

# 10 Halbleiter

Unter einem Halbleiter verstehen wir einen Festkörper, den wir hinsichtlich seiner elektrischen Leitfähigkeit sowohl als Leiter als auch als Nichtleiter betrachten können. Aufgrund der großen Variationsbreite ihrer elektrischen Eigenschaften und der Tatsache, dass ihre elektrische Leitfähigkeit von außen über elektrische Felder gesteuert werden kann, haben Halbleiter in unserer heutigen Elektronik eine enorme Bedeutung erlangt.

In Abschnitt 8.4 und 8.5 haben wir bereits diskutiert, dass wir Festkörper entsprechend der vorliegenden Bandstruktur in Metalle, Halbmetalle, Halbleiter und Isolatoren unterteilen können. Wir haben gelernt, dass vollkommen gefüllte und leere Bänder nicht zur elektrischen Leitfähigkeit beitragen und deshalb Materialien, die nur vollkommen gefüllte und leere Bänder besitzen, Isolatoren bzw. Halbleiter sind. Dabei haben wir gesehen, dass der Unterschied zwischen Isolatoren und Halbleitern nur ein quantitativer aber kein qualitativer ist. Ist die Bandlücke  $E_g$  zwischen dem obersten vollkommen gefüllten (*Valenzband*) und dem untersten vollkommen leeren Band (*Leitungsband*) nicht allzu groß, so sind bei Raumtemperatur aufgrund der Verschmierung der Fermi-Funktion einige Zustände im Leitungsband besetzt und einige Zustände im Valenzband leer. Wie wir später sehen werden, nimmt der Anteil der thermisch angeregten Ladungsträger proportional zu  $\exp(-E_g/2k_B T)$  zu und beträgt damit bei Raumtemperatur ( $k_B T \approx 25 \text{ meV}$ ) und einer Energielücke von  $E_g \approx 0.5 \text{ eV}$  etwa  $e^{-10} \approx 5 \times 10^{-5}$ . Sowohl die thermisch angeregten Elektronen im Leitungsband als auch die Löcher im Valenzband führen zu einer endlichen Leitfähigkeit und wir sprechen von einem Halbleiter. Ist die Bandlücke dagegen groß, so ist die Zahl der thermischen angeregten Ladungsträger und damit die elektrische Leitfähigkeit selbst bei Raumtemperatur verschwindend klein und wir sprechen von einem Isolator. Bei  $E_g = 3 \text{ eV}$  ist der Anteil der thermisch angeregten Ladungsträger bei Raumtemperatur nur noch etwa  $e^{-58} \approx 5 \times 10^{-26}$ .



Ist die Bandlücke  $E_g$  zwischen dem obersten vollkommen gefüllten (*Valenzband*) und dem untersten vollkommen leeren Band (*Leitungsband*) nicht allzu groß, so sind bei Raumtemperatur aufgrund der Verschmierung der Fermi-Funktion einige Zustände im Leitungsband besetzt und einige Zustände im Valenzband leer. Wie wir später sehen werden, nimmt der Anteil der thermisch angeregten Ladungsträger proportional zu  $\exp(-E_g/2k_B T)$  zu und beträgt damit bei Raumtemperatur ( $k_B T \approx 25 \text{ meV}$ ) und einer Energielücke von  $E_g \approx 0.5 \text{ eV}$  etwa  $e^{-10} \approx 5 \times 10^{-5}$ . Sowohl die thermisch angeregten Elektronen im Leitungsband als auch die Löcher im Valenzband führen zu einer endlichen Leitfähigkeit und wir sprechen von einem Halbleiter. Ist die Bandlücke dagegen groß, so ist die Zahl der thermischen angeregten Ladungsträger und damit die elektrische Leitfähigkeit selbst bei Raumtemperatur verschwindend klein und wir sprechen von einem Isolator. Bei  $E_g = 3 \text{ eV}$  ist der Anteil der thermisch angeregten Ladungsträger bei Raumtemperatur nur noch etwa  $e^{-58} \approx 5 \times 10^{-26}$ .

Der Übergang zwischen Halbleitern und Isolatoren ist fließend und nicht genau festgelegt. In den meisten Lehrbüchern werden üblicherweise diejenigen Materialien als Halbleiter bezeichnet, die bei Raumtemperatur einen spezifischen Widerstand im Bereich zwischen  $10^{-2}$  und  $10^9 \Omega \text{cm}$  haben. Materialien mit einem höheren spezifischen Widerstand werden als Isolatoren bezeichnet. Da die Ladungsträgerdichte und damit die elektrische Leitfähigkeit

mit sinkender Temperatur etwa proportional zu  $\exp(-E_g/2k_B T)$  abnimmt, sind bei einer Klassifizierung hinsichtlich des elektrischen Widerstands bei genügend tiefen Temperaturen alle Halbleiter natürlich Isolatoren. Eine vernünftige Klassifizierung ist z. B. diejenige, als Halbleiter alle Materialien mit einer endlichen Energielücke zu bezeichnen, die unterhalb ihrer Schmelztemperatur noch eine beobachtbare elektrische Leitfähigkeit besitzen. Die Tatsache, dass die elektrische Leitfähigkeit von Halbleitern mit der Temperatur abnimmt unterscheidet sie fundamental von Metallen, deren elektrische Leitfähigkeit mit sinkender Temperatur zunimmt (vergleiche Kapitel 7 und 9).

Der exponentielle Zusammenhang zwischen Ladungsträgerdichte und Bandlücke  $E_g$  gilt nur für so genannte *intrinsische Halbleiter*, bei denen freie Ladungsträger nur durch Anregung aus dem vollen Valenzband ins leere Leitungsband erzeugt werden können. In diesem Fall könnten wir alternativ Materialien mit einer Energielücke kleiner als etwa 3 eV als Halbleiter bezeichnen. Eine herausragende Eigenschaft von Halbleitern, die diese Materialklasse gegenüber Metallen auszeichnet, ist aber die Möglichkeit, ihre Ladungsträgerdichte und damit ihre elektrische Leitfähigkeit durch Verunreinigung mit kleinsten Mengen von Fremdatomen, man spricht hier von *Dotierung*, über mehrere Größenordnungen zu ändern. Durch die Wahl der Fremdatome kann ferner festgelegt werden, ob die erzielte Leitfähigkeit elektron- oder lochartig ist. Die meisten Halbleiterbauelemente basieren auf dieser spezifischen Eigenschaft von Halbleitern. In diesem Zusammenhang könnten wir auch alle Isolatoren, deren spezifischer elektrischer Widerstand sich durch Dotierung in den oben genannten Bereich zwischen etwa  $10^{-2}$  und  $10^9 \Omega\text{cm}$  bringen lässt, als Halbleiter bezeichnen. Dazu gehört z. B. auch Diamant, den wir ohne Dotierung, also als intrinsisches Material, aufgrund seiner großen Energielücke von etwa 5.5 eV den Isolatoren zuordnen müssen.

Halbleiter haben heute eine enorme Bedeutung für die Informations- und Kommunikationstechnik (integrierte Schaltkreise), aber auch für die Leistungselektronik (Transistoren, Thyristoren, Triacs), die Sensorik (Hall-Sensoren, Thermistoren, Photo-Detektoren, Drucksensoren), die Beleuchtungstechnik (Leuchtdioden), die Lasertechnik (Injektionslaser) oder die Photovoltaik (Solarzellen). Wir wollen in diesem Kapitel die spezifischen Eigenschaften von Halbleitern näher diskutieren. Dabei werden wir zunächst eine Klassifizierung von Halbleitern vornehmen und die grundlegenden Eigenschaften von intrinsischen und dotierten Halbleitern wie ihre Ladungsträgerdichte und ihren elektrischen Widerstand diskutieren. Anschließend werden wir uns mit räumlich inhomogenen Halbleitern und ihrer Anwendung in elektronischen Bauelementen beschäftigen. Zum Abschluss des Kapitels über Halbleiter werden wir niedrigdimensionale Elektronengase diskutieren, die mit Halbleitersystemen einfach realisiert werden können und in der heutigen Grundlagenforschung von großer Bedeutung sind. An einem mit einer Halbleiterstruktur realisierten zweidimensionalen Elektronengas wurde zum Beispiel der Quanten-Hall-Effekt entdeckt.

# 10.1 Grundlegende Eigenschaften von Halbleitern

## 10.1.1 Klassifizierung von Halbleitern

Unabhängig von der chemischen Zusammensetzung unterscheiden wir zwischen

- **intrinsischen und dotierten Halbleitern:**

Bei intrinsischen Halbleitern können wir freie Ladungsträger im Leitungsband nur durch Anregung aus dem Valenzband erzeugen. Bei dotierten Halbleitern ist dies nicht mehr der Fall. Wir werden sehen, dass durch das Einbringen von Fremdatomen in Halbleiter elektronische Niveaus in der Bandlücke erzeugt werden und wir dadurch freie Ladungsträger auch durch Anregung aus diesen Niveaus erzeugen können.

- **direkten und indirekten Halbleitern:**

Bei direkten Halbleitern (z. B. GaAs) liegt die Oberkante des Valenzbandes und die Unterkante des Leitungsbandes bei demselben Wellenvektor (z. B. am  $\Gamma$ -Punkt der Brillouin-Zone). Bei indirekten Halbleitern ist dies nicht der Fall. Zum Beispiel liegt bei Ge die Oberkante des Valenzbandes am  $\Gamma$ -Punkt, während die Unterkante des Leitungsbandes am  $L$ -Punkt liegt.

- **kristallinen und amorphen Halbleitern:**

Bezüglich ihrer kristallinen Ordnung können wir zwischen *kristallinen* und *amorphen Halbleitern* unterscheiden. Wir werden uns im Folgenden hauptsächlich auf die Eigenschaften von kristallinen Halbleitern konzentrieren.

In amorphem Silizium (a-Si) bestehen etwa ab der vierten Bindungslänge keinerlei Korrelationen in Abstand und Orientierung der Si-Atome mehr. Dadurch entstehen viele nicht abgesättigte Bindungen (engl. dangling bonds), die durch Wasserstoffatome gesättigt werden können. Man spricht dann von hydrogenisiertem amorphem Silizium (a-Si:H). Amorphes Silizium verfügt über ein hohes Absorptionsvermögen und kann daher bei Solarzellen mit besonders geringen Schichtdicken verwendet werden. Die üblichen Schichtdicken sind dabei etwa um einen Faktor 100 kleiner als bei kristallinem Silizium. Dies gleicht den durch die Defekte geringen Wirkungsgrad von nur etwa 6 bis 8% aus und macht a-Si für Anwendungen in der Photovoltaikindustrie wirtschaftlich interessant.

Bezüglich ihrer chemischen Zusammensetzung und ihrer Funktionalität können wir weitere Grundtypen von Halbleitern unterscheiden, die wir im Folgenden kurz vorstellen. Eine Zusammenstellung der wichtigsten Halbleitertypen ist in Tabelle 10.1 gezeigt.

### 10.1.1.1 Element-Halbleiter

Der zweifelsfrei am besten bekannte und für unser Alltagsleben sehr wichtige Halbleiter ist der Element-Halbleiter Silizium. Ohne den Halbleiter Si wäre unsere heutige Informations- und Kommunikationstechnologie undenkbar. Si kristallisiert wie andere Element-Halbleiter der IV. Hauptgruppe (C, Ge und  $\alpha$ -Sn) in der Diamantstruktur (siehe Abb. 1.24). In dieser Struktur ist jedes Atom von 4 nächsten Nachbaratomen umgeben, die einen Tetraeder bilden. Diamant ist wegen seiner großen Energielücke von etwa 5.5 eV als intrinsisches Material

ein guter Isolator. Seine Leitfähigkeit kann aber durch Einbringen von Dotieratomen in den Bereich von Halbleitern erhöht werden. Einige Elemente der V. und VI. Hauptgruppe des Periodensystems (z. B. P, S, Se, Te) sind ebenfalls Halbleiter. Die Atome in den jeweiligen Kristallstrukturen sind allerdings nur dreifach (P) oder zweifach (S, Se, Te) koordiniert.

Elementhalbleiter können heute in sehr großer Reinheit hergestellt werden. Die Herstellung von hochreinem Silizium mit Hilfe des *Zonenschmelzverfahrens*<sup>1</sup> wurde 1954 von einem Team um **Eberhard Spenke**<sup>2</sup> bei der Siemens & Halske AG ermöglicht. Dies brachte Mitte der 1950er Jahre den Durchbruch von Silizium als Halbleitermaterial für die Elektronikindustrie und in den 1980er Jahren auch für die ersten Produkte der Mikrosystemtechnik. Für die Herstellung von integrierten Schaltkreisen wird aber heute aus Kostengründen fast ausschließlich mit dem *Czochralski-Verfahren*<sup>3,4</sup> hergestelltes Silizium verwendet.

### 10.1.1.2 Verbindungshalbleiter

Die wichtigsten Vertreter der Verbindungshalbleiter sind die III-V-Halbleiter (z. B. GaAs, GaN, InP, InSb, GaSb, AlSb), die aus Elementen der III. und V. Hauptgruppe bestehen, und die II-VI-Halbleiter (z. B. ZnS, CdS, CdSe, CdTe, HgTe), die aus Elementen der II. und VI. Hauptgruppe aufgebaut sind. Die Eigenschaften der III-V-Halbleiter sind sehr

<sup>1</sup> Beim Zonenschmelzverfahren wird ein mit Hilfe einer Induktionsheizung erzeugter, aufgeschmolzener Bereich durch einen Stab mit noch polykristalliner Kristallstruktur bewegt. Damit die Zone gleichmäßig aufschmilzt, rotiert der Stab langsam. Die aufgeschmolzene Zone wird anfangs mit einem Impfkristall in Berührung gebracht und wächst unter Annahme seiner Kristallstruktur an ihm an. Diese Schmelzzone wird dann langsam durch den Stab bewegt, wobei die wieder erkalten- de Schmelze mit einer einheitlichen Kristallstruktur über die gesamte Materialbreite erstarrt und somit hinter der Schmelzzone den gewünschten Einkristall bildet. Das Zonenschmelzverfahren beruht auf der Tatsache, dass Verunreinigungen in der Schmelze eine energetisch günstigere chemische Umgebung (niedrigeres chemisches Potenzial) haben als im Festkörper und darum vom Festkörper in die Schmelze wandern. Die Menge, die im Kristall eingebaut wird, hängt u. a. von der Art der Verunreinigung und der Erstarrungsgeschwindigkeit ab. Fremdatome verbleiben also weitestgehend in der Schmelzzone und lagern sich schließlich am Ende der Säule an, die nach dem Erkalten entfernt wird. Durch mehrmaliges Zonenschmelzen kann die Reinheit weiter gesteigert werden.

<sup>2</sup> **Eberhard Spenke**, geboren am 5. Dezember 1905 in Bautzen, gestorben am 24. November 1992 in Pretzfeld. Er studierte an den Universitäten Bonn, Göttingen und Königsberg Physik, die Promotion erfolgte im Jahr 1929. Er war dann von 1929 bis 1946 als wissenschaftlicher Mitarbeiter im Berliner Zentrallaboratorium der Siemens & Halske AG tätig. Zusammen mit Walter Schottky (1886–1976) untersuchte er dort die Eigenschaften und Leitungsvorgänge von Halbleitermaterialien.

<sup>3</sup> Das Czochralski-Verfahren ist auch unter dem Begriff Tiegelziehverfahren bekannt. Im Tiegel wird die zu kristallisierende Substanz wenig über dem Schmelzpunkt gehalten. Darin taucht der Keim, z. B. ein kleiner Einkristall der zu züchtenden Substanz ein. Durch Drehen und langsames nach oben Ziehen wächst das erstarrende Material zu einem Einkristall, der das Kristallgitter des Keims fortsetzt. Das Czochralski-Verfahren wurde 1916 im Metall-Labor der AEG vom polnischen Chemiker Jan Czochralski (1885–1953, 1904–1929 in Deutschland) durch ein Versehen entdeckt: er tauchte seine Schreibfeder in einen Schmelztiegel mit flüssigem Zinn anstatt ins Tintenfass. Daraufhin entwickelte und verbesserte er das Verfahren und wies nach, dass damit Einkristalle hergestellt werden können.

<sup>4</sup> J. Czochralski, *Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle*, Zeitschrift für Physikalische Chemie **92**, 219–221 (1918).

Tabelle 10.1: Klassifizierung von Halbleitern.

Elementhalbleiter	Verbindungshalbleiter		organische Halbleiter
Ge, Si, $\alpha$ -Sn,	III-V	GaAs, GaP, InP, InSb, InAs, GaSb, GaN, AlN, InN, $Al_xGa_{1-x}As$	Tetracen, Pentacen, Phthalocyanine, Polythiophene,
C (Diamant, Fulleren),	II-VI	ZnO, ZnS, ZnSe, ZnTe, CdS, CdSe, CdTe, HgS, $Hg_{1-x}Cd_xTe$ , BeSe, BeTe,	PTCDA ( $C_{24}H_8O_6$ ), MePT-CDI ( $C_{26}H_{14}N_2O_4$ ),
B, Se, Te	III-VI	GaS, GaSe, GaTe, InS, InSe, InTe	Chinacridon, Acridon,
	IV-VI	PbS, PbTe, SnS	Flavanthron, Perinon,
<b>unter Druck:</b>	IV-IV	SiC, SiGe	Indanthron,
Bi, Ca, Sr, Ba,	I-VII	CuCl	Alq3 ( $C_{27}H_{18}AlN_3O_3$ )
Yb, P, S, I	I-III-VI	$CuInSe_2$ , $CuInGaSe_2$ , $CuInS_2$ , $CuInGaS_2$	

ähnlich zu denjenigen der Element-Halbleiter der IV. Hauptgruppe. Durch den Übergang von Elementen der IV. Hauptgruppe zu III-V Verbindungen erhält die chemische Bindung einen endlichen ionischen Charakter, da Ladung vom Gruppe-III zum Gruppe-V Element transferiert wird. Dieser ionische Bindungsanteil führt üblicherweise zu einer Erhöhung der Energielücke. Der Anteil und die Bedeutung der ionischen Bindung wird noch größer für die II-VI-Halbleiter. Deshalb haben die meisten II-VI-Halbleiter Energielücken oberhalb von 1 eV. Eine Ausnahme bilden hier die Systeme, die Hg enthalten. Diese haben kleine Energielücken oder sind sogar Halbmetalle wie z. B. HgTe. II-VI-Halbleiter mit großen Energielücken sind interessant für Displays und Laser, diejenigen mit kleiner Energielücke finden Anwendung in Infrarotdetektoren.

Weitere Verbindungshalbleiter sind IV-IV-Halbleiter (SiC, SiGe), IV-VI-Halbleiter (PbS, PbTe, PbSe, SnS) oder I-VII-Verbindungen (z. B. CuCl, AgBr). Letztere haben aufgrund des sehr starken ionischen Charakters der Bindung große Energielücken ( $> 3$  eV). Die binären Verbindungen aus Gruppe-IV und Gruppe-VI Elementen haben dagegen sehr kleine Bandlücken und kommen in Infrarotdetektoren zum Einsatz.

Zu den Verbindungshalbleitern zählen wir auch *ternäre Systeme* wie den I-III-VI-Halbleiter  $CuInSe_2$  (CIS) oder das II-IV-V-System  $CdSnAs_2$ . CIS wird zusammen mit seinen kompositionellen Abkömmlingen  $Cu(In,Ga)(S,Se)_2$  als Material für Dünnschichtsolarzellen benutzt, wobei hier aber die begrenzte Verfügbarkeit von In ein Problem darstellt. Weitere ternäre Systeme sind  $ZnCdTe$  oder  $HgCdTe$  (Anwendung in Infrarotdetektoren) sowie  $AlGaAs$  und  $GaAsP$  (Anwendung in der Lasertechnik). Für die Herstellung von Lasern für die optische Kommunikationstechnik bei Wellenlängen von 1.3 und 1.5  $\mu m$  werden *quaternäre Verbindungshalbleiter* wie  $(Ga,In)(As,P)$  verwendet.

### 10.1.1.3 Organische Halbleiter

Im Allgemeinen sind organische Materialien elektrisch isolierend. Besitzen allerdings Moleküle oder Polymere ein konjugiertes Bindungssystem, das aus Doppelbindungen, Dreifachbindungen und aromatischen Ringen besteht, können sie elektrisch leitend und als organische Halbleiter verwendet werden. Als erstes wurde dies 1976 bei Polyacetylen beobachtet.<sup>5</sup>

<sup>5</sup> C. K. Chiang *et al.*, *Electrical Conductivity in Doped Polyacetylene*, Phys. Rev. Lett. **39**, 1098–1101 (1977).

Bis heute wurden viele organische Verbindungen gefunden, die Halbleiter sind. Organische Halbleiter sind sehr viel versprechend für Anwendungen (z. B. für OLEDs – Organic Light Emitting Diodes oder OFETs – Organic Field Effect Transistors), da sie billig herzustellen sind, leicht hinsichtlich ihrer elektrischen und optischen Eigenschaften modifiziert werden können und biegsam sind. Zur Zeit haben allerdings Bauelemente basierend auf organischen Halbleitern häufig noch Haltbarkeitsprobleme.<sup>6</sup>

#### 10.1.1.4 Oxidische Halbleiter

Die meisten Oxide sind gute Isolatoren. Allerdings gibt es auch halbleitende Oxide wie CuO oder CuO<sub>2</sub>, das schon in den 1920er Jahren untersucht wurde und später in den 1950er Jahren für grundlegende Untersuchungen zur Exzitonenphysik verwendet wurde. Der zur Zeit wohl wichtigste Vertreter oxidischer Halbleiter ist ZnO. Es besteht hier die Hoffnung, dass ZnO aufgrund seiner physikalischen Eigenschaften besser für optoelektronische Bauelemente im blauen und ultravioletten Bereich geeignet ist als GaN.

Bei starker Dotierung werden einige oxidische Halbleiter metallisch und bei tiefen Temperaturen supraleitend. Der bekannteste Vertreter ist wohl das Materialsystem La<sub>2</sub>CuO<sub>4</sub> mit einer Energielücke von etwa 2 eV, das bei einer genügend großen Lochdotierung (partielle Substitution des dreiwertigen La durch zweiwertiges Ba oder Sr) metallisch und unterhalb von etwa 40 K supraleitend wird. An der Substanz La<sub>2-x</sub>Ba<sub>x</sub>CuO<sub>4</sub> wurde von **Bednorz** und **Müller** im Jahr 1986 die Hochtemperatur-Supraleitung entdeckt.<sup>7</sup>

#### 10.1.1.5 Schicht-Halbleiter

Materialien wie PbI<sub>2</sub>, MoS<sub>2</sub> oder GaSe besitzen eine ausgeprägte Schichtstruktur. Die Bindung innerhalb der Schichten ist kovalent, während zwischen den Schichten nur eine schwache van der Waals Bindung vorliegt. Die Materialsysteme haben Interesse gefunden, da sie intrinsisch quasi-zweidimensionale Elektronensysteme bilden und die Wechselwirkung der Schichten durch Einbringung von Fremdatomen (Interkalation) variiert werden kann.

#### 10.1.1.6 Magnetische Halbleiter

Durch die schnell wachsende Bedeutung des Arbeitsgebiets der *Spin-Elektronik*, in dem versucht wird, den Spin-Freiheitsgrad von Elektronen zur Verbesserung von elektronischen Bauelementen und zur Realisierung von neuartigen spintronischen Bauelementen auszunutzen, ist das Interesse an magnetischen Halbleitern stark gewachsen. Magnetische Halbleiter könnten z. B. für die Erzeugung und Detektion spinpolarisierter Ströme verwendet werden. Der wohl am längsten bekannte magnetische Halbleiter ist EuS. Heute werden ferromagnetische Halbleiter häufig dadurch erzeugt, dass ein kleiner Teil (typischerweise einige Prozent) der Halbleiteratome durch magnetische Atome wie Mn oder Co ersetzt werden. Typische Beispiele sind (Ga,Mn)As oder (Zn,Co)O. Wir nennen solche Systeme verdünnte magnetische Halbleiter (DMS: Diluted Magnetic Semiconductors).

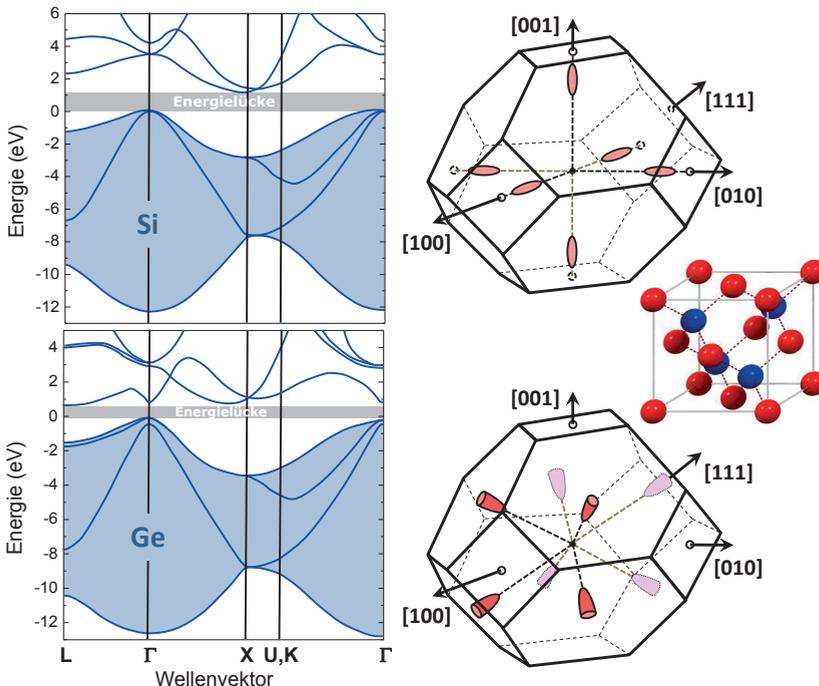
<sup>6</sup> M. Schwoerer, *Physikalische Blätter* **49**, 52 (1994); H. Sixl, H. Schenk, N. Yu, *Physikalische Blätter* **54**, 225 (1998); *Physik Journal* **7**, 29–32 (2008).

<sup>7</sup> G. Bednorz, K. A. Müller, *Possible high-T<sub>c</sub> superconductivity in the Ba-La-Cu-O system*, *Zeitschrift für Physik B* **64**, 189 (1986).

## 10.1.2 Intrinsische Halbleiter

### 10.1.2.1 Bandstruktur und effektive Masse

Wir haben in Abschnitt 8.5.2 bereits die Grundzüge der Bandstruktur von Element-Halbleitern der IV. Hauptgruppe diskutiert. In diesen Elementen liegt eine ausgeprägte  $sp^3$ -Hybridisierung vor, was in einer für die Diamantstruktur charakteristischen tetragonalen Bindungsstruktur resultiert. Die Bildung von  $sp^3$ -Hybridorbitalen führt zur Ausbildung von zwei energetisch getrennten  $sp^3$ -Subbändern (vergleiche hierzu Abb. 8.18 in Abschnitt 8.4.1). Wie Abb. 10.1 zeigt, ist das untere dieser Bänder mit den vier Valenzelektronen vollständig gefüllt, das obere ist vollkommen leer. Die Fermi-Energie liegt bei  $T = 0$  etwa in der Mitte der Bandlücke. Bei Germanium liegt die kleinste Lücke von  $E_g = 0.742$  eV zwischen dem  $\Gamma$ -Punkt ( $\mathbf{k} = [000]$ , Oberkante des Valenzbandes) und dem L-Punkt ( $\mathbf{k} = \frac{a}{\sqrt{3}}[111]$ , Unterkante des Leitungsbandes) vor. Die direkte Bandlücke am  $\Gamma$ -Punkt ist mit  $E_g = 1.1$  eV etwas größer. Wir sprechen deshalb von einem **indirekten Halbleiter**. Die Situation ist ähnlich für Si, wo die minimale Energielücke von  $E_g = 1.17$  eV zwischen dem  $\Gamma$ -Punkt ( $\mathbf{k} = [000]$ , Oberkante des Valenzbandes) und etwa dem 0.8-fachen des Abstandes zum X-Punkt ( $\mathbf{k}_0 = 0.82a[100]$ , Unterkante des Leitungsbandes) vorliegt.



**Abb. 10.1:** Berechnete Bandstrukturen von Si und Ge (nach J. R. Chelikowski und M. L. Cohen, Phys. Rev. B 14, 556 (1976)). Die vier Valenzbänder (unteres  $sp^3$ -Subband) sind farbig hinterlegt. Rechts sind die Flächen konstanter Energie in der Nähe des Minimums des Leitungsbandes und die Diamantstruktur gezeigt. Beide Materialien haben Diamantstruktur mit einem kubisch flächenzentrierten Bravais-Gitter, so dass die 1. Brillouin-Zone ein Rhombendodekaeder ist. Aus Gründen der Symmetrie treten immer mehrere Bandminima in äquivalente Richtungen des  $k$ -Raums auf. Für Si sind dies die  $\{001\}$ -, für Ge die  $\{111\}$ -Richtungen.

Wir haben in Abschnitt 8.3 auch bereits diskutiert, welche Abhängigkeit wir für die Größe der Energielücke vom Abstand der Atome im Gitter erwarten. Wir haben gesehen, dass mit wachsendem Abstand der Atome die Aufspaltung zwischen den Bändern abnimmt (vergleiche Abb. 8.15). Da sich Festkörper mit zunehmender Temperatur aufgrund von anharmonischen Effekte ausdehnen, erhalten wir also eine Zunahme des Abstands der Atome und damit eine Abnahme der Energielücke mit zunehmender Temperatur. Neben der thermischen Ausdehnung wirkt sich auch die Temperaturabhängigkeit der Phononenverteilung auf die Bandstruktur und die Größe der Energielücke aus, da Elektron-Phonon-Streuung zu einer Erniedrigung des effektiven Potenzials führt. Beide Effekte zusammen resultieren in einer Abnahme von  $E_g$  mit zunehmender Temperatur, die phänomenologisch mit der *Varshni-Formel*<sup>8</sup>

$$E_g(T) = E_g(0) - \frac{aT^2}{T + b} \quad (10.1.1)$$

beschrieben werden kann (Si:  $a = 4.73 \times 10^{-4} \text{ eV/K}$ ,  $b = 636 \text{ K}$ ; Ge:  $a = 4.774 \times 10^{-4} \text{ eV/K}$ ,  $b = 235 \text{ K}$ , GaAs:  $a = 5.405 \times 10^{-4} \text{ eV/K}$ ,  $b = 204 \text{ K}$ ). Bei tiefen Temperaturen ( $T \ll b$ ) ergibt sich ein etwa quadratischer, bei Raumtemperatur ein etwa linearer Temperaturverlauf der Energielücke. In Tabelle 10.2 sind die Werte der Bandlücke für einige Halbleiter bei 0 K und 300 K angegeben.

Die für die elektronischen Eigenschaften von Halbleitern relevanten Zustände liegen alle an der Oberkante des Valenz- bzw. der Unterkante des Leitungsbandes. Wir haben in Kapitel 8 bereits gesehen, dass wir den Bandverlauf in der Nähe der Bandkante gut durch eine Parabel annähern können:

$$E(\mathbf{k}) = E_c + \frac{\hbar^2}{2} \sum_{ij} k_i \left( \frac{1}{m^*} \right)_{ij} k_j \quad (\text{Elektronen}) \quad (10.1.2)$$

$$E(\mathbf{k}) = E_v + \frac{\hbar^2}{2} \sum_{ij} k_i \left( \frac{1}{m^*} \right)_{ij} k_j \quad (\text{Löcher}) \quad (10.1.3)$$

Hierbei ist  $E_c$  die Energie an der Unterkante des Leitungsbandes,  $E_v$  diejenige an der Oberkante des Valenzbandes und  $(m^{*-1})_{ij}$  der durch (9.1.13) gegebene effektive Massentensor.

**Tabelle 10.2:** Werte der Energielücken von einigen Halbleitern bei  $T = 0 \text{ K}$  (extrapolierte Werte) und bei 300 K. Quelle: *Handbook Series on Semiconductor Parameters*, Vol. 1 and 2, edited by M. Levinstein, S. Rumyantsev and M. Shur, World Scientific, London (1996, 1999).

Halbleiter	Typ	$E_g$ (0 K)	$E_g$ (300 K)	Halbleiter	Typ	$E_g$ (0 K)	$E_g$ (300 K)
Si	indir.	1.17	1.12	GaP	indir.	2.32	2.26
Ge	indir.	0.742	0.661	InP	direkt	1.421	1.344
GaAs	direkt	1.519	1.424	ZnO	direkt	3.44	3.2
InSb	direkt	0.24	0.17	ZnS	–	3.91	3.6
InAs	direkt	0.415	0.354	CdS	direkt	2.58	2.42
AlSb	indir.	1.65	1.58	CdTe	direkt	1.61	1.45
GaN (Wurzit)	indir.	3.47	3.39	GaN (ZnS)	direkt	3.28	3.20

<sup>8</sup> Y.P. Varshni, *Temperature dependence of the energy gap in semiconductors*, Physica **34**, 149 (1967).

Wir werden im Folgenden diese *parabolische Näherung* häufig verwenden. Es ist zu beachten, dass (10.1.3) eine nach unten geöffnete Parabel darstellt, da die effektive Masse der Löcher am  $\Gamma$ -Punkt negativ ist.

Da der effektive Massentensor reell und symmetrisch ist, können wir einen Satz von orthogonalen Hauptachsen finden, bezüglich der die Energien die diagonale Form

$$E(\mathbf{k}) = E_c + \hbar^2 \left( \frac{k_1^2}{2m_1} + \frac{k_2^2}{2m_2} + \frac{k_3^2}{2m_3} \right) \quad (\text{Elektronen}) \quad (10.1.4)$$

$$E(\mathbf{k}) = E_v + \hbar^2 \left( \frac{k_1^2}{2m_1} + \frac{k_2^2}{2m_2} + \frac{k_3^2}{2m_3} \right) \quad (\text{Löcher}) \quad (10.1.5)$$

haben. Die Flächen konstanter Energie sind somit Ellipsoide, die durch Angabe der drei Hauptachsen, der drei effektiven Massen und der Position im  $\mathbf{k}$ -Raum eindeutig definiert sind. Auf der rechten Seite von Abb. 10.1 sind diese Ellipsoide für Si und Ge skizziert. Beide Materialien haben Diamantstruktur mit einem kubisch flächenzentrierten Bravais-Gitter, so dass die 1. Brillouin-Zone ein Rhombendodekaeder ist.

Für Si hat das Leitungsband aus Symmetriegründen sechs Minima, die entlang der  $\{100\}$  Richtungen liegen und zwar bei etwa 80% des Abstandes zum Zonenrand. Jedes der sechs äquivalenten Ellipsoide muss rotationssymmetrisch bezüglich einer Rotation um die Würfelachsen sein. Wie Abb. 10.1 zeigt, sind die Ellipsoide zigarrenförmig in Richtung der Würfelachse gestreckt. Wir können zwei effektive Massen definieren. Während die *longitudinale effektive Masse*  $m_{el}^*$  entlang der Achse etwa der freien Elektronenmasse entspricht, ist die *transversale effektive Masse*  $m_{et}^*$  senkrecht zur Achse wesentlich kleiner und beträgt nur etwa  $0.2 m$ . Für Si gibt es zwei entartete Valenzbandmaxima bei  $\mathbf{k} = 0$ , die kugelsymmetrisch sind mit effektiven Massen  $m_{lh}^* = 0.16 m$  (lh = light holes) und  $m_{hh}^* = 0.49 m$  (hh = heavy holes).

Für Ge ist die Kristallstruktur und die Brillouin-Zone identisch zu Si. Allerdings treten die Minima des Leitungsbandes jetzt an den äquivalenten L-Punkten der Brillouin-Zone, das heißt, an den Zonenrändern entlang der  $\{111\}$  Richtungen auf (siehe hierzu Abb. 10.1). Die Flächen konstanter Energie sind wiederum Ellipsoide, die jetzt aber entlang der  $\{111\}$  Richtungen gestreckt sind mit effektiven Massen  $m_{el}^* = 1.57 m$  und  $m_{et}^* = 0.081 m$ . Da die Ellipsoide mit der jeweiligen Nachbarzelle geteilt werden, haben wir bei Ge nur  $8/2 = 4$  äquivalente Ellipsoide. Wie bei Si liegen zwei entartete Valenzbandmaxima bei  $\mathbf{k} = 0$  mit effektiven Massen  $m_{lh}^* = 0.043 m$  und  $m_{hh}^* = 0.33 m$  vor. Die verschiedenen effektiven Massen von Si und Ge und einiger anderer Halbleiter sind in Tabelle 10.3 zusammengestellt.

Halbleiter	$m_c^*/m$	$m_{et}^*/m$	$m_{el}^*/m$	$m_{lh}^*/m$	$m_{hh}^*/m$	$m_{soh}^*/m$	$\Delta$ (eV)
Si		0.19	0.98	0.16	0.49	0.24	0.044
Ge		0.081	1.59	0.043	0.33	0.084	0.295
GaAs	0.063			0.082	0.51	0.14	0.341
GaSb	0.041			0.04	0.4	0.15	0.80
GaP		1.12	0.22	0.14	0.79	0.25	0.08
InAs	0.023			0.026	0.41	0.16	0.41
InP	0.073			0.089	0.58	0.17	0.11
InSb	0.014			0.015	0.43	0.19	0.81

**Tabelle 10.3:** Effektive Massen von Elektronen und Löchern sowie Spin-Bahn-Aufspaltung  $\Delta$  für Si und Ge. Quelle: *Handbook Series on Semiconductor Parameters*, Vol. 1 and 2, edited by M. Levinstein, S. Rumyantsev and M. Shur, World Scientific, London (1996, 1999).

**Spin-Bahn-Aufspaltung** Eine detailliertere Analyse zeigt, dass die Bandstruktur von Si und Ge in der Nähe des Valenzbandmaximums beim  $\Gamma$ -Punkt etwas komplizierter ist. Neben den beiden Bändern mit unterschiedlicher Krümmung existiert ein weiteres Band, das von diesen Bändern um  $\Delta = 0.29$  eV für Ge und  $\Delta = 0.044$  eV für Si getrennt ist. Anschaulich kann dies dadurch verstanden werden, dass die Bänder im Rahmen einer Tight-Binding-Näherung aus den atomaren  $2p$ -Zuständen ( $L = 1, S = 1/2$ ) der freien Atome hervorgehen. Als Folge der Spin-Bahn-Kopplung  $\mathcal{H}_{so} = \lambda(\mathbf{L} \cdot \mathbf{S})$  kommt es zu einer Aufspaltung der Zustände mit Gesamtdrehimpuls  $J = L + S = 3/2$  und  $J = L - S = 1/2$  (vergleiche hierzu Abschnitt 8.3.3 und 12.5.4). Die Zustände mit  $J = 3/2$  ( $m_j = -3/2, -1/2, +1/2, +3/2$ ) werden um  $\frac{1}{3}\Delta$  nach oben verschoben und bilden die entarteten  $hh$ - ( $J = 3/2, m_j = \pm 3/2$ ) und  $lh$ -Bänder ( $J = 3/2, m_j = \pm 1/2$ ).<sup>9</sup> Die Zustände mit  $J = 1/2$  ( $m_j = \pm 1/2$ ) werden um  $-\frac{2}{3}\Delta$  nach unten verschoben und bilden das  $soh$ -Band. Die Valenzbandkante ist deshalb gegenüber dem Wert  $E_v(\lambda = 0)$  ohne Spin-Bahn-Kopplung um  $\frac{1}{3}\Delta$  nach oben verschoben und die Aufspaltung zwischen den  $hh, lh$ -Bändern und dem  $soh$ -Band beträgt  $\Delta$ . Der Energieunterschied  $\Delta$  ist ein Maß für die Stärke der Spin-Bahn-Wechselwirkung, die für Ge ( $\Delta = 290$  meV) und GaAs ( $\Delta = 350$  meV) aufgrund der größeren Kernladungszahlen wesentlich größer ist als für Si ( $\Delta = 44$  meV) oder Diamant ( $\Delta = 6$  meV). Das qualitative Verhalten der Bänder in der Nähe des  $\Gamma$ -Punktes ist in Abb. 10.2 zusammengefasst.

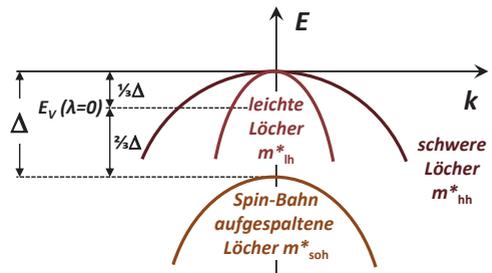


Abb. 10.2: Qualitativer Verlauf der Bandstruktur von Si und Ge in der Nähe des Valenzbandmaximums beim  $\Gamma$ -Punkt.  $\Delta$  ist die Spin-Bahn-Aufspaltung.

Im Gegensatz zu den Element-Halbleitern der IV. Hauptgruppe hat die chemische Bindung in den III-V Halbleitern einen ionischen Anteil. Den gemischten Charakter der Bindung können wir uns so vorstellen, dass ein Elektron vom Ga zum As Atom transferiert wird, so dass eine  $\text{Ga}^+ \text{As}^-$  Konfiguration vorliegt. Diese führt natürlich zu einer ionischen Bindung. Dieser überlagert ist eine kovalente Bindung, die zu keinem Ladungstransfer führt und beide Atome mit vier Elektronen belässt. Es kann sich dann wie für Si und Ge eine  $sp^3$ -Hybridisierung ausbilden. Wir können davon ausgehen, dass die kovalente Bindung dominiert, da der Kristall sonst nicht in der tetraedrisch koordinierten Zinkblende-Struktur vorliegen würde. Wie bei der Diamantstruktur sitzen hier die einzelnen Gitteratome im Mittelpunkt eines Tetraeders, das von vier nächsten Nachbarn gebildet wird. Die Zinkblende-Struktur unterscheidet sich nur dadurch von der Diamantstruktur, dass bei ihr die Basis zwei verschiedenartige Atome enthält (vergleiche Abschnitt 1.2.9).

<sup>9</sup> Die schweren und leichten Löcher resultieren dabei aus den  $\pi_{2p}$ - und  $\sigma_{2p}$ -Bindungen zwischen den beteiligten  $2p$ -Orbitalen. Da der Überlapp bei der  $\sigma_{2p}$ -Bindung größer ist, resultiert gemäß dem Tight-Binding-Modell eine größere Bandbreite und damit größere Banddispersion, woraus sich eine geringere effektive Masse ergibt.

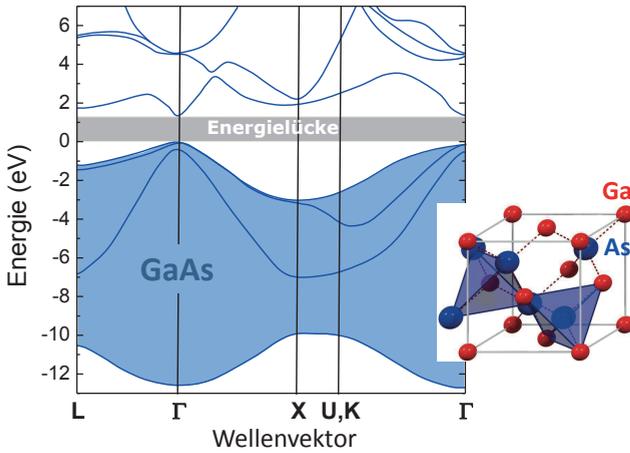


Abb. 10.3: Berechnete Bandstrukturen von GaAs (nach J. R. Chelikowski und M. L. Cohen, Phys. Rev. B 14, 556 (1976)). Die vier Valenzbänder (unteres  $sp^3$ -Subband) sind farbig hinterlegt. Rechts ist die Zinkblende-Struktur von GaAs gezeigt.

Der wichtigste Unterschied zwischen GaAs und Si bzw. Ge ist die Tatsache, dass GaAs genauso wie InAs, InSb oder InP ein *direkter Halbleiter* ist. Wie Abb. 10.3 zeigt, liegt das Maximum des Valenzbandes genauso wie das Minimum des Leitungsbandes beim  $\Gamma$ -Punkt. Ähnlich zu Si und Ge besitzt GaAs auch drei Valenzbänder ähnlicher Form mit den effektiven Massen  $m_{lh}^* = 0.08 m$ ,  $m_{hh}^* = 0.51 m$  und  $m_{soh}^* = 0.14 m$  sowie der Spin-Bahn-Aufspaltung  $\Delta = 0.34 eV$ . Die effektive Masse der Elektronen im Leitungsband beträgt  $m_e^* = 0.063 m$ .

10.1.2.2 Optische Absorption

Die Größe der Energielücke von Halbleitern kann durch Messung der optischen Absorption ermittelt werden. Bestrahlen wir einen Halbleiter mit Photonen, so kann durch Absorption eines Photons mit Energie  $\hbar\omega$  und Wellenvektor  $\mathbf{k}_\gamma$  ein Elektron aus dem Valenzband ins Leitungsband angeregt werden. Für solche *Interband-Übergänge* muss sowohl die Energie als auch die Impulserhaltung

$$E_g = \hbar\omega(\mathbf{k}_\gamma) \pm \hbar\Omega(\mathbf{q}) \tag{10.1.6}$$

$$\hbar\Delta\mathbf{k} = \hbar\mathbf{k}_\gamma \pm \mathbf{q} \tag{10.1.7}$$

gelten. Hierbei ist  $\Delta\mathbf{k}$  die Wellenvektordifferenz der beteiligten Zustände im Valenz- und Leitungsband. Wegen des sehr kleinen Wellenvektors  $k_\gamma = \omega/c$  von Photonen muss für Übergänge, an denen nur Photonen beteiligt sind,  $\Delta\mathbf{k}$  sehr klein sein. Solche Übergänge müssen

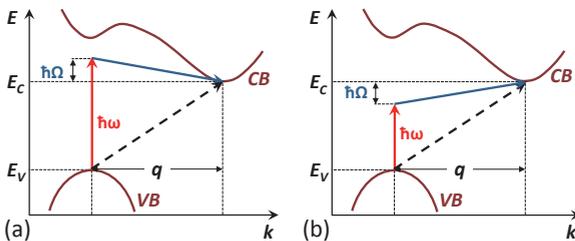
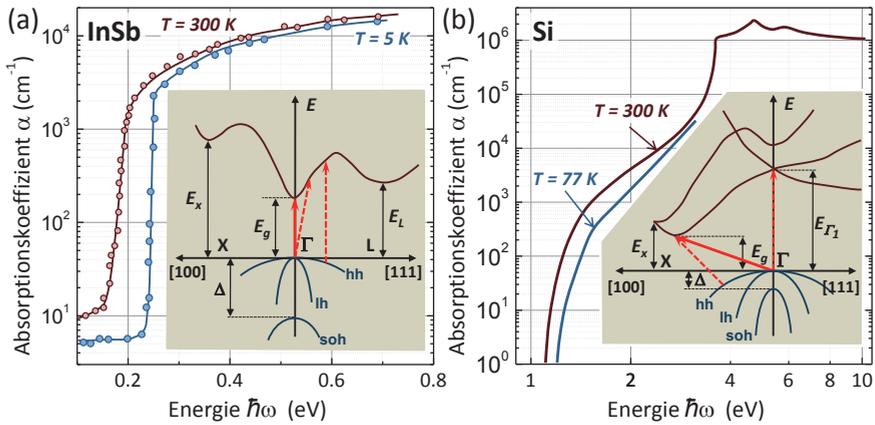


Abb. 10.4: Schematische Darstellung von Interbandübergängen zwischen Valenzbandmaximum und Leitungsbandminimum in indirekten Halbleitern bei (a) Absorption und (b) Emission eines Phonons mit Energie  $\hbar\Omega$  und Wellenvektor  $\mathbf{q}$ . Der Wellenvektor des Photons wurde vernachlässigt.



**Abb. 10.5:** (a) Schematische Bandstruktur und optische Absorption des direkten Halbleiters InSb bei 5 und 300 K. Der rote Pfeil markiert den Übergang mit kleinstmöglicher Energie. Bei höherer Photonenenergie können auch höher liegende Elektronenzustände angeregt werden (gestrichelte Pfeile). Quelle: E. J. Johnson, *Semiconductors and Semimetals*, R. K. Willardson und A. C. Beer, eds., Academic Press, N. Y., Vol. 3, 153–258 (1967). (b) Schematische Bandstruktur und optische Absorption des indirekten Halbleiters Si bei 77 und 300 K. Der rote Pfeil markiert wiederum den Übergang mit kleinstmöglicher Energie, der allerdings nur unter der Beteiligung eines Phonons möglich ist. Die schwache indirekte Absorption setzt bei  $\hbar\omega = E_g - \hbar\Omega$  ein und ist der bei höheren Photonenenergien (gestrichelte Pfeile) möglichen, viel stärkeren direkten Absorption vorgelagert. Quelle: S. M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons, N. Y. (1981).

im  $E(\mathbf{k})$  Diagramm quasi vertikal verlaufen. Übergänge mit größeren  $\Delta k$  sind nur durch die Beteiligung von weiteren Anregungen wie z. B. Phononen möglich. Da üblicherweise die Phononenenergie  $\hbar\Omega \ll \hbar\omega$  und der Phononenimpuls  $|\mathbf{q}| \gg |\mathbf{k}|$ , liefert dabei das Photon die Energie und das Phonon den Impuls. Optische Übergänge in indirekten Halbleitern mit Absorption und Emission eines Phonons sind in Abb. 10.4 schematisch dargestellt.

Abb. 10.5 zeigt typische Absorptionsspektren eines direkten (InSb) und indirekten Halbleiters (Si). Der Absorptionskoeffizient ist hierbei durch  $\alpha = -I^{-1}(dI/dx)$  gegeben, also als die relative Schwächung einer Lichtwelle mit Intensität  $I$  beim Durchlaufen eines Mediums in  $x$ -Richtung. Bei direkten Halbleitern ist die Situation einfach. Da die Unterkante des Leitungsbandes hier beim selben Wellenvektor wie die Oberkante des Valenzbandes liegt, sind vertikale Übergänge ohne Beteiligung von Phononen möglich, sobald die Photonenenergie  $\hbar\omega \geq E_g$  wird. Die Absorption steigt deshalb bei  $\hbar\omega = E_g$  stark an. Halbleiter sind deshalb für Frequenzen  $\omega \leq E_g/\hbar$ , die für viele Halbleiter im nahen Infraroten liegen, transparent. Durch Messung der optischen Absorption als Funktion der Photonenenergie (siehe Abb. 10.5a) kann der Wert der Energielücke einfach bestimmt werden. Bei höheren Temperaturen steigt die Absorption an, da jetzt Prozesse unter Absorption eines Phonons häufiger werden. Diese setzen bereits bei der Photonenenergie  $\hbar\omega = E_g - \hbar\Omega_q$  ein.

Bei indirekten Halbleitern (siehe Abb. 10.4 und 10.5b) ist die Situation etwas komplizierter. Da sich die Wellenvektoren für die Oberkante des Valenzbandes und die Unterkante des Leitungsbandes um  $\Delta\mathbf{k}$  unterscheiden, ist für Übergänge zwischen diesen Bereichen zur Impulserhaltung jetzt die Mitwirkung eines Phonons notwendig. Dies kann prinzipiell durch

Absorption oder Emission eines Phonons erfolgen. Führt man allerdings eine Absorptionsmessung bei tiefen Temperaturen durch, so sind aufgrund des Ausfrierens der Phononen fast nur Phononenemissionsprozesse möglich. Das absorbierte Photon muss also zusätzlich zur Energie  $E_g$  noch die Energie  $\hbar\Omega$  zur Erzeugung des Phonons liefern. Die optische Absorption setzt deshalb nicht wie bei direkten Halbleitern bei  $\hbar\omega = E_g$ , sondern erst bei  $\hbar\omega = E_g + \hbar\Omega$  ein. Ferner ist die Absorptionsstärke im Vergleich zu direkten Halbleitern stark reduziert. Dies liegt an der reduzierten Wahrscheinlichkeit des jetzt notwendigen Dreiteilchenprozesses. Dies ist in Abb. 10.5b gezeigt. Der Absorptionskoeffizient ist oberhalb von  $\hbar\omega = E_g + \hbar\Omega$  zunächst klein und steigt erst dann stark an, wenn die Photonenenergie ausreicht, um vertikale Übergänge ohne Beteiligung eines Phonons zu ermöglichen. Bei Si ist dies bei  $\hbar\omega = E_{\Gamma_1} = 3.4 \text{ eV}$  der Fall. Bei höheren Temperaturen werden Prozesse unter Absorption eines Phonons häufiger. Diese setzen bereits bei der Photonenenergie  $\hbar\omega = E_g - \hbar\Omega$  ein und erhöhen den Absorptionskoeffizienten. Für Übergänge vom  $\Gamma$ -Punkt zum X-Punkt werden Phononen mit dem Wellenvektor  $q_0$  benötigt. In der Nähe von  $q_0$  verläuft die Dispersionskurve der in Frage kommenden transversal-akustisch (TA) und transversal-optischen (TO) Phononen sehr flach, so dass ihre Energie in etwa konstant ist (TA: 18.7 meV, TO: 58.1 meV).

Zur quantitativen Erklärung des Absorptionskoeffizienten müssen nach Fermis goldener Regel die Übergangsmatrixelemente und die Zustandsdichten der Anfangs- und Endzustände bekannt sein (vergleiche hierzu Abschnitt 11.6.4). Bei direkten Übergängen werden wegen  $k_y \approx 0$  Anfangs- und Endzustände mit quasi gleichem  $\mathbf{k}$ -Wert verbunden, so dass

$$\Delta E_{i,f} = \hbar\omega \approx E_c(\mathbf{k}) - E_v(\mathbf{k}) = \left( \frac{1}{m_e^*} + \frac{1}{m_h^*} \right) \frac{\hbar^2 k^2}{2} = \frac{\hbar^2 k^2}{2m_{\text{komb}}^*} \quad (10.1.8)$$

gilt. Hierbei ist  $m_{\text{komb}}^*$  die so genannte kombinierte effektive Masse und wir können  $\Delta E_{i,f}(\mathbf{k})$  als kombinierte Bandstruktur für den optischen Übergang betrachten, der wir die kombinierte Zustandsdichte [vergleiche hierzu (7.1.18) sowie (10.1.15) und (10.1.16)]

$$D_{\text{komb}}(\Delta E_{i,f}) = \frac{V}{2\pi^2} \left( \frac{2m_{\text{komb}}^*}{\hbar^2} \right)^{3/2} \sqrt{\Delta E_{i,f} - E_g} \quad (10.1.9)$$

zuordnen können. Wir erwarten deshalb einen in etwa wurzelförmigen Verlauf

$$\alpha \propto (m_{\text{komb}}^*)^{3/2} \sqrt{\hbar\omega - E_g} \quad (10.1.10)$$

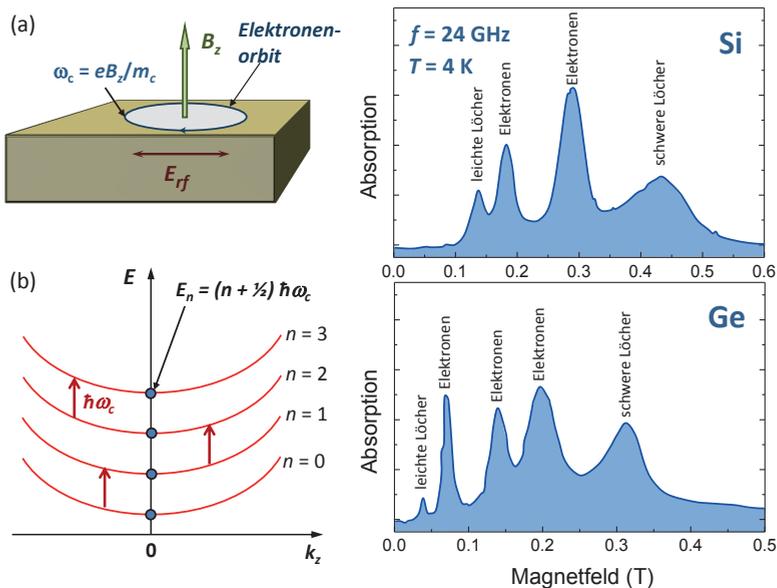
des Absorptionskoeffizienten.

Bei indirekten Übergängen ist die Situation schwieriger, da hier über alle Zustände im Valenz- und Leitungsband aufsummiert werden muss, die unter Phononenbeteiligung mit  $\hbar\omega$  verbunden werden können. Im Allgemeinen ergibt sich ein quadratischer Verlauf  $\alpha \propto (\hbar\omega - \hbar\Omega - E_g)^2$ , der sich aus der Faltung der Zustandsdichten der möglichen Anfangs- und Endzustände ergibt. Eine weitergehende Diskussion der optischen Eigenschaften von Halbleitern erfolgt in Abschnitt 11.6.4 im Rahmen der Diskussion der dielektrischen Eigenschaften von Festkörpern.

### 10.1.2.3 Zyklotron-Resonanz

Die effektiven Massen von Halbleitern können mit Hilfe der Zyklotron-Resonanz (siehe Abschnitt 9.8.3) bestimmt werden. Hierzu wird eine Halbleiterprobe in ein statisches magnetisches Feld gebracht und mit einem dazu senkrechten elektrischen Hochfrequenzfeld angeregt (siehe Abb. 10.6). Die Zyklotronresonanz tritt genau dann auf, wenn die Umlauffrequenz der Ladungsträger um das statische Magnetfeld, d. h. die Zyklotronfrequenz  $\omega_c = eB/m_c$ , mit der Frequenz des elektrischen Feldes übereinstimmt. Hierbei ist  $m_c$  die Zyklotronmasse [vergleiche (9.2.31)]. Typischerweise liegen die Resonanzfrequenzen bei Magnetfeldern von einigen 100 Gauss im Bereich von einigen GHz. Quantenmechanisch können wir uns die Resonanzabsorption durch die Erzeugung von elektrischen Dipolübergängen zwischen benachbarten Landau-Niveaus mit der Auswahlregel  $\Delta n = \pm 1$  vorstellen. Diese Übergänge sollten nicht mit der Resonanzabsorption von magnetischen Dipolen in der Elektronenspinresonanz verwechselt werden. Hier werden Übergänge durch Umklappen der Spinrichtung erzeugt.

Um die Zyklotronresonanz beobachten zu können, muss ferner  $\omega_c \tau \gg 1$  gelten. Das heißt, die Ladungsträger müssen innerhalb der Streuzzeit mehrere Umläufe ausführen können. Dies erreicht man durch Verwendung reiner Proben und tiefer Temperaturen. Da bei tiefen Temperaturen allerdings nur ganz wenige bewegliche Ladungsträger vorhanden sind, müssen



**Abb. 10.6:** Zur Messung der Zyklotronresonanz in Halbleitern. (a) Klassisches Bild: Das Hochfrequenzfeld  $E_{rf}$  steht senkrecht auf dem statischen Magnetfeld  $\mathbf{B}$  und induziert eine Kreisbewegung der Ladungsträger mit wachsendem Bahnradius und Umlaufgeschwindigkeit, aber konstanter Umlauffrequenz  $\omega_c$ . (b) Quantenmechanisches Bild: Das Hochfrequenzfeld induziert Dipolübergänge zwischen benachbarten Landau-Niveaus. Rechts sind die Absorptionsspektren für Si und Ge bei 24 GHz und 4 K gezeigt (Si: Feld in (110) Ebene unter  $30^\circ$  Winkel zu [100]-Richtung. Ge: Feld in (110) Ebene unter  $60^\circ$  Winkel zu [100]-Richtung, nach G. Dresselhaus *et al.*, Phys. Rev. **98**, 368 (1955)).

diese durch Bestrahlung mit Licht ( $\hbar\omega \geq E_g$ ) erzeugt werden. Im Gegensatz zu Metallen ist die Skin-Eindringtiefe des Hochfrequenzfeldes in Halbleitern aufgrund deren wesentlich kleineren elektrischen Leitfähigkeit üblicherweise größer als die Probenabmessung, so dass die Probleme, die mit einer kleinen Skin-Eindringtiefe verbunden sind, hier nicht auftreten.

Für ein homogenes Magnetfeld (z. B. entlang der  $z$ -Achse) kann leicht der Zusammenhang zwischen der Zyklotronmasse  $m_c$  und dem effektive Massentensor  $m^*$  hergestellt werden. Es gilt

$$m_c = \left( \frac{\det m^*}{m_{zz}} \right)^{1/2}. \quad (10.1.11)$$

Mit Hilfe der Eigenwerte und Hauptachsen des effektive Massentensors kann dies als

$$m_c = \sqrt{\frac{m_1 m_2 m_3}{\widehat{B}_1^2 m_1 + \widehat{B}_2^2 m_2 + \widehat{B}_3^2 m_3}} = \sqrt{\left( \frac{\cos^2 \theta}{m_t^2} + \frac{\sin^2 \theta}{m_l m_l} \right)^{-1}} \quad (10.1.12)$$

geschrieben werden. Hierbei sind  $\widehat{B}_i$  die Komponenten des Einheitsvektors in Feldrichtung entlang der drei Hauptachsen. Das zweite Gleichheitszeichen gilt für Materialien, für die die effektive Masse in einer Ebene senkrecht zu einer (longitudinalen) Hauptachsenrichtung gleich ist. Der Winkel  $\theta$  ist dabei der Winkel zwischen der Magnetfeldrichtung und der longitudinalen Hauptachse.

#### 10.1.2.4 Ladungsträgerdichte

Wir bezeichnen Halbleiter als *intrinsisch*, wenn freie Elektronen und Löcher nur durch Anregungen vom Valenzband ins Leitungsband erzeugt werden können. Wir wollen nun für intrinsische Halbleiter die Dichte  $n_c$  der Elektronen im Leitungsband und die Dichte  $p_v$  der Löcher im Valenzband ableiten. Diese Größen sind für die elektrischen Transporteigenschaften von Halbleitern von zentraler Bedeutung.

Wie in jedem Festkörper gehorcht die Besetzungswahrscheinlichkeit der elektronischen Zustände in einem Halbleiter der Fermi-Statistik  $f(E, T)$ , so dass wir schreiben können:

$$n_c = \frac{1}{V} \int_{E_c}^{\infty} D_c(E) f(E, T) dE \quad (10.1.13)$$

$$p_v = \frac{1}{V} \int_{-\infty}^{E_v} D_v(E) [1 - f(E, T)] dE. \quad (10.1.14)$$

Hierbei sind  $D_v(E)$  und  $D_c(E)$  die Zustandsdichten der Löcher im Valenzband und der Elektronen im Leitungsband. In der Nähe der Bandkante können wir die Bandstruktur gut durch eine Parabel annähern und erhalten [vergleiche (7.1.18)]

$$D_c(E) = \frac{V}{2\pi^2} \left( \frac{2m_{e,\text{DOS}}^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c} \quad (E \geq E_c) \quad (10.1.15)$$

$$D_v(E) = \frac{V}{2\pi^2} \left( \frac{2m_{h,\text{DOS}}^*}{\hbar^2} \right)^{3/2} \sqrt{E_v - E} \quad (E \leq E_v). \quad (10.1.16)$$

Die Zustandsdichte im Bereich der Energielücke,  $E_v < E < E_c$  ist natürlich null. Hierbei sind  $m_e^*$  und  $m_h^*$  die Zustandsdichtemassen des Leitungs- und Valenzbandes, die in die Berechnung der jeweiligen Zustandsdichten einfließen. Um die Form des isotropen Falls beibehalten zu können, definieren wir die effektive Zustandsdichtemasse als

$$m_{\text{DOS}}^* = p^{2/3} (m_1^* m_2^* m_3^*)^{1/3}, \tag{10.1.17}$$

wobei der Entartungsfaktor  $p$  die Zahl der äquivalenten Bandminima bzw. -maxima angibt und  $m_i^*$  die Eigenwerte des effektive Massentensors für die Elektronen und Löcher sind. Für das Leitungsband ist  $p = 6$  bzw.  $4$  für Si bzw. Ge, woraus sich  $m_{e,\text{DOS}}^* = 1.08m_e$  für Si und  $m_{e,\text{DOS}}^* = 0.55m_e$  für Ge ergibt. Die Zustandsdichten der Valenzbänder ( $hh$ ,  $lh$ ,  $soh$ ) addieren sich, woraus sich  $m_{h,\text{DOS}}^{*3/2} = m_{hh}^{*3/2} + m_{lh}^{*3/2} + m_{soh}^{*3/2}$  ergibt. Wir erhalten dann  $m_{h,\text{DOS}}^* = 0.65m_e$  für Si und  $m_{h,\text{DOS}}^* = 0.34m_e$  für Ge. Für Ge haben wir dabei nur den Beitrag der  $hh$ - und  $lh$ -Bänder berücksichtigt. Das  $soh$ -Band kann meist vernachlässigt werden, da es relativ weit (im Vergleich zu  $k_B T$  bei Raumtemperatur) unterhalb der Valenzbandkante liegt.

In einem intrinsischen Halbleiter stammen alle freien Elektronen im Leitungsband aus dem Valenzband. Deshalb muss die Zahl der Elektronen im Leitungsband und die der Löcher im Valenzband immer exakt gleich sein. Zur Veranschaulichung ist dies in Abb. 10.7 skizziert. Falls die effektiven Zustandsdichtemassen der Elektronen und Löcher und damit ihre Zustandsdichten gleich sind, so liegt das chemische Potenzial  $\mu$  in der Mitte der Energielücke. Falls aber z. B. die Zustandsdichte der Elektronen im Leitungsband kleiner ist, so verschiebt sich das chemische Potenzial in Richtung Leitungsbandkante, so dass die Besetzungsgleichungen (10.1.13) und (10.1.14) gleich sind. Wir werden weiter unten den genauen Zusammenhang zwischen  $\mu$  und den effektiven Massen herleiten.

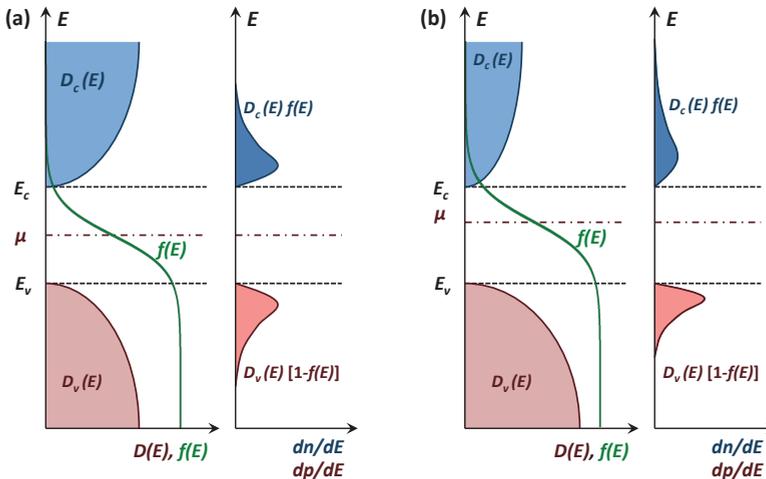


Abb. 10.7: Fermi-Funktion, Zustandsdichten sowie Elektronen- und Löcherkonzentrationen für einen intrinsischen Halbleiter für (a)  $D_v = D_c$  und (b)  $D_v \neq D_c$ .

Da die Temperaturverschmierung der Fermi-Funktion ( $\approx 2k_B T$ ) üblicherweise klein gegen die Energielücke  $E_g$  des Halbleiters ist, können wir  $f(E, T)$  innerhalb des Valenzbandes und des Leitungsbandes durch eine Boltzmann-Verteilung annähern:

$$\frac{1}{e^{(E-\mu)/k_B T} + 1} \simeq e^{-(E-\mu)/k_B T} \quad (E \geq E_c) \quad (10.1.18)$$

$$\frac{1}{e^{(\mu-E)/k_B T} + 1} \simeq e^{-(\mu-E)/k_B T} \quad (E \leq E_v). \quad (10.1.19)$$

Voraussetzung für die Gültigkeit dieser Näherung ist immer, dass das chemische Potenzial genügend weit von den Bandkanten entfernt ist. Mit dieser Näherung erhalten wir für die Ladungsträgerdichte

$$n_c = \frac{1}{2\pi^2} \left( \frac{2m_{e,\text{DOS}}^*}{\hbar^2} \right)^{3/2} e^{\mu/k_B T} \int_{E_c}^{\infty} \sqrt{E - E_c} e^{-E/k_B T} dE. \quad (10.1.20)$$

Substituieren wir  $x_c = (E - E_c)/k_B T$ , so ergibt sich

$$n_c = \frac{1}{2\pi^2} \left( \frac{2m_{e,\text{DOS}}^*}{\hbar^2} \right)^{3/2} (k_B T)^{3/2} e^{-(E_c-\mu)/k_B T} \int_0^{\infty} \sqrt{x_c} e^{-x_c} dx_c. \quad (10.1.21)$$

Mit  $\int \sqrt{x_c} e^{-x_c} dx_c = \sqrt{\pi}/2$  erhalten wir schließlich

$$n_c = 2 \left( \frac{m_{e,\text{DOS}}^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-(E_c-\mu)/k_B T} = n_c^{\text{eff}} e^{-(E_c-\mu)/k_B T}. \quad (10.1.22)$$

In analoger Weise erhalten wir für die Dichte der Löcher im Valenzband

$$p_v = 2 \left( \frac{m_{h,\text{DOS}}^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-(\mu-E_v)/k_B T} = p_v^{\text{eff}} e^{-(\mu-E_v)/k_B T}. \quad (10.1.23)$$

Die nur schwach temperaturabhängigen Faktoren vor den Exponentialfaktoren werden üblicherweise als effektive Ladungsträgerdichten  $n_c^{\text{eff}}$  und  $p_v^{\text{eff}}$  bezeichnet. Lassen wir deren Temperaturabhängigkeit außer Acht, so können wir (10.1.22) und (10.1.23) als Besetzungsdichten von zwei Energieniveaus interpretieren, die vom chemischen Potenzial die Abstände  $E_c - \mu$  bzw.  $\mu - E_v$  haben.

Wir können jetzt die Tatsache  $n_c = p_v$  benutzen, um den Verlauf des chemischen Potenzials zu bestimmen. Mit Hilfe von (10.1.22) und (10.1.23) erhalten wir:

$$\mu = E_v + \frac{1}{2} E_g + \frac{3}{4} k_B T \ln \frac{m_{h,\text{DOS}}^*}{m_{e,\text{DOS}}^*}. \quad (10.1.24)$$

Wir sehen also, dass für  $m_{h,\text{DOS}}^* = m_{e,\text{DOS}}^*$  das chemische Potenzial genau in der Mitte zwischen Valenz- und Leitungsbandkante liegt. Mit Gleichung (10.1.24) könnten wir nun das

chemische Potenzial aus (10.1.22) und (10.1.23) eliminieren. Wir erreichen dies aber auch einfach dadurch, dass wir die Ausdrücke für  $n_c$  und  $p_v$  miteinander multiplizieren. Wir erhalten dadurch

$$n_c \cdot p_v = 4 \left( \frac{k_B T}{2\pi\hbar^2} \right)^3 (m_{e,\text{DOS}}^* m_{h,\text{DOS}}^*)^{3/2} e^{-E_g/k_B T}. \quad (10.1.25)$$

Dieses Ergebnis entspricht dem Massenwirkungsgesetz der chemischen Reaktionskinetik.<sup>10</sup> Bei seiner Herleitung müssen wir keinen Gebrauch von der Bedingung  $n_c = p_v$  machen. Wir können diesen Ausdruck deshalb auch später bei der Diskussion der Störstellenleitung in dotierten Halbleitern benutzen. Eine direkte Folge aus dem Massenwirkungsgesetz (10.1.25) ist, dass wegen  $n_c \cdot p_v = \text{const}$  bei einer Änderung von  $n_c$  (z. B. durch Dotierung) sich auch  $p_v$  ändern muss. Wichtig ist, dass dabei zwar  $n_c \cdot p_v = \text{const}$ , sich aber  $n_c + p_v$  ändert. Da wir wegen  $n_c \cdot p_v = C$  für die Summe  $n_c + p_v = n_c + C/n_c$  erhalten, sehen wir durch Ableiten dieses Ausdrucks sofort, dass  $n_c + p_v$  für  $n_c^2 = C$  und damit  $n_c = p_v$  minimal wird.

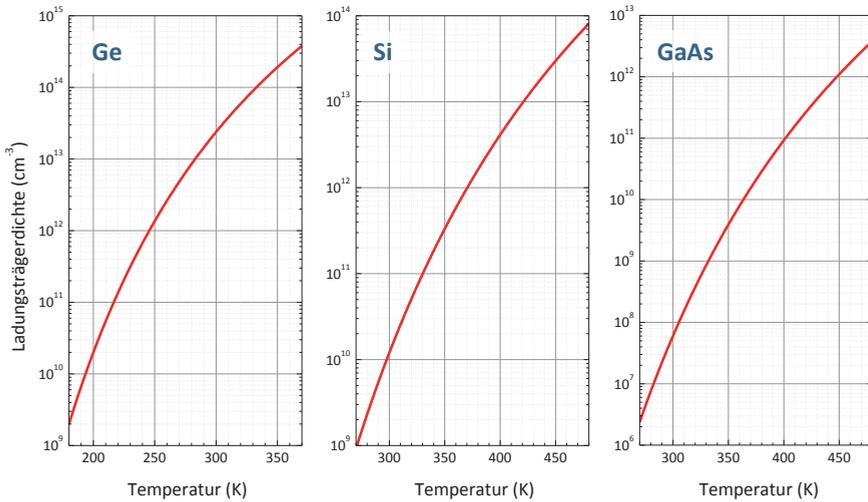


Abb. 10.8: Temperaturabhängigkeit der intrinsischen Ladungsträgerdichte von Si, Ge und GaAs.

<sup>10</sup> Um die Analogie zu einer chemischen Reaktionsgleichung zu verdeutlichen, können wir die Raten-gleichung für die Wechselwirkung eines intrinsischen Halbleiters mit dem Photonenfeld eines schwarzen Strahlers der Temperatur  $T$  betrachten. Die zeitliche Änderung der Dichten  $n_c$  und  $p_v$  der Elektronen im Leitungs- und der Löcher im Valenzband erfolgt durch eine Generationsrate  $A(T)$  und eine Rekombinationsrate  $B(T) n_c \cdot p_v$ . Das heißt, im Gleichgewicht können wir für die Reaktionsgleichung schreiben

$$\frac{dn_c}{dt} = A(T) - B(T)n_c \cdot p_v = \frac{dp_v}{dt} = 0$$

und damit

$$n_c \cdot p_v = \frac{A(T)}{B(T)} = \text{const} = h(T).$$

Mit (10.1.25) ergibt sich die Ladungsträgerkonzentration  $n_i = n_c = p_v$  in intrinsischen Halbleitern zu

$$n_i = \sqrt{n_c p_v} = 2 \left( \frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_{e,\text{DOS}}^* m_{h,\text{DOS}}^*)^{3/4} e^{-E_g/2k_B T} . \quad (10.1.26)$$

Der Verlauf der intrinsischen Ladungsträgerdichte ist in Abb. 10.8 für Si, Ge und GaAs dargestellt. Bei Raumtemperatur beträgt die intrinsische Ladungsträgerdichte von Ge ( $E_g = 0.67 \text{ eV}$ )  $2.4 \times 10^{13} \text{ cm}^{-3}$ , für Si ( $E_g = 1.12 \text{ eV}$ )  $1.5 \times 10^{10} \text{ cm}^{-3}$  und für GaAs ( $E_g = 1.42 \text{ eV}$ )  $5 \times 10^7 \text{ cm}^{-3}$ .

### 10.1.3 Dotierte Halbleiter

Die intrinsische Ladungsträgerdichte in Halbleitern ist für viele Anwendungen zu gering. Ladungsträgerdichten, die um mehrere Größenordnungen größer sind, können durch Einbringen von Verunreinigungen erzeugt werden. Für die meisten Halbleiter ist es sogar so, dass sie gar nicht in genügend reiner Form hergestellt werden können, um die intrinsische Ladungsträgerdichte beobachten zu können. Die geringste Verunreinigungskonzentration, die heute in Halbleitern erreicht werden kann, liegt im Bereich von  $10^{12} \text{ cm}^{-3}$ . Deshalb kann für Ge mit  $n_i = 2.4 \times 10^{13} \text{ cm}^{-3}$  bei Raumtemperatur Eigenleitung beobachtet werden, nicht aber für Si mit  $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$ . Die reinsten GaAs Einkristalle besitzen heute Verunreinigungskonzentrationen im Bereich von etwa  $10^{16} \text{ cm}^{-3}$ , was weit oberhalb der intrinsischen Ladungsträgerdichte von  $5 \times 10^7 \text{ cm}^{-3}$  liegt. Die gezielte Dotierung von Halbleitern, um deren Eigenschaften für Anwendungen maßzuschneidern, ist ein zentraler Bestandteil der heutigen Halbleitertechnologie. Im Gegensatz zu Halbleitern ist es für Metalle sehr schwierig, die elektrische Leitfähigkeit über mehrere Größenordnungen zu ändern.

#### 10.1.3.1 Donator- und Akzeptorniveaus

Elektrisch aktive Verunreinigungen in Halbleitern werden als *Donatoren*, wenn sie zusätzliche Elektronen im Leitungsband bereitstellen, bzw. als *Akzeptoren* bezeichnet, wenn sie zusätzliche Löcher im Valenzband bereitstellen. Für die Element-Halbleiter der IV. Hauptgruppe sind typische Donatoren Elemente der V. Hauptgruppe wie P, As oder Sb. Wird ein solches Donatoratom in einen Ge- oder Si-Kristall eingebaut, so werden von den fünf zur Verfügung stehenden Valenzelektronen nur vier für die kovalente Bindung benötigt. Das fünfte Elektron, welches nicht für die Bindung gebraucht wird, kann unter Aufwendung einer kleinen Energie vom Atomrumpf getrennt werden und steht dann als freies Elektron für den Ladungstransport zur Verfügung.

Um die Bindungsenergie des fünften Elektrons abzuschätzen, können wir das Bohrsche Atommodell benutzen und die Energieniveaus dieses Elektrons im Coulomb-Potenzial des einfach positiv geladenen Rumpfatoms berechnen. Dies entspricht formal gerade einem Wasserstoffatom. Für die Energieniveaus des Wasserstoffatoms gilt:

$$E_n^{\text{H}} = \frac{m_e e^4}{2(4\pi\epsilon_0\hbar)^2} \frac{1}{n^2} . \quad (10.1.27)$$

Für den Grundzustand ( $n = 1$ ) beträgt die Ionisierungsenergie 13.6 eV. Um von den Energieräumen des Wasserstoffatoms zu denjenigen des Donatoratoms zu gelangen müssen wir die freie Elektronenmasse  $m_e$  durch die effektive Bandmasse  $m_e^*$  eines Elektrons im Leitungsband ersetzen. Ferner müssen wir die Abschirmung des Coulomb-Potenzials durch die umgebenden Si-Atome berücksichtigen, indem wir im Ausdruck für die Energieniveaus des Wasserstoffatoms die Dielektrizitätskonstante  $\epsilon$  des Halbleitermaterials einsetzen. Es ergibt sich dann

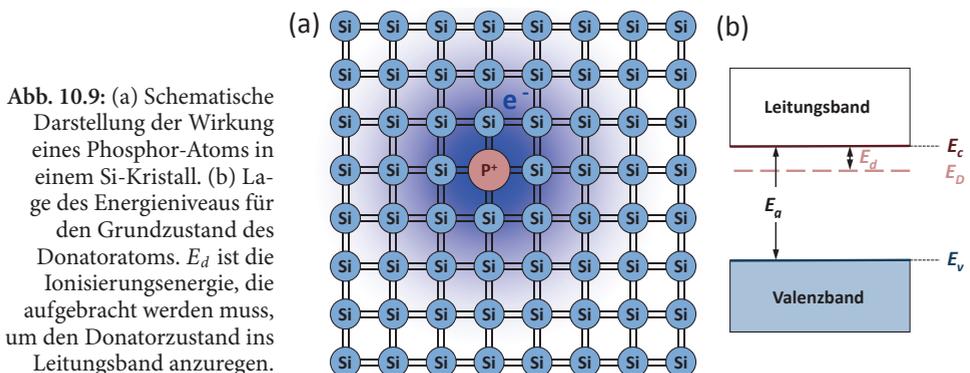
$$E_n = \frac{m_e^* e^4}{2(4\pi\epsilon\epsilon_0\hbar)^2} \frac{1}{n^2} = \frac{m_e^*}{m_e \epsilon^2} E_n^H. \quad (10.1.28)$$

Für Si erhalten wir mit  $m_e^* = (m_{et}^* m_{el})^{1/3} \simeq 0.3 m_e$  und  $\epsilon_{Si} = 11.7$  eine Ionisierungsenergie von  $E_1 \sim 30$  meV. Das heißt, der Donatorzustand  $E_D$  befindet sich nur  $E_d = E_c - E_D \simeq 30$  meV unterhalb der Leitungsbandkante, die wir mit dem Vakuumniveau des Wasserstoffatoms identifizieren. Da  $E_d$  in der Größenordnung der thermischen Energie bei Raumtemperatur ( $\sim 25$  meV) liegt, kann der Donatorzustand leicht thermisch ionisiert werden. Für Ge ist  $m_e^* = (m_{et}^* m_{el})^{1/3} \simeq 0.2 m_e$  und  $\epsilon_{Ge} = 15.8$ , so dass hier die Ionisierungsenergie mit  $E_1 \sim 10$  meV noch kleiner ist.

In Abb. 10.9 ist die Bandstruktur von Si schematisch zusammen mit dem Grundzustandsniveau ( $n = 1$ ) des Donatoratoms gezeigt. Zwischen dem Grundzustand und der Unterkante des Leitungsbandes befinden sich die angeregten Zustände ( $n > 1$ ) des Donatoratoms, deren Abstände mit zunehmender Quantenzahl stark abnehmen. Die Energieniveaus der angeregten Zustände können z. B. mit optischer Spektroskopie bestimmt werden. Tut man dies, so erkennt man, dass die Energieniveaus doch erheblich von denjenigen eines Wasserstoffspektrums abweichen. Dies wird durch das so genannte Kristallfeld der umgebenden Si-Atome verursacht, das zu einer Aufhebung der Entartung der wasserstoffähnlichen Zustände führt.

Wir können mit Hilfe des Wasserstoffmodells eines Donatoratoms auch den Bohrschen Radius berechnen. In Analogie zum Wasserstoffatom erhalten wir

$$r_d = \frac{4\pi\epsilon\epsilon_0\hbar^2}{m_e^* e^2}. \quad (10.1.29)$$



Wir sehen, dass der Bohrsche Radius des Donatorzustands um den Faktor  $\epsilon m_e / m_e^*$  größer als der Bohrsche Radius  $a_B = 0.525 \text{ \AA}$  eines Wasserstoffatoms ist. Er beträgt für Si etwa  $30 \text{ \AA}$  und ist damit wesentlich größer als der Atomabstand der Siliziumatome von  $2.35 \text{ \AA}$ . Das bedeutet, dass das an das Donatoratom gebundene Elektron mehr als 10 Gitterabstände ausgeschmiert ist. Dies rechtfertigt die Verwendung der Dielektrizitätskonstante von Silizium zur Berücksichtigung der Abschirmung.

Verwenden wir als Verunreinigungsatom in einem Si oder Ge-Kristall ein dreiwertiges Element wie z. B. B, Al, Ga oder In, so fehlt diesem für die tetraedrische Bindung ein Elektron. Das heißt, das Dotieratom kann sehr einfach ein Elektron aufnehmen, das dann in seiner Umgebung fehlt. Dieses fehlende Elektron können wir formal als Loch beschreiben, welches das einfach negativ geladene Akzeptoratom umkreist. Dies ist in Abb. 10.10 schematisch dargestellt.

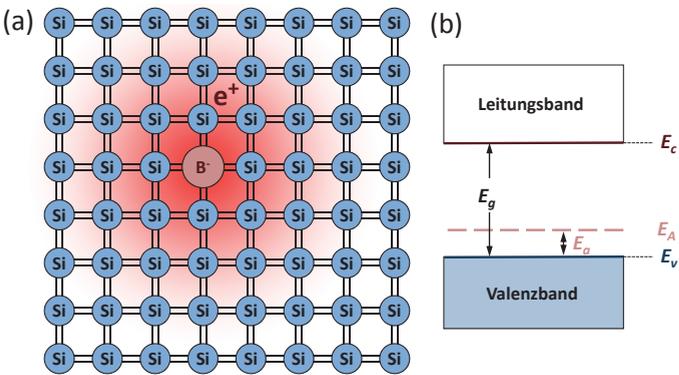


Abb. 10.10: (a) Schematische Darstellung der Wirkung eines Bor-Atoms in einem Si-Kristall. (b) Lage des Energieniveaus für den Grundzustand des Akzeptoratoms.  $E_a$  ist die Ionisierungsenergie, die aufgebracht werden muss, um den Akzeptorzustand ins Valenzband anzuregen.

Die Beschreibung der Akzeptorniveaus können wir unter Verwendung des Lochkonzepts formal gleich vornehmen. Wir müssen nur in Gleichung (10.1.28) die effektive Masse  $m_h^*$  eines Lochs im Valenzband verwenden. Eine Ionisierung des Akzeptoratoms ist gleichbedeutend mit der Freisetzung eines Loches. Hierzu muss ein Elektron aus dem Valenzband in ein Akzeptorniveau  $E_A$  angehoben werden. Das Akzeptorniveau liegt also oberhalb der Valenzbandkante, wobei sein Abstand von dieser Kante gerade der Ionisierungsenergie  $E_a$  entspricht (siehe Abb. 10.10). Der Wert von  $E_a$  liegt in der gleichen Größenordnung wie der von  $E_d$ . In Tabelle 10.4 sind die Ionisierungsenergien einiger Donator- und Akzeptoratomme für Si und Ge zusammengestellt.

Tabelle 10.4: Ionisierungsenergien  $E_d$  und  $E_a$  einiger Donatoren und Akzeptoren in Si und Ge. Quelle: *Handbook Series on Semiconductor Parameters*, Vol. 1 and 2, edited by M. Levinstein, S. Romyantsev and M. Shur, World Scientific, London (1996, 1999).

Halbleiter	Donatoren					Akzeptoren			
	P	As	Sb	Bi		B	Al	Ga	In
$E_d$	(meV)	(meV)	(meV)	(meV)	$E_a$	(meV)	(meV)	(meV)	(meV)
Si	45	54	43	69		45	72	74	157
Ge	13	14	9.6	13		11	11	11	12

Enthält ein Halbleiter sowohl Donatoren als auch Akzeptoren, so kann das Elektron eines Donators zu einem Akzeptor wandern und dort das fehlende Elektron ersetzen. Dadurch heben sich die Wirkung von Donatoren und Akzeptoren gegenseitig auf, wir sprechen von kompensierten Halbleitern. Überwiegt die Anzahl der Donatoren, so sprechen wir von einem *n-Halbleiter*, überwiegen die Akzeptoren, so sprechen wir von einem *p-Halbleiter*.

### 10.1.3.2 Ladungsträgerdichte und Fermi-Niveau

In einem dotierten Halbleiter kann ein Elektron im Leitungsband entweder aus dem Valenzband oder einem Donatorniveau stammen. Ein Loch im Valenzband entspricht damit entweder einem Elektron im Leitungsband oder einem ionisierten Akzeptorniveau. Zur Ermittlung der Ladungsträgerkonzentration in dotierten Halbleitern können wir nach wie vor die Ausdrücke (10.1.22) und (10.1.23) für  $n_c$  und  $p_v$  verwenden. Bei ihrer Herleitung haben wir lediglich vorausgesetzt, dass der Abstand des chemischen Potentials von der Leitungs- oder Valenzbandkante groß gegen  $k_B T$  sein soll. Das heißt, es gilt nach wie vor die dem Massenwirkungsgesetz entsprechende Beziehung

$$n_c \cdot p_v = n_c^{\text{eff}} p_v^{\text{eff}} e^{-E_g/k_B T}, \quad (10.1.30)$$

in der das chemische Potential  $\mu$  nicht mehr auftaucht. Im Gegensatz zu intrinsischen Halbleitern können wir jetzt aber die Lage des chemischen Potentials nicht mehr durch die einfache Neutralitätsbedingung  $n_c = p_v$  herleiten, sondern müssen eine kompliziertere Bedingung verwenden, die die Ladung der Verunreinigungen mit berücksichtigt.

Die Dichte der Donatoren lässt sich als

$$n_D = n_D^0 + n_D^+ \quad (10.1.31)$$

schreiben, wobei  $n_D^0$  die Dichte der neutralen und  $n_D^+$  diejenige der ionisierten Donatoren ist. Äquivalent gilt für die Akzeptoren

$$n_A = n_A^0 + n_A^- . \quad (10.1.32)$$

Da insgesamt Ladungsneutralität vorliegen muss, erhalten wir die Bedingung

$$n_c + n_A^- = p_v + n_D^+ . \quad (10.1.33)$$

Um diese Neutralitätsbedingung für die Bestimmung der Lage des chemischen Potentials verwenden zu können, benötigen wir noch Ausdrücke für  $n_A^-$  und  $n_D^+$ .

Die Wahrscheinlichkeit dafür, dass eine Störstelle nicht ionisiert ist, das heißt, die Wahrscheinlichkeit  $n_D^0/n_D$  bzw.  $n_A^0/n_A$  dafür, dass sie als neutrale Störstelle vorliegt, können wir mit Hilfe der Fermi-Dirac-Verteilung angeben:

$$\frac{n_D^0}{n_D} = 2 \frac{1}{e^{(E_D - \mu)/k_B T} + 1} \quad (10.1.34)$$

$$\frac{n_A^0}{n_A} = 4 \frac{1}{e^{(\mu - E_A)/k_B T} + 1} . \quad (10.1.35)$$

Hierbei berücksichtigt der Faktor 2 der rechten Seite die zweifache Spin-Entartung des Donatorzustands, d. h. effektiv liegt die Donatordichte  $2n_D$  vor. Der Faktor 4 resultiert aus den beiden  $hh$ - und  $lh$ -Valenzbandzuständen mit jeweils zweifacher Spin-Entartung, aus denen die Akzeptorzustände aufgebaut sind, d. h. effektiv liegt die Akzeptordichte  $4n_A$  vor. Bei der nachfolgenden Abschätzung vernachlässigen wir diese zusätzlichen Faktoren. Mit Hilfe von (10.1.34) und (10.1.35) können wir aus (10.1.31) und (10.1.32) Ausdrücke für  $n_D^+$  und  $n_A^-$  ableiten und in die Neutralitätsbedingung (10.1.33) einsetzen. Wir erhalten dadurch eine Bestimmungsgleichung für  $\mu$ , die allerdings bei gleichzeitiger Berücksichtigung von Donatoren und Akzeptoren nur numerisch lösbar ist.

Wir beschränken uns hier auf die Diskussion eines  $n$ -Halbleiters mit  $n_D \gg n_A$ . Entsprechende Ergebnisse können in analoger Weise für  $p$ -Halbleiter erhalten werden. Für  $n_D \gg n_A$  fangen alle Akzeptoren ein Elektron ein, das ursprünglich zu einem Donator gehörte. Das heißt, es kommen praktisch keine neutralen Akzeptoren vor, so dass wir  $n_A^0 \simeq 0$  und  $n_A^- \simeq n_A$  setzen können. Außerdem soll die Temperatur so niedrig sein, dass wir die Zahl der Elektronen, die aus dem Valenzband ins Leitungsband angeregt werden, gegenüber den durch Ionisierung von Donatoratomen erzeugten Elektronen vernachlässigen können, d.h.  $n_D^+ \gg n_c, p_v$ . Aus (10.1.33) folgt dann

$$n_c = p_v + n_D^+ - n_A^- \simeq n_D^+ - n_A = n_D - n_D^0 - n_A. \quad (10.1.36)$$

Benutzen wir jetzt noch (10.1.34), so erhalten wir

$$n_c \simeq n_D \left( 1 - \frac{1}{e^{(E_D - \mu)/k_B T} + 1} \right) - n_A. \quad (10.1.37)$$

Durch Umformen von

$$n_c = 2 \left( \frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-(E_c - \mu)/k_B T} = n_c^{\text{eff}} e^{-(E_c - \mu)/k_B T} \quad (10.1.38)$$

erhalten wir

$$\frac{n_c}{n_c^{\text{eff}}} e^{E_c/k_B T} = e^{\mu/k_B T} \quad (10.1.39)$$

und können damit in (10.1.37) das chemische Potenzial eliminieren. Wir erhalten

$$n_c \simeq n_D \left( 1 - \frac{n_c}{n_c^{\text{eff}} e^{-E_d/k_B T} + n_c} \right) - n_A = \frac{n_D}{1 + \frac{n_c}{n_c^{\text{eff}}} e^{E_d/k_B T}} - n_A. \quad (10.1.40)$$

Hierbei haben wir den Abstand  $E_d = E_c - E_D$  des Donatorniveaus vom Leitungsband benutzt. Durch weiteres Umformen von (10.1.40) erhalten wir das Ergebnis

$$\frac{n_c(n_c + n_A)}{n_D - n_A - n_c} = n_c^{\text{eff}} e^{-E_d/k_B T}, \quad (10.1.41)$$

aus dem wir den Temperaturverlauf von  $n_c$  bestimmen können. Für eine verschwindend kleine Akzeptorkonzentration  $n_A$  ergibt sich näherungsweise die quadratische Gleichung

$$\frac{n_c^2}{n_c^{\text{eff}}} e^{E_d/k_B T} + n_c - n_D \simeq 0 \quad (10.1.42)$$

mit der Lösung

$$n_c = \frac{2n_D}{1 + \sqrt{1 + 4 \frac{n_D}{n_c^{\text{eff}}} e^{E_d/k_B T}}} \quad (10.1.43)$$

Wir wollen im Folgenden die physikalische Bedeutung von (10.1.41) für verschiedene Temperaturbereich diskutieren.

- Sehr tiefe Temperaturen,  $k_B T \lll E_d$  (**Kompensationsbereich**):

Bei sehr tiefen Temperaturen sind so wenige Ladungsträger angeregt, dass  $n_c \ll n_A \ll n_D$  gilt. Damit ergibt sich aus (10.1.41)

$$n_c \simeq \frac{n_D n_c^{\text{eff}}}{n_A} e^{-E_d/k_B T}. \quad (10.1.44)$$

Setzen wir dies in (10.1.39) ein, so erhalten wir für die Lage des chemischen Potentials

$$\mu \simeq E_c - E_d + k_B T \ln \left( \frac{n_D}{n_A} \right). \quad (10.1.45)$$

Wir sehen, dass für  $T \rightarrow 0$  die Lage von  $\mu$  durch die Donatoren bestimmt wird. Durch die endliche Akzeptordichte werden alle von den Donatoren abgegebenen Elektronen von den Akzeptoren aufgenommen. Wir sprechen von einem Kompensationseffekt. Da mit zunehmender Temperatur die Donatoren auch Ladungsträger ins Leitungsband abgeben können, bewegt sich, wie Abb. 10.11 zeigt, für zunehmende Temperatur  $\mu$  von  $E_c - E_d$  nach oben in Richtung Leitungsbandkante. Gleichzeitig nimmt die Ladungsträgerdichte exponentiell zu.

- Tiefe Temperaturen,  $k_B T \ll E_d$  (**Bereich der Störstellenreserve**):

Wird die Temperatur weiter erhöht, so können immer mehr Donatoren ihre Ladungsträger ins Leitungsband abgeben, so dass schnell  $n_c \gg n_A$  gilt. Die Temperatur ist aber immer noch so gering, dass  $e^{-E_d/k_B T} \ll 1$  und somit  $n_c \ll n_D$ . Mit diesen Näherungen ergibt sich aus (10.1.41)

$$n_c \simeq \sqrt{n_D n_c^{\text{eff}}} e^{-E_d/2k_B T}. \quad (10.1.46)$$

Für die Lage des chemischen Potentials ergibt sich

$$\mu \simeq E_c - \frac{E_d}{2} - \frac{k_B T}{2} \ln \left( \frac{n_c^{\text{eff}}}{n_D} \right). \quad (10.1.47)$$

Das chemische Potenzial liegt also etwa in der Mitte zwischen Leitungsbandunterkante und Donatorniveau.

Gleichung (10.1.46) zeigt, dass im Temperaturbereich der Störstellenreserve die thermische Energie noch nicht ausreicht, um alle Donatoren zu ionisieren. Wie Abb. 10.11 zeigt, nimmt mit sinkender Temperatur die Ladungsträgerdichte exponentiell ab, wie wir es schon bei intrinsischen Halbleitern kennen gelernt haben. Wir sprechen von einem Ausfrieren der Ladungsträger. Allerdings ist im Vergleich zu intrinsischen Halbleitern die exponentielle Abnahme jetzt durch  $E_d$  und nicht durch  $E_g$  bestimmt. Wir sehen ferner,

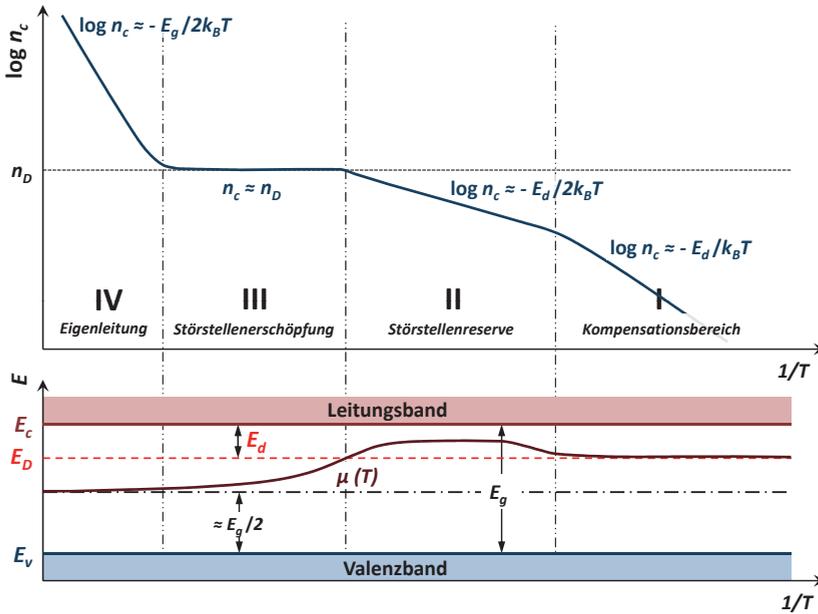


Abb. 10.11: Temperaturverlauf der Ladungsträgerdichte  $n_c$  und des chemischen Potentials  $\mu$  in einem dotierten  $n$ -Typ Halbleiter. Im Bereich I liegt Störstellenkompensation durch eine endliche Akzeptordichte vor, im Bereich II dominiert reine Störstellenleitung, im Bereich III der Störstellenerschöpfung sind sämtliche Störstellen ionisiert, so dass die Ladungsträgerdichte etwa konstant bleibt, und im Bereich IV tritt die Eigenleitung gegenüber der Störstellenleitung in den Vordergrund.

dass (10.1.46) formal dem Ausdruck (10.1.26) für die intrinsische Ladungsträgerdichte entspricht, wenn wir  $E_g$  durch  $E_d$  und die Zustandsdichte im Leitungsband durch  $n_D$  ersetzen. Die thermische Anregung von Ladungsträgern direkt vom Valenz- ins Leitungsband kann im Bereich der Störstellenreserve vollkommen vernachlässigt werden.

■ **Mittlere Temperaturen,  $k_B T \gtrsim E_d$  (Bereich der Störstellenerschöpfung):**

Bei genügend hohen Temperaturen  $k_B T \gtrsim E_d$  wird  $e^{-E_d/k_B T} \approx 1$  und wir erhalten mit  $n_c \gg n_A$  aus (10.1.41) die Beziehung  $n_c^2 \approx n_c^{\text{eff}}(n_D - n_c)$ . Da  $n_c \ll n_c^{\text{eff}}$ , folgt weiter  $n_D - n_c \approx 0$  und damit das einfache Ergebnis

$$n_c \approx n_D \tag{10.1.48}$$

und

$$\mu \approx E_c - k_B T \ln \left( \frac{n_c^{\text{eff}}}{n_D} \right). \tag{10.1.49}$$

Das Ergebnis ist einfach zu verstehen. Die Temperatur ist hoch genug, um alle Störstellen zu ionisieren, aber noch zu klein, um eine große Zahl von Ladungsträgern aus dem Valenz- ins Leitungsband anzuregen. Wir sprechen deshalb vom Bereich der Störstellenerschöpfung. Das chemische Potenzial bewegt sich mit zunehmender Temperatur nach unten in Richtung Bandmitte. Es sei hier noch erwähnt, dass sich für Si mit einer Phos-

phor-Dotierung von  $n_D = 3 \times 10^{14} \text{ cm}^{-3}$  der Bereich der Störstellenschöpfung von etwa 45 bis 500 K erstreckt. Das heißt, bei Raumtemperatur sind alle Donatoratome ionisiert.

■ Hohe Temperaturen,  $k_B T \gg E_d$  (**Bereich der Eigenleitung**):

Wenn wir zu noch höheren Temperaturen gehen, müssen wir natürlich wieder die Dichte der thermisch direkt aus dem Valenzband ins Leitungsband angeregten Ladungsträger berücksichtigen, sobald diese in die Größenordnung von  $n_D$  kommt. Die für (10.1.33) gemachte Annahme  $n_D^+ \gg n_c, p_v$  ist dann nicht mehr zulässig.

In dem Bereich sehr hoher Temperaturen verläuft die Temperaturabhängigkeit der Ladungsträgerdichte wieder wie bei intrinsischen Halbleitern. Es dominiert die Eigenleitung. Für die Ladungsträgerdichte und das chemische Potenzial gelten Gleichungen (10.1.22) und (10.1.24).

## 10.1.4 Elektrische Leitfähigkeit

Um die Temperaturabhängigkeit der elektrischen Leitfähigkeit zu diskutieren, müssen wir neben der Temperaturabhängigkeit der Ladungsträgerdichte auch noch etwas über die Temperaturabhängigkeit der Streuzeit  $\tau_e$  und  $\tau_p$  oder äquivalent der Beweglichkeiten  $\mu_e$  und  $\mu_p$  der Elektronen und Löcher wissen. Zum elektrischen Strom tragen sowohl die Elektronen im Leitungsband als auch die Löcher im Valenzband bei, so dass wir für die elektrische Stromdichte

$$\mathbf{J}_q = e(n_c \mu_e + p_v \mu_p) \mathbf{E} = \sigma \mathbf{E} \quad (10.1.50)$$

schreiben können. Im Gegensatz zu Metallen, wo wir nur Elektronen bei der Fermi-Energie berücksichtigen mussten, für deren Beweglichkeit  $\mu_e = \mu_e(E_F)$  gilt, sind die Beweglichkeiten  $\mu_e$  und  $\mu_p$  in einem Halbleiter die Mittelwerte für die Zustände, die Elektronen und Löcher im Leitungsband bzw. Valenzband besetzen. Diese Mittelwerte können wie folgt ausgedrückt werden:

$$\mu_e = \frac{e}{m_e^*} \frac{\langle \tau_e(\mathbf{k}) v^2(\mathbf{k}) \rangle}{\langle v^2(\mathbf{k}) \rangle}. \quad (10.1.51)$$

Hierbei ist  $v(\mathbf{k})$  die Geschwindigkeit eines Elektrons im Zustand  $\mathbf{k}$  und  $\tau_e(\mathbf{k})$  seine Streuzeit. Für  $\mu_h$  gilt ein entsprechender Ausdruck.

Wir wollen im Folgenden nur eine qualitative Betrachtung durchführen und auf eine detaillierte Behandlung mit Hilfe der Boltzmann-Transportgleichung verzichten.<sup>11</sup> Nach erheblicher Vereinfachung ergibt (10.1.51), dass die Beweglichkeit proportional zu einer mittleren Streuzeit  $\tau$  ist, d. h.  $\mu_e = e\tau_e/m_e^*$  bzw.  $\mu_h = e\tau_h/m_h^*$ . Da  $\tau$  die mittlere Zeit zwischen zwei Streuprozessen ist, können wir schreiben (vergleiche hierzu Abschnitt 9.3):

$$\frac{1}{\ell} = \frac{1}{\langle v \rangle \tau} \propto S. \quad (10.1.52)$$

<sup>11</sup> Eine detailliertere Betrachtung findet man zum Beispiel in *Fundamentals of Semiconductors*, P. Y. Yu, M. Cardona, Springer Verlag, Berlin (1996).

Hierbei ist  $S$  der Streuquerschnitt eines Streuzentrums für Elektronen bzw. Löcher. Im Gegensatz zu Metallen, wo wir  $v = v_F$  verwenden können, müssen wir für  $\langle v \rangle$  den thermischen Mittelwert über alle Elektronen im Leitungsband bzw. Löcher im Valenzband benutzen.<sup>12</sup> Gleichung (10.1.52) gibt dann die mittlere freie Weglänge von Elektronen bzw. Löchern an. Da wir für Halbleiter eine Boltzmann-Statistik verwenden dürfen, gilt für die mittlere Geschwindigkeit  $\langle v \rangle = \sqrt{3k_B T/m^*}$  und wir erhalten

$$\langle v \rangle \propto \sqrt{T}. \tag{10.1.53}$$

Wir wollen im Folgenden verschiedene Streuprozesse diskutieren.

**Streuung an akustischen Phononen:** Wir beginnen mit der Streuung an akustischen Phononen. Wir haben in Abschnitt 9.3.2 abgeleitet, dass

$$S_{\text{ph}} \propto T \quad \text{für} \quad T \gg \Theta_D. \tag{10.1.54}$$

Aus (10.1.52) und (10.1.53) erhalten wir dann [vergleiche hierzu (9.3.28)]

$$\mu_{\text{ph}} \propto T^{-3/2}. \tag{10.1.55}$$

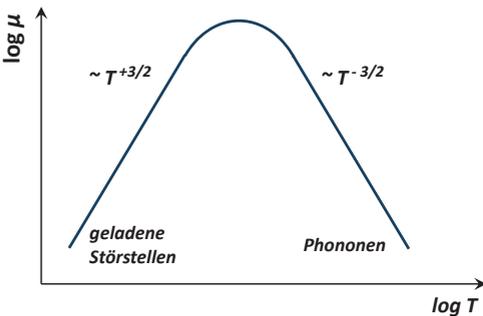
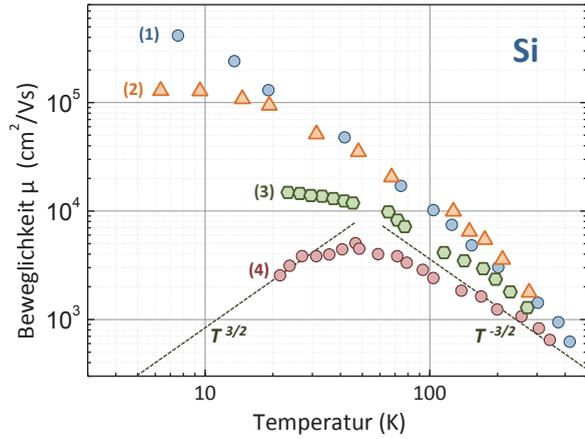


Abb. 10.12: Qualitativer Temperaturverlauf der Beweglichkeit in einem Halbleiter, in dem hauptsächlich Streuung an akustischen Phononen und geladenen Verunreinigungen vorliegt.

**Streuung an optischen Phononen:** In polaren Kristallen mit ionischem Bindungsanteil (z. B. GaAs) spielt die Streuung von Ladungsträgern an longitudinal-optischen (LO) Phononen eine große Rolle. Diese Phononen erzeugen lokale elektrische Felder, an welche die Ladungsträger ankoppeln können und zur Streuung führen. Ist die Energie der Ladungsträger groß gegen die Phononen-Energie,  $E(\mathbf{k}) \gg \hbar\Omega_{\text{LO}}$ , ergibt sich  $\mu \propto T^{-1/2} \exp(-\hbar\Omega_{\text{LO}}/k_B T)$ , während sich im anderen Grenzfall,  $E(\mathbf{k}) \ll \hbar\Omega_{\text{LO}}$ , in etwa  $\mu \propto \exp(-\hbar\Omega_{\text{LO}}/k_B T)$  ergibt. Für eine Herleitung dieser Zusammenhänge wird auf die Fachliteratur verwiesen.

<sup>12</sup> Für Halbleiter mit nicht allzu hoher Dotierung und bei nicht allzu tiefen Temperaturen gilt  $k_B T > E_F$ . Es liegt deshalb kein entartetes Fermi-Gas wie in Metallen sondern ein klassisches Teilchengas vor, dessen Geschwindigkeitsverteilung wir mit der Maxwell-Boltzmann-Verteilung beschreiben können.



**Abb. 10.13:** Temperaturabhängigkeit der Beweglichkeit der Ladungsträger in Si. (1) Hochreines Si,  $n_D < 10^{12} \text{ cm}^{-3}$ ; (2) hochreines Si,  $n_D < 4 \times 10^{13} \text{ cm}^{-3}$ ; (3)  $n_D = 1.75 \times 10^{16} \text{ cm}^{-3}$ ,  $n_A = 1.48 \times 10^{15} \text{ cm}^{-3}$ ; (4)  $n_D = 1.3 \times 10^{17} \text{ cm}^{-3}$ ,  $n_A = 2.2 \times 10^{15} \text{ cm}^{-3}$ . Quellen: P. Norton *et al.*, Phys. Rev. B 8, 5632 (1973); C. Canali *et al.*, Phys. Rev. B 12, 2265 (1975).

**Streuung an geladenen Störstellen:** Als nächstes wollen wir die Streuung an geladenen Defekten diskutieren. Da nichtionisierte Donatoren und Akzeptoren solche Defekte darstellen, ist dieser Streuprozess für Halbleiter insbesondere bei tiefen Temperaturen sehr wichtig. Für dieses Rutherford-artige Streuproblem erhalten wir<sup>13</sup>

$$S_{\text{def}} \propto \langle v \rangle^{-4} \propto T^{-2}, \tag{10.1.56}$$

wobei wir den thermischen Mittelwert  $\langle v \rangle \propto \sqrt{T}$  für die Geschwindigkeit benutzt haben. Der Zusammenhang  $S_{\text{def}} \propto \langle v \rangle^{-4}$  ergibt sich sofort, wenn wir uns klar machen, dass der Streuquerschnitt in einfachster Näherung eine kreisförmige Scheibe ist, deren Radius die Entfernung  $r$  von dem geladenen Streuzentrum ist, bei der die Coulomb-Energie ( $\propto 1/r$ ) gerade gleich der kinetischen Energie ( $\propto \langle v \rangle^2$ ) der Ladungsträger ist. Daraus ergibt sich der Streuquerschnitt  $S_{\text{def}} \propto r^2 \propto \langle v \rangle^{-4}$ . Aus (10.1.52) und (10.1.53) erhalten wir dann wiederum [vergleiche hierzu auch (9.3.16)]

$$\mu_{\text{def}} \propto T^{3/2}. \tag{10.1.57}$$

**Mehrere Streuprozesse:** Nach der Matthiessen-Regel erhalten wir  $1/\mu_{\text{tot}}$ , indem wir die reziproken Mobilitäten addieren. Das Ergebnis ist in Abb. 10.12 schematisch dargestellt.

<sup>13</sup> Für die Streuung an einer geladenen Störstelle mit  $Z = 1$  erhalten wir den Wirkungsquerschnitt [vergleiche (9.3.12) und (9.3.14)]

$$\sigma(v, \vartheta) = \left( \frac{e^2}{2\epsilon\epsilon_0 m^* \langle v \rangle^2} \right)^2 \sin^{-4}(\vartheta/2),$$

woraus sich eine vom Streuwinkel  $\vartheta$  abhängige Streurrate  $1/\tau(\vartheta) \propto \langle v \rangle \sigma(\vartheta)$  ergibt. Die gesamte Streurrate erhalten wir durch Integration über alle Streuwinkel. Das Ergebnis dieser Integration ist die *Conwell-Weißkopf-Formel*

$$\mu \propto \frac{1}{\tau} = \frac{2n_{D,A} \pi e^4}{(\epsilon\epsilon_0 m^*)^2 \langle v \rangle^3} \ln \left[ 1 + \left( \frac{\epsilon\epsilon_0 m^* \langle v \rangle^2 (3/4 \pi n_{D,A})^{1/3}}{2e^2} \right)^2 \right].$$

Hierbei ist  $n_{D,A}$  die Dichte der geladenen Donatoren bzw. Akzeptoren.

**Tabelle 10.5:** Beweglichkeiten von einigen Halbleitern bei 300 K. Quelle: *Handbook Series on Semiconductor Parameters*, Vol. 1 and 2, edited by M. Levinstein, S. Rumyantsev and M. Shur, World Scientific, London (1996, 1999).

Halbleiter	Si	Ge	C	GaAs	InAs	InSb	InP
$\mu_e$ (cm <sup>2</sup> /Vs)	1 400	3 900	1 800	8 500	40 000	77 000	4 500
$\mu_h$ (cm <sup>2</sup> /Vs)	500	1 900	1 400	400	500	850	100

Die Beweglichkeit nimmt bei tiefen Temperaturen mit ansteigender Temperatur zunächst proportional zu  $T^{3/2}$  zu, durchläuft ein Maximum und nimmt dann proportional zu  $T^{-3/2}$  ab. Dies ist am Beispiel von Si in Abb. 10.13 gezeigt. In polaren Halbleitern nimmt die Beweglichkeit bei hohen Temperaturen aufgrund der hier dominierenden Streuung an LO-Phononen deutlich stärker ab. Als weitere Streuprozesse kommen die Streuung an neutralen Störstellen und die Streuung an transversal-akustischen (TA) Phononen hinzu, die in Viertal-Halbleitern wie Si und Ge wichtig ist. Diese sind aber in den meisten Fällen nicht dominant und werden deshalb hier nicht näher erläutert. Die Beweglichkeiten von einigen Halbleitern bei 300 K sind in Tabelle 10.5 aufgelistet.

Der in Abb. 10.12 gezeigte charakteristische Verlauf der Beweglichkeit spiegelt sich auch häufig in der elektrischen Leitfähigkeit  $\sigma$  wider, da die Ladungsträgerdichte in dem Bereich der Ladungsträgersättigung über ein großes Temperaturintervall konstant bleibt. Nur bei sehr hohen und sehr tiefen Temperaturen dominiert die exponentielle Temperaturabhängigkeit der Ladungsträgerdichte (vergleiche Abb. 10.11) die Leitfähigkeit.

### 10.1.5 Hall-Effekt

Bei Halbleitern erfolgt der Ladungstransport sowohl durch die Elektronen im Leitungsband als auch durch die Löcher im Valenzband. Für den Hall-Koeffizienten müssen wir deshalb den Zweiband-Ausdruck [vergleiche (9.6.18)]

$$R_H = \frac{\sigma_1 \frac{e\tau_1}{m_1} + \sigma_2 \frac{e\tau_2}{m_2}}{(\sigma_1 + \sigma_2)^2} . \quad (10.1.58)$$

verwenden, den wir in Abschnitt 9.6.2 für ein Zweiband-Modell hergeleitet haben. Hierbei müssen wir berücksichtigen, dass die Elektronen die Ladung  $-e$  und die Löcher die Ladung  $+e$  transportieren. Unter Benutzung von  $\sigma_1 = n_c e \mu_e$  und  $\sigma_2 = p_v e \mu_h$ , sowie  $\mu_e = e\tau_e/m_e^*$  und  $\mu_h = e\tau_h/m_h^*$  erhalten wir den Ausdruck

$$R_H = \frac{p_v \mu_h^2 - n_c \mu_e^2}{e(p_v \mu_h + n_c \mu_e)^2} . \quad (10.1.59)$$

Bei reiner Eigenleitung ist  $n_c = p_v = n_i$  und für die Hall-Konstante ergibt sich

$$R_{H,i} = \frac{1}{n_i e} \frac{\mu_h - \mu_e}{\mu_h + \mu_e} . \quad (10.1.60)$$

Wir sehen, dass die Hall-Konstante positiv oder negativ sein kann, je nachdem ob  $\mu_h > \mu_e$  oder  $\mu_h < \mu_e$ .

Bei reiner Störstellenleitung können wir jeweils eine Ladungsträgersorte vernachlässigen und es ergibt sich aus (10.1.59)

$$R_{H,e} = -\frac{1}{n_c e} \quad \text{oder} \quad R_{H,h} = +\frac{1}{p_v e}, \quad (10.1.61)$$

je nachdem ob reine  $n$ -Leitung ( $p_v = 0$ ) oder reine  $p$ -Leitung ( $n_c = 0$ ) vorliegt. Dieser Ausdruck entspricht dem bereits bekannten Ergebnis, das wir bei Vorliegen nur einer Ladungsträgersorte (Einband-Modell) abgeleitet haben (vergleiche hierzu Abschnitt 7.3.4).

Wir wollen nun noch diskutieren, wie wir durch Messung von  $R_H$  wichtige Parameter von Halbleitern bestimmen können:

- Zur Bestimmung der Energielücke  $E_g$  messen wir  $R_{H,i}$  als Funktion der Temperatur im Bereich hoher Temperaturen, wo die Eigenleitung dominiert. Für die Temperaturabhängigkeit der intrinsischen Ladungsträgerdichte gilt

$$n_i(T) \propto T^{3/2} e^{-E_g/2k_B T}. \quad (10.1.62)$$

Der Term  $\frac{\mu_h - \mu_e}{\mu_h + \mu_e}$  in (10.1.60) zeigt keine Temperaturabhängigkeit, da sich diese durch die Quotientenbildung heraushebt. Wir erhalten dann

$$\ln(|R_{H,i}| T^{3/2}) = \text{const} + \frac{E_g}{2k_B} \frac{1}{T}. \quad (10.1.63)$$

Tragen wir also  $\ln(|R_{H,i}| T^{3/2})$  gegen  $1/T$  auf, so erhalten wir eine Gerade mit der Steigung  $E_g/2k_B$ .

- Zur Bestimmung der Ionisierungsenergie  $E_d$  in einem  $n$ -Halbleiter mit Hilfe des Hall-Effekts müssen wir den Hall-Effekt in dem Temperaturbereich messen, der durch das Ausfrieren der Ladungsträger dominiert wird (Störstellenreserve). Wir dürfen für  $n_c$  dann den Ausdruck (10.1.46)

$$n_c \simeq \sqrt{n_D n_c^{\text{eff}}} e^{-E_d/2k_B T} \propto T^{3/4} e^{-E_d/2k_B T} \quad (10.1.64)$$

verwenden und erhalten

$$\ln(|R_{H,e}| T^{3/4}) = \text{const} + \frac{E_d}{2k_B} \frac{1}{T}. \quad (10.1.65)$$

Tragen wir wiederum  $\ln(|R_{H,e}| T^{3/4})$  gegen  $1/T$  auf, so erhalten wir eine Gerade mit der Steigung  $E_d/2k_B$ . Eine analoge Betrachtung gilt für die Bestimmung der Ionisierungsenergie  $E_a$  in einem  $p$ -Halbleiter.

- In einem  $n$ -Halbleiter, der keine Akzeptoren enthält, ist in einem weiten Temperaturbereich die Ladungsträgerdichte  $n_c = n_D = \text{const}$  (Bereich der Störstellenschöpfung). Nach (10.1.61) gilt für diesen Bereich dann

$$n_D = -\frac{1}{R_{H,e} e}. \quad (10.1.66)$$

Für einen  $p$ -Halbleiter ohne Donatoren gilt entsprechend

$$n_A = +\frac{1}{R_{H,h} e}. \quad (10.1.67)$$

- Die Beweglichkeiten  $\mu_e$  und  $\mu_h$  hängen über die Streuzeiten  $\tau_e$  und  $\tau_h$  von der Temperatur ab. Für den Temperaturbereich, in dem reine Eigenleitung vorliegt, erhält man die Beweglichkeiten durch eine kombinierte Messung von  $R_{H,e}$  bzw.  $R_{H,h}$  und  $\sigma$ . Aus (10.1.61) folgt

$$\mu_e = R_{H,e}\sigma \quad \text{und} \quad \mu_h = R_{H,h}\sigma, \quad (10.1.68)$$

Bei reiner Eigenleitung gilt ferner

$$\sigma = e(n_c\mu_e + p_v\mu_p), \quad (10.1.69)$$

woraus sich mit Hilfe von (10.1.60) die Beziehung

$$R_{H,i}\sigma = \mu_h - \mu_e \quad (10.1.70)$$

ergibt. Um aus den beiden Gleichungen (10.1.69) und (10.1.70) die Beweglichkeiten  $\mu_e$  und  $\mu_h$  zu berechnen, benötigen wir außer den gemessenen Größen  $R_{H,i}$  und  $\sigma$  noch die Elektronendichte  $n_i$  bei Eigenleitung. Diese können wir nach (10.1.26) berechnen, wenn wir neben der Energielücke  $E_g$  noch die effektiven Massen  $m_e^*$  und  $m_h^*$  kennen. Letztere können z. B. mit Hilfe der Zyklotron-Resonanz bestimmt werden.

Insgesamt sehen wir, dass wir durch Messung der elektrischen Leitfähigkeit und des Hall-Effekts sowie durch die Bestimmung der effektiven Massen mit Hilfe der Zyklotron-Resonanz alle relevanten Halbleiterparameter wie  $E_g$ ,  $E_a$ ,  $E_d$ ,  $n_D$ ,  $n_A$ ,  $\mu_e$  oder  $\mu_h$  bestimmen können.

## 10.1.6 Vertiefungsthema: Seebeck- und Peltier-Effekt



Nach Gleichung (9.5.42) gilt  $\mathbf{J}_h = \Pi \mathbf{J}_q$ . Um den Peltier-Koeffizienten  $\Pi$  von Halbleitern abzuschätzen, müssen wir uns überlegen, welche Wärmemenge ein Ladungsträger in einem Halbleiter transportiert. Ein Elektron im Leitungsband hat bezogen auf das chemische Potenzial die Energie  $(E_c - \mu) + \frac{3}{2}k_B T$  und transportiert deshalb nicht nur die Wärmemenge  $\frac{3}{2}k_B T$  sondern  $(E_c - \mu) + \frac{3}{2}k_B T$ . Ein Loch im Valenzband transportiert entsprechend die Energie  $(\mu - E_v) + \frac{3}{2}k_B T$ . Das heißt, der mit dem Ladungsfluss  $\mathbf{J}_q = n_c(-e)\mu_e \mathbf{E}$  der Elektronen bzw.  $\mathbf{J}_q = p_v(+e)\mu_h \mathbf{E}$  der Löcher verbundene Wärmefluss  $\mathbf{J}_h$  ist gegeben durch

$$\mathbf{J}_h = n_c \left( E_c - \mu + \frac{3}{2}k_B T \right) \mu_e \mathbf{E} \quad (10.1.71)$$

$$\mathbf{J}_h = p_v \left( \mu - E_v + \frac{3}{2}k_B T \right) \mu_h \mathbf{E}. \quad (10.1.72)$$

Damit erhalten wir mit  $\mathbf{J}_h = \Pi \mathbf{J}_q$  die Peltier-Koeffizienten

$$\Pi_e = - \frac{E_c - \mu + \frac{3}{2}k_B T}{e} \quad (10.1.73)$$

$$\Pi_h = + \frac{\mu - E_v + \frac{3}{2}k_B T}{e}. \quad (10.1.74)$$

Die Thermokraft bestimmen wir über die Kelvin-Beziehung  $\Pi = S \cdot T$ . Der Temperaturverlauf des Peltier-Koeffizienten ist durch die Temperaturabhängigkeit des chemischen Potentials gegeben, die für einen  $n$ -Halbleiter in Abb. 10.11 gezeigt ist. Für nicht allzu hohe Temperaturen ist  $(E_c - \mu) \gg k_B T$  bzw.  $(\mu - E_v) \gg k_B T$ . Das bedeutet, dass der Peltier-Koeffizient bzw. die Thermokraft für einen  $n$ -Halbleiter negativ bzw. für einen  $p$ -Halbleiter positiv ist. Wir können deshalb durch Messung der Thermokraft sehr einfach bestimmen, ob wir einen  $n$ - oder einen  $p$ -Halbleiter vorliegen haben. Außerdem sind Peltier-Koeffizient und Thermokraft von Halbleitern üblicherweise wesentlich größer als bei Metallen.

## 10.2 Inhomogene Halbleiter

Viele Entwicklungen im Bereich der modernen Festkörperphysik sind mit dem großen Erfolg von Halbleiterbauelementen und der halbleiterbasierten Festkörperelektronik verbunden. Da die Funktionsweise der Mehrzahl der Halbleiterbauelemente auf Phänomenen beruht, die mit räumlich inhomogenen Halbleiterstrukturen zusammenhängen, wollen wir uns in diesem Abschnitt mit solchen Halbleitersystemen beschäftigen. Dabei konzentrieren wir uns auf Systeme, die eine räumlich inhomogene Konzentration von Donatoren und Akzeptoren enthalten. Wir werden uns dabei auf die Diskussion der grundlegenden Aspekte und der zentralen Strukturen wie des  $p$ - $n$ -Übergangs beschränken. Die theoretischen Grundlagen zum  $p$ - $n$ -Übergang wurden von **William B. Shockley**<sup>14</sup> erarbeitet.<sup>15,16</sup> Eine umfassende Diskussion verschiedener Halbleiter-Bauelemente kann in speziellen Lehrbüchern gefunden werden.<sup>17,18</sup>

Wir betrachten einen Halbleiter (z. B. Silizium) der auf der linken Seite  $p$ - und auf der rechten Seite  $n$ -dotiert sein soll (siehe Abb. 10.14). Die Konzentration der Dotieratome soll sich an der Grenzfläche, die senkrecht zur  $x$ -Richtung verläuft, abrupt ändern, so dass die Konzentration von Donatoren und Akzeptoren wie folgt geschrieben werden kann:

$$n_A(x) = \begin{cases} n_A & \text{für } x < 0 \\ 0 & \text{für } x > 0 \end{cases}, \quad n_D(x) = \begin{cases} 0 & \text{für } x < 0 \\ n_D & \text{für } x > 0 \end{cases}. \quad (10.2.1)$$

Solche abrupten Übergänge sind nicht nur konzeptionell interessant, sondern haben auch große Anwendungsrelevanz. Die Herstellung eines solchen abrupten Übergangs ist allerdings schwierig. Wir werden aber später sehen, dass im Rahmen einer physikalischen Modellierung abrupt nur bedeutet, dass die Breite des Bereichs, innerhalb dessen sich die Dotierung ändert, kleiner als die am Übergang auftretende Ladungsträgerverarmungszone sein

<sup>14</sup> **William Bradford Shockley**, siehe Kasten auf Seite 502.

<sup>15</sup> W. Shockley, *The Theory of  $p$ - $n$  Junctions in Semiconductors and  $p$ - $n$  Junction Transistors*, Bell. Syst. Tech. J. 28, 435 (1949).

<sup>16</sup> C. T. Sah, R. N. Noyce, W. Shockley, *Carrier generation and recombination in  $p$ - $n$  junctions and  $p$ - $n$  junction characteristics*, Proc. IRE 45, 1228 (1957).

<sup>17</sup> *Semiconductor Devices: Physics and Technology*, S. M. Sze, Wiley, New York (1985).

<sup>18</sup> *The Physics of Semiconductor Devices*, S. M. Sze and K. Ng Kwok, John Wiley & Sons, New York (1981).

muss. Diese Anforderung bedeutet, dass die Übergangsbreite im Bereich von 10 nm und mehr sein darf.

Um die Reaktion eines inhomogenen Halbleiters auf ein äußeres elektrostatisches Potenzial zu beschreiben oder um einfach die Ladungsverteilung in Abwesenheit eines elektrostatischen Potenzials zu berechnen, wird üblicherweise ein semi-klassisches Modell benutzt. Wie wir in Kapitel 9 ausführlich diskutiert haben, dürfen wir eine semi-klassische Beschreibung immer dann vornehmen, wenn sich das elektrostatische Potenzial  $\phi(\mathbf{r})$  auf einer Längenskala ändert, die groß gegenüber dem Gitterabstand der Atome ist, also groß gegenüber der Längenskala auf der sich das periodische Potenzial des Festkörpers ändert (vergleiche hierzu Abschnitt 9.1.2).

### 10.2.1 $p$ - $n$ Übergang im thermischen Gleichgewicht

Wir betrachten zunächst einen  $p$ - $n$  Übergang im thermischen Gleichgewicht ohne ein von außen angelegtes elektrisches Potenzial. Wir beginnen unsere Diskussion mit zwei völlig getrennten  $p$ - und  $n$ -dotierten Halbleitern (Abb. 10.14a). In diesem Fall liegt das chemische Potenzial in beiden Materialien auf einer gemeinsamen Energieskala bei unterschiedlichen Werten. Stellen wir den Kontakt zwischen den beiden Seiten her, so muss im thermischen Gleichgewicht das elektrochemische Potenzial über die ganze Struktur konstant sein. Das bedeutet, dass sich in einer Übergangszone auf beiden Seiten der Kontaktfläche die Valenz- und Leitungsbänder der beiden Halbleiter verbiegen müssen (Abb. 10.14b). Im Rahmen der semiklassischen Beschreibung kann diese Situation durch ein inneres elektrostatisches Potenzial oder *Makropotenzial*  $\phi(x)$  beschrieben werden, welches der Bandverbiegung Rechnung trägt. Voraussetzung ist, dass sich das Potenzial nur wenig auf der Längenskala der Gitterkonstante ändert.<sup>19</sup> Mit dem Makropotenzial  $\phi(x)$  erhalten wir die potenzielle Energie der Löcher (Ladung  $+e$ ) zu  $+e\phi(x)$  und die potenzielle Energie der Elektronen (Ladung  $-e$ ) zu  $-e\phi(x)$ .

Nach der Poisson-Gleichung ist das Makropotenzial  $\phi(x)$  mit einer Raumladung  $\rho(x)$  verbunden:

$$-\nabla^2 \phi = -\frac{\partial^2 \phi(x)}{\partial x^2} = \frac{\rho(x)}{\epsilon \epsilon_0}. \quad (10.2.2)$$

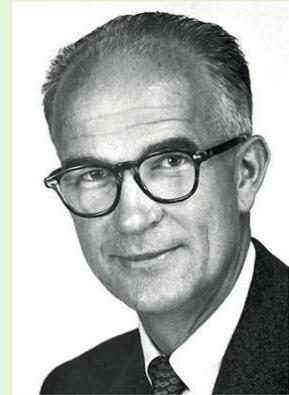
Die Entstehung der Raumladungszone können wir qualitativ dadurch verstehen, dass ein Teil der im  $n$ -Halbleiter überwiegend vorhandenen Elektronen und der im  $p$ -Halbleiter überwiegend vorhandenen Löcher in das Gebiet mit der jeweils anderen Dotierung diffundiert und dort teilweise rekombiniert. Wir sprechen in diesem Zusammenhang deshalb von einem *Diffusionsstrom* oder auch *Rekombinationsstrom*. Wir bezeichnen Elektronen im  $n$ -Gebiet bzw. Löcher im  $p$ -Gebiet als *Majoritätsladungsträger*, da sie dem vorherrschenden Ladungstyp entsprechen. Da aber Elektronen ins  $p$ - und Löcher ins  $n$ -Gebiet diffundieren

<sup>19</sup> Das bei der Einstellung des thermodynamischen Gleichgewichts erzeugte Makropotenzial  $\phi(x)$  wird üblicherweise in das chemische Potenzial einbezogen, so dass  $\mu = \text{const}$ . Im Gegensatz dazu wird ein von außen angelegtes elektrostatisches Potenzial (z. B. angelegte Spannung  $U$ ) explizit hinzugefügt, so dass  $\mu + eU(x) = \text{const}$ . Das chemische Potenzial  $\mu$  wird dann als verbogene Kurve gezeichnet.

### William Bradford Shockley (1910–1989)

William Bradford Shockley wurde am 13. Februar 1910 in London. Er starb am 12. August 1989 in Stanford.

William Shockley machte seine Ausbildung in Kalifornien und erhielt 1932 seinen B. Sc. vom California Institute of Technology (Caltech). Er promovierte dann 1936 bei John C. Slater am Massachusetts Institute of Technology (MIT) über die Struktur der Energiebänder in NaCl. Anschließend ging er zu den Bell Telephone Laboratories, wo er bis auf kurze Unterbrechungen arbeitete, z. B. in der Gruppe von Clinton Davisson. Im Jahr 1946 war er Gastprofessor an der Princeton University und 1954 am California Institute of Technology. Er war ferner für ein Jahr (1954/55) stellvertretender Direktor der Weapon Systems Evaluation Group des US-Verteidigungsministeriums.

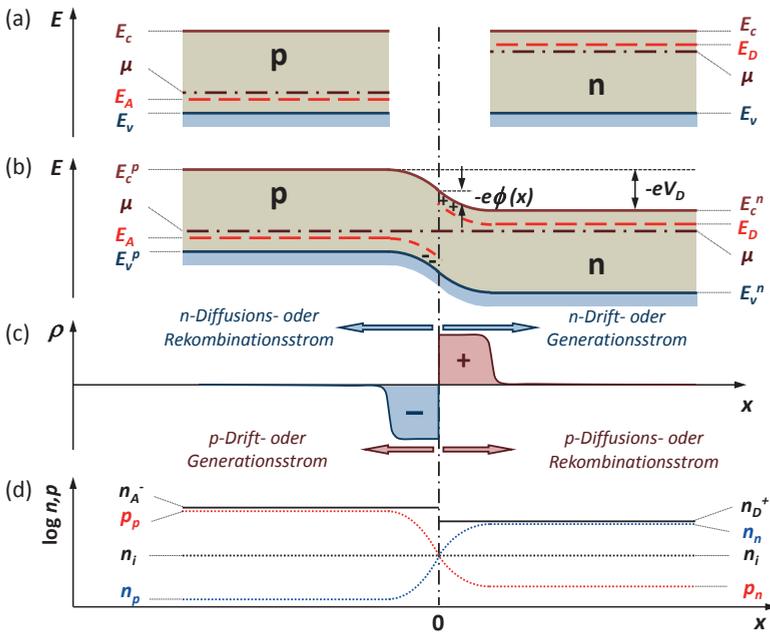


©The Nobel Foundation.

Im Jahr 1955 wechselte Shockley als Direktor zum Shockley-Halbleiterlaboratorium bei Beckman Instruments in Mountain View, Kalifornien, um dort den neuen Transistor und weitere Halbleiterbauelemente weiterzuentwickeln und zu produzieren. Er wurde 1963 zum Alexander M. Poniatoff Professor für Ingenieurwissenschaften an der Stanford University ernannt. Seit 1951 war er Mitglied des wissenschaftlichen Beraterstabes der US Army und ab 1958 der US Air Force. Im Jahr 1962 wurde er in den wissenschaftlichen Beraterstab des US-Präsidenten berufen.

In seiner wissenschaftlichen Arbeit beschäftigte sich Shockley mit den Energiebändern von Festkörpern, mit Legierungen, der Theorie der Vakuumröhren, mit theoretischen Modellen zu Versetzungen und Korngrenzen, mit ferromagnetischen Domänen, sowie mit Photoelektronen in Silberchlorid. Nach der Entwicklung des Transistors (kurz vor Weihnachten 1947) beschäftigte er sich mit den verschiedenen Aspekten der Transistorphysik. Daneben betrieb er „Operations Research“ über den Einfluss des Gehaltes auf die individuelle Produktivität in Forschungslaboratorien. Nach 1963 widmete sich Shockley, obwohl er keine Ausbildung im Fach Psychologie genossen hatte, der Erforschung von Zusammenhängen zwischen Rasse und Intelligenz sowie Themen aus dem Bereich Eugenik. Diese unrühmlichen Arbeiten mündeten in einigen unhaltbaren rassistischen Thesen. So sah Shockley in der größeren Kinderzahl der schwarzen US-Bürger eine Bedrohung für die Zukunft der USA, da diese Bevölkerungsgruppe weniger intelligent sei als die weiße US-Bevölkerung und forderte die Sterilisation für Menschen mit einem niedrigeren IQ als 100 und die verstärkte Fortpflanzung Intelligenter.

Shockley erhielt für seine physikalischen Arbeiten mehrere Auszeichnungen, u. a. die Medal for Merit (1946), den Morris Leibmann Memorial Prize des Institute of Radio Engineers (1952), den Oliver E. Buckley Solid State Physics Prize der American Physical Society (1953) und den Cyrus B. Comstock Award der National Academy of Sciences (1954). Zusammen mit Walter H. Brattain und John Bardeen wurde Shockley 1956 mit dem Nobelpreis für Physik „für ihre Untersuchungen über Halbleiter und ihre Entdeckung des Transistoreffekts“ ausgezeichnet.



**Abb. 10.14:** Schematische Darstellung eines  $p$ - $n$  Übergangs im thermischen Gleichgewicht: (a) Bänderschema im  $p$ - und  $n$ -Halbleiter bei völliger Trennung, (b) Bandverlauf im  $p$ - $n$  Übergang im thermischen Gleichgewicht nach Herstellung des Kontakts, (c) Verlauf der Raumladungszone  $\rho(x)$  im Bereich des  $p$ - $n$  Übergangs und (d) qualitativer Verlauf der Konzentration der Donatoren  $n_D^+$  und Akzeptoren  $n_A^-$  sowie der Elektronen im Leitungsband und Löcher im Valenzband. Wir nehmen an, dass alle Donatoren und Akzeptoren ionisiert sind, so dass  $n_D = n_D^+$  und  $n_A = n_A^-$ . Auf der  $n$ -Seite ist  $V_D$  positiv, so dass die potenzielle Energie der Elektronen auf der  $n$ -Seite um  $-eV_D$  abgesenkt ist.

können, erhalten wir eine Konzentration  $n_p$  von Elektronen im  $p$ -Gebiet und  $p_n$  von Löchern im  $n$ -Gebiet. Wir bezeichnen diese Ladungsträger als **Minoritätsladungsträger**.<sup>20</sup>

Durch das Abwandern von Elektronen aus der Grenzschicht des  $n$ -Halbleiters entsteht dort eine positive Raumladungszone, da die ortsfesten ionisierten Donatoren dort zurückbleiben. Umgekehrt entsteht durch das Abwandern von Löchern aus der Grenzschicht des  $p$ -Halbleiters dort eine negative Raumladungszone, da die ortsfesten ionisierten Akzeptoren dort zurückbleiben. Die resultierende Raumladungszone  $\rho(x)$  ist in Abb. 10.14c gezeigt. Auf diese Weise wird in der Grenzschicht ein elektrisches Feld erzeugt, welches dem von dem Konzentrationsgradienten der Elektronen und Löcher verursachten Diffusionsstrom entgegenwirkt. Die gesamte zwischen dem  $p$ - und  $n$ -Bereich resultierende Potentialdifferenz  $\phi(\infty) - \phi(-\infty)$  wird als **Diffusionsspannung**  $V_D$  bezeichnet. Sie führt zu einem **Driftstrom**, der im thermischen Gleichgewicht den Diffusionsstrom gerade kompensiert. Der Driftstrom setzt sich aus einem Strom von Elektronen und Löchern zusammen, die jeweils aus dem  $p$ - und  $n$ -Halbleiter kommen. Der Driftstrom wird also von den jeweiligen Minoritätsladungsträgern getragen. Da diese in den jeweiligen Halbleitertypen ständig neu erzeugt werden müs-

<sup>20</sup> Generell geben wir mit dem Index an, in welchem Halbleitertyp sich der jeweilige Ladungsträger befindet.

sen, bezeichnet man den Driftstrom auch als *Generationsstrom*. Weit außerhalb der Raumladungszone werden die ionisierten Donatoren  $n_D^+$  bzw. Akzeptoren  $n_A^-$  im  $n$ - und  $p$ -Gebiet durch die hohe Konzentration  $n_n$  der Elektronen im  $n$ -Gebiet bzw. die hohe Konzentration  $p_p$  der Löcher im  $p$ -Gebiet kompensiert, so dass hier  $\rho(x) = 0$  (siehe Abb. 10.14c und d).

Wir wollen nun Ausdrücke für die Ladungskonzentrationen sowie den Verlauf und die Größe des Makropotenzials ableiten. Wir benutzen dazu die Tatsache, dass im thermischen Gleichgewicht das Massenwirkungsgesetz ( $n_i^2 = n \cdot p$ ) erfüllt sein muss. Für die Konzentrationen der Majoritätsladungsträger weit weg von der Grenzfläche gilt mit den in Abschnitt 10.1.3 benutzten Argumenten

$$n_n(\infty) = n_c^{\text{eff}} \exp\left(-\frac{E_c^n - \mu}{k_B T}\right) \quad (10.2.3)$$

$$p_p(-\infty) = p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v^p}{k_B T}\right). \quad (10.2.4)$$

Hierbei ist  $E_c^n = E_c - e\phi(\infty)$  und  $E_v^p = E_v - e\phi(-\infty)$ ,  $n_c^{\text{eff}}$  und  $p_v^{\text{eff}}$  sind die durch (10.1.22) und (10.1.23) definierten effektiven Ladungsträgerdichten. Mit dem Massenwirkungsgesetz folgt daraus (siehe Abb. 10.14b)

$$\begin{aligned} n_i^2 &= n_n \cdot p_p = n_c^{\text{eff}} p_v^{\text{eff}} \exp\left(-\frac{E_c^n - E_v^p}{k_B T}\right) \\ &= n_c^{\text{eff}} p_v^{\text{eff}} \exp\left(-\frac{[E_c - E_v] - [e\phi(\infty) - e\phi(-\infty)]}{k_B T}\right) \\ &= n_c^{\text{eff}} p_v^{\text{eff}} \exp\left(-\frac{E_g - eV_D}{k_B T}\right). \end{aligned} \quad (10.2.5)$$

Wir erhalten somit

$$eV_D = E_g + k_B T \ln\left(\frac{n_i^2}{n_c^{\text{eff}} p_v^{\text{eff}}}\right) = E_g + k_B T \ln\left(\frac{n_n p_p}{n_c^{\text{eff}} p_v^{\text{eff}}}\right). \quad (10.2.6)$$

Die Ladungsträgerdichte für einen beliebigen Ort  $x$  erhalten wir, indem wir in (10.2.3) und (10.2.4) die ortsabhängigen Werte  $E_c^n(x) = E_c - e\phi(x)$  und  $E_v^p(x) = E_v - e\phi(x)$  für die Bandkanten einsetzen. Alternativ können wir ein ortsabhängiges elektrochemisches Potenzial  $\tilde{\mu}(x) = \mu + e\phi(x)$  verwenden. Mit  $\tilde{\mu}(x)$  werden die Ausdrücke (10.2.3) und (10.2.4) äquivalent zu den Ausdrücken (10.1.22) und (10.1.23) für die Ladungsträgerdichten in einem homogenen Halbleiter:

$$n(x) = n_c^{\text{eff}} \exp\left(-\frac{E_c - e\phi(x) - \mu}{k_B T}\right) = n_c^{\text{eff}} \exp\left(-\frac{E_c - \tilde{\mu}(x)}{k_B T}\right) \quad (10.2.7)$$

$$p(x) = p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v + e\phi(x)}{k_B T}\right) = p_v^{\text{eff}} \exp\left(-\frac{\tilde{\mu}(x) - E_v}{k_B T}\right). \quad (10.2.8)$$

Wir haben oben bereits qualitativ argumentiert, dass sich im thermischen Gleichgewicht die Diffusions- und Driftströme an der Grenzfläche kompensieren müssen. Wir wollen diesen

Zusammenhang jetzt quantitativ diskutieren. Mit den Diffusionskonstanten  $D_n$  und  $D_p$  für die Elektronen und Löcher erhalten wir für die Ströme

$$J^{\text{diff}} = J_n^{\text{diff}} + J_p^{\text{diff}} = e \left( D_n \frac{\partial n}{\partial x} - D_p \frac{\partial p}{\partial x} \right) \quad (10.2.9)$$

$$J^{\text{drift}} = J_n^{\text{drift}} + J_p^{\text{drift}} = e (n\mu_n + p\mu_p) E_x. \quad (10.2.10)$$

Aus  $J^{\text{diff}} + J^{\text{drift}} = 0$  folgt, dass die Beiträge der Elektronen und Löcher einzeln verschwinden müssen, da sich weder Elektronen noch Löcher in irgendeinem Teilgebiet ansammeln können. Es muss deshalb sowohl für Elektronen als auch für Löcher gelten:

$$D_n \frac{\partial n}{\partial x} = n\mu_n \frac{\partial \phi(x)}{\partial x} \quad - D_p \frac{\partial p}{\partial x} = p\mu_p \frac{\partial \phi(x)}{\partial x}. \quad (10.2.11)$$

Hierbei haben wir  $E_x = -\partial\phi/\partial x$  verwendet. Mit den durch (10.2.7) und (10.2.8) gegebenen Ladungsträgerdichten erhalten wir

$$\frac{\partial n}{\partial x} = n \frac{e}{k_B T} \frac{\partial \phi(x)}{\partial x} \quad \frac{\partial p}{\partial x} = -p \frac{e}{k_B T} \frac{\partial \phi(x)}{\partial x} \quad (10.2.12)$$

und damit nach Substitution in (10.2.11)

$$D_n = \frac{k_B T}{e} \mu_n \quad D_p = \frac{k_B T}{e} \mu_p. \quad (10.2.13)$$

Diese Beziehungen werden *Einstein-Relationen* genannt und gelten immer dann, wenn Diffusions- und Driftströme durch denselben Ladungsträgertyp getragen werden.

Wir können nun die Poisson-Gleichung (10.2.2) dazu benutzen, eine Beziehung zwischen  $\phi(x)$  und  $\rho(x)$  herzustellen. Nehmen wir an, dass alle Akzeptoren und Donatoren ionisiert sind ( $n_A = n_A^-$  und  $n_D = n_D^+$ ), so ergibt sich die Ladungsdichte durch die Dotieratome und die Ladungsträgerdichte zu

$$\rho(x) = e [n_D(x) - n_A(x) - n(x) + p(x)]. \quad (10.2.14)$$

Setzen wir die entsprechenden Ausdrücke (10.2.1), (10.2.7) und (10.2.8) in diese Gleichung ein und substituieren das Ergebnis in die Poisson-Gleichung, so erhalten wir eine nicht-lineare Differentialgleichung für  $\phi(x)$ , die sich nur numerisch lösen lässt. Wir wollen im Folgenden eine näherungsweise Betrachtung machen, die als *Schottky-Modell* der Raumladungszone bekannt ist.

Mit der Voraussetzung eines abrupten  $p$ - $n$ -Übergangs können wir die Ladungsträgerdichte schreiben als

$$\rho(x < 0) = e [-n_A - n(x) + p(x)] \quad (10.2.15)$$

$$\rho(x > 0) = e [+n_D - n(x) + p(x)]. \quad (10.2.16)$$

Die ortsabhängigen Ladungsträgerkonzentrationen  $n(x)$  und  $p(x)$  hängen vom Abstand der jeweiligen Bandkante vom chemischen Potenzial  $\mu$  ab. Obwohl sich dieser Abstand nur

langsam ändert (siehe Abb. 10.14b), bewirkt die Fermi-Verteilungsfunktion, dass sich die Besetzungswahrscheinlichkeit innerhalb eines schmalen Energiefensters von etwa  $2k_B T \simeq 50 \text{ meV}$ , das viel kleiner als der Bandabstand  $E_g$  ist, von null auf den maximalen Wert ändert. Vernachlässigen wir nun diese schmalen Übergangsbereiche, so können wir die Konzentrationen  $n_D^+$  und  $n_A^-$  der geladenen Donatoren und Akzeptoren, die nicht durch freie Ladungsträger kompensiert werden, durch Stufenfunktionen annähern. In dieser Näherung können wir die Raumladungsdichte schreiben als

$$\rho(x) = \begin{cases} 0 & \text{für } x < -d_p \\ -en_A & \text{für } -d_p < x < 0 \\ +en_D & \text{für } 0 < x < d_n \\ 0 & \text{für } x > d_n \end{cases} \quad (10.2.17)$$

Hierbei geben die Längen  $d_p$  und  $d_n$  die Ausdehnung der Raumladungszone im  $p$ - und  $n$ -Halbleiter an. Mit dieser stückweise konstanten Raumladungsdichte erhalten wir die Poisson-Gleichung zu

$$\frac{\partial^2 \phi}{\partial x^2} = \begin{cases} 0 & \text{für } x < -d_p \\ \frac{+en_A}{\epsilon\epsilon_0} & \text{für } -d_p < x < 0 \\ \frac{-en_D}{\epsilon\epsilon_0} & \text{für } 0 < x < d_n \\ 0 & \text{für } x > d_n \end{cases} \quad (10.2.18)$$

Durch Integration ergibt sich

$$\phi(x) = \begin{cases} \phi(-\infty) & \text{für } x < -d_p \\ \phi(-\infty) + \left(\frac{en_A}{2\epsilon\epsilon_0}\right)(d_p + x)^2 & \text{für } -d_p < x < 0 \\ \phi(+\infty) - \left(\frac{en_D}{2\epsilon\epsilon_0}\right)(d_n - x)^2 & \text{für } 0 < x < d_n \\ \phi(+\infty) & \text{für } x > d_n \end{cases} \quad (10.2.19)$$

Der Verlauf von  $\phi(x)$  sowie seiner 1. (elektrisches Feld) und 2. Ableitung (Raumladungsdichte) sind in Abb. 10.15 dargestellt.

Die Randbedingungen (Stetigkeit von  $\phi(x)$  und seiner 1. Ableitung) werden von der Lösung bei  $x = d_n$  und  $x = -d_p$  explizit erfüllt. Damit die 1. Ableitung von  $\phi(x)$  auch bei  $x = 0$  stetig ist, muss

$$n_D d_n = n_A d_p \quad (10.2.20)$$

gelten. Diese Forderung stellt sicher, dass die negative Raumladung im  $p$ -Halbleiter mit der positiven im  $n$ -Halbleiter übereinstimmt. Damit  $\phi(x)$  bei  $x = 0$  stetig ist, muss

$$\frac{e}{2\epsilon\epsilon_0} (n_D d_n^2 + n_A d_p^2) = \phi(+\infty) - \phi(-\infty) = V_D \quad (10.2.21)$$

gelten.

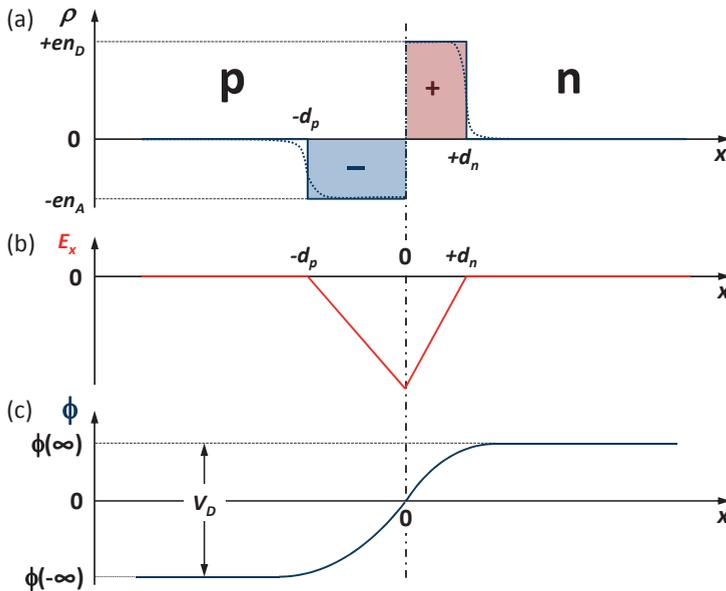


Abb. 10.15: Schottky-Modell der Raumladungszone eines  $p$ - $n$ -Übergangs: (a) Raumladungsdichte  $\rho(x)$  (gestrichelt ist die realistische Form von  $\rho(x)$  angegeben, die im Rahmen des Schottky-Modells durch eine Stufenfunktion approximiert wird). (b) Verlauf des elektrischen Feldes und (c) Verlauf des Makropotenzials  $\phi(x)$ . Die potenzielle Energie der Elektronen (Ladung  $-e$ ) beträgt  $-e\phi(x)$ .

Aus (10.2.20) und (10.2.21) können wir bei bekannten Verunreinigungskonzentrationen die Ausdehnung der Raumladungszone berechnen. Wir erhalten

$$d_n = \left( \frac{2\epsilon\epsilon_0 V_D}{e} \frac{n_A/n_D}{n_A + n_D} \right)^{1/2} \tag{10.2.22}$$

$$d_p = \left( \frac{2\epsilon\epsilon_0 V_D}{e} \frac{n_D/n_A}{n_A + n_D} \right)^{1/2} . \tag{10.2.23}$$

Üblicherweise ist für die meisten Halbleitermaterialien  $eV_D \simeq E_g$ . Mit  $E_g \sim 1$  eV und typischen Verunreinigungskonzentrationen im Bereich von  $10^{14}$  bis  $10^{18}$  cm<sup>3</sup> liegen  $d_p$  und  $d_n$  zwischen 10 und 1000 nm. Die Feldstärke innerhalb der Raumladungszone ist  $V_D/(d_p + d_n)$  und liegt bei einer Energielücke von 1 eV im Bereich zwischen  $10^6$  und  $10^8$  V/cm.

### 10.2.2 $p$ - $n$ Übergang mit angelegter Spannung

Wir betrachten nun den Fall, dass eine zeitunabhängige externe Spannung  $U$  am  $p$ - $n$  Übergang anliegt. Wir nehmen  $U$  als positiv an, wenn sie das Potenzial auf der  $p$ -Seite bezüglich der  $n$ -Seite anhebt. In den Abbildungen zeichnen wir allerdings immer die potenzielle Energie der Elektronen ein. Für eine positive Spannung über den  $p$ - $n$ -Kontakt ist die potenzielle Energie der Elektronen im  $p$ -Bereich um  $-eU$  abgesenkt bzw. im  $n$ -Bereich um  $+eU$  ange-

hoben. Wir haben im vorangegangenen Abschnitt gesehen, dass im thermischen Gleichgewicht am  $p$ - $n$  Übergang eine Verarmungszone mit einer Breite zwischen 10 und 1000 nm entsteht. Aufgrund der sehr geringen Ladungsträgerdichte in dieser Zone können wir in guter Näherung annehmen, dass die gesamte angelegte Spannung über die Verarmungszone abfällt. Das bedeutet, dass sich das in Abb. 10.14b gezeigte Bandschema nur in dem Bereich der Raumladungszone ändert. Außerhalb der Raumladungszone verlaufen die Bänder und das Potenzial  $\phi(x)$  nach wie vor horizontal. Die Potenzialänderung über die Raumladungszone erhält mit der angelegten Spannung  $U$  den Wert

$$\phi(\infty) - \phi(-\infty) = V_D - U. \quad (10.2.24)$$

Die angelegte Spannung  $U$  verändert nun die Breite der Raumladungszone, da die Größe  $V_D$  in (10.2.22) und (10.2.23) durch  $V_D - U$  ersetzt werden muss. Wir erhalten damit

$$d_n = d_n(U=0) \left(1 - \frac{U}{V_D}\right)^{1/2} \quad (10.2.25)$$

$$d_p = d_p(U=0) \left(1 - \frac{U}{V_D}\right)^{1/2}. \quad (10.2.26)$$

Wir sehen, dass die Raumladungszone für positive Spannungen, wir nennen diese Richtung die *Durchlassrichtung*, abnimmt, während sie für negative Spannungen, wir nennen diese Richtung die *Sperrrichtung*, zunimmt.

Mit der Ausdehnung  $d_n$  der Raumladungszone ändert sich auch die in dieser Zone gespeicherte Ladung

$$Q_R = en_D d_n(U) A. \quad (10.2.27)$$

Um die Kapazität der Raumladungszone bei einer angelegten Gleichspannung  $U_0$  abzuschätzen, müssen wir überlegen, welche Ladungsmenge durch eine kleine Wechselfspannung  $\delta U$  an den Rändern der Verarmungszone hinzugefügt und entfernt wird. Die Ladungsmenge ist gegeben durch

$$\delta Q_R = en_D A \left. \frac{d d_n}{d U} \right|_{U_0} \delta U + en_A A \left. \frac{d d_p}{d U} \right|_{U_0} \delta U. \quad (10.2.28)$$

Hierbei können wir  $\delta Q_R$  als diejenige Ladungsmenge betrachten, die durch die Wechselfspannung mit Amplitude  $\delta U$  auf einen Plattenkondensator der Fläche  $A$  geschoben wird. Wir können dann die spannungsabhängige Kapazität des  $p$ - $n$ -Übergangs schreiben als

$$C_R(U_0) = \left| \frac{d Q_R}{d U} \right| = [en_D A d_n(0) + en_A A d_p(0)] \left| \frac{d}{d U} \left(1 - \frac{U}{V_D}\right)^{1/2} \right|_{U_0}. \quad (10.2.29)$$

Mit den Ausdrücken (10.2.22) und (10.2.23) sowie (10.2.25) und (10.2.26) erhalten wir

$$C_R(U_0) = A \left( \frac{n_A n_D}{n_A + n_D} \frac{e \epsilon \epsilon_0}{(V_D - U_0)} \right)^{1/2}. \quad (10.2.30)$$

Messen wir die Raumladungskapazität  $C_R$  als Funktion der angelegten Spannung, so können wir Informationen über die Verunreinigungskonzentrationen gewinnen. Ferner können wir durch Auftragung von  $1/C_R^2$  gegen  $U$  und Extrapolation auf  $U = 0$  die Diffusionsspan-

nung  $V_D$  bestimmen. Für  $n_A \gg n_D$  erhalten wir

$$\frac{1}{C_R^2} = \frac{1}{A^2} \frac{1}{n_D} \frac{(V_D - U_0)}{e\epsilon\epsilon_0}. \quad (10.2.31)$$

### 10.2.2.1 Strom-Spannungs-Charakteristik

Wir wollen jetzt die Strom-Spannungs-Charakteristik eines  $p$ - $n$  Kontakts diskutieren. Im letzten Abschnitt haben wir gesehen, dass für  $U = 0$  die Drift- und Diffusionsströme sich gegenseitig kompensieren, so dass der Gesamtstrom verschwindet. Im Fall einer von außen angelegten Spannung trifft dies nicht mehr zu. Wir betrachten im Folgenden den Elektronenstrom. Wir müssen einerseits den Driftstrom der Minoritätsladungsträger (Elektronen im  $p$ -Material) berücksichtigen, der vom  $p$ - in den  $n$ -Bereich fließt. Da die Minoritätsladungsträger im  $p$ -Bereich durch thermische Aktivierung ständig neu erzeugt werden müssen, nennen wir diesen Strom auch den **Generationsstrom**  $J_n^{\text{gen}}$ . Für eine genügend dünne Raumladungszone wird jedes Elektron, das aus dem  $p$ -Material in den Raumladungsbereich gelangt, durch das elektrische Feld innerhalb der Raumladungszone in den  $n$ -Bereich getrieben. Dieser Strom wird in erster Näherung unabhängig von der Größe des elektrischen Feldes und damit auch von der angelegten Spannung sein.

Andererseits müssen wir den Diffusionsstrom der Majoritätsladungsträger (Elektronen im  $n$ -Bereich) berücksichtigen, der vom  $n$ - in den  $p$ -Bereich fließt. Diesen Strom nennen wir **Rekombinationsstrom**  $J_n^{\text{rec}}$ , da die Elektronen nach Diffusion in den  $p$ -Bereich mit den dort vorhandenen zahlreichen Löchern rekombinieren. In der Richtung von  $n$  nach  $p$  bewegen sich die Elektronen gegen die Potentialschwelle der Diffusionsspannung. Je nach Vorzeichen wird diese Potentialschwelle durch die angelegte Spannung erhöht oder erniedrigt. Der Anteil der Elektronen, der die Potentialschwelle überwinden kann, wird durch einen Boltzmann-Faktor  $\exp(-e(V_D - U)/k_B T)$  bestimmt und hängt somit stark von der angelegten Spannung ab. Aufgrund unserer Diskussion können wir folgende Spannungsabhängigkeit der Rekombinationsstromdichte angeben:

$$J_n^{\text{rec}}(U) \propto e^{-e(V_D - U)/k_B T}. \quad (10.2.32)$$

Benutzen wir nun die Tatsache, dass der Generationsstrom in etwa unabhängig von  $U$  ist und für  $U = 0$  der Rekombinations- und Generationsstrom gleich sein müssen,  $J_n^{\text{rec}}(U = 0) = J_n^{\text{gen}}(U = 0)$ , so erhalten wir wegen  $J_n^{\text{rec}}(0) \propto e^{-eV_D/k_B T} = J_n^{\text{gen}}(0) = J_n^{\text{gen}}(U)$

$$J_n^{\text{rec}}(U) = J_n^{\text{gen}} e^{eU/k_B T}. \quad (10.2.33)$$

Die Gesamtstromdichte ergibt sich dann zu

$$J_n = J_n^{\text{rec}} - J_n^{\text{gen}} = J_n^{\text{gen}} (e^{eU/k_B T} - 1). \quad (10.2.34)$$

Die gleiche Analyse können wir für den Löcherstrom durchführen und erhalten

$$J_p = J_p^{\text{rec}} - J_p^{\text{gen}} = J_p^{\text{gen}} (e^{eU/k_B T} - 1). \quad (10.2.35)$$

Für den Gesamtstrom aus Elektronen und Löchern ergibt sich somit

$$J(U) = (J_p^{\text{gen}} + J_n^{\text{gen}}) (e^{eU/k_B T} - 1). \quad (10.2.36)$$

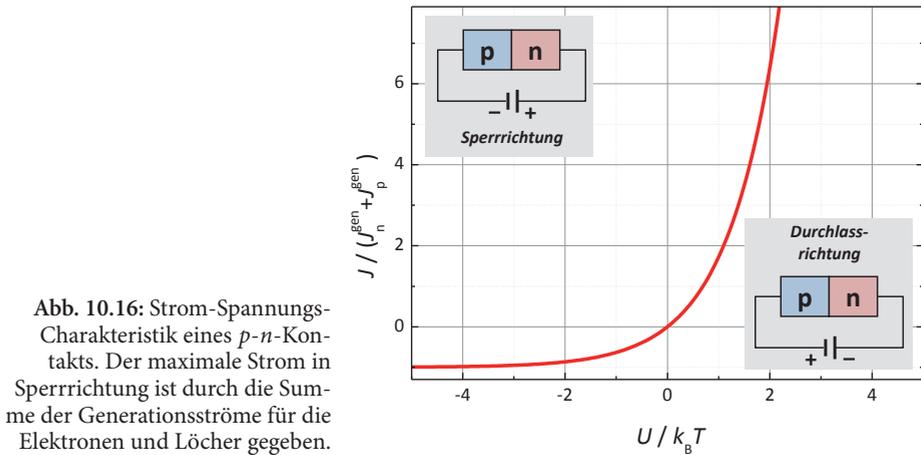


Abb. 10.16: Strom-Spannungs-Charakteristik eines  $p$ - $n$ -Kontakts. Der maximale Strom in Sperrichtung ist durch die Summe der Generationsströme für die Elektronen und Löcher gegeben.

Die entsprechende Strom-Spannungs-Charakteristik ist in Abb. 10.16 dargestellt. Sie ist bezüglich der beiden Polaritäten der angelegten Spannung extrem asymmetrisch und resultiert in einem gleichrichtenden Verhalten des  $p$ - $n$  Kontakts.



### 10.2.2.2 Vertiefungsthema: Sättigungsstrom

Um einen quantitativen Ausdruck für den Sättigungsstrom in Sperrichtung abzuleiten, müssen wir eine detailliertere Betrachtung des stationären Zustandes bei Anlegen einer Sperrspannung machen. Unsere obige Betrachtung hat gezeigt, dass eine Störung des Gleichgewichts hauptsächlich durch die Diffusionsströme erfolgt, wogegen der Einfluss der Spannung auf die Driftströme in guter Näherung vernachlässigt werden kann. Wir werden im Folgenden die so genannte *Diffusionsstrom-Näherung* diskutieren, in der es ausreicht, nur die Diffusionsströme unter dem Einfluss einer äußeren Spannung zu diskutieren.

Wir betrachten einen  $p$ - $n$ -Kontakt in Durchlassrichtung (siehe Abb. 10.17). Die Ladungsträgerdichte nimmt durch die anliegende Spannung im Bereich der Raumladungszone zu. Das Massenwirkungsgesetz  $n_i^2 = n \cdot p$  ist in diesem Fall nicht mehr gültig. Die elektrochemischen Potentiale weit außerhalb der Raumladungszone unterscheiden sich genau um die der anliegenden Spannung  $U$  entsprechende Energie  $-eU$ . Da im Bereich der Raumladungszone die Ladungsträger nicht im Gleichgewicht sind, können wir hier kein gemeinsames elektrochemisches Potenzial mehr definieren. Elektronen und Löcher stehen allerdings jeweils untereinander im Gleichgewicht, so dass wir zwei getrennte Quasi-Potentiale  $\mu^e(x)$  und  $\mu^p(x)$  definieren können, die wir unabhängig voneinander behandeln können. Diese sind in Abb. 10.17 durch gepunktete und gestrichelte Linien gezeichnet. Falls ferner der stationäre Zustand nicht stark vom Gleichgewichtszustand abweicht (dieses wollen wir im Folgenden annehmen), können wir nach wie vor die Situation näherungsweise mit Hilfe der Boltzmann-Statistik beschreiben.

In der Näherung, dass wir die Rekombination von Elektronen und Löchern im Bereich der Raumladungszone vernachlässigen können, ist es ausreichend, die Änderung der Diffusionsstromdichten am Rand der Raumladungszone bei  $-d_p$  und  $+d_n$  zu betrachten. Da die

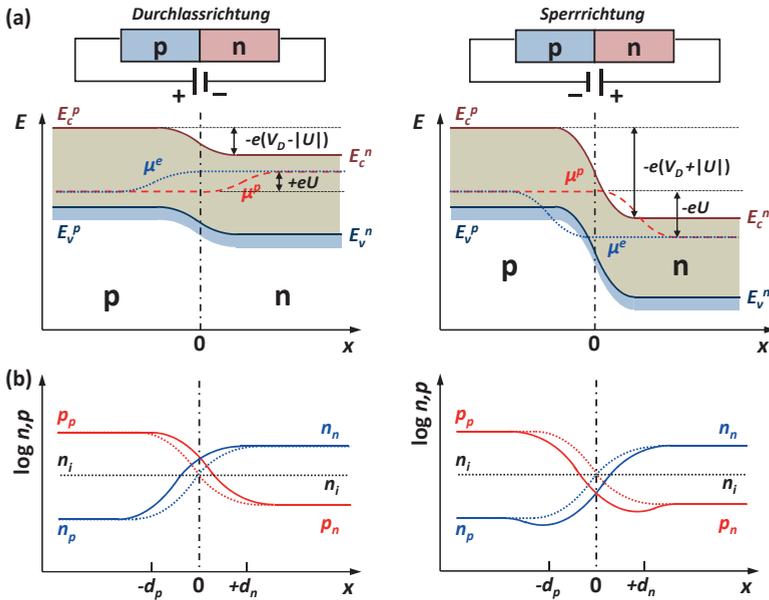


Abb. 10.17:  $p$ - $n$ -Kontakt in Durchlass- (links) und Sperrrichtung (rechts). (a) Verlauf der Leitungs- und Valenzbandkante sowie des elektrochemischen Quasi-Potenzials für Elektronen (gepunktet) und Löcher (gestrichelt). In Durchlassrichtung ist  $U$  negativ auf der  $n$ -Seite, so dass die potenzielle Energie  $(-e)(-U)$  der Elektronen positiv ist. (b) Räumliche Variation der Elektronenkonzentration  $n$  und der Löcherkonzentration  $p$  im Fall einer anliegenden Spannung (durchgezogene Linien) und im Fall  $U = 0$  (gepunktete Linien). Die Konzentrationen weit weg von der Raumladungszone werden mit  $p_p, n_p, n_n$  und  $p_n$  bezeichnet.

Berechnung für Elektronen und Löcher völlig analog ist, beschränken wir uns im Folgenden auf die Berechnung der Löcherstromdichte.

Für den Diffusionsstrom bei  $x = d_n$  folgt aus (10.2.9)

$$J_p^{\text{diff}}(x = d_n) = -eD_p \left. \frac{\partial p}{\partial x} \right|_{x=d_n}. \quad (10.2.37)$$

Wir werden im Folgenden zeigen, dass die Diffusionstheorie einen einfachen Ausdruck zwischen dem Konzentrationsgradienten  $\frac{\partial p}{\partial x}$  und der Zunahme der Lochkonzentration bei  $x = -d_n$  liefert. Die Lochkonzentration  $p(x = d_p)$  und  $p(x = d_n)$  können wir mit Hilfe der Boltzmann-Statistik erhalten. Wir benutzen hierzu Gleichung (10.2.8) und verwenden  $\phi(-d_p) = V_D - U$  sowie  $\phi(d_n) = -U$ . Wir erhalten dann

$$\begin{aligned} p(x = -d_p) &= p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v + e(V_D - U)}{k_B T}\right) \\ &= p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v}{k_B T}\right) \exp\left(-\frac{e(V_D - U)}{k_B T}\right) = p_p \exp\left(-\frac{e(V_D - U)}{k_B T}\right) \end{aligned} \quad (10.2.38)$$

sowie

$$\begin{aligned} p(x = d_n) &= p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v - eU}{k_B T}\right) \\ &= p_v^{\text{eff}} \exp\left(-\frac{\mu - E_v}{k_B T}\right) \exp\left(\frac{eU}{k_B T}\right) = p_n \exp\left(\frac{eU}{k_B T}\right). \end{aligned} \quad (10.2.39)$$

Hierbei sind  $d_p$  und  $d_n$  die Breiten der Raumladungszonen im thermischen Gleichgewicht ( $U = 0$ ). Die für die Durchlassrichtung um den Faktor  $e^{eU/k_B T}$  erhöhte Löcherkonzentration im  $n$ -Bereich führt zu einer erhöhten Rekombinationsrate, so dass diese schnell abklingt. Fern vom  $p$ - $n$ -Übergang wird der Strom deshalb im  $n$ -Gebiet von Elektronen und umgekehrt im  $p$ -Gebiet von Löchern getragen.

Um das Abklingen der Löcherkonzentration im  $n$ -Gebiet zu berechnen, benutzen wir die Kontinuitätsgleichung. Diese erfordert, dass die Löcherkonzentration in einem bestimmten Volumenelement sich nur durch Zu- oder Abfluss, durch thermische Generation sowie durch Rekombination ändert. Die Rekombinationsrate beschreiben wir mit Hilfe einer Rekombinationszeit  $\tau_p$  und erhalten damit

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \nabla \cdot \mathbf{J}_p^{\text{diff}} - \frac{p - p_n}{\tau_p}. \quad (10.2.40)$$

Hierbei beschreibt der 1. Term auf der rechten Seite den Zu- bzw. Abfluss und der 2. Term die Rekombination ( $p > p_n$ ) bzw. die Generation ( $p < p_n$ ). Im stationären Zustand ist  $\frac{\partial p}{\partial t} = 0$ , so dass wir unter Benutzung von (10.2.37)

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p - p_n}{\tau_p} = 0 \quad (10.2.41)$$

und somit

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{D_p \tau_p} (p - p_n) \quad (10.2.42)$$

erhalten. Die Lösung dieser Differentialgleichung ergibt das Diffusionsprofil

$$p(x) = p_n - p_n e^{-x/\sqrt{D_p \tau_p}} = p_n (1 - e^{-x/L_p}). \quad (10.2.43)$$

Hierbei ist  $L_p = \sqrt{D_p \tau_p}$  die Diffusionslänge für die Löcher im  $n$ -Material.

Für die Ableitung des Diffusionsprofils an der Stelle  $x = d_n$  erhalten wir unter Benutzung von (10.2.39):

$$\left. \frac{\partial p}{\partial x} \right|_{x=d_n} = -\frac{p(x = d_n) - p_n}{L_p} = -\frac{p_n (e^{eU/k_B T} - 1)}{L_p}. \quad (10.2.44)$$

Setzen wir diesen Ausdruck in (10.2.37) ein, so ergibt sich

$$J_p^{\text{diff}}(x = d_n) = \frac{e D_p}{L_p} p_n (e^{eU/k_B T} - 1). \quad (10.2.45)$$

Eine völlig analoge Rechnung können wir für den Diffusionsstrom der Elektronen im  $p$ -Gebiet durchführen:

$$J_n^{\text{diff}}(x = -d_p) = \frac{eD_n}{L_n} n_p (e^{eU/k_B T} - 1). \quad (10.2.46)$$

Wie oben bereits diskutiert wurde, müssen wir die Driftströme nicht berücksichtigen. Ihre Komponenten bleiben in der gemachten Näherung unverändert gegenüber dem thermischen Gleichgewicht und kompensieren gerade die Gleichgewichtsanteile der Diffusionsströme. Der gesamte Strom über den  $p$ - $n$ -Kontakt ergibt sich aus der Summe der beiden Diffusionsströme (10.2.45) und (10.2.46) zu

$$J(U) = \left( \frac{eD_p}{L_p} p_n + \frac{eD_n}{L_n} n_p \right) (e^{eU/k_B T} - 1). \quad (10.2.47)$$

Wir sehen, dass wir jetzt die Generationsströme in Gleichung (10.2.36) als Funktionen der Diffusionskonstanten und Diffusionslängen der Elektronen und Löcher sowie der Minoritätsladungsträgerdichten  $p_n$  und  $n_p$  ausgedrückt haben.

### 10.2.3 Schottky-Kontakt

Wir wollen in diesem Abschnitt einen Kontakt zwischen einem Halbleiter und einem Metall diskutieren, den wir als **Schottky-Kontakt** bezeichnen.<sup>21</sup> Solche Kontakte sind wichtige Elemente von elektronischen Schaltkreisen. Bei der Diskussion eines solchen Kontakts können wir konzeptionell ähnlich vorgehen, wie bei der Behandlung eines  $p$ - $n$ -Kontakts. Abb. 10.18 zeigt, was passiert, wenn wir einen  $n$ -Halbleiter und ein Metall in Kontakt bringen. Da die Abstände  $\Phi_H$  und  $\Phi_M$  zwischen chemischem Potenzial und Vakuumniveau für einen Halbleiter und ein Metall nicht gleich sein müssen, sind die chemischen Potenziale im Halbleiter und Metall um den Betrag  $\Phi_M - \Phi_H$  gegeneinander verschoben. Im Metall entspricht  $e\Phi_M$  gerade der **Austrittsarbeit**. Im Halbleiter ist  $e\Phi_H = e\chi + eV_n$  der Abstand des chemischen Potenzials vom Vakuumniveau. Der Abstand  $e\chi$  der Leitungsbandkante vom Vakuumniveau ist die **Elektronenaffinität**.

Wir betrachten zuerst den Fall  $\Phi_M > \Phi_H$ . Bringen wir die beiden Materialien in Kontakt, so muss im thermischen Gleichgewicht das chemische Potenzial horizontal verlaufen, was zu der in Abb. 10.18 gezeigten Bandverbiegung führt. Aus den gleichen Gründen wie beim  $p$ - $n$ -Kontakt erhalten wir auch beim Schottky-Kontakt eine Raumladungszone. In dem in Abb. 10.18 gezeigten Beispiel ist die Raumladungszone im Halbleiter positiv und im Metall negativ geladen. Aufgrund der hohen Ladungsträgerdichte im Metall ist allerdings ihre räumliche Ausdehnung im Metall verschwindend klein (typischerweise kleiner als 1 nm). Die Entstehung der Raumladungszone können wir qualitativ dadurch verstehen, dass ein Teil der im  $n$ -Halbleiter vorhandenen Elektronen in das Metall diffundiert. In der Grenzschicht des  $n$ -Halbleiters entsteht dann eine positive Raumladungszone, da die ortsfesten ionisierten Donatoren dort zurückbleiben. Umgekehrt können aber die Elektronen aus dem Metall nicht in den Halbleiter diffundieren, da dort auf der Höhe des chemischen Potenzials keine Zustände vorhanden sind.

<sup>21</sup> Walter H. Schottky, siehe Kasten auf Seite 516.

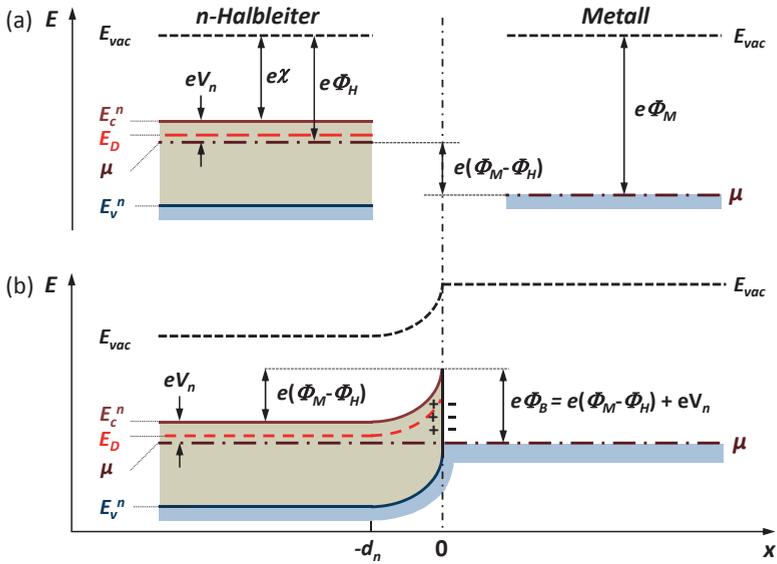


Abb. 10.18: Schematische Darstellung eines Schottky-Kontakts im thermischen Gleichgewicht für  $\Phi_M > \Phi_H$ : (a) Bänderschema im Metall (rechts) und *n*-Halbleiter (links) bei völliger Trennung, (b) Bandverlauf im Schottky-Kontakt im thermischen Gleichgewicht nach Herstellung des Kontakts.

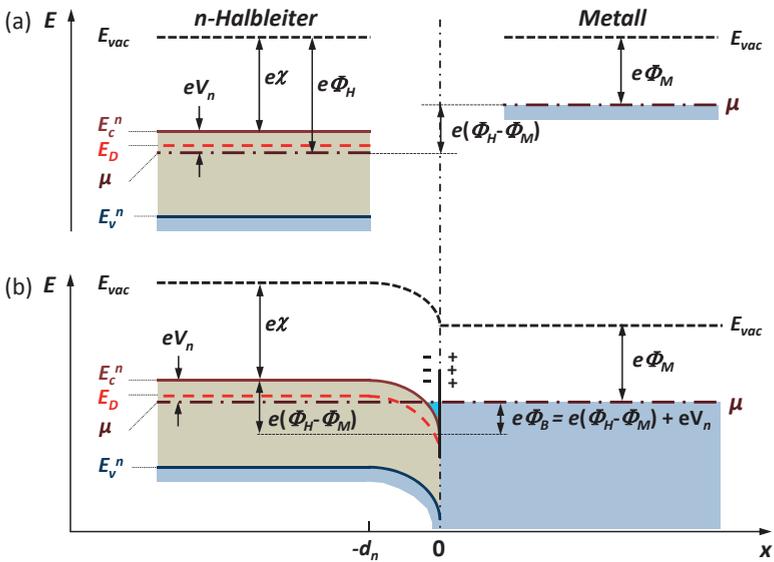


Abb. 10.19: Schematische Darstellung eines Schottky-Kontakts im thermischen Gleichgewicht für  $\Phi_M < \Phi_H$ : (a) Bänderschema im Metall (rechts) und *n*-Halbleiter (links) bei völliger Trennung, (b) Bandverlauf im Schottky-Kontakt im thermischen Gleichgewicht nach Herstellung des Kontakts. Im Halbleiter bildet sich an der Grenzfläche eine starke Anreicherung von Elektronen.

Im Rahmen des auch für den  $p$ - $n$ -Kontakt verwendeten Schottky-Modells erhalten wir im Bereich  $-d_n < x < 0$  [vergleiche (10.2.18)]

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{en_D}{\epsilon\epsilon_0}. \quad (10.2.48)$$

Durch Integration erhalten wir  $\phi(x) = \frac{n_D e}{2\epsilon\epsilon_0} x^2 + Bx + C$ . Die Integrationskonstanten  $B$  und  $C$  ergeben sich aus den Randbedingungen  $\phi(x = -d_n) = \Phi_M - \Phi_H$  und  $\left. \frac{\partial \phi}{\partial x} \right|_{x=-d_n} = 0$  zu  $C = \Phi_M - \Phi_H + \frac{n_D e}{2\epsilon\epsilon_0} d_n^2$  und  $B = \frac{n_D e}{\epsilon\epsilon_0} d_n$ . Damit erhalten wir

$$\phi(x) = (\Phi_M - \Phi_H) - \frac{en_D}{2\epsilon\epsilon_0} (x + d_n)^2. \quad (10.2.49)$$

Hierbei entspricht  $(\Phi_M - \Phi_H)$  der Diffusionsspannung  $V_D$  beim  $p$ - $n$ -Kontakt. Die Ausdehnung der Raumladungszone im Halbleiter erhalten wir wegen  $\Phi(x = 0) = 0$  zu

$$d_n = \left( \frac{2\epsilon\epsilon_0}{en_D} (\Phi_M - \Phi_H) \right)^{1/2}. \quad (10.2.50)$$

Wichtig ist, dass die Breite der Raumladungszone umgekehrt proportional zur Verunreinigungsdichte  $n_D$  des Halbleiters ist.

Für den Fall  $\Phi_M < \Phi_H$  ist die Situation anders. Jetzt fließen aufgrund der kleineren Austrittsarbeit des Metalls Elektronen vom Metall in den Halbleiter. Durch die negative Raumladung im Halbleiter werden die Leitungsbandkanten des Halbleiters nach unten gebogen. Das Fermi-Niveau liegt dadurch im Leitungsband des  $n$ -Halbleiters und es kommt im Halbleiter zu einer starken Anreicherung von Elektronen. Legt man eine äußere Spannung an, so können die Elektronen ohne Behinderung über die Grenzfläche fließen. Der Kontakt zeigt dadurch ohmsches Verhalten.

## 10.2.4 Schottky-Kontakt mit angelegter Spannung

Abb. 10.20 zeigt den Bandverlauf für einen Schottky-Kontakt mit angelegter Spannung für den Fall  $\Phi_M > \Phi_H$ . Für eine positive Spannung (Durchlassrichtung) wird das Potenzial auf der Seite des Metalls angehoben, wodurch die potenzielle Energie der Elektronen abgesenkt wird. Für eine negative Spannung (Sperrrichtung) ist es genau umgekehrt. In Durchlassrichtung wird die Potentialschwelle, welche die Elektronen im Leitungsband des  $n$ -Halbleiters überwinden müssen, um ins Metall zu gelangen, abgesenkt, so dass der Strom exponentiell mit der Spannung ansteigt. Umgekehrt wird in Sperrrichtung die Potentialschwelle vergrößert, so dass der Strom mit zunehmender negativer Spannung einen Sättigungswert erreicht, weil die Elektronen die Potentialschwelle nicht mehr überwinden können. Der Wert des Sättigungsstroms  $I_s$  wird durch den Tunnelstrom durch die Potenzialbarriere bestimmt und soll hier nicht näher diskutiert werden.

Mit der gleichen Argumentation wie in Abschnitt 10.2.2 erhalten wir für die Stromspannungs-Charakteristik den Ausdruck [vergleiche (10.2.36)]

$$J(U) = I_s \left( e^{eU/k_B T} - 1 \right). \quad (10.2.51)$$

### Walter Hermann Schottky (1886–1976)

Walter Hermann Schottky wurde am 23. Juli 1886 in Zürich geboren. Er starb am 4. März 1976 in Pretzfeld.

Walter Schottky studierte Physik an der Humboldt Universität in Berlin, wo er 1912 mit einer Doktorarbeit zur Speziellen Relativitätstheorie promovierte, die von Albert Einstein sieben Jahre vorher veröffentlicht wurde. Schottky's Tutor war Max Planck. Nach Abschluss seiner Doktorarbeit ging Schottky nach Jena, wo er mit Max Wien arbeitete. Dort wechselte er sein Arbeitsgebiet von der Relativitätstheorie hin zu dem Gebiet, das sein Lebenswerk werden sollte: die Wechselwirkung von Elektronen und Ionen im Vakuum und in Festkörpern.

Für etwa 15 Jahre wechselte Schottky zwischen universitärer und industrieller Forschung. Er begann mit einigen Jahren bei Max Wien in Jena, wonach er zu den Siemens Forschungslaboratorien in Berlin wechselte und dort bis 1919 blieb. Im Jahr 1920 kehrte er wieder an die Universität zurück und arbeitet mit Wilhelm Wien in Würzburg. Dort qualifizierte er sich auch zum Hochschullehrer. Nach 3 Jahren mit Wilhelm Wien wurde Schottky zum Professor für Theoretische Physik an die Universität Rostock berufen. Im Jahr 1927, also im Alter von 41 Jahren, ging Schottky zum letzten mal zurück in die industrielle Forschung, indem er zur Siemens AG wechselte. Dort blieb er bis zu seiner Pensionierung im Jahr 1958.

Walter Schottky beeinflusste durch seine theoretischen Untersuchungen ab 1912 maßgeblich die Entwicklung der Nachrichtentechnik. Die Erfindung der Raumladungsgitterröhre und der Schirmgitterröhre, das Aufstellen der Theorie des Schroteffekts und die Erfindung des Überlagerungsempfängers gehen auf Schottky zurück, der sich von 1915 bis 1919 als wissenschaftlicher Mitarbeiter im Schwachstromkabel-Laboratorium von Siemens & Halske mit diesen Arbeiten beschäftigte. 1915 erfand Schottky die Tetrode, eine Schirmgitterröhre. Im Jahr 1918 entwickelte er das Superhet-Prinzip, einen besonders hochwertigen Rundfunkempfangskreis, der mit einer Zwischenfrequenz arbeitet. Nach seiner Lehrtätigkeit an den Universitäten Würzburg und Rostock war Schottky von 1927 bis 1951 wieder im Forschungslaboratorium bei Siemens & Halske und den Siemens-Schuckertwerken in Berlin und Pretzfeld tätig. Neben anderen wichtigen Arbeiten entwickelte er 1938 seine Randschichttheorie (auch Raumladungstheorie der Sperrschichten genannt), die sich als bahnbrechend für die Halbleitertechnik erwies. Nach Schottky benannt wurde der Schottky-Effekt (Glühemission, wichtig für die Röhrentechnik), die Schottky-Diode, die Schottky-Barriere, die Schottky-Fehlstellen und die Schottky-Gleichung (auch Langmuir-Schottkysches Raumladungsgesetz genannt).

Schon zu Lebzeiten erhielt der 1976 gestorbene Wissenschaftler zahlreiche Ehrungen. Nach ihm ist der Walter-Schottky-Preis der Deutschen Physikalischen Gesellschaft für hervorragende Leistungen in der Festkörperphysik sowie das Walter Schottky Institut der Technischen Universität München benannt.



Quelle Wikimedia Commons.

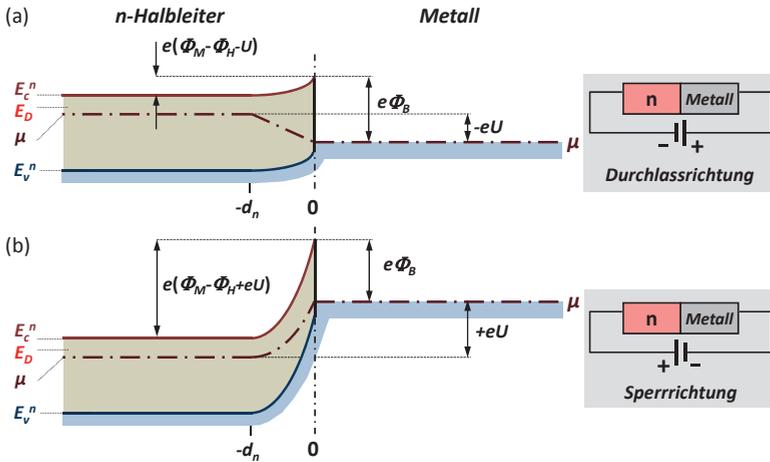


Abb. 10.20: Schematische Darstellung eines Schottky-Kontakts mit angelegter Spannung. Oben: Durchlassrichtung, unten: Sperrrichtung. In Durchlassrichtung ist die Spannung am Metall positiv, so dass die potenzielle Energie  $-eU$  der Elektronen auf der Seite des Metalls abgesenkt ist.

Die Dicke der Raumladungszone (Schottky-Randschicht) wird durch die angelegte Spannung verkleinert (Durchlassrichtung) bzw. vergrößert (Sperrrichtung). Analog zu (10.2.50) erhalten wir

$$d_n = \left( \frac{2\epsilon\epsilon_0}{en_D} (\Phi_M - \Phi_H \pm eU) \right)^{1/2}. \tag{10.2.52}$$

Die spannungsabhängige Dicke der Schottky-Randschicht führt wie beim  $p$ - $n$ -Kontakt zu einer spannungsabhängigen Kapazität

$$C_R \propto \frac{1}{d_n} \propto \sqrt{\frac{en_D}{2\epsilon\epsilon_0} \frac{1}{\Phi_M - \Phi_H \pm eU}}. \tag{10.2.53}$$

Tragen wir  $1/C_R^2$  als Funktion von  $U$  auf, so können wir durch Extrapolation auf  $U = 0$  den Wert von  $\Phi_M - \Phi_H$  bestimmen. Der gemessene Wert weicht allerdings oft von dem erwarteten Ergebnis ab. Die Ursache dafür sind Oberflächenladungen, die wir in unserer Analyse nicht berücksichtigt haben, in der Praxis aber meistens nicht zu vermeiden sind.

## 10.3 Halbleiter-Bauelemente

Wir wollen in diesem Abschnitt einige einfache Halbleiter-Bauelemente diskutieren, die auf dem  $p$ - $n$ -Übergang basieren. Im Einzelnen werden dies die Zener-Diode, die Esaki-Diode, die Solarzelle und der bipolare Transistor sein.

### 10.3.1 Zener-Diode

In Sperrrichtung eines  $p$ - $n$ -Kontakts kann keine beliebig hohe Spannung angelegt werden. Aufgrund der bei hohen Spannungen auftretenden hohen elektrischen Feldstärken treten neue Effekte auf:

■ **Lawinendurchbruch:**

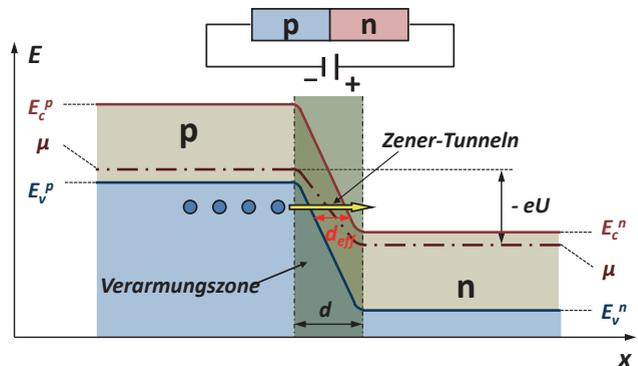
Falls die elektrische Feldstärke in der Verarmungszone groß genug wird, können die Ladungsträger durch die Beschleunigung im elektrischen Feld so viel Energie gewinnen, dass sie in der Verarmungszone Elektron-Loch-Paare durch einen Ionisationsprozess erzeugen. Da diese zusätzlich erzeugten Ladungsträger wiederum selbst beschleunigt werden und Elektron-Loch-Paare erzeugen können, kommt es zu einem lawinenartigen Anstieg der Ladungsträgerdichte und damit des Stroms, den wir Lawinendurchbruch nennen.

■ **Zener-Tunneln:**

Falls die Breite der Verarmungszone nicht allzu groß ist, können Elektronen aus dem Valenzband des  $p$ -Halbleiters direkt ins Leitungsband des  $n$ -Halbleiters tunneln. Man spricht vom **Zener-Effekt** bzw. vom **Zener-Tunneln**.<sup>22,23,24</sup> Die effektive Breite  $d_{\text{eff}}$  der Zone, die durchtunnelt werden muss, ist durch den Abstand des Valenzbandes des  $p$ -Halbleiters vom Leitungsband des  $n$ -Halbleiters gegeben. Da  $d_{\text{eff}}$  mit wachsender Sperrspannung immer dünner wird, tritt dieser Effekt ab einer genügend hohen Spannung auf (siehe Abb. 10.21). Die Tunnelwahrscheinlichkeit durch die Verarmungszone der Breite  $d_{\text{eff}}$  ist

$$T \propto e^{-2\kappa d_{\text{eff}}}, \quad (10.3.1)$$

Abb. 10.21: Schematischer Bandverlauf bei einer Zener-Diode. Bei genügend hoher Spannung in Sperrrichtung können die Ladungsträger die grün hinterlegte Verarmungszone durchtunneln, was zu einem starken Anstieg des Sperrstroms führt. In Sperrrichtung liegt der Pluspol der Spannungsquelle am  $n$ -Gebiet, so dass dort die potenzielle Energie der Elektronen um  $(-e)U$  abgesenkt wird.



<sup>22</sup> Clarence Melvin Zener, US-amerikanischer Physiker und Elektrotechniker, geboren am 1. Dezember 1905 in Indianapolis, Indiana, USA; gestorben am 2. Juli 1993 in Pittsburgh, USA.

<sup>23</sup> C. M. Zener *Non-adiabatic Crossing of Energy Levels*, Proceedings of the Royal Society of London A 137 (6), 696–702 (1932).

<sup>24</sup> E. C. G. Stückelberg, *Theorie der unelastischen Stöße zwischen Atomen*, Helvetica Physica Acta 5, 369–422 (1932).

wobei die Abklingkonstante

$$\kappa \propto \sqrt{\frac{2m^*(V_0 - E)}{\hbar^2}} \tag{10.3.2}$$

von der Barrierenhöhe  $V_0$  und der Energie  $E$  der Ladungsträger abhängt. Wir sehen, dass die Dicke der Barriere exponentiell in die Tunnelwahrscheinlichkeit eingeht, so dass aufgrund der Abnahme der effektiven Barrierendicke  $d_{\text{eff}}$  mit zunehmender Sperrspannung die Tunnelwahrscheinlichkeit und damit der Strom exponentiell ansteigt.<sup>25</sup>

Es stellt sich die Frage, unter welchen Bedingungen bei hohen Sperrspannungen ein Lawinendurchbruch und wann Zener-Tunneln auftritt. Um diese Frage zu diskutieren, nehmen wir an, dass der  $p$ - und  $n$ -Halbleiter die gleiche Konzentration von Verunreinigungen haben, so dass  $n_D = n_A = n$ . In diesem Fall erhalten wir nach (10.2.22) und (10.2.23)

$$d_n = d_p = \sqrt{\frac{\epsilon\epsilon_0 V_D}{en}} \tag{10.3.3}$$

Wir sehen, dass bei schwacher Dotierung  $d = d_n + d_p$  groß und damit die Tunnelwahrscheinlichkeit sehr klein wird. In diesem Fall werden wir bei großen Spannungen einen Lawinendurchbruch erhalten. Mit zunehmender Dotierung wird  $d = d_n + d_p$  immer kleiner und die Tunnelrate wird irgendwann bereits bei Spannungen, die kleiner als die für einen Lawinendurchbruch erforderliche Spannung sind, stark ansteigen. Wir nennen diese Spannung **Zener-Spannung**. Wir erwarten also, dass bei hoher Dotierung der Sperrstrom durch das Zener-Tunneln und nicht durch den Lawinendurchbruch ansteigt. Dies ist in Abb. 10.22 gezeigt. Mit abnehmender Dotierung verschiebt sich der Zener-Durchbruch zu höheren Spannungswerten und es tritt dann wieder ein Lawinendurchbruch auf. Für

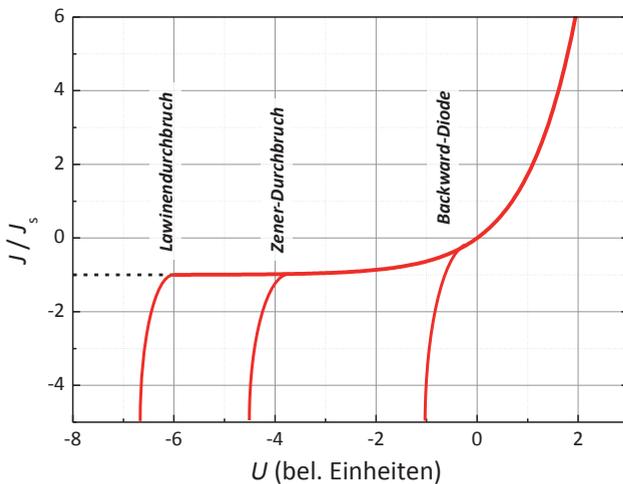


Abb. 10.22: Strom-Spannungs-Charakteristiken für den Fall des Lawinen- und des Zener-Durchbruchs. Ebenfalls gezeigt ist die Strom-Spannungs-Charakteristik einer Rückwärtsdiode (Backward Diode).

<sup>25</sup> *The Physics of Semiconductor Devices*, S. M. Sze and K. Ng Kwok, John Wiley & Sons, New York (1981).

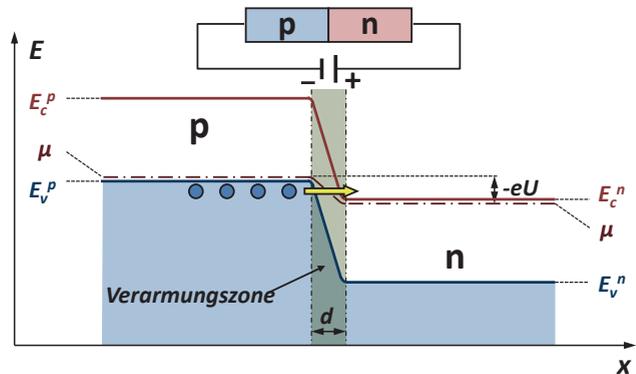
Silizium kann die Zener-Spannung bei der Herstellung durch Variation der Dotierung im Bereich zwischen etwa 2 bis 600 V eingestellt werden. Die Zener-Diode wird bei Anwendungen überwiegend in Sperrrichtung betrieben. In Durchlassrichtung arbeitet sie wie eine normale Diode.

### 10.3.1.1 Rückwärtsdiode

Bei extrem hoher Dotierung nähert sich das chemische Potenzial im  $p$ -Halbleiter der Valenzbandkante und im  $n$ -Halbleiter der Leitungsbandkante an. Außerdem wird die Breite der Verarmungszone extrem klein. Diese Situation ist in Abb. 10.23 gezeigt. Wir sehen, dass bereits bei sehr kleinen Spannungen in Sperrrichtung ein starker Tunnelstrom einsetzen kann. Dieser Tunnelstrom kann größer sein als der Strom bei der entsprechenden Spannung in Durchlassrichtung, d. h.  $I(U < 0) > I(U > 0)$ . Wir sprechen in diesem Fall von einer *Rückwärtsdiode* (engl.: Backward Diode).

Abb. 10.23: Schematischer Bandverlauf bei einer Rückwärtsdiode.

Der Tunnelstrom in Sperrrichtung setzt bereits bei sehr kleinen Spannungen ein und führt zu  $I(U < 0) > I(U > 0)$ . Für die gezeigte Sperrrichtung liegt der Pluspol der Spannungsquelle am  $n$ -Gebiet an, so dass die potenzielle Energie der Elektronen dort um  $(-e)U$  abgesenkt wird.



### 10.3.2 Esaki- oder Tunneldiode

Die Tunneldiode oder Esaki-Diode ist nichts anderes als ein  $p$ - $n$ -Kontakt, bei dem sowohl der  $n$ - als auch der  $p$ -Halbleiter so stark dotiert sind, dass das chemische Potenzial im Leitungs- bzw. Valenzband liegt (siehe Abb. 10.24a). Wir bezeichnen solche Halbleiter als entartet. Das chemische Potenzial liegt typischerweise einige  $k_B T$  von der Leitungs- bzw. Valenzbandkante entfernt und die Breite der Verarmungszone beträgt aufgrund der extrem hohen Dotierung nur etwa 10 nm.

Wie bei der Rückwärtsdiode tritt bei der Esaki-Diode für die Sperrrichtung ein hoher Tunnelstrom auf (siehe Abb. 10.24(2)). Aufgrund der Tatsache, dass das chemische Potenzial auf der  $p$ - bzw. der  $n$ -Seite im Valenz- bzw. Leitungsband liegt, erhalten wir im Gegensatz zur Rückwärtsdiode auch für die Durchlassrichtung bei kleinen Spannungen einen hohen Tunnelstrom (siehe Abb. 10.24(3)). In diesem Spannungsbereich können Elektronen aus dem Leitungsband des  $n$ -Halbleiters direkt in freie Zustände des  $p$ -Halbleiters tunneln. Für größere Spannungen in Durchlassrichtung ist dies nicht mehr der Fall, so dass der Strom zunächst wieder abnimmt und erst bei höheren Spannungswerten aufgrund der Abnahme der Breite der Potenzialbarriere für den Strom der Elektronen aus dem  $n$ - in den  $p$ -Bereich wie-

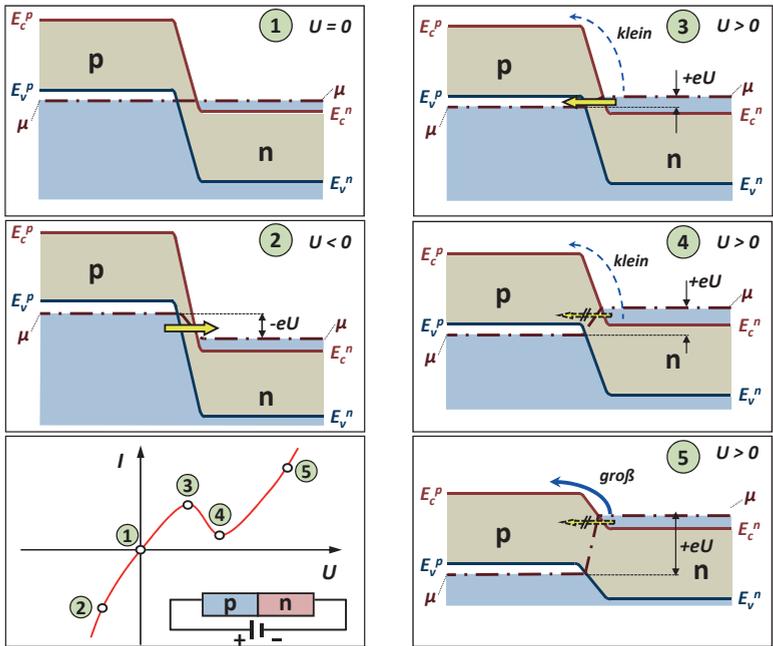


Abb. 10.24: Schematischer Bandverlauf bei einer Esaki-Diode für verschiedene Spannungen und die daraus resultierende Strom-Spannungs-Kennlinie. In Durchlassrichtung ist der Plus-Pol der Spannungsquelle mit dem *p*-Halbleiter verbunden. Dadurch wird für eine positive Spannung die potenzielle Energie der Elektronen im *p*-Gebiet um  $(-e)U$  abgesenkt, oder äquivalent im *n*-Gebiet um  $eU$  angehoben.

der exponentiell ansteigt. Der daraus resultierende negativ differentielle Widerstand wird in vielen Bauelementen ausgenutzt. Insbesondere erlaubt er die einfache Realisierung von Mikrowellenoszillatoren. Dabei wird die Esaki-Diode z. B. parallel zu einem LC-Schwingkreis geschaltet. Die Verluste des Schwingkreises, die mit einem ohmschen Widerstand charakterisiert werden können, werden dabei durch den negativ differentielle Widerstand der Esaki-Diode gerade kompensiert.

### 10.3.3 Solarzelle

Die Solarzelle wurde erstmals von **Chapin, Fuller und Pearson** im Jahr 1954 entwickelt.<sup>26</sup> Sie verwendeten diffundierte Si *p-n*-Kontakte. Bis heute wurden Solarzellen aus verschiedenen anderen Halbleitermaterialien hergestellt, wobei verschiedene Bauelementkonfigurationen auf der Basis von einkristallinen, polykristallinen und amorphen Dünnschichtstrukturen verwendet wurden. Solarzellen werden seit vielen Jahren erfolgreich für die Energieversorgung von Satelliten und Raumfahrzeugen eingesetzt. Aufgrund des weltweit steigenden

<sup>26</sup> D. M. Chapin, C. S. Fuller, G. L. Pearson, *A New Silicon p-n Junction Photocell for Converting Solar Radiation into Electrical Power*, J. Appl. Phys. 25, 676 (1954).

### Leo Esaki (geb. 1925), Nobelpreis für Physik 1973

Leo Esaki wurde am 12. März 1925 in Osaka, Japan, geboren. Bekannt wurde er vor allem durch die Erfindung der nach ihm benannten Esaki-Diode.

Esaki studierte Physik auf der Universität Tokio und machte dort 1947 seinen Bachelor of Science, sowie anschließend 1959 seine Doktorarbeit. Esaki arbeitete von 1960 an am IBM Thomas J. Watson Research Center, Yorktown Heights, New York, im Bereich der Halbleiterforschung. Im Jahr 1967 wurde er zum IBM Fellow ernannt. Bevor er zu IBM ging, arbeitete er bei der Sony Corp., wo seine Forschungsarbeiten an stark dotiertem Ge und Si in der Entdeckung der Esaki-Tunnel-Diode resultierten. Seit etwa 1969 beschäftigte sich Esaki hauptsächlich mit Halbleiter-Quantenstrukturen auf der Basis von künstlichen Halbleiter-Heterostrukturen.

Leo Esaki erhielt 1973 zusammen mit Ivar Giaever den Nobelpreis für Physik für experimentelle Entdeckungen, die das Tunnel-Phänomen in Halbleitern betrafen. Weitere Auszeichnungen, die für diese Entdeckung an Esaki verliehen wurden sind der Nishina Memorial Award (1959), der Morris N. Liebmann Memorial Prize (1961), die Stuart Ballantine Medal des Franklin Instituts (1961), der Japan Academy Award (1965), und der Order of Culture der japanischen Regierung (1974).

Im Jahr 1993 verließ Esaki IBM und wurde Präsident der University of Tsukuba in Japan.

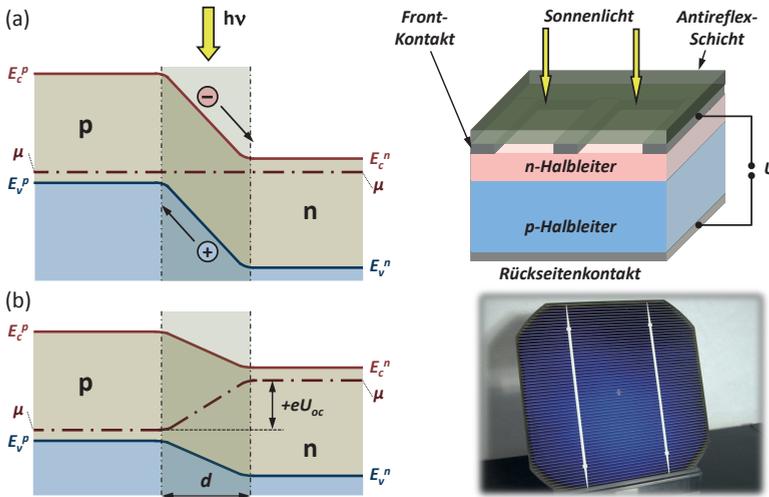


© The Nobel Foundation.

Energiebedarfs und den knapper werdenden fossilen Brennstoffen werden Solarzellen auch zunehmend für terrestrische Anwendungen interessant. Für eine breite Anwendung ist aber eine kostengünstigere und mit geringerem Energieaufwand auskommende Herstellung notwendig.

Die klassische Silizium-Solarzelle besteht aus einer ca.  $1\ \mu\text{m}$  dicken  $n$ -Schicht, welche in das ca.  $0.6\ \text{mm}$  dicke  $p$ -leitende Si-Substrat eingebracht wurde. Bei der monokristallinen Silizium-Solarzelle wird die  $n$ -Schicht durch oberflächennahes Einbringen (Dotieren) von ca.  $10^{19}$  Phosphor-Atomen pro  $\text{cm}^3$  in das  $p$ -leitende Si-Substrat erzeugt. Die  $n$ -Schicht ist so dünn, damit das Sonnenlicht hauptsächlich in der Raumladungszone am  $p$ - $n$ -Übergang absorbiert wird. Das  $p$ -leitende Si-Substrat muss dick genug sein, um die tiefer eindringenden Sonnenstrahlen absorbieren zu können und um der Solarzelle mechanische Stabilität zu geben.

Eine Solarzelle ist nichts anderes als ein  $p$ - $n$ -Kontakt, in dessen Verarmungszone durch Lichteinstrahlung Elektron-Loch-Paare erzeugt werden. Die Funktionsweise einer Solarzelle können wir uns anhand von Abb. 10.25 veranschaulichen. Ohne Bestrahlung liegt das in Abb. 10.25a gezeigte Bandschema vor. Wird die Solarzelle nun beleuchtet, werden in dem in der Verarmungszone des  $p$ - $n$ -Kontakts existierenden elektrischen Feld die durch die Lichteinwirkung erzeugten Elektron-Loch-Paare getrennt, wobei die Löcher ins  $p$ - und die Elektronen ins  $n$ -Material driften. Dadurch lädt sich das  $p$ -Gebiet positiv und das  $n$ -Gebiet ne-



**Abb. 10.25:** Zur Funktionsweise einer Solarzelle. In (a) ist der schematische Bandverlauf vor Einstrahlung von Licht gezeigt. Durch Lichteinstrahlung werden in der Verarmungszone des  $p$ - $n$ -Übergangs Elektron-Loch-Paare erzeugt. Diese werden im elektrischen Feld der Verarmungszone getrennt, wobei die Löcher ins  $p$ - und die Elektronen ins  $n$ -Material driften. Dies führt zu einer positiven bzw. negativen Aufladung des  $p$ - bzw.  $n$ -Gebiets (b) und damit zu einer Spannung  $U_{oc}$  (open circuit), die von außen abgegriffen werden kann. Rechts ist schematisch der Aufbau und ein Bild einer monokristallinen Si-Solarzelle gezeigt (Photo: Stephan Kambor).

gativ auf. Die potenzielle Energie der Elektronen im  $n$ -Gebiet wird um  $(-e)(-U)$ , diejenige der Löcher im  $p$ -Gebiet ebenfalls um  $(+e)(+U)$  angehoben. Dies führt zu einer positiven Spannung über den  $p$ - $n$ -Kontakt, die wir von außen abgreifen können. Ohne äußere Last baut sich die „open circuit“ Spannung  $U_{oc}$  auf.

Prinzipiell können nicht mehr Elektron-Loch-Paare erzeugt werden, als Lichtquanten auf die Zelle treffen. Von den Lichtquanten kann auch nur der Anteil genutzt werden, dessen Energie  $h\nu$  größer als der Bandabstand  $E_g$ , also groß genug ist, ein Elektron-Loch-Paar durch Anregung eines Elektrons aus dem Valenzband ins Leitungsband zu erzeugen. Weiterhin kann ein Lichtquant relativ hoher Energie nur den Teil in nutzbare elektrische Energie umwandeln, der dazu benötigt wird, das Elektron-Loch-Paar zu erzeugen. Daraus resultiert für kommerziell verfügbare, mit vertretbarem Aufwand herstellbare Solarzellen – z. B. polykristalline Siliziumzellen – ein Wirkungsgrad von etwa 15%. Von den maximal etwa 1000 W Sonneneinstrahlung pro Quadratmeter können also nur 150 W in Form elektrischer Leistung verfügbar gemacht werden. Die Angabe des Wirkungsgrades kann auf zwei Flächenangaben bezogen werden: einmal auf die Zellenfläche, zum anderen auf die Modulfläche. Im ersten Fall wird nur die Zellentechnik berücksichtigt, im zweiten Fall zusätzlich die durch die Zellengeometrie benötigte Fläche.<sup>27</sup> Eine grobe Einteilung von Solarzellen können wir nach dem in Tabelle 10.6 gezeigten Schema vornehmen.

<sup>27</sup> So kann der zellenbezogene Wirkungsgrad für eine kreisförmige Zelle sehr hoch sein, in einem Modul wird er niedriger, weil kreisförmige Zellen nicht flächendeckend angeordnet werden können.

**Tabelle 10.6:** Einteilung von Solarzellen nach Halbleitermaterial und Kristallinität bzw. Schichtstruktur. Angegeben sind auch Wirkungsgrade, die im Labor und in der Produktion erreicht werden (Stand 2012).

Halbleitermaterial	Kristallinität/ Schichtstruktur	Wirkungsgrad Labor (%)	Wirkungsgrad Produktion (%)
Si	amorph	13	5–8
	polykristallin	20	13–16
	monokristallin	25	14–17
GaAs	Einschicht	25	15–22
GaAs/GaInP/GaInAs	Mehrschicht	40	20–28
Cu(In,Ga)Se <sub>2</sub>	Einschicht	20	13–15
CdTe	Einschicht	16.5	5–12
Organische Halbleiter	Einschicht	6.5	—

Um den Wirkungsgrad einer Solarzelle abzuschätzen, betrachten wir das in Abb. 10.26 gezeigte Ersatzschaltbild. Der durch das Generieren von Elektron-Loch-Paaren erzeugte elektrische Strom  $I_L$  kann durch eine Stromquelle beschrieben werden, die zum  $p$ - $n$ -Kontakt parallel geschaltet ist und einen Strom erzeugt, der parallel zum Sperrstrom gerichtet ist. Ohne äußere Last ( $R_L \rightarrow \infty$ ) ist der Strom im äußeren Kreis null und wir können den Spannungsabfall  $U_{oc}$  dadurch bestimmen, dass wir den Strom des  $p$ - $n$ -Kontakt dem durch die Lichtbestrahlung erzeugten Strom  $I_L$  gleichsetzen. Wir erhalten mit (10.2.51)

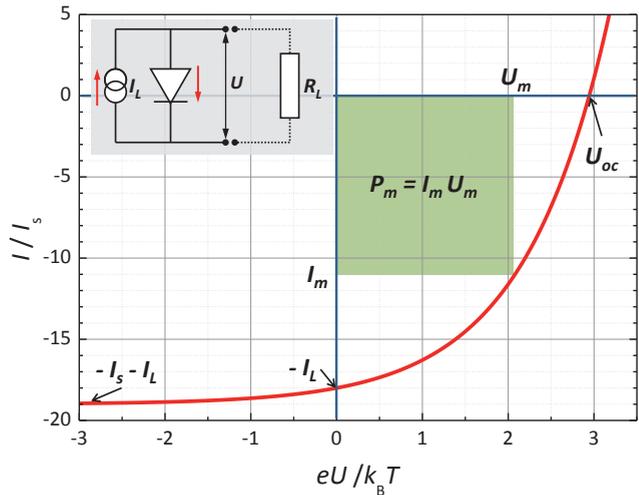
$$I = 0 = I_s (e^{eU_{oc}/k_B T} - 1) - I_L. \tag{10.3.4}$$

Auflösen nach  $U_{oc}$  ergibt die open circuit Spannung

$$U_{oc} = \frac{k_B T}{e} \ln \left( \frac{I_L}{I_s} + 1 \right) \approx \frac{k_B T}{e} \ln \left( \frac{I_L}{I_s} \right). \tag{10.3.5}$$

Schließen wir die Solarzelle von außen kurz ( $R_L \rightarrow 0$ ), so ist der Spannungsabfall über den  $p$ - $n$ -Kontakt null und es muss  $I_s = -I_L$  gelten. Insgesamt erhalten wir die in Abb. 10.26 gezeigte Strom-Spannungs-Kennlinie

$$I = I_s (e^{eU/k_B T} - 1) - I_L. \tag{10.3.6}$$



**Abb. 10.26:** Ersatzschaltbild und Strom-Spannungs-Kennlinie einer Solarzelle. Die maximale Leistung der Solarzelle ergibt sich aus der maximalen Fläche  $P_m = I_m \cdot U_m$ .

Die maximale Leistung, die aus der Solarzelle gewonnen werden kann, ergibt sich aus der maximalen Fläche  $P_m = I_m \cdot U_m$  des in Abb. 10.26 eingezeichneten Rechtecks. Setzen wir  $dP/dU = 0$ , so erhalten wir

$$I_m = I_s \frac{eU_m}{k_B T} e^{eU_m/k_B T} \quad (10.3.7)$$

$$U_m = \frac{k_B T}{e} \ln \left( \frac{I_L/I_s + 1}{\frac{eU_m}{k_B T} + 1} \right) = U_{oc} - \frac{k_B T}{e} \ln \left( \frac{eU_m}{k_B T} + 1 \right) \quad (10.3.8)$$

und damit  $P_m = I_L \cdot (E_m/e)$ . Hierbei ist

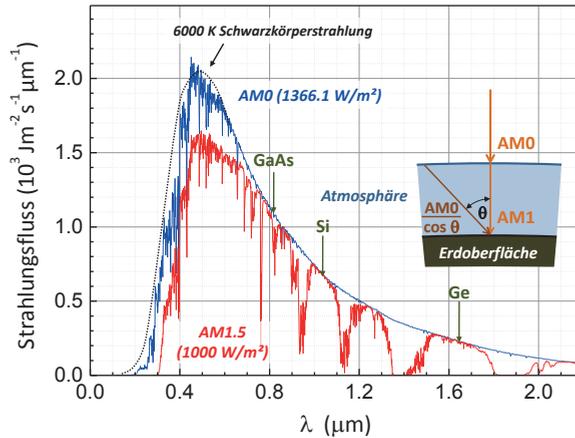
$$E_m = e \left\{ U_{oc} - \frac{k_B T}{e} \ln \left( \frac{eU_m}{k_B T} + 1 \right) \right\} \quad (10.3.9)$$

diejenige Energie, die pro erzeugten Ladungsträger an die Last abgegeben wird. Die Gleichungen (10.3.7) bis (10.3.9) zeigen, dass wir ein großes Verhältnis  $I_L/I_s$  benötigen, um  $E_m$  groß zu machen. Für ein bestimmtes Halbleitermaterial kann die Sättigungsstromdichte nach Gleichung (10.2.47) berechnet werden. Bei 300 K liegt für Silizium die kleinste erreichbare Sättigungsstromdichte im Bereich  $J_s \sim 10^{-15} \text{ A/cm}^2$ . Eine Erhöhung von  $I_L/I_s$  bei vorgegebenem Material und Beleuchtungsstärke kann durch Erniedrigung der Temperatur oder Verwendung von Konzentratorzellen erzielt werden.

Um den Wirkungsgrad einer Solarzelle anzugeben, müssen wir überlegen, wie effizient die ankommende Strahlungsleistung in elektrische Leistung umgesetzt wird. Dabei gehen unter anderem folgende Faktoren ein: (i) Photonen mit  $h\nu < E_g$  können keine Elektron-Loch-Paare anregen und tragen deshalb nicht zur elektrischen Leistung bei. (ii) Photonen mit  $h\nu > E_g$  regen zwar Elektron-Loch-Paare an, die pro absorbiertes Photon an die Last abgegebene Energie ist aber  $E_m < h\nu$ . (iii) Photonen können an der Oberfläche der Zelle reflektiert werden und tragen dann nicht zur elektrischen Leistung bei. (iv) Durch Kontaktierungsschichten und Montagevorrichtungen ist die effektive Zellfläche kleiner als 100 %. Um den Einfluss der Bandlücke  $E_g$  zu diskutieren, betrachten wir den Kurzschlussstrom  $I_L$ . Dieser hängt davon ab, wie der von der Sonne kommende Photonenfluss (Photonen pro Fläche und Zeit) mit spektraler Verteilung  $n_{ph}(\nu)$  (siehe Abb. 10.27) in einen elektrischen Strom  $I_L$  umgesetzt wird. Dazu müssen wir berücksichtigen, dass nur Photonen mit  $h\nu \geq E_g$  Elektron-Loch-Paare generieren können. Um die vom Photonenfluss erzeugte elektrische Stromdichte zu erhalten, integrieren wir den Photonenfluss der Sonne von  $E_g$  bis  $\infty$  und erhalten

$$J_L(E_g) = e \int_{h\nu=E_g}^{\infty} \frac{dn_{ph}(\nu)}{d(h\nu)} d(h\nu). \quad (10.3.10)$$

Das Ergebnis der Integration ist in Abb. 10.28 gezeigt. Nachdem wir den Sättigungsstrom  $J_s$  und  $J_L$  kennen, können wir  $E_m$  durch numerische Lösung der Gleichungen (10.3.5), (10.3.8) und (10.3.9) erhalten. Da  $J_s$  und damit  $E_m$  von den Materialeigenschaften des Halbleiters abhängt, müssen zum Erreichen der optimalen Effizienz einer Solarzelle die Materialparameter so optimiert werden, dass  $J_s$  minimal wird.



**Abb. 10.27:** Von der Sonne abgestrahlter Strahlungsfluss als Funktion der Wellenlänge. Die Pfeile markieren die maximale Wellenlänge  $\lambda_m = hc/E_g$ , bis zu der bei den verschiedenen Halbleitern Band-Band-Übergänge möglich sind. Das langwellige Spektrum rechts der Pfeile kann in Solarzellen aus dem jeweiligen Material nicht ausgenutzt werden. Das extraterrestrische Spektrum ohne Abschwächung durch die Erdatmosphäre wird mit AM0 bezeichnet. Es besitzt eine integrierte Strahlungsleistung von  $1366.1 \text{ W/m}^2$ . Die Strahlungsleistung kann in einen Photonenfluss  $dn_{\text{ph}}/d(h\nu)$  umgerechnet werden, der in Photonen/ $\text{eV m}^2 \text{ s}$  angegeben wird. Im Maximum der AM0-Kurve hat man etwa  $4 \times 10^{21}$  Photonen/ $\text{eV m}^2 \text{ s}$ . Durch die Drehung der Erde um die Sonne ändert sich der Einfallswinkel der Strahlung und damit auch die Länge des Weges durch die Atmosphäre. Um diese Änderung zu charakterisieren, wurde der Begriff der „Air Mass“ (AM) eingeführt. AM1 kennzeichnet den Strahlungsfluss auf der Erdoberfläche (Meereshöhe) bei senkrechtem Einfall des Sonnenlichtes und besitzt durch die Absorption der Atmosphäre eine integrierte Strahlungsleistung von nur etwa  $925 \text{ W/m}^2$  (AM1). Für einen Zenitwinkel  $\theta$  der Sonne ergibt sich ein entsprechendes AMX-Spektrum mit  $X = 1/\cos \theta$ , z. B. AM1.5 für  $\theta = 48.19^\circ$ . Für Testzwecke wurde das AM1.5 Spektrum (ISO 9845-1) definiert. Es besitzt die typische spektrale Verteilung auf der Erdoberfläche mit einer integrierten Strahlungsleistung von  $1000 \text{ W/m}^2$ . Diese ist höher als der Wert von  $844 \text{ W/m}^2$ , der für die tatsächliche terrestrische Sonnenstrahlung bei einer Luftmasse von 1.5 gemessen wird.

Die charakteristische Energie  $E_m$  ist ebenfalls in Abb. 10.28 als Funktion der Photonenenergie gezeigt. Die ideale Konversionseffizienz einer Solarzelle ist gegeben durch

$$\eta = \frac{P_m}{P_{\text{in}}} = \frac{I_L \frac{E_m}{e}}{P_{\text{in}}}, \quad (10.3.11)$$

das heißt, durch das Verhältnis der maximalen Ausgangsleistung  $P_m$  und der einfallenden Strahlungsleistung  $P_{\text{in}}$ . Da  $P_m$  durch die maximal mögliche Rechtecksfläche unter der Kurve  $E_m(E_g)$  und  $P_{\text{in}}$  durch die gesamte Fläche unter der Kurve  $J_L(E_g)$  gegeben ist, kann die ideale Konversionseffizienz grafisch mit Hilfe von Abb. 10.28 ermittelt werden. Für  $E_g \rightarrow 0$  geht die rote Fläche und damit  $\eta$  gegen null, da zwar alle ankommenden Photonen Elektron-Loch-Paare erzeugen können, die pro absorbiertes Photon an die Last abgegebene Energie  $E_m$  aber gegen null geht. Für sehr große  $E_g$  Werte wird zwar  $E_m$  groß, aber nur wenige Photonen können jetzt noch Elektron-Loch-Paare anregen und damit wird  $\eta$  auch hier klein. Die maximal mögliche Rechtecksfläche ergibt sich für  $E_g \approx 1.35 \text{ eV}$ , was dem Energielückenwert von GaAs nahe kommt. Man erhält für diesen Wert eine maximale Effizienz

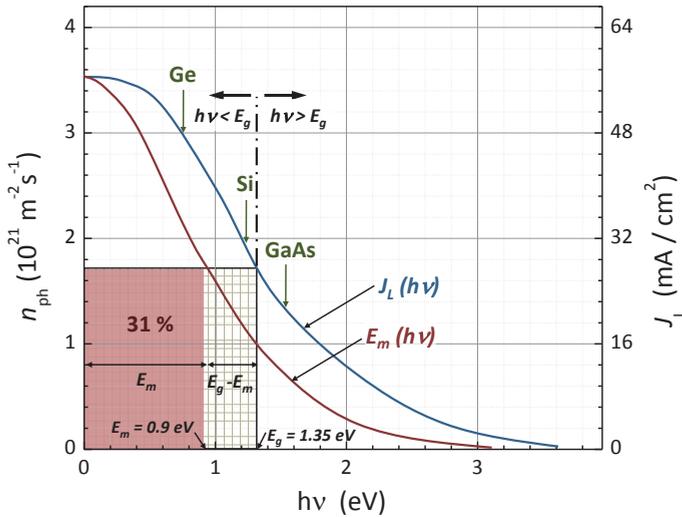


Abb. 10.28: Schematische Darstellung des Photonenflusses  $n_{ph}$  bzw. äquivalent von  $J_L$  als Funktion der Photonenenergie  $h\nu$  berechnet nach (10.3.10) für AM1.5.  $J_L$  ist äquivalent zur Anzahl  $n_{ph}$  der Photonen pro Fläche und Zeit in der auf die Erdoberfläche einfallenden Sonnenstrahlung. Ebenfalls gezeigt ist charakteristische Energie  $E_m$  als Funktion der Photonenenergie  $h\nu$ . Die eingezeichneten Rechtecke veranschaulichen die grafische Bestimmung der Konversionseffizienz  $\eta$ . Die Pfeile markieren die Bandlücke von Ge, Si und GaAs.

von etwa 31 %. Die ideale Konversionseffizienz besitzt ein breites Maximum zwischen  $E_g = 1$  und  $1.5$  eV, hängt also über einen weiten Bereich nicht kritisch von  $E_g$  ab. Glücklicherweise gibt es mehrere Halbleitermaterialien mit Energielücken in diesem Bereich. Die Effizienz, die in der Praxis erhalten wird, liegt mehr oder weniger weit unter der erreichbaren idealen Effizienz. Einige Ursachen dafür wurden oben bereits genannt, sollen hier aber nicht im Einzelnen diskutiert werden.

Aufgrund der großen Bedeutung des Wirkungsgrades von Solarzellen für die Anwendung wurden verschiedene Strategien entwickelt, mit denen die Effizienz von Solarzellen optimiert werden kann:

- Oberflächenstrukturierung zur Verminderung von Reflexionsverlusten: Die Zelloberfläche kann z. B. in einer Pyramidenstruktur aufgebaut werden, damit einfallendes Licht mehrfach auf die Oberfläche trifft.
- Neue Materialien: Neben Silizium kommen zum Beispiel Galliumarsenid (GaAs), Cadmiumtellurid (CdTe), Kupfer-Indium-Diselenid ( $\text{CuInSe}_2$ )<sup>28</sup> oder organische Halbleiter<sup>29</sup> zum Einsatz. Mit GaAs-Zellen, die auf Ge-Substraten gewachsen wurden, konnte

<sup>28</sup> Miguel Contreras *et al.*, 19.9%-efficient ZnO/CdS/CuInGaSe<sub>2</sub> solar cell with 81.2% fill factor, Progress in Photovoltaics: Research and Applications 16, 235 (2008).

<sup>29</sup> J.-Y. Kim *et al.*, Efficient tandem polymer solar cells fabricated by all-solution processing, Science 317, 222 (2007).

vor kurzem ein Wirkungsgrad von 24.7 % mit  $U_{oc} = 999$  mV und  $J_L = 29.7$  mA/cm<sup>2</sup> erreicht werden.

- Tandem- oder Stapelzellen: Um ein breiteres Strahlungsspektrum nutzen zu können, werden unterschiedliche Halbleitermaterialien, die für verschiedene Spektralbereiche geeignet sind, übereinander angeordnet.
- Konzentrazorzellen: Durch die Verwendung von Spiegel- und Linsensystemen wird eine höhere Lichtintensität auf die Solarzellen fokussiert. Diese Systeme werden der Sonne nachgeführt, um stets die direkte Strahlung auszunutzen. Durch die Kombination von Konzentrazor- und Stapelzellen konnte vor kurzem ein Wirkungsgrad oberhalb von 40% erreicht werden.<sup>30</sup> Dieser Wirkungsgrad liegt oberhalb des theoretisch maximalen Wirkungsgrades von 31 % von Einzelzellen.
- MIS-Inversionsschicht-Zellen: Das innere elektrische Feld wird nicht durch einen  $p$ - $n$ -Übergang erzeugt, sondern durch den Übergang einer dünnen Oxidschicht (I) zu einem Halbleiter (S).
- Grätzel-Zelle: Es wird eine elektrochemische Flüssigkeitszelle mit Titandioxid als Elektrolyten und einem Farbstoff zur Verbesserung der Lichtabsorption verwendet.

### 10.3.4 Bipolarer Transistor

Der bipolare Transistor wurde im Jahr 1947 von **John Bardeen**,<sup>31</sup> **Walter H. Brattain**<sup>32</sup> und **William Bradford Shockley**<sup>33</sup> entwickelt (siehe Abb. 10.29).<sup>34</sup> Sie erhielten dafür im Jahr 1956 den Nobelpreis für Physik.

Der schematische Aufbau und der Bandverlauf eines bipolaren  $npn$ -Transistors ist in Abb. 10.30 gezeigt. Ein bipolarer Transistor besteht aus zwei  $p$ - $n$ -Übergängen, wobei der eine (Emitter-Basis-Kontakt) in Durchlassrichtung und der andere (Basis-Kollektor-Kontakt)

<sup>30</sup> R. R. King *et al.*, *40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells*, Applied Physics Letters **90**, 183516 (2007).

<sup>31</sup> **John Bardeen**, geboren am 23. Mai 1908 in Madison Wisconsin, gestorben am 30.01.1991 in Boston.

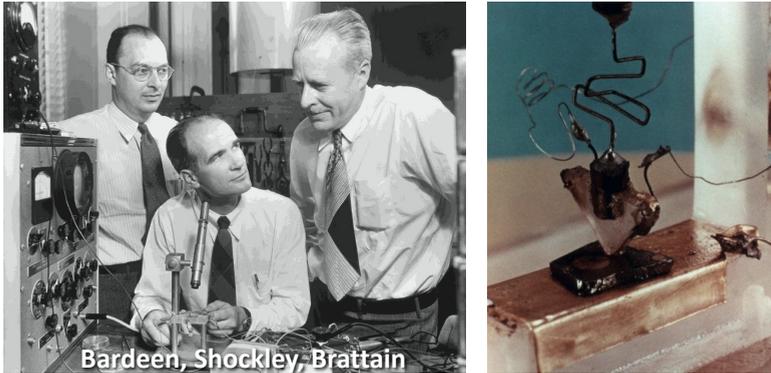
<sup>32</sup> **Walter H. Brattain**, geboren am 10. Februar 1902 in Amoy, China, gestorben am 13. Oktober 1987 in Seattle.

<sup>33</sup> **William Bradford Shockley**, geboren am 13. Februar 1910 in London, gestorben am 12. August 1989 in London.

<sup>34</sup> Die Erfindung des bipolaren Transistors wird allgemein auf Dezember 1947 in den Bell Laboratories datiert. Beteiligt an der Erfindung waren William B. Shockley, John Bardeen und Walter Brattain, die 1956 den Nobelpreis dafür erhielten. In den 1950er Jahren gab es einen Wettlauf zwischen Röhre und Transistor, in dessen Verlauf die Chancen des Transistors häufig eher skeptisch beurteilt wurden.

Zuerst wurden Transistoren aus Germanium hergestellt und ähnlich wie Röhren in winzige Glasröhrchen eingeschmolzen. Das Germanium wurde später durch Silizium ersetzt. Es werden auch Mischmaterialien benutzt, diese sind aber seltener vertreten.

Wenn man alle Transistoren in sämtlichen bislang hergestellten Schaltkreisen (Speicher, Prozessoren usw.) zusammenzählt, ist der Transistor inzwischen diejenige technische Funktionseinheit, die bis heute von der Menschheit mit der höchsten Gesamtstückzahl produziert wurde. Heute werden pro Jahr weit mehr als eine Trillion Transistoren produziert.



**Abb. 10.29:** Links: John Bardeen, William B. Shockley und Walter Brattain (von links nach rechts im Jahr 1956) entwickelten den Transistor im Dezember 1947 in den Bell Laboratories, Murray Hill (Photo: Nick Lazarnick, Bell Laboratories, mit freundlicher Genehmigung durch AIP Emilio Segre Visual Archives). Rechts: Photographie des Versuchsaufbaus (Quelle: Lucent Technologies).

in Sperrrichtung geschaltet ist. Bei dem in Durchlassrichtung geschalteten Emitter-Basis-Kontakt bewirkt eine kleine Änderung  $\delta U_{EB}$  der Emitter-Basis-Spannung eine große Änderung des Emitter-Basis-Stroms  $I_{EB}$ . Bei dem in Sperrrichtung geschalteten Basis-Kollektor-Kontakt bewirkt dagegen eine Veränderung  $\delta U_{BC}$  der Basis-Kollektor-Spannung kaum eine Veränderung des Basis-Kollektor-Stroms  $I_{BC}$ . Der über den Emitter-Basis-Kontakt in die Basisschicht injizierte Strom (in dem in Abb. 10.30 gezeigten *npn*-Transistor sind dies überwiegend Elektronen) kann entweder über die Basis oder den Kollektor abfließen. Da die Breite der Basis viel kleiner als die Diffusionslänge der Ladungsträger im Basismaterial ist, werden fast alle über den Emitter-Basis-Kontakt injizierten Ladungsträger über den Basis-Kollektor-Kontakt abgesaugt. Das bedeutet, dass der Emitterstrom  $I_{EB}$  etwa gleich dem Kollektorstrom  $I_{BC}$  ist und somit

$$I_{EB} \simeq I_{BC} \quad \Rightarrow \quad I_B \simeq 0 \quad (10.3.12)$$

gilt. Wir können daraus sofort ersehen, dass die Leistung im Eingangskreis  $P_{EB} = I_B \cdot U_{EB} \simeq 0$  und die Leistung im Ausgangskreis aufgrund der großen Basis-Kollektor-Spannung  $P_{BC} = I_{BC} \cdot U_{CB} \gg 0$ .

Auf der Basis der bisherigen Diskussion können wir einfach verstehen, wie wir einen bipolaren Transistor als verstärkendes Bauelement verwenden können. Das zu verstärkende Spannungssignal wird an die Emitter-Basis-Diode angelegt und resultiert in einer großen Änderung  $\delta I_{EB}$  des Emitter-Basis-Stroms. Da der injizierte Strom fast vollständig über die Basis-Kollektor-Diode abgesaugt wird, d. h.  $\delta I_{EB} \simeq \delta I_{BC}$ , erfolgt die Steuerung im Eingangskreis wegen  $I_B \simeq 0$  quasi leistungslos. Im Ausgangskreis erhalten wir dagegen eine große Leistungsänderung  $\delta P_{BC} = \delta I_{EB} \cdot U_{BC}$ .

Für eine optimale Funktion des bipolaren Transistors sollte die Basisschicht möglichst dünn sein (typischerweise 5–25  $\mu\text{m}$ ), damit in dieser keine Rekombination der injizierten Elektronen mit den dort in großer Dichte vorhandenen Löchern erfolgt. Bei einer endlichen Rekombinationsrate wird  $I_B > 0$ , das heißt, wir erhalten eine endliche Verlustleistung im Ein-

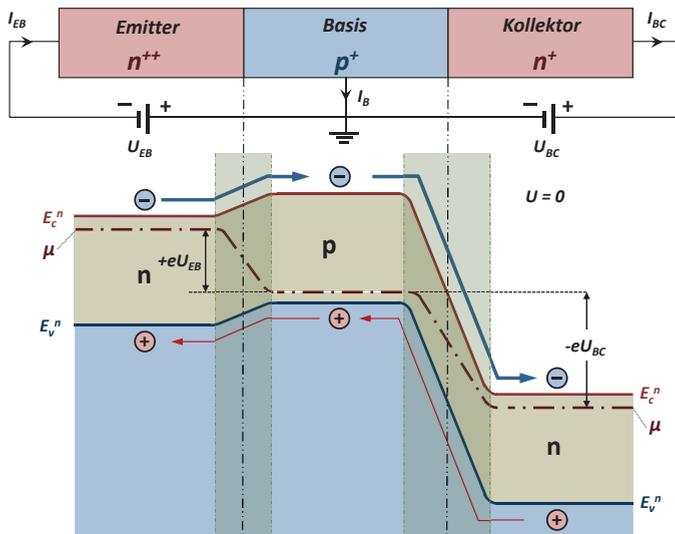


Abb. 10.30: Schematischer Aufbau (oben) eines  $npn$ -Transistors und Bandverlauf für die Situation, dass der Emitter-Basis-Kontakt in Durchlass- und der Basis-Kollektor-Kontakt in Sperrrichtung geschaltet ist (unten). Die Breite der Basisschicht ist klein gegenüber der Diffusionslänge der Ladungsträger im Basismaterial, so dass die über die Emitter-Basis-Diode injizierten Ladungsträger fast alle zum Basis-Kollektor-Übergang diffundieren können und dort durch die anliegende Sperrspannung abgesaugt werden. Mit den dicken blauen Pfeilen bzw. dünnen roten Pfeilen wird der Majoritäts- bzw. Minoritätsladungsträger- bzw. Minoritätsladungsträgerstrom angedeutet.

gangskreis. Die Dicke  $d_B$  der Basis bestimmt ferner die obere Grenzfrequenz des Transistors. Mit der charakteristischen Driftgeschwindigkeit  $v_D$  der Ladungsträger kann die Größenordnung der Grenzfrequenz zu  $f_G = v_D/d_B$  abgeschätzt werden. Durch Reduktion der Dicke der Basisschicht kann deshalb auch die Grenzfrequenz des Bauelements erhöht werden.

#### 10.3.4.1 Dreitor-Bauelemente

Der bipolare Transistor gehört zu den so genannten *Dreitor-Bauelementen*, die in sehr allgemeiner und anschaulicher Weise mit dem Landauerschen Flüssigkeitsmodell<sup>35</sup> beschrieben werden können (siehe Abb. 10.31). In diesem Modell wird eine Flüssigkeit A, die sich entlang einer Röhre bewegt (z. B. Elektronenflüssigkeit vom Emitter zum Kollektor) durch einen

<sup>35</sup> **Rolf Landauer**, geboren am 4. Februar 1927 in Stuttgart, gestorben am 27. April 1999 in Briarcliff Manor, New York.

Landauer musste aufgrund seiner jüdischen Herkunft Deutschland bereits in früher Jugend verlassen. Er studierte Physik an der Harvard University. Nach seinem Studium arbeitete er für etwa 2 Jahre bei dem National Advisory Committee for Aeronautics (der späteren NASA) bevor er 1952 zu IBM ans Thomas J Watson Research Center in Yorktown Heights, New York wechselte. Dort arbeitete er 47 Jahre lang, im Jahr 1967 wurde er zum IBM Fellow ernannt. Landauer interessierte sich unter anderem für die Grenzen der Datenverarbeitung. Bereits 1961 zeigte er, dass Rechenoperationen im Prinzip keine Energie verbrauchen, dass allerdings das Löschen von Information mit einer Dissipation verbunden ist. Landauer beschäftigte sich außerdem intensiv mit dem Verhalten von Elektronen in Festkörpern, insbesondere in Halbleitern und Nanostrukturen.

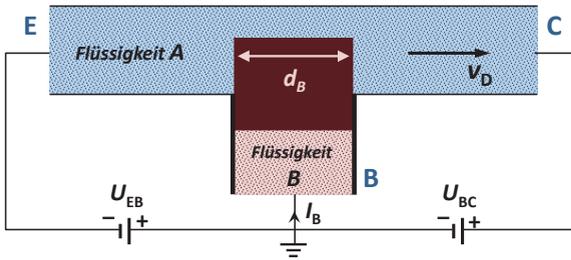


Abb. 10.31: Veranschaulichung der prinzipiellen Funktionsweise eines Dreitor-Bauelements mit Hilfe des Landauerschen Flüssigkeitsmodells.

Kolben gesteuert (z. B. durch Emitter-Basis-Spannung). Falls der Kolben dicht ist, d. h. falls die Steuerflüssigkeit B sich nicht mit der in der Röhre strömenden Flüssigkeit A vermischen kann, ist der Fluss der Steuerflüssigkeit null und die Steuerung erfolgt leistungslos.

Als Flüssigkeit kann eine gewöhnliche Flüssigkeit wie Wasser, aber auch Elektronen, wie z. B. in Halbleiter-Transistoren oder Elektronenröhren, oder Flussquanten, wie z. B. in den fluxonischen Bauelementen der Supraleitungselektronik verwendet werden. Entscheidend für die gute Funktionsweise und die maximale Geschwindigkeit eines Dreitor-Bauelements ist, dass sich die Flüssigkeit gut steuern lässt und sich die Flüssigkeitsteilchen schnell bewegen. Wie oben bereits diskutiert wurde, ist die obere Grenzfrequenz des Bauelements durch  $f_G = v_D/d_B$  gegeben, wobei  $v_D$  die Driftgeschwindigkeit der Flüssigkeitsteilchen und  $d_B$  die Kanallänge unter dem Kolben ist. Beide Voraussetzungen sind für elektronische Bauelemente gut erfüllt. Elektronen können durch elektrische Felder leicht beeinflusst werden und besitzen ferner hohe Geschwindigkeiten. Das gleiche gilt für Fluxonen in supraleitenden Bauelementen, die sich gut mit Magnetfeldern steuern lassen und sich mit Geschwindigkeiten bis etwa  $1/10$  der Lichtgeschwindigkeit bewegen können. Hinsichtlich der Geschwindigkeit wären Photonen ideal, die sich mit Lichtgeschwindigkeit bewegen. Allerdings lassen sich Photonen nur schwer beeinflussen, d. h. es ist schwierig, einen einfachen Kolben zur Steuerung des Photonensflusses zu realisieren.

## 10.4 Realisierung von niedrigdimensionalen Elektronengassystemen

Halbleiterstrukturen werden heute häufig dazu verwendet, niedrigdimensionale Elektronengassysteme herzustellen. Dabei werden nulldimensionale Systeme als *Quantenpunkte*, eindimensionale Systeme als *Quantendrähte* und zweidimensionale Systeme als *Zweidimensionale Elektronengase* (2DEG: two dimensional electron gas) bezeichnet. Wie in Abschnitt 7.4 bereits gezeigt wurde, erhalten wir niedrigdimensionale Elektronengassysteme dadurch, dass wir ein dreidimensionales Elektronengas durch Potenzialwälle in einer, zwei oder allen drei Raumrichtungen geometrisch einschränken. Für die Realisierung von niedrigdimensionalen Elektronengassystemen werden häufig Halbleiter-Heterostrukturen und -Übergitter verwendet, die aus unterschiedlich dotierten Halbleitern oder aus Halbleitern mit unterschiedlicher Energielücke aufgebaut sind. Wir wollen im Folgenden den Aufbau und den Bandverlauf in einigen Halbleiter-Heterostrukturen diskutieren.

## 10.4.1 Zweidimensionale Elektronengase

Mit Hilfe von Depositionstechniken wie der Molekularstrahlepitaxie (MBE: molecular beam epitaxy) oder der metallorganischen Gasphasenepitaxie (MOCVD: metal organic chemical vapor deposition) ist es heute möglich, verschiedene Halbleitermaterialien in einkristalliner Form übereinander aufzuwachsen. Man spricht von heteroepitaktischem Wachstum.<sup>36</sup> Voraussetzung dafür ist, dass die Halbleitermaterialien ähnliche Gitterkonstanten besitzen, um epitaxiale Verspannungseffekte zu minimieren. Wie wir bereits in Abschnitt 4.4 diskutiert haben, gibt es verschiedene Familien von Halbleitermaterialien, deren Gitterparameter sehr gut zusammenpassen (siehe Abb. 4.5). Der interessante Aspekt ist nun, dass wir Halbleitermaterialien mit unterschiedlichen physikalischen Parametern wie Größe der Energielücke und Größe der Elektronenaffinität übereinander aufwachsen können und damit die Form des Verlaufs der Leitungs- und Valenzbandkante senkrecht zur Wachstumsrichtung gezielt beeinflussen können. Die ersten Vorschläge zur Erzeugung von Übergittern durch periodische Modulation der Zusammensetzung von Halbleitern stammen von R. Tsu und L. Esaki aus dem Jahr 1970.<sup>37,38,39</sup> Das gezielte Einstellen des Bandverlaufs wird heute bereits in Bauelementen ausgenutzt und wird in Zukunft sicherlich an Bedeutung gewinnen. Sowohl die Gitterfehlpassung als auch der Unterschied in der elektronischen Bandstruktur wird dabei häufig kontinuierlich durch Verwendung von ternären oder quaternären Legierungen eingestellt.

Um die grundlegenden Aspekte von Halbleiter-Heterostrukturen zu veranschaulichen, betrachten wir eine Struktur aus Ge und GaAs. Beide Halbleiter haben etwa die gleiche Gitterkonstante von 5.65 Å (siehe Abb. 4.5), wobei Ge in der Diamant- und GaAs in der Zinkblende-Struktur kristallisiert. Es besteht aber ein großer Unterschied bezüglich ihrer Energielücken, die 0.67 eV für Ge und 1.43 eV für GaAs betragen. Es ist insbesondere dieser Energielückenunterschied, der die Kombination der beiden Materialien in Heterostrukturen interessant macht.

Neben der Größe der Energielücke ist auch die Elektronenaffinität der Materialien wichtig, da der Unterschied der Elektronenaffinität die Größe der Banddiskontinuität

$$\Delta E_c = e(\chi_1 - \chi_2) = e\Delta\chi \quad (10.4.1)$$

bestimmt. Je nach Größe von  $\Delta\chi$  erhalten wir die in Abb. 10.32 gezeigten prinzipiellen Typen von Banddiskontinuitäten in Halbleiter-Heterostrukturen. Für Ge/GaAs Heterostrukturen erwartet man aufgrund von Bandstrukturrechnungen, dass die Valenzbandkante von Ge etwa 0.42 eV oberhalb und die Leitungsbandkante etwa 0.35 eV unterhalb der von GaAs liegt. Dies entspricht dem mit „normal“ bezeichneten Typ einer Banddiskontinuität.

<sup>36</sup> Von Epitaxie spricht man allgemein, wenn eine definierte Beziehung zwischen der kristallographischen Ausrichtung eines Substrats und einer Beschichtung vorliegt. Bei der Homoepitaxie sind Substrat- und Schichtmaterial gleich, bei der Heteroepitaxie unterschiedlich.

<sup>37</sup> L. Esaki und R. Tsu, *Superlattice and Negative Differential Conductivity in Semiconductors*, IBM J. Res. Develop. 14, 61 (1970).

<sup>38</sup> L. Esaki, *Compositional Superlattices*, in *The Technology and Physics of Molecular Beam Epitaxy*, herausgegeben von E. H. C. Parker, Plenum Press, New York (1985).

<sup>39</sup> *Two-dimensional Systems: Physics and New Devices*, herausgegeben von G. Bauer, F. Kuchar, H. Heinrich, Springer Series on Solid State Science, Vol. 67, Springer Berlin, Heidelberg (1986).

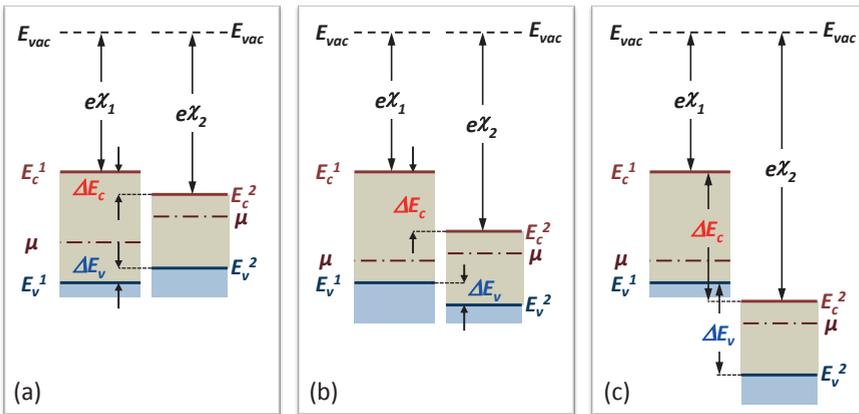


Abb. 10.32: Prinzipielle Typen von Banddiskontinuitäten in Halbleiter-Heterostrukturen: (a) normale (z. B. GaAs/(Al,Ga)As), (b) gestapelte und (c) gebrochene Banddiskontinuität (z. B. GaSb/InAs).

Die Banddiskontinuitäten wirken als Potenzialbarrieren für Elektronen und Löcher und zwar mit entgegengesetztem Vorzeichen. Dies führt an den Grenzflächen zu einer Diffusion von Elektronen und Löchern in den jeweiligen Halbleitertyp, in dem ihre Energie niedriger ist. Da die ionisierten Störstellen zurückbleiben, führt dies zu einer Raumladungszone und zu Bandverbiegungen, wie wir sie für den  $p$ - $n$ -Übergang bereits diskutiert haben. Wollen wir den Bandverlauf in einer Halbleiter-Heterostruktur berechnen, so müssen wir im Prinzip die beiden folgenden Probleme lösen:

- Wie sieht die Verschiebung  $\Delta E_c$  der Leitungsbandkante und die Verschiebung  $\Delta E_v$  der Valenzbandkante aus?
- Welche Bandverbiegung resultiert bei einem Kontakt der beiden Halbleiter?

In den meisten Fällen können wir diese beiden Probleme getrennt behandeln, da sie mit unterschiedlichen Energie- und Längenskalen verbunden sind. Die Anpassung der beiden Bandstrukturen geschieht auf einer atomaren Längenskala. Hier sind atomare Kräfte und Energieskalen maßgebend und die resultierenden elektrischen Felder sind in der Größenordnung atomarer Felder ( $\sim 10^8$  V/cm). Die Verbiegung der Bandstruktur erfolgt dagegen auf einer Längenskala von einigen 10 nm und wird durch die Konzentration der Dotieratome in den beteiligten Halbleitern bestimmt. Die aus den Raumladungszonen resultierenden elektrischen Felder sind nur in der Größenordnung von  $10^5$  V/cm.

Bei der Bestimmung der Banddiskontinuitäten spielen Oberflächen- bzw. Grenzflächeneffekte eine große Rolle. Wir können deshalb zur Bestimmung der Banddiskontinuitäten nicht einfach die Bulk-Werte der Elektronenaffinitäten heranziehen. Eine theoretische Beschreibung ist bei einer Berücksichtigung von Grenzflächenzuständen und Ladungstransfervorgängen schwierig.<sup>40</sup> In Tabelle 10.7 sind die experimentell bestimmten Diskontinuitäten der

<sup>40</sup> G. Margaritondo, P. Perfetti, *The Problem of Heterojunction Band Discontinuities*, in *Heterojunction Band Discontinuities, Physics and Device Applications*, herausgegeben von F. Capasso und G. Margaritondo, North-Holland, Amsterdam (1987).

Tabelle 10.7: Experimentell bestimmte Werte der Banddiskontinuitäten für einige Paare von Halbleitern (nach H. Morcoc, in *The Technology and Physics of Molecular Beam Epitaxy*, herausgegeben von E. H. C. Parker, Plenum Press, New York (1985)).

Heterostruktur	$\Delta E_v$ (eV)	Heterostruktur	$\Delta E_v$ (eV)
Si-Ge	0.28	InAs-Ge	0.33
AlAs-Ge	0.86	InAs-Si	0.15
AlAs-GaAs	0.34	InP-Ge	0.64
AlSb-GaSb	0.4	InP-Si	0.57
GaAs-Ge	0.49	InSb-Ge	0.0
GaAs-Si	0.05	InSb-Si	0.0
GaAs-InAs	0.17	CdS-Ge	1.75
GaP-Ge	0.80	CdS-Si	1.55
GaP-Si	0.80	CdSe-Ge	1.30
GaSb-Ge	0.20	CdSe-Si	1.20
GaSb-Si	0.05	CdTe-Ge	0.85
ZnSe-Ge	1.40	ZnTe-Ge	0.95
ZnSe-Si	1.35	ZnTe-Si	0.85

Valenzbandkante für einige Paare von Halbleitern aufgelistet. Mit Hilfe von  $\Delta E_v$  kann bei bekannter Energielücke der Halbleitermaterialien die Diskontinuität  $\Delta E_c$  der Leitungsbandkante einfach berechnet werden.

#### 10.4.1.1 Isotypische Heterostrukturen und Modulationsdotierung

Als erstes Beispiel einer Halbleiter-Heterostruktur betrachten wir *isotypische Halbleiter-Heterostrukturen*, die aus Halbleitermaterialien mit dem gleichen Dotierungstyp aufgebaut sind. Für praktische Anwendungen sind vor allem Systeme interessant, die aus einem Material II mit einer kleinen Energielücke und gleichzeitig sehr niedriger Dotierung und einem Material I mit großer Energielücke bei gleichzeitig hoher Dotierung bestehen. Der Verlauf des Leitungs- und Valenzbandes einer solchen Struktur ist in Abb. 10.33 gezeigt. Da Elektronen aus dem Material I mit der größeren Energielücke in das Material II mit der kleineren diffundieren, entsteht im Material I eine positive und im Material II eine negative Raumladungszone. Dies führt dazu, dass das Valenz- und Leitungsband im Material I nach oben und im Material II nach unten gebogen werden. Im Material II entsteht eine Randschicht mit einer sehr hohen Ladungsträgerkonzentration. Das chemische Potenzial liegt hier im Leitungsband.

Der interessante Aspekt der in Abb. 10.33 gezeigten Struktur ist, dass die hohe Dichte der Elektronen in der Randschicht von Material II von den ionisierten Donatoren in Material I räumlich getrennt ist. Dadurch wird die Streuung an Verunreinigungen, die ja bei niedrigen Temperaturen dominiert (vergleiche Abschnitt 10.1.4 und Abb. 10.12), stark reduziert. In einem homogen dotierten Halbleiter ist dagegen eine hohe Ladungsträgerdichte immer mit einer hohen Verunreinigungskonzentration und damit einer starken Verunreinigungsstreuung verbunden. Diese Art der Dotierung nennen wir *Modulationsdotierung*.<sup>41</sup> Die La-

<sup>41</sup> In einigen Materialsystemen liegt eine intrinsische Modulationsdotierung vor. Als Beispiel sollen hier die Kuprat-Supraleiter genannt werden (vergleiche Abschnitt 13.9.2). In diesen Materialien findet die elektrische Leitung in den zweidimensionalen  $\text{CuO}_2$ -Ebenen statt, wobei die Ladungsträger von zwischen den  $\text{CuO}_2$ -Ebenen liegenden Ladungsreservoirschichten in diese Ebenen dotiert werden.

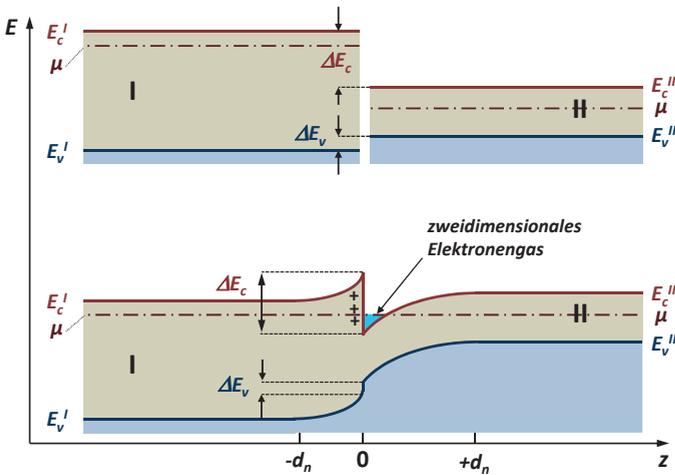


Abb. 10.33: Schematischer Verlauf der Leitungs- und Valenzbandkanten in einer isotypischen Halbleiter-Heterostruktur ohne Kontakt (oben) und im thermischen Gleichgewicht (unten). Im Material II bildet sich eine Randschicht mit einem zweidimensionalen Elektronengas aus.

dungsträger bewegen sich hier ungestört in einem Bereich, in dem sich keine bzw. nur sehr wenige streuende Dotieratome befinden. Dies können wir anschaulich mit einer Bahnlinie vergleichen, bei der ein Zug durch eine Oberleitung, die von neben der Bahnlinie stehenden Masten getragen wird, mit Strom versorgt wird. Der Zug kann sich ungestört auf den Schienen bewegen, ohne mit den Masten zusammenzustoßen.

Die Elektronenmobilität in homogen dotiertem GaAs hat bei einer Donorkonzentration von  $n_D \approx 10^{17} \text{ cm}^{-3}$  ein Maximum von etwa  $8 \times 10^3 \text{ cm}^2/\text{Vs}$  bei einer Temperatur von etwa 150 K. Oberhalb und unterhalb dieser Temperatur nimmt die Beweglichkeit aufgrund der Streuung an Phononen und Verunreinigungen ab (vergleiche Abb. 10.12). In einer modulationsdotierten Heterostruktur beobachtet man dagegen keine Abnahme der Beweglichkeit bei tiefen Temperaturen. Die Beweglichkeit steigt mit sinkender Temperatur weiter an und erreicht unterhalb von 10 K Werte oberhalb von  $10^6 \text{ cm}^2/\text{Vs}$ . Eine weitere Steigerung kann durch Einbringen einer undotierten  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ -Trennschicht zwischen die hochdotierte AlGaAs- und die schwach dotierte GaAs-Schicht erzielt werden, da dadurch Streuprozesse an Verunreinigungen in der unmittelbaren Nähe zur Grenzschicht vermieden werden.

Für eine  $n$ -Dotierung in AlGaAs im Bereich von  $10^{18} \text{ cm}^{-3}$  liegt die Dicke der Anreicherungsschicht im Bereich zwischen 50 und 100 Å. Die Elektronen sind dadurch in  $z$ -Richtung in einen schmalen, dreieckförmigen Potenzialwall eingesperrt und können sich nur noch in der  $xy$ -Ebene frei bewegen. Wir erhalten somit in der Randschicht ein zweidimensionales Elektronengas.

#### 10.4.1.2 Kompositionsübergitter

Ähnliche Effekte zu denjenigen in einer einfachen isotypischen Heterostruktur werden in so genannten *Kompositionsübergittern* gefunden. Diese erhält man durch heteroepitaktisches Wachstum von Übergittern, die aus zwei Halbleitermaterialien (z. B. AlGaAs und GaAs) bestehen. Wählen wir Materialien ähnlich zu denen in Abb. 10.33, so erhalten wir eine Serie von so genannten *Quantentrögen*, in denen sich die Leitungselektronen ansammeln (siehe Abb. 10.34). Wir erhalten wiederum eine Modulationsdotierung, da die hohe Dichte der

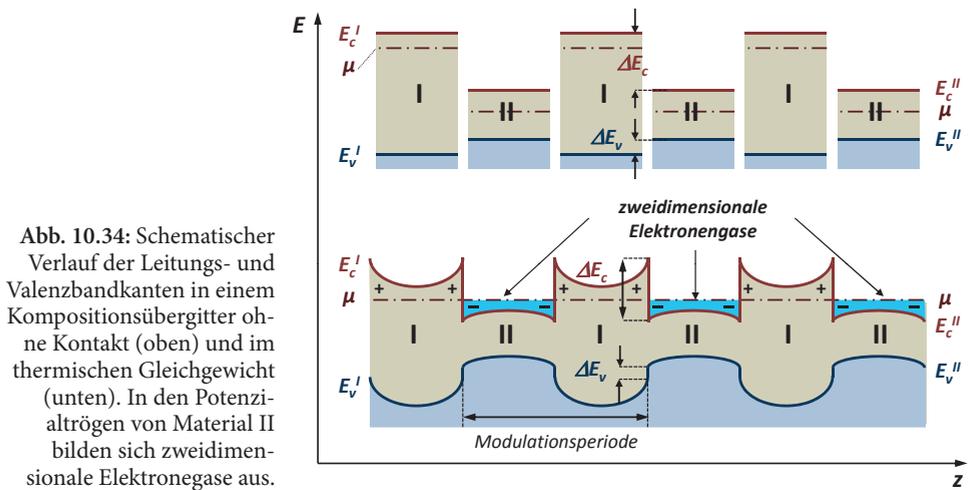


Abb. 10.34: Schematischer Verlauf der Leitungs- und Valenzbandkanten in einem Kompositionsübergitter ohne Kontakt (oben) und im thermischen Gleichgewicht (unten). In den Potentialrögen von Material II bilden sich zweidimensionale Elektronengase aus.

Leitungselektronen in Material II wiederum von der hohen Dichte der Dotieratome in Material I räumlich getrennt ist. Dadurch wird Streuung an Verunreinigungen stark unterdrückt und wir erhalten senkrecht zur Wachstumsrichtung der Übergitter ( $z$ -Richtung) bei tiefen Temperaturen eine sehr hohe Elektronenbeweglichkeit. Falls die Breite der einzelnen Schichten des Übergitters sehr klein wird (kleiner als etwa 10 nm), erhalten wir Quantisierungseffekte in den einzelnen Quantentrögen.

Das in Abb. 10.34 gezeigte Übergitter stellt eine Serie von aneinandergrenzenden Potentialwällen endlicher Höhe dar. Da die Wellenfunktionen bei einer endlichen Höhe der Potentialwälle mit einer charakteristischen Abklinglänge  $1/\kappa$  in den verbotenen Zonen abklingen [vergleiche (7.4.15)], erhalten wir bei einer kleinen Breite der Potentialwälle eine endliche Überlagerung der Wellenfunktionen der einzelnen Potentialmulden. Ähnlich wie die Überlagerung von Atomorbitalen zu Bändern führt (vergleiche hierzu Abschnitt 8.3), wenn wir die Atome einander annähern, spalten die Energieniveaus  $E_n$  der einzelnen Potentialtöpfe durch die Überlagerung in Bänder auf. Wir nennen die entstehenden Bänder *Minibänder*.

### 10.4.1.3 Dotierungsübergitter

Wachsen wir eine große Anzahl von zum Beispiel  $n$ - und  $p$ -dotiertem GaAs epitaktisch aufeinander auf, so erhalten wir so genannte *Dotierungsübergitter*. Diese bestehen also aus dem gleichen Halbleitermaterial, wobei aber die einzelnen Schichten einen unterschiedlichen Dotierungstyp haben. In diesen Übergittern erhalten wir dann eine große Zahl von  $p$ - $n$ -Übergängen. Da an den  $p$ - $n$ -Übergängen eine Verarmungszone mit einer quasi-intrinsischen Ladungsträgerdichte auftritt, werden diese Strukturen auch als *nipi*-Strukturen bezeichnet.

In den Dotierungsübergittern nähert sich das chemische Potenzial alternierend der Leitungsbandkante ( $n$ -Halbleiter) und der Valenzbandkante ( $p$ -Halbleiter) an. Da im thermischen Gleichgewicht das chemische Potenzial horizontal verläuft, werden die Leitungsband- bzw. Valenzbandkante periodisch nach unten und oben gebogen. Dies ist in Abb. 10.35

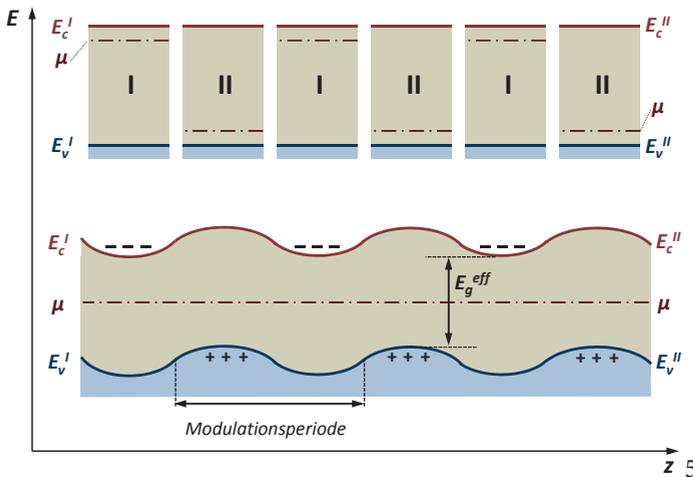


Abb. 10.35: Schematischer Verlauf der Leitungs- und Valenzbandkanten in einem Dotierungsübergitter ohne Kontakt (oben) und im thermischen Gleichgewicht (unten).

dargestellt. Wir erhalten durch die periodische Modulation der Leitungs- und Valenzbandkante eine periodische Modulation der Ladungsträgerdichte. Interessant ist dabei die Tatsache, dass die Elektronen und Löcher räumlich getrennt sind, was zu einer längeren Rekombinationslebensdauer führt. Trotz dieser Trennung liegt aber immer noch ein gewisser Überlapp der Wellenfunktionen im  $i$ -Gebiet vor. Die effektive Bandlücke  $E_g^{\text{eff}}$  kann über die Dotierung gezielt eingestellt werden, wodurch optische Übergänge abgestimmt werden können.

#### 10.4.1.4 MOSFET

Eine weitere Möglichkeit, ein zweidimensionales Elektronengas zu erhalten, ist der MOSFET (Metal-Oxide-Semiconductor Field Effect Transistor). Historisch gesehen ist das MOSFET-Prinzip wesentlich länger bekannt als der Bipolartransistor. Die ersten Patentanmeldungen stammen aus den Jahren 1926<sup>42</sup> und 1934.<sup>43</sup> Die ersten MOSFETs wurden allerdings erst 1960 gefertigt, als mit Silizium/Siliziumdioxid ein Materialsystem zur Verfügung stand, mit dem sich reproduzierbar eine gute Halbleiter-Isolator-Grenzfläche herstellen ließ.<sup>44,45</sup> Damit verbunden war die Abkehr vom Germanium als Basismaterial und steigende Anforderungen an die Fertigungsbedingungen. Der Feldeffekt-Transistor ist ein unipolarer Transistor. Unipolar daher, weil im Gegensatz zum bipolaren Transistor, je nach Typ, entweder nur Defektelektronen (Löcher) oder Elektronen am Stromtransport beteiligt sind.

<sup>42</sup> Julius Edgar Lilienfeld, U.S. Patent 1 745 175, *Method and apparatus for controlling electric currents* (1930); eingereicht 1926.

<sup>43</sup> Oskar Heil, British Patent 439 457, *Improvements in or relating to electrical amplifiers and other control arrangements and devices* (1935); zuerst eingereicht in Deutschland am 2. März 1934.

<sup>44</sup> D. Kahng, M. M. Atalla, *Silicon-silicon dioxide field induced surface devices*, IRE Solid State Res. Conf., Pittsburgh (1960).

<sup>45</sup> D. Kahng, *A Historical Perspective on the Development of MOS Transistors and Related Devices*, IEEE Trans. El. Dev. ED-23, 655 (1976).

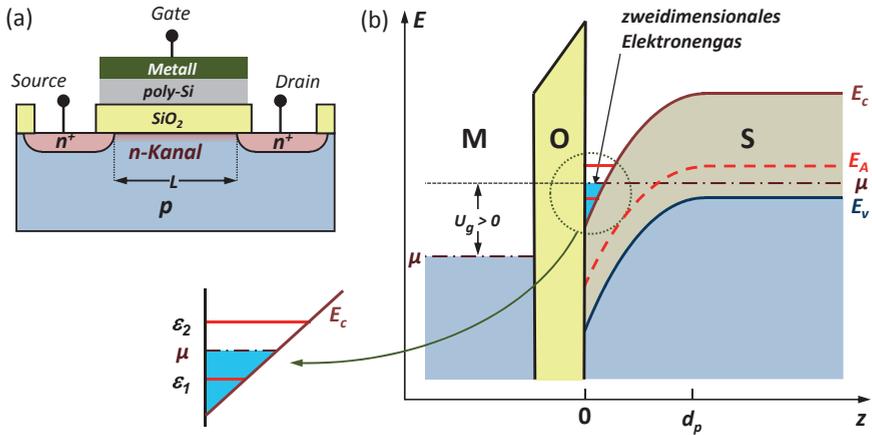


Abb. 10.36: (a) Schematischer Aufbau eines Si-MOSFET. (b) Verlauf der Leitungs- und Valenzbandkanten in einem MOSFET bei einer positiven Gate-Spannung  $U_g$ , was einer Absenkung der potenziellen Energie der Elektronen um  $(-e)U_g$  in der metallischen Gate-Elektrode entspricht. Aufgrund der angelegten Gate-Spannung entsteht an der Grenzfläche zwischen  $p$ -Halbleiter und Oxidschicht eine zweidimensionale Ladungsträgerschicht in einem dreieckförmigen Potenzialtopf. In dem gezeigten Fall ist nur das unterste Subband besetzt.

Der schematische Aufbau eines Si-MOSFETs ist in Abb. 10.36a gezeigt. Wird keine Spannung an die Gate-Elektrode angelegt, so wirkt das darunterliegende, leicht  $p$ -dotierte Silizium (Substrat) aufgrund der beiden  $p$ - $n$ -Kontakte im Source-Drain-Kanal sperrend. Der einzige Strom, der zwischen den beiden Kontakten Source und Drain fließen kann, ist der Sperrstrom des  $p$ - $n$ -Kontakts. Wenn über dem Gate eine positive Spannung angelegt wird, wandern Minoritätsladungsträger (bei  $p$ -Silizium Elektronen) im Substrat an die Grenzschicht und rekombinieren dort mit den Majoritätsladungsträgern (bei  $p$ -Silizium Löcher). Dies wirkt sich wie eine Verdrängung der Majoritätsladungsträger aus und wird Verarmung genannt. Ab einer bestimmten Spannung  $U_{th}$  (=Threshold-Spannung) ist die Verdrängung der Majoritätsladungsträger so groß, dass diese nicht mehr für die Rekombination zur Verfügung stehen. Es kommt zu einer Ansammlung der Minoritäten an der Grenzfläche zur Oxidschicht, die einen  $n$ -leitenden Kanal erzeugen, so dass jetzt Ladungsträger zwischen Source und Drain fließen können. Dieser Zustand wird *starke Inversion* genannt.

Der Verlauf der Leitungs- und Valenzbandkanten in einem MOSFET bei einer angelegten positiven Gate-Spannung  $U_g$  ist in Abb. 10.36b gezeigt. Ab einer bestimmten Spannung  $U_{th}$  sinkt die Leitungsbandkante an der Grenzfläche zur Oxidschicht unter das chemische Potenzial ab und wir erhalten eine sehr hohe Elektronenkonzentration unmittelbar an der Grenzfläche. Wir sprechen von einem *Inversionskanal*. Falls die Breite des Inversionskanals sehr klein ist (kleiner als etwa 10 nm) erhalten wir Quantisierungseffekte in dem dreieckförmigen Potenzialtopf. Die Elektronen befinden sich also in  $z$ -Richtung in einem Potenzialtopf (siehe Abb. 10.36) und können sich nur noch in der  $xy$ -Ebene frei bewegen. Die Eigenenergien sind dann durch  $E_n = E_{||} + \epsilon_n$  gegeben.<sup>46</sup> Auf jeden Energieeigenwert  $\epsilon_n$  baut sich ein Quasi-

<sup>46</sup> Die Ausdrücke für  $\epsilon_n$  weichen natürlich für einen dreieckförmigen Potenzialtopf von den durch (7.4.12) gegebenen Werten für einen rechteckförmigen Potenzialtopf ab.

Kontinuum, ein so genanntes Subband auf, das durch die kinetische Energie der Elektronen parallel zur Grenzfläche bedingt ist. Bei genügend tiefen Temperaturen lässt sich erreichen, dass nur noch das unterste Subband besetzt ist. In diesem Fall haben wir in der Randschicht ein zweidimensionales Elektronengas vorliegen. An einem derartigen Elektronengas hat **Klaus von Klitzing** im Jahr 1980 den Quanten-Hall-Effekt entdeckt.<sup>47,48</sup> Für die Elektronenzahl in der Randschicht gilt  $n_{2D} \propto U_g - U_{th}$ , wobei  $U_g$  die angelegte Gate-Spannung ist.

Die Kanallänge  $L$  von MOSFETs wurde durch die immer höhere Packungsdichte von Transistoren in integrierten Schaltungen seit 1960 von einigen  $10 \mu\text{m}$  bis auf heute weit unter  $100 \text{ nm}$  reduziert. Diese dem *Mooreschen Gesetz*<sup>49</sup> folgende exponentielle Abnahme der Kanallänge  $L$  mit der Zeit wird wohl aber in einigen Jahren an eine physikalische Grenze stoßen.

## 10.4.2 Vertiefungsthema: Halbleiter-Laser



Ein weites Anwendungsgebiet von Halbleiter-Heterostrukturen sind Halbleiterlaser, die heute z. B. zum Beschreiben und Auslesen von optischen Datenspeichern oder bei der optischen Datenübermittlung verwendet werden. Wir diskutieren im Folgenden den in Abb. 10.37 gezeigten *Double Heterostructure Injection Laser*, der von **Herbert Kroemer**<sup>50</sup> vorgeschlagen wurde. Die optisch aktive Zone (z. B. GaAs) ist hier zwischen zwei Halbleitermaterialien mit größerer Energielücke (Al,Ga)As und unterschiedlicher Dotierung eingebettet. Mit dem in dieser Struktur erzielten Verlauf der Leitungs- und Valenzbandkante wird erreicht, dass die Elektronen, die von rechts in die optische aktive GaAs-Schicht fließen, in dieser eingefangen werden. Das gleiche gilt für die Löcher, die von links in diese Schicht fließen. Das heißt, sowohl Elektronen als auch Löcher werden in der optisch aktiven Schicht festgehalten und es bildet sich dort ein entartetes Elektronen- bzw. Lochgas aus. Die injizierten Elektronen und Löcher können dann unter Emission von Photonen rekombinieren. Für Halbleiter-Laser

<sup>47</sup> K. von Klitzing, G. Dorda, M. Pepper, *New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance*, Phys. Rev. Lett. **45**, 494–497 (1980).

<sup>48</sup> K. von Klitzing, G. Dorda, M. Pepper, *High-magnetic-field transport in a dilute two-dimensional electron gas*, Phys. Rev. B **28**, 4886–4888 (1983).

<sup>49</sup> Als Mooresches Gesetz wird die empirische Beobachtung bezeichnet, dass bei der technologischen Entwicklungen von integrierten Schaltkreisen sich die Komplexität eines integrierten Schaltkreises etwa alle 2 Jahre verdoppelt. Diese Gesetzmäßigkeit wurde erstmals von Gordon E. Moore, einem der Mitbegründer der Firma Intel, im April 1965 formuliert und wird deshalb heute nach ihm benannt: „*The complexity for minimum component costs has increased at a rate of roughly a factor of two per year . . . Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer.*“

<sup>50</sup> **Herbert Kroemer** (Herbert Krömer), geboren am 25. August 1928 in Weimar. Er wurde im Jahr 2000 zusammen mit **Schores Iwanowitsch Alforow** mit der Hälfte des Nobelpreises für Physik für die Entwicklung von Halbleiterheterostrukturen für Hochgeschwindigkeits- und Optoelektronik ausgezeichnet, die andere Hälfte ging an **Jack Kilby** für die Entwicklung Integrierter Schaltkreise.

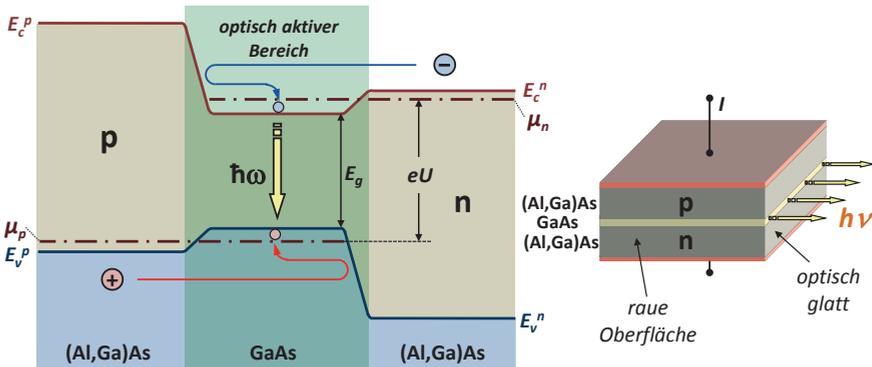


Abb. 10.37: Double Heterostructure Injection Laser: Elektronen fließen von rechts in die grün hinterlegte optisch aktive Zone, Löcher von links. Beide werden dort eingefangen und bilden entartete Elektronen- bzw. Lochgase.

werden ausschließlich Materialien mit direkter Bandlücke verwendet, da es bei Materialien mit indirekter Lücke auch nichtstrahlende Rekombination unter Emission von Phononen als Konkurrenzprozess gibt.

Üblicherweise sind die Rekombinationszeiten für die Überschussladungsträger weit größer als die Zeit für die Thermalisierung der Elektronen im Leitungsband bzw. der Löcher im Valenzband. Das heißt, Elektronen und Löcher erreichen jeweils getrennt sehr schnell einen Gleichgewichtszustand, aber nicht untereinander. Wir können dann sowohl den Elektronen als auch den Löchern getrennte chemische Potentiale  $\mu_n$  und  $\mu_p$  (Quasi-Potentiale) zuordnen. Diese Quasi-Potentiale verlaufen horizontal, da jeweils die Elektronen und Löcher im Gleichgewicht sind. Um eine Besetzungsinversion zu erhalten, muss

$$\mu_n - \mu_p > E_g \quad (10.4.2)$$

gelten.<sup>51</sup> Das heißt, um den Laserbetrieb zu erhalten, muss der Abstand der Quasi-Potentiale größer als die Energielücke sein. Diese Bedingung wird in Abb. 10.37 dadurch erreicht, dass eine Spannung  $eU > E_g$  angelegt wird, wobei  $E_g$  die Energielücke der optisch aktiven Zone ist.

Die in Abb. 10.37 gezeigte Diodenstruktur bildet gleichzeitig einen optischen Resonator, da die Reflektivität an der Halbleiterkristall-Luft Grenzfläche groß ist. Die Halbleiterkristalle werden planparallel poliert und die Strahlung wird parallel zur Ebene des Heterokontakts emittiert. Verwendet man GaAs als optisch aktive Schicht, so erhält man Laserstrahlung im nahen Infrarotbereich bei einer Wellenlänge von 838.3 nm oder 1.43 eV. Der Wirkungsgrad solcher GaAs-Laserdioden ist sehr groß. Das Verhältnis von Lichtleistung zu elektrischer Leistung beträgt fast 50 %. Die Wellenlänge dieser Laser kann über einen weiten Bereich durch Verwendung von Legierungen des Typs  $\text{Ga}_x\text{In}_{1-x}\text{P}_y\text{As}_{1-y}$  variiert werden und dadurch dem Minimum der optischen Absorption von Glasfasern angepasst werden. Die Kom-

<sup>51</sup> siehe z. B. *The Physics of Semiconductor Devices*, S. M. Sze, John Wiley & Sons, New York (1981).

bination von Double Heterostructure Injection Lasern mit Glasfasern bildet das Kernstück unserer heutigen optischen Datenkommunikation.<sup>52</sup>

## 10.5 Zweidimensionales Elektronengas: Quanten-Hall-Effekt

Wir haben uns in Abschnitt 7.5 bereits mit den Transporteigenschaften von niederdimensionalen Elektronengasen beschäftigt. Dabei haben wir allerdings zweidimensionale Elektronengase ausgelassen. Dies wollen wir jetzt nachholen.

### 10.5.1 Zweidimensionales Elektronengas im Magnetfeld

In Abschnitt 10.4.1 haben wir gelernt, dass wir mit Hilfe von Halbleiter-Heterostrukturen oder einem MOSFET zweidimensionale Elektronengase erzeugen können. Wir betrachten nun ein System aus *freien Elektronen*, in dem nur das unterste Subband besetzt sein soll, in einem in  $z$ -Richtung, d. h. senkrecht zum zweidimensionalen System angelegten Magnetfeld  $\mathbf{B} = (0, 0, B)$ . Durch die Wirkung des Magnetfeldes werden die ursprünglich gleichmäßig im zweidimensionalen  $k$ -Raum verteilten Zustände auf Landau-Kreise gezwungen (vergleiche hierzu Abschnitt 9.7 und Abb. 9.38). Die Eigenenergien ergeben sich gemäß (9.7.11) zu

$$E_n = \epsilon_1 + \left(n + \frac{1}{2}\right) \hbar\omega_c + g_s \mu_B s B. \quad (10.5.1)$$

Hierbei ist  $\epsilon_1$  die Energie des untersten Subbandes. Wir haben zusätzlich berücksichtigt, dass wir zwei mögliche Spin-Einstellungen  $s = \pm \frac{1}{2}$  vorliegen haben und die Zustände deshalb im Magnetfeld aufspalten ( $\mu_B$  ist das Bohrsche Magneton und  $g_s$  ist der Landé-Faktor für Elektronen). Den Spin-Freiheitsgrad werden wir bei der nachfolgenden Diskussion nicht berücksichtigen.

Die Quantisierung der Elektronenenergien in Landau-Niveaus mit Abstand  $\hbar\omega_c$  führt zu einer Aufspaltung des parabelförmigen Subbandes in diskrete Energieniveaus (siehe

<sup>52</sup> Heute wird eine Vielzahl von unterschiedlichen Halbleiterlasern mit unterschiedlichen Wellenlängen hergestellt, die wichtige Anwendungsgebiete haben. Neben der optischen Datenkommunikation sind vor allem die Anwendung in Blu-ray-Disc- und in HD-DVD-Laufwerken (405 nm, blauviolett, basierend auf dem Halbleitermaterial InGaN), der Einsatz als Pump Laser für Nd:YAG-Laser z. B. bei grünen Laser-Pointern oder bei Diodenlasern (808 nm, infrarot, basierend GaAlAs) sowie die Verwendung in kostengünstigen roten Laser-Pointern oder Barcode-Lesegeräten (670 nm, rot, basierend auf InGaAlP). Der erreichbare Wirkungsgrad von Laserdioden lag im Jahr 2012 zwischen 10% (grün, 530 bis 540 nm), 20% (blau, 440 nm) und 70% (rot und IR, ab 650 nm). Blaue Laserdioden erreichen mittlerweile einen Wirkungsgrad von 27% und Lebensdauer von 10 000 Stunden. Die Herstellung geeigneter InGaN-Halbleitermaterialien für leistungsfähige grüne Laser ist noch immer schwierig. Für Beleuchtungszwecke ist es deshalb preiswerter, mit kurzwelligem blauem Licht geeignete Leuchtstoffe im langwelligeren Bereich anzuzugeln

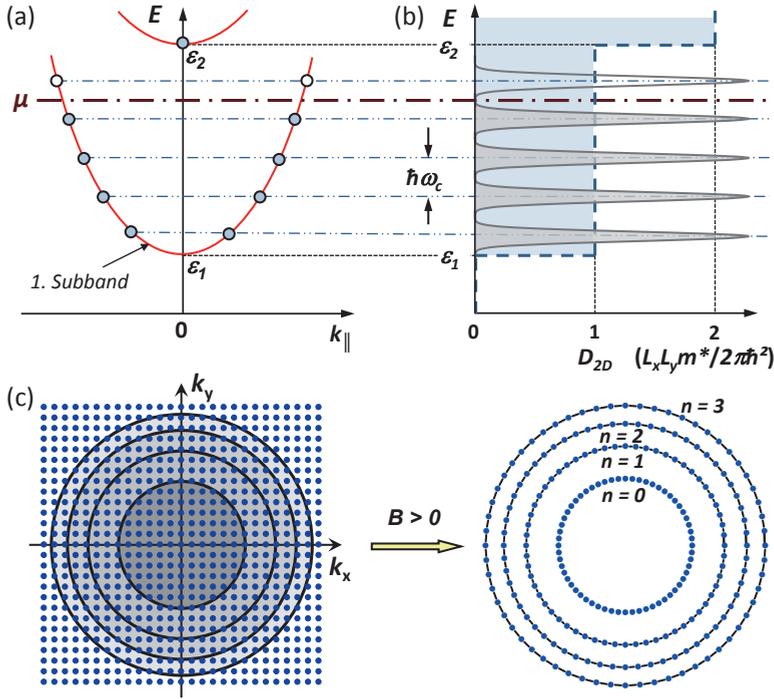


Abb. 10.38: Qualitative Darstellung der Zustandsquantisierung in einem zweidimensionalen Elektronengas. (a) Energieparabeln des 1. und 2. Subbandes. Das chemische Potenzial liegt zwischen 1. und 2. Subband. Durch Anlegen des Magnetfeldes erhalten wir eine zusätzliche Quantisierung in Landau-Niveaus (Punkte). (b) Zustandsdichte des zweidimensionalen Elektronengases ohne (gestrichelte Linie) und mit Magnetfeld (schraffierte Flächen). (c) Quantisierung der ohne Feld gleichförmig im zweidimensionalen  $k$ -Raum verteilten Zustände auf Landau-Kreise.

Abb. 10.38). Der in Abb. 7.21 für  $B = 0$  gezeigte stufenförmige Verlauf der Zustandsdichte spaltet in einem angelegten Magnetfeld in eine Reihe von  $\delta$ -funktionsartigen Peaks auf, die einen Abstand  $\hbar\omega_c$  voneinander haben. Wir können sagen, dass die Zustände in scharfe Landau-Niveaus „kondensieren“. Da keine Zustände verloren gehen, müssen die Landau-Niveaus den Entartungsgrad

$$p = \hbar\omega_c D_{2D} = \hbar\omega_c \frac{m}{2\pi\hbar^2} L_x L_y = \frac{eB}{h} L_x L_y \quad (10.5.2)$$

haben, den wir bereits in Abschnitt 9.7 hergeleitet haben [vergleiche (9.7.17)]. Hierbei ist  $D_{2D} = \frac{m}{2\pi\hbar^2} L_x L_y$  die zweidimensionale Zustandsdichte für eine Spin-Richtung [vergleiche (7.1.20) oder (7.4.13)]. Setzen wir Zahlenwerte ein, so erhalten wir

$$p = \frac{eB}{h} L_x L_y = 2.42 \times 10^{10} \text{ cm}^{-2} \times B[\text{T}] \times L_x L_y [\text{cm}^2]. \quad (10.5.3)$$

Wir können die Entartung  $p$  auch schreiben als

$$p = \frac{L_x L_y}{\ell_B^2} \quad \text{mit} \quad \ell_B^2 = \frac{2\pi\hbar}{eB}. \quad (10.5.4)$$

Die magnetische Landau-Länge  $\ell_B$  entspricht dabei dem Radius des klassischen Elektronen-orbits für das Niveau  $n = 1$ . Sie beträgt etwa 25 nm für  $B = 1$  T. Entsprechend können wir im reziproken Raum  $p$  schreiben als

$$p = \frac{k_B^2}{k_x k_y}, \quad (10.5.5)$$

wobei  $k_B^2 = (2\pi/\ell_B)^2 = S = 2\pi eB/\hbar$  die Fläche pro Zustand im  $\mathbf{k}$ -Raum ist und  $k_{x,y} = 2\pi/L_{x,y}$ .

Mit dem magnetischen Fluss  $\Phi = BL_x L_y$ , der die Probe durchsetzt, und dem Flussquant  $\tilde{\Phi}_0 = h/e$  können wir die Entartung auch als

$$p = \frac{eB}{h} L_x L_y = \frac{\Phi}{\tilde{\Phi}_0} = N_\Phi \quad (10.5.6)$$

schreiben. Der Entartungsgrad entspricht also der Zahl  $N_\Phi$  der magnetischen Flussquanten in der Probe. Liegt das chemische Potenzial zwischen zwei Landau-Niveaus, so ist die Gesamtzahl der Elektronen  $N_e = n \cdot p = n \cdot N_\Phi$  und damit  $N_e/N_\Phi = n =$  ganzzahlig. Allgemein bezeichnen wir das Verhältnis  $\nu = N_e/N_\Phi$  als den **Füllfaktor**. Liegt also das chemische Potenzial zwischen zwei Landau-Niveaus, so ist  $\nu = n$ , d. h. der Füllfaktor ist ganzzahlig und entspricht dem obersten besetzten Landau-Niveau.

Die Gesamtenergie des Elektronensystems als Funktion des angelegten Feldes haben wir bereits ausführlich im Zusammenhang mit dem de Haas-van Alphen-Effekt in Abschnitt 9.8.1 diskutiert. Wir haben gesehen, dass die freie Energie als Funktion des Magnetfeldes oszilliert und zwar periodisch in  $1/B$ . Dies führt zu charakteristischen Oszillation der Magnetisierung (de Haas-van Alphen-Effekt) oder des elektrischen Widerstands (Shubnikov-de Haas-Effekt).

Gehen wir von freien Elektronen zu **Kristallelektronen** über, so müssen wir Folgendes beachten:

1. Wir müssen in den Ausdrücken für die Zustandsdichte, die Energieeigenwerte oder die Zyklotronfrequenz die freie Elektronenmasse  $m$  durch die effektive Bandmasse  $m^*$  ersetzen.
2. Durch das Magnetfeld wird nicht nur die Spin-Entartung aufgehoben, sondern in Si auch noch die durch die Struktur des Leitungsbandes bedingte zweifache Valley-Entartung aufgehoben. Das bedeutet, dass jedes Landau-Niveau insgesamt in vier Niveaus aufspaltet.
3. Durch die Streuung der Ladungsträger an Gitterdefekten und Phononen werden die Landau-Niveaus verbreitert. Aus der Unschärferelation ergibt sich für eine mittlere Streuzeit  $\tau$  eine Energieverschmierung  $\Delta E = \hbar/\tau$ . Effekte, die mit der Aufspaltung in die Landau-Niveaus verbunden sind, können nur dann experimentell gut beobachtet werden, wenn  $\Delta E \ll \hbar\omega_c$ , also

$$\omega_c \tau \gg 1 \quad \text{oder} \quad \mu B \gg 1. \quad (10.5.7)$$

Wir benötigen also hohe Felder und lange Streuzeiten  $\tau$  bzw. hohe Beweglichkeiten  $\mu$ . Letzteres erreichen wir durch tiefe Temperaturen und saubere Proben.

## 10.5.2 Transporteigenschaften des zweidimensionalen Elektronengases

Elektrische Transportexperimente an zweidimensionalen Elektronengasen werden üblicherweise mit Proben durchgeführt, welche die in Abb. 10.39 gezeigte „Hall-Bar-Geometrie“ besitzen. Der Leitfähigkeitstensor für das zweidimensionale System lautet (vergleiche (7.3.43) in Abschnitt 7.3.4)

$$\begin{pmatrix} J_x \\ J_y \end{pmatrix} = \frac{\sigma_0}{1 + \omega_c^2 \tau^2} \begin{pmatrix} 1 & +\omega_c \tau \\ -\omega_c \tau & 1 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad (10.5.8)$$

mit

$$\sigma_0 = \frac{n_{2D} e^2 \tau}{m^*}. \quad (10.5.9)$$

Hierbei ist  $n_{2D}$  die Zahl der Elektronen pro Flächeneinheit.

Da der Leitfähigkeitstensor und der Widerstandstensor invers zueinander sind, können wir durch Invertieren der obigen Matrix folgende Beziehungen ableiten:

$$\begin{aligned} \sigma_{xx} &= \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2} & \sigma_{xy} &= \frac{-\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2} \\ \sigma_{yy} &= \sigma_{xx} & \sigma_{yx} &= -\sigma_{xy} \end{aligned} \quad (10.5.10)$$

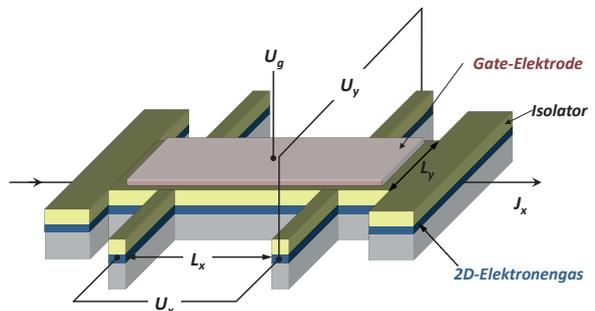
$$\begin{aligned} \rho_{xx} &= \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} & \rho_{xy} &= \frac{-\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \\ \rho_{yy} &= \rho_{xx} & \rho_{yx} &= -\rho_{xy} \end{aligned} \quad (10.5.11)$$

Im Experiment wird ein Strom in Längsrichtung ( $x$ -Richtung) vorgegeben und die Längs- und Querspannung gemessen. Es gilt

$$U_x = \rho_{xx} L_x \cdot J_x \Rightarrow \rho_{xx} = \frac{U_x}{J_x L_x} \quad (10.5.12)$$

$$U_y = \rho_{xy} L_y \cdot J_x \Rightarrow \rho_{xy} = \frac{U_y}{J_x L_y}. \quad (10.5.13)$$

Abb. 10.39: Schematische Darstellung der für die Messung der elektrischen Transporteigenschaften von zweidimensionalen Elektronengasen verwendeten „Hall-Bar“-Geometrie. Die Elektronendichte  $n_{2D}$  im Elektronengas kann durch Variation der Gate-Spannung  $U_g$  variiert werden.



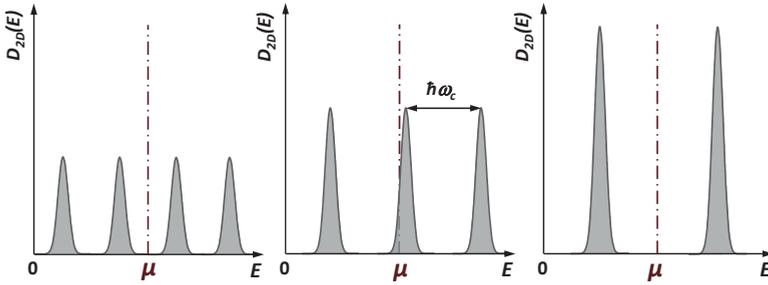


Abb. 10.40: Besetzung der Landau-Niveaus für unterschiedliche Magnetfelder. Das Magnetfeld nimmt von links nach rechts zu. Mit zunehmendem  $B$  nimmt die Entartung der Landau-Niveaus zu, was durch zunehmende Flächen der Zustandsdichtepeaks berücksichtigt ist.

Messen wir die Querspannung  $U_y$  in einer Konfiguration, in der  $J_y = 0$ , so müssen sich die Kraft  $eE_y = eU_y/L_y$  durch das Hall-Feld und die Lorentz-Kraft  $e\mathbf{v}B$  gerade kompensieren.<sup>53</sup> Aus  $e\mathbf{v}_x B - eU_y/L_y = 0$  folgt  $U_y = v_x B \cdot L_y$ . Setzen wir dies in (10.5.13) ein und benutzen  $J_x = n_{2D}e v_x$ , so erhalten wir

$$\rho_{xy} = \frac{B}{n_{2D}e} = R_H \cdot B. \tag{10.5.14}$$

Hierbei ist  $R_H$  die Hall-Konstante. Für  $\omega_c \tau \rightarrow \infty$  gilt allgemein:  $\sigma_{xx}, \sigma_{yy} \rightarrow 0$  und  $\sigma_{xy} \rightarrow n_{2D}e/B$ . Die Tatsache, dass  $\sigma_{xy} \rightarrow n_{2D}e/B$  ist eine allgemeine Eigenschaft von zweidimensionalen Elektronengasen in gekreuzten elektrischen und magnetischen Feldern.<sup>54</sup>

Abb. 10.40 zeigt, dass die Zustandsdichte beim chemische Potenzial  $\mu$  als Funktion des angelegten Magnetfeldes variiert. Sie wird null, wenn der Füllfaktor  $\nu = n$ , also ganzzahlig ist und

<sup>53</sup> Dies gilt für Löcher mit Ladung  $+e$ . Für Elektronen ist die Kraft durch das Hall-Feld  $-eU_H/L_y$ . Die Lorentz-Kraft behält ihr Vorzeichen, da wir  $e$  durch  $-e$ , aber auch  $v_x$  durch  $-v_x$  ersetzen müssen, weil sich die Elektronen bei gleicher Stromrichtung in die entgegengesetzte Richtung bewegen.

<sup>54</sup> Aus (10.5.10) und (10.5.11) folgt, dass für  $\sigma_{xx} = 0$  auch  $\rho_{xx} = 0$ . Das heißt,  $J_x$  kann ohne  $E_x$  fließen, so dass die effektive Leitfähigkeit gegen unendlich geht. Paradoxiere Weise erhalten wir diesen Grenzfall gerade dann, wenn  $\sigma_{xx}, \sigma_{yy} \rightarrow 0$ . Wir können aber leicht zeigen, dass die effektive Leitfähigkeit immer noch endlich ist. Es gilt nämlich

$$\begin{aligned} J_x &= \sigma_{xx}E_x + \sigma_{xy}E_y \\ J_y &= \sigma_{yy}E_y + \sigma_{yx}E_x \end{aligned}$$

und aus geometrischen Gründen gilt  $J_y = 0$ , so dass

$$E_y = \frac{\sigma_{xy}}{\sigma_{yy}} E_x.$$

Hierbei haben wir  $\sigma_{yx} = -\sigma_{xy}$  verwendet. Setzen wir dies in den Ausdruck für  $J_x$  ein, so erhalten wir

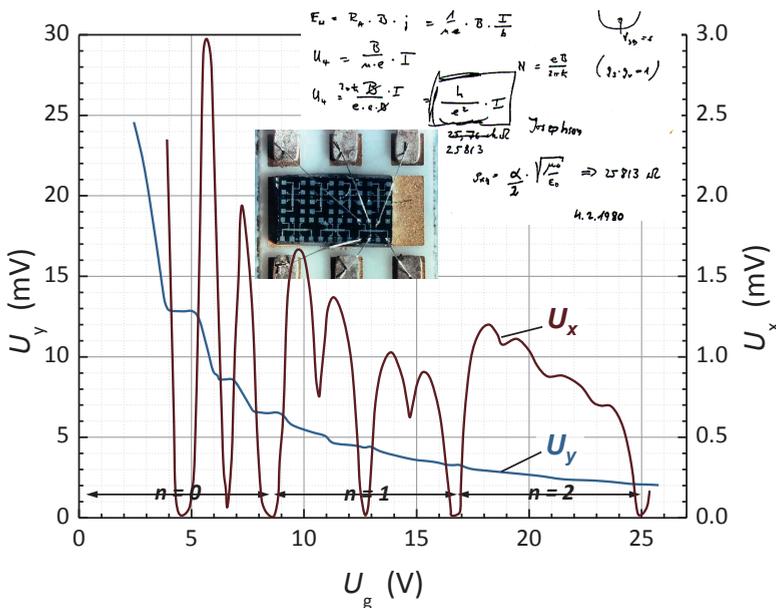
$$J_x = \left( \sigma_{xx} + \frac{\sigma_{xy}^2}{\sigma_{yy}} \right) E_x = \sigma_{\text{eff}} E_x.$$

Wir sehen, dass  $\sigma_{\text{eff}} \rightarrow \infty$  für  $\sigma_{xx}, \sigma_{yy} \rightarrow 0$ .

hat ein Maximum für  $\nu \approx n + \frac{1}{2}$ . Diese Variation kann in vielen physikalischen Eigenschaften beobachtet werden. So zeigt der elektrische Widerstand starke Oszillationen als Funktion des angelegten Magnetfeldes, die *Shubnikov-de Haas Oszillationen* genannt werden. Die elektrische Leitfähigkeit  $\sigma_{xx}$  wird null für  $\nu = n$ , wenn die Zustandsdichte bei  $\mu$  verschwindet, und maximal für  $\nu \approx n + \frac{1}{2}$ , wenn die Zustandsdichte bei  $\mu$  maximal ist. Nach (10.5.11) wird für  $\sigma_{xx} = 0$  auch  $\rho_{xx} = 0$ .

### 10.5.3 Ganzzahliger Quanten-Hall-Effekt

Im Jahr 1980 haben von Klitzing, Dorda und Pepper die elektrische Leitfähigkeit und den Hall-Effekt in einem zweidimensionalen Elektronengas untersucht, das sie mit Hilfe eines Si-MOSFET (siehe Abb. 10.36) realisiert haben.<sup>55</sup> Die dabei verwendete Probengeometrie entsprach der in Abb. 10.39 gezeigten Hall-Bar-Geometrie. In ihren Experimenten, die 1980 am Max-Planck-Hochfeldmagnetlabor in Grenoble durchgeführt wurden, wurde der Strom  $J_x$  vorgegeben und die Längs- und Querspannung als Funktion der Gate-Spannung, also der Ladungsträgerdichte, bei konstantem Magnetfeld gemessen. Das Ergebnis ist in



**Abb. 10.41:** Längsspannung  $U_x$  und Hall-Spannung  $U_y$  in einem zweidimensionalen Elektronengas ( $L_x = 130 \mu\text{m}$ ,  $L_y = 50 \mu\text{m}$ ) als Funktion der Gate-Spannung  $U_g$  bei  $T = 1.5 \text{ K}$  und  $B = 18 \text{ T}$  (nach K. von Klitzing, G. Dorda, M. Pepper, Phys. Rev. Lett. 45, 494 (1980)). Oben sind handschriftliche Notizen von Klaus von Klitzing vom 04.02.1980 und ein Bild der Originalprobe gezeigt (Quelle: PTB Braunschweig). Die Landau-Niveaus sind aufgrund der Spin- und Valley-Entartung vierfach aufgespalten.

<sup>55</sup> K. von Klitzing, G. Dorda, M. Pepper, *New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance*, Phys. Rev. Lett. 45, 494–497 (1980).

Abb. 10.41 gezeigt. Aus den gemessenen Spannungswerten wurde nach (10.5.12) und (10.5.13) der spezifische Längs- und Querwiderstand  $\rho_{xx}$  und  $\rho_{xy}$  bzw. die spezifischen Leitfähigkeiten  $\sigma_{xx}$  und  $\sigma_{xy}$  bestimmt. Sie beobachteten, dass  $\sigma_{xx}$  periodisch als Funktion der Gate-Spannung verschwindet. Das Verschwinden von  $\sigma_{xx}$  kann dabei mit dem Auffüllen der Landau-Niveaus mit zunehmender Gate-Spannung erklärt werden. Wenn das chemische Potenzial zwischen zwei Landau-Niveaus zu liegen kommt, kann die Probe bei tiefen Temperaturen ( $k_B T \ll \hbar \omega_c$ ) als Isolator betrachtet werden. In diesem Fall erwarten wir, dass die Diagonalelemente  $\sigma_{xx}$  und  $\sigma_{yy}$  des Leitfähigkeitstensors verschwinden. Nach unserer obigen Betrachtung erwarten wir auch ein Verschwinden von  $\rho_{xx}$ .

Das unerwartete Ergebnis war jedoch, dass der Querwiderstand Plateaus bei genau denjenigen Gate-Spannungen zeigte, an denen  $\rho_{xx}$  verschwand. Die Werte des Querwiderstands betragen  $25\,813 \, \Omega/\nu$  mit den Füllfaktoren  $\nu = 1, 2, 3, \dots$ . Bereits in den ursprünglichen handschriftlichen Aufzeichnungen weist von Klitzing darauf hin, dass dieser Quantenwiderstand durch  $R_K = h/e^2$  ausgedrückt werden kann, also durch Naturkonstanten. Heute kann der Wert dieses Quantenwiderstands mit der sehr hohen relativen Genauigkeit etwa  $10^{-9}$  gemessen werden. Er wird deshalb von den Eichanstalten als Widerstandsnormale verwendet. Der heutige Wert der *von Klitzing-Konstante* lautet

$$R_K = \frac{h}{e^2} = 25\,812.807\,4434(84)\,\Omega. \quad (10.5.15)$$

Außerdem kann die Messung von  $R_K$  zur Präzisionsbestimmung der *Sommerfeldschen Feinstruktur-Konstanten*

$$\alpha = \frac{\mu_0 c}{2} \frac{e^2}{h} = \frac{1}{2} \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{e^2}{h} = 7.297\,352\,5698(24) \times 10^{-3} \quad (10.5.16)$$

herangezogen werden.

Wir wollen zunächst eine stark vereinfachte Erklärung für das Auftreten der Oszillationen im Längswiderstand und der Plateaus im Querwiderstand geben. Wir setzen  $k_B T \ll \hbar \omega_c$  voraus und nehmen an, dass wir die Gate-Spannung auf einen Wert eingestellt haben, dass die Dichte der Elektronen gerade  $N_e = n \cdot p$  ist. Das bedeutet, dass das chemische Potenzial zwischen dem  $n$ -ten und  $(n+1)$ -ten Landau-Niveau liegen muss. Für die Elektronendichte gilt dann

$$n_{2D} = \frac{N_e}{L_x L_y} = \frac{n \cdot p}{L_x L_y} = n \frac{eB}{h}. \quad (10.5.17)$$

Formal entspricht die Situation derjenigen eines Isolators, bei dem das chemische Potenzial zwischen einem vollständig gefüllten und einem vollkommen leeren Band liegt. Da in einem vollständig gefüllten Band die Transportstromdichte verschwindet (vergleiche hierzu die Diskussion in Abschnitt 9.2), und auch wegen  $k_B T \ll \hbar \omega_c$  Anregungen ins nächst höhere Landau-Niveau sehr unwahrscheinlich sind, erwarten deshalb

$$\sigma_{xx} = \sigma_{yy} \rightarrow 0, \quad \rho_{xx} \rightarrow 0, \quad (10.5.18)$$

$$\rho_{xy} = \frac{B}{n_{2D} e} = \frac{h}{e^2} \frac{1}{n}. \quad (10.5.19)$$

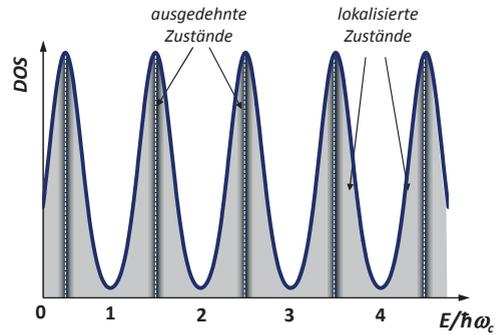


Abb. 10.42: Zustandsdichte eines zweidimensionalen Elektronengases in einem starken Magnetfeld bei Vorhandensein von Unordnung. Die im Idealfall  $\delta$ -funktionsartigen Zustandsdichtepeaks sind stark verbreitert und an den Flanken der Kurven treten lokalisierte Zustände auf. Nur die Zustände um das Maximum der Kurve sind delokalisiert.

Das heißt, wir haben das beobachtete Ergebnis erklärt, aber eben nur für eine bestimmte Gate-Spannung. Würden wir die Gate-Spannung nur infinitesimal ändern, so hätten wir ein teilweise gefülltes Landau-Niveau vorliegen. Die elektrische Leitfähigkeit  $\sigma_{xx}$  wäre dann endlich und die Beziehungen (10.5.18) und (10.5.19) würden nicht mehr gelten. Wir würden damit keine Plateaus im Querwiderstand erhalten, die Länge der Plateaus als Funktion der Gate-Spannung würde vielmehr gegen null gehen.

Auch eine detailliertere Betrachtung zeigt, dass für eine perfekte Probe die Breite der Quanten-Hall-Plateaus in der Tat null sein sollte. Allerdings besitzen reale Proben immer Defekte und diese resultieren in einer Verbreiterung der Landau-Niveaus. Wie Abb. 10.42 zeigt, liegen in der Zustandsdichte keine  $\delta$ -Funktionen bei  $(n + \frac{1}{2})\hbar\omega_c$ , sondern stark verbreiterte Glockenkurven vor. In den Flanken dieser Kurven sind die Zustände an Störstellen lokalisiert und tragen deshalb nicht zum Transport bei. Nur im zentralen Bereich liegen delokalisierte Zustände vor. Das Vorliegen lokalisierter und delocalisierter Zustände ist für die Erklärung der endlichen Breite der Hall-Plateaus von essentieller Bedeutung.

In Abb. 10.43 ist das Gesamtpotenzial eines zweidimensionalen Elektronengases mit Unordnung (z. B. durch Verunreinigungen, Kristalldefekte) gezeigt. Das Unordnungspotenzial  $\phi_{\text{dis}}$  führt zu räumlichen Fluktuationen des Landau-Niveaus. Für eine räumlich eingeschränkte Probe müssen wir noch das Einschlusspotenzial hinzufügen, das im Probeninneren verschwindet und am Probenrand steil zum Vakuumniveau ansteigt. Da die Landau-Niveaus am Probenrand durch das Einschlusspotenzial steil nach oben gebogen werden, schneidet das chemische Potenzial die Landau-Niveaus am Probenrand. Um die Auswirkung von Unordnung auf die Bewegung von Elektronen im Magnetfeld qualitativ zu diskutieren, nehmen wir an, dass  $|\nabla\phi_{\text{dis}}| \cdot \ell_B \ll \hbar\omega_c$  an, so können wir die Bewegung der Elektronen in zwei Anteile zerlegen. Erstens eine schnelle Zyklotronbewegung und zweitens eine langsame Driftbewegung des Schwerpunkts der Zyklotronbahn entlang der Äquipotenziallinien des Gesamtpotenzials  $\phi = \phi_{\text{dis}} + \phi_{\text{el}}$ , wobei  $\phi_{\text{el}}$  das angelegte elektrische Potenzial ist. Dies ist schematisch in Abb. 10.43 gezeigt. Die Bewegung der Elektronen erfolgt in so genannten *Skipping Orbits* entlang der Äquipotenziallinien von  $\phi$ . Die Bewegungsrichtung der Driftbewegung wird durch die Lorentz-Kraft bestimmt, die proportional zu  $\nabla\phi \times \mathbf{B}$  ist. Am Probenrand stoßen die Elektronen an das Einschlusspotenzial und werden elastisch reflektiert. Dabei unterdrückt das starke Magnetfeld die Rückstreuung. Die Elektronen können sich auf ihren Skipping Orbits nur in eine Richtung bewegen, die auf

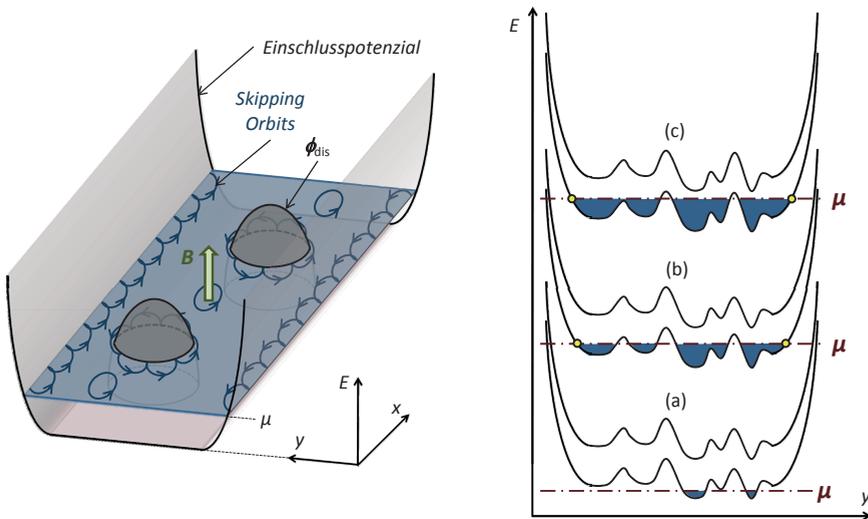


Abb. 10.43: Links: Schematische Darstellung des Potenzialverlaufs in einem zweidimensionalen Elektronengas. Aus Gründen der Übersichtlichkeit ist nur ein Landau-Niveau gezeigt. Das endliche Unordnungspotenzial  $\phi_{\text{dis}}$  bewirkt räumliche Fluktuationen des Gesamtpotenzials, die durch zwei Potentialhügel angedeutet sind. Am Rand der Probe steigt das Gesamtpotenzial durch das Einschlusspotential stark an. Rechts sind Schnitte des Gesamtpotenzials für zwei benachbarte Landau-Niveaus  $n$  und  $n + 1$  senkrecht zur Probe ( $y$ -Richtung) gezeigt. Von (a) nach (c) wird die Ladungsträgerdichte erhöht und damit das chemische Potenzial  $\mu$  nach oben geschoben. Dies entspricht den in Abb. 10.44(a) bis (c) gezeigten Situationen. Die gelben Punkte markieren die Randkanäle.

den gegenüberliegenden Probenseiten entgegengesetzt gerichtet ist. Wir können uns leicht überlegen, dass selbst Streuprozesse an Defekten diese bevorzugte Bewegungsrichtung nicht ändern können. Am Probenrand bilden sich also durch die Skipping Orbits eindimensionale **Randkanäle** aus. Jedes Elektron, das in den Randkanal eintritt, wird mit der Transmissionswahrscheinlichkeit  $t = 1$  durch den Kanal transmittiert. Aufgrund der in Abschnitt 7.5.1 geführten Diskussion können wir diesem Transportkanal den Leitwert  $e^2/h$  zuordnen. Wir werden auf die Randkanäle weiter unten nochmals zurückkommen. Wir wollen noch darauf hinweisen, dass die Elektronen, die um Erhebungen des Unordnungspotenzials über das chemische Potenzial umlaufen, räumlich lokalisiert sind (siehe hierzu Abb. 10.43, links).

Benutzen wir die Bohr-Sommerfeld-Quantisierung, so können wir das gerade vorgestellte klassische in ein quantenmechanisches Bild überführen. Wie in Abschnitt 9.7 diskutiert wurde, ist eine Konsequenz der Bohr-Sommerfeld-Quantisierung die Quantisierung der von einem Zyklotronorbit im  $\mathbf{k}$ -Raum umschlossenen Fläche in ganzzahligen Vielfachen der Fläche  $S = (2\pi e/\hbar)B$ . Im Ortsraum umschließen die Bahnen ganzzahlige Vielfache eines Flussquants  $\Phi_0 = h/e$ . Innerhalb der gemachten Näherung werden diese Quantisierungsbedingungen nicht tangiert.

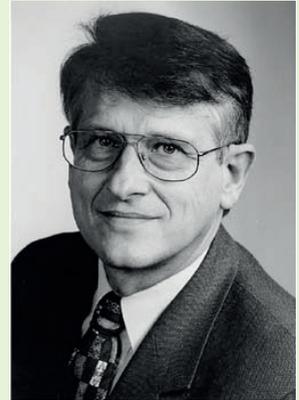
Wir wollen nun ein Quanten-Hallsystem mit Unordnung betrachten und diskutieren, was passiert, wenn wir bei  $T = 0$  die Gate-Spannung  $U_g$  bei konstantem Magnetfeld variieren. Der Abstand der Landau-Niveaus bleibt dabei wegen  $B = \text{const}$  gleich und wir schieben das

### Klaus von Klitzing (geb. 1943), Nobelpreis für Physik 1985

Klaus von Klitzing wurde am 28. Juni 1943 in Schroda (Posen) geboren.

Nach dem Abitur in Quakenbrück im Februar 1962 studierte er Physik an der Technischen Universität Braunschweig. Das Studium schloss er mit dem Diplom (Diplomarbeit bei F.R. Kessler) im März 1969 ab. Bis November 1980 war er dann bei Prof. Dr. Gottfried Landwehr an der Julius-Maximilians-Universität Würzburg tätig. Dort schrieb er 1972 seine Doktorarbeit zum Thema „Galvanomagnetische Eigenschaften von Tellur in starken Magnetfeldern“. Im Jahr 1978 folgte die Habilitation. Er war zu Forschungsarbeiten 1975 bis 1978 am Clarendon Laboratory in Oxford und als Heisenberg-Stipendiat von 1979 bis 1980 am Hochfeld-Magnetlabor in Grenoble tätig, wo er die entscheidende Entdeckung für den Quanten-Halleffekt machte. Nach seiner Entdeckung des Quanten-Hall-Effektes kam er 1980 als C3-Professor an das Physik-Department der Technischen Universität München und führte am Lehrstuhl von Prof. Frederick Koch detaillierte Untersuchungen zum weiteren Verständnis des Effektes durch. Seit 1985 ist er Direktor am Max-Planck-Institut für Festkörperforschung in Stuttgart.

Klaus von Klitzing erhielt 1985 den Nobelpreis für Physik für seine Arbeiten zum quantisierten Hall-Effekt, den er am 5. Februar 1980 im Grenobler Hochfeld-Magnetlabor entdeckte.

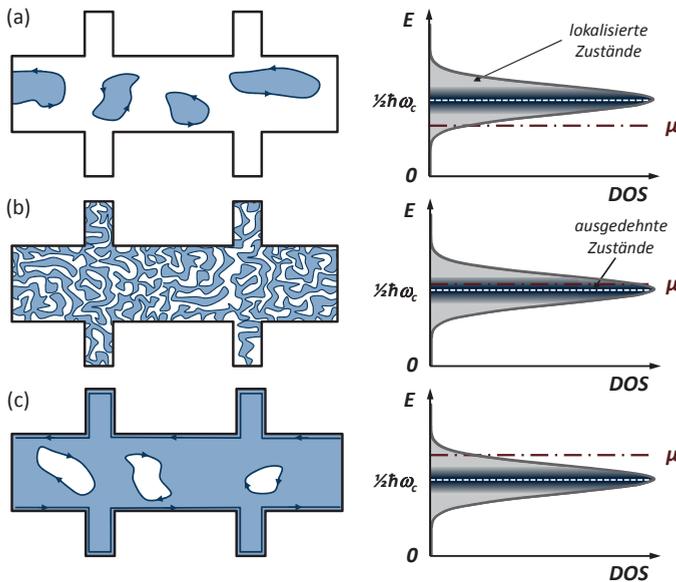


Mit freundlicher Genehmigung von Klaus von Klitzing.

chemische Potenzial  $\mu$  durch die Gate-induzierte Erhöhung der Ladungsträgerdichte nach oben.<sup>56</sup>

Wir starten mit einer Situation, in der das chemische Potenzial  $\mu$  irgendwo zwischen zwei Landau-Niveaus liegt [siehe hierzu Abb. 10.43(a) und 10.44(a)]. Aufgrund der Unordnung kann  $\mu$  jetzt irgendwo zwischen den beiden Landau-Niveaus liegen und nicht nur genau in der Mitte wie im Fall ohne Unordnung. Für  $T = 0$  sind Konturlinien der Schwerpunktsbewegung mit  $E < \mu$  mit Elektronen besetzt und diejenigen mit  $E > \mu$  sind leer. Das bedeutet, dass in der Umgebung eines Minimums des Gesamtpotenzials  $\phi$  alle Äquipotenzialkonturlinien mit  $E < \mu$  besetzt sind. Diese Gebiete bezeichnen wir als *Quantum-Hall-Droplets*. Die Elektronen, die die äußerste Bahn in diesen Droplets einnehmen, nennen wir *Randelektronen*. Da diese Elektronen sich bei  $\mu$  befinden, sind es gerade diese Elektronen, die zum Transport beitragen. In Abb. 10.44 haben wir die Droplets als schattierte Flächen eingezeichnet und die Randelektronen sind durch die durchgezogenen Ränder der Droplets gekennzeichnet. Wir haben oben bereits diskutiert, dass wir die Bewegung der Randelektronen mit so genannten

<sup>56</sup> Im vielen Experimenten wird auch bei konstanter Gate-Spannung, d. h. konstanter Ladungsträgerdichte, das Magnetfeld variiert. Physikalisch kommt es immer auf eine Variation des Füllfaktors  $\nu = N_e/N_\phi$  an, wobei  $N_e$  über die Gate-Spannung und  $N_\phi$  über das Magnetfeld variiert wird. Eine Variation der Gate-Spannung oder des Magnetfeldes führen also prinzipiell zum gleichen Ergebnis.



**Abb. 10.44:** Schematische Darstellung von Quantum-Hall-Droplets (links) und der Zustandsdichte (rechts) für unterschiedliche Lagen des chemischen Potentials  $\mu$ . Die Droplets sind als gefüllte Fläche dargestellt, die Elektronenbahnen als durchgezogene Linien mit Pfeilen. (a)  $\mu$  liegt unterhalb des 1. Landau-Niveaus im Bereich der lokalisierten Zustände (hellgrau). (b)  $\mu$  liegt im Bereich der delokalisierten Zustände (dunkelgrau). (c)  $\mu$  liegt oberhalb des 1. Landau-Niveaus im Bereich der lokalisierten Zustände (hellgrau).

Skipping Orbits beschreiben können. Elektronen weit weg vom Rand des Droplets können Kreisbahnen ausführen. Am Rand wird aber die Kreisbewegung durch die Streuung mit dem Rand unterbrochen und wir erhalten eine springende Bewegung entlang des Randes.

Abb. 10.44(a) zeigt, dass wir bei der gewählten Ladungsträgerdichte nur lokalisierte Zustände haben, die in den Mulden des Unordnungspotenzial gebunden sind. Erhöhen wir jetzt die Ladungsträgerdichte durch Variation der Gate-Spannung, so schieben wir das chemische Potential  $\mu$  nach oben. Wie Abb. 10.43(b) zeigt, nimmt die Zahl der Zustände mit  $E < \mu$  dadurch zu. Die Größe der vorhandenen Quantum-Hall-Droplets wächst und neue kommen hinzu. Haben wir die Ladungsträgerdichte soweit erhöht, dass die Droplets überlappen [siehe Abb. 10.44(b)], so können Elektronen über die zusammenhängenden Droplets von der einen zur anderen Probenseite gelangen. Wir haben es jetzt mit ausgedehnten Zuständen zu tun und  $\mu$  liegt im Bereich der delokalisierten Zustände. Erhöhen wir die Ladungsträgerdichte weiter, so liegt  $\mu$  schließlich im Bereich der lokalisierten Zustände oberhalb des delokalisierten Bereichs. Wie Abb. 10.44(c) zeigt, erhalten wir hier Quantum-Hall-Droplets, die keine Zustände enthalten. Wir bezeichnen diese als Loch-Droplets.<sup>57</sup> Die Bewegungs-

<sup>57</sup> Anschaulich entspricht die Situation einem Potenzialgebirge, das wir in einer Badewanne langsam mit Wasser überfluten. Am Anfang bilden sich einige kleinere Seen an den tiefsten Stellen des Potenzialgebirges (Droplets). Steigt der Wasserspiegel, so überlappen sich irgendwann die Seen und bilden eine zusammenhängende Wasserfläche, die vom einen bis zum anderen Rand der Badewanne

richtung der Skipping Orbits um die Loch-Droplets herum ist der Bewegungsrichtung der Skipping Orbits innerhalb der Droplets in Abb. 10.44(a) genau entgegengesetzt.

Wir können nun das in Abb. 10.43 und 10.44 gezeigte Bild verwenden, um das Entstehen der Hall-Plateaus zu erklären. Für kleine Ladungsträgerdichten liegen nur wenige Droplets vor, die klein und wohl getrennt voneinander sind. Das heißt, dass die Orbits der Randelektronen nicht überlappen und somit lokalisiert sind. Somit ist  $\sigma_{xx} = \rho_{xx} = 0$ . Erhöhen wir die Ladungsträgerdichte, so überlappen einige Droplets,  $\sigma_{xx}$  und  $\rho_{xx}$  bleiben aber null, solange die Droplets noch keinen geschlossenen Pfad zwischen den Spannungskontakten bilden können. Erst bei weiterer Erhöhung der Ladungsträgerdichte bildet sich ein Perkolationspfad von der einen zur anderen Probenseite aus. Wir erhalten dann ausgedehnte Zustände, wodurch  $\sigma_{xx}$  und  $\rho_{xx}$  endlich werden. Bei weiterer Erhöhung der Ladungsträgerdichte rückt das äußerste Skipping Orbit an den Rand der Probe, wo das Potenzial aufgrund des Einschusses der Elektronen in der Probe steil ansteigt (Rand der Badewanne). Es bleiben nur noch einige kleine Loch-Droplets übrig [siehe Abb. 10.44(c)]. In diesem Fall wird  $\sigma_{xx} = \rho_{xx} = 0$  und  $\sigma_{xy} = e^2/h$ . Das Verhältnis  $\sigma_{xy}/(e^2/h)$  gibt die Zahl der Randkanäle im Perkolationsprozess an. Jeder Randkanal stellt einen eindimensionalen Leitungskanal dar und trägt genau die Leitfähigkeit  $e^2/h$  bei (vergleiche hierzu Abschnitt 7.5.1). Jedes Mal, wenn wir das chemische Potenzial über ein weiteres Landau-Niveau anheben, entsteht ein neuer Randkanal und wir erhalten  $\sigma_{xy} = n \cdot (e^2/h)$ . Wir erhalten also Hall-Plateaus, wenn sich das chemische Potenzial im Bereich der lokalisierten Zustände befindet, und der Wert des Hall-Widerstands hängt von der Zahl  $n$  der Randkanäle ab. Wir sehen, dass wir bei einer starken Unordnung (breiter Bereich der lokalisierten Zustände), also in Proben mit einer hohen Störstellenkonzentration, breite Hall-Plateaus bekommen. Wir dürfen die Störstellenkonzentration aber auch nicht zu hoch machen, da dann die Streuzeit  $\tau$  sehr klein und die Breite  $\hbar/\tau$  der Landau-Niveaus größer als ihr Abstand  $\hbar\omega_c$  wird, d. h. wir erfüllen dann die Bedingung  $\omega_c \tau \gg 1$  nicht mehr.

Die bereits in Abschnitt 7.5.1 geführte Diskussion und Gleichung (7.5.6) sind nur für einen einzelnen Transportkanal gültig. Fassen wir die gerade diskutierten Randkanäle als eindimensionale Transportkanäle auf, so haben wir es mit einem System aus mehreren Transportkanälen und mehreren Kontakten zu tun. Diese Situation wird durch den *Landauer-Büttiker-Formalismus* beschrieben, auf den wir hier kurz eingehen wollen.<sup>58</sup> Für die in Abb. 10.45 gezeigte Konfiguration erhalten wir

$$I_i = \frac{e}{h} \left\{ (n - r_{ii})\mu_i - \sum_{j=1}^6 t_{ij}\mu_j \right\}, \quad (10.5.20)$$

wobei  $I_i$  der Strom in den  $i$ -ten Kontakt,  $r_{ii}$  die Rückstreuwahrscheinlichkeit von Kontakt  $i$  nach Kontakt  $i$  und  $t_{ij}$  die Transmissionswahrscheinlichkeit von Kontakt  $i$  nach Kontakt  $j$  ist. Mit  $r_{ii} = 0$  und  $t_{ij} = 1$  folgt für den Hall-Widerstand bei in unserem Beispiel zwei Rand-

---

ne reicht (ausgedehnte Zustände). Schließlich werden nach weiterem Anstieg des Wasserpegels nur noch einige Bergspitzen aus dem Wasser ragen, die quasi Löcher in der ansonsten geschlossenen Wasserfläche bilden (Loch-Droplets).

<sup>58</sup> siehe z. B. *The Physics of Low Dimensional Semiconductors*, J. H. Davies, Cambridge University Press, Cambridge (1998).

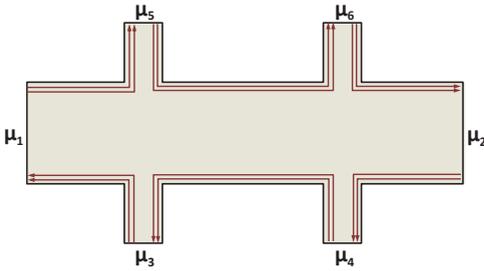


Abb. 10.45: Zur Veranschaulichung des Quanten-Hall-Widerstands im Randkanalbild.

kanälen

$$R_{35} = \frac{\mu_3 - \mu_5}{eI} = \frac{h}{2e^2}, \quad (10.5.21)$$

da

$$I_1 = I = \frac{2e}{h}(\mu_1 - \mu_5) \quad (10.5.22)$$

und

$$I_3 = 0 = \frac{2e}{h}(\mu_3 - \mu_1) \quad (10.5.23)$$

$$I_5 = 0 = \frac{2e}{h}(\mu_5 - \mu_1). \quad (10.5.24)$$

Um uns dieses etwas abstrakte Ergebnis zu veranschaulichen, nehmen wir an, dass wir ein negatives Potenzial  $-\mu_1$  an Kontakt 1 angelegt haben und  $\mu_2 = 0$ . Es fließt dann ein technischer Strom von Kontakt 2 nach 1, bzw. ein Elektronenstrom von Kontakt 1 nach 2. In Abb. 10.45 geben die Pfeile die Elektronenrichtung an. Aus Kontakt 1 werden also Elektronen in die beiden Randkanäle injiziert und verlassen diesen Kontakt. Sie können sich ohne Streuung am oberen Rand der Probe entlangbewegen und treten in Kontakt 5 ein. Dieser Kontakt kann aber keinen Strom aufnehmen. Sein chemisches Potenzial muss deshalb soweit ansteigen, damit er einen gleich großen Strom wieder in den auslaufenden Randkanal abgeben kann. Dies erfordert  $\mu_5 = \mu_1$ . Das gleiche gilt für Kontakt 6, d. h.  $\mu_6 = \mu_1$ . Somit haben alle Kontakte entlang der oberen Probenseite das gleiche Potenzial. Mit der gleichen Argumentation können wir folgern, dass alle Kontakte auf der Probenunterseite auf dem gleichen Potenzial  $\mu_2 = 0$  liegen (damit tragen die unteren Randkanäle keinen Strom).

Der gesamte aus Kontakt 1 injizierte Strom ist für die beiden Randkanäle  $I = (2e/h)(\mu_1 - \mu_3) = -(2e/h)\mu_1$  und der Hall-Widerstand ist  $R_{35} = (\mu_3 - \mu_5)/eI = -\mu_5/eI = h/2e^2$ . Wir erhalten also den quantisierten Wert. Ferner ist der longitudinale Widerstand  $(\mu_6 - \mu_5)/eI = 0$ . Wir haben also in relativ einfacher Weise die experimentellen Beobachtungen erklärt. Eine wichtige Voraussetzung ist, dass die Transmissionswahrscheinlichkeit in den Randkanälen gleich eins ist. Dies ist gegeben, wenn keine Streuung von der einen Probenseite auf die andere erfolgt. In der in Abb. 10.44b dargestellten Situation ist dies offensichtlich nicht der Fall. Hier befinden wir uns gerade im Bereich zwischen den Hall-Plateaus.

Wir möchten schließlich noch darauf hinweisen, dass eine allgemeinere Interpretation des Quanten-Hall-Effekts von **Robert B. Laughlin** gegeben wurde.<sup>59,60</sup> Dabei wird der Quanten-Hall-Effekt als direkte Folge des allgemeinen Prinzips der Eichinvarianz ausgedrückt. In einer moderneren Betrachtungsweise wird der Quanten-Hall-Zustand als topologischer Isolator klassifiziert (siehe hierzu Abschnitt 10.6). Für eine tiefergehendere Diskussion des Quanten-Hall-Effekts wird auf die Spezialliteratur verwiesen.<sup>61,62,63,64,65</sup>



## 10.5.4 Vertiefungsthema: Fraktionaler Quanten-Hall-Effekt

In Halbleiter-Heterostrukturen mit hohen Beweglichkeiten sind die Plateaus des ganzzahligen Quanten-Hall-Effekts viel schmaler als in Proben mit geringer Beweglichkeit, wie sie in dem ursprünglichen Experiment von von Klitzing, Dorda und Pepper verwendet wurden. Zwischen diesen schmalen Plateaus wurden nun weitere Plateaus bei gebrochen-rationalen Vielfachen  $f$  des Quantenwiderstands  $R_K$  gefunden (siehe Abb. 10.46). Die Plateaus

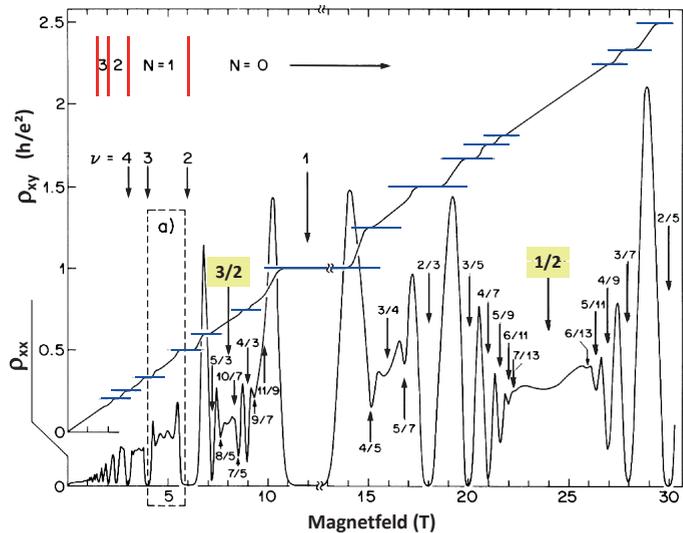


Abb. 10.46: Fraktionaler Quanten-Hall-Effekt (nach R. Willett, J. P. Eisenstein, H. L. Störmer, D. C. Tsui, A. C. Gossard, Phys. Rev. Lett. **59**, 1776 (1987), © (2012) American Physical Society).

- <sup>59</sup> Robert B. Laughlin, *Quantized Hall conductivity in two dimensions*, Phys. Rev. B **23**, 5632 (1981).  
<sup>60</sup> R. B. Laughlin, in *McGraw-Hill Book of Science and Technology 1984*, McGraw-Hill, New York (1984).  
<sup>61</sup> H. L. Störmer, D. C. Tsui, *The Quantized Hall Effect*, Science **220**, 1241–1246 (1983).  
<sup>62</sup> M. Janssen, O. Viehewer, U. Fastenrath, J. Hajdu, *Introduction to the Theory of the Integer Quantum Hall Effect*, VCH Weinheim (1994).  
<sup>63</sup> R. E. Prange, S. M. Girvin, *The Quantum Hall Effect*, Springer Verlag, New York, Berlin (1990).  
<sup>64</sup> T. Chakraborty, P. Pietiläinen, *The Quantum Hall Effect – Fractional and Integral*, Springer Series on Solid-State Science, Vol. 85, Springer Berlin, Heidelberg (1995).  
<sup>65</sup> A. H. McDonald, *The Quantum Hall Effect: A Perspective*, Kluwer, Boston (1989).

treten für  $f = q/(2pq \pm 1)$  und  $f = 1 - q/(2pq \pm 1)$  ( $p, q =$  ganze Zahl), also für  $f = 1/3, 2/3, 2/5, 3/5, \text{etc.}$  auf. Die Werte von  $f$  sind also gekennzeichnet durch einen ungeraden Nenner und eine Symmetrie um  $\nu = 1/2$ . Diesen Effekt, der 1983 von **Horst L. Störmer** und **Daniel C. Tsui** entdeckt wurde, nennen wir *gebrochenzahligen* oder *fraktionalen Quanten-Hall-Effekt*.<sup>66</sup> Für diese Entdeckung erhielten Störmer und Tsui im Jahr 1998 zusammen mit **Robert B. Laughlin**, dem es gelang, die experimentellen Beobachtungen theoretisch zu erklären,<sup>67</sup> den Nobelpreis für Physik.

Wie beim ganzzahligen QHE entstehen auch beim fraktionalen QHE Plateaus im Hall-Widerstand, wenn das chemische Potenzial in einer Lücke der Zustandsdichte bzw. in einem Bereich lokalisierter Zustände liegt. Beim fraktionalen QHE ist allerdings die Ursache der Energielücke nicht einfach die Quantisierung von Einteilchenzuständen. Die Energielücke entsteht vielmehr durch die kollektive Bewegung aller Elektronen im System. Es handelt sich um einen Vielteilcheneffekt und ist theoretisch schwierig zu beschreiben.<sup>68</sup> Für  $\nu \leq 1$  befinden sich alle Elektronen im untersten Landau-Niveau, sie haben alle dieselbe kinetische Energie  $\frac{1}{2}\hbar\omega_c$  und ihre Spins sind alle parallel ausgerichtet. Die Vielteilcheneffekte kommen nun dadurch zustande, dass wir die Coulomb-Wechselwirkung zwischen den Elektronen berücksichtigen müssen.

In jüngster Zeit wurde zur Beschreibung des fraktionalen QHE das Konzept der *zusammengesetzten Fermionen* (Composite Fermions) eingeführt. Diese Quasiteilchen bestehen aus einem Elektron (oder Loch), das mit einer geraden Zahl von magnetischen Flussquanten verbunden ist. Jedes Elektron bildet also mit Quanten des magnetischen Flusses ein zusammengesetztes (composite) Fermion. Die an die Elektronen angehängten Flussquanten bilden ein homogenes Magnetfeld, das dem äußeren Magnetfeld entgegenwirkt. Das zusammengesetzte Fermion, das eine ganzzahlige Ladung trägt, bewegt sich in einem durch die Flussquanten reduzierten, effektiven Magnetfeld. Mit diesen Quasiteilchen kann der fraktionale QHE auf den ganzzahligen QHE abgebildet werden. Das heißt, der fraktionale QHE von Elektronen in einem äußeren Feld wird dann der ganzzahlige QHE der neuen zusammengesetzten Fermionen in einem effektiven Feld. Die zusammengesetzten Fermionen haben eine ganzzahlige Ladung. Da sie sich aber in einem effektiven Feld bewegen, erscheint es so, als ob sie eine fraktionale topologische Ladung besäßen. Das Bild der zusammengesetzten Fermionen kann alle experimentellen Beobachtung Rechnung tragen. Insbesondere erklärt es, dass  $\nu = 1/2$  ein besonderer Zustand ist, da hier für ein Quasiteilchen aus einem Elektron und zwei Flussquanten das effektive Feld null ist. Abb. 10.46 zeigt in der Tat, dass sich sowohl  $\rho_{xy}$  als auch  $\rho_{xx}$  um  $\nu = \frac{1}{2}$  ähnlich zu den um  $B = 0$  gemessenen Größen ist.

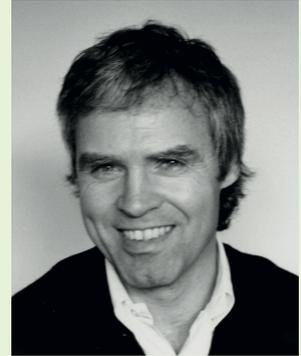
<sup>66</sup> D. C. Tsui, H. L. Störmer, A. C. Gossard, *Two-Dimensional Magnetotransport in the Extreme Quantum Limit*, Phys. Rev. Lett. **48**, 1559 (1982).

<sup>67</sup> R. B. Laughlin, *Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations*, Phys. Rev. Lett. **50**, 1395 (1983).

<sup>68</sup> T. Chakraborty, P. Pietiläinen, *The Quantum Hall Effect – Fractional and Integral*, Springer Series on Solid-State Science, Vol. 85, Springer Berlin, Heidelberg (1995).

### Horst Ludwig Störmer (geb. 1949), Nobelpreis für Physik 1998

Horst Ludwig Störmer wurde am 6. April 1949 in Frankfurt am Main geboren. Hier studierte er zunächst Mathematik, da er den Anmeldezeitraum für das Physikstudium verpasst hatte, und wechselte erst 1968 zur Physik. Nach dem Diplom 1974 am Lehrstuhl von Werner Martienssen – zur gleichen Zeit arbeitete auch Gerd Binnig in der Arbeitsgruppe – wechselte er an das Hochfeldmagnetlabor in Grenoble in die Arbeitsgruppe um Prof. Queisser. Er promovierte 1977 an der Universität Stuttgart und ging anschließend als Postdoktorand an die Bell Laboratories in Murray Hill, New Jersey. Dort erhielt er 1978 eine Festanstellung und wurde 1983 zuerst zum Leiter der Abteilung für elektrische und optische Eigenschaften von Festkörpern und später 1992 zum Direktor des Physikalischen Forschungslabors der Bell Labs ernannt. Im Jahr 1997 wechselte Störmer an die Columbia University in New York, wo er 1998 zum Professor für Angewandte Physik ernannt wurde. Hier lehrt und forscht er vor allem auf dem Gebiet der Nanowissenschaften.



©The Nobel Foundation.

Den Nobelpreis für Physik erhielt Horst L. Störmer 1998 zusammen mit den US-amerikanischen Forschern Robert B. Laughlin und Daniel C. Tsui für ihre Entdeckung, dass Elektronen in starken Magnetfeldern und bei extrem niedrigen Temperaturen eine neue Art von Quantenflüssigkeit mit fraktionell geladenen Anregungen bilden.

## 10.6 Topologische Quantenmaterialien

Kondensierte Materie kann in vielfältigen Erscheinungsformen mit unterschiedlichen Eigenschaften auftreten. Verschiedene Zustandsformen können dabei durch die Symmetrien charakterisiert werden, die sie brechen. Beispiele sind:

- kristalline Festkörper → gebrochene Translationssymmetrie,
- Flüssigkristalle → gebrochene Rotationssymmetrie,
- ferroelektrische Materialien → gebrochene Inversionssymmetrie,
- ferromagnetische Materialien → gebrochene Rotationssymmetrie,
- supraleitende Materialien → gebrochene Eichsymmetrie.

Der 1980 entdeckte Quanten-Hall (QH) Zustand war das erste Beispiel für einen Quantenzustand, der nicht durch eine spontan gebrochene Symmetrie klassifiziert werden konnte. Seine fundamentalen physikalischen Eigenschaften wie die quantisierte Hall-Leitfähigkeit werden vielmehr durch die topologische Struktur des zugrunde liegenden quantenmechanischen Zustands bestimmt. Wir bezeichnen ihn deshalb heute als topologischen Isolator. Unter topologischen Quantenmaterialien verstehen wir heute allgemein solche Zustandsfor-

men der kondensierter Materie, die durch eine neuartige Ordnung charakterisiert werden, die nicht in das allgemeine Paradigma der Symmetriebrechung passt. Sie werden stattdessen durch eine globale Größe beschrieben, die wir als topologische Ordnung bezeichnen und die nicht von den Details des jeweiligen Systems abhängt. Um uns dies anhand eines Beispiels klarzumachen, betrachten wir einen Torus und ein Krug mit Henkel. Obwohl beide sehr unterschiedlich ausschauen, besitzen sie die gleiche Topologie, da sie durch eine kontinuierliche Umformung ineinander übergeführt werden können. Ein Krug ohne Henkel hätte dagegen eine andere Topologie. Wir werden im Folgenden sehen, dass die Quantenphysik im Zusammenspiel mit der Topologie der elektronischen Bandstruktur zu einer neuen Klasse von Materialien – den *topologischen Quantenmaterialien* – führt.

Das Feld der topologischen Quantenmaterialien hat mit der Entdeckung der *topologischen Isolatoren* großes Interesse gefunden. Lange Zeit dachte man, dass alle Bandisolatoren (vergleiche Abschnitt 8.4.3) prinzipiell äquivalent sind. Eine topologische Bandtheorie sagt nun aber für zweidimensionale Systeme zwei fundamental unterschiedliche Klassen und für dreidimensionale Systeme sogar 16 unterschiedliche Klassen von Bandisolatoren voraus. Die Existenz einer nichttrivialen topologischen Ordnung in einem Isolator führt zu charakteristischen physikalischen Eigenschaften. Die wohl bemerkenswerteste Konsequenz ist die Existenz von Oberflächenzuständen, die keine Energielücke besitzen. Diese führen dazu, dass topologische Isolatoren in ihrem Inneren zwar elektrisch isolierend, an ihrer Oberfläche aber elektrisch leitend sind. Außerdem gibt es Materialien, in denen Elektronen mit entgegengesetzter Bewegungsrichtung eine entgegengesetzte Spin-Richtung besitzen. Bewegen sich also z.B. gleich viele Elektronen in Vorwärts- und Rückwärtsrichtungen, so führt das zu einem verschwindenden Ladungstrom. Aufgrund der entgegengesetzten Spin-Richtungen resultiert aber ein endlicher Spin-Strom. Diese Tatsache ist für das Gebiet der Spin-Elektronik von großem Interesse. Die topologischen Isolatoren wurden 2005 zuerst für zweidimensionale<sup>69</sup> und wenig später im Jahr 2007 auch für dreidimensionale Materialien<sup>70</sup> theoretisch vorhergesagt. Eine erste experimentelle Bestätigung folgte für zweidimensionale Systeme bereits 2007.<sup>71</sup> Der in den Experimenten an zweidimensionalen Systemen gemessene elektrische Widerstand war immer derselbe, unabhängig von der Breite der Probe und zwar halb so groß wie der Quantenwiderstand  $R_K = h/e^2$ . Dies deutet darauf hin, dass die Randströme in den topologischen Isolatoren nur halb so viele Elektronen transportieren wie die Randkanäle in Quanten-Hall-Proben. Wie wir später sehen werden ist der Grund hierfür, dass die Bewegungsrichtung der Elektronen mit der Orientierung ihres Spins gekoppelt ist: Spin- $\uparrow$  Elektronen können nur in die eine Richtung, Spin- $\downarrow$  Elektronen nur in die entgegengesetzte Richtung fließen.

<sup>69</sup> C. L. Kane, E. J. Mele, *Z<sub>2</sub> Topological Order and the Quantum Spin Hall Effect*, Phys. Rev. Lett. **95**, 146802 (2005); siehe auch Phys. Rev. Lett. **95**, 226801 (2005).

<sup>70</sup> L. Fu, C. L. Kane, E. J. Mele, *Topological Insulators in Three Dimensions*, Phys. Rev. Lett. **98**, 106803 (2007).

<sup>71</sup> M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, S.-C. Zhang, *Quantum Spin Hall Insulator State in HgTe Quantum Wells*, Science **318**, 766-770 (2007).

Das Feld der topologischen Quantenmaterialien ist noch jung und entwickelt sich sehr schnell.<sup>72,73,74</sup> Neben topologischen Isolatoren werden heute auch topologische Supraleiter oder topologische nodale Halbmetalle diskutiert. Wir wollen im Folgenden einige Grundzüge dieser sehr interessanten Materialien zusammenfassen.

## 10.6.1 Topologie und Bandstruktur

Zum Verständnis von topologischen Quantenmaterialien benötigen wir grundlegendes Wissen zur Topologie und Bandstruktur. Wir werden hier das Konzept der topologischen Äquivalenz einführen und seine Bedeutung bezüglich der elektronischen Bandstruktur von Festkörpern erläutern. Wir werden dabei die Berry-Phase einführen, die ein wichtiges konzeptionelles Werkzeug für die Analyse von topologischen Zuständen ist.

### 10.6.1.1 Klassifizierung von geometrischen Körpern

Die Topologie ist ein Teilgebiet der Mathematik. Sie untersucht jene Eigenschaften von geometrischen Körpern, die sich bei einer stetigen Verformung nicht verändern. Ein bekanntes Beispiel ist die topologische Äquivalenz eines Krugs mit Henkel und eines Torus. Beide geometrische Körper können dann mit dem gleichen topologischen Index klassifiziert werden. Die gleiche topologische Eigenschaft der beiden geometrischen Körper besteht darin, dass wir auf ihren Oberflächen geschlossene Wege finden können, die durch ein Loch (beim Krug ist dies der Henkel) hindurchgehen, und solche, die das nicht tun. Wenn wir den Torus kontinuierlich in einen Krug verformen, ändert sich daran nichts. Würden wir jedoch den Torus zu einer Kugel verformen, dann verschwindet das Loch, und die besagten Wege durch das Loch existieren dann nicht mehr. Der Torus und die Kugel sind also topologisch nicht äquivalent und werden durch zwei unterschiedliche topologische Invarianten klassifiziert.

Um uns das anhand eines einfachen Beispiels klar zu machen, betrachten wir die Euler-Poincaré-Charakteristik von geschlossenen Flächen. Wir können diese Flächen  $A$  mit einer Kennzahl  $\chi(A)$ , die auch als topologische Invariante bezeichnet wird, kennzeichnen (Gauß-Bonnet-Theorem):

$$\chi(A) = \frac{1}{2\pi} \int_A G dA. \quad (10.6.1)$$

Das Integral läuft über die gesamte geschlossene Fläche  $A$  und  $G$  ist deren Gaußsche Krümmung. Für die Kugel, für die  $G = 1/R^2$ , erhalten wir  $\chi = 2$ , für den Torus  $\chi = 0$  und den Doppeltorus  $\chi = -2$  (siehe Abb. 10.47). Zwei Flächen besitzen die gleiche Euler-Poincaré-

<sup>72</sup> B. A. Bernevig, T. L. Hughes, *Topological Insulators and Topological Superconductors*, Princeton University Press (2013).

<sup>73</sup> M. Franz, L. Molenkamp (Hg.), *Topological Insulators*, in Contemporary Concepts of Condensed Matter Series, Vol. 6, Elsevier, Amsterdam (2013).

<sup>74</sup> M. Z. Hasan, C. L. Kane, *Topological Insulators*, Rev. Mod. Phys. **82**, 3045-3067 (2010).

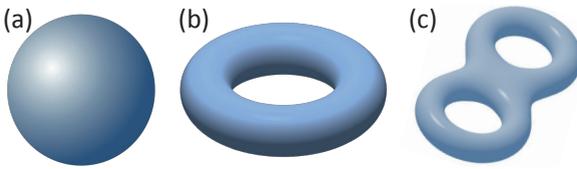


Abb. 10.47: Zur Veranschaulichung der Euler-Poincaré-Charakteristik geschlossener Flächen. Die Kugel (a) besitzt die topologische Invariante  $\chi = 2$ , der Torus (b)  $\chi = 0$  und der Doppeltorus (c)  $\chi = -2$ .

Charakteristik,<sup>75</sup> wenn sie stetig ineinander umgeformt werden können. Für solche mit unterschiedlichen Charakteristiken ist dies nicht möglich.<sup>76</sup>

Um uns den unten eingeführten Begriff der geometrischen Phasen zu veranschaulichen, wollen wir hier zuerst das geometrische Beispiel des so genannten Parallel- oder Levi-Civita-Transports diskutieren. Wir betrachten dazu eine Kugel und transportieren auf der Kugel einen Vektor entlang des in Abb. 10.48 gezeigten Weges. Der Vektor  $\mathbf{R}$  soll dabei immer parallel zur Kugeloberfläche sein ( $\mathbf{R} \cdot \hat{\mathbf{n}} = 0$ ) und darf auch nicht um die lokale Normalenrichtung  $\hat{\mathbf{n}}$  gedreht werden ( $d\mathbf{R} \times \hat{\mathbf{n}} = 0$ ). Diesen Transport nennen wir Paralleltransport. Nachdem wir den geschlossenen Weg  $\Gamma$  durchlaufen haben und an den Nordpol zurückgekehrt sind, stellen wir fest, dass der Vektor seine Richtung um den Winkel  $\alpha$  geändert hat. Wir erhalten also eine so genannte Holonomie, die den Vektor vor und nach dem Transport unterscheidet und zwar nur aufgrund der Krümmung der Kugel. Würden wir nämlich den umfahrenen Sektor der Kugel auf eine ebene Unterlage abflachen (siehe Abb. 10.48, rechts), so würde der Normalenvektor  $\hat{\mathbf{n}}$  immer aus der Papierebene heraus und damit wegen  $d\mathbf{R} \times \hat{\mathbf{n}} = 0$  auch  $\mathbf{R}$  immer in die gleiche Richtung zeigen. Der Winkel  $\alpha$  beim Paralleltransport auf der Kugeloberfläche ist gerade durch den von  $\Gamma$  umschlossenen halben Raumwinkel  $\Omega$  gegeben. Das Konzept des Paralleltransports kann auf beliebige Oberflächen verallgemeinert werden.

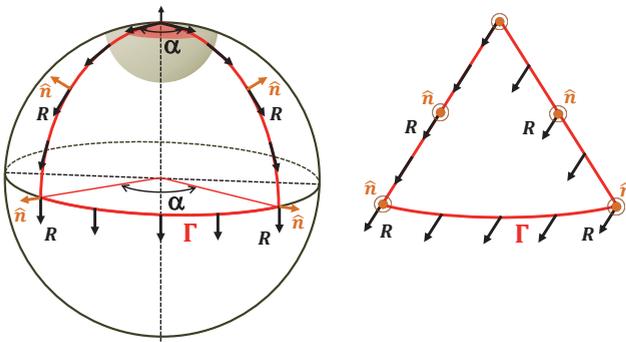


Abb. 10.48: Paralleltransport eines Vektors auf einer Kugeloberfläche. Rechts haben wir den Kugelsektor auf eine ebene Unterlage abgeflacht.

<sup>75</sup> Eine noch intuitivere Invariante ist das Geschlecht  $g = \frac{1}{2}(\chi - 2)$  der Fläche, das gerade die Zahl der Löcher angibt.

<sup>76</sup> M. Nakahara, *Geometry, Topology and Physics*, 2. Auflage, Taylor & Francis (2003).

### 10.6.1.2 Elektronische Bandstruktur und topologische Invarianten

Wir müssen uns nun fragen, welche Rolle die Topologie bei der Klassifizierung von Quantenmaterialien spielt. Die Antwort lautet: Es sind die Symmetrieeigenschaften der elektronischen Bandstruktur im reziproken Raum, welche die topologische Äquivalenz oder eben Nicht-Äquivalenz verschiedener Materialien festlegen. Es hat sich herausgestellt, dass die Gesamtheit aller Valenzbänder eines Materials eine topologische Eigenschaft besitzt, die wir wiederum mit einer geeigneten topologischen Invarianten klassifizieren können. Bei topologischen Isolatoren übernimmt die Rolle des Lochs bei unserem geometrischen Beispiel (Torus, Krug) die Bandlücke, durch die das Material zum Isolator wird. Wollen wir z.B. verschiedene isolierende Phasen hinsichtlich ihrer topologischen Eigenschaften klassifizieren, müssen wir untersuchen, ob wir ihre Valenzbänder kontinuierlich ineinander transformieren können, ohne dabei die Energielücke zu schließen.<sup>77</sup> Falls das möglich ist, würden alle isolierende Phasen die gleichen topologischen Eigenschaften besitzen. Es hat sich aber gezeigt, dass dies nicht der Fall ist. Wir bezeichnen nun Isolatoren, deren Valenzbänder nicht-triviale topologische Eigenschaften besitzen, als topologische Isolatoren. Solche mit trivialer Topologie sind die bekannten Bandisolatoren. Zur Klassifizierung von Isolatoren können wir wie bei der Klassifizierung geometrischer Objekte eine geeignete topologische Kennzahl verwenden.

Die Bandstruktur von Festkörpern haben wir bereits ausführlich in Kapitel 8 diskutiert. Für kristalline Festkörper können wir aufgrund der Translationssymmetrie Einteilchenzustände mit dem Kristallimpuls  $\mathbf{k}$  bezeichnen und sie durch Bloch-Wellen  $\Psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r})$  mit einer gitterperiodischen Funktion  $u_{\mathbf{k}}(\mathbf{r})$  beschreiben. Ein wesentlicher Aspekt der topologischen Bandtheorie ist es nun, topologisch verschiedene Hamilton-Operatoren  $\mathcal{H}$  mit Hilfe von topologischen Invarianten zu klassifizieren und damit unterschiedliche topologische Phasen herauszuparieren.

## 10.6.2 Berry-Phase und Chern-Zahl

Es ist wohlbekannt, dass die Energieabhängigkeit der Bloch-Wellenfunktionen viele Eigenschaften von Festkörpern bestimmt. Weniger bekannt ist die mit der *Berry-Phase*<sup>78</sup> verbundene Physik, die mit der  $\mathbf{k}$ -Abhängigkeit der Bloch-Wellenfunktionen verbunden ist. Den Ursprung der Berry-Phase können wir uns anhand des in Abb. 10.48 gezeigten Beispiels des Paralleltransports klar machen. Der Vektor  $\mathbf{R}$  erhielt dabei entlang einem geschlossenen Pfad eine Drehung. Wir können diese Drehung nur dann vermeiden, wenn wir den Vektor auf dem gleichen Weg in umgekehrter Richtung an den Ausgangspunkt zurücktransportieren würden. Dieses Bild können wir auf Bloch-Wellen in einem Festkörper übertragen. Wenn wir einen Bloch-Zustand langsam (adiabatisch) durch den Phasenraum bewegen und ihn

<sup>77</sup> Hierzu benutzen wir das Prinzip der adiabatischen Kontinuität. Zwei Isolatoren sind äquivalent, wenn wir sie durch eine langsame Änderung des Hamilton-Operators adiabatisch ineinander transformieren können. Dieser Prozess ist möglich, falls wir eine endliche Energielücke vorliegen haben, welche die Zeitskala dafür angibt, wie langsam wir die adiabatische Transformation machen müssen.

<sup>78</sup> M. V. Berry, *Quantal Phase Factors Accompanying Adiabatic Changes*, Proc. R. Soc. Lond. A 392, 45 (1984).

irgendwann wieder an seinen Ausgangspunkt zurückbringen, erwarten wir intuitiv, dass wir (bis auf einen Phasenfaktor, welcher der Zeitentwicklung Rechnung trägt) wieder den gleichen Zustand vorliegen haben. Wie beim Beispiel des Paralleltransports stellen wir nun aber fest, dass der Endzustand vom ursprünglichen Zustand abweicht, wenn wir eine geschlossene Schleife im Parameterraum durchlaufen haben. Insbesondere können wir einen geometrischen Phasenfaktor erhalten, den wir als Berry-Phase bezeichnen. Die Berry-Phase spielt eine zentrale Rolle bei der Diskussion der Topologie der elektronischen Bandstruktur.

Wir betrachten ein System, das wir mit dem Parametervektor  $\mathbf{R}$  beschreiben und sich adiabatisch entlang einem Pfad  $\Gamma$  in seinem Parameterraum bewegt. Befindet sich das System zur Zeit  $t_0$  in einem bestimmten Energiezustand  $E_m(t_0)$  ( $m$  ist der Bandindex) und ändern wir die Systemparameter adiabatisch, so bleibt das System in dem sich zeitlich ändernden Energiezustand  $E_m(t)$ . Wir nehmen ferner an, dass keine Entartung vorliegt, so dass sich die Energieniveaus nicht schneiden. Vergleichen wir die Phase des System im Anfangs- und Endpunkt des Pfades, so stellen wir eine Phasendifferenz fest. Diese setzt sich zum einen aus der dynamischen Phase

$$\phi_m = -\frac{1}{\hbar} \int_{t_0}^t E_m(t') dt', \quad (10.6.2)$$

die von einem Zustand mit Energie  $E_m(t)$  bei seiner Bewegung entlang dem Pfad  $\Gamma$  im betrachteten Parameterraum aufgesammelt wird, und zum anderen aus einer geometrischen Phase

$$\gamma_m(\Gamma) = i \int_{\Gamma} \langle \Psi_m | \nabla_{\mathbf{R}} | \Psi_m \rangle \cdot d\mathbf{R} = \int_{\Gamma} \mathbf{A}_m(\mathbf{R}) \cdot d\mathbf{R} \quad (10.6.3)$$

zusammen, die wir als Linienintegral über die vektorwertige Funktion  $\mathbf{A}_m(\mathbf{R})$ , die als *Mead-Berry-Vektorpotenzial* bezeichnet wird, ausdrücken können.

Um uns die Bedeutung der geometrischen Phase klar zu machen, betrachten wir jetzt geschlossene Pfade. Das heißt, nach der Umlaufzeit  $T$  gilt  $\mathbf{R}(t_0 + T) = \mathbf{R}(t_0)$ . Um zu sehen, ob wir die geometrische Phase wegeichen können, führen wir eine Transformation der Form

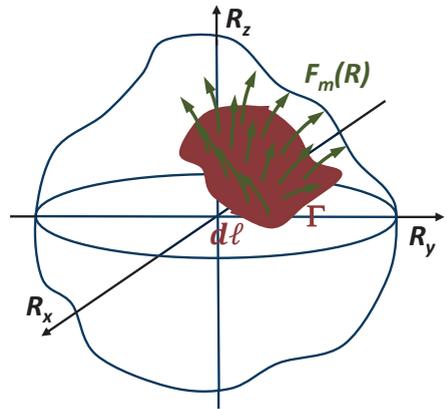
$$|\Psi_m(t)\rangle \rightarrow |\Psi'_m(t)\rangle = e^{-i\varphi(t)} |\Psi_m(t)\rangle \quad (10.6.4)$$

durch. Für den geschlossenen Weg muss  $e^{-i\varphi(t_0)} = e^{-i\varphi(t_0+T)}$  und damit  $\varphi(t_0) = \varphi(t_0 + T) + 2\pi n$  gelten, wobei  $n$  eine ganze Zahl ist. Mit (10.6.3) folgt dann sofort

$$\gamma'_m(\Gamma) = \oint_{\Gamma} \mathbf{A}'_m[\mathbf{R}(t)] \cdot d\mathbf{R} = \gamma_m(\Gamma) + 2\pi n. \quad (10.6.5)$$

Wir erkennen sofort, dass sich die Berry-Phase für einen geschlossenen Pfad nicht wegeichen lässt. Es ist klar, dass das Berry-Potenzial  $\mathbf{A}_m$  von der Eichung abhängt und somit keine physikalische Observable ist. Sein Integral über einen geschlossenen Pfad, die Berry-Phase  $\gamma_m$ , ist aber eichinvariant bis auf Vielfache von  $2\pi$ . Deshalb ist  $e^{i\gamma_m}$  absolut eichinvariant und besitzt Bezug zu physikalischen Observablen. Da für einen geschlossenen Pfad  $\mathbf{R}(T) = \mathbf{R}(0)$  gilt, hängt die Berry-Phase im Gegensatz zur dynamischen Phase nicht von der Umlaufzeit

Abb. 10.49: Zur Veranschaulichung der Herleitung der Berry-Phase und des Berry-Flusses. Wir bewegen uns entlang einer geschlossenen Schleife  $\Gamma$  im Parameterraum. Die Berry-Phase erhalten wir durch das Linienintegral des Berry-Potenzials  $\mathbf{A}_m(\mathbf{R})$  entlang  $\Gamma$ . Der Berry-Fluss  $\mathbf{F}_m = \nabla_{\mathbf{R}} \times \mathbf{A}_m$  durchsetzt die von  $\Gamma$  umschlossene rote Fläche. Falls diese Fläche geschlossen ist, schrumpft das Linienintegral auf einen Punkt und ergibt ein ganzzahliges Vielfaches von  $2\pi$ .



$T$  ab, sondern nur von der Geometrie des Parameterraums und des darin gewählten Pfades. Deshalb bezeichnet man sie als „geometrische Phase“.

Das Linienintegral (10.6.3) können wir mit dem Stokesschen Theorem in ein Oberflächenintegral über den **Berry-Fluss** (siehe hierzu Abb. 10.49)

$$\mathbf{F}_m(\mathbf{R}) = \nabla_{\mathbf{R}} \times \mathbf{A}_m(\mathbf{R}) \quad (10.6.6)$$

überführen. Falls die Oberfläche eine geschlossene Mannigfaltigkeit darstellt, verschwindet der Randterm. Da der Randterm aber bis auf ganzzahlige Vielfache von  $2\pi$  unbestimmt ist, erhalten wir das Chern-Theorem, das besagt, dass das Integral des Berry-Flusses (auch Berry-Krümmung genannt) über eine geschlossene Mannigfaltigkeit in Einheiten von  $2\pi$  quantisiert ist. Die damit verbundene Zahl ist die **Chern-Zahl**, die für das Verständnis von Quantisierungseffekten eine wichtige Rolle spielt.

Das Bloch-Theorem impliziert, dass der reziproke Raum geschlossen ist. Wie in Abb. 10.50 veranschaulicht ist, besitzt die 2D-Brillouin-Zone dieselbe Topologie wie ein zweidimensionaler Torus.<sup>79</sup> Aufgrund der Gitterperiodizität sind die Anfangs- und Endpunkte der roten und blauen Vektoren äquivalent, da sie sich um einen reziproken Gittervektor unterscheiden. Um der Gitterperiodizität Rechnung zu tragen, müssten wir deshalb die Spitzen der Vektoren an ihre Fußpunkte zurückbiegen. Dies ist beim Torus erfüllt. Der Integration über die Brillouin-Zone (BZ) entspricht also formal die Integration über eine geschlossene Fläche

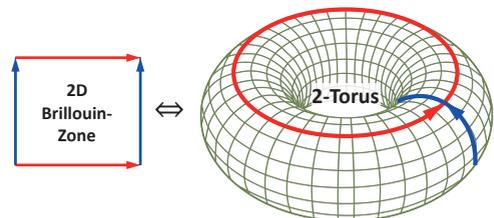


Abb. 10.50: Zur Veranschaulichung der topologischen Äquivalenz der 2D-Brillouin-Zone mit einem zweidimensionalen Torus.

<sup>79</sup> In  $d$ -Dimensionen besitzt die Brillouin-Zone die Topologie eines  $d$ -dimensionalen Torus  $T^d$ , der kurz  $d$ -Torus genannt wird.

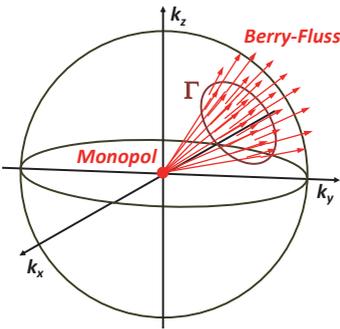


Abb. 10.51: Vektorfluss eines Monopolfeldes. Der Monopol im Ursprung erzeugt einen Berry-Fluss durch die von  $\Gamma$  umschlossene Fläche.

und wir erhalten in Analogie zu (10.6.1) die Chern-Zahl

$$n_m = \frac{1}{2\pi} \int_{\text{BZ}} \mathbf{F}_m(\mathbf{R}) \cdot d\mathbf{S}. \tag{10.6.7}$$

Wir können diesen Ausdruck in Analogie zur Elektrodynamik so interpretieren, dass  $\mathbf{F}_m(\mathbf{R})$  (die Krümmung einer Zusammenhangsform) einer „magnetischen Flussdichte“ entspricht und  $\mathbf{A}_m(\mathbf{R})$  (die Zusammenhangsform selbst) seinem Vektorpotenzial. Die Chern-Zahl  $n_m$  entspricht dann dem gesamten Fluss von  $\mathbf{F}_m(\mathbf{R})$  durch die Oberfläche der Brillouin-Zone. Diesen können wir uns durch einen magnetischen Monopol im Zentrum der Brillouin-Zone erzeugt denken (siehe Abb. 10.51). Die Berry-Phase und die Chern-Zahl sind im Allgemeinen sehr nützlich zur Klassifizierung von geschlossenen Schleifen im Parameterraum.

Summieren wir  $n_m$  über alle Valenzbänder auf, erhalten wir die Gesamt-Chern-Zahl

$$n = \sum_{m=1}^N n_m. \tag{10.6.8}$$

Wir erkennen natürlich sofort die Analogie zwischen Gleichung (10.6.1) und (10.6.7). Genauso wie das Integral der Gaußschen Krümmung eine quantisierte topologische Invariante (Euler-Poincaré-Charakteristik) zur Klassifizierung geometrischer Körper ist, trifft dies auch für das Integral der Berry-Krümmung (Chern-Zahl) zu, die wir zur Klassifizierung der Topologie der elektronischen Bandstruktur verwenden können. Das Beispiel zeigt, dass wir das von der Mathematik entwickelte, zugegebenermaßen für Physiker vielleicht etwas gewöhnungsbedürftige Konzept der topologischen Invarianten zur eindeutigen Klassifizierung von topologischer Materie benutzen können.

**Beispiel Spin-1/2-System:** Wir wenden die gerade eingeführten Konzepte auf ein Spin- $\frac{1}{2}$ -System an, das sich in einem Magnetfeld  $\mathbf{B}(t)$  befindet. Die zeitliche Variation des Feldes soll dabei so langsam sein, dass das Spin- $\frac{1}{2}$ -System dem Feld adiabatisch folgen kann. Wir nehmen an, dass das Magnetfeld mit der Kreisfrequenz  $\omega$  unter einem Polarwinkel  $\theta$  um die  $z$ -Achse rotiert (siehe Abb. 10.52):

$$\mathbf{B}(t) = B_0 \begin{pmatrix} \sin \theta \cos(\omega t) \\ \sin \theta \sin(\omega t) \\ \cos \theta \end{pmatrix}. \tag{10.6.9}$$

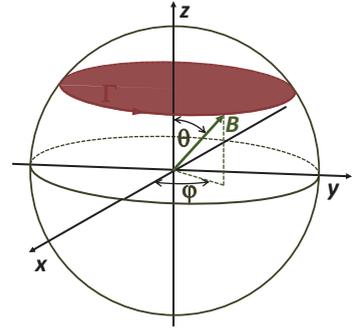


Abb. 10.52: Zur Berry-Phase eines Spin- $\frac{1}{2}$ -Systems im Magnetfeld. Die Berry-Phase ist proportional zum halben Raumwinkel, der durch den Pfad  $\Gamma$  umschlossen wird.

Da das Spin- $\frac{1}{2}$ -System dem Feld adiabatisch folgt, erhalten wir einen zeitunabhängigen Wechselwirkungsterm  $\mu_B \mathbf{B}_0 \cdot \boldsymbol{\sigma}$ , wobei  $\mu_B$  das Bohrsche Magneton und  $\boldsymbol{\sigma}$  der Vektor der Pauli-Matrizen ist. Wir erhalten die normalisierten Eigenzustände ( $\varphi = \omega t$ )

$$|m_{-}\rangle = \begin{pmatrix} \sin \frac{\theta}{2} e^{-i\varphi} \\ -\cos \frac{\theta}{2} \end{pmatrix} \quad |m_{+}\rangle = \begin{pmatrix} \cos \frac{\theta}{2} e^{-i\varphi} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad (10.6.10)$$

mit den zugehörigen Eigenenergien  $E_{\pm} = \pm \mu_B B_0$ . Sie stellen Punkte auf der Oberfläche der so genannten Bloch-Kugel dar. Berechnen wir die  $\theta$ - und  $\varphi$ -Komponente des Berry-Potenzials in Kugelkoordinaten ( $B_0, \theta, \varphi = \omega t$ ), so erhalten wir

$$\mathbf{A}_{-} = \begin{pmatrix} i \langle m_{-} | \nabla_{\theta} | m_{-} \rangle \\ i \langle m_{-} | \nabla_{\varphi} | m_{-} \rangle \end{pmatrix} = \frac{1}{B_0 \sin \theta} \begin{pmatrix} 0 \\ \sin^2 \frac{\theta}{2} \end{pmatrix} \quad (10.6.11)$$

$$\mathbf{A}_{+} = \begin{pmatrix} i \langle m_{+} | \nabla_{\theta} | m_{+} \rangle \\ i \langle m_{+} | \nabla_{\varphi} | m_{+} \rangle \end{pmatrix} = \frac{1}{B_0 \sin \theta} \begin{pmatrix} 0 \\ \cos^2 \frac{\theta}{2} \end{pmatrix}. \quad (10.6.12)$$

Integrieren wir dies entlang einem geschlossenen Pfad  $\Gamma$  ( $B_0 = \text{const}$ ,  $\theta = \text{const}$ ) im Parameterraum auf (siehe hierzu Abb. 10.52), ergibt sich

$$\gamma_{\pm}(\Gamma) = \oint A_{\varphi, \pm} B_0 \sin \theta d\varphi = W(\Gamma) \pi(1 \mp \cos \theta) = W(\Gamma) \frac{1}{2} \Omega(\Gamma). \quad (10.6.13)$$

Hierbei ist  $W(\Gamma)$  die Windungszahl des Pfades  $\Gamma$  und die Größe  $\pi(1 \mp \cos \theta)$  entspricht gerade dem halben Raumwinkel  $\Omega(\Gamma)$ , der durch den Pfad  $\Gamma$  umschlossen wird. Die bei einem Umlauf aufgesammelte dynamische Phase beträgt  $\phi_{\pm} = \pm \frac{1}{\hbar} \int_0^T \mu_B B_0 dt = \mu_B B_0 T / \hbar$  und hängt im Gegensatz zur Berry-Phase von der Umlaufzeit  $T$  ab.

Wir können mit (10.6.11) und (10.6.12) auch den Berry-Fluss  $\mathbf{F}_{\pm} = \nabla \times \mathbf{A}_{\pm}$  berechnen, für den nur die radiale Komponente ( $\mathbf{B}$ -Richtung)

$$F_{B, \pm} = \mp \frac{1}{2B_0^2}. \quad (10.6.14)$$

einen endlichen Wert hat. Wie wir anhand von Abb. 10.51 bereits veranschaulicht haben, können wir uns diesen Fluss durch einen magnetischen Monopol im Zentrum der Kugel erzeugt denken, der die Ladung  $\mp \frac{1}{2}$  besitzt. Da die betrachtete Kugel den Radius  $B_0$  hat, entspricht  $F_{B, \pm}$  bis auf einen Faktor  $\frac{1}{2}$  der Krümmung der Kugel, weshalb der Berry-Fluss auch häufig als Berry-Krümmung bezeichnet wird.

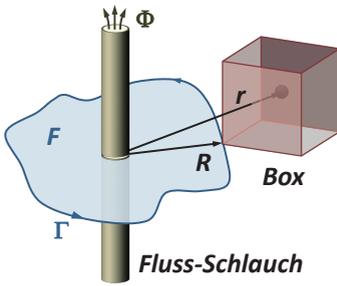


Abb. 10.53: Zum Aharonov-Bohm-Effekt: Ein magnetischer Fluss  $\Phi$  ist in einem Schlauch eingeschlossen und in seiner Nachbarschaft befindet sich eine Box mit Teilchen der Ladung  $q$ . Die Box wird adiabatisch entlang dem Pfad  $\Gamma$  um den Fluss-Schlauch bewegt.

**Beispiel Aharonov-Bohm-Effekt:** Als weiteres Beispiel betrachten wir die in Abb. 10.53 gezeigte Konfiguration, bei der ein magnetischer Fluss  $\Phi$  in einem Schlauch eingeschlossen ist und sich in der Nachbarschaft an der Stelle  $\mathbf{R}$  eine Box mit Teilchen der Ladung  $q$  befindet. Die magnetische Flussdichte  $\mathbf{B}$  verschwindet außerhalb des Schlauchs, während das zugehörige Vektorpotenzial  $\mathbf{a}$  nicht verschwindet, außer wenn  $\Phi = n\tilde{\Phi}_0 = n\frac{h}{e}$ .<sup>80</sup> Der Hamilton-Operator, der die Teilchen in der Box beschreibt, sei  $\mathcal{H}(\mathbf{p} - q\mathbf{a}, \mathbf{r} - \mathbf{R})$ . Die Wellenfunktionen für  $\mathbf{a} = 0$  sind  $\psi_m(\mathbf{r} - \mathbf{R})$  mit Eigenenergien  $E_m$ , die unabhängig von  $\mathbf{R}$  sind. Für  $\Phi \neq 0$  können wir die neuen Basiszustände  $|m(\mathbf{R})\rangle$  verwenden, die

$$\mathcal{H}(\mathbf{p} - q\mathbf{a}, \mathbf{r} - \mathbf{R}) |m(\mathbf{R})\rangle = E_m |m(\mathbf{R})\rangle \tag{10.6.15}$$

erfüllen und deren Lösungen durch

$$\langle \mathbf{r} | m(\mathbf{R}) \rangle = \exp \left[ i \frac{q}{\hbar} \int_{\mathbf{R}}^{\mathbf{r}} d\mathbf{r}' \mathbf{a}(\mathbf{r}') \right] \psi_m(\mathbf{r} - \mathbf{R}) \tag{10.6.16}$$

gegeben sind. Das Integral verläuft dabei entlang einem Pfad innerhalb der Box. Da wir immer eine Eichtransformation finden können, für die  $\mathbf{a}$  in der Box verschwindet, sind die Eigenenergien  $E_m$  unabhängig vom Vektorpotenzial.

Wir sehen, dass der Hamilton-Operator über das Vektorpotenzial von der Position  $\mathbf{R}$  der Box abhängt. Das bedeutet, dass der Parameterraum in unserem Beispiel der Ortsraum ohne den Bereich des Fluss-Schlauches ist. Wir transportieren jetzt die Box adiabatisch entlang einem geschlossenen Pfad  $\Gamma$  um den Schlauch herum. Mit dem Berry-Potenzial

$$\mathbf{A}(\mathbf{R}) = i \langle m(\mathbf{R}) | \nabla_{\mathbf{R}} | m(\mathbf{R}) \rangle = \frac{q}{\hbar} \mathbf{a}(\mathbf{R}) \tag{10.6.17}$$

erhalten wir die Berry-Phase

$$\gamma_m(\Gamma) = \oint_{\Gamma} \mathbf{A}(\mathbf{R}) \cdot d\boldsymbol{\ell} = \frac{q}{\hbar} \oint_{\Gamma} \mathbf{a}(\mathbf{R}) \cdot d\boldsymbol{\ell} = \frac{q}{\hbar} \int_F \mathbf{B} \cdot d\mathbf{F}. \tag{10.6.18}$$

Die Berry-Krümmung  $\mathbf{F}(\mathbf{R}) = \nabla \times \mathbf{A} = \frac{q}{\hbar} \mathbf{B}$  ist proportional zur magnetischen Flussdichte  $\mathbf{B}$  und verschwindet außerhalb des Schlauchs. Da wir den Bereich des Schlauchs aber aus

<sup>80</sup> Wir benutzen in diesem Abschnitt  $\mathbf{a}$  für das Vektorpotenzial, um Verwechslungen mit dem Berry-Potenzial  $\mathbf{A}$  zu vermeiden.

dem Parameterraum ausgeschlossen haben, ist dieser mehrfach verbunden. Die Berry-Phase ist deshalb rein topologisch und entspricht der Windungszahl  $W(\Gamma)$  des Pfades  $\Gamma$  um den Schlauch. Wir erhalten dann

$$\gamma_m(\Gamma) = W(\Gamma) \frac{q}{\hbar} \Phi = 2\pi W(\Gamma) \frac{\Phi}{\tilde{\Phi}_0}, \quad (10.6.19)$$

also eine ganzzahlige Vielfaches von  $2\pi$ , falls der eingeschlossene Fluss  $\Phi$  eine Vielfaches des Flussquants  $\tilde{\Phi}_0 = h/e$  ist.

### 10.6.3 Klassifizierung von Topologischen Isolatoren

Wenn wir heute ein Schema entwerfen sollten, nach dem wir topologische Isolatoren (TI) klassifizieren können, so wäre sicherlich eine Einteilung nach ihrer Raum-Zeit-Dimension und den erhaltenen Symmetrien (Zeitumkehrinvarianz, Ladungskonjugation, chirale Symmetrien) sinnvoll. Heute stehen vor allem zwei- und dreidimensionale, zeitumkehrinvariante Systeme im Fokus, da sie bereits in der Natur gefunden wurden. Prinzipiell muss es aber für jede diskrete Symmetrie eine topologisch isolierende Phase mit physikalischen Eigenschaften geben, die sich grundlegend von den dazu äquivalenten, topologisch trivialen Phasen unterscheiden. Neue Entdeckungen sind hier zu erwarten.

Historisch gesehen kann der in Abschnitt 10.5.3 diskutierte Quanten-Hall (QH) Zustand als der zuerst entdeckte TI betrachtet werden. Wir werden sehen, dass die gemessene QH-Leitfähigkeit direkt durch die Chern-Zahl gegeben ist. Im QH-Zustand wird allerdings die Zeitumkehrsymmetrie durch das von außen angelegte Magnetfeld gebrochen. Lange Zeit glaubte man, dass Zeitumkehrsymmetriebrechung und Zweidimensionalität grundlegende Voraussetzungen für topologische Isolatoren sind. Dies änderte sich allerdings, als 2005 TI zuerst für zweidimensionale und kurz darauf für dreidimensionale Systeme vorhergesagt wurden, in denen die Zeitumkehrsymmetrie nicht gebrochen ist. Die Eigenschaften dieser TI können sowohl mit einer topologischen Bandtheorie als auch mit einer topologischen Feldtheorie beschrieben werden. In der topologischen Bandtheorie wird der Isolator mit nichtwechselwirkenden Elektronen beschrieben, die eine bestimmte Anzahl von Bändern füllen. Sie erlaubt dann die Berechnung einer topologischen Invarianten, die z.B. für zweidimensionale Systeme nur die binären Werte 0 oder 1 haben kann. Wir sprechen von einer  $\mathbb{Z}_2$ -Klassifizierung. Mit der topologischen Feldtheorie wurde die Erweiterung auf wechselwirkende Elektronensysteme gemacht. Es stellte sich insgesamt heraus, dass für die Realisierung von TI ohne gebrochene Zeitumkehrsymmetrie eine starke Spin-Bahn-Wechselwirkung eine wichtige Ingredienz ist. Letztere ist bei schweren Elementen wie z.B. Hg, Pb oder Bi besonders stark ausgeprägt.

#### 10.6.3.1 Oberflächen und Grenzflächen

Wir haben bereits erwähnt, dass sich TI dadurch auszeichnen, dass sie zwar im Inneren isolierend, an der Oberfläche aber elektrisch leitend sind. Es stellt sich damit sofort die Frage, woher dieser fundamentale Unterschied zu den üblichen Bandisolatoren kommt, die ja sowohl im Innern als auch an der Oberfläche isolierend sind. Die Antwort ist erwartungsgemäß mit der nichttrivialen Topologie der Bandstruktur von TI verknüpft. Bringen wir einen

TI nun mit einem normalen Isolator (das kann auch Luft oder Vakuum sein) in Kontakt, der eine triviale Topologie besitzt, so entsteht an der Grenzfläche ein Problem. Anschaulich ist dies das gleiche Problem, das auftritt, wenn wir einen zu einer Brezel verschlungenen Draht mit einem völlig unverknöteten geraden Drahtstück verbinden wollen. Aufgrund der unterschiedlichen Topologie gelingt uns das nicht, es sei denn wir schneiden die Brezel auf. Das Gleiche passiert nun an der Grenzfläche zwischen einem TI und einem gewöhnlichen Isolator. Die verknötete Komponente ist hier die elektronische Wellenfunktion im  $\mathbf{k}$ -Raum. Da die beiden Isolatoren topologisch nicht äquivalent sind (die topologische Invariante ändert sich an der Grenzfläche), müssen sich an der Grenzfläche metallisch leitende Zustände ausbilden, was anschaulich gleichbedeutend mit dem Aufschneiden der Brezel ist. Übertragen auf das bereits oben verwendete Beispiel von Torus (er symbolisiert den TI) und Kugel (normaler Isolator) bedeutet dies, dass wir die beiden Formen nur dann miteinander verbinden können, wenn wir die Öffnung im Torus irgendwie schließen, was bei einem TI dem Schließen der Bandlücke entspricht. Auf eine mathematische Formulierung dieser Zusammenhänge wollen wir hier verzichten, da dies einige Zeit in Anspruch nehmen würde.

## 10.6.4 Zweidimensionale Topologische Isolatoren

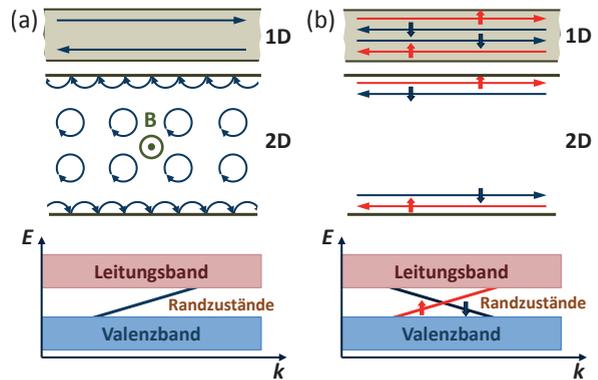
### 10.6.4.1 TI mit gebrochener Zeitumkehrsymmetrie

Wir betrachten zuerst zweidimensionale TI, bei denen die Zeitumkehrsymmetrie gebrochen ist. Der einfachste topologisch geordnete Zustand liegt in zweidimensionalen Elektronengasen in einem starken senkrechten Magnetfeld vor. Die Lorentz-Kraft resultiert in einer kreisförmigen Bewegung der Elektronen, die mit quantisierten Energien (Landau-Niveaus) verbunden ist. Dies entspricht der Situation in einem gewöhnlichen Isolator, in dem sich die lokalisierten Elektronen auf geschlossenen Bahnen um die Atomrümpfe bewegen. Wir erhalten aber einen isolierenden Zustand nur im Inneren der Probe, da sich am Rand durch Randstreuung so genannte "Skipping Orbits" ausbilden (siehe Abb. 10.54a). Dies führt zu einer Propagation der Ladungsträger entlang den beiden Rändern in entgegengesetzte Richtungen. Dies resultiert, wie bereits in Abschnitt 10.5.3 diskutiert wurde, in eindimensionalen Transportkanälen mit quantisierten Leitwerten, die Rückstreuung verboten ist. Vergleichen wir dies mit der Situation in einem eindimensionalen Leiter mit Elektronen, die in beide Richtungen propagieren (siehe Abb. 10.54a, oben), so sehen wir, dass die beiden Ränder der QH-Probe nur jeweils den halben Freiheitsgrad enthalten. Diese Eigenschaft macht den Quanten-Hall-Effekt (QHE) topologisch robust. Falls ein Elektron in einem Randkanal gestreut wird, macht es nur einen kleinen Umweg, wandert dann aber wieder in die gleiche Richtung.

Den bereits im Jahr 1980 von Klaus von Klitzing entdeckten Quanten-Hall-Zustand können wir heute als erstes Beispiel für einen topologischen Isolator betrachten. Bereits 1982 zeigten Thouless, Kohmoto, Nightingale und de Nijs (TKNN),<sup>81</sup> dass der Quanten-Hall-Effekt in einem zweidimensionalen Elektronengas auf eine topologische Eigenschaft der gefüllten Bänder (Landau-Niveaus) zurückgeführt werden kann. Sie zeigten, dass die Hall-Leitfähigkeit

<sup>81</sup> D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Quantized Hall Conductance in a Two-dimensional Periodic Potential*, Phys. Rev. Lett. **49** 405-408 (1982).

Abb. 10.54: Zur Veranschaulichung von zweidimensionalen topologischen Isolatoren (a) mit und (b) ohne Brechung der Zeitumkehrsymmetrie. Oben ist jeweils ein eindimensionaler Quantendraht und in der Mitte ein zweidimensionaler topologischer Isolator gezeigt, dessen oberer und unterer Rand jeweils als die Hälfte des Quantendrahts betrachtet werden können.



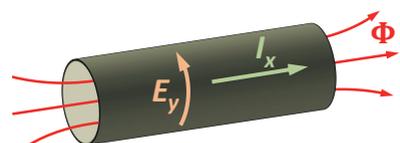
quantisiert ist und proportional zu einer topologischen Invarianten, der Chern-Zahl ist. Um uns dies klar zu machen, betrachten wir den in Abb. 10.55 gezeigten Zylinder mit Radius  $R$  und erhöhen adiabatisch den Fluss  $\Phi$  durch den Zylinder von  $\Phi = 0$  auf  $\Phi = \tilde{\Phi}_0 = h/e$ . Die Flussänderung erzeugt gemäß Faradayschem Induktionsgesetz eine elektrische Spannung  $U_y = -d\Phi/dt$  (bzw. ein elektrisches Feld  $E_y$ ) um den Zylinder, die wiederum einen Hall-Strom  $I_x = \sigma_{xy}U_y$  entlang dem Zylinder erzeugt. Haben wir den Fluss von 0 auf  $h/e$  erhöht, so haben wir insgesamt die Ladung  $Q = \sigma_{xy}h/e$  transportiert. Wenn wir  $\Phi = \tilde{\Phi}_0$  erreicht haben, können wir aber das Vektorpotenzial durch eine Eichtransformation eliminieren und erhalten wieder den gleichen Ausgangszustand wie für  $\Phi = 0$ . Von TKNN wurde gezeigt, dass die bei dem Vorgang entlang dem Zylinder erzeugte Polarisation  $\Delta P$  nur modulo  $e$  definiert ist. Deshalb kann sich die Polarisation nur um  $\Delta P = ne$  ändern (dies wird auch als Thouless Ladungspumpe bezeichnet). Das heißt, dass wir durch die gesamte Flussänderung eine Ladung  $Q = ne$ , also ein ganzzahliges Vielfaches der Elementarladung transportiert haben. Daraus ergibt sich sofort  $\sigma_{xy} = n \cdot (e^2/h)$ .

Die Chern-Zahl, die das System charakterisiert, können wir durch Aufsummierung über alle besetzten Subbänder (Landau-Niveaus) erhalten, die wir mit der azimuthalen Wellenzahl  $k_{m,y} = (m + \Phi/\tilde{\Phi}_0)/R$  indizieren können:

$$n = \sum_{m=1}^N \frac{1}{2\pi} \int_0^{\tilde{\Phi}_0} d\Phi \int dk_x \mathbf{F}[k_x, k_{m,y}(\Phi)]. \tag{10.6.20}$$

Wechseln wir von der Integrationsvariablen  $\Phi$  zu  $k_{m,y}$ , können wir zeigen, dass die Summe über die Integrale in ein einzelnes Integral über die 2D-Brillouin-Zone, also in ein Integral über eine geschlossene Fläche übergeht und somit der Chern-Zahl (10.6.8) entspricht.

Abb. 10.55: Eine Flussänderung von  $\tilde{\Phi}_0 = h/e$  durch den Zylinder resultiert in einer Polarisationsänderung, die der Verschiebung von einer Ladung  $Q = ne$  entlang dem Zylinders entspricht. Wir können den Zylinder durch eine kontinuierliche Transformation auch in eine Corbino-Scheibe überführen.



### 10.6.4.2 TI ohne gebrochene Zeitumkehrsymmetrie

Wir können uns nun fragen, ob wir das Magnetfeld los werden können und dadurch einen topologischen Zustand in einem System ohne Zeitumkehrsymmetriebrechung erhalten können. Wir betrachten zuerst einen eindimensionalen Leiter (siehe Abb. 10.54b, oben). In diesem hätten wir jeweils zwei in Vorwärts- und zwei in Rückwärtsrichtung propagierende Moden mit jeweils entgegengesetzter Spin-Richtung. Wir können nun eine zum Quanten-Hall-System analoge Situation schaffen, indem wir die Spin- $\uparrow$  Elektronen an einem Probenrand nur nach links und die Spin- $\downarrow$  Elektronen nur nach rechts laufen lassen und am gegenüberliegenden Rand genau umgekehrt (siehe Abb. 10.54b, Mitte). In diesem Fall erhalten wir an beiden Rändern keinen elektrischen Strom sondern einen reinen Spin-Strom, da die beiden Spin-Spezies in entgegengesetzte Richtungen laufen. Da der Spin-Strom die Rolle des elektrischen Stroms übernimmt, bezeichnen wir ein System mit solchen Randkanälen als Quanten-Spin-Hall (QSH) System.

Wir müssen jetzt noch klären, warum erstens sich in entgegengesetzte Richtung bewegende Elektronen entgegengesetzten Spin besitzen und warum zweitens in den Randkanälen keine Rückstreuung stattfindet. Die Antwort auf die erste Frage gibt die Spin-Bahn-Wechselwirkung, die wir im Detail in Abschnitt 12.5.4 diskutieren. Sie führt zu einer Kopplung zwischen der Spin-Richtung und der orbitalen Bewegung der Elektronen und ist besonders stark in Materialien ausgeprägt, die aus schweren Elementen aufgebaut sind.

Die Antwort auf die zweite Frage gibt die Zeitumkehrinvarianz. Um uns klarzumachen, warum die Rückstreuung durch nichtmagnetische Streuer verboten ist, obwohl in den Randkanälen eines QSH-Systems sich die Elektronen in beide Richtungen bewegen, betrachten wir den in Abb. 10.56 skizzierten Prozess, bei dem ein Elektron an einer Verunreinigung zurückgestreut wird. Das Elektron kann bei der Streuung entweder links- oder rechtsherum laufen. Der Spin rotiert dabei um den Winkel  $\pm\pi$ . Der Unterschied der Spin-Drehung für beide Pfade, die über die Zeitumkehrsymmetrie miteinander verknüpft sind, beträgt also gerade  $2\pi$ . Eine volle  $2\pi$ -Drehung des Spins erhalten wir durch eine zweimalige Anwendung des Zeitumkehroperators  $\mathcal{T}$ , die aber ein zusätzliches Minuszeichen liefert ( $\mathcal{T}$  ist ein antiunitärer Operator,  $\mathcal{T}^2 = -1$ ). Das heißt, die Wellenfunktion eines Spin-1/2-Teilchens erhält bei einer  $2\pi$ -Rotation des Spins ein Minuszeichen. Das bedeutet wiederum, dass die

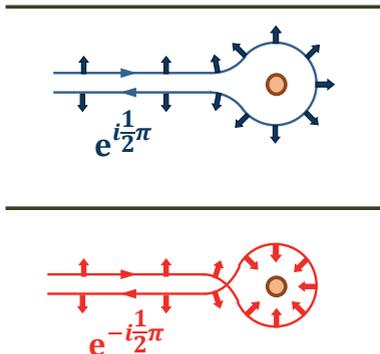
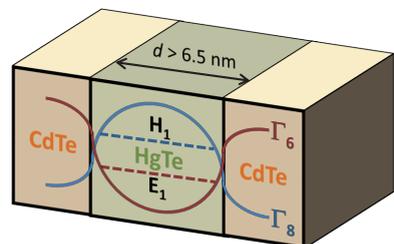


Abb. 10.56: Zur Veranschaulichung der Unterdrückung der Rückstreuung in einem Quanten-Spin-Hall-Randkanal. Die Streuung an einer nichtmagnetischen Störstelle kann in zwei unterschiedliche Richtungen erfolgen. Bei Streuung im Uhrzeigersinn (blaue Kurve) rotiert der Spin um den Winkel  $\pi$ , bei Streuung gegen den Uhrzeigersinn (rote Kurve) um  $-\pi$ . Der mit der relativen Spin-Drehung von  $2\pi$  verbundene Phasenfaktor von  $-1$  führt zu einer destruktiven Interferenz der beiden Pfade.

beiden Rückstreupfade destruktiv interferieren und somit die Rückstreuung verboten ist. Wir können deshalb sagen, dass die Robustheit der QSH-Randkanäle durch die Zeitumkehrsymmetrie gewährleistet wird. Das eben vorgestellte Bild gilt nur für einzelne Paare von QSH-Randkanälen. Würden wir die oberen und unteren Randkanäle in Abb. 10.54 in einem einzigen eindimensionalen Leiter zusammenbringen, so könnte ein Elektron von einem vorwärts- in einen rückwärtslaufenden Kanal gestreut werden, ohne den Spin zu ändern. Es würde dann keine destruktive Interferenz auftreten und wir hätten durch die endliche Rückstreuung eine endliche Dissipation vorliegen. Damit der QSH-Zustand robust ist, muss er aus einer ungeraden Zahl von vorwärts- und rückwärtslaufenden Moden bestehen. Dieser gerade-ungerade Effekt wird durch die  $\mathbb{Z}_2$ -Topologie charakterisiert. Da wir also den QSH-Zustand mit einer topologischen Invariante charakterisieren können, können wir ihn als topologischen Isolator bezeichnen.

**Realisierungsmöglichkeiten:** Eine allgemeine Vorgehensweise bei der Realisierung von zweidimensionalen TI ist die Verwendung von Halbleiter-Heterostrukturen, die eine Bandinversion zeigen. Hierbei heißt Bandinversion, dass die Anordnung von Valenzband und Leitungsband invertiert ist, was durch eine starke Spin-Bahn-Kopplung erreicht werden kann. In den meisten Halbleitermaterialien wird das Leitungsband aus Elektronen in  $s$ -Orbitalen und das Valenzband aus solchen in  $p$ -Orbitalen aufgebaut. In Materialien mit schweren Elementen wie HgTe ist die Spin-Bahn-Wechselwirkung nun so stark, dass das Valenzband oberhalb dem Leitungsband zu liegen kommt, so dass eine invertierte Bandstruktur entsteht. In CdTe liegt dagegen eine normale Bandfolge vor. Wächst man eine CdTe/HgTe/CdTe-Heterostruktur, so kann man den in Abb. 10.57 gezeigten Bandverlauf konstruieren. Bandstrukturberechnungen zeigen nun, dass bei genügender Dicke ( $d > 6.5$  nm) der HgTe-Schicht ein Paar von Randzuständen vorliegt, die unterschiedlichen Spin tragen. Die Randzustände dispergieren vom Valenzband bis zum Leitungsband (siehe Abb. 10.54b, unten) und schneiden sich. Der Schnittpunkt ist eine direkte Folge der Zeitumkehrsymmetrie und kann nicht entfernt werden – er stellt die topologische Signatur eines Quanten-Spin-Hall-Isolators dar.

Der in Abb. 10.57 skizzierte Quantentopf zeigt aufgrund der perfekt leitenden Randkanäle einen quantisierten Widerstand  $R = \frac{1}{2} R_K$ . In den Experimenten konnte gezeigt werden, dass sich dieser Widerstand nicht mit der Probenbreite ändert. Dieser Befund belegt, dass in der Tat nur die Randkanäle zum Transport beitragen, wogegen die Probe im Inneren isolierend ist. Im Vergleich zum Quanten-Hall-Effekt wird ein um den Faktor  $\frac{1}{2}$  reduzierter Quantenwiderstand gemessen, da nur eine Spin-Richtung beiträgt.



**Abb. 10.57:** Stark vereinfachter Bandverlauf in einer CdTe/HgTe-Heterostruktur. Gezeigt sind die Energieniveaus des niedrigsten Leitungsbandniveaus ( $\Gamma_6$ , rot) und des obersten Valenzbandniveaus ( $\Gamma_8$ , blau).  $E_1$  und  $H_1$  sind die niedrigsten Quantentopf-Zustände.

### 10.6.5 Dreidimensionale Topologische Isolatoren

Zweidimensionale TI besitzen ein Paar von eindimensionalen Randkanälen, die sich bei  $k = 0$  kreuzen (siehe Abb. 10.54b, unten). In der Nähe des Kreuzungspunktes verläuft die Dispersion dieser Zustände linear und entspricht deshalb derjenigen, die man in der Quantenfeldtheorie aus der Dirac-Gleichung für masselose, relativistische Fermionen in einer Dimension erhält. Dieses Bild können wir auf dreidimensionale TI verallgemeinern. Hier besteht der Oberflächenzustand aus zweidimensionalen masselosen Dirac-Fermionen, deren Dispersion einen so genannten Dirac-Kegel bildet (siehe Abb. 10.58). Der Kreuzungspunkt – die Spitze des Kegels – ist wie im zweidimensionalen Fall durch die Zeitumkehrinvarianz topologisch geschützt. Wir weisen darauf hin, dass es für eine einzelne 2D-Oberfläche eines 3D-TI immer eine ungerade Zahl von Bändern (beide Spin-Richtungen werden separat gezählt) gibt, welche die Fermi-Energie  $E_F$  schneiden. Im Gegensatz dazu ist es bei einem 2D-Elektronengas mit Rashba-Spin-Bahn-Kopplung (vergleiche hierzu Abb. 12.16) immer eine gerade Zahl.

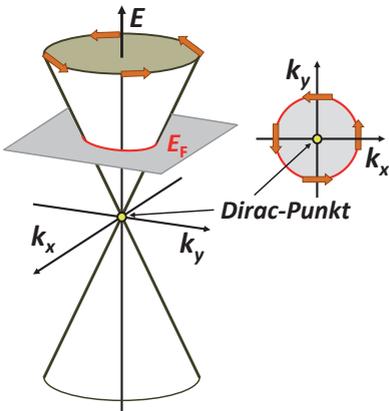


Abb. 10.58: Die Oberfläche eines dreidimensionalen topologischen Isolators erlaubt die Bewegung der Elektronen in einer beliebigen Richtung parallel zur Oberfläche. Die Spin-Richtung ist aber aufgrund der Spin-Bahn-Wechselwirkung immer an die Bewegungsrichtung der Elektronen gekoppelt. Die zweidimensionale Dispersionsrelation ist durch einen Dirac-Kegel gegeben. Rechts ist ein Schnitt durch den Kegel bei der Fermi-Energie  $E_F$  gezeigt. Die Pfeile geben die Spin-Richtung an.

Es wurde theoretisch vorhergesagt, dass die Legierung  $\text{Bi}_{1-x}\text{Sb}_x$  für einen bestimmten  $x$ -Bereich einen TI bildet.<sup>82</sup> In der Tat gibt es bereits experimentelle Evidenz für die Existenz von topologischen Oberflächenzuständen in diesem System.<sup>83</sup> Allerdings sind diese Oberflächenzustände und die zugrundeliegende Physik komplex und noch Gegenstand aktueller Forschung. Weitere viel versprechende Materialsysteme sind  $\text{Bi}_2\text{Te}_3$ ,  $\text{Bi}_2\text{Se}_3$  und  $\text{Sb}_2\text{Te}_3$ .

### 10.6.6 Topologische Supraleiter

Topologische Supraleiter sind eine weitere interessante Klasse von topologischen Quantenmaterialien. Topologische Supraleitung ist insbesondere deshalb interessant, da sie mit Quasiteilchenanregungen assoziiert ist, die *Majorana-Fermionen* sind (Majorana-Fermionen sind Teilchen, die ihre eigenen Anti-Teilchen sind). Es wurde vor kurzem

<sup>82</sup> L. Fu, C. L. Kane, *Topological Insulators with Inversion Symmetry*, Phys. Rev. B 76, 045302 (2007).

<sup>83</sup> D. Hsieh et al., *A Topological Dirac Insulator in a Quantum Spin Hall Phase*, Nature 452, 970 (2008).

vorgeschlagen, topologische Supraleiter dadurch zu erzeugen, dass man einen gewöhnlichen  $s$ -Wellensupraleiter über den Proximity-Effekt an einen topologischen Isolator koppelt und dieses System in ein externes Magnetfeld bringt. Eine Umsetzung und experimentelle Überprüfung der Vorschläge steht aber noch weitgehend aus.

## 10.6.7 Zukunftsperspektiven

Topologische Quantenmaterialien bieten faszinierende Möglichkeiten zur Konstruktion exotischer Quantenzustände und für die Realisierung exotischer Quasiteilchen. Für zweidimensionale QSH-Isolatoren erwarten wir zum Beispiel gebrochen-ganzzahlige Ladungen an den Probenrändern. Für eine Punktladung über der Oberfläche eines dreidimensionalen TI wird vorhergesagt, dass sie nicht nur eine Bildladung sondern auch das Bild eines magnetischen Monopols an der Oberfläche erzeugt. In topologischen magnetischen Isolatoren erwarten wir die Existenz von so genannten Axionen, die in der Teilchenphysik postuliert wurden, um einige Rätsel des Standardmodells der Teilchenphysik zu lösen. Wenn ein Supraleiter in Kontakt mit der Oberfläche eines TI gebracht wird, verhalten sich die Flussschläuche im Supraleiter wie Majorana-Fermionen. Das Austauschen oder Verflechten dieser Flussschläuche führt zu einer nicht-abelschen Teilchenstatistik. Schließlich denkt man bereits an Anwendungen von topologischer Quantenmaterie. Magnetische Bildmonopole könnten zur Datenspeicherung oder Majorana-Fermionen für einen topologischen Quantenrechner benutzt werden.<sup>84</sup>



## Literatur

- G. Bauer, F. Kuchar, H. Heinrich (Hg.), *Two-dimensional Systems: Physics and New Devices*, Springer Series on Solid State Science, Vol. 67, Springer Berlin, Heidelberg (1986).
- G. Bednorz, K. A. Müller, *Possible high- $T_c$  superconductivity in the Ba-La-Cu-O system*, Zeitschrift für Physik **B 64**, 189 (1986).
- B. A. Bernevig, T. L. Hughes, *Topological Insulators and Topological Superconductors*, Princeton University Press (2013).
- T. Chakraborty, P. Pietiläinen, *The Quantum Hall Effect – Fractional and Integral*, Springer Series on Solid-State Science, Vol. 85, Springer Berlin, Heidelberg (1995).
- D. M. Chapin, C. S. Fuller, G. L. Pearson, *A New Silicon  $p$ - $n$  Junction Photocell for Converting Solar Radiation into Electrical Power*, J. Appl. Phys. **25**, 676 (1954).
- C. K. Chiang *et al.*, *Electrical Conductivity in Doped Polyacetylene*, Phys. Rev. Lett. **39**, 1098–1101 (1977).
- M. Contreras *et al.*, *19.9%-efficient ZnO/CdS/CuInGaSe<sub>2</sub> solar cell with 81.2% fill factor*, Progress in Photovoltaics: Research and Applications **16**, 235 (2008).

<sup>84</sup> C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Non-Abelian Anyons and Topological Quantum Computation*, Rev. Mod. Phys. **80**, 1083–1159 (2008).

- J. Czochralski, *Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle*, Zeitschrift für Physikalische Chemie **92**, 219–221 (1918).
- J. H. Davies, *The Physics of Low Dimensional Semiconductors*, Cambridge University Press, Cambridge (1998).
- L. Esaki, *Compositional Superlattices*, in *The Technology and Physics of Molecular Beam Epitaxy*, herausgegeben von E. H. C. Parker, Plenum Press, New York (1985).
- L. Esaki, R. Tsu, *Superlattice and Negative Differential Conductivity in Semiconductors*, IBM J. Res. Develop. **14**, 61 (1970).
- M. Franz, L. Molenkamp (Hg.), *Topological Insulators*, in *Contemporary Concepts of Condensed Matter Series*, Vol. 6, Elsevier, Amsterdam (2013).
- M. Z. Hasan, C. L. Kane, *Topological Insulators*, Rev. Mod. Phys. **82**, 3045–3067 (2010).
- M. Grundmann, *The Physics of Semiconductors: An Introduction Including Nanophysics and Application*, Springer-Verlag, Berlin (2010).
- M. Janssen, O. Viehweger, U. Fastenrath, J. Hajdu, *Introduction to the Theory of the Integer Quantum Hall Effect*, VCH Weinheim (1994).
- D. Kahng, *A Historical Perspective on the Development of MOS Transistors and Related Devices*, IEEE Trans. El. Dev. ED-**23**, 655 (1976).
- D. Kahng, M. M. Atalla, *Silicon-silicon dioxide field induced surface devices*, IRE Solid State Res. Conf., Pittsburgh (1960).
- J.-Y. Kim *et al.*, *Efficient tandem polymer solar cells fabricated by all-solution processing*, Science **317**, 222 (2007).
- R. R. King *et al.*, *40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells*, Applied Physics Letters **90**, 183516 (2007).
- R. B. Laughlin, *Quantized Hall conductivity in two dimensions*, Phys. Rev. B **23**, 5632 (1981).
- R. B. Laughlin, *Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations*, Phys. Rev. Lett. **50**, 1395 (1983).
- R. B. Laughlin, in *McGraw-Hill Book of Science and Technology 1984*, McGraw-Hill, New York (1984).
- G. Margaritondo, P. Perfetti, *The Problem of Heterojunction Band Discontinuities*, in *Heterojunction Band Discontinuities, Physics and Device Applications*, herausgegeben von F. Capasso und G. Margaritondo, North-Holland, Amsterdam (1987).
- A. H. McDonald, *The Quantum Hall Effect: A Perspective*, Kluwer, Boston (1989).
- R. E. Prange, S. M. Girvin, *The Quantum Hall Effect*, Springer Verlag, New York, Berlin (1990).
- C. T. Sah, R. N. Noyce, W. Shockley, *Carrier generation and recombination in p-n junctions and p-n junction characteristics*, Proc. IRE **45**, 1228 (1957).
- M. Schwoerer, Physikalische Blätter **49**, 52 (1994); H. Sixl, H. Schenk, N. Yu, Physikalische Blätter **54**, 225 (1998); Physik Journal **7**, 29–32 (2008).

- W. Shockley, *The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors*, Bell. Syst. Tech. J. **28**, 435 (1949).
- H. L. Störmer, D. C. Tsui, *The Quantized Hall Effect*, Science **220**, 1241–1246 (1983).
- E. C. G. Stückelberg, *Theorie der unelastischen Stöße zwischen Atomen*, Helvetica Physica Acta **5**, 369–422 (1932).
- S. M. Sze, *Semiconductor Devices: Physics and Technology*, Wiley, New York (1985).
- S. M. Sze, K. Ng Kwok, *The Physics of Semiconductor Devices*, John Wiley & Sons, New York (1981).
- D. C. Tsui, H. L. Störmer, A. C. Gossard, *Two-Dimensional Magnetotransport in the Extreme Quantum Limit*, Phys. Rev. Lett. **48**, 1559 (1982).
- Y.Ā. Varshni, *Temperature dependence of the energy gap in semiconductors*, Physica **34**, 149 (1967).
- K. von Klitzing, G. Dorda, M. Pepper, *New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance*, Phys. Rev. Lett. **45**, 494–497 (1980).
- K. von Klitzing, G. Dorda, M. Pepper, *High-magnetic-field transport in a dilute two-dimensional electron gas*, Phys. Rev. B **28**, 4886–4888 (1983).
- P. Y. Yu, M. Cardona, *Fundamentals of Semiconductors*, Springer Verlag, Berlin (1996).
- C. M. Zener *Non-adiabatic Crossing of Energy Levels*, Proceedings of the Royal Society of London A **137** (6), 696–702 (1932).