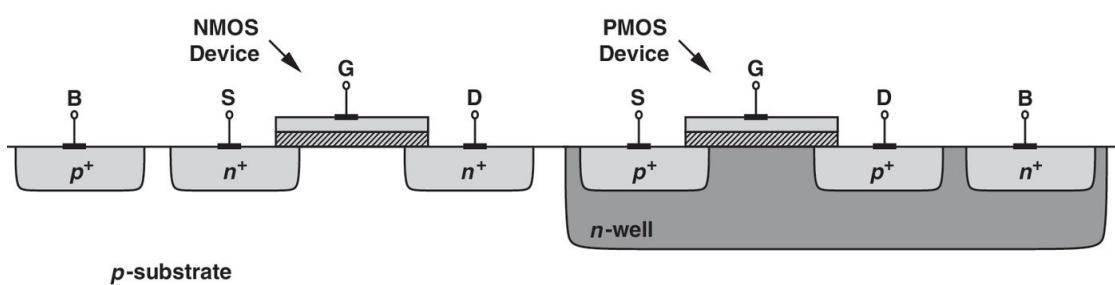


Quelle: B. Razavi, Fundamentals of Microelectronics

Wird die logische 0 durch eine Spannung 0V repräsentiert und eine logische 1 durch die Versorgungsspannung V_{DD} , so bewirkt eine logische 0 am Eingang eine Gate-Source des NMOS (M1) von 0V, d.h. M1 sperrt und entspricht einem sehr hochohmigen Widerstand. Die Source-Gate Spannung des PMOS (M2) ist hingegen gleich der Versorgungsspannung, d.h. der Kanal des PMOS ist leitend. Der Ausgang wird daher vom PMOS in Richtung der logischen 1 gezogen. Analog bewirkt eine logische 1 am Eingang, d.h. $V_{in} = V_{DD}$ einen leitenden NMOS-Kanal und einen sperrenden PMOS, d.h. eine logische 0, d.h. $V_{out} = 0V$ am Ausgang.

In digitalen CMOS-Schaltungen wird der logische Zustand immer durch einen leitenden Kanal in die eine Richtung und einen komplementären hochohmigen Kanal repräsentiert. Dadurch fließt im statischen Zustand kein Strom (im Gegensatz zu Bipolartransistoren, welche im leitenden Zustand immer einen Basis-Emitter-Strom benötigen) und es wird im statischen Zustand (abgesehen von Leckströmen) keine Wärme dissipiert. Ströme werden lediglich für die Änderung der logischen Zustände, d.h. für das Umladen der Gate-Kapazität, benötigt.

Wie kann nun auf dem gleichen Wafer sowohl ein NMOS (mit einem p-dotierten Substrat) und ein PMOS (mit einem n-dotierten Substrat) gefertigt werden?



Quelle: B. Razavi, Fundamentals of Microelectronics

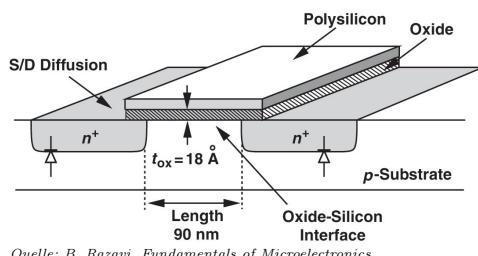
- Typische Si Wafer sind aus n-dotiertem Si
- Für PMOS Transistoren wird lokal (z.B. durch Diffusion) eine n-dotierte Region erzeugt, eine sogenannte **n-Wanne** oder **n-Well**
- Die n-Wanne wird typischerweise an die Versorgungsspannung angeschlossen, wodurch den PN-Übergang zum p-dotierten Wafer immer sperrt
- In der n-Wanne werden die PMOS-Transistoren angeordnet

Die meisten modernen integrierten Schaltungen werden in CMOS-Technologie realisiert.

Aufgabe 68: Wie kann bei PMOS-Transistoren in CMOS Technologie der Body Effekt vermieden werden?

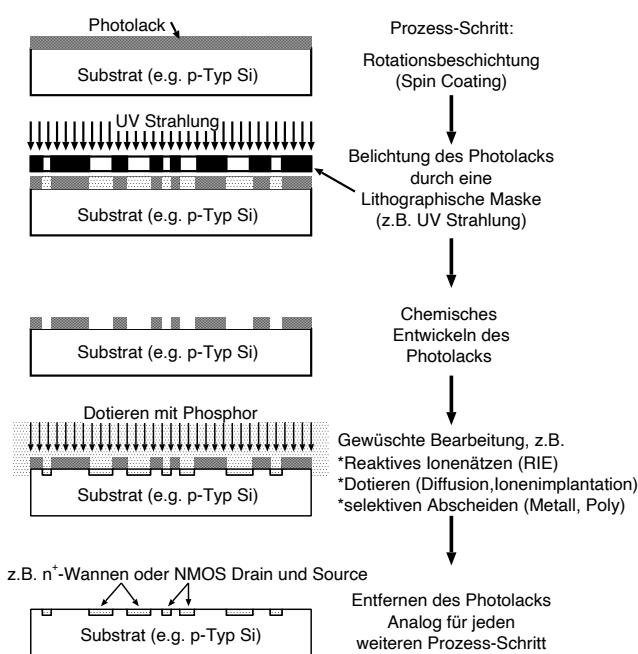
8 – Halbleiterfertigung

Die Methode zur Fertigung von Halbleiterbauelementen wird **Lithographie** genannt (von altgriechisch *lithos* „Stein“ und *graphein* „schreiben“). Betrachten wir zuerst die geometrischen Ausmaße eines typischen modernen CMOS-Transistors mit minimaler Kanallänge von 90nm , siehe Abbildung.



Auf einer *Si*-Scheibe (einem Wafer) mit Durchmessern bis 30cm (Stand 2024) werden typischerweise viele hundert oder mehr Chips gleichzeitig gefertigt. Wenn man bedenkt, dass jeder dieser Chips viele Millionen Transistoren beinhaltet, welche alle ohne Fehlverhalten gefertigt werden müssen ($> 95\%$ Ausbeute pro Wafer als Ziel) und minimale Geometrien von wenigen 10nm sowie Schichtdicken von wenigen Atomdurchmessern mit gleichbleibender Qualität erforderlich sind, wird klar, dass die Chipfertigung höchste Anforderungen an die Reinheit der Materialien (*Si*-Wafer sind Einkristalle mit einer Reinheit von 99.99999999% und mehr) und der Prozesse stellt. Ein hohes Maß an Reproduzierbarkeit und Parallelisierung in der Fertigung wird durch die **Photolithographie** erreicht, die für einen Prozess-Schritt im Folgenden kurz skizziert wird.

Photolack (ein durch Belichtung chemisch und physikalisch veränderliches Material) wird gleichmäßig über den Wafer verteilt. Für jeden Prozessschritt wurde eine wiederverwendbare Maske (also ein UV-undurchsichtiges Material, mit den gewünschten Bereichen als Ausnehmungen geschrieben. Der Photolack wird über diese Maske mit UV-Licht bestrahlt, wodurch der Photolack an den Ausnehmungen bestrahlt und dadurch verändert wird (entweder verhärtet oder, wie dargestellt, löslich wird). Die gewünschten Strukturen sind in den chemisch oder physikalisch veränderten Bereichen des Photolacks abgebildet, und können chemisch entfernt werden, d.h. der Wafer liegt an den zu fertigenden Strukturen frei.



Für digitale Anwendungen werden typischerweise Transistoren minimaler Kanallänge L verwendet, in der Abbildung 90nm (es gibt schon Technologien mit $< 20\text{nm}$). Die Breite der Transistoren richtet sich nach dem benötigten Treiberstrom um das folgende Gatter zu schalten. Die Dicke des Oxsids t_{OX} liegt im Bereich weniger Nanometer, in der Abbildung sind es 1.8nm , d.h. wenige Atomdurchmesser!

Nun erfolgt der eigentliche Fertigungsschritt, indem der Wafer (je nach Prozess-Schritt) an den offenliegenden Stellen entweder geätzt (z.B. für Durchstiche) oder dotiert wird (z.B. durch Diffusion geeigneter Dotieratome in den Wafer oder durch Einschießen via Ionen-Implantation). Es kann aber auch z.B. für die Oxidbereiche, die Gates oder Metallbahnen entsprechendes Material abgeschieden werden. In einem letzten Prozess-Schritt wird der restliche Photolack chemisch entfernt und der entsprechende Fertigungsschritt ist abgeschlossen - d.h. der Wafer hat an den Ausnehmungen der Photomaske z.B. ein geätztes Loch, oder einen dotierten Bereich oder aber ein abgeschiedenes Material (also ein Oxid, Gate oder eine Leiterbahn). Der Ablauf wird für jeden Prozess-Schritt wiederholt

Aufgabe 68: Wie groß darf die Wahrscheinlichkeit für einen falsch gefertigten Transistor sein, wenn ein Wafer mit 100 Chips von jeweils 100 Millionen Transistoren eine Ausbeute von 97% erzielen soll?