

В. А. Гольдаде

ВВЕДЕНИЕ В ФИЗИКУ ПОЛУПРОВОДНИКОВ

Допущено

*Министерством образования Республики Беларусь
в качестве учебного пособия для студентов
учреждений высшего образования по специальности
«Физическая электроника»*

Минск
РИВШ
2022

УДК 537.222.22(075.8)
ББК 22.379я73
Г63

Рецензенты:
кафедра физики полупроводников и наноэлектроники
Белорусского государственного университета;
заведующий лабораторией контактно-динамических методов контроля
ГНУ «Институт прикладной физики НАН Беларуси»,
доктор технических наук, доцент *А. П. Крень*

Гольдаде, В. А.
Г63 Введение в физику полупроводников : учебное пособие / В. А. Гольдаде. – Минск : РИВШ, 2022. – 172 с.
ISBN 978-985-586-600-9.

В учебном пособии изложены модельные представления о структуре полупроводников и основы зонной теории твердых тел, а также рассмотрены основные свойства полупроводников, в том числе электропроводность в сильных электрических полях, кинетические явления и механизмы рассеяния носителей заряда, генерация и рекомбинация носителей заряда, контактные явления в полупроводниках. В основу издания положен курс лекций, читаемый автором в течение последних 10 лет на факультете физики и информационных технологий ГГУ им. Франциска Скорины. Учебное пособие разработано в соответствии с Типовой учебной программой «Физика полупроводников и полупроводниковых приборов» и с требованиями образовательного стандарта высшего образования по специальности 1-31 04 03 «Физическая электроника». Оно может быть использовано в качестве методической поддержки при проведении лекционных, практических и лабораторных занятий для закрепления и углубления знаний, а также для самостоятельной подготовки студентов к зачету и экзамену, включая дистанционные формы обучения.

Адресовано студентам учреждений высшего образования, обучающимся по специальности 1-31 04 03 «Физическая электроника».

УДК 537.222.22(075.8)
ББК 22.379я73

ISBN 978-985-586-600-9

© Гольдаде В. А., 2022
© Оформление. ГУО «Республиканский институт высшей школы», 2022

СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

АСМ – атомно-силовая микроскопия
БИС – большая интегральная схема
БЦК – базоцентрированный куб
ВАХ – вольт-амперная характеристика
ГЦК – гранецентрированный куб
ИС – интегральная схема
ОЦК – объемно-центрированный куб
РЭА – радиоэлектронная аппаратура
СБИС – сверхбольшие интегральные схемы
 D_n, D_p – коэффициенты диффузии электронов и дырок
 a, b, c – периоды кристаллической ячейки
 \mathbf{a}, \mathbf{b} – основной и эмиттерный векторы трансляции
 e – заряд электрона
 E – энергия
 E_g – ширина запрещенной зоны
 \mathcal{E} – напряженность электрического поля
 J, j – плотность тока
 \mathbf{k} – волновой вектор
 k – волновое число
 m^* – эффективная масса электрона
 n, p – концентрация электронов и дырок
 R – сопротивление
 \mathbf{r} – координаты электронов
 \mathbf{R} – координаты атомных ядер
 $T_{пл}$ – температура плавления
 U – напряжение
 v – скорость движения носителей заряда
 V_k – контактная разность потенциалов
 γ – удельная электропроводность
 ϵ_0 – диэлектрическая постоянная
 ϵ_r – диэлектрическая проницаемость
 λ – длина волны
 μ – подвижность носителей заряда
 ρ – удельное электрическое сопротивление
 σ – эффективное сечение рассеяния
 τ – время жизни носителя заряда
 ϕ_0 – потенциал, потенциальный барьер
 ϕ_T – тепловой потенциал
 ψ, Φ – волновая функция

ВВЕДЕНИЕ

Физика полупроводников – раздел физики твердого тела, посвященный изучению особенностей физических свойств полупроводников и происходящих в них физических явлений. Предметом изучения являются структурные, электрофизические, оптические свойства полупроводников, многие из которых используются при создании полупроводниковых приборов. Полупроводники относятся к твердым телам, т. е. к веществам, состояние которых характеризуется стабильностью формы и тепловым движением атомов, совершающих малые колебания около положений равновесия. Изучением свойств твердого тела занимается самостоятельная область науки – физика твердого тела, развитие которой привело к появлению фундаментальных свойств материального мира и стимулируется потребностями техники. Около половины физиков мира работают в области физики твердого тела, почти половина всех научных публикаций по физике относится к исследованию твердого тела.

Систематизация данных о свойствах твердого тела началась с XVII в. Был установлен ряд эмпирических законов, описывающих воздействие на твердое тело механических сил, света, электрических и магнитных полей: законы Гука (R. Hooke, английский естествоиспытатель, 1678 г.), Дюлонга и Пти (P. L. Dulong, A. Th. Petit, французские физики, 1819 г.), Ома (G. Ohm, немецкий физик, 1826 г.), Видемана – Франца (G. Wiedeman, R. Franz, немецкие физики, 1853 г.) и др. В первой половине XIX в. были созданы основные концепции теории упругости.

Представление о кристалле как совокупности атомов, упорядоченно расположенных в пространстве и удерживаемых около положения равновесия силами взаимодействия, заложено в трактатах Н. Стено (N. Steno, датский естествоиспытатель, 1669 г.) и И. Ньютона (I. Newton, английский физик, создатель классической механики, 1686 г.), развито в работах Бернулли (D. Bernoulli, швейцарский математик, 1727 г.) и Коши (A. L. Cauchy, французский математик, 1830 г.) и было в окончательном виде сформулировано швейцарским математиком О. Браве (A. Bravais) в 1848 г. В 1890 г. российский кристаллограф Е. С. Федоров и немецкий химик А. Шёнфлис (A. Schoenflies) показали возможность существования 230 пространственных групп симметрии кристаллов, т. е. вариантов упорядоченного расположения частиц в твердом теле.

В 1912 г. немецкие физики лауреат Нобелевской премии М. Лауэ (M. Laue), П. Книппинг (P. Knipping) и В. Фридрих (W. Friedrich) открыли дифракцию рентгеновских лучей на кристаллах, окончательно утвердив

представление о кристаллическом твердом теле как упорядоченной дискретной структуре. В 1913 г. английский физик лауреат Нобелевской премии У. Л. Брэгг (W. L. Bragg) и российский кристаллограф Г. В. Вульф установили соотношение, связывающее период кристаллической решетки и длину волны рентгеновского излучения с направлениями дифракционных максимумов на рентгенограммах. Это положило начало рентгеноструктурному анализу и другим дифракционным методам исследования структуры твердых тел. Впоследствии была обнаружена дифракция на кристаллах электронов, нейтронов и других частиц.

Дисциплина «Физика полупроводников» является базовой для твердотельной и микроэлектроники, а также содержит научные знания, которые необходимы для понимания большинства специальностей электронно-технического профиля.

Эпоха полупроводников насчитывает более 100 лет. Еще в первой половине XIX в. М. Фарадей (M. Faraday, английский физик-экспериментатор и химик) обнаружил, что электропроводность некоторых тел растет по экспоненциальному закону с ростом температуры, в отличие от большинства проводников, у которых электропроводность уменьшается с ростом температуры по линейному закону. Спустя несколько лет А. Беккерель (A. H. Becquerel, французский физик, лауреат Нобелевской премии по физике и один из первооткрывателей радиоактивности) обнаружил, что электропроводность «плохих» проводников, которые изучал М. Фарадей, увеличивается при их освещении, в них появляется ЭДС. Эти «плохие» проводники, имеющие к тому времени загадочные свойства, называли *полупроводниками*.

Исследования полупроводников активизировались в начале XX в. после создания квантовой механики, которая позволила разработать зонную теорию, объясняющую поведение электронов в твердых телах. Согласно этой теории, электроны в кристаллах не могут иметь любую энергию. Разрешенные значения энергии образуют определенные интервалы, которые получили название разрешенных зон. Разрешенные зоны отделены промежутками запрещенных значений энергии.

Начиная с 1930-х гг. интенсивно развивалась полупроводниковая электроника. Ученые исследовали физические процессы в полупроводниках, влияние примесей на эти процессы, термоэлектрические и фотоэлектрические свойства полупроводников, выпрямление переменного тока полупроводниковыми приборами. Была разработана квантовая теория полупроводников, введено понятие подвижности свободных мест кристаллической решетки полупроводника, получивших впоследствии название дырок, создана теория генерации пар «электрон – дырка». Была экспериментально подтверждена теория полупроводников, созданная школой советского академика А. Ф. Иоффе.

Зонная теория твердых тел предложила строгий критерий их разделения на металлы, диэлектрики и полупроводники. Если самая верхняя зона, содержащая электроны при $T = 0$ К, заполнена не полностью, а частично, то такое твердое тело обладает высокими проводящими свойствами. Если самая верхняя зона, содержащая электроны при $T = 0$ К, полностью заполнена, а следующая за ней зона полностью пуста и отделена небольшим промежутком запрещенных энергий ΔE_g , то такое твердое тело относят к полупроводникам, а если ΔE_g достаточно велико, то это диэлектрик ($\Delta E_g \geq 3$ эВ). Самая верхняя зона называется зоной проводимости (C -зона), она отделена от валентной (V -зоны) запрещенной зоной.

При возбуждении полупроводника электроны из V -зоны могут переходить в C -зону и в ней они будут участвовать в переносе электрического тока (это электронная проводимость). В этом случае в V -зоне остаются свободные места (дырки), которые также будут участвовать в переносе электрического тока (дырочная проводимость). Выяснилось, что можно создать полупроводники, у которых основными носителями заряда будут являться электроны, это полупроводники электронного, или n -типа. Также можно создать полупроводники, у которых основными носителями заряда будут дырки (полупроводники p -типа).

Зонная теория твердых тел позволила раскрыть к концу 1930-х гг. природу явлений на контакте полупроводников n - и p -типа, т. е. физическую теорию выпрямления. В результате были созданы p - n -переходы – важнейшие приборы твердотельной электроники, а в 1948 г. – транзисторы на основе n - p - n - и p - n - p -переходов.

С появлением транзистора в 1948 г. (за это изобретение американским ученым Д. Бардину, У. Браттейну и У. Шокли в 1956 г. была присуждена Нобелевская премия по физике) начался новый этап в развитии радиоэлектроники – этап *микроминиатюризации* радиоэлектронной аппаратуры. Применение транзисторов вместо электронных ламп позволило существенно сократить размеры радиокомпонентов, уменьшить массу и объем РЭА, снизить потребление электроэнергии и повысить надежность аппаратуры.

Развитие транзисторной электроники, совершенствование технологии изготовления транзисторов привели к возникновению нового направления в электронике – микроэлектроники. В 1960-х гг. были созданы интегральные схемы, в которых все элементы создаются в едином технологическом процессе, нераздельно связаны и электрически соединены между собой как единое целое, что позволило резко повысить надежность РЭА, еще более уменьшить габариты и массу РЭА, повысить экономичность.

Это стало возможным благодаря разработке *планарной технологии*, которая представляет собой высокопроизводительный метод группового изготовления полупроводниковых приборов и интегральных схем. Основой

планарной технологии является нанесение тонкой диэлектрической пленки на поверхность полупроводникового кристалла или полупроводниковой пленки. Затем методом фотолитографии производится удаление этой пленки с отдельных участков полупроводника. Рисунок, отражающий схему расположения элементов на кристалле, проецируется на него световыми или электронными лучами. После этого через незащищенные участки полупроводника вводится специальная примесь (лигатура), которая переводит полупроводник в *n*- или *p*-тип. В результате этих операций образуются области с *p-n*-переходами.

Интегральные схемы в середине 1960-х гг. содержали до 100 элементов на полупроводниковом кристалле при размере элементов около 100 мкм. В начале 1970-х гг. появились большие интегральные схемы, имевшие на кристалле от 100 до 104 элементов при размере элементов от 3 до 100 мкм. В конце 1970-х гг. созданы сверхбольшие интегральные схемы со 104–106 элементами на кристалле при размере элементов от 1 до 3 мкм. Дальнейшее развитие микроэлектроники привело к освоению субмикронных размеров элементов микросхем. Исследования показали, что пределом уменьшения размеров элементов является значение 0,2 мкм. Однако достижение такого размера связано с преодолением определенных технологических трудностей. Иначе говоря, существуют физические пределы развития интегральной микроэлектроники.

Параллельно с интегральной микроэлектроникой в 1980-е гг. развивалась функциональная электроника, позволяющая реализовать определенную функцию аппаратуры без применения стандартных базовых элементов (диодов, резисторов, транзисторов и т. д.), базируясь непосредственно на физических явлениях в твердом теле. В функциональной электронике используются такие механизмы, как оптические явления (оптоэлектроника), взаимодействие потока электронов с акустическими волнами в твердом теле (акустоэлектроника) и др.

Вступление в третье тысячелетие электроника отметила зарождением нового направления – нанoeлектроники. В начале 1990-х гг. были созданы атомно-силовые микроскопы, позволяющие не только наблюдать атомы, но и манипулировать ими.

Нанотехнологии позволяют, последовательно размещая нужные атомы и атомные структуры в четком порядке и в точно определенном месте, конструировать такие технологические диковинки, которым пока еще и название не придумали. Сейчас разработки в области нанотехнологий продолжают развиваться во многих странах мира.

Глава 1

МОДЕЛЬНЫЕ ПРЕДСТАВЛЕНИЯ О СТРУКТУРЕ ПОЛУПРОВОДНИКОВ

По своему структурному состоянию полупроводники относятся к твердым телам. Твердое тело – агрегатное состояние вещества, которое характеризуется стабильностью формы и тепловым движением атомов, совершающих малые колебания около положений равновесия. Изучением свойств твердого тела занимается самостоятельная область науки – физика твердого тела, развитие которой привело к появлению фундаментальных свойств материального мира и стимулируется потребностями техники.

В настоящее время методы и теория твердого тела, развитые для описания свойств и структуры монокристаллов, широко применяются для получения и исследования новых материалов: композитов и наноструктур, квазикристаллов и аморфных тел. Физика твердого тела служит основой для изучения явлений высокотемпературной сверхпроводимости, гигантского магнетосопротивления и многих других перспективных современных наукоемких технологий.

1.1. Типы химических связей

Наименьшей устойчивой частицей вещества, обладающей его основными химическими и физическими свойствами и состоящей из атомов одного или разных химических элементов, является молекула. Атомы объединяются в молекулы с помощью химических связей, в образовании которых принимают участие в основном внешние (валентные) электроны.

Разделение химически активных элементов на металлы и неметаллы позволяет подразделить химические связи на три основных типа: металлическую, ковалентную и ионную. Связь между сильно электроположительными атомами металлов и электроотрицательными атомами неметаллов трактуют как ионную связь. Поскольку она существует между противоположно заряженными ионами, ее называют *гетерополярной*. Металлическую и ковалентную связи относят к *гомеополярным*. Металлическая связь реализуется между металлом и неметаллом, ковалентная – между неметаллом и неметаллом. Названные типы связей являются предельными случаями химического взаимодействия.

Ионная связь возникает из-за притяжения противоположно заряженных ионов. Наиболее ярким примером конденсированных тел, в которых этот тип связи доминирует, являются щелочно-галогидные кристаллы: NaCl,

LiF и др. Образование ионов в таких соединениях энергетически выгодно, поскольку потенциал ионизации атома щелочного металла превосходит сродство к электрону атома галогена. Энергия электростатического (кулоновского) притяжения $E_{\text{пр}}$ для пары ионов с зарядом $\pm e$, находящихся на расстоянии r , составляет $E_{\text{пр}} = -e^2 / r$. В кристалле, где ионы образуют упорядоченную решетку, полная энергия кулоновского взаимодействия (отнесенная к паре ионов) становится равной

$$E_{\text{пр}} = -\alpha \frac{e^2}{r}, \quad (1.1)$$

где α – постоянная Маделунга (E. Madelung – немецкий физик-теоретик), зависящая только от структуры кристалла. Для NaCl $\alpha = 1,748$.

В формировании стабильной структуры твердых тел существенную роль играет энергия отталкивания $E_{\text{от}}$, резко возрастающая при уменьшении расстояния между атомами и ионами. Энергия $E_{\text{от}}$ имеет квантовомеханическую природу и обусловлена прежде всего отталкиванием ядер взаимодействующих атомов (ионов), экранированных окружающими их электронами. Эта энергия, отнесенная к паре ионов, равна

$$E_{\text{от}} = \frac{\beta}{r^m}, \quad (1.2)$$

где β – постоянная; $6 < m < 10$; r – расстояние между частицами.

Сумма энергий $E_{\text{пр}}$ и $E_{\text{от}}$ дает полную энергию взаимодействия пары ионов. Как видно из рис. 1.1, график $E(r)$ имеет минимум при некотором расстоянии r_0 между ионами, которое определяет равновесный параметр структуры твердого тела, а соответствующая энергия $E(r_0)$ является энергией связи ионов в нем.

Ковалентная связь образуется путем обменного электронного взаимодействия между атомами, которое реализуется в виде сгущений электронной плотности при обобществлении внешних электронов у атомов с перекрывающимися орбитами. В отличие от ионной связи, ковалентно могут быть связаны и одинаковые, и разные атомы. Ковалентная связь имеет место: в гомеоплярных молекулах H_2 , Cl_2 , I_2 и др.; в веществах, атомы которых расположены в периодической системе элементов в отдалении от инертных газов (кристаллы алмаза, Si, Ge, Te); в молекулах органических соединений; в полупроводниковых кристаллах типа $\text{A}^{\text{III}}\text{B}^{\text{V}}$

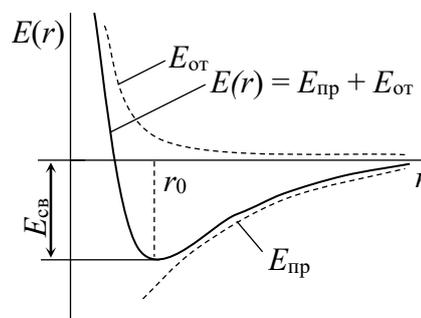


Рис. 1.1. Зависимость энергии взаимодействия двух ионов от расстояния между ними

(A^{III} – элемент третьей группы в таблице Менделеева, B^{V} – элемент пятой группы).

Расчет энергии ковалентных связей затруднен необходимостью привлечения представлений квантовой механики для описания электронной плотности. Это вызвано тем, что движение электронов в изолированных и взаимодействующих атомах существенно отличается. Ковалентная связь обычно образуется парой электронов с антипараллельными спинами, по одному от каждого из соединяющихся атомов.

В молекуле метана CH_4 (рис. 1.2) атом углерода с электронной конфигурацией $1s^2 2s^2 2p^2$ окружен четырьмя атомами водорода с электронной конфигурацией $1s^1$. Орбиты внешних гибридных $2s-2p$ электронов углерода и электронов атомов водорода перекрываются, и эти электроны обобществляются. В органической химии это трактуется как образование «связей» между атомами С и Н (рис. 1.2, а).

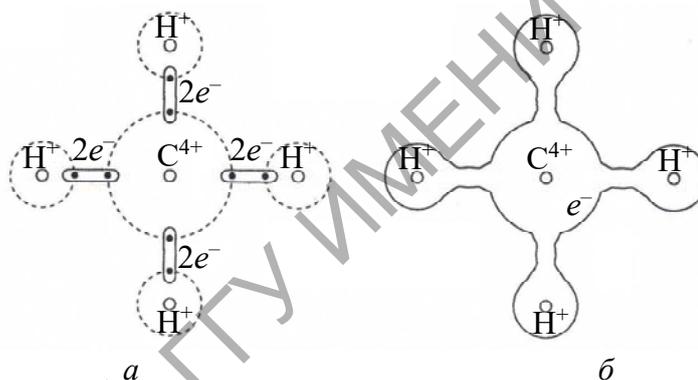


Рис. 1.2. Ковалентная связь в молекуле метана:
а – модель образования связей в результате обобществления электронов;
б – сгущения электронной плотности

Однако согласно принципам квантовой механики четыре гибридных $2s^2$ и $2p^2$ электрона углерода индивидуально не различимы, и можно говорить лишь об образовании сгущений электронной плотности между атомами водорода и углерода (рис. 1.2, б). С позиций квантовой теории химической связи образование таких локализованных сгущений энергетически выгодно, что подтверждается образованием устойчивых молекул.

Важными свойствами ковалентной связи являются насыщенность и направленность. *Насыщенность* – свойство атома образовывать ковалентную связь лишь с определенным числом своих соседей. Так, атом водорода может устанавливать связь только с одним соседом. В алмазе или графите несколько валентных электронов являются общими для атома углерода и его соседей, и потому невозможно считать какую-либо группу атомов

химически насыщенной (рис. 1.3). С этой точки зрения идеальный кристалл алмаза представляет собой огромную молекулу.

Направленность ковалентной связи состоит в том, что связь образуется в тех направлениях, по которым локализована максимальная электронная плотность. Вследствие направленности связей ковалентные кристаллы обладают высокой твердостью и хрупкостью.

Ковалентные связи между парами атомов можно рассматривать как короткодействующие силы. В молекуле H_2 энергия химической связи зависит только от расстояния между атомами, а в алмазе, где атом углерода находится в тетраэдрическом окружении аналогичных атомов (рис. 1.3, б), для расчета энергии надо знать еще и значения валентных углов (примерно 109 градусов).

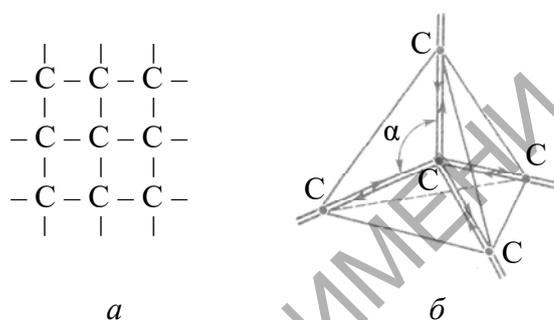


Рис. 1.3. Схемы связей (а) и тетраэдрического расположения атомов углерода (б) в структуре алмаза. Стрелками обозначены направления обмена электронами

В конденсированных телах ковалентные связи в чистом виде реализуются редко и имеют частично ионную природу. Веществ с единственно ионными связями также чрезвычайно мало. Степень «ионности» связи в химических соединениях атомов A и B можно оценить, исходя из электроотрицательности X_A и X_B . Относительная плотность связи равна

$$\xi = 1 - \exp\left[-0,25(X_A - X_B)^2\right].$$

Если $\xi = 1$, т. е. ионность составляет 100 %, то связь между атомами чисто ионная, если $\xi = 0$, то связь чисто ковалентная. В таблице 1.1 приведены приблизительные значения ионности связи для некоторых соединений. Видно, что $NaCl$ можно считать ионным кристаллом, а SiC и $GaAs$ – преимущественно ковалентными. Атомы с почти заполненными оболочками (F, Cl, Br, I) обнаруживают тенденцию к ионной связи, тогда как атомы III, IV и V групп периодической системы элементов – к ковалентной связи. В таблице 1.2 приведены значения энергии ковалентной связи для некоторых пар атомов.

Таблица 1.1

Степень ионности связи в кристаллах бинарных соединений

Кристалл	Степень ионной связи	Кристалл	Степень ионной связи
Si	0,00	GaSb	0,26
SiC	0,18	GaAs	0,32
Ge	0,00	CuBr	0,74
ZnTe	0,61	CuCl	0,75
ZnO	0,62	AgI	0,77
ZnS	0,62	AgBr	0,85
ZnSe	0,63	AgCl	0,86
CdTe	0,67	MgS	0,79
CdS	0,69	MgSe	0,79
CdSe	0,70	MgO	0,84
CdO	0,79	LiF	0,92
InSb	0,32	NaCl	0,94
InAs	0,35	RbF	0,96
InP	0,44		

Таблица 1.2

Энергии ковалентной связи

Связь	Энергия связи	
	эВ	кДж/моль
O–O	1,4	138
Te–Te	1,4	138
Ge–Ge	1,6	159
Si–Si	1,8	176
P–P	2,2	213
Cl–Cl	2,5	243
C–C	3,6	347
H–H	4,5	435

Энергия ионной связи в полупроводниках $A^{III}B^V$ возрастает при переходе к более тяжелым соединениям – от GaAs к InSb. С другой стороны, в кристаллах, которые называют «ионными», в частности, в состав которых входят ионы элементов VI группы (O^{2-} , S^{2-} и др.), всегда наряду с ионной реализуется ковалентная связь, т. е. некоторая доля электронной плотности «обобществлена» соседними атомами. Даже в галогенидах щелочных

металлов вклад ковалентной связи хотя и мал, но не равен нулю. Таким образом, твердых тел, в которых связь была бы чисто ионной, не существует. В отличие от ионной, только ковалентная связь имеет место в диэлектрических монокристаллах, например в алмазе.

Металлическая связь – разновидность гомеоплярной химической связи, характерна для металлов и сплавов. Металлическая связь так же, как и ковалентная, осуществляется обобществленными электронами, но принадлежащими не соседним атомам, а всему объему конденсированного тела. Внешние валентные электроны в атомах металлов связаны с ядрами довольно слабо. В металле они коллективизированы и образуют «*электронный газ*» (или «*электронную жидкость*»), заполняющий межионное пространство. Положительно заряженные ионы «стягиваются» отрицательно заряженным электронным газом. Таким образом, металлическая связь возникает вследствие взаимодействия положительных ионов с электронным газом.

С уменьшением расстояния между ионами увеличивается плотность электронного газа, вследствие чего растет сила, стягивающая ионы. С другой стороны, по мере сближения ионов увеличивается сила отталкивания между ними. Расстояние между ионами, при котором силы притяжения уравновешены силами отталкивания, характеризует устойчивое состояние кристаллической решетки. Это принципиально верное представление недостаточно точно. В действительности металлическая связь имеет более сложную природу, и методы ее расчета основаны на зонной теории твердого тела. В наиболее простом варианте энергия металлической связи определяется двумя факторами. С одной стороны, при сближении металлических атомов волновые функции электронов перекрываются, и электрон получает возможность перемещаться в более широкой (чем в изолированном атоме) зоне, где он имеет более низкую потенциальную энергию. С другой стороны, при «сжатии» электронного газа возрастает средняя кинетическая энергия электронов. Равновесная плотность электронов соответствует минимуму полной энергии взаимодействия двух атомов (ионов). Расстояние между ними, при котором это условие реализуется, можно считать атомным (ионным) радиусом металла.

Наличие свободных электронов в металле обуславливает *ненаправленный* и *ненасыщенный* характер металлической связи. Чистая ненаправленная металлическая связь наблюдается у одновалентных металлов (Na, Li и др.) с кубической, плотно упакованной структурой. У металлов с несколькими электронами на внешней оболочке не все электроны делокализуются. Поэтому определенную составляющую в связь вносит ковалентное взаимодействие ионов. Металлическая связь определяет электрические и тепловые

свойства металлов, обуславливая высокие электро- и теплопроводность. Металлическая связь является довольно сильной, о чем свидетельствует прочность и высокая температура плавления многих (хотя и не всех) металлов.

Помимо упомянутых химических связей – ионной, ковалентной и металлической – атомы связываются посредством ван-дер-ваальсовой и водородной связей.

Ван-дер-ваальсовая связь имеет место в кристаллах, состоящих из молекул с насыщенными связями (O_2 , N_2 , Cl_2 , Br_2 и др.) или из атомов инертных элементов. Название связи происходит от имени нидерландского физика Я. Д. ван дер Ваальса (J. D. van der Waals), который для объяснения свойств газов и жидкостей предложил в 1873 г. уравнение состояния, учитывающее межмолекулярные взаимодействия. Механизм этого вида связи состоит в мгновенном возникновении электрических диполей в молекуле вследствие флуктуаций межатомных расстояний (отсюда еще одно название – *флуктуационная связь*). Квантовомеханический расчет энергии взаимодействия диполей показал, что она обратно пропорциональна шестой степени расстояния между атомами. Ван-дер-ваальсовую связь называют также *молекулярной*, поскольку она связывает частицы в молекулярных кристаллах. Энергия этой связи очень мала, о чем свидетельствуют чрезвычайно низкие температуры плавления веществ, в которых они реализуются.

Водородная связь осуществляется в результате кулоновского взаимодействия иона H^+ (протона) с «дополнительным» отрицательным ионом азота, кислорода, хлора и др. Она не образуется с ионами, с которыми водород вступает в ковалентную связь. Энергия водородной связи приблизительно равна 0,1 эВ, т. е. на порядок меньше энергии ковалентной связи.

Образование водородной связи можно объяснить следующим образом. Электроотрицательный атом, «оттягивая» электрон от ядра атома водорода, приобретает положительный заряд, равный заряду протона. Этот протон и осуществляет связь между атомами. Из-за малого размера ядра атома водорода водородная связь осуществляется только между двумя атомами.

Типичным примером водородной связи является связь между молекулами воды (рис. 1.4). Атом кислорода одной молекулы воды и атом водорода другой молекулы образуют диполь с зарядами $+q$ и $-q$ на полюсах. Притяжение между этими зарядами приводит к возникновению водородной связи. Длина водородной связи ($r_{OH} = 2,76 \text{ \AA}$) значительно больше длины ковалентной связи $H-O$ в молекуле воды ($r_0 = 0,96 \text{ \AA}$).

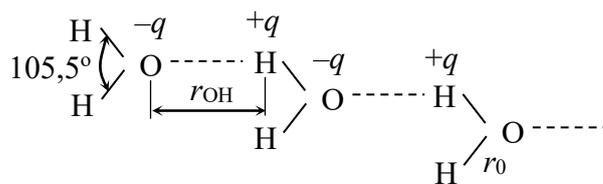


Рис. 1.4. Водородная связь (пунктир) между молекулами воды

Наличием водородных связей объясняется своеобразие структуры и физических свойств воды и водных растворов. Кристаллическая структура льда представляет собой тетраэдрическую сетку водородных связей, в которой имеется большое количество пустот. При плавлении льда эти пустоты частично заполняются молекулами воды, и поэтому плотность воды выше плотности льда.

Водородные связи могут быть не только межмолекулярными, но и внутримолекулярными. Внутримолекулярные водородные связи: стабилизируют глобулярную структуру атомов белков; влияют на свойства древесины и бумаги, построенных из волокон целлюлозы; ответственны за уникальную структуру молекул нуклеиновых кислот, содержащихся во всех живых организмах.

На рис. 1.5 приведены схемы основных видов межатомного взаимодействия в конденсированных телах.

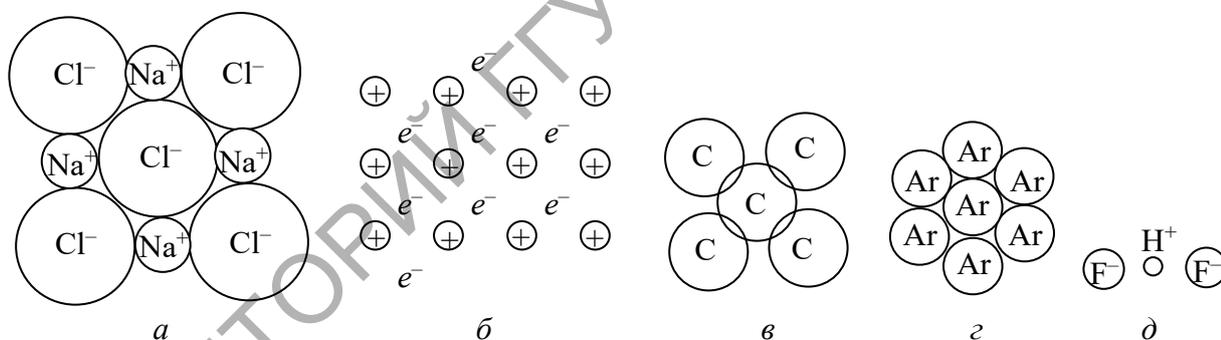


Рис. 1.5. Основные виды связи между атомами (ионами) в конденсированных телах: *a* – ионная связь в кристалле хлористого натрия; *b* – металлическая связь; *в* – ковалентная связь между атомами алмаза; *г* – силы ван-дер-ваальса между атомами аргона; *д* – водородная связь в молекуле HF₂

Химические связи классифицируют по ряду других признаков. Важнейшим из них является энергетический. Энергию связи многоатомных молекул определяют как энергию диссоциации. Для многоатомных молекул эта величина является условной и отвечает энергии такого процесса, при котором данная химическая связь исчезает, а все остальные остаются без изменения. По энергетическому признаку связи подразделяют на сильные (≥ 500 кДж/моль), слабые (от 100 до 15 кДж/моль) и ван-дер-ваальсовы

(≤ 5 кДж/моль). Параметрами связей являются: равновесные межъядерные расстояния (длины связей); валентные углы; частоты колебаний, отнесенные к паре связанных атомов; поляризуемость; электрический дипольный и магнитный моменты химической связи. Значения параметров получают экспериментально, например, оценивая распределение электронной плотности рентгенографическими методами.

1.2. Кристаллическая решетка

Существование кристаллической решетки в твердом теле объясняется тем, что равновесие сил притяжения и отталкивания между атомами, соответствующее минимуму его потенциальной энергии, достигается при условии трехмерной периодичности.

Кристалл есть однородное, но в то же время анизотропное тело. *Однородность* выражается в том, что любые элементарные объемы кристаллического вещества, одинаково ориентированные в пространстве, абсолютно одинаковы по всем своим свойствам: имеют один и тот же цвет, массу, твердость и т. д. *Анизотропия* выражается в том, что физические свойства кристаллов (твердость, прочность, теплопроводность, электропроводность, скорость распространения света) неодинаковы по разным направлениям.

Для описания кристаллической решетки нужно знать координаты центров атомов (или *узлы решетки*) в элементарной ячейке кристалла (рис. 1.6). *Элементарная ячейка* – часть атомной структуры кристалла, параллельными переносами которой (трансляциями) в трех измерениях можно построить всю кристаллическую решетку. Длины ребер элементарной ячейки a , b , c называют постоянными, или *периодами* кристаллической решетки, а в векторной форме – *векторами трансляции*. Элементарная ячейка имеет форму параллелепипеда, ее выбор определяется симметрией кристаллов.

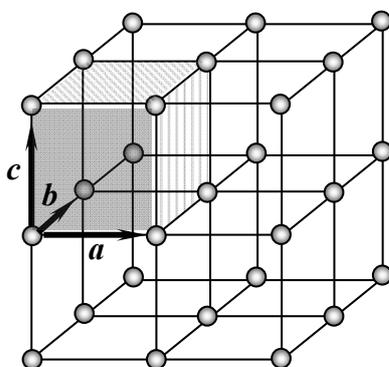


Рис. 1.6. Элементарная ячейка кристаллической решетки (заштрихована); кружками обозначены узлы решетки

В физике твердого тела наиболее часто в качестве элементарной ячейки используют так называемую *ячейку Вигнера – Зейтца* (рис. 1.7, а). Ю. Вигнер (Eu. Wigner) – американский физик, лауреат Нобелевской премии, Ф. Зейтц (F. Seitz) – его ученик, физик-теоретик. Для построения ячейки узел решетки соединяют линиями со всеми соседними узлами и через середины полученных отрезков перпендикулярно к ним проводят линии (в случае объемной ячейки – плоскости). Многоугольник, содержащий выбранный узел и ограниченный этими линиями, и есть элементарная ячейка. Все точки внутри многоугольника лежат ближе к центру ячейки, чем к любой другой трансляционно эквивалентной центру точке кристалла. Пример ячейки Вигнера – Зейтца для кубического объемноцентрированного кристалла приведен на рис. 1.7, б.

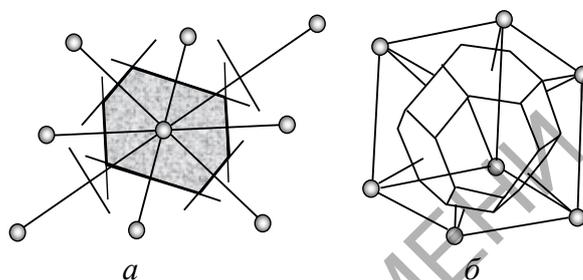


Рис. 1.7. Ячейки Вигнера – Зейтца: а – плоская (заштрихованный шестиугольник); б – для объемноцентрированного куба (усеченный октаэдр)

Ячейка Вигнера – Зейтца полностью определяет трансляционную структуру кристалла, и при смещении на векторы трансляции решетки заполняет собой весь кристалл. Ячейка Вигнера – Зейтца обратной решетки кристалла представляет собой первую зону Бриллюэна (см. п. 2.2).

В элементарной ячейке кристалла может размещаться от одного атома (химические элементы) до сотен (химические соединения) и миллионов атомов (белки, вирусы). Соответственно периоды кристаллических решеток составляют от нескольких десятых до 10^2 нм (таблица 1.3).

Таблица 1.3

Параметры элементарных ячеек некоторых кристаллов

Тип кристаллов	Периоды кристаллической решетки, нм	Число атомов в элементарной ячейке
Химические элементы, простейшие соединения	0,5–1,0	1–10
Неорганические и простые молекулярные соединения	1,0–2,0	до сотен
Сложные органические соединения	2,0–4,0	до тысяч
Белки	до 10–30	10^3 – 10^5
Вирусы	до 200	10^6 – 10^9

По признаку точечной симметрии элементарной ячейки все кристаллы делят на семь групп, получивших название кристаллографических *сингоний*, которые различаются соотношением между длинами ребер a , b , c ячейки и углами α , β , γ между ними (рис. 1.8):

- 1) кубическая ($a = b = c$, $\alpha = \beta = \gamma = 90^\circ$);
- 2) тетрагональная ($a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$);
- 3) гексагональная ($a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$);
- 4) тригональная ($a = b = c$, $\alpha = \beta = \gamma \neq 90^\circ$);
- 5) ромбическая ($a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$);
- 6) моноклинная ($a \neq b \neq c$, $\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$);
- 7) триклинная ($a \neq b \neq c$, $\alpha \neq \beta \neq \gamma \neq 90^\circ$).

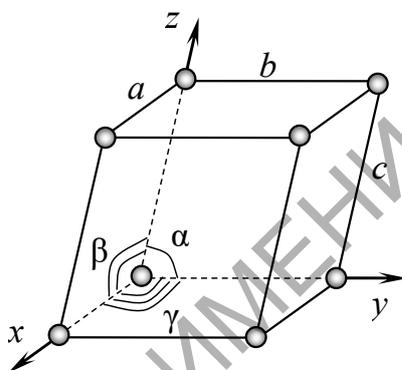


Рис. 1.8. Конфигурация элементарной ячейки с параметрами a , b , c , α , β , γ ; x , y , z – кристаллографическая система координат

Каждую кристаллическую структуру можно охарактеризовать определенным набором трансляций. Разным соотношениям значений векторов \mathbf{a} , \mathbf{b} , \mathbf{c} и взаимной ориентации основных трансляций соответствуют решетки, отличающиеся друг от друга по симметрии.

В 1848 г. О. Браве (A. Bravais – французский физик) математическим путем доказал, что существует 14 типов трехмерных геометрических фигур, характеризующих все возможные типы трансляционной симметрии кристаллической решетки (рис. 1.9).

Различают *примитивные* решетки Браве, в которых узлы расположены только в вершинах элементарных параллелепипедов, *базоцентрированные* (узлы в вершинах и в центрах двух противоположных граней), *объемноцентрированные* (узлы в вершинах и в центре параллелепипедов) и *гранецентрированные* (узлы в вершинах и в центрах всех граней). Понятие «решетки Браве» используют при описании атомной структуры кристаллов. В простейших случаях (например, у металлов) структура кристалла описывается одной решеткой Браве. Сложную кристаллическую структуру можно описывать несколькими решетками Браве, «вдвинутыми» одна

в другую. Полное описание кристаллической решетки содержит характеристику пространственной группы симметрии, параметры элементарной ячейки, координаты атомов в ячейке. В этом случае понятие «кристаллическая решетка» эквивалентно понятию атомной структуры кристалла.

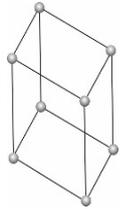
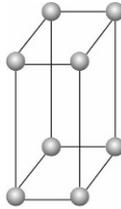
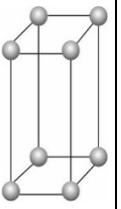
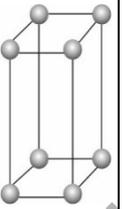
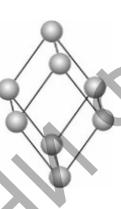
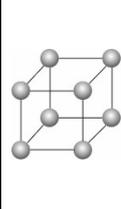
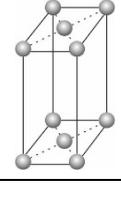
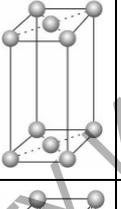
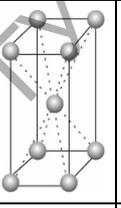
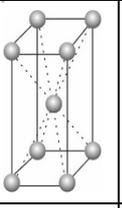
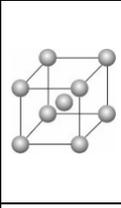
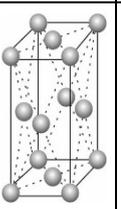
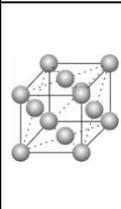
Сингония Тип решетки	Триклинная $a \neq b \neq c,$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$	Моноклиная $a \neq b \neq c,$ $\alpha = \gamma = 90^\circ,$ $\beta \neq 90^\circ$	Ромбическая $a \neq b \neq c,$ $\alpha = \beta = \gamma = 90^\circ$	Тетрагональная $a = b \neq c,$ $\alpha = \beta = \gamma = 90^\circ$	Тригональная (ромбоэдрическая) $a = b = c,$ $\alpha = \beta = \gamma \neq 90^\circ$	Гексагональная $a = b \neq c,$ $\alpha = \beta = 90^\circ,$ $\gamma = 120^\circ$	Кубическая $a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
Примитивный							
Базоцентри- рованный							
Объемно- центриро- ванный							
Гранцентри- рованный							

Рис. 1.9. Трансляционные решетки Браве

Структура кристаллических твердых тел отличается от идеализированной схемы, описываемой понятием «кристаллическая решетка». Идеализацией является представление о дискретности кристаллической решетки. В действительности электронные оболочки атомов, составляющих кристаллическую решетку, перекрываются, образуя непрерывное периодическое распределение заряда с максимумами около дискретно расположенных ядер. Идеализацией является также неподвижность атомов. Атомы и молекулы кристаллической решетки колеблются около положений равновесия,

а параметры колебаний зависят от взаимодействия атомов. С повышением температуры амплитуда колебаний увеличивается, что в конечном итоге приводит к разрушению кристаллической решетки и переходу твердого вещества в жидкое состояние. Кроме того, в кристаллах всегда имеются дефекты (см. рис. 1.6), искажающие кристаллическую решетку.

1.3. Кристаллографические обозначения

В кристаллографии используют несколько методов описания симметрии кристаллов. Хотя большая часть этих методов создана еще до появления современной техники изучения атомов, они не утратили своей практической ценности.

Одна из самых важных задач кристаллографии – определение направления атомных рядов. Простой метод ее решения основан на применении векторных обозначений.

Пусть проекции вектора на оси x , y , z куба представляют собой числа x_0 , y_0 , z_0 . Эти проекции удобно измерять в единицах длины ребра куба a . Находят число r , для которого величины x_0/r , y_0/r , z_0/r составляют ряд наименьших целых чисел, пропорциональных проекциям. Эти величины называются *индексами направлений*, или индексами Вейса (P. Weiss – французский физик). Индексы направлений обычно заключают в квадратные скобки $[uvw]$. Отрицательные значения проекций на любую координатную ось имеют отрицательные индексы, которые обозначают чертой над соответствующим индексом. Направления, одинаково симметричные относительно всех трех координатных осей, образуют *группу*. Обобщенные индексы симметричных направлений заключают в угловые скобки $\langle uvw \rangle$.

В качестве примера рассмотрим вектор, проекции которого на оси координат равны: $x_0 = 2a$, $y_0 = a$, $z_0 = 1,5a$. Разделим эти числа на $r = 0,5a$ и получим три целых числа: 4, 2 и 3. Они относятся друг к другу так же, как исходные размеры проекций, и в то же время представляют собой последовательность из наименьших целых чисел, обладающих этим свойством. То есть полученные числа – это индексы направлений исходного вектора $[423]$. Ось $+x$ в прямоугольной системе координат имеет индексы $[100]$, ось $-x$ – индексы $[\bar{1}00]$. Диагональ грани xu обозначают индексами $[110]$, а диагональ куба в положительном квадранте – индексами $[111]$ (рис. 1.10). Восемь диагоналей куба $[111]$, $[\bar{1}\bar{1}\bar{1}]$, $[\bar{1}11]$, $[1\bar{1}1]$, $[11\bar{1}]$, $[\bar{1}\bar{1}1]$, $[\bar{1}1\bar{1}]$, $[1\bar{1}\bar{1}]$ образуют группу $\langle 111 \rangle$, все они различаются положением чисел 1 и $\bar{1}$ в индексах.

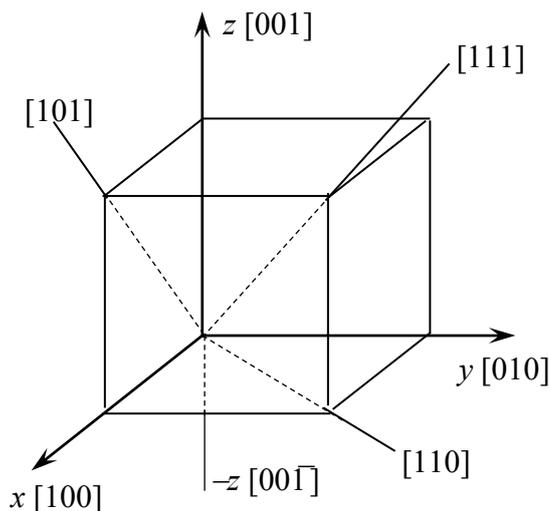


Рис. 1.10. Индексы направлений в прямоугольной системе координат

В отличие от индексов направлений набор параллельных атомных плоскостей в кристалле характеризуют с помощью трех чисел, которые связаны с длиной отрезков, отсекаемых какой-либо плоскостью на осях координат: x_1 , y_1 , z_1 . Существует такое число s , которое при умножении на *обратные* значения длин этих отрезков дает ряд наименьших целых чисел, относящихся друг к другу как $1/x_1$, $1/y_1$ и $1/z_1$. Числа $h = s/x_1$, $k = s/y_1$ и $l = s/z_1$ называют *индексами Миллера* (W. Miller – английский кристаллограф) для плоскости и записывают в круглых скобках (hkl) .

При обозначении индексов плоскостей действуют те же общие правила, что и для индексов направлений. Отрицательные индексы соответствуют плоскостям, пересекающим отрицательные направления координатных осей. Так, плоскость с отрицательным x записывается как $(\bar{h}kl)$. Плоскости кристалла, эквивалентные по признаку симметрии, образуют группу, в которой индексы заключаются в фигурные скобки: $\{hkl\}$. При обозначении кубических решеток эта схема имеет следующие особенности (рис. 1.11):

1) индексы параллельных плоскостей одинаковы (hkl) ; аналогичное правило справедливо и для индексов направлений, поэтому у параллельных атомных рядов индексы одинаковы $[uvw]$;

2) плоскость, параллельная одной из осей координат, имеет индекс 0 для этого направления;

3) угол θ между двумя кристаллографическими направлениями $[u_1v_1w_1]$ и $[u_2v_2w_2]$ рассчитывают по формуле:

$$\cos \theta = \frac{u_1u_2 + v_1v_2 + w_1w_2}{\left[(u_1^2 + v_1^2 + w_1^2)(u_2^2 + v_2^2 + w_2^2) \right]^{1/2}}; \quad (1.3)$$

4) нормаль к плоскости с индексами (hkl) лежит в направлении $[hkl]$, это позволяет применять формулу (1.3) для вычисления углов между плоскостями;

5) расстояние d между соседними плоскостями с индексами (hkl) можно вычислить в единицах длины ребра куба по формуле:

$$d = \frac{a}{(h^2 + k^2 + l^2)^{1/2}}. \quad (1.4)$$

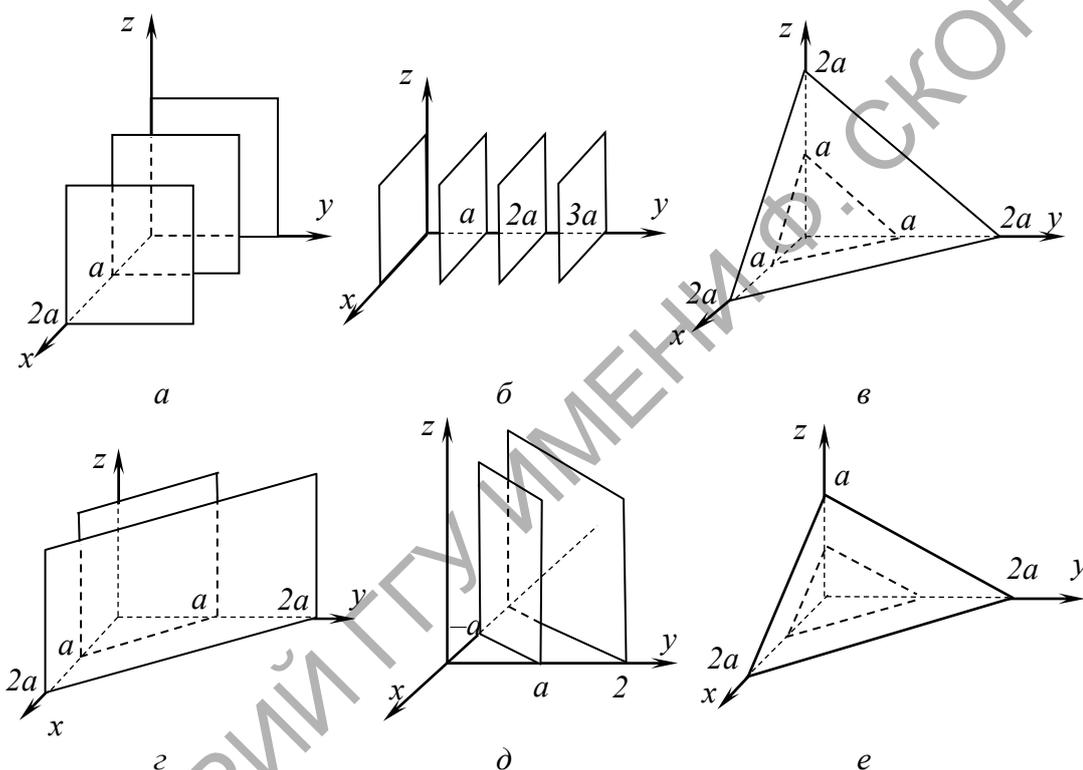


Рис. 1.11. Примеры обозначения кристаллографических плоскостей в кубической структуре: $a - (100)$; $б - (010)$; $в - (111)$; $г - (110)$; $д - (\bar{1}10)$; $e - (112)$

Из формулы (1.4) следует, что для кубической ячейки с ребром a расстояния между плоскостями с большими индексами Миллера меньше расстояний между плоскостями, имеющими малые индексы. В атомных плоскостях с большими индексами более высокая плотность расположения атомов (количество атомов на единице площади). При описании физических и химических свойств твердого тела в большинстве случаев наиболее важны плоскости с небольшими индексами. Плоскости с индексами более 3 анализируют очень редко. Для описания основных свойств кристалла представляют интерес лишь примерно 15 групп плоскостей.

1.4. Обратная решетка

В физике твердого тела при анализе явлений, связанных с периодическим положением атомных частиц, используют понятие «обратная решетка». Она реально не существует в кристалле, а представляет собой удобную абстракцию, позволяющую математически просто и точно описывать условия, в которых протекает то или иное явление в кристаллическом теле.

Обратная решетка – это периодическая решетка в абстрактном пространстве, элементарные векторы трансляции \mathbf{b}_i которой связаны с основными векторами трансляции \mathbf{a}_j исходной атомной решетки условиями

$$\mathbf{b}_i \mathbf{a}_j = \begin{cases} 2\pi, & i = j \\ 0, & i \neq j \end{cases}. \quad (1.5)$$

Координаты узлов обратной решетки задают вектором трансляции $\mathbf{G} = \sum_i L_i \mathbf{b}_i$, где L_i – произвольные целые числа, $i = 1, 2, 3$ для трехмерной решетки или $i = 1, 2$ – для двухмерной. Размерность обратной решетки совпадает с размерностью прямой решетки. Так, для трехмерной прямой решетки обратная решетка является также трехмерной с элементарными векторами трансляции, равными в соответствии с условиями

$$\begin{aligned} \mathbf{b}_1 &= 2\pi[\mathbf{a}_2 \mathbf{a}_3] / V; \\ \mathbf{b}_2 &= 2\pi[\mathbf{a}_3 \mathbf{a}_1] / V; \\ \mathbf{b}_3 &= 2\pi[\mathbf{a}_1 \mathbf{a}_2] / V. \end{aligned} \quad (1.6)$$

Здесь $V = (\mathbf{a}_1[\mathbf{a}_2 \mathbf{a}_3])$ – объем элементарной ячейки прямой решетки. Объем элементарной ячейки обратной решетки равен $(2\pi)^3 / V$.

В обратной решетке вектор $\mathbf{G}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ перпендикулярен плоскости с кристаллографическими индексами hkl .

Симметрия обратной решетки полностью определяется симметрией решетки Браве. Например, обратная решетка для простой кубической решетки – простая кубическая, для гранецентрированной кубической – объемноцентрированная кубическая (и наоборот) и т. д.

Обратная решетка является важным математическим образом, находящим многочисленные применения в геометрической кристаллографии, в теории дифракции и структурном анализе кристаллов, в физике твердого тела.

Например, понятие обратной решетки используется для описания периодического распределения отражающей способности кристалла по отношению к рентгеновским лучам. Отражение рентгеновских лучей от плоскостей структуры кристалла описывается формулой Вульфа – Брэгга, из которой следует, что при постоянной длине волны рентгеновского

излучения λ большому межплоскостному расстоянию для семейства параллельных отражающих плоскостей d отвечает малый угол падения θ , т. е., чем больше межплоскостное расстояние, тем ближе направления отраженных лучей к направлению падающего пучка.

Каждый узел обратной решетки соответствует возможному отражению от плоскостей прямой решетки кристалла. Направление вектора обратной решетки G_{hkl} совпадает с направлением отражения от плоскостей hkl в прямой решетке, а n -й узел обратной решетки в этом ряду отвечает отражению n -го порядка от этих плоскостей.

1.5. Кристаллические структуры твердых тел

Структура кристалла – это конкретное расположение частиц в пространстве. Описывая структуру, необходимо указать вид и размер частиц, а также расстояния между ними. Если вещества кристаллизуются в сходных друг с другом решетках, то говорят, что они относятся к одному *структурному типу*. У кристаллов, принадлежащих к одному структурному типу, элементарные ячейки одинаковы с точностью до подобия. Опишем некоторые наиболее часто встречающиеся структурные типы.

Опытные данные показывают, что большинство металлов и полупроводников кристаллизуются в объемно- и гранецентрированных кубических, а также в гексагональных плотноупакованных решетках.

Металлы с ГЦК-решеткой относятся к *структурному типу меди*, например золото, серебро, никель, алюминий, кальций, свинец, α -кобальт и др. Все эти металлы сравнительно мягкие, пластичные, легко обрабатываются. Структурой типа меди обладают некоторые твердые растворы, например твердый раствор меди в золоте (Cu–Au), а также многие так называемые интерметаллические соединения (MgZn, NaSn, Au₄Al и др.). Атомы в веществах со структурным типом меди располагаются в вершинах и центрах граней куба с ребрами, равными по длине периоду элементарной ячейки. На элементарную ячейку приходится 4 атома. Каждый атом окружен 12 ближайшими атомами, т. е. координационное число $n = 12$. Структура характеризуется плотной кубической трехслойной упаковкой *ABCABC...*

ОЦК-структура характерна для *структурного типа вольфрама*. К этому типу, кроме самого вольфрама, относятся тугоплавкие металлы: хром, ванадий, молибден, ниобий, тантал, β -кобальт, α -железо и др. Атомы располагаются по вершинам и в центре ячейки, т. е. на ячейку приходится 2 атома. Эта структура не относится к плотноупакованным. Расстояния между атомами равны $a\sqrt{3}/2$, где a – период элементарной ячейки.

Другой структурный тип, характерный для некоторых металлических кристаллов, – *тип магния*. Элементарная ячейка магния – гексагональная примитивная. К структурному типу относятся, например, кадмий, бериллий, таллий, титан, никель и др., интерметаллические соединения, AgCd , AuCd , CuCd_3 и т. д. Элементарная ячейка построена на трех трансляциях, две из которых лежат в плотноупакованном слое атомов и составляют между собой угол $\gamma = 120^\circ$, а третья перпендикулярна этому слою. В структуре магния на элементарную ячейку приходится 2 атома магния. Каждый атом окружен 12 ближайшими атомами: 6 в том же слое и 3 в соседних слоях сверху и снизу. Следовательно, координационное число, характерное для данного структурного типа, $n = 12$. В идеальных плотноупакованных гексагональных решетках металлов отношение высоты элементарной ячейки c к расстоянию a между соседними атомами равно 1,633, хотя сами периоды c и a для разных веществ различны.

Многие щелочногалогидные кристаллы относятся к *структурному типу каменной соли* NaCl (рис. 1.12). Крупные ионы Cl (анионы) образуют кубическую плотнейшую упаковку, в которой все октаэдрические пустоты заселены катионами Na . Тетраэдрические пустоты не заполнены. Для структуры NaCl координационное число $n = 6$ для обоих сортов атомов, т. е. каждый атом Na окружен 6 атомами Cl , и наоборот.

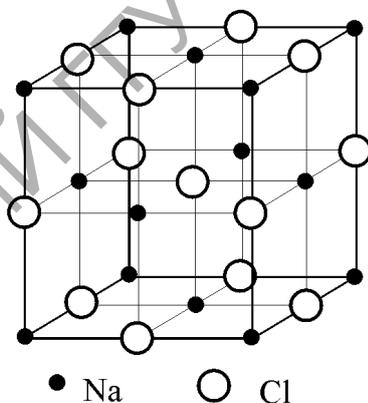


Рис. 1.12. Структура NaCl (каменной соли)

Коэффициент упаковки в структуре NaCl $\eta = 52,3\%$, т. е. около половины объема ячейки приходится на пустоты между атомами. Период ячейки NaCl $a = 5,64 \text{ \AA}$. Решетку такого типа можно рассматривать как совокупность двух гранецентрированных решеток, смещенных друг относительно друга на $\frac{1}{2}$ пространственной диагонали куба, каждая из которых содержит ионы лишь одного знака. Структурным типом NaCl характеризуются, например, окислы переходных элементов: титана (TiO), железа (FeO), никеля (NiO).

На рис. 1.13 изображены *структуры сульфида цинка ZnS*. Этими структурными типами характеризуются многие бинарные соединения (например, арсенид галлия GaAs, окись цинка ZnO). В природе ZnS встречается в двух полиморфных модификациях:

1) *обычная цинковая обманка, или сфалерит* (рис. 1.13, *а*), имеет в основе кубическую упаковку анионов цинка Zn, одна половина тетраэдрических пустот которой занята катионами серы S. Структуру сфалерита можно описать как две ГЦК-решетки – серы и цинка, смещенные друг относительно друга на четверть пространственной диагонали куба.

2) *гексагональная модификация – вюрцит* (рис. 1.13, *б*) – имеет в основе плотнейшую гексагональную упаковку анионов серы, а катионы цинка занимают половину тетраэдрических пустот.

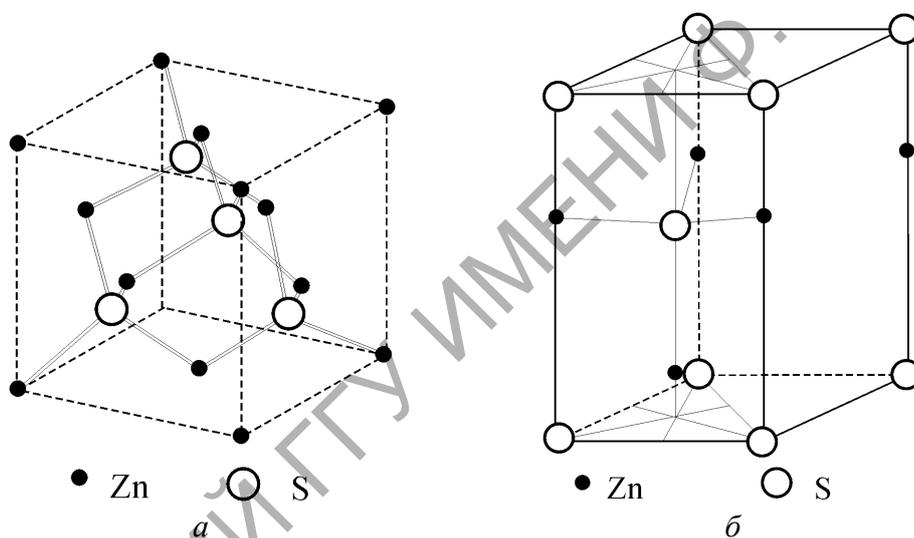


Рис. 1.13. Структура сфалерита (*а*) и вюрцита (*б*)

ГЦК-структура цинковой обманки идентична структуре алмаза, если в ней атомы цинка и серы заменить атомами углерода (рис. 14, *а*). К структурному типу алмаза относятся такие полупроводники, как арсенид галлия GaAs (рис. 1.14, *б*), кремний (Si), германий (Ge) и др.

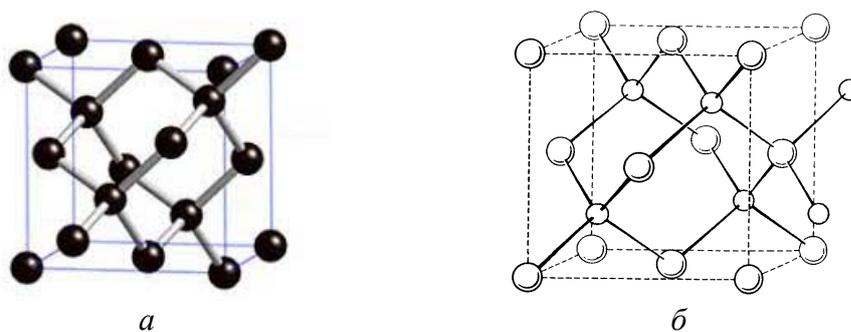


Рис. 1.14. Структура алмаза (*а*) и арсенида галлия (*б*)

Перечисленными структурными типами, конечно, не исчерпывается все их многообразие. Более полно эти сведения изложены в учебниках по кристаллографии.

1.6. Дефекты в кристаллах

Все реальные твердые тела (монокристаллические и поликристаллические) содержат дефекты структуры, являющиеся нарушениями периодичности пространственного расположения атомов. Влияние дефектов на физические свойства кристаллов чрезвычайно разнообразно. Оно определяется характером сил связи в кристаллах, их энергетической структурой (металлы, полупроводники или диэлектрики). Если фундаментальные физические свойства вещества определяются его химическим составом и идеальной структурой, то некоторые изменения этих свойств и придание новых оптических, электронных, механических и других характеристик можно осуществить введением или изменением концентрации дефектов в них.

Некоторый беспорядок в расположении атомов в кристалле вносит уже тепловое движение. Поскольку атомы колеблются в произвольных направлениях, их мгновенное расположение в какой-либо момент времени отличается от такового в другой момент и является в некоторой степени неупорядоченным. Однако само по себе тепловое движение оказывается причиной только слабого отклонения свойств реального кристалла от идеального, поскольку статистически в среднем центры колебаний расположены в узлах идеальной решетки.

В реальном кристалле неизбежны другого рода нарушения порядка, характерного для идеального, и они представляют собой уже настоящие дефекты. По своей природе дефекты решетки идеального кристалла разделяют на *собственные* (или *структурные*) и *примесные* (*химические*). Под химическими дефектами понимают связанные с наличием примесей отклонения от правильной решетки идеального кристалла. К структурным, собственным, дефектам относят геометрические отклонения от регулярного расположения атомов в идеальном кристалле. Геометрическая классификация структурных дефектов основывается на их пространственной протяженности, по отношению к числу направлений, в которых нарушено периодическое расположение атомов в решетке. На этой основе выделяют четыре класса дефектов:

- точечные (нульмерные), размеры которых не превышают одного или нескольких межатомных расстояний. К ним относятся вакансии, межузельные атомы, дефекты Френкеля и Шоттки;

- линейные (одномерные), представляющие собой нарушение периодичности в одном измерении, например, дислокации, микротрещины;
- поверхностные (двумерные), к которым относятся границы зерен и двойников, дефекты упаковки, межфазные границы, стенки доменов, поверхность кристалла;
- объемные (трехмерные) – это микропустоты.

Точечный дефект – это локальное нарушение кристаллической структуры, размеры которого во всех трех измерениях сравнимы с одним или несколькими (немногими) межатомными расстояниями. К простейшим точечным дефектам относятся *вакансии* – отсутствие атома или иона в узле кристаллической решетки и *межузельные* внедренные атомы, находящиеся в пустотах кристаллической решетки. Важнейшей особенностью точечных дефектов в ионных и ковалентных кристаллах является то, что они могут быть как электрически нейтральны, так и иметь заряд. Однако в целом кристалл остается электрически нейтрален. Условие электронейтральности обеспечивается образованием равного количества положительно и отрицательно заряженных дефектов, сложных дефектов или свободных электронов и дырок.

Возникновение точечных дефектов в твердом теле может быть обусловлено различными процессами, происходящими как за счет флуктуаций тепловой энергии внутри самого твердого тела, так и при внешнем воздействии на него. Один из методов получения избыточного при данной температуре количества точечных дефектов состоит в резком охлаждении от более высокой температуры – *закалке*. Ее часто применяют для получения необходимых свойств материала. Другой способ создания избыточных дефектов заключается в сильной деформации кристаллической решетки *ковкой* или *прокаткой*. Третий метод – облучение твердого тела атомами или частицами с высокой энергией, например, путем облучения в циклотроне или быстрыми нейтронами в ядерном реакторе. Кроме того, дефекты возникают при встречной диффузии с разными парциальными коэффициентами диффузии (эффект Киркендала).

Согласно основным принципам статистической физики, даже в том случае, когда средняя кинетическая энергия атомов очень мала, всегда найдется некоторое количество атомов с большей энергией, достаточной для выхода атома из узла кристаллической решетки. Перемещаясь по кристаллу и отдавая часть своей энергии другим атомам, такой атом может разместиться в междоузлиях. Совокупность атома в междоузлии и вакансии называется *дефектом по Френкелю* (или *парой Френкеля*) (рис. 1.15). Вакансия и межузельный атом связаны значительными упругими силами.

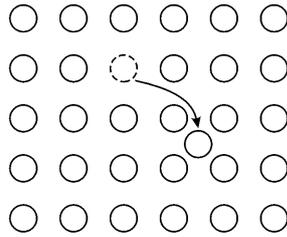


Рис. 1.15. Образование дефекта по Френкелю

Дефекты по Френкелю легко возникают в кристаллах, содержащих значительные межатомные пустоты. Примером таких кристаллов могут служить вещества со структурой алмаза или каменной соли. В кристаллической решетке некоторых металлов с плотной упаковкой такими дефектами могут быть только примесные атомы с меньшими радиусами.

В ковалентных кристаллах, таких как алмаз, вюрцит, цинковая обманка, размеры межузельных пустот достаточно велики, что позволяет размещаться в них атомам примесей. В ионных кристаллах размеры пустот уменьшаются, но все же упаковка ионов оставляет некоторое свободное пространство. Однако часто ионы сильно отличаются по размеру, и в результате упаковка получается довольно плотная. Поэтому вероятность появления межузельных атомов в ионных соединениях существенно изменяется от одного вещества к другому.

Вакансия и межузельный атом вызывают локальные упругие искажения за счет симметричного смещения окружающих их атомов в направлении от центра или к центру дефекта (рис. 1.16). В случае межузельного атома всегда имеет место эффект сжатия окружающей этот атом решетки. Это приводит к увеличению среднего периода решетки. Для вакансий наблюдается противоположный эффект.

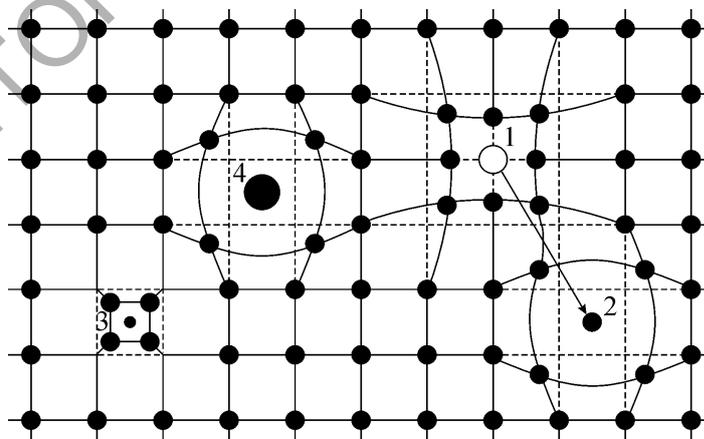


Рис. 1.16. Образование упругих смещений, вызываемых точечными дефектами: 1 – вакансия; 2 – собственный межузельный атом; 3 – атом примеси внедрения меньший, чем атомы в узлах решетки; 4 – атом примеси внедрения больший, чем атомы в узлах решетки

При введении в решетку кристалла атомов примеси также происходят упругие смещения атомов кристаллической решетки в результате того, что размеры примесных атомов всегда отличаются от размеров атомов кристаллической решетки. Если атомы примеси располагаются в узлах кристаллической решетки, то образуется твердый раствор замещения, если же в междоузлиях – твердый раствор внедрения.

Точечные *дефекты по Шоттки* в основном встречаются в кристаллах с плотной упаковкой, где образование межузельных атомов затруднено или энергетически невыгодно. Некоторые атомы из приповерхностного слоя в результате теплового движения могут выйти из кристалла на поверхность (рис. 1.17). Вакансия в освободившемся узле может затем мигрировать в объем кристалла. Образование дефектов по Шоттки уменьшает плотность кристалла, так как растет его объем при постоянной массе, тогда как при образовании дефектов Френкеля плотность остается неизменной, поскольку объем всего тела не меняется. Измерения плотности показывают, что, например, в чистых щелочно-галогенидных кристаллах преобладают дефекты по Шоттки, тогда как для чистых кристаллов галогенидов серебра характерны дефекты Френкеля.

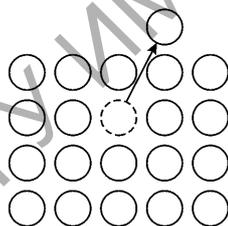


Рис. 1.17. Образование дефекта по Шоттки

Точечные дефекты, возникающие при облучении кристалла быстрыми частицами, называются *радиационными*. Основное отличие этого вида точечных дефектов от рассмотренных выше тепловых состоит в том, что они термодинамически неравновесны, т. е. после прекращения бомбардировки быстрыми частицами состояние кристалла не стационарно. Основные процессы, приводящие к появлению радиационных дефектов, следующие:

- упругие столкновения быстрых частиц с ядрами атомов кристалла;
- возбуждение электронных оболочек атомов и их ионизация;
- ядерные превращения – переход части атомов в кристалле в радиоактивное состояние и превращение их после распада в примесные центры.

В зависимости от природы излучения, энергии быстрых частиц и свойств самого кристалла одни процессы могут преобладать над другими.

Для возникновения радиационных точечных дефектов наибольшее значение имеет первый из перечисленных процессов. Быстрая частица,

испытывающее столкновение с атомом твердого тела, передает ему некоторую кинетическую энергию. Если эта энергия превышает некоторое *пороговое значение* E_d , то атом, выбитый из узла решетки, движется через кристалл. В результате на его месте возникает вакансия. Если же энергия, полученная атомом при столкновении, меньше E_d , то возникают лишь упругие волны небольших смещений, передающиеся от атома к атому, энергия которых затем переходит в энергию теплового движения атомов.

Для кристаллов, энергия связи в которых составляет приблизительно 10 эВ, значение пороговой энергии $E_d \approx 25$ эВ. Если значение энергии выбитого в междоузлии атома значительно превышает пороговое, то эти атомы сами могут выбить следующие атомы с их мест. Такие атомы называются *атомами отдачи*. Первичные атомы отдачи могут вызвать появление вторичных, третичных и т. д. до тех пор, пока энергия следующего выбитого атома не станет меньше значения E_d для данного материала. ♦

Радиационные точечные дефекты, в отличие от тепловых, всегда парные (по Френкелю). Итог облучения твердого тела быстрыми частицами может состоять в том, что на пути движущейся в кристалле быстрой частицы образуется сильно разупорядоченная область (рис. 1.18), размеры и форма которой зависят от энергии, массы и природы быстрой частицы, массы атомов кристалла, его температуры и структуры. Образуется каскад смещений.

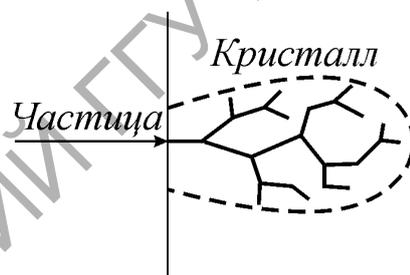


Рис. 1.18. Схема образования разупорядоченной области в кристалле (выделена пунктиром) за счет каскада смещений атомов (сплошные линии)

Этот процесс происходит за очень короткое время (10^{-13} – 10^{-15} с). После этого наступает длительный процесс релаксации, по окончании которого тело приходит в термодинамическое равновесие.

В некоторых твердых телах даже в условиях комнатных или более низких температур первичные радиационные дефекты могут перемещаться по кристаллу, встречаясь с другими дефектами. При этом вероятны как процессы аннигиляции вакансий и межузельных атомов, так и их объединение в более устойчивые вторичные дефекты. Окончательное распределение дефектов определяется количеством первоначально смещенных атомов и природой кристалла.

Некоторая критическая концентрация радиационных дефектов может приводить к нарушению устойчивости кристаллического состояния, его переходу в аморфное состояние, что негативно отражается на механических и других свойствах материала. Поэтому, например, при выборе конструкционных материалов для реакторов большой мощности необходимо учитывать изменение свойств материала под действием излучения.

К **линейным дефектам** в кристаллах относят *дислокации* – нарушения регулярности решетки вдоль линии, представляющие собой линейные искажения типа обрыва или сдвига атомных слоев, нарушающие правильность их чередования. Поперечные размеры линейного дефекта не превышают одного или нескольких межатомных расстояний, а длина может достигать размеров кристалла. Понятие о дислокации было введено в 1940-е гг. Френкелем и Тейлором для объяснения механизма процесса пластической деформации.

Линейную дислокацию удобно рассмотреть на примере однородного изотропного (например, резинового) упругого цилиндра (рис. 1.19).

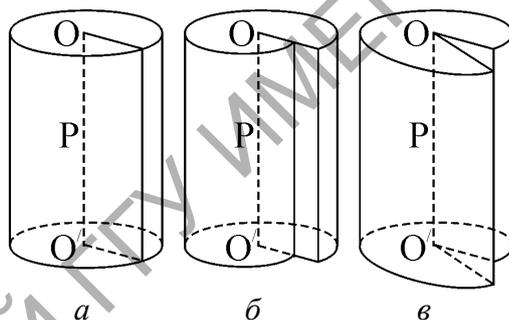


Рис. 1.19. Образование линии дислокации в упругом однородном цилиндре

Разрежем цилиндр вдоль плоскости P (рис. 1.19, *a*), сместим края так, как показано на рис. 1.19, *б* и *в*, и склеим. Линия OO' , определяющая область, в которой сдвиг произошел, от области, где сдвига нет (рис. 1.19, *в*), получила название *дислокации*.

В отличие от рассмотренного примера, кристалл имеет дискретное строение и в общем случае анизотропен, поэтому накладываются ограничения на характер возможных смещений. Допускаются только те смещения, которые соответствуют свойствам симметрии кристаллической решетки.

Рассмотрим сдвиговую дислокацию. На $n + 1$ атомных плоскостях выше плоскости скольжения приходится n плоскостей ниже плоскости скольжения. Дислокация OO' (или *ось дислокации*) представляет собой край полуплоскости $MNO'O$ и получила название *краевой* (рис. 1.20). *Краевая дислокация* перпендикулярна вектору сдвига. Если «лишняя» полуплоскость

находится в верхней части кристалла, то дислокацию называют *положительной* и обозначают значком \perp , если в нижней части – *отрицательной* и обозначают значком \top . Область вблизи оси дислокации, составляющая 3–4 межатомных расстояния, представляющая собой некую трубку вокруг линии дислокации, называют *ядром дислокации*.

Приложение напряжения сдвига приводит к перемещению краевой дислокации, причем смещение ее оси OO' на одну трансляцию означает смену полуплоскости, образующей в данный момент дислокацию. Перемещение краевой дислокации через весь кристалл приведет к сдвигу части кристалла на одно межатомное расстояние. Результатом этого является пластическая деформация кристалла (рис. 1.21), т. е. части кристалла оказываются смещены друг относительно друга на одну трансляцию.

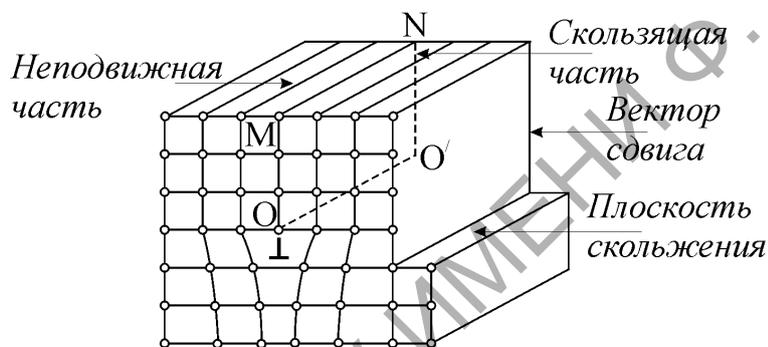


Рис. 1.20. Краевая дислокация OO' , возникшая в результате сдвига

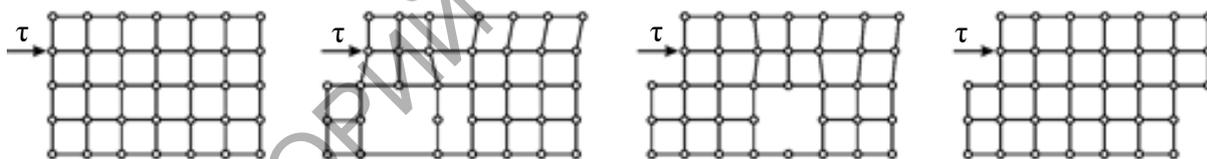


Рис. 1.21. Пластическая деформация как результат движения краевой дислокации под действием напряжения сдвига τ

Еще одним типом линейных дефектов являются *винтовые дислокации*. И. Бюргерсом (J. M. Burgers – нидерландский физик) было дано представление о *винтовой дислокации*. Пусть в кристалле произведен такой сдвиг, при котором линия дислокации OO' (рис. 1.22, а), отделяющая область, где он произошел, от области, где сдвига нет, параллельна вектору сдвига. В этом случае кристалл можно представить в виде атомной плоскости, «закрученной» вокруг оси дислокации OO' винтом. Такая дислокация названа винтовой (рис. 1.22, б).

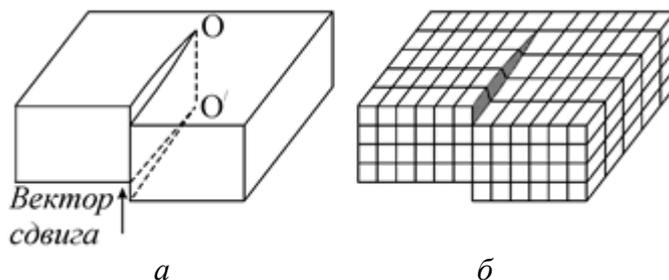


Рис. 1.22. Винтовая дислокация

Таким образом, и винтовая, и краевая дислокации – это граница между сдвинутой и несдвинутой частями кристалла (область незавершенного сдвига) или нарушение правильности структуры вдоль некоторой линии, которая не может оборваться внутри кристалла. Линия дислокации должна либо выходить на поверхность кристалла, либо разветвляться на другие дислокации, либо образовывать внутри кристалла замкнутые петли или взаимосвязанную сетку. Именно последняя возможность чаще всего реализуется в кристаллах.

Вопросы

1. Какую природу имеют силы притяжения и силы отталкивания атомов в твердом теле?
2. Какая химическая связь является самой сильной?
3. Чем отличается металлическая связь от ковалентной?
4. С какой целью в физике твердого тела используют ячейку Вигнера – Зейтца?
5. Что характеризуют в кристалле индексы Вейса и Миллера?
6. Где используется понятие обратной решетки?
7. К какому структурному типу относятся алмаз, кремний и германий?
8. В чем состоит отличие дефектов Френкеля и Шоттки?
9. Какой вид дислокации чаще всего реализуется в кристаллах?

Глава 2

ОСНОВЫ ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ

Зонная теория – один из основных разделов квантовой теории твердых тел – объяснила имеющиеся противоречия классической теории и предсказала ряд новых явлений. Согласно ей, разрешенные значения энергии электронов в твердом теле образуют определенные интервалы – разрешенные зоны, которые могут быть отделены друг от друга запрещенными зонами. Зонная теория объясняет ряд физических свойств, в частности, механизмы электропроводности твердых тел. Основы зонной теории созданы немецким физиком Нобелевским лауреатом Ф. Блохом (F. Bloch, 1928 г.) и французским физиком Л. Бриллюэном (L. Brillouin, 1930 г.). Блох сформулировал фундаментальную теорему (1929 г.), устанавливающую вид волновой функции электрона, находящегося в поле с периодически распределенным потенциалом, в частности в кристаллической решетке. Волновые функции этих (так называемых блоховских) электронов представляют собой решение уравнения Шрёдингера (1926 г., E. Schrödinger, австрийский физик, лауреат Нобелевской премии) – основного уравнения квантовой механики. В основу квантовой механики легла гипотеза французского физика Л. де Бройля (L. de Broglie, лауреат Нобелевской премии) о двойственной природе микрочастиц – корпускулярно-волновом дуализме (1924 г.). Подтвержденная на опыте идея де Бройля принципиально изменила представления о физической структуре микромира. Итальянский физик Нобелевский лауреат Э. Ферми (E. Fermi) ввел понятие изоэнергетической поверхности (Ферми-поверхность), соответствующей энергии (Ферми-энергия, или уровень Ферми), ниже которой при $T = 0$ К все энергетические состояния частиц или квазичастиц заполнены, а выше – пусты. При определении Ферми-поверхностей используют представления о зонах Бриллюэна – областях обратной решетки, в пределах которых энергия частиц (например, электронов в кристалле) меняется непрерывно, а следовательно, непрерывно и движение волнового вектора в пределах зоны Бриллюэна. При определенных допущениях зонная теория применима для описания физических свойств некристаллических твердых тел.

Зонная теория – «сердце» квантовой теории твердых тел, объясняет физическую суть перемещений электронов в полупроводниках. Ее основной результат достаточно прост: энергии электронов в полупроводниках могут принимать определенные значения в пределах разрешенных зон, которые отделены друг от друга запрещенными зонами. Зонная теория представляет собой фундаментальный результат применения представлений квантовой механики в физике твердого тела, позволивший ввести объяснение атомных явлений в русло классической физики.

2.1. Квантово-механические предпосылки зонной теории

2.1.1. Корпускулярно-волновой дуализм

Квантовые свойства электромагнитного поля были открыты после исследования М. Планком законов теплового излучения (1900 г). Он предложил понятие «кванты электромагнитного поля». Квант электромагнитного поля, или *фотон*, – нейтральная элементарная частица с нулевой массой и спином, равным 1. Фотоны являются переносчиками электромагнитного взаимодействия между заряженными частицами и сами во многом похожи на частицы (корпускулы), т. е. обладают определенной энергией и импульсом. В то же время фотоны обуславливают волновые свойства электромагнитного излучения, которые проявляются в явлениях дифракции и интерференции света. Таким образом, возникло представление о двойственной природе фотона, или о корпускулярно-волновом дуализме.

В 1924 г. Л. де Бройль высказал гипотезу о том, что корпускулярно-волновой дуализм присущ всем без исключения движущимся материальным частицам – электронам, протонам, атомам и т. п. Количественные соотношения между волновыми и корпускулярными характеристиками частиц те же, что и установленные ранее для фотонов. Частице с энергией E и импульсом p соответствует волна с частотой $\nu = E / h$ и длиной $\lambda = h / p$, где h – постоянная Планка. Эти волны получили название *волн де Бройля*.

Для частиц, движущихся с не очень высокой скоростью ($v \ll c$), длина волны равна $\lambda = h / (mv)$, где m и v – масса и скорость частицы. Следовательно, длина волны де Бройля тем меньше, чем больше масса частицы и ее скорость. Например, частице с массой в 1 г, движущейся со скоростью 1 м/с, соответствует волна де Бройля с $\lambda \approx 10^{-18}$ Å, что лежит за пределами доступной наблюдению области длин волн. Поэтому волновые свойства несущественны в механике макроскопических тел. Для электронов с энергиями от 1 эВ до 10 кэВ длины волн де Бройля лежат в диапазоне от 10 до 0,1 Å, т. е. в интервале длин волн рентгеновского излучения. Поэтому волновые свойства электронов проявляются при рассеянии на тех же кристаллах, на которых наблюдается дифракция рентгеновских лучей.

Гипотеза де Бройля была подтверждена экспериментально в 1927 г. Американские физики К. Дэвиссон (K. Davisson) и Л. Джермер (L. Germer) обнаружили, что пучок электронов, рассеивающийся на естественной дифракционной решетке (кристалл никеля), дает отчетливую дифракционную картину. Дифракционные максимумы соответствовали формуле Вульфа – Брэгга ($2d \sin \theta = m\lambda$, где d – расстояние между кристаллографическими плоскостями; θ – угол скольжения лучей; $m = 1, 2, 3 \dots$), а длина волны λ

оказалась точно равной длине волны де Бройля. В дальнейшем формула де Бройля была подтверждена многочисленными опытами.

Согласно корпускулярно-волновой природе вещества, для описания микрочастиц используют как волновые, так и корпускулярные представления (таблица 2.1). Однако корпускулярно-волновая двойственность свойств частиц, изучаемых в квантовой механике, приводит к принципиальному вопросу о границе применимости понятий классической физики в микромире. Ответ на него дает *принцип неопределенности* (соотношение неопределенностей) – фундаментальное положение квантовой теории, утверждающее, что любая физическая система не может находиться в состояниях, в которых координаты ее центра инерции и импульс одновременно принимают вполне определенные, точные значения.

Таблица 2.1

Связь между корпускулярными и волновыми свойствами свободных частиц, обладающих массой m и скоростью v

Корпускулярные свойства	Волновые свойства
Скорость v	Длина волны де Бройля $\lambda = h / (mv) = h / p$
Импульс p	Частота волны де Бройля $\nu = E / h$
Энергия свободной частицы $E = p^2 / (2m)$	Групповая скорость волн де Бройля $u = v$. Фазовая скорость волн де Бройля $v_{\text{фаз}} = c^2 / v$

Согласно соотношению неопределенностей, микрочастица (микрообъект) не может иметь одновременно и определенную координату (x, y, z), и определенную соответствующую проекцию импульса (p_x, p_y, p_z). Неопределенность значений этих характеристик удовлетворяет условиям:

$$\Delta x \Delta p_x \geq h, \Delta y \Delta p_y \geq h, \Delta z \Delta p_z \geq h, \quad (2.1)$$

где $\Delta x, \Delta y$ и Δz – неточности в определении координат частицы; $\Delta p_x, \Delta p_y, \Delta p_z$ – неточности в определении проекции импульса p частицы на эти координаты.

Из соотношений (2.1) следует, что чем точнее определена одна из входящих в неравенство величин, тем менее определено значение другой. Ввиду малости величины h соотношения неопределенностей существенны в основном для микрообъектов атомных (и меньших) масштабов и не проявляются в опытах с конденсированными телами.

Принцип неопределенности, открытый в 1927 г. немецким физиком В. Гейзенбергом (W. Heisenberg, лауреат Нобелевской премии 1932 г.), сыграл ключевую роль в выяснении закономерностей внутриатомных явлений и построении квантовой механики.

2.1.2. Волновые функции и уравнение Шрёдингера

Поскольку всем микрообъектам (по традиции за ними сохранился термин «частицы») присущи и корпускулярные, и волновые свойства, то очевидно, любую из этих «частиц» нельзя считать ни частицей, ни волной в классическом понимании. Возникла потребность в теории, согласно которой волновые и корпускулярные свойства материи выступали бы не как исключаяющие, а как взаимно дополняющие друг друга. Такая теория – *волновая*, или *квантовая механика*, – появилась благодаря концепции де Бройля. Последняя нашла отражение даже в названии «волновая функция» для величины, описывающей в этой теории состояние системы.

Исторически название «волновая функция» возникло из-за того, что уравнение, определяющее эту функцию, похоже на уравнение, описывающее волновые процессы. Волновая функция характеризует состояние квантовых объектов, изменения которых во времени описываются уравнением Шрёдингера.

Уравнение Шрёдингера – основное динамическое уравнение нерелятивистской квантовой механики, которое играет в квантовой механике такую же фундаментальную роль, как уравнения движения Ньютона в классической механике и уравнения Максвелла в классической электродинамике. Термин «нерелятивистская механика» означает, что частицы движутся со скоростями, несоизмеримыми со скоростью света. Уравнение Шрёдингера постулируется. Оно не может быть выведено из более простых представлений, как не могут быть выведены законы Ньютона. Уравнение Шрёдингера, следовательно, является просто законом физики, объясняющим физические явления. Его справедливость доказана тем, что выводы квантовой механики, полученные с помощью этого уравнения в атомной и ядерной физике, находятся в хорошем согласии с опытом.

Решение уравнения Шрёдингера позволяет определить стационарные состояния и энергетический спектр совокупности большого числа атомных ядер и электронов в кристалле. В общем виде уравнение Шрёдингера имеет вид:

$$\hat{H}\Phi = \varepsilon\Phi, \quad (2.2)$$

где \hat{H} – гамильтониан кристалла; Φ – его волновая функция; ε – энергия кристалла. Волновая функция кристалла зависит от координат всех электронов r_i и всех атомных ядер R_a :

$$\Phi = \Phi(r_1, r_2, \dots, r_n, R_1, R_2, \dots, R_n).$$

Оператор Гамильтона включает в себя:

1) оператор кинетической энергии электронов

$$\sum_i \left(-\frac{\hbar^2}{2m_0} \Delta_i \right),$$

где \hbar – постоянная Планка h , деленная на 2π ($\hbar = \frac{h}{2\pi}$); m_0 – масса электрона;

$\Delta_i = \frac{d^2}{dx_i^2} + \frac{d^2}{dy_i^2} + \frac{d^2}{dz_i^2}$ – оператор Лапласа для i -го электрона;

2) оператор кинетической энергии ядер

$$\sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \Delta_\alpha \right),$$

где M_α – масса ядра; $\Delta_\alpha = \frac{d^2}{dX_\alpha^2} + \frac{d^2}{dY_\alpha^2} + \frac{d^2}{dZ_\alpha^2}$;

3) потенциальную энергию парного взаимодействия электронов

$$\frac{1}{2} \sum_i \sum_{\substack{j \\ i \neq j}} \frac{e^2}{r_{ij}};$$

4) потенциальную энергию парного взаимодействия ядер

$$V_0(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N);$$

5) потенциальную энергию взаимодействия электронов с ядрами

$$U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n).$$

С учетом этих составляющих гамильтониана уравнение Шрёдингера примет вид:

$$\left[\sum_i \left(-\frac{\hbar^2}{2m_0} \Delta_i \right) + \sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \Delta_\alpha \right) + \frac{1}{2} \sum_i \sum_j \frac{e^2}{r_{ij}} + V_0(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) + \right. \\ \left. + U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n) \right] \Phi = \varepsilon \Phi. \quad (2.3)$$

Число независимых переменных в (2.3) определяется полным числом частиц в кристалле, которые в 1 см^3 вещества составляют около 10^{23} . Такая задача в настоящее время не может быть решена в общем виде. Возможно лишь ее приближенное решение, которое достигается с помощью ряда последовательных приближений.

Приближенное решение может быть сведено к задаче движения одного (любого) электрона в поле периодически расположенных в кристалле ионов. Такой подход носит название *одноэлектронного приближения*. В этом приближении вводится ряд упрощающих условий:

1) большая разница масс атомных ядер и электронов приводит к очень большой разнице в скоростях движения. Поэтому при описании движения электрона можно не учитывать движение ядер, а рассматривать движение электрона в поле неподвижных ядер. Это условие получило название *адиабатического приближения*;

2) медленное движение ядер можно рассматривать не в поле, создаваемом мгновенным расположением электронов, а в поле, создаваемом средним пространственным распределением зарядов электронов;

3) взаимодействие каждого электрона с остальными, зависящее от мгновенного расположения всех электронов, рассматривается как взаимодействие электрона с самосогласованным полем, создаваемым средним пространственным распределением зарядов электронов.

Разделим всю систему частиц на легкие (электроны) и тяжелые (атомные ядра). В равновесном состоянии средние значения кинетической энергии этих частиц одного порядка. Так как масса ядра намного больше массы электрона, т. е. $M_\alpha \gg m_0$, то скорость движения электронов намного превосходит скорость ядер (приблизительно на два порядка). При каждом изменении положения атомных ядер практически мгновенно устанавливается пространственное распределение электронов, соответствующее новому положению ядер. Поэтому можно считать, что движение электронов не зависит от движения ядер, а значит, движение электронов и ядер происходит без обмена между ними энергией. Это позволяет в первом приближении рассматривать движение электронов в потенциальном поле фиксированных ядер. В этом случае волновая функция и энергия электронов будут некоторыми функциями, адиабатически меняющимися с изменением расположения ядер, координаты которых будут входить в эти функции как параметры. При изучении движения ядер, напротив, следует учитывать не мгновенное положение электронов, а поле, создаваемое их средним пространственным распределением. Такое приближенное рассмотрение называют *адиабатическим, или приближением Борна – Оппенгеймера*.

Воспользуемся наиболее грубым приближением, полагая, что ядра покоятся. Для этого случая координаты ядер $\mathbf{R}_1, \dots, \mathbf{R}_N$ уже не переменные, а параметры $\mathbf{R}_{10}, \dots, \mathbf{R}_{N0}$, представляющие собой координаты узлов кристаллической решетки. Для случая покоящихся ядер, когда $\mathbf{R}_\alpha, \dots, \mathbf{R}_{\alpha 0}$, кинетическая энергия ядер обращается в нуль, а энергия их взаимодействия V_0 принимает постоянное значение. Выбором начала отсчета энергии ее

можно обратить в нуль. С учетом этого уравнение (2.3) упрощается и теперь описывает движение электронов в поле покоящихся ядер. Для этого случая энергии электронов E_e и их волновая функция ψ_e уже параметрически зависят от расположения ядер $\mathbf{R}_{\alpha 0}$. В результате уравнение Шрёдингера для электронов, движущихся в поле покоящихся ядер, запишется в виде:

$$\left[\sum_i \left(-\frac{\hbar^2}{2m_0} \Delta_i \right) + \frac{1}{2} \sum_i \sum_{\substack{j \\ i \neq j}} \frac{e^2}{r_{ij}} + U(\mathbf{r}_1, \dots, \mathbf{r}_n, \mathbf{R}_{10}, \dots, \mathbf{R}_{N0}) \right] \Psi_e = E_e \Psi_e. \quad (2.4)$$

Для дальнейшего сокращения числа переменных в уравнении (2.4) используется *валентная аппроксимация*. Принимается, что все электроны атома, кроме валентных, вместе с ядром образуют неподвижный атомный остаток (ион). При этом (2.4) записывается только для валентных электронов, движущихся в потенциальном поле фиксированных ионов. Несмотря на проведенные упрощения, уравнение (2.4) не может быть решено в общем виде: по-прежнему это задача для многих частиц, не поддающаяся непосредственному решению.

Одним из наиболее распространенных методов решения многоэлектронной задачи для кристалла является *метод Хартри – Фока*, позволяющий многоэлектронную задачу свести к одноэлектронной. Его идея заключается в том, что энергия попарного взаимодействия электронов заменяется взаимодействием каждого электрона с усредненным полем всех остальных электронов. Пусть потенциальная энергия i -го электрона в этом поле будет Ω_i . Она зависит не только от движения всех остальных электронов, но и данного электрона, так как его движение оказывает влияние на движение остальных электронов. Это поле получило название *самосогласованного*. Введение самосогласованного поля позволяет в уравнении (2.4) двойную сумму заменить суммой, каждый член которой зависит от координат одного электрона, т. е.

$$\frac{1}{2} \sum_i \sum_{\substack{j \\ i \neq j}} \frac{e^2}{r_{ij}} = \sum_i \Omega_i(\mathbf{r}_i), \quad (2.5)$$

где $\Omega_i(\mathbf{r}_i)$ – потенциальная энергия i -го электрона в поле всех остальных электронов.

Аналогично потенциальную энергию взаимодействия электронов с ядрами можно также представить в виде суммы:

$$U(\mathbf{r}_1, \mathbf{r}_2, \dots) = \sum_i U_i(\mathbf{r}_i), \quad (2.6)$$

где $U_i(\mathbf{r}_i)$ – потенциальная энергия i -го электрона в поле всех ядер.

С учетом (2.5) и (2.6) уравнение (2.4) преобразуется к виду

$$\left\{ \sum_i \left[-\frac{\hbar^2}{2m_0} \Delta_i + \Omega_i(\mathbf{r}_i) + U_i(\mathbf{r}_i) \right] \right\} \Psi_e = E_e \Psi_e. \quad (2.7)$$

Это уравнение можно записать как

$$\hat{\mathbf{H}} \Psi_e = \left(\sum_i \hat{\mathbf{H}}_i \right) \Psi_e = E_e \Psi_e, \quad (2.8)$$

где $\hat{\mathbf{H}}_i$ – гамильтониан i -го электрона, равный:

$$\hat{\mathbf{H}}_i = -\frac{\hbar^2}{2m_0} \Delta_i + \Omega_i(\mathbf{r}_i) + U_i(\mathbf{r}_i).$$

Из соотношения (2.8) следует, что гамильтониан кристалла представляет собой сумму гамильтонианов, каждый из которых зависит от координат одной частицы. Это означает, что электроны ведут себя независимо друг от друга (как бы не взаимодействуя), а полная энергия системы частиц равна сумме энергий отдельных электронов:

$$E_s = E_1 + E_2 + \dots = \sum_i E_i. \quad (2.9)$$

Таким образом, введение согласованного поля позволяет рассматривать электроны в кристалле как *невзаимодействующие частицы*, т. е. движущиеся независимо друг от друга. (Это служит основанием для представления электронов проводимости в виде идеального газа.) Иными словами, введение самосогласованного поля позволяет задачу многих частиц свести к задаче для одного электрона:

$$\hat{\mathbf{H}}_i \Psi = E \Psi, \quad (2.10)$$

где $\hat{\mathbf{H}}_i$, $\Psi(\mathbf{r})$, E – соответственно гамильтониан, волновая функция и энергия электрона в кристалле.

Если ввести обозначение для потенциальной энергии электрона в кристалле через функцию $V(\mathbf{r})$, равную:

$$V(\mathbf{r}) = U(\mathbf{r}) + \Omega(\mathbf{r}),$$

то уравнение Шрёдингера для электрона в кристалле запишется в виде:

$$\left[-\frac{\hbar^2}{2m_0} \Delta + V(\mathbf{r}) \right] \Psi(\mathbf{r}) = E \Psi(\mathbf{r}). \quad (2.11)$$

Существуют различные способы рационального выбора вида функции $V(\mathbf{r})$ для решения одноэлектронной задачи. Наиболее часто для этого

используют состояния электрона, находящегося в потенциальном поле всех ионов решетки, заряд которых в среднем скомпенсирован зарядом валентных электронов, т. е. в поле периодического потенциала. Иными словами, принимается, что член $V(\mathbf{r})$ в уравнении (2.11) есть *полный потенциал кристалла*, обладающий трехмерной периодичностью решетки.

Уравнение Шрёдингера является математическим выражением фундаментального свойства микрочастиц – корпускулярно-волнового дуализма. В предельном случае, когда длины волн де Бройля значительно меньше размеров, характерных для рассматриваемого движения, уравнение Шрёдингера позволяет описать движение частиц по законам классической механики. Переход от уравнения Шрёдингера к уравнениям классической механики, описывающим движения частиц по траекториям, подобен переходу от волновой оптики к геометрической. Аналогия между классической механикой и геометрической оптикой, которая является предельным случаем волновой оптики, сыграла важную роль в установлении уравнения Шрёдингера.

С математической точки зрения уравнение Шрёдингера есть волновое уравнение, по структуре подобное уравнению, описывающему колебания нагруженной струны. Однако, в отличие от уравнения колебания струны, которое дает решения, описывающие геометрическую форму струны в данный момент времени, решения $\Psi_n(x, y, z, t)$ уравнения Шрёдингера не имеют прямого физического смысла. Смысл имеет квадрат модуля волновой функции, а именно величина $|\Psi_n(x, y, z, t)|^2$, равная вероятности нахождения частицы (системы) в момент времени t в квантовом состоянии n в точке пространства с координатами x, y, z . Эта вероятностная интерпретация волновой функции – один из основных постулатов квантовой механики.

2.1.3. Волна (функция) Блоха

В идеальном кристалле атомы периодически расположены в пространстве. Такое расположение атомов означает, что при смещении кристалла на вектор

$$\mathbf{a}_n = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3,$$

где $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ – периоды идентичности решетки по трем произвольным направлениям, а n_1, n_2, n_3 – произвольные целые числа, кристалл совмещается сам с собой. Из этого следует, что точка с радиусом-вектором \mathbf{r} и точка с радиусом-вектором $\mathbf{r} + \mathbf{a}_n$ физически эквивалентны, поэтому

$$V(\mathbf{r} + \mathbf{a}_n) = V(\mathbf{r}). \quad (2.12)$$

Соотношение (2.12) выражает условие периодичности потенциального поля кристалла.

В периодическом поле кристалла волновая функция электрона $\Psi(\mathbf{r} + \mathbf{a}_n)$ отличается от волновой функции $\Psi(\mathbf{r})$ только постоянным множителем, т. е.

$$\Psi(\mathbf{r} + \mathbf{a}_n) = C\Psi(\mathbf{r}). \quad (2.13)$$

Но из условия нормировки волновой функции следует, что

$$|C|^2 = 1,$$

поэтому можно положить, что

$$C = e^{ika_n}, \quad (2.14)$$

где \mathbf{k} – постоянный вектор, называемый *волновым вектором*, характеризует квантовое состояние электрона в кристалле.

Поскольку произведение ka_n должно быть безразмерным, то волновой вектор имеет размерность обратной длины. Модуль волнового вектора, называемый волновым числом, выражает собой число длин волн, укладывающихся на отрезке 2π , и поэтому равен $k = 2\pi/\lambda$.

С учетом (2.14) выражение (2.13) запишется в виде

$$\Psi(\mathbf{r} + \mathbf{a}_n) = e^{ika_n}\Psi(\mathbf{r}),$$

откуда

$$\Psi(\mathbf{r}) = e^{-ika_n}\Psi(\mathbf{r} + \mathbf{a}_n) = e^{ikr}U_k(\mathbf{r}),$$

где

$$U_k(\mathbf{r}) = e^{-ik(\mathbf{r} + \mathbf{a}_n)}\Psi(\mathbf{r} + \mathbf{a}_n).$$

Функция $U_k(\mathbf{r})$ обладает трехмерной периодичностью кристаллической решетки. Таким образом, стационарная волновая функция электрона в периодическом поле кристалла зависит от волнового вектора \mathbf{k} и имеет вид:

$$\Psi_k(\mathbf{r}) = e^{ikr}U_k(\mathbf{r}), \quad (2.15)$$

где e^{ikr} – плоская волна, бегущая в направлении вектора \mathbf{k} ; $U_k(\mathbf{r})$ – некая функция координат, зависящая от волнового вектора \mathbf{k} и имеющая периодичность решетки. Выражение (2.15) для $\Psi_k(\mathbf{r})$ носит название *волны (или функции) Блоха*. Следовательно, волновая функция электрона в периодическом поле кристалла представляет собой плоскую волну, модулированную по амплитуде с периодичностью решетки кристалла и бегущую в направлении волнового вектора.

Если функцию Блоха (2.15) подставить в (2.10), то будем иметь:

$$\hat{H}_i\Psi_k(\mathbf{r}) = E\Psi_k(\mathbf{r}). \quad (2.16)$$

Из (2.16) следует, что энергия электрона в кристалле должна зависеть от волнового вектора \mathbf{k} , т. е. $E = E(\mathbf{k})$. Следовательно, решением уравнения Шрёдингера для электрона в периодическом поле кристалла является бегущая плоская волна, модулированная с периодичностью решетки, а закон дисперсии для энергии электрона в кристалле зависит от волнового вектора \mathbf{k} .

Электроны в периодическом поле кристаллической решетки, волновые функции которых являются блоховскими функциями (2.15), называют *блоховскими электронами*. Периодический потенциал $V(\mathbf{r})$, определяющий свойства блоховских электронов, получил название *самосогласованный потенциал*, так как его величина зависит от взаимодействия между всеми электронами и ионами, образующими кристаллическую решетку. Поэтому блоховский электрон представляет собой *квазичастицу*, т. е. частицу, находящуюся в самосогласованном поле окружающих частиц.

Волновые функции блоховских электронов обнаруживают сходство с волновыми функциями свободных электронов; их можно представить как промодулированные по амплитуде плоские волны. Роль сохраняющегося импульса \mathbf{p} , определяющего поведение волновой функции свободного электрона при трансляции на вектор \mathbf{a}_n , у блоховского электрона играет квазиимпульс $\hbar\mathbf{p}$. Истинного сохраняющегося импульса у блоховского электрона нет, так как в силовом поле закон сохранения импульса не выполняется – квазиимпульс сохраняется с точностью до вектора обратной решетки.

Если бы полученные результаты на основе введения самосогласованного поля зависели от численного значения потенциала периодического поля $V(\mathbf{r})$, то в настоящее время, по-видимому, не существовало бы теории твердого тела, так как поля эти неизвестны и не могут быть определены ни теоретически, ни экспериментально. Однако для того, чтобы получить фундаментальные результаты теории, нет необходимости знать численные значения силовых полей, достаточно знать, что это поле периодически в пространстве и что его периоды совпадают с периодами решетки.

2.2. Зоны Бриллюэна

Введенный при обсуждении блоховской функции волновой вектор \mathbf{k} играет в задаче о движении электрона в периодическом поле кристалла такую же роль, какую играет волновой вектор в задаче о движении свободного электрона. Состояние свободно движущегося электрона характеризуется энергией E и импульсом $\mathbf{p} = m\mathbf{v}$, которые связаны между собой соотношением:

$$E = \frac{p^2}{2m}. \quad (2.17)$$

Этому электрону соответствует волна де Бройля длиной

$$\lambda = \frac{h}{p} = \frac{h}{mv}. \quad (2.18)$$

Учитывая, что $|\mathbf{k}| = 2\pi/\lambda$, а $\hbar = h/(2\pi)$, получаем из (2.18), что $p = \hbar k$ или в векторной форме

$$\mathbf{p} = \hbar \mathbf{k}, \quad (2.19)$$

т. е. волновой вектор пропорционален импульсу электрона, а энергия свободного электрона связана с волновым вектором соотношением:

$$E = \frac{\hbar^2 k^2}{2m}. \quad (2.19, a)$$

Когда на электрон не действуют никакие силы, его энергия остается постоянной, т. е. $E(\mathbf{k}) = \text{const}$. Это значит, что не меняется \mathbf{k} и остается постоянным импульс \mathbf{p} , т. е. выполняются законы сохранения энергии и импульса.

Электрон, движущийся в кристалле (блоховский электрон), подвергается воздействию периодического поля решетки. Энергия этого взаимодействия является периодической функцией координат. Следовательно, энергия и импульс блоховского электрона изменяются со временем под действием этого поля. Пользуясь понятием волнового вектора \mathbf{k} , введенного для электрона в кристалле, т. е. входящего в функцию Блоха (2.16), можно ввести характеристику, аналогичную импульсу, но постоянную во времени:

$$\mathbf{P} = \hbar \mathbf{k}. \quad (2.20)$$

Чтобы подчеркнуть сходство и одновременно отметить отличие фигурирующей в (2.20) величины $\hbar \mathbf{k}$ от истинного импульса, эту величину называют *квазиимпульсом* электрона.

Волновой вектор электрона в кристалле, в отличие от волнового вектора свободного электрона, может принимать множество значений. Условие Блоха (2.16) не нарушается, если волновой вектор \mathbf{k} заменить на вектор $\mathbf{k} + 2\pi\mathbf{G}$, где $\mathbf{G} = hb_1 + kb_2 + lb_3$ – вектор обратной решетки (см. п. 1.4). Состояния электрона, характеризующиеся волновыми векторами \mathbf{k} и $\mathbf{k} + 2\pi\mathbf{G}$, физически эквивалентны. Другими словами, и волновая функция, и энергия электрона в кристалле являются периодическими функциями волнового вектора \mathbf{k} с периодом $2\pi\mathbf{G}$ (или квазиимпульса \mathbf{P} с периодом $2\pi\hbar\mathbf{G}$):

$$\begin{aligned} E(\mathbf{k}) &= E(\mathbf{k} + 2\pi\mathbf{G}), \\ E(\mathbf{P}) &= E(\mathbf{P} + 2\pi\hbar\mathbf{G}). \end{aligned} \quad (2.21)$$

Если в \mathbf{k} -пространстве (или в \mathbf{P} -пространстве) построить обратную решетку, растянутую во всех направлениях в 2π раз, то всё \mathbf{k} - (или \mathbf{P} -) пространство можно разделить на области, в которых имеются физически

эквивалентные состояния электронов. Эти области называют *зонами Бриллюэна*. Многогранник минимального объема, построенный вокруг начала координат в k - (или P -) пространстве, содержащий все возможные состояния электрона, называется *первой*, или *основной*, *зоной Бриллюэна*.

Первая зона Бриллюэна представляет собой элементарную ячейку Вигнера – Зейтца обратной решетки. Для определения вида первой зоны Бриллюэна в обратной решетке с параметрами ячейки $2\pi b_1$, $2\pi b_2$, $2\pi b_3$ строят ячейку Вигнера – Зейтца, пользуясь правилами, описанными в п. 1.2.

Вторая зона Бриллюэна строится аналогичным образом. В обратной решетке, параметры которой увеличены в 2π раз, узел, выбранный за начало отсчета при построении первой зоны Бриллюэна, соединяют прямыми отрезками с ближайшими эквивалентными узлами, лежащими на поверхности второй координационной сферы. Затем, как и при построении первой зоны, строят плоскости, перпендикулярные этим отрезкам и проходящие через их середины. В результате получают замкнутый многогранник, объем которого, образовавшийся после вычитания из него объема первой зоны Бриллюэна, представляет собой вторую зону. Третью и последующие зоны строят таким же образом, используя ближайшие эквивалентные узлы обратной решетки, лежащие на третьей, четвертой и последующих координационных сферах.

В отличие от первой, вторая и последующие зоны Бриллюэна состоят из нескольких отделенных друг от друга областей (рис. 2.1, *а*). В так называемой *приведенной зоне* различные участки одной зоны Бриллюэна, сдвинутые на векторы трансляции обратной решетки, соединены вместе (рис. 2.1, *б*). В результате «приведения», т. е. соединения всех областей одной зоны, становится очевидным, что каждая зона совпадает с элементарной ячейкой обратной решетки, т. е. фактически с первой зоной Бриллюэна, а объемы всех зон Бриллюэна равны.

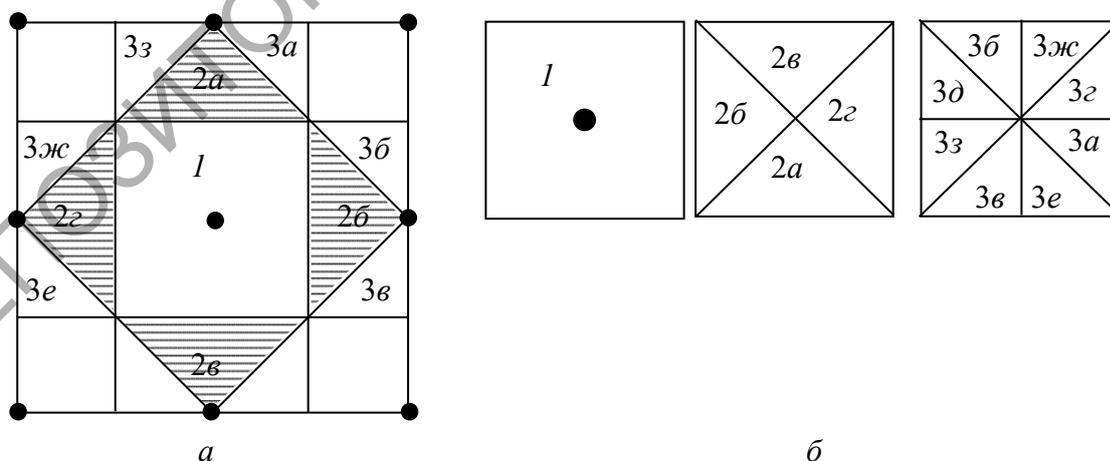


Рис. 2.1. Зоны Бриллюэна: *а* – для плоской квадратной кристаллической решетки; *б* – приведенной зоны той же решетки. Цифрами обозначены номера зон, буквами при цифрах – участки соответствующих зон

Первая зона Бриллюэна для разных кристаллических решеток имеет форму правильного многогранника различной сложности (рис. 2.2). Объем обратного пространства, заключенный в первой зоне Бриллюэна, равен $(2\pi)^3/V_0$, где V_0 – объем элементарной ячейки для решетки Браве. Эквивалентность состояний электрона в разных зонах Бриллюэна позволяет ограничить рассмотрение движения электрона только пределами первой зоны Бриллюэна. Понятие зон Бриллюэна используют при описании электронной структуры твердых тел, в частности при определении Ферми-поверхности металлов.

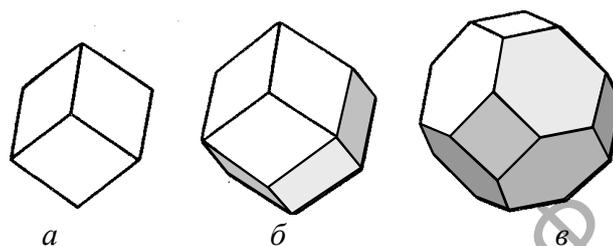


Рис. 2.2. Первая зона Бриллюэна для простой кубической (а), ОЦК (б) и ГЦК (в) кристаллических решеток

2.3. Поверхность и энергия Ферми

Поверхность Ферми (или *Ферми-поверхность*) – изоэнергетическая поверхность, ограничивающая в пространстве квазиимпульсов (p -пространстве) область энергетических состояний, занятых электронами проводимости при $T = 0$ К. Энергию, соответствующую Ферми-поверхности, называют *Ферми-энергией*, или *уровнем Ферми*. Смысл этих абстрактных понятий поясним на примере системы частиц «свободные электроны в металле», которая характеризуется наиболее простой формой Ферми-поверхности – сферической.

Энергия свободных электронов в металле квантована (т. е. принимает определенные значения). Согласно принципу Паули, в состоянии с одинаковой энергией может находиться не более двух электронов с противоположно направленными спинами.

Принцип Паули позволяет объяснить распределение электронов в твердом теле по энергиям. При $T = 0$ К они располагаются на ступенях «энергетической лестницы» по два электрона на каждом уровне, от самого нижнего до самого высокого, который определяется числом имеющихся в твердом теле свободных электронов (рис. 2.3, а).

Если в металле имеется N свободных электронов, то минимальное число занятых электронами уровней будет $N/2$ (по два электрона с разными спинами на каждом уровне). В этом случае говорят, что электронный газ

полностью «вырожден». Уровень энергии, который отделяет полностью заполненные уровни от полностью не заполненных, называется *уровнем Ферми*, или *энергией Ферми*, и обозначается E_F .

Повышение температуры ($T > 0$ К) оказывает влияние только на электроны, которые находятся вблизи уровня Ферми. Они возбуждаются и переходят в более высокие незанятые энергетические состояния (рис. 2.3, б), и «вырождение» электронного газа постепенно снимается.

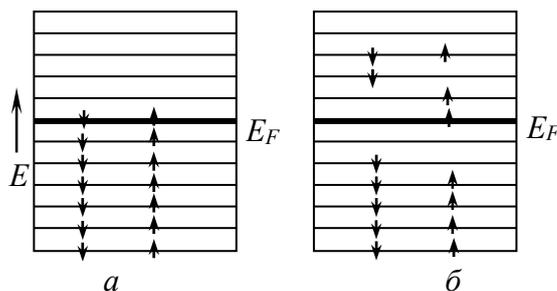


Рис. 2.3. Схема распределения электронов в твердом теле по энергетическим уровням при $T = 0$ К (а) и $T > 0$ К (б). Маленькими стрелками показаны направления спинов электронов, E_F – уровень Ферми

Если занятые электронные состояния описывать точками в k -пространстве, то для системы из N свободных электронов с массой m эти заполненные состояния образуют сферическую область с радиусом k_F . Радиус k_F сферы, называемой *сферой Ферми*, определяется из соотношения:

$$E_F = \frac{\hbar^2 k_F^2}{2m}. \quad (2.22)$$

Энергия E_F , соответствующая поверхности этой сферы, является энергией Ферми, а поверхность сферы Ферми, отделяющей в k -пространстве заполненные состояния от незаполненных, называют *поверхностью Ферми*.

Поверхность Ферми – одно из фундаментальных понятий современной теории металлов. Каждый металл характеризуется своей Ферми-поверхностью, и в общем случае она не является сферической. Геометрические характеристики Ферми-поверхности (форма, кривизна, площадь сечений и т. п.) связаны с физическими свойствами металлов, что позволяет строить Ферми-поверхности по экспериментальным данным.

Понятие о Ферми-энергии используют в физике твердого тела, в ядерной и астрофизике. Величина энергии Ферми существенно зависит от физических свойств системы. Так, в металлах она обычно составляет единицы электрон-вольт, в полупроводниках – тысячные доли эВ, а в звездном веществе может достигать сотен эВ.

2.4. Энергетический спектр электронов в кристалле

Зонная теория твердых тел является квантовой теорией энергетического спектра электронов в кристалле. Согласно этой теории, электронный спектр состоит из чередующихся зон (полос) разрешенных и запрещенных энергий.

В основе зонной теории лежит так называемое *одноэлектронное приближение*, которое базируется на следующих упрощениях:

1) атомные ядра в узлах идеальной кристаллической решетки неподвижны, их масса велика по сравнению с массой электронов;

2) электрон движется в поле периодического потенциала $U(\mathbf{r})$ решетки, которое складывается из полей, создаваемых ядрами и всеми остальными электронами;

3) периодическое поле обладает трансляционной инвариантностью. *Инвариантность* – фундаментальное физическое понятие, неизменность какой-либо величины при изменении физических условий или по отношению к преобразованиям координат и времени. Трансляционная инвариантность поля кристаллической решетки – это неизменность потенциала решетки в точках, находящихся вдоль прямой на расстояниях, пропорциональных периоду решетки a . Условие инвариантности:

$$U(\mathbf{r} + n\mathbf{a}) = U(\mathbf{r}), \quad (2.23)$$

где \mathbf{r} – пространственная координата произвольно выбранной точки в кристалле, \mathbf{a} – период решетки, n – целое число.

В такой модели для волновой функции ψ электрона в решетке выполняется теорема Блоха (2.15). Это означает, что волновая функция свободного электрона $\psi_k(\mathbf{r})$, имеет амплитуду, которая промодулирована (изменяется в пространстве) с периодом, равным периоду решетки.

2.4.1. Одномерная модель электронных зон

Спектр, т. е. совокупность всех значений энергии электронов, можно определить, подставляя волновую функцию в виде (2.15) в стационарное уравнение Шрёдингера (2.10) и вводя те или иные граничные условия и упрощения. Так, в одномерной *модели Кронига – Пенни* (R. Kronig, W. G. Penney, немецкий и английский физики, 1931 г.) периодический потенциал кристаллической решетки $U(x)$ аппроксимируют периодической последовательностью прямоугольных потенциальных ям глубиной U_0 и шириной a , разделенных потенциальными барьерами шириной b , так что постоянная решетки равна

$a + b$ (рис. 2.4). Модель Кронига – Пенни позволяет получить аналитическое решение уравнения Шрёдингера для движения электрона в периодическом поле и наглядно иллюстрирует механизм возникновения энергетических зон в кристалле.

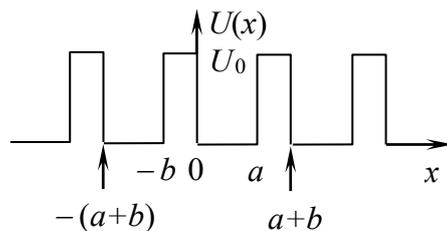


Рис. 2.4. Одномерный периодический потенциал $U(x)$ в модели Кронига – Пенни

Решение уравнения Шрёдингера дает энергетический спектр электронов в виде серии полос разрешенных энергий $E_l(k)$ (l – номера разрешенных зон), разделенных полосами запрещенных энергий. На рис. 2.5 показаны зависимости $E_l(k)$ для электрона, находящегося в одномерной решетке ($l = 3$, т. е. показаны три первые энергетические зоны). Видно, что в пределах каждой зоны для всех k , отличающихся на $\frac{2\pi}{a}n$ (a – период решетки,

n – целое число), энергия одна и та же. Интервал значений k от $-\frac{\pi}{a}$ до $\frac{\pi}{a}$

представляет собой первую зону Бриллюэна, два отрезка от $-\frac{2\pi}{a}$ до $-\frac{\pi}{a}$

и от $\frac{\pi}{a}$ до $\frac{2\pi}{a}$ – вторую зону Бриллюэна и т. д. Значения энергии электрона

в любой энергетической зоне можно получить, изменяя k в пределах первой зоны Бриллюэна. Значения E в разных полосах разрешенных энергий можно сопоставить, «приведя» их в первую зону. Такой способ изображения $E(k)$, иллюстрируемый рис. 2.5, б, получил название *схемы приведенных зон*, в отличие от зависимости, показанной на рис. 2.5, а и называемой *периодической зонной схемой*.

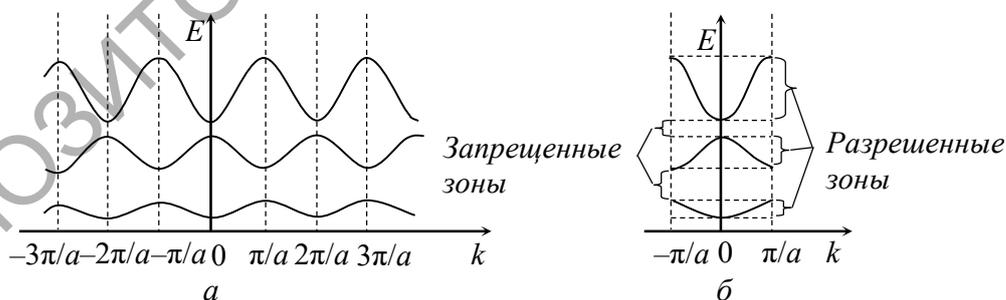


Рис. 2.5. Зависимости $E(k)$ для электрона в одномерной решетке: a – периодическая зонная схема; b – схема приведенных зон. Кривые соответствуют полосам разрешенных энергий

Другой способ описания структуры энергетической зоны основан на законе дисперсии, т. е. зависимости энергии электрона от его квазиимпульса

$E(\mathbf{p})$. Если зафиксировать координаты p_y и p_z квазиимпульса, то зависимость $E(p_x)$ представляет собой *дисперсионную кривую* на плоскости (рис. 2.6),

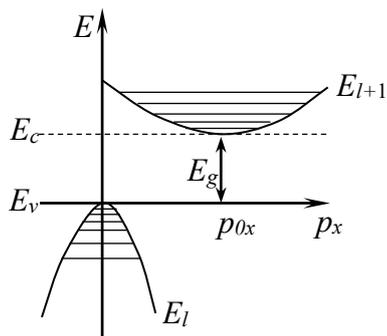


Рис. 2.6. Дисперсионные кривые $E_l(p_x)$ и $E_{l+1}(p_x)$ энергетического спектра электрона при фиксированных p_y и p_z : E_c – дно $l + 1$ зоны (зоны проводимости); E_v – потолок l -й зоны (валентной зоны); E_g – ширина запрещенной зоны; заштрихованные области – уровни, заполненные электронами и дырками

характеризующую уровни энергии в разрешенных энергетических зонах. Повторяя эту операцию для $E(p_y)$ и $E(p_z)$, получают набор дисперсионных кривых, полностью характеризующих функцию $E(\mathbf{p})$.

Зонная структура в трехмерном измерении может быть значительно сложнее, чем в рассмотренной выше одномерной модели. Зависимости $E(\mathbf{k})$ в трехмерном кристалле могут отличаться для разных направлений в зоне Бриллюэна. Это обусловлено тем, что трехмерный периодический потенциал $U(\mathbf{r})$, зависящий от структуры кристалла, в различных направлениях не одинаков. Следствием этого может быть перекрытие разрешенных зон, что для рассмотренной одномерной модели невозможно.

Следствием этого может быть перекрытие разрешенных зон, что для рассмотренной одномерной модели невозможно.

2.4.2. Заполнение зон электронами в идеальном кристалле

Зонная структура энергетического спектра электронов возникает в кристалле, образованном из N атомов, потому что каждый из них в свободном состоянии обладает дискретным электронным энергетическим спектром. Кристалл можно трактовать как гигантскую молекулу, в которой электроны всех атомов обобществлены и которую следует рассматривать как единую квантовомеханическую систему. В ней происходят квантовые переходы между уровнями энергии атомов, из которых состоит кристалл. Эти переходы связаны с изменением энергетического состояния электронов. Электронные уровни атомов образуют энергетические зоны. Если в каждом атоме Z электронов, то полное число электронов в кристалле равно NZ , и они занимают уровни в разрешенных зонах, начиная с нижних и заполняя более высокие. Нижние зоны целиком заполнены электронами внутренних оболочек атомов.

Физические свойства кристаллов определяются в основном электронами верхних разрешенных зон. Энергетический интервал E_g (рис. 2.6) между «дном» E_c (минимум энергии) самой верхней содержащей электроны зоны и «потолком» E_v (максимум энергии) соседней нижней целиком

заполненной зоны называется *запрещенной зоной*. Ниже по шкале энергии может быть еще несколько запрещенных и разрешенных зон. Зона, заполненная электронами частично (может стать пустой при $T = 0$ К), называется *зоной проводимости*. Самая верхняя зона, целиком заполненная электронами даже при $T = 0$ К, называется *валентной зоной*.

Кристаллы, у которых нижние зоны (в том числе валентная) полностью заполнены электронами, а зона проводимости пуста, являются диэлектриками или полупроводниками (рис. 2.7). Вещества с широкой запрещенной зоной, разделяющей валентную зону и зону проводимости ($E_g > 3$ эВ), условно относят к *диэлектрикам*, а вещества с более узкой запрещенной зоной ($E_g < 3$ эВ) – к *полупроводникам*.

Необходимым условием электрической проводимости твердого тела является наличие в зоне проводимости свободных энергетических уровней, на которые под действием внешнего электрического поля могли бы перейти электроны. Если зона проводимости заполнена частично и содержит свободные верхние уровни, твердое тело будет проводником. *Металлы* – это кристаллы с частично заполненной зоной проводимости. Проводниками являются также *полуметаллы*, у которых имеется перекрытие зон с образованием гибридной зоны, причем нижняя зона заполнена, а верхняя пуста, но перекрывается нижней (рис. 2.7, г).

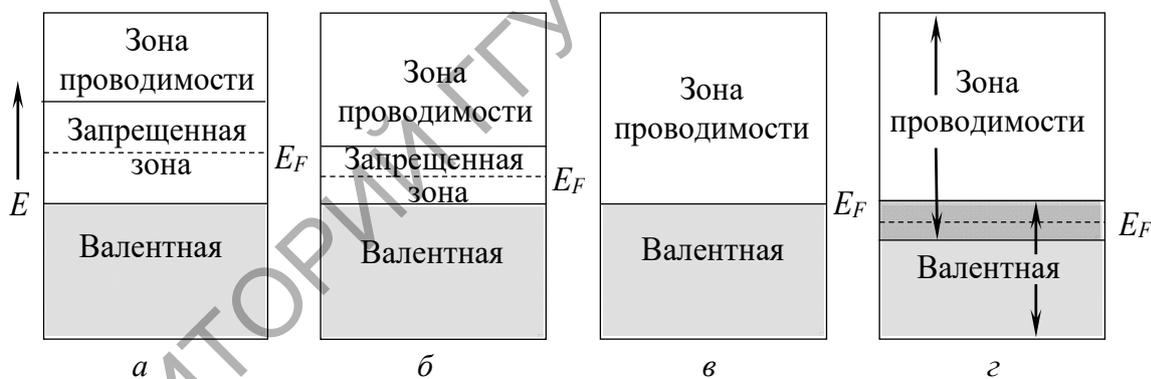


Рис. 2.7. Схема энергетических зон в диэлектриках (а), полупроводниках (б), металлах (в) и полуметаллах (г). E_F – уровень Ферми

Большинство металлов состоит из атомов с не полностью заполненными наружными электронными оболочками. Например, одиннадцать электронов в атоме Na распределены по состояниям следующим образом: $1s^2 2s^2 2p^6 3s^1$. При объединении N атомов в кристалл энергетические уровни атомов группируются в зоны. Электроны внутренних оболочек атома полностью заполняют зоны, образованные из уровней $1s$ -, $2s$ - и $2p$ -оболочек. Зону проводимости образуют электроны в $3s$ -состоянии. В ней имеется $2N$ энергетических состояний, в которых находятся N электронов из

3s-оболочки. Таким образом, в кристалле натрия зона проводимости заполнена лишь наполовину.

Кристаллы, составленные из атомов или ионов с полностью заполненными оболочками, обычно являются диэлектриками или полупроводниками. Например, щелочно-галогидные кристаллы типа NaCl, у которых все s-электроны катиона переходят на p-оболочку аниона, полностью заполняя ее, – диэлектрики. Однако многие из таких кристаллов в результате перекрытия зон приобретают свойства металлов. Типичный пример металла с подобной зонной структурой – магний. У каждого атома Mg ($1s^2 2s^2 2p^6 3s^2$) в валентной оболочке имеются два электрона. В кристалле магния валентные электроны полностью заполняют 3s-зону. Однако эта зона перекрывается со следующей разрешенной зоной, образованной из 3p-уровней.

Принципиальным и существенным является *качественное различие* между металлами, полупроводниками и диэлектриками. В металлах электронная проводимость является природным свойством кристаллов. В полупроводниках проводимость (электронная или дырочная) является вынужденной. В диэлектриках приложение внешнего поля приводит к его поляризации или к пробое.

В полупроводниках при $T > 0$ К заметное число электронов перебросено в зону проводимости. При температурах, близких к абсолютному нулю, любой полупроводник становится хорошим диэлектриком. Таким образом, между металлами и диэлектриками существует принципиальное различие, а между диэлектриками и полупроводниками – только количественное. Иначе говоря, полупроводники и диэлектрики отличаются от металлов тем, что валентная зона у них заполнена электронами, а ближайшая свободная зона (зона проводимости) отделена от валентной зоны запрещенной зоной. При этом ширина запрещенной зоны у полупроводников составляет от десятых долей до 3 эВ, а у диэлектриков больше – от 3 до 5 эВ. В таблице 2.2 приведены значения ширины запрещенной зоны для некоторых диэлектриков и полупроводников.

Таблица 2.2

Ширина запрещенной зоны в кристаллах

Кристалл	Al ₂ O ₃	C (алмаз)	BN	GaAs	Si	Ge	InSb
E_g , эВ	7,0	5,2	4,6	1,43	1,08	0,66	0,17

Электронная структура атомов, образующих твердое тело, – не единственный фактор, обуславливающий различие в заполнении зон. Степень заполнения энергетических зон зависит также от структуры кристалла

и природы химической связи. Так, например, углерод в структуре алмаза – диэлектрик, а углерод в структуре графита обладает свойствами проводника.

При $T = 0$ К уровень Ферми E_F определяет границу между заполненными и незаполненными электронными уровнями энергии. В чистых веществах – полупроводниках и диэлектриках E_F находится в запрещенной зоне, разделяющей валентную зону и зону проводимости; а в металлах и полуметаллах – в разрешенной зоне (рис. 2.7). При повышении температуры в полупроводниках и диэлектриках электроны переходят из валентной зоны в зону проводимости, образуя пустые места в валентной зоне, называемые *дырками*. Движение носителей заряда в валентной зоне обычно описывают как движение дырок. Каждой дырке приписывают заряд и волновой вектор, равные с обратным знаком заряду и волновому вектору отсутствующего электрона.

При $T > 0$ К степень заполнения электронами состояния с энергией E характеризуется Ферми-распределением, или *функцией распределения Ферми – Дирака* (Р. А. М. Дирак, английский физик, один из создателей квантовой механики, лауреат Нобелевской премии 1933 г.), описывающей распределение по энергетическим уровням фермионов (частиц с полуцелым спином) при условии, что взаимодействие между ними слабое и им можно пренебречь:

$$f_s(E) = \left[\exp\left(\frac{E - E_F}{kT}\right) + 1 \right]^{-1}. \quad (2.24)$$

График функции (2.24) при различных температурах изображен на рис. 2.8. Как видно из рисунка, кривая распределения в виде ступеньки при $T = 0$ К с повышением температуры изменяется вблизи $E = E_F$, т. е. появляется вероятность заселения электронами состояний, находящихся выше E_F .

Вероятность заполнения дырочных состояний описывают аналогичной функцией Ферми для дырок:

$$f_d(E) = 1 - f_s(E) = \left[\exp\left(\frac{E + E_F}{kT}\right) + 1 \right]^{-1}. \quad (2.25)$$

В полуметаллах движение носителей заряда в нижней из перекрывающихся зон также описывают как движение дырок.

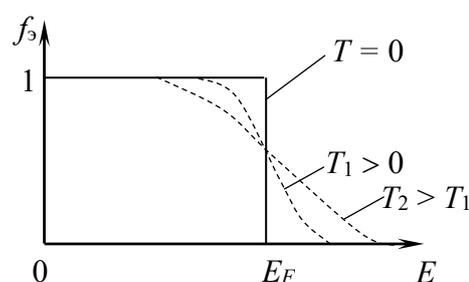


Рис. 2.8. Графическое изображение функции распределения Ферми – Дирака при различных T

2.4.3. Эффективная масса электрона

Закономерности движения свободного электрона и его движения в кристаллической решетке различны. В кристалле электрон находится в периодическом поле решетки и приобретает физические свойства, отличающие его от классической частицы. Однако в ряде случаев движение электронов в кристалле может быть описано как движение свободного электрона, на который действует только сила F (по закону Ньютона), но имеющего массу, отличную от массы свободного электрона.

Пусть на кристалл наложено внешнее электрическое поле напряженностью \mathcal{E}_0 . Оно действует на электрон в кристалле так же, как на свободный электрон, с силой $F = -e\mathcal{E}_0$, направленной против поля. Кроме нее на электрон действуют значительные внутренние силы, создаваемые периодическим полем решетки. Движение электрона в кристалле можно описать с помощью волнового пакета блоховских функций (2.16). Средняя скорость движения электрона равна групповой скорости распространения волн

$$v = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk}, \quad (2.26)$$

где ω – частота волны, k – волновой вектор, $E = \hbar\omega$ – энергия электрона.

За время dt внешняя сила F совершит работу, которая идет на приращение энергии электрона:

$$dE = v dt \cdot F = \frac{F}{\hbar} \cdot \frac{dE}{dk} dt. \quad (2.27)$$

Отсюда

$$F = \hbar \cdot \frac{dk}{dt}. \quad (2.28)$$

Продифференцировав выражение (2.26) для скорости волн v по времени, получим ускорение электрона

$$a = \frac{dv}{dt} = \frac{1}{\hbar} \cdot \frac{d}{dt} \left(\frac{dE}{dk} \right) = \frac{1}{\hbar} \cdot \frac{d^2 E}{dk^2} \frac{dk}{dt}. \quad (2.29)$$

Подставим сюда $\frac{dk}{dt}$ из формулы (2.28) и получим:

$$a = \frac{1}{\hbar^2} \cdot \frac{d^2 E}{dk^2} \cdot F, \text{ или } F = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1} \cdot a. \quad (2.30)$$

Уравнение (2.30) связывает ускорение электрона с внешней силой $F = -e\mathcal{E}_0$. Если предположить, что величина $\hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1}$ имеет смысл массы, то уравнение (2.30) приобретает вид второго закона Ньютона $F = m^* \cdot a$, где

$$m^* = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1}. \quad (2.31)$$

Величина m^* , имеющая размерность массы и характеризующая динамические свойства квазичастиц, получила название *эффективной массы электрона*. Эффективная масса отражает влияние поля решетки на движение электрона в кристалле под действием внешней силы. Из (2.30) следует, что электрон в периодическом поле кристаллической решетки движется под действием внешней силы так, как двигался бы свободный электрон под действием этой силы, если бы он обладал массой m^* .

Эффективная масса не является массой в ее обычном понимании, т. е. она не определяет ни гравитационные, ни инерционные свойства электрона. Она является лишь коэффициентом в уравнении движения (2.30) и отражает меру взаимодействия электрона с кристаллической решеткой. По величине эффективная масса может быть как больше, так и меньше массы покоящегося электрона (таблица 2.3), а по знаку – как положительной, так и отрицательной (рис. 2.9).

Таблица 2.3

Эффективные массы электронов и энергии Ферми (для некоторых металлов) и запрещенной зоны (для полупроводников)

Металл	m^*/m	E_F , эВ	Полупроводник	m^*/m	E_g , эВ
Be	1,60	12,0	Si – электроны	1,06	1,12
Cu	1,00	7,0	Si – дырки	0,56	
Al	0,97	11,8	Ge – электроны	0,22	0,67
Zn	0,85	11,0	Ge – дырки	0,39	

На рис. 2.9 приведены зависимости $E(k)$, $v(k)$ и $m^*(k)$ для одной из зон Бриллюэна (см. рис. 2.5, б). Минимум энергии соответствует центру зоны ($k = 0$, дно зоны), а максимум – ее границы ($k = \pm \frac{\pi}{a}$, потолок зоны). При

небольших значениях волнового числа k , пока кривая $E(k)$ остается параболой, скорость движения электрона растет с ростом k , а эффективная масса остается практически постоянной и равной массе покоя электрона. В точке A [точка перегиба зависимости $E(k)$] первая производная энергии по

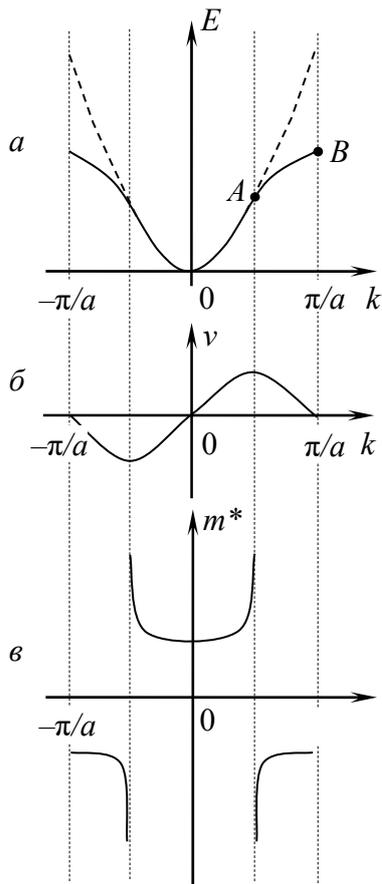


Рис. 2.9. Зависимость энергии (а), скорости (б) и эффективной массы (в) от волнового числа. Пунктир на рис. (а) – зависимость $E(k)$ для свободного электрона – формула (2.19, а)

перейдет электрон с более низкого энергетического уровня, а дырка опустится. Затем дырку займет следующий электрон и т. д. В результате дырка будет смещаться вниз по энергетической шкале. Таким образом, ток в кристаллах может быть обусловлен не только электронами в зоне проводимости, но и положительно заряженными дырками в валентной зоне. Дырочная проводимость наиболее характерна для полупроводников.

В анизотропных кристаллах кривизна зависимости $E(k)$ во многих точках k -пространства неодинакова в разных направлениях. В этом случае имеет место анизотропия эффективной массы, т. е. величина m^* является тензором. Так, в германии наибольшая величина анизотропии эффективной массы в зоне проводимости составляет: $m^* = 1,6m_0$ в направлении (111) и $0,08m_0$ – в перпендикулярном направлении, где m_0 – масса покоя электрона.

Понятие эффективной массы обобщают на другие типы квазичастиц (фононы, фотоны, экситоны и др.).

волновому числу $\frac{dE}{dk}$ максимальна, а вторая

производная обращается в нуль. Поэтому при $k = k_A$ скорость движения электрона (2.16) максимальна, а эффективная масса m^* (2.21) становится неопределенной. При значениях k , близких к границе зоны Бриллюэна, электроны обладают отрицательной эффективной массой. Это означает, что ускорение электрона направлено против внешней силы. Такой результат является следствием брэгговского отражения электрона от поверхности кристалла. В точке $k = \pm \frac{\pi}{a}$ движение электрона соответствует уже не бегущей, а стоячей волне, и его скорость $v = 0$.

Состояния с $m^* < 0$ обычно расположены вблизи «потолка» валентной зоны полупроводников, так называемые акцепторные уровни (см. п. 2.4.4). Для электронов с отрицательной эффективной массой оказалось удобным ввести в рассмотрение понятие дырки как квазичастицы с положительной эффективной массой и положительным электрическим зарядом. Если в валентной зоне все уровни, кроме одного, заняты электронами, валентное состояние вблизи «потолка» зоны является дыркой. Под действием внешнего поля на это вакантное состояние

2.4.4. Примесные уровни и поверхностные состояния

Приведенные в п. 2.4.2 закономерности касаются распределения электронов по энергиям в кристаллах, имеющих идеальную периодичность. Однако все твердые тела содержат дефекты и примеси. Присутствие в кристалле атома примеси или дефекта структуры приводит к тому, что на периодический потенциал решетки $U(\mathbf{r})$ накладывается достаточно сильное возмущение $U_0 = U(\mathbf{r}_0)$, локализованное в малой области объемом V_0 с центром в точке \mathbf{r}_0 , где расположен примесный атом или дефект. В этом случае одноэлектронное уравнение Шрёдингера (2.4) принимает вид:

$$\left[-\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) + U(\mathbf{r}_0) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (2.32)$$

Рассчитать положение уровней примесей или дефектов, даже если известен вид возмущения U_0 , практически невозможно, так как неизвестна точная функция $U(\mathbf{r})$. Однако, пользуясь понятием эффективной массы, уравнение (2.32) можно переписать в виде:

$$\left[-\frac{\hbar^2}{2m^*} \Delta + U(\mathbf{r}_0) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (2.33)$$

Здесь отсутствует периодический потенциал, а появившаяся эффективная масса электрона m^* может быть определена экспериментально.

Решая уравнение (2.33), находят разрешенные значения энергии электрона. Наложение на потенциал $U(\mathbf{r})$ возмущения приводит к смещению энергетических уровней электронов из разрешенных зон в запрещенную. На рис. 2.10 этот эффект схематически показан в виде отклонения уровней энергии от прямолинейности. При $U_0V_0 > 0$ уровень, соответствующий потолку разрешенной зоны, поднимается вверх, а все остальные уровни практически не меняют своего положения. Если $U_0V_0 < 0$, то уровень минимальной энергии опускается вниз. Таким образом, в запрещенной зоне появляются локализованные разрешенные уровни $E_{л1}$ и $E_{л2}$, возникшие из-за наличия примесей или дефектов.

Уровни $E_{л1}$, расположенные ниже дна зоны проводимости, называются *донорными*, а уровни $E_{л2}$ выше потолка валентной зоны – *акцепторными* (см. п. 3.3).

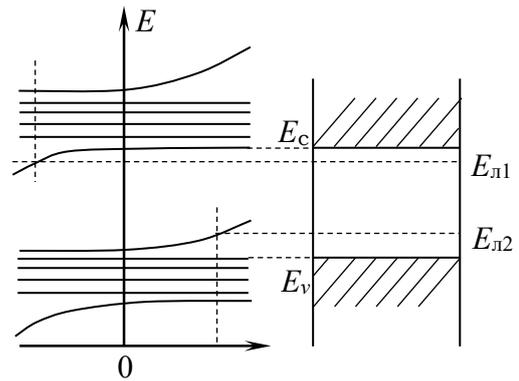


Рис. 2.10. Образование локализованных разрешенных уровней в запрещенной зоне кристалла под действием возмущений

Заполнение примесных уровней также определяется функцией распределения (2.24). В металлах, имеющих большое число свободных электронов, переходы носителей заряда с примесных уровней в зоны энергетического спектра не играют заметной роли. В полупроводниках и диэлектриках концентрация носителей заряда при не очень высоких температурах определяется числом электронов, перешедших с донорных уровней в зону проводимости, или числом электронов валентной зоны, захваченных акцепторами с образованием дырок.

Расстояние от примесного уровня до ближайшей разрешенной зоны может быть мало или сравнимо с шириной запрещенной зоны. В зависимости от этого различают *мелкие* и *глубокие* примесные уровни.

Мелкие уровни, образованные *примесью замещения* (замещение атома кристалла примесным атомом), являются донорными, если валентность примесного атома больше валентности основного элемента на единицу, и акцепторными – при обратном соотношении. Например, примесные атомы As (V группа периодической системы Менделеева) образуют в кристаллах Ge (IV группа) донорные уровни, а примесные атомы В (III группа) образуют в том же Ge акцепторные примесные уровни.

Глубокие примесные уровни обычно образуются при замещении атомов основного вещества атомами, отличающимися по валентности более чем на ± 1 . Такие примеси иногда способны создавать несколько уровней, соответствующих различным зарядовым состояниям образовавшихся примесных ионов. Например, атомы меди в решетке германия создают три примесных уровня, соответствующих ионам Cu^{1-} , Cu^{2-} и Cu^{3-} . Глубокие примесные уровни, образуемые разными ионами одного элемента, могут быть как донорными, так и акцепторными.

В случае *примесей внедрения* (дополнительный атом в междоузлии) донорная или акцепторная природа примесных уровней не зависит от валентности элементов, а определяется величиной их электроотрицательности. Если электроотрицательность примесных атомов больше, чем у основных атомов решетки, то примесный уровень является акцепторным, в обратном случае – донорным. Одна и та же примесь может быть источником донорных уровней при замещении и акцепторных – при внедрении (например, кислород в кремнии), либо наоборот.

Наряду с объемными уровнями в любом кристалле имеются электронные состояния, локализованные вблизи его поверхности, – *поверхностные состояния*. Различают *собственные* и *несобственные* поверхностные состояния. Собственные возникают вследствие обрыва кристаллической решетки на границе кристалла, несобственные – локализованы на примесях или дефектах, находящихся на поверхности кристалла, например, в слое

оксида, покрывающем кристалл. Собственные поверхностные состояния образуют разрешенные энергетические зоны, которые разделены запрещенными зонами, причем поверхностные разрешенные зоны могут располагаться в области энергий, соответствующих как запрещенным, так и разрешенным объемным энергетическим зонам.

На возможность существования поверхностных состояний впервые (1932 г.) указал И. Е. Тамм (советский физик-теоретик, лауреат Нобелевской премии 1958 г.). Поэтому собственные поверхностные состояния называются таммовскими (или *уровнями Тамма*). И. Е. Тамм определил электронный спектр ограниченной одномерной решетки, состоящей из прямоугольных потенциальных ям, разделенных прямоугольными барьерами. В отличие от модели Кронига – Пенни (см. п. 2.4.1) зависимость $U(x)$ в модели Тамма является периодической с периодом решетки внутри кристалла (т. е. в направлении $x > 0$), а вне кристалла ($x < 0$) $U(x) = U_0$ (рис. 2.11). Плоскость $x = 0$ соответствует поверхности кристалла, U_0 – потенциальная энергия электрона у поверхности. Появление таммовских состояний вызвано разной высотой потенциальных барьеров в объеме и на поверхности кристалла.

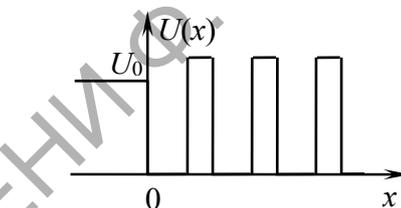


Рис. 2.11. Потенциальная энергия электрона в одномерной ограниченной цепочке атомов внутри кристалла

В кристаллических веществах таммовские состояния соответствуют ненасыщенным валентным связям поверхностных атомов. Обрыв этих связей обуславливает перестройку (*реконструкцию*) поверхности кристалла, т. е. смещение поверхностных атомов в плоскостях, касательной к поверхности и расположенной по нормали к ней. В поверхностном слое кристалла образуются структуры с периодом, равным нескольким периодам объемной решетки. Степень реконструкции поверхностного слоя зависит от кристаллографической ориентации поверхности, условий ее возникновения (например, от температуры отжига), а также от вида и концентрации адсорбированной примеси или наличия на поверхности кристалла оксидного слоя.

Несобственные поверхностные уровни электронов в кристалле связаны с наличием дефектов и чужеродных атомов на поверхности. Адсорбция чужеродных атомов или оксиды на поверхности кристалла изменяют спектр поверхностных состояний и, в частности, приводят к исчезновению собственных поверхностных состояний в области запрещенных зон полупроводников и появлению в этой области несобственных поверхностных состояний.

Для экспериментального изучения собственных поверхностных состояний используют поверхности кристаллов, полученные сколом в высоком

вакууме или жидком гелии, очищенные бомбардировкой ионами инертных газов с последующим отжигом в вакууме, а также пленки, получаемые методом молекулярной эпитаксии. Плотность поверхностных уровней в кристалле определяется числом одномерных цепочек атомов, выходящих на единицу площади поверхности кристалла. Она достигает $10^{15} - 10^{16} \text{ см}^{-2}$.

Поверхностные состояния обнаружены у многих металлов, полупроводников и диэлектриков. Поверхностные состояния, находящиеся в запрещенных зонах полупроводников, существенно влияют на электрические свойства последних: концентрацию и подвижность электронов в приграничных слоях, работу выхода полупроводника, поверхностную рекомбинацию электронов и т. п.

2.5. Энергетический спектр некристаллических твердых тел

Экспериментальные данные свидетельствуют о том, что несмотря на отсутствие дальнего порядка в некристаллических твердых телах, отдельные положения зонной теории могут быть применимы для их описания. Однако, в отличие от кристалла, рассеяние носителей заряда в неперриодическом поле некристаллического твердого тела столь интенсивно, что квазиимпульс \mathbf{p} – векторная характеристика состояния квазичастицы – в неперриодической среде не сохраняет признаков закономерного изменения даже приближенно. В связи с этим теряет смысл представление о законе дисперсии как функциональной зависимости между энергией и квазиимпульсом $E = f(\mathbf{p})$. Это значит, что при описании неупорядоченных твердых тел не «работают» понятия энергии Ферми и зоны Бриллюэна. Единственное понятие, одинаково пригодное для описания свойств кристаллических и некристаллических веществ, – это плотность состояний $N(E)$.

2.5.1. Плотность состояний

Плотность состояний – число возможных физических неэквивалентных состояний системы частиц в малом интервале энергии E , отнесенное к ширине интервала ΔE :

$$N(E) = \lim_{\Delta E \rightarrow 0} \Delta N(E) / \Delta E = dN(E) / dE, \quad (2.34)$$

где $\Delta N(E)$ – число состояний с энергиями между E и $E + \Delta E$.

Величина $dN(E) = N(E)dE$ представляет собой число состояний в единичном объеме твердого тела, в которых может находиться электрон

с заданным спином и с энергией в интервале от E до $E + dE$. В некристаллических твердых телах, так же как и в кристаллах, состояния могут быть свободны или заняты электронами, а число занятых состояний в единичном объеме вещества есть произведение $N(E)f(E)dE$, где $f(E)$ – функция распределения Ферми – Дирака (2.24).

Для кристаллического твердого тела, обладающего идеальной периодичностью, плотность состояний на краях зон резко уменьшается до нуля (рис. 2.12, *a*). Кроме того, следствием периодичности поля кристалла является то, что состояния не локализованы в пространстве, т. е. волновая функция электрона распространяется по всей решетке. Локальные нарушения периодичности, связанные с введением в кристалл атомов примесей или дефектов, приводят к появлению отдельных разрешенных состояний в запрещенной зоне (см. п. 2.4.4). В отличие от «зонных» состояний эти состояния локализованы в пространстве. Волновая функция электрона, находящегося в области одного из примесных центров, не «расплывается» по другим центрам, а экспоненциально спадает до нуля, т. е. остается локализованной.

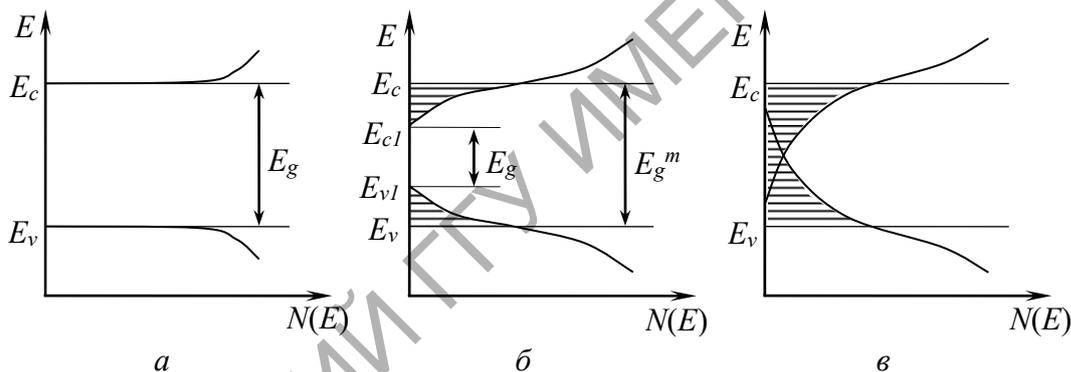


Рис. 2.12. Плотности состояний в энергетическом спектре кристалла (*a*) и некристаллических твердых тел (*б, в*). Области локализованных состояний заштрихованы. E_{cl} и E_{vl} – границы областей с высокими плотностями состояний

Анализ электронных свойств неупорядоченных твердых тел показал, что благодаря существованию ближнего порядка возможно их приближенное описание в терминах, разрешенных и запрещенных энергетических зон. В некристаллических телах локализованные электронные состояния возникают в запрещенной зоне. Их плотность не обращается в нуль на границе разрешенных зон, а монотонно убывает вглубь запрещенной зоны, образуя «хвосты» плотности состояний (рис. 2.12, *б, в*).

Представление о возможности локализации частиц и квазичастиц в неупорядоченных системах было впервые предложено в 1958 г. Ф. У. Андерсоном (Ph. W. Anderson, американский физик, лауреат Нобелевской премии 1977 г.). С его именем и именем Н. Ф. Мотта (N. F. Mott, английский

физик, лауреат Нобелевской премии совместно с Андерсоном) связано введение этих понятий в физику полупроводников, обусловившее дальнейшее развитие теории неупорядоченных систем.

Ф. Андерсон решил уравнение Шрёдингера для электрона в поле неупорядоченного потенциала $U(r)$. Оказалось, что при достаточно больших флуктуациях ΔU все электронные состояния являются локализованными (отсюда название локализованных состояний – *флуктуационные*). Н. Мотт постулировал, что существует критическое значение энергии E_c в зоне проводимости некристаллических тел, разделяющее локализованные и нелокализованные состояния. Аналогичное значение энергии E_v существует в валентной зоне.

В зависимости от природы некристаллического вещества возможны два вида распределения плотности состояний на энергии: 1) флуктуационные состояния, имеющие точные границы E_{c1} и E_{v1} (рис. 2.12, б); 2) спектры флуктуационных состояний накладываются друг на друга (рис. 2.12, в). Очевидно, что в первом случае представление о запрещенной зоне сохраняет точный смысл: имеется область энергий $E_g = E_{c1} - E_{v1}$, где плотность состояний равна нулю. Таким энергетическим спектром обладает ряд оптически прозрачных некристаллических веществ. Во втором случае весь энергетический интервал $E_c - E_v$ заполнен дискретными уровнями, т. е. запрещенная зона в том смысле, как она трактуется для кристаллических тел, здесь отсутствует. Тем не менее область $E_c - E_v$ принципиально отличается от разрешенных зон. Электроны, локализованные в этой зоне, могут участвовать в переносе заряда только путем перескоков. При $T \rightarrow 0$ К вероятность последних стремится к нулю, так что их вклад в электропроводность полностью исчезает. Поэтому область энергий, занятую локализованными состояниями, называют также запрещенной зоной.

Для обозначения этого интервала энергий используют еще один термин – «щель подвижности». Он обусловлен существованием локализованных и нелокализованных состояний, на которых подвижность электронов сильно отличается. При $E > E_c$ перенос зарядов осуществляется по нелокализованным состояниям, и подвижность электронов возрастает на несколько порядков. Резкие границы между этими состояниями E_c и E_v называют *краями подвижности*, а расстояние между ними – запрещенной зоной, или *щелью подвижности* E_g^m . Термины «зона проводимости» и «валентная зона» также используют в физике некристаллических веществ. Их относят к областям энергий, занятым нелокализованными состояниями электронов и дырок.

Наряду с флуктуационными состояниями в энергетических зонах некристаллических тел могут существовать локализованные состояния, связанные с примесными атомами и дефектами структуры. При их наличии

плотность состояний $N(E)$ немонотонно зависит от энергии частиц. Максимум плотности локализованных состояний, связанных с дефектами структуры, расположен обычно вблизи середины щели подвижности (рис. 2.13). При высокой плотности локализованных состояний уровень Ферми E_F расположен в зоне состояний, обусловленных дефектами.

Рассмотренные представления позволяют перенести представления зонной теории на некристаллические металлы, полупроводники и диэлектрики. Если уровень Ферми лежит в области нелокализованных состояний, то вещество представляет собой металл. Его сопротивление при $T \rightarrow 0$ К приближается к некоторому конечному значению. Если же

уровень Ферми при низких температурах находится в интервале энергий, занятом локализованными состояниями, то твердое тело представляет собой полупроводник или диэлектрик. Здесь возможны два механизма переноса носителей заряда: 1) по нелокализованным состояниям, если электроны возбуждены и перешли из локализованных состояний в нелокализованные; 2) путем «перескоков» электронов (дырок) из одного локализованного состояния в другое (прыжковая проводимость).

Таким образом, плотность состояний является основной характеристикой энергетического спектра неупорядоченных систем (жидкостей, стекол, аморфных веществ и др.). Плотность состояний можно определить экспериментально, например по фотоэлектронной эмиссии. Информация о плотности состояний важна при определении термодинамических характеристик твердых тел (теплоемкости, магнитной восприимчивости и др.), задаваемых интегралами по энергии от соответствующих микроскопических величин, умноженных на функцию распределения и плотность состояний. Плотность состояний влияет на кинетические характеристики твердых тел (электропроводность, теплопроводность и др.). Так, для электронов в металлах плотность состояний определяет положение и форму поверхности Ферми $N(E_F)$, для полупроводников – положения дна зоны проводимости и потолка валентной зоны.

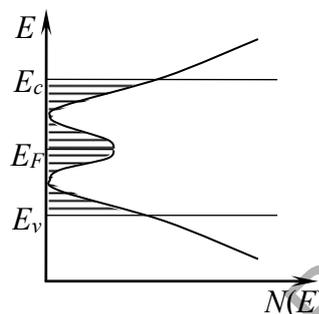


Рис. 2.13. Плотность состояний в некристаллических телах при наличии примесей или дефектов структуры. Области локализованных состояний заштрихованы

2.5.2. Границы применимости зонной теории

Зонная теория твердых тел исходит из следующих предположений:

а) потенциал кристаллической решетки строго периодичен;

б) взаимодействие между свободными электронами может быть описано с помощью одноэлектронного самосогласованного потенциала (одноэлектронное приближение);

в) взаимодействие электронов с фононами слабое и может быть рассмотрено в рамках теории возмущений, которая описывает взаимодействие носителей заряда в твердых телах с колебаниями кристаллической решетки – фононами.

Условие а) не выполняется в неупорядоченных системах, т. е. в конденсированных веществах, у которых отсутствует дальний порядок (жидкие, аморфные и стеклообразные вещества, а также твердые растворы). Однако размытие атомных уровней связано с перекрытием волновых функций соседних атомов, поэтому и в неупорядоченных средах образуются разрешенные зоны и квазизапрещенные зоны с резко пониженной плотностью электронных состояний.

В аморфном твердом теле спектр энергий электронов можно разделить на две области, в которых подвижность электронов $\mu \gg 0$ (нелокализованные или проводящие состояния) и $\mu \approx 0$ (локализованные или непроводящие состояния). Граница между этими областями является *порогом подвижности* (см. п. 2.5.1). Так, в аморфных полупроводниках локализованы электронные состояния, лежащие в запрещенной зоне, где плотность состояний относительно мала. Электроны, находящиеся в локализованном состоянии, могут переносить ток лишь путем «прыжков» из одного состояния в другое. При $T = 0$ К этот механизм не работает и электроны в локализованном состоянии вообще не могут создавать электрический ток. При более высоких температурах электропроводность неупорядоченных конденсированных сред осуществляется путем теплового «заброса» электронов в состояния, лежащие выше порога подвижности. Если уровень Ферми расположен выше порога подвижности, т. е. в области нелокализованных состояний, то механизм проводимости таких веществ такой же, как в металлах (например, в сильно легированных полупроводниках).

Условие б) справедливо для полупроводников и диэлектриков с малым числом свободных электронов, когда взаимодействие между ними слабое и представляет собой *межэлектронное* (электрон-электронное) *рассеяние* – процесс, при котором происходит передача энергии и импульса от одного электрона к другому, но их суммарная энергия и импульс сохраняются.

В металлах, где число свободных электронов велико, их взаимодействие характеризуют с помощью так называемого самосогласованного одноэлектронного потенциала. В любой многоэлектронной системе в результате электрон-электронного взаимодействия возникают элементарные возбуждения – квазичастицы. Это, во-первых, заряженные частицы – фермионы,

с помощью которых описывают самосогласованное движение всей системы электронов, и, во-вторых, бозоны, не связанные с переносом заряда (плазмоны, спиновые волны и др.). В полупроводниках и диэлектриках в результате взаимодействия электрона зоны проводимости и дырки валентной зоны образуются связанные состояния – экситоны Ванье – Мотта. В молекулярных кристаллах и диэлектриках возбужденные состояния атомов и молекул, перемещающиеся по кристаллу, образуют экситоны Френкеля. Волновая функция экситона удовлетворяет формуле (2.15), области разрешенных энергий экситона названы экситонными зонами.

Условие в) выполняется практически во всех металлах, где электрон-фононное взаимодействие ослаблено в результате экранирования свободными электронами, а также во многих полупроводниках. В полярных диэлектриках и полупроводниках с достаточно большой степенью ионной связи между частицами носители заряда, поляризуя решетку, образуют *автолокализованные состояния* (*автолокализация* – возникновение сильной деформации кристаллической решетки вокруг квазичастицы, приводящее к ее локализации в потенциальной яме, созданной деформацией). Обычно автолокализованные состояния имеют малый радиус, размер области локализации соизмерим с постоянной решетки. Это – поляроны в оксидах переходных металлов, автолокализованные дырки в щелочно-галогидных кристаллах, экситоны в кристаллах инертных газов и т. д.

Свободные и автолокализованные состояния квазичастиц в кристалле сосуществуют. Они разделены энергетическим барьером, связанным с затратой энергии на образование потенциальной ямы, которая может «захватить» квазичастицу. Превращение свободных квазичастиц в автолокализованные происходит при низких температурах путем туннелирования (перехода) через автолокализационный барьер, при более высоких температурах – по механизму термоактивационных переходов.

Вопросы

1. Каким объектам присущ корпускулярно-волновой дуализм?
2. Какая величина в уравнении Шрёдингера имеет физический смысл и какой?
3. На каких упрощениях базируется так называемое одноэлектронное приближение?
4. Какие электроны называют блоховскими?
5. Что представляет собой «первая зона Бриллюэна» и где используется это понятие?

6. Какие геометрические характеристики Ферми-поверхности связаны со свойствами металлов и полупроводников?

7. С точки зрения зонной теории в чем принципиальное различие металлов и полупроводников и в чем принципиальное сходство полупроводников и диэлектриков?

8. Что отражает эффективная масса электрона и в чем ее отличие от массы покоящегося электрона?

9. Какое понятие одинаково пригодно для описания свойств кристаллических и некристаллических твердых тел?

Глава 3

ОСНОВНЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

К полупроводникам относят класс веществ, характеризующихся значениями удельной электропроводности γ , промежуточными между удельной электропроводностью металлов и хороших диэлектриков. В отличие от металлов, электропроводность полупроводников всегда вынужденная и увеличивается с ростом температуры:

$$\gamma = \gamma_0 e^{\left[\frac{E_A}{kT} \right]},$$

где E_A – энергия активации проводимости, γ_0 – коэффициент (изменяется с температурой значительно медленнее, чем экспоненциальный множитель).

Эта формула означает, что в полупроводниках энергия связи электронов с атомами имеет порядок E_A . При повышении температуры тепловое движение разрывает связи электронов и часть их в количестве, пропорциональном экспоненте $\exp[-E_A/kT]$, становится свободными носителями заряда. Кроме того, удельная электропроводность имеет отрицательный температурный коэффициент и сильно зависит от концентрации примесных атомов и от их вида.

Связь электронов с атомами может быть разрушена также внешними воздействиями: светом, потоком быстрых частиц, сильным электрическим полем и др. Поэтому электропроводность полупроводниковых кристаллов чувствительна к внешним воздействиям, а также к содержанию дефектов и примесей в решетке. Во многих случаях энергия E_A электронов, локализованных вблизи примесей и дефектов, существенно меньше, чем в идеальном полупроводниковом кристалле. Возможность управлять в широких пределах электропроводностью полупроводников изменением температуры, напряженности внешних полей и введением примесей является основой их многочисленных и разнообразных применений.

3.1. Зонная структура полупроводников

Движение носителей заряда в полупроводниках описывает зонная теория твердого тела. В полупроводниках верхняя из заполненных разрешенных зон (валентная) отделена от самой низкой из незаполненных зон (зоны проводимости) энергетической щелью E_g (запрещенная зона). Тепловое движение «забрасывает» часть электронов из валентной зоны в зону проводимости, что обуславливает появление в валентной зоне дырок (рис. 3.1).

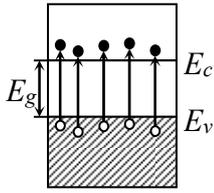


Рис. 3.1. Зонная структура полупроводника:

E_c – дно зоны проводимости;
 E_v – потолок валентной зоны;
 E_g – ширина запрещенной зоны;
 белые кружки – дырки; черные кружки – электроны проводимости

Электроны и дырки обычно сосредоточены вблизи дна зоны проводимости или потолка валентной зоны на энергетических расстояниях от них $\sim kT$, что гораздо меньше ширины разрешенных зон.

Ширина запрещенной зоны полупроводников меняется в широких пределах (таблица 3.1). Так, при $T \rightarrow 0$ К $E_g = 0,165$ эВ в PbSe и 5,6 эВ в алмазе, а серое олово – пример так называемого бесщелевого полупроводника, у которого $E_g = 0$.

Таблица 3.1

Ширина запрещенной зоны некоторых полупроводников при $T = 300$ К

Полупроводник	E_g , эВ	Полупроводник	E_g , эВ
Химические элементы		Соединения $A^{II}B^{VI}$	
Si	1,107	ZnS	3,6
Ge	0,66	CdS	2,42
Te	0,33	CdSe	1,8
Se	1,79	ZnSe	2,7
Соединения $A^{III}B^V$		Соединения $A^{IV}B^{VI}$	
GaAs	1,428	PbS	0,41
InAs	0,356	SnS	1,08
GaSb	0,78	PbTe	0,32
GaP	2,24	SnTe	0,18

3.2. Дрейфовая и диффузионная электропроводность

В полупроводниках процесс переноса зарядов может наблюдаться при наличии электронов в зоне проводимости и при неполном заполнении электронами валентной зоны. При выполнении этих условий и при отсутствии градиента температуры перенос носителей зарядов возможен либо под действием электрического поля, либо под действием градиента концентрации носителей заряда.

Дрейфовый ток. В полупроводниках свободные электроны и дырки находятся в состоянии хаотического движения. Поэтому, если выбрать произвольное сечение внутри объема полупроводника и подсчитать число носителей заряда, проходящих через это сечение за единицу времени слева

направо и справа налево, значения этих чисел окажутся одинаковыми. Это означает, что электрический ток в данном объеме полупроводника отсутствует.

При помещении полупроводника в электрическое поле напряженностью E на хаотическое движение носителей зарядов накладывается составляющая направленного движения. Направленное движение носителей зарядов в электрическом поле обуславливает появление тока, называемого дрейфовым (рис. 3.2, а). Из-за столкновения носителей зарядов с атомами кристаллической решетки их движение в направлении действия электрического поля прерывисто и характеризуется подвижностью μ . Подвижность равна средней скорости \bar{v} , приобретаемой носителями заряда в направлении действия электрического поля напряженностью $\mathcal{E} = 1$ В/м, т. е.

$$\mu = \frac{\bar{v}}{\mathcal{E}}. \quad (3.1)$$

Подвижность носителей зарядов зависит от механизма их рассеивания в кристаллической решетке (см. гл. 4). Исследования показывают, что подвижности электронов μ_n и дырок μ_p имеют различное значение ($\mu_n > \mu_p$) и определяются температурой и концентрацией примесей. Рост температуры приводит к уменьшению подвижности, что обусловлено увеличением числа столкновений носителей зарядов в единицу времени.

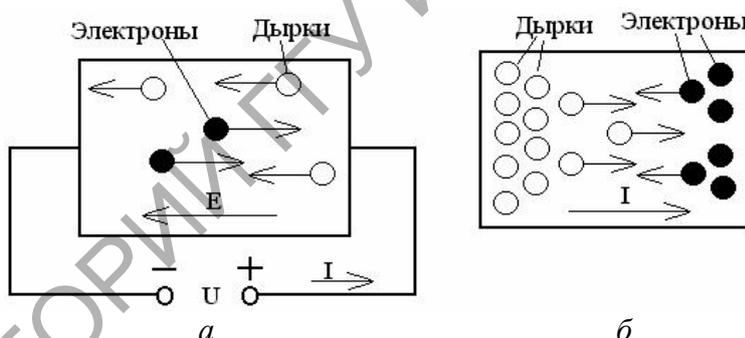


Рис. 3.2. Дрейфовый (а) и диффузионный (б) токи в полупроводнике

Плотность тока в полупроводнике, обусловленного дрейфом свободных электронов под действием внешнего электрического поля со средней скоростью \bar{v}_n , определяется выражением $j_n = en\bar{v}_n$. Перемещение (дрейф) дырок в валентной зоне со средней скоростью \bar{v}_p создает в полупроводнике дырочный ток, плотность которого $j_p = e\bar{v}_p$. Следовательно, полная плотность тока в полупроводнике содержит электронную j_n и дырочную j_p составляющие и равна их сумме (n и p – концентрации соответственно электронов и дырок). Подставляя в выражение для плотности тока соотношение для средней скорости электронов и дырок из (3.1), получаем

$$j = e(n\mu_n + p\mu_p)\mathcal{E}. \quad (3.2)$$

Если сравнить выражение (3.2) с законом Ома $j = \gamma \mathcal{E}$, то удельная электропроводность полупроводника определяется соотношением

$$\gamma = \gamma_n + \gamma_p = e(n\mu_n + p\mu_p). \quad (3.3)$$

У полупроводника с собственной электропроводностью концентрации электронов и дырок равны ($n_i = p_i$), и его удельная электропроводность определяется выражением

$$\gamma = en_i(\mu_n + \mu_p) = ep_i(n\mu_n + p\mu_p). \quad (3.4)$$

В полупроводнике n -типа $n_0 > p_0$, и его удельная электропроводность с достаточной степенью точности может быть определена выражением

$$\gamma \approx \gamma_n = e\mu_n n_0. \quad (3.5)$$

В полупроводнике p -типа $p_0 > n_0$, и удельная электропроводность такого полупроводника

$$\gamma \approx \gamma_p = e\mu_p p_0. \quad (3.6)$$

В области высоких температур концентрация электронов и дырок значительно возрастает за счет разрыва ковалентных связей и, несмотря на уменьшение их подвижности, электропроводность полупроводника увеличивается по экспоненциальному закону.

Диффузионный ток. Кроме теплового возбуждения, приводящего к возникновению равновесной концентрации зарядов, равномерно распределенных по объему полупроводника, обогащение полупроводника электронами до концентрации n_p и дырками до концентрации p_n может осуществляться его освещением, облучением потоком заряженных частиц, введением их через контакт (инжекцией) и т. д. В этом случае энергия возбудителя передается непосредственно носителям заряда, и тепловая энергия кристаллической решетки остается практически постоянной. Следовательно, избыточные носители заряда не находятся в тепловом равновесии с решеткой и поэтому называются *неравновесными*. В отличие от равновесных они могут неравномерно распределяться по объему полупроводника (рис. 3.2, б). После прекращения действия возбудителя за счет рекомбинации электронов и дырок концентрация избыточных носителей быстро убывает и достигает равновесного значения.

Скорость рекомбинации неравновесных носителей пропорциональна избыточной концентрации дырок ($p_n - p_{n0}$) или электронов ($n_p - n_{p0}$):

$$\frac{dp}{dt} = -(p_n - p_{n0})/\tau_p, \quad \frac{dn}{dt} = -(n_p - n_{p0})/\tau_n,$$

где τ_p – время жизни дырок; τ_n – время жизни электронов.

За время жизни концентрация неравновесных носителей уменьшается в $e = 2,7$ раза. Время жизни избыточных носителей составляет $0,01-0,001$ с. Носители зарядов рекомбинируют в объеме полупроводника и на его поверхности. Неравномерное распределение неравновесных носителей зарядов сопровождается их диффузией в сторону меньшей концентрации. Это движение носителей зарядов обуславливает прохождение электрического тока, называемого *диффузионным* (рис. 3.2, б).

Рассмотрим одномерный случай движения носителей заряда. Пусть в полупроводнике концентрации электронов $n(x)$ и дырок $p(x)$ являются функциями координаты. Это приведет к диффузионному движению дырок и электронов из области с большей их концентрацией в область с меньшей концентрацией. Диффузионное движение носителей зарядов обуславливает прохождение диффузионного тока электронов и дырок, плотности которых определяются из соотношений:

$$j_{n\text{диф}} = eD_n \frac{dn(x)}{dx}, \quad (3.7)$$

$$j_{p\text{диф}} = -eD_p \frac{dp(x)}{dx}, \quad (3.8)$$

где $dn(x)/dx$, $dp(x)/dx$ – градиенты концентраций электронов и дырок; D_n , D_p – коэффициенты диффузии электронов и дырок.

Градиент концентрации характеризует степень неравномерности распределения зарядов (электронов и дырок) в полупроводнике вдоль какого-то выбранного направления (в данном случае вдоль оси x). Коэффициенты диффузии показывают количество носителей заряда, пересекающих в единицу времени единичную площадку, перпендикулярную к выбранному направлению, при градиенте концентрации в этом направлении, равном единице. Коэффициенты диффузии связаны с подвижностями носителей зарядов соотношениями Эйнштейна:

$$D_n = \mu_n kT / e, \quad D_p = \mu_p kT / e. \quad (3.9)$$

Знак «минус» в выражении (3.8) означает противоположную направленность электрических токов в полупроводнике при диффузионном движении электронов и дырок в сторону уменьшения их концентраций.

Если в полупроводнике существует и электрическое поле, и градиент концентрации носителей, проходящий ток будет иметь дрейфовую и диффузионную составляющие. В таком случае плотности токов рассчитываются по следующим уравнениям:

$$j_n = en\mu_n E + eD_n \frac{dn(x)}{dx}, \quad j_p = ep\mu_p E + eD_p \frac{dp(x)}{dx}. \quad (3.10)$$

3.3. Собственная и примесная проводимость

Электропроводность полупроводников может быть обусловлена собственными электронами атомов данного вещества (собственная проводимость) и электронами примесных атомов (примесная проводимость). Источниками носителей тока могут быть также дефекты кристаллической структуры (вакансии, междоузельные атомы и др.).

Собственная проводимость. Электропроводность полупроводника, не содержащего примесей и дефектов, при $T = 0$ К равна нулю, поскольку в нем нет свободных носителей заряда. Действительно, его валентная зона полностью заполнена электронами и не дает никакого вклада в проводимость, а зона проводимости пуста. При $T > 0$ К появляется вероятность заброса электронов из валентной зоны в зону проводимости (см. рис. 3.1). В результате в валентной зоне образуются дырки, причем концентрации электронов n и дырок p будут одинаковы. Одновременно с процессом образования свободных носителей (*генерация*) идет процесс их исчезновения (*рекомбинация*), т. е. возврат части электронов из зоны проводимости в валентную зону и заполнение дырок. При температуре, отличной от абсолютного нуля, в полупроводниках устанавливается некоторая равновесная концентрация носителей заряда. Так, например, при комнатной температуре концентрация свободных электронов и дырок составляет в Si примерно 10^{10} см⁻³, а в Ge $\sim 10^{13}$ см⁻³.

При наложении электрического поля E в полупроводнике возникает ток, состоящий из электронной и дырочной составляющих. Полупроводники, в которых за счет перехода некоторого количества электронов из валентной зоны в зону проводимости образуется такое же количество дырок, называются *собственными*. Соответственно их проводимость, состоящую из электронной и дырочной составляющих, называют *собственной проводимостью*.

Результирующая электропроводность собственного полупроводника определяется суммой электронной и дырочной компонент:

$$\gamma = en\mu_n + ep\mu_p, \quad (3.11)$$

где μ_n и μ_p – подвижности электронов и дырок соответственно.

В уравнение (3.11) входят два важнейших параметра полупроводника – концентрация (n , p) и подвижность носителей зарядов.

Концентрация носителей (электронов и дырок) в собственном полупроводнике увеличивается с ростом температуры по экспоненциальному закону:

$$n_i = A \cdot \exp\left(-\frac{E_g}{2kT}\right), \quad (3.12)$$

где A – коэффициент, слабо зависящий от температуры.

Энергия активации этого процесса равна половине ширины запрещенной зоны $E_g/2$. Температурная зависимость концентрации собственных носителей, построенная в координатах $\ln n_i$ от $1/T$, имеет вид:

$$\ln n_i = \text{const} - \frac{E_g}{2k} \cdot \frac{1}{T} \quad (3.13)$$

и представляет собой практически прямую линию (рис. 3.3). Тангенс угла наклона этой прямой к оси абсцисс равен половине ширины запрещенной зоны: ($\text{tg}\varphi = -E_g/2$), т. е. численно равен энергии активации.

Подвижность носителей заряда в собственном полупроводнике, где нет никаких примесей и дефектов, определяется рассеянием носителей на фонах. Поскольку средняя длина свободного пробега электрона обратно пропорциональна концентрации фононов, которая, в свою очередь, в области высоких температур пропорциональна температуре, то $\mu_n \sim 1/T$.

При низких температурах подвижность носителей практически не зависит от температуры. Таким образом, электропроводность собственных полупроводников изменяется от температуры по экспоненциальному закону.

Примесная проводимость. Примеси и дефекты в полупроводниках делятся на донорные и акцепторные. *Доноры* отдают в объем полупроводника избыточные электроны и создают электронную проводимость (*n*-типа). *Акцепторы* захватывают валентные электроны вещества, в результате чего создаются дырки и возникает дырочная проводимость (*p*-типа). Наличие в полупроводниках примесных атомов и структурных дефектов обуславливает существование в запрещенной зоне локальных разрешенных энергетических состояний – *примесных уровней*.

Если в полупроводник введена донорная или акцепторная примесь, то при низких температурах, когда энергии тепловых колебаний недостаточно для переброса электронов из валентной зоны в зону проводимости, свободные носители заряда могут появиться за счет ионизации примесей или дефектов. Энергия ионизации мелких доноров и акцепторов незначительна по сравнению с шириной запрещенной зоны полупроводника, поэтому связанный с донорным атомом электрон может быть оторван от этого атома, т. е. переведен с донорного уровня в зону проводимости (рис. 3.4, а). Чем выше температура, тем больше доноров отдают свои электроны. Правда, с повышением температуры увеличивается и вероятность межзонных переходов. Поэтому количество электронов в зоне проводимости полупроводника, содержащего доноры, может значительно превышать количество

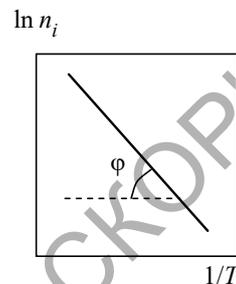


Рис. 3.3. Зависимость концентрации носителей в собственном полупроводнике от температуры

дырок в валентной зоне, и электропроводность будет электронной. Такой полупроводник называют *электронным (донорным)*, или полупроводником *n-типа*. Электроны являются в нем *основными носителями заряда*, а дырки – *неосновными*.

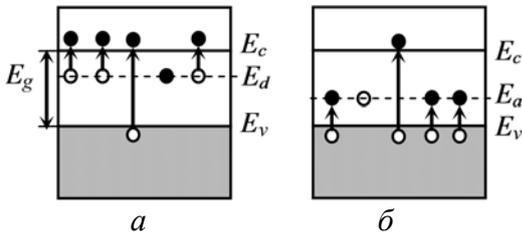


Рис. 3.4. Зонная структура (энергетическая диаграмма) электронного (а) и дырочного (б) полупроводников

В полупроводнике, содержащем акцепторную примесь, электроны легко переходят из валентной зоны на акцепторные уровни, а в валентной зоне образуются дырки (рис. 3.4, б). Их количество значительно больше, чем количество свободных электронов, образовавшихся за счет переходов из валентной зоны в зону проводимости. Поэтому дырки

являются основными носителями заряда, а электроны – неосновными. Проводимость полупроводника, содержащего акцепторную примесь, будет дырочной, а сам полупроводник в соответствии с этим называется *дырочным (акцепторным)*, или полупроводником *p-типа*.

Положение примесных уровней в запрещенной зоне полупроводника определяется энергией ионизации атомов примесей. В качестве примера рассмотрим элементарные полупроводники IV группы таблицы Менделеева – Ge и Si. Донорные примесные уровни в них создают элементы V группы (P, As, Sb, Bi), а акцепторные уровни – элементы III группы (B, Al, Ga, In).

Предположим, что в одном из кристаллов германия находится атом мышьяка (As), имеющий пять электронов в валентной оболочке. Четыре валентных электрона участвуют в образовании ковалентных связей с четырьмя атомами германия. Поскольку ковалентная связь является насыщенной, пятый электрон новую связь образовывать не может. Он сравнительно слабо взаимодействует с большим числом окружающих мышьяк атомов германия, вследствие чего его связь с атомом мышьяка уменьшается, и он движется по орбите большого радиуса (рис. 3.5, а). Его движение можно сравнить с движением электрона в атоме водорода.

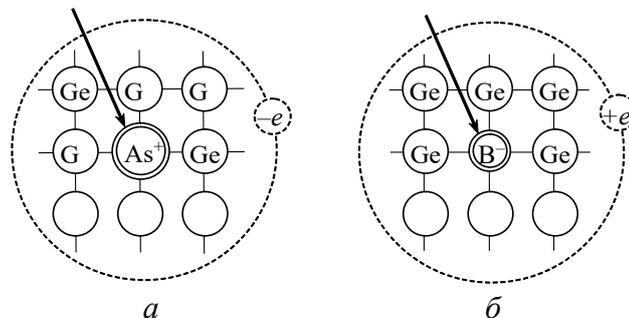


Рис. 3.5. Схема кристаллической структуры германия с атомом донорной (а) и акцепторной (б) примеси

Таким образом, задача определения энергии ионизации примесного атома сводится к отысканию уровней энергии водородоподобного атома. При ее решении необходимо учесть следующие обстоятельства.

Поскольку электрон движется не только в кулоновском поле иона As^+ , но и в периодическом поле кристаллической решетки полупроводника, он обладает эффективной массой m^* (см. п. 2.4.3). Взаимодействие электрона с ионом мышьяка, имеющим заряд $Z \cdot e$ (Z – кратность заряженного иона), происходит в твердом теле с диэлектрической проницаемостью ϵ . В этом случае решение уравнения Шрёдингера для пятого электрона мышьяка дает собственные значения энергии этого электрона (в эВ):

$$E_n = E_c - \frac{E_{\text{ион}} Z^2}{\epsilon^2} \left(\frac{m^*}{m} \right) \frac{1}{n^2}. \quad (3.14)$$

Здесь энергия отсчитывается от дна зоны проводимости E_c ; $n \geq 1$ – квантовое число; $E_{\text{ион}} = 13,52$ эВ – энергия ионизации атома водорода.

Энергия ионизации примесного атома E_d (рис. 3.3, а) равна по абсолютному значению энергии основного состояния электрона ($n = 1$):

$$E_d = \frac{E_{\text{ион}} Z^2}{\epsilon^2} \left(\frac{m^*}{m} \right). \quad (3.15)$$

Из формулы (3.15) следует, что: 1) E_d значительно меньше энергии ионизации атома водорода; 2) E_d растет с увеличением Z , поэтому уровень двукратно заряженного иона примеси лежит в запрещенной зоне полупроводника ниже уровня однократно заряженного иона.

Расчетное значение энергии ионизации примесных атомов в кристалле германия ($\epsilon \approx 16$, $m^* = 0,25 m_0$) составляет $E_d \approx 0,01$ эВ, в кристалле кремния – $\sim 0,04$ эВ. Таким образом, достаточно совсем незначительной энергии, чтобы перевести пятый электрон мышьяка из связанного состояния в свободное, т. е. в зону проводимости. Экспериментально определенные значения энергии ионизации доноров в германии и кремнии (таблица 3.2) достаточно хорошо согласуются с расчетными величинами E_d .

Таблица 3.2

Энергия ионизации доноров в кремнии и германии

Донор	E_d , эВ	
	в Si	в Ge
P	0,045	0,0128
As	0,053	0,0140
Sb	0,043	0,0098
Bi	0,069	0,0125

Пользуясь моделью водородоподобного атома, можно оценить размеры области, в которой локализована волновая функция электрона, связанного с примесным атомом:

$$a_n = a_0 \left(\frac{m}{m^*} \right) \varepsilon n^2, \quad (3.16)$$

где $a_0 = 0,53 \cdot 10^{-10}$ м – радиус первой боровской орбиты атома водорода.

Для донорной примеси в германии получаем $a_1 = 64a_0 \approx 34 \cdot 10^{-10}$ м. Поскольку постоянная решетки германия равна примерно 0,56 нм, первая боровская орбита электрона примесного атома охватывает около 200 узлов решетки. Радиус первой боровской орбиты электрона примеси V группы в кремнии несколько меньше, однако тоже достаточно велик ($a_1 \approx 30 a_0$).

Положение акцепторных уровней в запрещенной зоне полупроводника оценивают аналогичным образом. Предположим, что один из узлов решетки германия замещен примесным атомом III группы таблицы Менделеева, например атомом бора (рис. 3.5, б). Три валентных электрона атома бора образуют три ковалентные связи с соседними атомами германия, а четвертая связь остается незавершенной. Она представляет собой дырку с положительным зарядом $+e$, что обеспечивает электронейтральность области кристалла, окружающей примесь. Атом бора может захватить незавершенной связью электрон основного вещества и превратиться в отрицательно заряженный ион.

Собственные значения энергии для акцепторной примеси можно вычислить таким же путем, каким их определяют для донорной примеси:

$$E_n = E_v + \frac{E_{\text{ион}} Z^2}{\varepsilon^2} \left(\frac{m^d}{m} \right) \frac{1}{n^2}, \quad (3.17)$$

где m^d – эффективная масса дырки.

Основной уровень акцепторной примеси ($n = 1$)

$$E_a = \frac{E_{\text{ион}} Z^2}{\varepsilon^2} \left(\frac{m^d}{m} \right) \quad (3.18)$$

находится в запрещенной зоне германия примерно на 0,01 эВ выше потолка валентной зоны. В таблице 3.3 приведены экспериментальные значения E_a примесей III группы в германии и кремнии, отсчитанные от потолка валентной зоны.

Из формул (3.14) и (3.17) следует, что наряду с основными примесными уровнями в запрещенной зоне могут существовать возбужденные примесные состояния, соответствующие значениям квантового числа $n = 2, 3, 4, \dots$. Они расположены выше основного донорного состояния или ниже

основного акцепторного (рис. 3.6). Экспериментально такие состояния были обнаружены в запрещенной зоне кремния при температуре жидкого гелия. Однако при обычных температурах возбужденные состояния можно не учитывать, так как из-за малой величины E_d и E_a они не оказывают существенного влияния на электрические свойства полупроводников.

Таблица 3.3

Энергия ионизации акцепторов в кремнии и германии

Акцептор	E_a , эВ	
	в Si	в Ge
B	0,045	0,0104
Al	0,057	0,0102
Ga	0,065	0,0108
In	0,016	0,0102

В электронном полупроводнике, содержащем донорную примесь с концентрацией N_d , при $T = 0$ К уровень Ферми E_F лежит посередине между дном зоны проводимости и примесным уровнем. С ростом температуры уровень Ферми смещается к середине запрещенной зоны (рис. 3.7). Это снижение приводит к тому, что концентрация электронов в зоне проводимости увеличивается. Постепенно концентрация n электронов в зоне проводимости становится сравнимой с N_d . Равенство $n = N_d$ означает, что вся донорная примесь ионизована. Область температур, в которой выполняется это равенство, называется *областью истощения примеси*. Полная ионизация примеси наступает, когда E_F опускается на несколько kT ниже E_d .

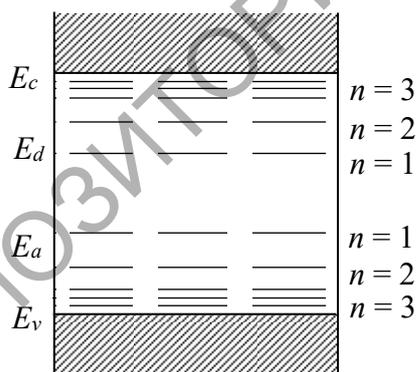


Рис. 3.6. Основные и возбужденные примесные состояния в запрещенной зоне полупроводника

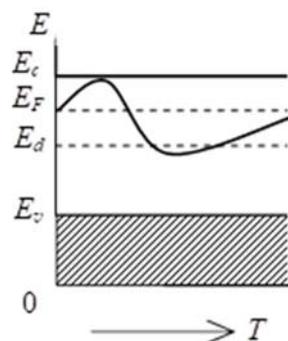


Рис. 3.7. Положение уровня Ферми в донорном полупроводнике в зависимости от температуры

Дальнейшее повышение температуры вызывает увеличение концентрации электронов в зоне проводимости за счет переходов в нее электронов

из валентной зоны. Температурная зависимость концентрации электронов приведена на рис 3.8. В области высоких температур (область 3) она такая же, как у собственных полупроводников (см. рис. 3.3).

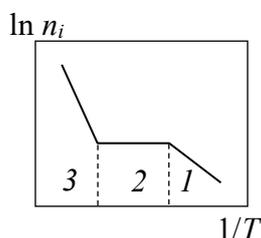


Рис. 3.8. Зависимость концентрации электронов от температуры в донорном полупроводнике: 1 – область ионизации примесей; 2 – область истощения примесей; 3 – переходы электронов из валентной зоны в зону проводимости (собственная проводимость)

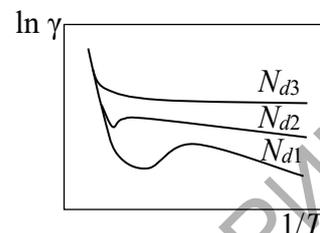


Рис. 3.9. Зависимость удельной электропроводности примесного полупроводника от температуры при концентрациях донорных примесей $N_{d1} < N_{d2} < N_{d3}$

В примесных полупроводниках подвижность носителей заряда ограничена рассеянием не только на фононах, но и на ионизованных атомах примесей. При высоких температурах доминирует рассеяние на фононах. Подвижность μ , обусловленная рассеянием на ионизованной примеси, пропорциональна $T^{3/2}$.

Таким образом, температурная зависимость электропроводности примесного полупроводника немонотонна: в интервале температур, где концентрация носителей экспоненциально зависит от температуры, $\gamma(T)$ также практически является экспоненциальной функцией, а в области истощения примесей ход кривой $\gamma(T)$ определяется температурной зависимостью подвижности носителей. Типичный вид температурной зависимости электропроводности примесного полупроводника с разной концентрацией донорных примесей N_{di} показан на рис. 3.9.

3.4. Основные электрофизические свойства полупроводников

Основные электрофизические свойства важнейших полупроводниковых материалов (ширина запрещенной зоны, подвижность носителей тока, температура плавления и т. д.) представлены в таблице 3.4. Ширина запрещенной зоны E_g является одним из фундаментальных параметров полупроводниковых материалов. Чем больше E_g , тем выше допустимая рабочая температура и тем более сдвинут в коротковолновую область спектра рабочий диапазон приборов, создаваемых на основе соответствующих полупро-

водниковых материалов. Например, максимальная рабочая температура германиевых приборов не превышает 50–60 °С, для кремниевых приборов она возрастает до 150–170 °С, а для приборов на основе GaAs достигает 250–300 °С; длинноволновая граница собственной фотопроводимости составляет: для InSb – 5,4 мкм (при 77 К), InAs – 3,2 мкм (195 К), Ge – 1,8 мкм (300 К), Si – 1 мкм (300 К), GaAs – 0,92 мкм (300 К). Величина E_g хорошо коррелирует с температурой плавления. Обе эти величины возрастают с ростом энергии связи атомов в кристаллической решетке, поэтому для широкозонных полупроводниковых материалов характерны высокие температуры плавления, что создает большие трудности на пути создания чистых и структурно совершенных монокристаллов таких полупроводниковых материалов. Подвижность носителей тока в значительной мере определяет частотные характеристики полупроводниковых приборов. Для создания приборов СВЧ диапазона необходимы полупроводниковые материалы, обладающие высокими значениями μ . Аналогичное требование предъявляется и к полупроводниковым материалам, используемым для изготовления фотоприемников. Температура плавления и период кристаллической решетки, а также коэффициент линейного термического расширения играют первостепенную роль при конструировании гетероэпитаксиальных композиций. Для создания совершенных гетероструктур желательно использовать полупроводниковые материалы, обладающие одинаковым типом кристаллической решетки и минимальными различиями в величинах ее периода и коэффициентах термического расширения. Плотность полупроводниковых материалов определяет такие важные технические характеристики, как удельный расход материала, масса прибора.

Таблица 3.4

Основные электрофизические свойства полупроводников

Элемент, тип соединения (по номеру группы)	Наименование полупроводника	Ширина запрещенной зоны, эВ		Подвижность носителей заряда при 300 К, см ² /(В·сек)		Кристаллическая структура	Постоянная решетки, Å	$T_{пл}$, °С
		при 300 К	при 0 К	электроны	дырки			
Элемент	C (алмаз)	5,47	5,51	2800	2100	алмаз	3,56679	4027
	Ge	0,661	0,89	3900	1900	типа алмаза	5,65748	937
	Si	1,12	1,16	1500	600	типа алмаза	5,43086	1420
	α -Sn		~0,08			типа алмаза	6,4892	

Окончание таблицы 3.4

Элемент, тип соединения (по номеру группы)	Наименование полупроводника	Ширина запрещенной зоны, эВ		Подвижность носителей заряда при 300 К, см ² /(В·сек)		Кристаллическая структура	Постоянная решетки, Å	T _{пл} , °С
		при 300 К	при 0 К	электроны	дырки			
IV—IV	α-SiC	3	3,1	400	50	типа сфалерита	4,358	3100
III—V	AlSb	1,63	1,75	200	420	типа сфалерита	6,1355	1050
	BP	6				типа сфалерита	4,538	>1300
	GaN	3,39		440	200	типа вюртцита	3,186 (по оси а) 5,176 (по оси с)	>1700
	GaSb	0,726	0,80	2500	680	типа сфалерита	6,0955	706
	GaAs	1,424	1,52	8500	400	типа сфалерита	5,6534	1239
	GaP	2,27	2,40	110	75	типа сфалерита	5,4505	1467
	InSb	0,17	0,26	78 000	750	типа сфалерита	6,4788	525
	InAs	0,354	0,46	33 000	460	типа сфалерита	6,0585	943
	InP	1,34	1,34	4600	150	типа сфалерита	5,8688	1060
II—VI	CdS	2,42	2,56	300	50	типа вюртцита	4,16 (по а) 6,756 (по с)	1750
	CdSe	1,7	1,85	800		типа сфалерита	6,05	1258
	ZnO	3,36		200		кубическая	4,58	1975
	ZnS	3,6	3,7	165		типа вюртцита	3,82 (по а) 6,26 (по с)	1700
IV—VI	PbS	0,41	0,34	600	700	кубическая	5,935	1103
	PbTe	0,32	0,24	1700	840	кубическая	6,460	917

3.5. Получение и легирование полупроводников

Получение. Необходимым условием достижения высоких электрофизических характеристик полупроводниковых материалов является их глубокая очистка от посторонних примесей. В случае Ge и Si эта проблема решается путем синтеза их летучих соединений (хлоридов, гидридов) и последующей глубокой очистки этих соединений с применением методов ректификации, сорбции, частичного гидролиза и специальных термических обработок. Хлориды особой чистоты подвергают затем высокотемпературному восстановлению водородом, прошедшим предварительную глубокую очистку, с осаждением восстановленных продуктов на кремниевых или германиевых прутках. Из очищенных гидридов Ge и Si выделяют путем термического разложения. В результате получают Ge и Si с суммарным содержанием остаточных электрически активных примесей на уровне 10^{-7} – 10^{-9} %. Особо чистые полупроводниковые соединения получают синтезом из элементов, прошедших глубокую очистку. Суммарное содержание остаточных примесей в исходных материалах не превышает обычно 10^{-4} – 10^{-5} %. Синтез разлагающихся соединений проводят либо в запаянных кварцевых ампулах при контролируемом давлении паров летучего компонента в рабочем объеме, либо под слоем жидкого флюса (например, особо чистого обезвоженного B_2O_3). Синтез соединений, имеющих большое давление паров летучего компонента над расплавом, осуществляют в камерах высокого давления. Часто процесс синтеза совмещают с последующей дополнительной очисткой соединений путем направленной или зонной кристаллизации расплава.

Наиболее распространенный способ получения монокристаллов полупроводниковых материалов – вытягивание из расплава по *методу Чохральского*. Этим методом получают монокристаллы Ge, Si, соединения типа $A^{III}B^V$, $A^{II}B^{VI}$, $A^{IV}B^{VI}$ и т. д. Вытягивание монокристаллов неразлагающихся полупроводниковых материалов проводят в атмосфере H_2 , инертных газов или в условиях глубокого вакуума. При выращивании монокристаллов разлагающихся соединений (InAs, GaAs, InP, GaP, CdTe, PbTe и др.) расплав герметизируют слоем жидкого флюса (B_2O_3) и вытягивают монокристаллы, погружая затравку в расплав через флюс и поддерживая в рабочем объеме над расплавом определенное давление инертного газа. Часто вытягивание осуществляют в камерах высокого давления, совмещая выращивание монокристаллов с предварительным синтезом соединений под слоем флюса (GaAs, InP, GaP и др.).

Для выращивания монокристаллов полупроводниковых материалов также широко используют методы *направленной и зонной кристаллизации*

расплава в контейнере. В случае разлагающихся соединений для получения монокристаллов требуемого стехиометрического состава процесс проводят в запаянных кварцевых ампулах, поддерживая равновесное давление паров летучего компонента над расплавом; часто для этого требуются камеры высокого давления, в которых поддерживается противодействие инертного газа. При получении монокристаллов необходимой кристаллографической ориентации используют ориентированные соответствующим образом монокристаллические затравки.

Для выращивания монокристаллов полупроводниковых материалов, обладающих подходящим сочетанием плотности и поверхностного натяжения расплава, используют *метод бестигельной зонной плавки*. Наибольшее распространение этот метод получил в технологии получения монокристаллов Si, имеющего сравнительно невысокую плотность и достаточно большое поверхностное натяжение расплава. Отсутствие контакта расплава со стенками контейнера позволяет получать этим методом наиболее чистые монокристаллы. Обычно процесс выращивания монокристалла совмещают с предварительной дополнительной очисткой полупроводниковых материалов зонной плавкой.

Для получения монокристаллов ряда тугоплавких разлагающихся полупроводниковых соединений (например, CdS, ZnS, SiC, AlN и др.) используют *кристаллизацию из газовой фазы* (методы сублимации и химических транспортных реакций). В случае если при выращивании монокристаллов не удастся получить соединение требуемого стехиометрического состава, кристаллы разрезают на пластины, которые подвергают дополнительному отжигу в парах недостающего компонента. Наиболее часто этот прием используют в технологии получения монокристаллов узкозонных соединений типа $A^{II}B^{VI}$ и $A^{IV}B^{VI}$, где собственные точечные дефекты сильно влияют на концентрацию и подвижность носителей тока, т. е. проявляют высокую электрическую активность (PbTe, $Pb_xSn_{1-x}Te$, $Cd_xHg_{1-x}Te$ и др.). При этом удается снизить концентрацию носителей заряда в кристаллах на несколько порядков. Для выращивания профилированных монокристаллов полупроводниковых материалов (ленты, прутки, трубы и т. д.) используют метод Степанова.

Широко распространено получение полупроводниковых материалов в виде монокристаллических пленок на разного рода монокристаллических подложках. Такие пленки называют эпитаксиальными, а процессы их получения – *эпитаксиальным наращиванием*. Возможности получения тонких и сверхтонких однослойных и многослойных структур разнообразной геометрии с широкой вариацией состава и электрофизических свойств по толщине и поверхности наращиваемого слоя, с резкими границами *p-n*-переходов

и гетеропереходов обуславливают широкое использование методов эпитаксиального наращивания в микроэлектронике и интегральной оптике, в практике создания больших и быстродействующих интегральных схем, а также оптоэлектронных приборов.

Для получения эпитаксиальных структур полупроводниковых материалов применяют методы жидкостной, газофазной и молекулярно-пучковой эпитаксии. В качестве растворителя обычно используют расплав нелетучего компонента соответствующего соединения. Наращивание эпитаксиального слоя проводят либо в режиме программируемого снижения температуры, либо из предварительно переохлажденного расплава. Этим методом можно воспроизводимо получать многослойные структуры с толщинами отдельных слоев до $\sim 0,1$ мкм при толщинах переходных слоев на гетерограницах порядка десятков нм.

Легирование. Для получения полупроводниковых материалов электронного типа проводимости (*n*-типа) с изменяющейся в широких пределах концентрацией носителей заряда (электронов) обычно используют донорные примеси, образующие «мелкие» энергетические уровни в запрещенной зоне вблизи дна зоны проводимости (энергия ионизации $\leq 0,05$ эВ). Для полупроводниковых материалов дырочного типа проводимости (*p*-типа) аналогичная задача решается путем введения акцепторных примесей, образующих «мелкие» энергетические уровни в запрещенной зоне вблизи потолка валентной зоны. Такие примеси при комнатной температуре практически полностью ионизованы, так что их концентрация приблизительно равна концентрации носителей заряда. Для Ge и Si основными донорными легирующими примесями являются элементы V группы периодической системы: P, As, Sb, а акцепторными – элементы III группы: B, Al, Ga. Для соединений типа $A^{III}B^V$ – соответственно примеси элементов VI группы (S, Se, Te), а также Sn, и элементов II группы (Be, Mg, Zn, Cd). Элементы IV группы (Si, Ge) в зависимости от условий получения кристаллов и эпитаксиальных слоев соединений типа $A^{III}B^V$ могут проявлять как донорные, так и акцепторные свойства. В соединениях типа $A^{II}B^{VI}$ и $A^{IV}B^{VI}$ поведение вводимых примесей сильно осложняется присутствием собственных точечных структурных дефектов. Необходимые тип и величина проводимости в них обычно достигаются прецизионным регулированием отклонения состава от стехиометрического, обеспечивающего заданную концентрацию определенного типа собственных точечных дефектов структуры в кристаллах.

Перечисленные выше легирующие примеси образуют, как правило, твердые растворы замещения и обладают достаточно высокой растворимостью (10^{18} – 10^{20} атомов/см³) в широком интервале температур. При этом максимум растворимости приходится на температурный интервал 700–900 °С

в Ge, 1200–1350 °C в Si и 1100–1200 °C в GaAs. Представляется важным знание *предельной растворимости* примесей в материале (в твердой фазе). Под этим термином подразумевается концентрация примеси в насыщенном твердом растворе, образованном основным веществом и данной примесью. Если концентрация примеси в полупроводнике меньше параметра предельной растворимости, то примесь распределяется в кристаллической решетке моноатомно; если превышает этот параметр, то в выращиваемом кристалле появляются структурные нарушения, например макроскопические частицы инородной фазы, что сопровождается в первую очередь резким ростом плотности дислокаций. Указанные примеси являются малоэффективными центрами рекомбинации носителей и сравнительно слабо влияют на величину их времени жизни.

Примеси тяжелых и благородных металлов (Fe, Ni, Cr, W, Cu, Ag, Au и др.) в большинстве полупроводниковых материалов образуют глубокие, часто многозарядные донорные или акцепторные уровни в запрещенной зоне, имеют большие сечения захвата носителей заряда и являются эффективными центрами рекомбинации носителей, приводя к значительному снижению их времени жизни. Эти примеси обладают малой и обычно ярко выраженной растворимостью в полупроводниковых материалах и имеют очень малые значения коэффициента распределения между кристаллом и расплавом. Легирование ими производят в тех случаях, когда надо получить полупроводниковые материалы с малым временем жизни носителей или с высоким удельным электрическим сопротивлением, достигаемым компенсацией мелких энергетических уровней противоположной природы. Последнее часто используют для получения полуизолирующих кристаллов широкозонных полупроводниковых материалов типа $A^{III}B^V$ (GaAs, GaP, InP); легирующими примесями служат Cr, Fe, Ni.

Легирование полупроводниковых материалов обычно осуществляют непосредственно в процессах получения монокристаллов и эпитаксиальных структур. Примесь вводится в расплав либо в виде элемента, либо в виде сплава с данным полупроводниковым материалом (лигатуры). Часто легирование осуществляют из газовой фазы (паров) данного элемента или его легколетучих соединений. Это основной способ легирования в процессах эпитаксии при кристаллизации из газовой фазы. При молекулярно-пучковой эпитаксии источником легирующей добавки обычно является сама элементарная примесь. Расчет необходимого содержания легирующей примеси требует знания точной количественной связи между ее концентрацией и заданными свойствами полупроводниковых материалов, а также основных физико-химических характеристик примеси: коэффициента распределения между газовой фазой и кристаллом (K), упругости паров и скорости

испарения в широком интервале температур, растворимости в твердой фазе и т. п.

Одна из главных задач легирования – обеспечение равномерного распределения вводимой примеси в объеме кристалла и по толщине эпитаксиального слоя. При направленной кристаллизации из расплава равномерное распределение примеси по длине слитка достигается либо путем поддержания ее постоянной концентрации в расплаве за счет его подпитки из твердой, жидкой или газовой фазы, либо путем программированного изменения эффективного коэффициента распределения примеси при соответствующем изменении параметров процесса роста. Эффективный способ повышения объемной однородности монокристаллов – воздействие на массоперенос в расплаве наложением магнитного поля. Однородного распределения примеси по толщине слоя в процессе жидкофазной эпитаксии достигают кристаллизацией при постоянной температуре в условиях подпитки расплава, а при газофазной эпитаксии – поддержанием постоянной концентрации легирующей примеси в газовой фазе над подложкой на протяжении всего процесса наращивания.

Легирование полупроводниковых материалов может быть осуществлено также путем *радиационного воздействия* на кристалл, когда в результате ядерных реакций с участием собственных атомов вещества образуются электрически активные примеси. Наибольший интерес для радиационного легирования представляет воздействие тепловыми нейтронами, которые обладают большой проникающей способностью, что обеспечивает повышенную однородность легирования. Легирование облучением тепловыми нейтронами обеспечивает строго контролируемое введение заданных концентраций примеси и равномерное ее распределение в объеме кристалла. Однако в процессе облучения в кристалле образуются радиационные дефекты, для устранения которых необходим последующий высокотемпературный отжиг. Кроме того, может появиться наведенная радиоактивность, требующая выдержки образцов после облучения. Легирование облучением тепловыми нейтронами обычно используют для получения однородно легированных фосфором монокристаллов Si с высоким удельным электрическим сопротивлением.

При создании структур с *p-n*-переходами для полупроводниковых приборов широко используют *легирование путем диффузионного введения примеси*. В основе любой теории диффузии лежат законы Фика, описывающие неравновесный процесс перемещения (молекул и атомов в газах, ионов в плазме, электронов в полупроводниках и т. п.) вещества из области с высокой концентрацией в область с низкой концентрацией, приводящий к самопроизвольному выравниванию концентраций по всему занимаемому

объему. Профиль концентрации примеси при диффузии описывается обычно функцией ошибок и имеет вид плавной кривой, характер которой определяется следующими факторами: температурой и временем проведения процесса; толщиной слоя, из которого осуществляется диффузия; концентрацией и формой нахождения примеси в источнике, а также ее электрическим зарядом и возможностью взаимодействовать с сопутствующими примесями и дефектами в полупроводниковом материале. Из-за малых значений коэффициента диффузии основных легирующих примесей диффузионное легирование обычно проводят при высоких температурах (для Si, например, при 1100–1350 °С) и в течение длительного времени; при этом оно, как правило, сопровождается генерированием в кристалле значительного количества структурных дефектов, в частности дислокаций. При диффузионном легировании возникают трудности в получении тонких легированных слоев и достаточно резких *p-n*-переходов.

Для получения тонких легированных слоев перспективны процессы *ионного легирования (ионной имплантации)*, при которых введение примесных атомов в приповерхностный слой материала осуществляется путем бомбардировки соответствующими ионами с энергией от нескольких кэВ до нескольких МэВ. Возможность введения практически любой примеси в любой полупроводниковый материал, низкие рабочие температуры процесса, гибкое управление концентрацией и профилем распределения вводимой примеси, возможность легирования через диэлектрические покрытия с получением тонких, сильно легированных слоев обеспечили широкое распространение этого метода в технологии полупроводниковых приборов. Однако в процессе ионного легирования генерируются собственные точечные дефекты структуры, возникают области разупорядочения решетки, а при больших дозах – аморфизованные слои. Поэтому для получения качественных легированных слоев необходим последующий отжиг введенных дефектов. Отжиг проводят при температурах существенно более низких, чем при диффузии (для Si, например, не выше 700–800 °С). После отжига свойства имплантированных слоев близки к свойствам материала, легированного до тех же концентраций традиционными методами.

Вопросы

1. Какое строение имеет зонная структура (энергетический спектр) полупроводников?
2. При каких условиях возможен перенос носителей заряда в полупроводниках?

3. Чем отличается электропроводность собственных полупроводников от электропроводности примесных полупроводников?

4. При каких условиях ток в полупроводниках имеет диффузионную и дрейфовую составляющие?

5. Какие полупроводники называют электронными и дырочными?

6. Как изменяется положение уровня Ферми в собственных и примесных полупроводниках в зависимости от температуры?

7. Какой способ получения монокристаллов полупроводниковых материалов является наиболее распространенным?

8. Какие методы легирования применяют при создании структур с *p-n*-переходами для полупроводниковых приборов?

Глава 4

РАССЕЯНИЕ НОСИТЕЛЕЙ ЗАРЯДА В ПОЛУПРОВОДНИКАХ

Под действием внешнего электрического поля носители заряда приобретают некоторую скорость направленного движения (скорость дрейфа) и создают электрический ток. Подвижность носителей заряда, равная средней скорости носителей заряда в полупроводнике в электрическом поле с напряженностью I В/см

$$\mu = \frac{v_0}{E} \quad (4.1)$$

зависит от длины их свободного пробега,

$$\mu = f(l), \quad (4.2)$$

а следовательно, определяется процессами рассеяния движущихся в полупроводнике электронов (дырок).

Процесс рассеяния представляет собой искривление траектории движения носителя заряда под влиянием сил, действующих на электрон или дырку со стороны рассеивающего центра. Если таким центром является ион, то рассеивающей силой будет кулоновский потенциал; если рассеивающим центром является нейтральный атом примеси, рассеиваемый электрон останется в атоме, а выбитый, получив энергию, движется по измененной траектории. Так как электроны неразличимы, акт обмена электронами рассматривается как акт изменения траектории электрона, т. е. рассеяние. Характерной особенностью рассеяния на нейтральных атомах является независимость времени релаксации от энергии рассеиваемых носителей заряда и температуры. Процесс рассеяния электронов на тепловых колебаниях решетки рассматривается как столкновение с фононами. Поскольку концентрация фононов определяется температурой, то и рассеяние носителей заряда зависит от температуры. Рассеивающими центрами при движении электрона являются и структурные дефекты кристаллической решетки – дислокации, вакансии, имеет место также электрон-электронное рассеяние.

В реальных полупроводниках действуют одновременно несколько механизмов рассеяния, причем вклад каждого из них может сильно меняться с изменением температуры и концентрации примеси.

4.1. Механизмы рассеяния электронов и дырок

Рассеяние носителей заряда в кристаллических твердых телах – процесс взаимодействия электрона проводимости (дырки) с нарушениями идеальной периодичности кристалла. В реальном кристалле электроны и дырки совершают сложные траектории движения вследствие соударения с дефектами решетки. Поскольку дефекты, искажающие периодичность поля решетки, являются центрами рассеяния, имеющими разную природу, то они будут обуславливать различные механизмы рассеяния носителей заряда. В полупроводниках центрами рассеяния могут быть тепловые колебания решетки и статические дефекты, такие как атомы и ионы примеси, вакансии, дислокации, границы двойников и кристаллитов. Для количественной оценки процесса рассеяния вводят параметр σ , называемый *эффективным сечением рассеяния*.

Предположим, что имеется n свободных электронов, которые со средней тепловой скоростью v_0 движутся в данном направлении. Тогда nv_0 есть плотность тока электронов, т. е. количество электронов, проходящих в единицу времени через единичную площадку образца, перпендикулярную направлению скорости. Допустим, что на пути потока электронов в единичном сечении образца имеется N одинаковых центров рассеяния. Каждый центр характеризуется эффективным сечением, равным σ . Это, по существу, то пространство вокруг центра, в области которого имеет место рассеяние электронов. Поэтому количество электронов n_1 , рассеянных в единицу времени, определяется эффективным сечением рассеяния σ , количеством центров рассеяния N и плотностью падающего потока электронов nv_0 :

$$n_1 = \sigma N n v_0. \quad (4.3)$$

Если W – вероятность рассеяния одной частицы в единицу времени, то количество рассеянных электронов в единицу времени n_1 есть

$$n_1 = W n. \quad (4.4)$$

Тогда на основании (4.3) и (4.4) можно написать:

$$\sigma = \frac{n_1/N}{n v_0} = \frac{W}{N v_0}. \quad (4.5)$$

Таким образом, *эффективное сечение рассеяния σ есть отношение числа электронов, удаленных из пучка в результате рассеяния на одном центре в единицу времени, к плотности падающего пучка частиц*. Эффективное сечение рассеяния имеет размерность площади [см^2]:

$$[\sigma] = \frac{[W]}{[N v_0]} = \frac{T^{-1}}{L^{-3} L T^{-1}} = L^2.$$

Из формулы (4.5) следует, что вероятность рассеяния

$$W = \sigma N v_0. \quad (4.6)$$

Следовательно, вероятность рассеяния или вероятность столкновения определяется эффективным сечением рассеяния, количеством центров рассеяния и скоростью движения носителей заряда.

В то же время по определению вероятность столкновения обратно пропорциональна времени свободного пробега:

$$W = 1 / \tau. \quad (4.7)$$

Поэтому время свободного пробега и длину свободного пробега можно выразить через эффективное сечение рассеяния. Из (4.6) и (4.7) следует, что

$$\tau = \frac{1}{\sigma N v_0}. \quad (4.8)$$

Но $\tau = \frac{l}{v_0}$, тогда

$$l = \frac{1}{\sigma N}. \quad (4.9)$$

Поскольку роль дефектов в процессе рассеяния носителей заряда различна, то различные дефекты должны иметь разное эффективное сечение. Для их количественной оценки за эффективное сечение рассеяния σ примем площадку, в пределах которой возможно взаимодействие между носителем заряда и дефектом.

Такие дефекты, как вакансии и межузельные атомы, во многих отношениях сходны с примесями замещения. Эти дефекты называются точечными дефектами. Для них за σ можно принять площадь квадрата со стороной, равной постоянной решетки, т. е.

$$\sigma_A = (5 \cdot 10^{-8})^2 \approx 3 \cdot 10^{-15} \text{ см}^2.$$

Если предположить, что концентрация атомов примеси $N_A = 10^{16} \text{ см}^{-3}$, то длина свободного пробега при рассеянии носителей заряда на атомах примеси будет составлять

$$l_A = (3 \cdot 10^{-15} \cdot 10^{16})^{-1} \approx 3 \cdot 10^{-2} \text{ см} = 300 \text{ мкм}.$$

Для иона примеси будем считать, что его диаметр в 10 раз больше диаметра примесного центра, т. е. $\sigma_I = (5 \cdot 10^{-8} \cdot 10)^2 \approx 3 \cdot 10^{-13} \text{ см}^2$. В случае если $N_I = 10^{16} \text{ см}^{-3}$, то $l_I = 3 \cdot 10^{-4} \text{ см} = 3 \text{ мкм}$.

Дислокации являются линейными дефектами, простирающимися на большие области кристалла. Предположим, что линейная дислокация имеет

длину 0,1 см, а ее диаметр измеряется сотней периодов решетки. В этом случае площадь ее осевого сечения равна $5 \cdot 10^{-8} \times 100 \cdot 10^{-1} = 5 \cdot 10^{-7} \text{ см}^2$. При объемной плотности дислокаций $N_D = 10^8 \text{ см}^{-3}$ длина свободного пробега будет порядка $l_D = 2 \cdot 10^{-2} \text{ см} = 200 \text{ мкм}$.

Эффективное сечение рассеяния на тепловых колебаниях решетки определяются площадью сечения области, которую занимает колеблющийся атом за вычетом площади сечения самого атома. Если считать, что амплитуда колебаний порядка $r = 0,05 \text{ нм} = 5 \cdot 10^{-9} \text{ см}$, а диаметр атома $d = 10^{-8} \text{ см}$, то $\sigma_T = (d + r)^2 - d^2 \approx 2rd = 10^{-16} \text{ см}^2$. Это значение много меньше, чем для других центров рассеяния, но число колеблющихся атомов велико: $N_T = 10^{22} \text{ см}^{-3}$, поэтому $l_T = 10^{-6} \text{ см} = 0,01 \text{ мкм}$.

4.2. Рассеяние на ионах примеси

В примесном полупроводнике каждый ион примеси создает вокруг себя электрическое поле. Под воздействием этого поля движущийся носитель заряда отклоняется от своего первоначального направления (рис. 4.1). При этом носитель заряда отклоняется тем сильнее, чем медленнее и ближе к иону он движется.

Задача сводится фактически к нахождению траектории движения заряженной частицы в центральном поле иона. Она может быть решена чисто классическими методами. Для этого будем считать носитель заряда классической частицей, движущейся в кулоновском поле положительного иона примеси с зарядом Ze . В этом случае потенциальную энергию взаимодействия можно представить в виде

$$U(r) = \pm \frac{Ze^2}{4\pi\epsilon_r\epsilon_0 r}, \quad (4.10)$$

где плюс соответствует дырке, минус – электрону.

Как показывает расчет, в кулоновском поле иона носитель заряда движется по гиперболе. На рис. 4.2 изображены траектории движения электрона и дырки в поле положительного иона. Оба носителя заряда отклоняются одинаково, хотя в одном случае действует сила притяжения, в другом – отталкивания. Через b обозначено *прицельное расстояние* – минимальное расстояние между ионом и траекторией движения носителя заряда, а θ – *угол рассеяния*, т. е. угол между первоначальным направлением движения носителя заряда и направлением движения после рассеяния. При этом угол рассеяния зависит от прицельного расстояния следующим образом:

$$b = \frac{Ze^2}{\epsilon_r m^* v^2} \operatorname{ctg} \frac{\theta}{2}, \quad (4.11)$$

где m^* и v – эффективная масса и скорость движения носителя заряда.

Процесс рассеяния является случайным процессом, поэтому различные носители заряда будут отклоняться на различные углы от направления своего движения. В силу этого при подсчете времени релаксации необходимо учитывать усредненное значение сечения рассеяния по углам рассеяния.

Допустим, что угол рассеяния меняется от θ до $\theta + d\theta$. Поскольку в нашем случае рассеивающий центр обладает осевой симметрией, эти углы являются углами двух конусов, как это изображено на рис. 4.3. Здесь через $d\Omega$ обозначен телесный угол, заключенный между этими конусами, он равен:

$$d\Omega = 2\pi \sin \theta d\theta. \quad (4.12)$$

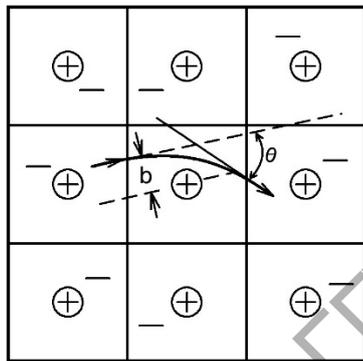


Рис. 4.1. Отклонение движущегося электрона в поле положительного иона

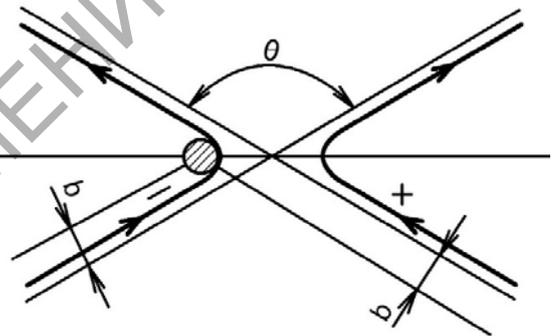


Рис. 4.2. Траектории электрона и дырки в поле положительного иона

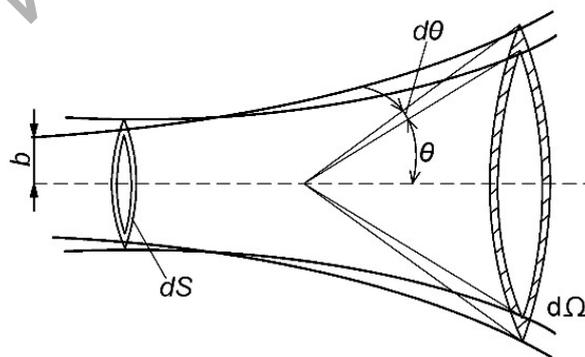


Рис. 4.3. Рассеяние частицы в телесном угле $d\Omega$

Процессы рассеяния в данном случае характеризуются дифференциальным сечением рассеяния $\sigma(\theta)$, которое равно отношению числа частиц, отклоненных одним центром за 1 с на угол $d\theta$ в единичный телесный угол, к потоку падающих частиц на 1 см^2 в 1 с.

Э. Резерфордом, при изучении рассеяния α -частиц на ядрах элементов, была получена формула:

$$\sigma(\theta) = \left(\frac{Ze^2}{2\epsilon_r m^* v^2} \right)^2 \frac{1}{\sin^4 \frac{\theta}{2}}. \quad (4.13)$$

Если выразить скорость через энергию $v = (2E/m^*)^{1/2}$, то время релаксации при рассеянии носителей заряда на ионах примеси и длина свободного пробега будут пропорциональны соответственно:

$$\begin{aligned} \tau_I &\sim E^{3/2} \cdot m^{*1/2}, \\ l_I &\sim \frac{E^2}{m^{*1/2}}. \end{aligned} \quad (4.14)$$

Таким образом, при рассеянии носителей заряда на ионах примеси время релаксации зависит от энергии в степени 3/2 и пропорционально эффективной массе в степени 1/2, а длина свободного пробега пропорциональна квадрату энергии и обратно пропорциональна эффективной массе в степени 1/2.

4.3. Рассеяние на атомах примеси и дислокациях

Рассеяние носителей заряда на нейтральных атомах примеси проявляется гораздо слабее, чем рассеяние на ионах примеси. Однако при низких температурах, когда концентрация ионизованных атомов примеси меньше концентрации нейтральных атомов примеси, этот механизм рассеяния играет существенную роль. В этом случае рассеяние носителей заряда может осуществляться двумя путями: 1) при помощи прямого упругого соударения и 2) обменом падающего электрона с электроном примесного атома. Определить время релаксации для этих двух механизмов рассеяния достаточно сложно. Однако на основе расчета, при котором рассеяние носителей заряда на нейтральных атомах примеси рассматривалось как рассеяние медленных электронов на атоме водорода, находящемся в среде с относительной диэлектрической проницаемостью ϵ_r , получена зависимость:

$$\tau_A = \frac{e^2 m^{*2}}{20\epsilon_r \hbar^3} \frac{1}{N_A} \sim \frac{m^{*2}}{N_A}, \quad (4.15)$$

где N_A – концентрация нейтральных атомов примеси при данной температуре.

Как следует из (4.15), время релаксации при рассеянии носителей заряда на нейтральной примеси не зависит ни от температуры, ни от энергии рассеиваемого носителя заряда, но определяется концентрацией атомов примеси. Поэтому этот механизм рассеяния особенно важен при очень низких температурах, когда примесь не ионизирована.

Рассеяние носителей заряда может происходить на нарушениях решетки, вносимых дислокациями. Вокруг дислокации существует область, в которой кристалл деформирован, поэтому в окрестности дислокации зависимость энергии от волнового вектора отличается от закона дисперсии в идеальном кристалле. Это приводит к рассеянию свободных носителей заряда. В полупроводниках, которые кристаллизуются в структуре типа алмаза, дислокации действуют как цепочки акцепторов, которые захватывают свободные электроны. Поэтому они представляют собой положительный центр с отрицательно заряженной линией в середине; т. е. дислокацию можно уподобить линейному отрицательному заряду в виде бесконечно длинного цилиндра радиусом R , вокруг которого имеется положительный пространственный заряд. Электроны, движущиеся перпендикулярно дислокациям, не могут свободно перемещаться в направлении поля, так как должны обойти дислокации и рассеяться на них, испытывая отталкивание. Расчет показывает, что *при рассеянии носителей заряда на дислокациях время релаксации не зависит от температуры кристалла, а определяется плотностью дислокаций на единицу поверхности N_D и скоростью рассеиваемого электрона v согласно выражению вида:*

$$\tau_D = \frac{3}{8Rv} \frac{1}{N_D}. \quad (4.16)$$

Если принять $v = 10^7$ см/с; $R = 3 \cdot 10^{-5}$ см; $N_D = 10^6$ см⁻², то время релаксации $\tau_D = 1,25 \cdot 10^{-9}$ с. Эта величина на 4–5 порядков больше, чем время релаксации для рассеяния на тепловых колебаниях решетки при комнатной температуре. Поэтому рассеяние на дислокациях незначительно при комнатных температурах, но оно может быть существенно при низких температурах.

4.4. Рассеяние на тепловых колебаниях решетки (фононах)

Рассеяние электронов проводимости при взаимодействии с фононами – основной механизм электрического сопротивления металлов и полупроводников. Способность электронов проводимости излучать и поглощать фононы приводит к образованию так называемых куперовских пар, что при низких

температурах является причиной перехода ряда металлов в сверхпроводящее состояние. Излучения фононов возбужденными атомами и молекулами тел обеспечивает возможность безызлучательных электронных переходов. В релаксационных процессах в твердых телах фононы обычно служат стоком для энергии, запасенной другими степенями свободы кристалла, например электронами.

Различают акустические и оптические фононы.

Акустические фононы характеризуются при малых волновых векторах линейным законом дисперсии и параллельным смещением всех атомов в элементарной ячейке. Такой закон дисперсии описывает звуковые колебания решетки, поэтому фонон и называется акустическим.

Оптические фононы существуют только в кристаллах, элементарная ячейка которых содержит два и более вида атомов. Эти фононы характеризуются при малых волновых векторах такими колебаниями атомов, при которых центр тяжести элементарной ячейки остается неподвижным.

Взаимодействие электрона (или дырки) с колебаниями решетки осуществляется двояким образом:

1) электрон передает часть своей энергии решетке, так что определенное нормальное колебание с частотой ω_q увеличивает свое квантовое число на единицу. В результате такого процесса рассеяния образуется фонон с энергией $\hbar\omega_q$ и квазиимпульсом $Q = \hbar q$ и число фононов возрастает на единицу; $Q = \hbar q$ есть импульс, который приобретает решетка вследствие соударения с электронами;

2) электрон при взаимодействии получает часть энергии от решетки, так что квантовое число определенного нормального колебания с частотой ω_q уменьшается на единицу. В этом случае исчезает фонон с энергией $\hbar\omega_q$ и квазиимпульсом $Q = \hbar q$, а число фононов уменьшается на единицу.

Следовательно, в любом из этих двух случаев электрон сталкиваясь с фононом, обменивается с ним энергией и квазиимпульсом. Такой механизм рассеяния называется *однофононным*. Так как число фононов определяется температурой, то и рассеяние электронов на тепловых колебаниях решетки должно зависеть от температуры.

Рассмотрим процесс рассеяния электронов в атомном полупроводнике кубической структуры, энергетическая схема которого представлена на рис. 4.4, где a – постоянная решетки в отсутствии деформации. Будем считать, что зона проводимости такого кристалла простая, так что энергия электрона в зоне проводимости равна

$$E = \frac{\hbar^2 k^2}{2m^*}. \quad (4.17)$$

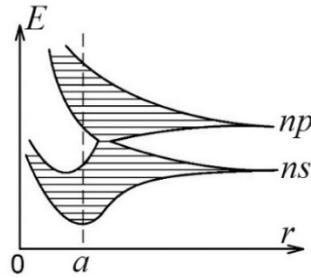


Рис. 4.4. Энергии электронов в кристалле полупроводника

В твердом теле поперечные волны представляют собой волны деформации сдвига и в кубическом кристалле не приводят к изменению объема. Продольные волны, являющиеся волнами деформации сжатия и растяжения, вызывают изменение объема кристалла. Как следует из рис. 4.4, при сжатии кристалла, сопровождающемся уменьшением постоянной решетки a , нижний край зоны проводимости смещается вверх, а верхний край валентной зоны – вниз, в результате этого ширина запрещенной зоны увеличивается. При растяжении, приводящем к увеличению постоянной a , ширина запрещенной зоны уменьшается. Следовательно, в таком кристалле локальная деформация, создаваемая продольной акустической волной, приводит к волнообразному смещению дна зоны проводимости и потолка валентной зоны, как это схематически представлено на рис. 4.5. Поэтому движущийся электрон, сталкиваясь с волной смещения, обусловленной тепловыми колебаниями решетки, будет рассеиваться только на продольных колебаниях. Это рассеяние упругое и происходит на длинноволновых акустических фонах.



Рис. 4.5. Изменение ширины запрещенной зоны под действием продольной акустической волны

Расчеты показывают, что время релаксации при рассеянии электронов на длинноволновых акустических колебаниях решетки в атомах полупроводника обратно пропорционально температуре в первой степени и энергии в степени $1/2$, а длина свободного пробега ее зависит от энергии носителей заряда:

$$\tau_l \sim T^{-1} E^{-1/2}, \quad (4.18)$$

$$l \sim T^{-1}.$$

Оценим время релаксации для германия в случае рассеяния электронов на колебаниях решетки. В чистом германии при комнатной температуре подвижность электронов $\mu_n = 0,39 \text{ м}^2/(\text{В}\cdot\text{с})$. Если принять $m_n^* = 0,3 m_0$, то $e/m_n^* = 6 \cdot 10^{11} \text{ Кл/кг}$, следовательно $\tau_1 = m_n^* \mu_n / e = 6,5 \cdot 10^{-13} \text{ с}$.

Из сравнения соотношений (4.18) и (4.14) следует, что при повышении энергии носителей заряда время релаксации при рассеянии их на ионизированной примеси увеличивается, а при рассеянии на акустических колебаниях уменьшается. Это означает, что в случае примесного рассеяния преобладающее влияние оказывают медленные электроны и дырки, и поэтому этот механизм рассеяния должен сильнее проявляться при низких температурах. При рассеянии на акустических фононах основной вклад вносят сравнительно быстрые электроны, и этот механизм рассеяния должен быть значительнее при высоких температурах.

В полярных полупроводниках, а также в полупроводниковых соединениях типа $A^{\text{III}}B^{\text{V}}$, в которых связь между атомами носит частично ионный характер, электроны проводимости гораздо сильнее взаимодействуют с оптическими колебаниями, чем с акустическими. Поскольку при оптических колебаниях смещение ионов в ячейке происходит в противоположные стороны, то разноименные заряды создадут электрические поля, перемещающиеся по кристаллу в виде плоских волн. Длинноволновые оптические колебания получили название *поляризационных волн*. Взаимодействие носителей заряда с поляризационными волнами и приводит к их рассеянию. При этом продольные колебания сильнее рассеивают, чем поперечные. Каждое столкновение электрона с оптическим фононом приводит к возникновению или исчезновению фонона с энергией $\hbar\omega_{\text{оп}}$. Частота оптического фонона слабо зависит от квазиимпульса p , поэтому при соударении энергия электрона либо увеличивается, либо уменьшается на одинаковую величину, равную:

$$\hbar\omega_{\text{оп}} = \pm\hbar\omega_0, \quad (4.19)$$

где ω_0 – максимальная частота продольной оптической ветви колебания.

Энергия, передаваемая носителем заряда решетке за одно столкновение, определяется энергией возникшего фонона, умноженной на отношение разности вероятностей испускания и поглощения к сумме этих вероятностей. Но вероятность поглощения фонона пропорциональна концентрации фононов N_q , а вероятность испускания фонона пропорциональна $N_q + 1$. Следовательно, на одно столкновение электрона, сопровождающееся передачей энергии фонону, приходится $2N_q + 1$ столкновений.

При высоких температурах, когда $T \gg \theta_D$ (где θ_D – температура Дебая) или $kT \gg \hbar\omega_0$, т. е. когда энергия оптического фонона $\hbar\omega_0$ много меньше энергии электрона kT , тогда

$$\delta_{\text{оп}} = \frac{\Delta E}{E} = \frac{(\hbar\omega_0)^2}{2(kT)^2} \ll 1. \quad (4.20)$$

В этом случае у большинства электронов изменение энергии при поглощении или испускании оптического фонона будет незначительно. Для случая рассеяния электронов на оптических фононах в ионном кристалле при высоких температурах, как показывает расчет, время релаксации и длина свободного пробега соответственно пропорциональны:

$$\begin{aligned} \tau_{\text{оп}} &\sim T^{-1}E^{1/2}, \\ l_{\text{оп}} &\sim T^{-1}E. \end{aligned} \quad (4.21)$$

При низких температурах, когда $T \ll \theta_D$ или $kT \ll \hbar\omega_0$, электроны могут только поглощать оптические фононы, в результате чего энергия их сильно увеличивается. Поскольку вероятность испускания фонона в $(N_q + 1)/N_q$ раз превышает вероятность поглощения фонона, время пребывания электрона в состоянии с большой энергией будет очень мало. Можно считать, что электрон сразу после поглощения фонона испускает фонон той же энергии. Фактически рассеяние можно рассматривать как процесс обмена фононами, а поэтому энергия электрона практически остается неизменной, но направление его квазиимпульса сильно изменится. Это позволяет считать, что и при низких температурах, значительно меньших характеристической температуры Дебая $\theta_D = \hbar\omega_0/k$, электрон, взаимодействуя с оптическими колебаниями решетки, рассеивается упруго.

Расчет, в котором учитывается только поглощение фонона, дает выражение для времени релаксации электрона

$$\tau_{\text{оп}} = \exp\left(\frac{1}{T}\right), \quad (4.22)$$

которое не зависит от энергии электрона, но экспоненциально зависит от температуры.

Длина свободного пробега электронов в случае рассеяния на оптических фононах при низких температурах определяется соотношением вида

$$l_{\text{оп}} \sim E^{1/2}, \quad (4.23)$$

т. е. зависит от энергии носителей заряда в степени $1/2$.

Все рассмотренные случаи и характеристики рассеяния носителей заряда в полупроводниках обобщены в таблице 4.1.

Таблица 4.1

Характеристики рассеяния носителей заряда в полупроводниках

На ионах примеси	На атомах примеси	На дислокациях	В атомных п/п на длинноволновых акустических колебаниях	В полярных п/п на оптических колебаниях
$\tau_I \sim E^{3/2} \cdot m^{*1/2}$ $l_I \sim E^2 / m^{*-1/2}$	$\tau_A \sim \frac{m^{*2}}{N_A}$	$\tau_D = \frac{1}{vN_D}$	$\tau_l \sim T^{-1}E^{-1/2}$ $l \sim T^{-1}$	1) Высокие T $T \gg \theta_D, kT \gg \hbar\omega_0$ $\tau_{оп} \sim T^{-1}E^{1/2}$ $l_{оп} \sim T^{-1}E$
Механизм важен при низких и высоких T	При очень низких T	При низких T	При высоких T	2) Низкие T $T \ll \theta_D, kT \ll \hbar\omega_0$ $\tau_{оп} = \exp\left(\frac{1}{T}\right)$ $l_{оп} \sim E^{1/2}$

Вопросы

1. Чем определяется вероятность рассеяния электронов?
2. На каких примесях эффективное сечение рассеяния больше – на атомах или ионах?
3. Какой механизм рассеяния называется однофононным?
4. Какие механизмы рассеяния носителей заряда преобладают на низких/высоких температурах?

Глава 5

ГЕНЕРАЦИЯ И РЕКОМБИНАЦИЯ ЭЛЕКТРОНОВ И ДЫРОК

Генерация носителей заряда – образование свободных электронов и дырок – происходит при воздействии теплового хаотического движения атомов кристаллической решетки (тепловая генерация), при воздействии квантов света (световая генерация), сильных электрических полей и других энергетических факторов. Так как полупроводник всегда находится под действием всех этих факторов или хотя бы одного ($T \neq 0$), генерация носителей происходит непрерывно. Одновременно с генерацией в полупроводнике непрерывно идет и обратный процесс – рекомбинация носителей заряда, т. е. возвращение электронов из зоны проводимости в валентную зону, в результате чего исчезает (рекомбинирует) пара носителей заряда. В состоянии термодинамического равновесия процессы генерации и рекомбинации носителей заряда взаимно уравновешены. При этом в полупроводнике существуют равновесные концентрации электронов n_0 и дырок p_0 . При воздействии на полупроводник нетеплового внешнего энергетического фактора (света, сильного электрического поля и др.) из-за генерации новых носителей заряда общая концентрация n и p будет превышать равновесную концентрацию n_0, p_0 на величину Δn (или Δp), которую называют неравновесной, или избыточной, концентрацией.

5.1. Равновесные и неравновесные носители заряда

Рассмотрим донорный полупроводник, находящийся в состоянии термодинамического равновесия. В результате тепловой генерации электроны с донорной примеси переходят в зону проводимости. При достаточно высоких температурах будет существенным переброс электронов непосредственно из валентной зоны в зону проводимости.

На рис. 5.1 изображены плотность квантовых состояний $N(E)$ в зоне проводимости и валентной зоне, функция распределения Ферми – Дирака $f_0(E)$ и концентрация равновесных электронов n_0 и дырок p_0 , которые занимают состояния вблизи краев соответствующих зон (заштрихованные области). Стрелками показаны переходы электронов при тепловом возбуждении. Свободные носители зарядов, возникающие в результате тепловой генерации и находящиеся в термодинамическом равновесии с кристаллической решеткой, называют *равновесными* (n_0, p_0).

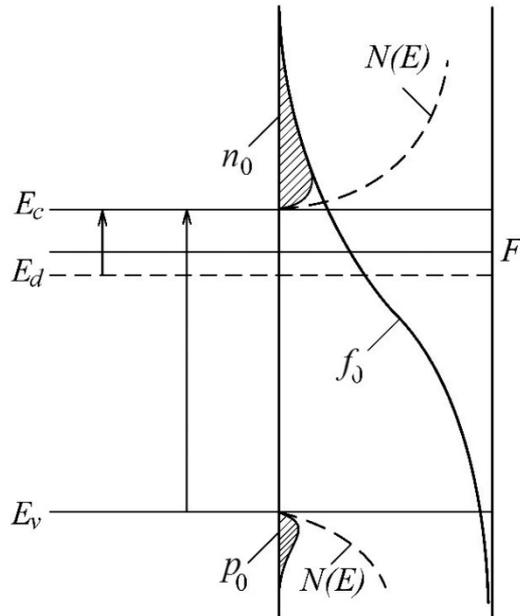


Рис. 5.1. Тепловая генерация свободных носителей заряда и их концентрация в состоянии термодинамического равновесия

Одновременно с генерацией свободных носителей заряда идет процесс рекомбинации: электроны возвращаются в свободные состояния в валентной зоне, в результате чего исчезают свободный электрон и свободная дырка. В условиях термодинамического равновесия процессы генерации и рекомбинации взаимно уравниваются. Обозначим через G_0 число генерируемых, а через R_0 – число рекомбинирующих электрон-дырочных пар в единице объема кристалла за одну секунду. Вероятность рекомбинации пропорциональна произведению концентрации свободных носителей заряда, поэтому

$$R_0 = \gamma_r n_0 p_0, \quad (5.1)$$

где γ_r – коэффициент пропорциональности, называемый коэффициентом рекомбинации.

Для равновесного состояния полупроводника справедливо равенство

$$G_0 = R_0, \quad (5.2)$$

которое является выражением принципа детального равновесия.

Помимо тепловой генерации имеются другие механизмы, приводящие к возникновению свободных носителей заряда. Например, они могут образоваться в результате облучения полупроводника светом, разрыва валентных связей в сильных электрических полях или инжекции с помощью p - n -перехода. Во всех этих случаях возникает некоторая концентрация электронов n и дырок p , которая отличается от термодинамически равновесной. Подвижные носители заряда, не находящиеся в термодинамическом

равновесии как по концентрации, так и по энергетическому распределению, являются *неравновесными носителями заряда*, а их концентрацию Δn и Δp называют *неравновесной*, или *избыточной*.

При оптической генерации носители заряда могут иметь кинетическую энергию, значительно превышающую среднюю тепловую энергию равновесных частиц. В результате рассеяния на дефектах кристаллической решетки носители заряда довольно быстро передают ей свою избыточную энергию. Этот процесс сводится к тому, что неравновесные электроны, рассеивая свою избыточную энергию, «опускаются» к нижнему краю зоны проводимости, а неравновесные дырки, рассеивая избыточную энергию, «поднимаются» к верхнему краю валентной зоны. Если при этом концентрация неравновесных носителей заряда мало отличается от равновесной, то передача кристаллу кинетической энергии неравновесными носителями заряда практически не изменит энергию решетки, а следовательно, и температуру кристалла, поэтому не изменится концентрация равновесных носителей заряда. В этом случае общее количество электронов и дырок будет равно соответственно:

$$n = n_0 + \Delta n; \quad p = p_0 + \Delta p, \quad (5.3)$$

5.2. Биполярная оптическая генерация носителей заряда

Рассмотрим *собственный* полупроводник, в котором оптической генерацией создается некоторая концентрация электронов и дырок. Генерацию носителей заряда, при которой в результате возбуждения возникают пары – электрон и дырка, будем называть *биполярной генерацией*. В этом случае поглощение кванта света сопровождается разрывом ковалентной связи и количество генерируемых избыточных электронов и дырок одинаково:

$$\Delta n = \Delta p. \quad (5.4)$$

Параллельно с процессом генерации идет процесс рекомбинации, и в стационарном состоянии число электронов и дырок, возбуждаемых за единицу времени в результате разрыва ковалентных связей, равно числу электронов и дырок, рекомбинирующих за это же время. Поскольку неравновесные носители заряда через очень малое время становятся фактически неотличимыми по энергии от равновесных, можно считать, что они имеют тот же коэффициент рекомбинации γ_r , что и равновесные носители заряда.

После выключения возбуждающего света концентрации электронов и дырок уменьшаются в результате рекомбинации. При этом скорость

убывания числа свободных электронов и дырок определяется разностью скоростей рекомбинации и тепловой генерации

$$-\left(\frac{dn}{dt}\right)_r = -\left(\frac{dp}{dt}\right)_r = \gamma_r np - G_0. \quad (5.5)$$

Принимая во внимание (5.1)–(5.4), равенство (5.5) можно переписать следующим образом:

$$-\left(\frac{dn}{dt}\right)_f = \gamma_r (np - n_0 p_0) = \gamma_r (n_0 + p_0 + \Delta n) \Delta n. \quad (5.6)$$

Учитывая, что в случае малого уровня возбуждения $\Delta n \ll (n_0 + p_0)$, получим:

$$\left(\frac{dn}{dt}\right)_r = -\gamma_r (n_0 + p_0) \Delta n = -\gamma_r (n_0 + p_0) (n + n_0). \quad (5.7)$$

Введем обозначение

$$\tau = \frac{1}{\gamma_r (n_0 + p_0)}. \quad (5.8)$$

Тогда (5.7) примет вид:

$$\left(\frac{dn}{dt}\right)_r = -\frac{n - n_0}{\tau} = -\frac{\Delta n}{\tau}, \quad (5.9)$$

откуда

$$\Delta n = \Delta n(0) e^{-t/\tau}, \quad (5.10)$$

где $\Delta n(0)$ – избыточная концентрация электронов в момент выключения возбуждающего света.

Таким образом, в случае малого уровня биполярной генерации избыточная концентрация электронов и дырок после прекращения возбуждения уменьшается по экспоненциальному закону и за время τ количество их в результате рекомбинации убывает в e раз. Следовательно, τ представляет собой среднее время существования избыточной концентрации электронов и дырок, а поэтому его называют *временем жизни неравновесных носителей заряда*. Для собственного полупроводника скорости убывания числа электронов и дырок равны и значение τ определяет *время жизни электронно-дырочных пар*. Значение времени жизни неравновесных носителей заряда в зависимости от типа полупроводникового материала и от степени его чистоты может изменяться в очень широких пределах – от 10^{-2} до 10^{-8} с.

5.3. Монополярная оптическая генерация носителей заряда. Максвелловское время релаксации

Процессы генерации неравновесных носителей заряда в *примесных* полупроводниках могут существенно отличаться от явлений, происходящих в материалах с чисто собственной проводимостью. Рассмотрим донорный полупроводник, в котором облучение светом приводит к перебросу электронов с донорных уровней в зону проводимости. Такой процесс возбуждения носителей заряда будем называть *монополярной оптической генерацией*. Она характеризуется образованием избыточной концентрации основных носителей заряда (в данном случае электронов). При этом электронейтральность полупроводника не нарушается, так как избыточный заряд свободных электронов, генерируемых светом, скомпенсирован зарядом образовавшихся положительных ионов донорной примеси.

Если неравновесные электроны генерируются в некоторой области образца, то концентрация электронов в этой области будет повышенной: $n = n_0 + \Delta n$. Наличие избыточной концентрации электронов вызовет перемещение их в неосвещенную область полупроводника. Поэтому в неосвещенной области образца, куда продиффундируют электроны, возникает отрицательный объемный заряд избыточных электронов, а в области, откуда уйдут электроны, будет сосредоточен положительный заряд, обусловленный ионами донорной примеси. Если в момент времени $t = 0$ генерация электронов прекратится, то электрическое поле \mathcal{E} , порожденное объемными зарядами, вызовет ток проводимости, который в течение некоторого времени приведет к уничтожению объемного заряда.

Изменение плотности пространственного заряда ρ в результате протекания тока, плотность которого равна \mathbf{J} , как известно, подчиняется уравнению непрерывности электрического заряда

$$\operatorname{div} \mathbf{J} = \operatorname{div}(\gamma \mathcal{E}) = -\frac{d\rho}{dt}, \quad (5.11)$$

а напряженность электрического поля и объемный заряд связаны уравнением Пуассона

$$\operatorname{div} \mathcal{E} = \rho / \varepsilon_r \varepsilon_0, \quad (5.12)$$

где ε_r – относительная диэлектрическая проницаемость полупроводника; ε_0 – электрическая постоянная, ρ – плотность объемного заряда.

Из уравнений (5.11) и (5.12) следует, что

$$\frac{d\rho}{dt} = \frac{\gamma}{\varepsilon_r \varepsilon_0} \rho. \quad (5.13)$$

Отсюда найдем закон изменения объемного заряда во времени

$$\rho = \rho_0 e^{-t/\tau_\mu}, \quad (5.14)$$

где ρ_0 – плотность объемного заряда в момент $t = 0$, а

$$\tau_\mu = \epsilon_r \epsilon_0 / \gamma \quad (5.15)$$

есть *диэлектрическое*, или *максвелловское*, время релаксации.

Оценим численное значение максвелловского времени релаксации. Для германия, имеющего $\gamma = 1 \text{ Ом}^{-1}\text{см}^{-1} = 10^2 \text{ См/м}$ и $\epsilon_r = 16$, $\tau_\mu = 10^{-12} \text{ с}$.

Как следует из (5.14), в случае монополярной оптической генерации носителей заряда возникает объемный заряд, который после выключения возбуждающего света со временем уменьшается *по экспоненциальному закону с постоянной времени τ_μ* . Другими словами, объемный заряд, созданный избыточными основными носителями заряда, в результате тока проводимости после выключения возбуждающего света исчезнет в среднем через τ_μ .

5.4. Механизмы рекомбинации

Процессы рекомбинации неравновесных носителей заряда в полупроводниках можно разделить на два основных типа: межзонная рекомбинация и рекомбинация через ловушки.

Межзонная рекомбинация осуществляется при переходе электрона из состояния в зоне проводимости в пустое состояние валентной зоны, что равносильно уничтожению свободного электрона и свободной дырки. Этот процесс совершается при соблюдении законов сохранения энергии и квазиимпульса. Если энергия и квазиимпульс электрона в зоне проводимости (до рекомбинации) были E' и p' , а после рекомбинации равны соответственно E и p , то законы сохранения энергии и квазиимпульса запишутся в виде

$$E' = E + \Delta E; \quad p' = p + Q, \quad (5.16)$$

где Q – квазиимпульс, передаваемый электроном кристаллической решетке; ΔE – энергия, которая выделяется при процессе рекомбинации.

В зависимости от того, каким образом рассеивается избыток энергии ΔE , межзонная рекомбинация делится на три типа:

1) если энергия, освобождающаяся в процессе рекомбинации, излучается в виде кванта света, рекомбинацию называют *излучательной*, или *фотонной*;

2) если энергия, освобождающаяся в процессе рекомбинации, передается решетке, т. е. расходуется на образование фононов, рекомбинация носит название *безызлучательной*, или *фононной*;

3) если в акте рекомбинации выделяющаяся энергия передается третьему свободному носителю заряда, рекомбинацию называют *ударной*, или *Оже-рекомбинацией*.

В полупроводниковом кристалле всегда имеются дефекты, энергетические уровни которых находятся в запрещенной зоне. Поэтому наряду с межзонной рекомбинацией может идти процесс рекомбинации через локальные уровни энергии. Дефект решетки, способный захватить электрон из зоны проводимости и дырку из валентной зоны, осуществляя их рекомбинацию, называют *рекомбинационной ловушкой*. В этом случае исчезновение пары – электрона и дырки проводимости – осуществляется следующим образом. Нейтральная рекомбинационная ловушка захватывает электрон из зоны проводимости, который затем через некоторое время перейдет в валентную зону. В зависимости от способа расходования энергии ΔE при переходе электрона на более низкий энергетический уровень рекомбинация через ловушки также может быть фотонной или фононной.

5.4.1. Межзонная излучательная рекомбинация

При межзонной излучательной рекомбинации электрон из зоны проводимости переходит в валентную зону, при этом излучается квант света, энергия которого равна ширине запрещенной зоны:

$$h\nu = E_g. \quad (5.17)$$

При термодинамическом равновесии количество носителей заряда, рекомбинирующих в единице объема за единицу времени с испусканием фотонов, в соответствии с принципом детального равновесия равно числу носителей заряда, возбуждаемых при поглощении фотонов. Иными словами, плотность равновесного излучения R_n^0 равна числу поглощаемых квантов R_p^0 в единице объема за одну секунду. Но интенсивность процесса излучательной рекомбинации пропорциональна произведению концентраций равновесных электронов и дырок:

$$R_n^0 = R_p^0 = \gamma_r n_0 p_0 = \gamma_r n_i^2, \quad (5.18)$$

где γ_r – коэффициент межзонной излучательной рекомбинации.

При отклонении от равновесного состояния концентрации носителей заряда равны $n = n_0 + \Delta n$ и $p = p_0 + \Delta p$. Они, как отмечалось в п. 5.2, характеризуются тем же коэффициентом рекомбинации, что и равновесные. При этом, если не наступает вырождение носителей заряда, сохраняется пропорциональность скорости рекомбинации R произведению концентраций

электронов и дырок. Поэтому, определяя γ_r из соотношения (5.18), можно написать:

$$R = \gamma_r np = R_{\text{и}}^0 \frac{np}{n_i^2}. \quad (5.19)$$

Теоретически показано, что время жизни неравновесных носителей зарядов Δn и Δp в собственном полупроводнике при межзонной излучательной рекомбинации тем меньше, чем выше температура и чем меньше ширина запрещенной зоны. В примесном полупроводнике время жизни неравновесных носителей заряда меньше, чем τ_{ir} в собственном полупроводнике, и с ростом степени легирования и температуры оно уменьшается. Время жизни неравновесных носителей заряда в случае межзонной излучательной рекомбинации снижается с уменьшением ширины запрещенной зоны.

5.4.2. Межзонная ударная рекомбинация

Если происходит столкновение одновременно двух свободных электронов и одной дырки или двух дырок и одного свободного электрона, может иметь место рекомбинация двух из них с передачей энергии третьему носителю заряда, который переходит на более высокий энергетический уровень в соответствующей зоне, становясь «горячим». «Горячий» носитель в результате столкновения передаст свою энергию фононам и придет в равновесное состояние с решеткой. Этот процесс рекомбинации является процессом, обратным ударной ионизации, при котором переход свободного носителя заряда в зоне на более низкий энергетический уровень в результате выделившейся энергии происходит при возникновении пары «электрон – дырка». На рис. 5.2 схематично изображены процессы ударной рекомбинации с участием в качестве третьего носителя заряда электрона (рис. 5.2, а) или дырки (рис. 5.2, б).

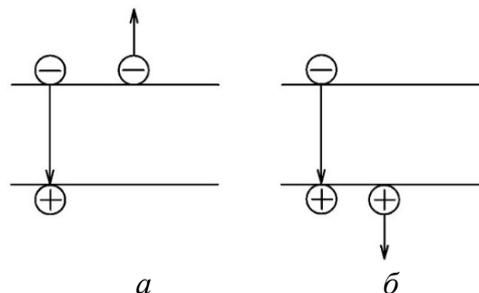


Рис. 5.2. Схема процессов межзонной ударной рекомбинации

Вероятность столкновения пары «электрон – дырка» со свободным электроном пропорциональна $n^2 p$, а с дыркой – $p^2 n$. Таким образом, уменьшение концентрации свободных электронов или дырок в результате межзонной ударной рекомбинации определяется выражением

$$-\frac{dn}{dt} = -\frac{dp}{dt} = \eta^{(n)} n^2 p + \eta^{(p)} p^2 n - \eta^{(n)} n_0^2 p_0 - \eta^{(p)} p_0^2 n_0, \quad (5.20)$$

где $\eta^{(n)}$ и $\eta^{(p)}$ – коэффициенты ударной рекомбинации с участием в качестве третьего носителя заряда электрона или дырки. Два последних члена в (5.20) учитывают интенсивность генерации при термодинамическом равновесии. Следует отметить, что коэффициенты ударной рекомбинации отличаются по размерам от соответствующих коэффициентов в случае рекомбинации другого типа.

Время жизни неравновесных носителей заряда для собственного полупроводника при межзонной ударной рекомбинации максимально. В примесном полупроводнике время жизни при этом типе рекомбинации сильно зависит от концентрации основных носителей заряда. При ее увеличении возрастает число столкновений неравновесных электронно-дырочных пар с основными носителями заряда и время жизни снижается. Мгновенное время жизни для ударной межзонной рекомбинации при высоком уровне возбуждения очень сильно зависит от концентрации избыточных носителей заряда.

5.4.3. Рекомбинация носителей заряда через ловушки

Реальные полупроводниковые материалы содержат обычно примеси нескольких типов, каждая из которых может создавать один или несколько уровней в запрещенной зоне полупроводника.

Дефекты решетки, обычно нейтральные в состоянии термодинамического равновесия и способные захватывать подвижные носители заряда одного знака и освобождать их, называют *ловушками*, или *центрами захвата*. Ограничимся рассмотрением простейшего случая, когда в полупроводнике имеется один тип ловушек, создающих единичный энергетический уровень.

Пусть концентрация ловушек равна N_t , а их энергетический уровень E_t лежит вблизи середины запрещенной зоны. Процесс рекомбинации через эти ловушки проиллюстрирован на рис 5.3.

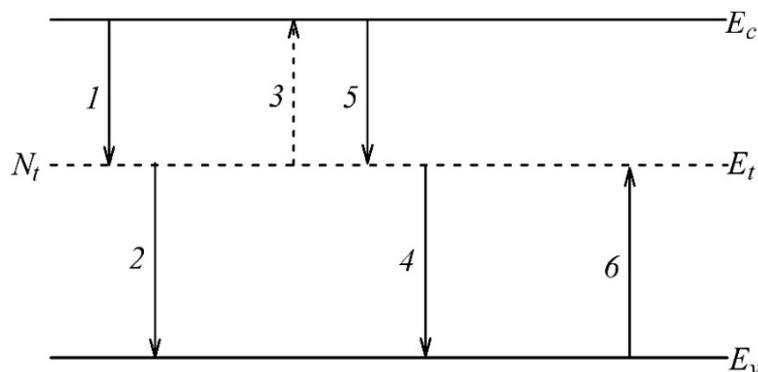


Рис. 5.3. Переходы электронов, связанные с рекомбинацией через ловушки (описание в тексте)

В полупроводнике донорного типа, имеющем большое число свободных электронов, рекомбинация электрона и дырки происходит следующим образом. Нейтральная ловушка захватывает электрон из зоны проводимости, приобретая отрицательный заряд (переход 1), а затем электрон с уровня ловушки переходит в валентную зону, что эквивалентно захвату дырки отрицательно заряженной ловушкой (переход 2). При снижении концентрации свободных дырок вероятность второго процесса уменьшается и возможна эмиссия электрона, т. е. обратный переход электрона с ловушки в зону проводимости (переход 3). Если материал содержит большое количество свободных дырок, они интенсивно захватываются ловушками, т. е. электроны переходят с ловушки в валентную зону (переход 4), а затем происходит захват электрона из зоны проводимости (переход 5). При малой концентрации электронов проводимости возможен процесс освобождения дырки, т. е. возвращение дырки в валентную зону (переход 6).

Перейдем к количественному описанию рассмотренных процессов. Если через $f_i = f(E_i)$ обозначим вероятность заполнения ловушки электроном, то $1 - f_i$ будет вероятностью того, что на ловушке нет электрона, т. е. она свободна. Тогда интенсивность захвата ловушками электронов $R_{\text{зах}}^n$, которая пропорциональна концентрации электронов в зоне проводимости n и числу свободных мест на уровне ловушек $N_t(1 - f_i)$, равна:

$$R_{\text{зах}}^n = \gamma_n n N_t (1 - f_i), \quad (5.21)$$

где γ_n – коэффициент захвата неравновесных электронов ловушками.

Количество электронов $R_{\text{воз}}^n$, возвращающихся с ловушек в зону проводимости в результате эмиссии, пропорционально концентрации электронов на ловушках $N_t f_i$:

$$R_{\text{воз}}^n = \beta_n f_i N_t, \quad (5.22)$$

где β_n – коэффициент ионизации электронов с ловушек.

Можно показать, что скорость изменения концентрации *неравновесных электронов* в зоне проводимости в результате процессов захвата электронов ловушками и освобождения их из ловушек в зону проводимости равна

$$-\frac{dn}{dt} = \gamma_n N_t [n(1 - f_i) - n_1 f_i], \quad (5.23)$$

где

$$n_1 = N_t e^{-(E_c - E_i)/kT}. \quad (5.24)$$

Скорость изменения концентрации *неравновесных дырок* в валентной зоне запишется следующим образом:

$$-\frac{dp}{dt} = \gamma_p p N_t f_t - \beta_p N_t (1 - f_t). \quad (5.25)$$

Здесь первое слагаемое определяет количество дырок, захватываемых ловушками, а второе – число дырок, возвращающихся из ловушек в валентную зону. Величины γ_p и β_p представляют собой соответственно коэффициенты захвата и ионизации для дырок.

Запишем условие локальной электронейтральности для однородного полупроводника, содержащего некоторое количество ионов донорной примеси N_d^+ . Для равновесного случая условие электронейтральности имеет вид:

$$n_0 + N_t f_{0t} = p_0 + N_d^+. \quad (5.26)$$

Будем считать, что число ионов донорной примеси не меняется и в неравновесном состоянии. Тогда

$$n + N_t f_t = p + N_d^+. \quad (5.27)$$

Вычитая (5.26) из (5.27), получаем соотношение для избыточной концентрации носителей заряда

$$n + n_0 + N_t (f_t - f_{0t}) = p - p_0$$

или

$$\Delta n + N_t \Delta f_t = \Delta p. \quad (5.28)$$

Если концентрация ловушек в полупроводнике мала по сравнению с плотностью избыточных носителей заряда, т. е. $N_t \ll \Delta n, \Delta p$, то количеством электронов $N_t \Delta f_t$, оседающих на ловушках, можно пренебречь. При этом условии избыточная концентрация электронов равна избыточной концентрации дырок, т. е.

$$\Delta n = \Delta p. \quad (5.29)$$

В случае *малого уровня возбуждения*, когда $\Delta n \ll (n_0 + p_0)$, время жизни неравновесных носителей заряда в зависимости от положения уровня Ферми при постоянной температуре, т. е. от степени легирования материала, будет равно:

$$\tau = \tau_{p0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n0} \frac{p_0 + p_1}{n_0 + p_0}, \quad (5.30)$$

где введены обозначения:

$$\tau_{p0} = 1/\gamma_p N_t; \quad \tau_{n0} = 1/\gamma_n N_t.$$

Итак, при малом уровне возбуждения время жизни электронно-дырочных пар не зависит от концентрации избыточных носителей заряда, а определяется значениями равновесных концентраций электронов и дырок и положением энергетического уровня рекомбинационной ловушки, задающим величины n_1 и p_1 .

На основании формулы (5.30) для зависимости $\ln \tau$ от положения уровня Ферми (F) можно выделить четыре основных области (рис. 5.4).

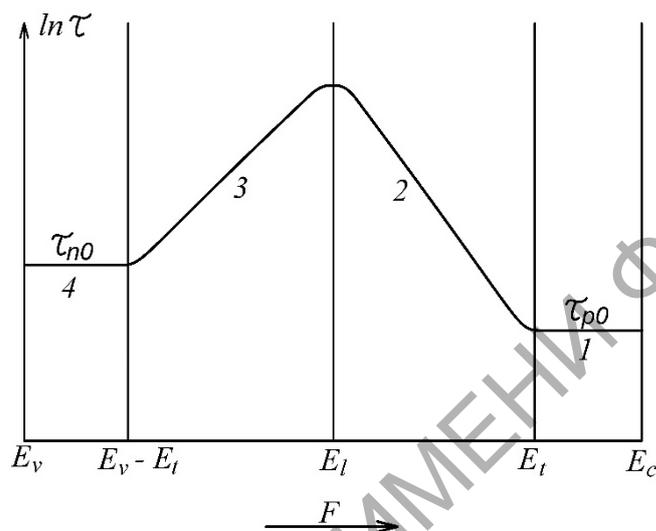


Рис. 5.4. Зависимость времени жизни неравновесных носителей заряда от положения уровня Ферми

Область 1. Сильно легированный полупроводник n -типа. Уровень Ферми расположен ниже дна зоны проводимости, но выше энергетического уровня рекомбинации ловушки, т. е. $E_t < F < E_c$. В этом случае справедливы следующие соотношения концентраций носителей заряда:

$$n_0 \ll p_0; \quad n_0 \gg n_1; \quad n_0 \gg p_1.$$

Учитывая эти неравенства, из (5.30) получаем:

$$\tau = \tau_{p0} = 1/\gamma_p N_t. \quad (5.31)$$

Таким образом, в сильно легированном полупроводнике n -типа время жизни неравновесных электронов и дырок постоянно и равно $\tau = \tau_{p0}$, т. е. определяется временем жизни неосновных равновесных носителей заряда – дырок. В этом случае скорость процесса рекомбинации определяется числом актов захвата дырок ловушками, занятыми электронами.

Равенство (5.31) выражает собой тот факт, что каждая дырка в материале n -типа, перемещаясь в валентной зоне, обладает определенной вероятностью быть захваченной отрицательно заряженной рекомбинационной ловушкой.

Область 2. Полупроводник n -типа легирован слабо, так что уровень Ферми лежит в запрещенной зоне ниже энергетического уровня ловушек, но выше середины запрещенной зоны, т. е. $E_i < F < E_t$. В этом случае имеют место следующие соотношения для концентрации носителей заряда:

$$n_0 \gg p_0; \quad n_0 \gg p_1; \quad n_0 < n_1.$$

Соответственно выражение для τ приобретает вид:

$$\tau \approx \tau_{p0} \frac{n_1}{n_0} = \tau_{p0} e^{(E_t - F)/kT}. \quad (5.32)$$

Из (5.32) следует, что по мере понижения уровня Ферми время жизни неравновесных носителей заряда τ растет по экспоненциальному закону. При этом чем ниже опускается уровень Ферми, тем меньше степень заполнения ловушек электронами, поэтому вероятность захвата дырки ловушкой снижается, что ведет к увеличению времени жизни дырки. Это подтверждается данными, представленными на рис. 5.4.

Область 3. В слабо легированном полупроводнике p -типа положение уровня Ферми определяется неравенством $(E_v - E_t) < F < E_i$, что дает следующие соотношения для концентраций носителей заряда:

$$p_0 \gg n_0; \quad p_0 \gg p_1; \quad n_1 > p_0.$$

На основании этого из (5.30) получаем формулу для времени жизни:

$$\tau \approx \tau_{p0} \frac{n_1}{n_0} = \tau_{p0} \frac{N_c}{N_v} e^{-(E_c + E_v - E_t - F)/kT}. \quad (5.33)$$

Видно, что для акцепторного полупроводника по мере понижения уровня Ферми от середины запрещенной зоны в результате легирования время жизни неравновесных носителей заряда уменьшается по экспоненциальному закону. Это происходит потому, что в слабо легированном полупроводнике p -типа почти все ловушки свободные и охотно захватывают электроны из зоны проводимости. С увеличением количества дырок проводимости растет вероятность их рекомбинации с электронами, попавшими на ловушки. При этом интенсивность процесса освобождения электронов с ловушек снижается и время жизни неравновесных носителей заряда уменьшается (рис. 5.4).

Область 4. Сильно легированный материал p -типа, для которого положение уровня Ферми определяется неравенством вида $E_v < F < (E_v - E_t)$.

В этом случае

$$p_0 \gg n_0; \quad p_0 \gg p_1; \quad p_0 \ll n_1,$$

а из (5.30) следует:

$$\tau = \tau_{n0} = \frac{1}{\gamma_n N_t} = 1 / \sigma_n N_t v_0, \quad (5.34)$$

где σ_n – эффективное сечение захвата электрона.

Таким образом, в сильно легированном полупроводнике *p*-типа время жизни пары «электрон – дырка» постоянно, равно $\tau = \tau_{n0}$, определяется временем жизни равновесных электронов – неосновных носителей заряда и не зависит от положения уровня Ферми. У такого полупроводника все ловушки свободны от электронов и каждый электрон, захваченный ловушкой, немедленно рекомбинирует с дыркой, так как количество дырок в валентной зоне очень велико. Процесс обратного переброса электронов с ловушек в зону проводимости практически полностью отсутствует и не влияет на время жизни.

Рассмотрим случай **большого уровня возбуждения** избыточных носителей заряда, когда $\Delta n \gg n_0$; $\Delta p \gg p_0$, а значит $\Delta n \gg n_1$ и $\Delta p \gg p_1$. В этом случае время жизни неравновесных носителей заряда будет равно:

$$\tau_\infty \approx \tau_{p0} + \tau_{n0} = \frac{\gamma_n + \gamma_p}{\gamma_n \gamma_p N_t}, \quad (5.35)$$

где $\tau_{p0} = 1 / \gamma_p N_t$; $\tau_{n0} = 1 / \gamma_n N_t$.

Как следует из этого соотношения, при большом уровне возбуждения время жизни неравновесных электронов и дырок τ_∞ при рекомбинации через ловушки не зависит от концентраций электронов и дырок, а определяется лишь количеством и свойствами ловушек.

Вопросы

1. Какие носители заряда называются равновесными/неравновесными?
2. В каких полупроводниках возможна биполярная оптическая генерация носителей заряда?
3. В каком временном диапазоне может изменяться время жизни неравновесных носителей заряда в собственных полупроводниках?
4. Какие процессы генерации неравновесных носителей заряда могут возникать в примесных полупроводниках?
5. Какие виды межзонной рекомбинации носителей заряда могут существовать в полупроводниках?
6. Какие дефекты решетки называют ловушками, или центрами захвата?

Глава 6

КИНЕТИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

Процессы, связанные с перемещением носителей тока в кристалле под действием внешних сил (градиента температуры, электрического поля, магнитного поля и т. д.), электропроводность, теплопроводность, термоэлектрические, гальваномагнитные и термомагнитные эффекты называются явлениями переноса или кинетическими явлениями.

Все кинетические явления в полупроводниках подразделяют на термоэлектрические, гальваномагнитные и фотоэлектрические. К термоэлектрическим относятся эффекты Зеебека, Пельтье и Томсона, они имеют место в проводнике (полупроводнике), когда на него действуют только внешнее электрическое поле и градиент температуры. Гальваномагнитные (термомагнитные) явления возникают в проводнике, помещенном в магнитное поле, когда вдоль проводника течет электрический (тепловой) ток, причем в зависимости от относительной ориентации магнитного поля \mathbf{H} и тока (теплого или электрического) все гальвано- и термомагнитные эффекты также подразделяют на группы: 1) продольные эффекты в продольном магнитном поле, т. е. эффекты, возникающие в направлении тока, когда магнитное поле параллельно току; 2) продольные эффекты в поперечном магнитном поле, т. е. эффекты, возникающие в направлении тока, когда магнитное поле перпендикулярно этому направлению; 3) поперечные эффекты в поперечном магнитном поле, т. е. эффекты, возникающие в направлении, перпендикулярном к плоскости тока и магнитного поля.

Особую группу эффектов составляют явления, происходящие в полупроводниках под действием электромагнитного излучения. К ним относятся фотоэлектрические эффекты (фотоэффекты), внешний и внутренний.

6.1. Эффект Холла

Если проводник (или полупроводник), по которому протекает электрический ток, поместить в магнитное поле, то в нем возникает ряд эффектов: уменьшается его электро- и теплопроводность; в направлении, перпендикулярном магнитному полю и току, возникает разность потенциалов. Явления, происходящие в проводниках первого рода (т. е. в материалах, в которых ток переносится электронами) при одновременном воздействии электрического и магнитного полей, называются *гальваномагнитными явлениями (эффектами)*. Наиболее информативным из гальваномагнитных эффектов является эффект Холла. Остановимся на нем более подробно.

Случай примесной проводимости. Рассмотрим образец примесного (для определенности – электронного) полупроводника, по которому протекает электрический ток, помещенный в магнитное поле, перпендикулярное направлению тока (рис. 6.1). На движущийся электрон действует сила Лоренца

$$F_{\text{Л}} = e \left\{ E + \frac{1}{c} [vH] \right\},$$

где e – заряд электрона, v – его скорость, c – скорость света, E и H – напряженности электрического и магнитного полей.

Под действием магнитной составляющей силы Лоренца

$$F = \frac{e}{c} [vH] \quad (6.1)$$

электроны будут отклоняться на боковую грань образца (направление их движения указано на чертеже стрелками). Таким образом, на одной из граней будет скапливаться отрицательный заряд, а на противоположной будет оставаться нескомпенсированный положительный заряд. Возникшее вследствие этого поперечное электрическое поле и поперечная разность потенциалов и называются эффектом Холла.



Рис. 6.1. Схема возникновения ЭДС Холла

Процесс накопления заряда будет продолжаться до тех пор, пока поперечное электрическое поле не уравновесит «магнитную силу»; после этого электроны будут двигаться вдоль образца, накопление заряда прекратится и установится стационарное состояние. Таким образом, поперечное холловское поле E_x может быть вычислено из условия равенства магнитной и электрической сил, действующих на электрон

$$\frac{ev_y H}{c} = eE_x. \quad (6.2)$$

Откуда

$$E_x = \frac{v_y H}{c}. \quad (6.3)$$

Или, так как $v_y = \mu E_y$,

$$E_x = \frac{\mu H}{c} E_y. \quad (6.4)$$

При опыте измеряют не поперечное поле, а разность потенциалов

$$V_x = E_x b \text{ или } E_x = \frac{V_x}{b} \quad (6.5)$$

и не скорость электронов v_y , а силу тока I , связанную с v_y соотношениями

$$j_y = env_y, I = j_y S = j_y bd,$$

т. е.

$$v_y = \frac{j_y}{en} = \frac{I}{S \cdot en} = \frac{I}{bd \cdot en}, \quad (6.6)$$

где $S = bd$ – поперечное сечение образца (рис. 6.1).

Выразив E_x через V_x и v_y через I и подставив E_x и v_y в (6.3), получим окончательно

$$V_x = \frac{1}{enc} \frac{HI}{d} = R_H \frac{HI}{d}, \quad (6.7)$$

где введено обозначение

$$R_H = \frac{1}{enc}, \quad (6.8)$$

а d – размер образца в направлении магнитного поля. Величина R_H , являющаяся коэффициентом, связывающим поперечную разность потенциалов с напряженностью магнитного поля и силой тока, называется *постоянной Холла*.

Измерив четыре величины, входящие в состав (6.7) – V_x , H , I и d , – можно вычислить R_H :

$$R_H = V_x \cdot d \cdot H \cdot I, \quad (6.9)$$

а зная R_H , мы можем определить из (6.7) концентрацию свободных электронов в исследуемом материале:

$$n = \frac{1}{eR_H c}. \quad (6.10)$$

Измеряя одновременно постоянную Холла и электропроводность, можно, кроме того, вычислить и подвижность электронов. Действительно, в случае примесного полупроводника

$$\gamma = en\mu \quad (6.11)$$

и согласно (6.8)

$$R_H \gamma = \frac{\mu}{c}. \quad (6.12)$$

Таким образом, одновременное измерение электропроводности и эффекта Холла позволяет получить важные сведения о примесном полупроводнике: знак носителей тока (по знаку ЭДС Холла), их число и подвижность. В этом и заключается значение измерения эффекта Холла как одного из основных методов исследования электрических свойств проводников и полупроводников.

В заключение этого параграфа необходимо сделать несколько замечаний.

Математический вывод ЭДС Холла и постоянной Холла, приведенный выше, весьма нестрог. Действительно, в выражение для магнитной силы (6.1) входит полная скорость электрона, состоящая в полупроводниках из двух частей: скорости беспорядочного теплового движения v_0 и дрейфовой скорости за счет электрического поля v_d ($v = v_0 + v_d$); в выражение же для плотности тока входит только скорость дрейфа в электрическом поле. Отождествив в нашем выводе v_d и v , мы тем самым пренебрегли хаотическим тепловым движением электронов и их распределением по скоростям, поэтому значение постоянной Холла (6.8), полученное нами, не точно. Более строгий вывод (с учетом теплового движения электронов) дает несколько другое значение для постоянной Холла:

$$R_H = \frac{A}{enc}, \quad (6.13)$$

где постоянная A зависит от механизма рассеяния электронов, т. е. от показателя степени в выражении для зависимости длины свободного пробега электрона от его энергии.

Перечислим наиболее часто встречающиеся на практике случаи:

- в ковалентных кристаллах при рассеянии электронов на акустических колебаниях длина свободного пробега электрона зависит от энергии и

$$A = \frac{3\pi}{8} \text{ или } R_H = \frac{3\pi}{8} \frac{1}{enc} \approx 1,17 \cdot \frac{1}{enc}; \quad (6.14)$$

- в полупроводниках с ионной решеткой и при рассеянии электронов на оптических колебаниях при температуре выше температуры Дебая длина свободного пробега пропорциональна энергии и

$$A = 1,11 \text{ или } R_H \approx 1,11 \cdot \frac{1}{enc}; \quad (6.15)$$

- при рассеянии на ионах примеси длина свободного пробега электрона пропорциональна квадрату энергии и

$$R_H \approx 1,93 \cdot \frac{1}{enc}; \quad (6.16)$$

• в металлах и сильно вырожденных полупроводниках в электропроводности могут принимать участие только электроны, находящиеся на самых верхних уровнях распределения Ферми; энергия и скорость у этих электронов вполне определимы, в данном случае никакого распределения по скоростям учитывать не нужно; в соответствии с этим вывод, приведенный нами, оказывается строгим для металлов и вырожденных полупроводников, т. е. $A = 1$ и постоянная Холла в этом случае равна

$$R_H = \frac{1}{e n c}. \quad (6.17)$$

Случай смешанной проводимости. Если в переносе электричества участвуют и дырки, и электроны, то картина эффекта Холла значительно усложняется. В выражение для магнитной силы входит произведение скорости частицы на заряд; так как для электронов и дырок, участвующих в переносе тока, и знак (направление) скорости, и знак заряда противоположны, то направления сил, действующих на них, будут одни и те же и они будут отклоняться в одну и ту же сторону.

Если и подвижность, и концентрации электронов и дырок одинаковы, то их заряды будут полностью компенсировать друг друга и холловское поле будет равно нулю. Если же это равенство не имеет места и концентрация или подвижность носителей одного знака (для определенности предположим, что электронов) больше, чем другого, то на одной грани будет скапливаться отрицательный заряд, а на другой будет оставаться нескомпенсированный положительный заряд. Таким образом, возникает поперечное холловское поле, тормозящее движение электронов в этом направлении и ускоряющее движение дырок. Процесс накопления заряда будет продолжаться до тех пор, пока это поле не уравнивает потоки электронов и дырок. Расчет дает в этом случае следующее выражение для постоянной Холла:

$$R_H = \frac{A}{e c} \frac{\mu_p^2 p - \mu_n^2 n}{(\mu_p p + \mu_n n)^2}, \quad (6.18)$$

где μ_p , p , μ_n , n – соответственно подвижности и концентрации дырок и электронов.

Постоянная A в (6.18) так же, как и в случае одного знака носителей, определяется механизмом рассеяния: $A = 3\pi/8$ в атомной решетке и т. д.

Нетрудно убедиться, что: а) если $n = 0$ или $p = 0$, (6.18) переходит в (6.17); б) если $n = p$ и $\mu_p = \mu_n$, то постоянная Холла, а следовательно, и ЭДС Холла равны нулю.

6.2. Магниторезистивный эффект

В 1856 г. Уильям Томсон (Лорд Кельвин) впервые описал магниторезистивный эффект, который заключается в увеличении удельного сопротивления полупроводника в магнитном поле с ростом индукции B .

Под действием силы Лоренца траектория движения носителей в магнитном поле искривляется. В результате накопления заряда на одной из граней пластинки в ней возникает электрическое поле Холла E_x (рис. 6.2), перпендикулярное направлению приложенного (внешнего) электрического поля E .

В пластинке возникает суммарное электрическое поле E_Σ , направление которого определяется векторной суммой

$$E_\Sigma = E + E_x.$$

Угол между векторами E и E_x носит название *угол Холла*. Значение угла Холла φ определяется из соотношения

$$\operatorname{tg} \varphi = \frac{E_x}{E}. \quad (6.19)$$

Из соотношения (6.4) следует, что величина электрического поля Холла E_x равна

$$E_x = \frac{1}{en} B j. \quad (6.20)$$

Напряженность поля, создаваемого внешним источником питания в пластинке полупроводника, определяется по формуле

$$E = \frac{j}{\gamma} = \frac{j}{en\mu_n}. \quad (6.21)$$

Подставляя эти значения в формулу (6.19), получим $\operatorname{tg} \varphi = \mu_n B$. Для небольших значений индукции B величина $\operatorname{tg} \varphi \approx \varphi$. Следовательно,

$$\varphi = \mu B. \quad (6.22)$$

Рассмотрим случай неограниченного полупроводника, для которого выполняется условие $a > l$. В такой «короткой» пластине ЭДС Холла не образуется. Однако траектория носителей тока по-прежнему оказывается сдвинутой от направления внешнего поля на угол Холла φ (рис. 6.3).

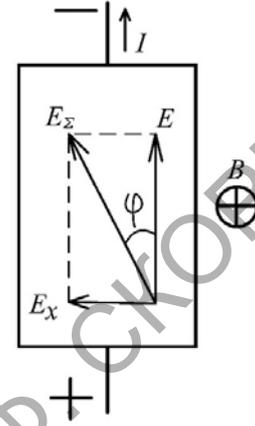


Рис. 6.2. Объяснение магниторезистивного эффекта (φ – угол Холла)

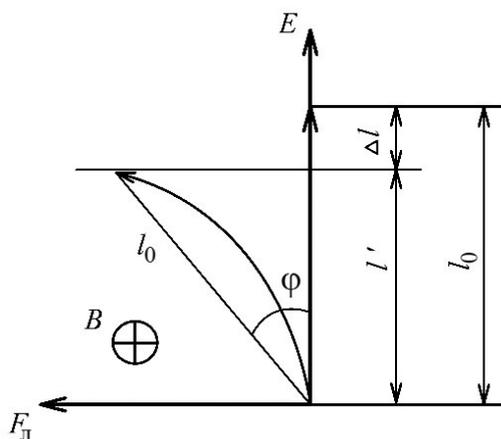


Рис. 6.3. Изменение длины свободного пробега носителя заряда вдоль вектора электрического поля E в полупроводнике, находящемся в магнитном поле с индукцией B

Отклонение траектории движения носителей тока от направления поля E равносильно уменьшению длины свободного пробега носителей l_0 в направлении поля E :

$$\Delta l = l_0 - l' = l_0 - l_0 \cos \varphi, \quad (6.23)$$

где l' – проекция длины свободного пробега на направление внешнего электрического поля E .

Для малых φ значение $\cos \varphi$ можно разложить в ряд:

$$\cos \varphi \approx 1 - \frac{\varphi^2}{2!} + \dots$$

Тогда

$$\Delta l = l_0 - l_0 + l_0 \frac{\varphi^2}{2} = l_0 \frac{\varphi^2}{2}.$$

Подставляя в это выражение значение φ из формулы (4.21), получим

$$\Delta l \approx l_0 \frac{\mu^2 B^2}{2}.$$

Уменьшение l_0 эквивалентно уменьшению скорости дрейфа носителей $v_{др}$, которая, в свою очередь, пропорциональна величине проводимости полупроводника γ . Следовательно,

$$\frac{\Delta l}{l_0} = \frac{\Delta \gamma}{\gamma} = \frac{\Delta \rho}{\rho} \sim \frac{\mu^2 B^2}{2}.$$

В результате относительное изменение удельного сопротивления полупроводника определяется выражением

$$\frac{\Delta \rho}{\rho} = C \mu^2 B^2, \quad (6.24)$$

где C – коэффициент, зависящий от геометрических размеров пластинки полупроводника.

Магниторезистивный эффект применяется в магниторезисторах и магнитодиодах.

Магниторезисторы – это резисторы переменного сопротивления, величина которого зависит от напряженности приложенного магнитного поля. Магниторезистор представляет собой пластинку полупроводника, на

поверхность которого нанесены металлические полосы (рис. 6.4). Каждая часть пластины полупроводника между двумя металлическими полосами представляет собой отдельный магниторезистор. Металлические полосы выполняют роль шунтов, уменьшающих ЭДС Холла, которая возникает на боковых гранях пластинки полупроводника.

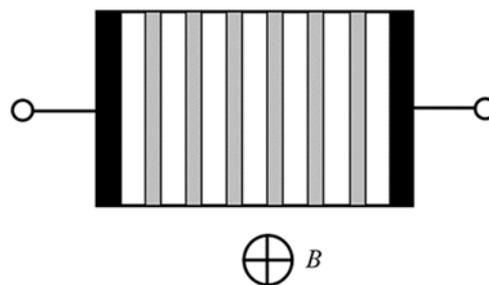


Рис. 6.4. Вариант конструкции магниторезистора

Основным полупроводниковым материалом для магниторезисторов является антимонид индия InSb и арсенид индия InAs – материалы с большой подвижностью электронов, соответственно $7,6 \text{ м}^2/(\text{В} \cdot \text{с})$ и $3,3 \text{ м}^2/(\text{В} \cdot \text{с})$. Серийно изготавливаемые магниторезисторы типа MR, CM имеют характеристики: номинальное сопротивление 50–220 Ом, рассеиваемая мощность 0,15–0,25 Вт.

Магнитодиоды (рис. 6.5) – это диоды с толстой базой, сопротивление которой увеличивается в поперечном магнитном поле в результате уменьшения подвижности основных и неосновных носителей заряда, как и в обычном магниторезисторе. Увеличение сопротивления базы диода с толстой базой может быть связано также с уменьшением времени жизни неосновных носителей, если из-за искривления траектории движения неосновные носители будут достигать поверхности базовой области, где велика скорость их рекомбинации. В качестве материала для изготовления магнитодиодов обычно используют монокристаллический германий или кремний, имеющие достаточно большую подвижность носителей заряда. Прямые ветви ВАХ германиевого магнитодиода в магнитных полях с различной магнитной индукцией показаны на рис. 6.5, б.

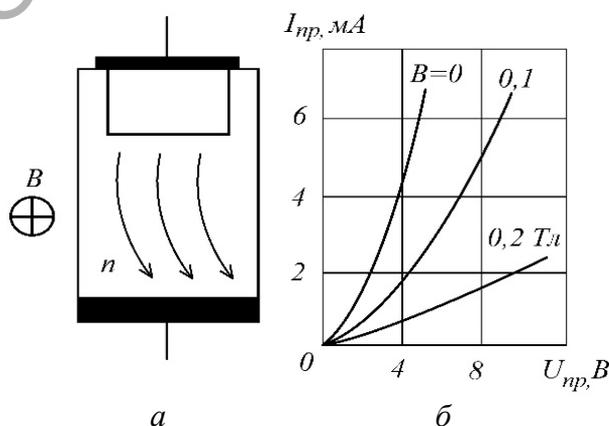


Рис. 6.5. Магнитодиод: а – схема конструкции; б – прямые ветви ВАХ магнитодиода на основе Ge, находящегося в магнитных полях с различной магнитной индукцией B

Для оценки чувствительности магнитодиода к магнитному полю, по аналогии с преобразователями Холла, используют *вольтовую чувствительность* γ_v , выражение для которой задается в виде

$$\gamma_v = \frac{\Delta U}{BI_{\text{пр}}}, \quad (6.25)$$

где ΔU – изменение напряжения на магнитодиоде при внесении его в магнитное поле, В; $I_{\text{пр}}$ – значение прямого тока, А; B – значение магнитной индукции, Тл.

Вольтовая чувствительность магнитодиодов может быть значительно выше вольтовой чувствительности преобразователей Холла из того же материала.

6.3. Термоэлектрические явления

Термоэлектрические явления – совокупность физических явлений, обусловленных взаимосвязью между тепловым и электрическими процессами в металлах и полупроводниках. К термоэлектрическим явлениям относятся: Эффект Зеебека, эффект Пельтье, эффект Томсона. В некоторой степени все эти эффекты взаимосвязаны, поскольку причина всех термоэлектрических явлений – нарушение теплового равновесия в потоке носителей (т. е. отличие средней энергии электронов в потоке от энергии Ферми). Абсолютные значения всех термоэлектрических коэффициентов растут с уменьшением концентрации носителей; поэтому в полупроводниках они в десятки и сотни раз больше, чем в металлах и сплавах.

Эффект Зеебека – явление возникновения ЭДС в замкнутой электрической цепи, состоящей из последовательно соединенных разнородных проводников (полупроводников), контакты между которыми находятся при

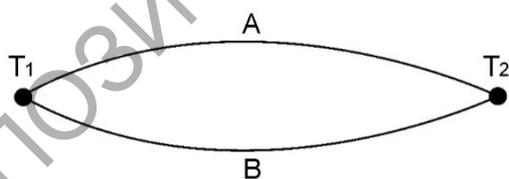


Рис. 6.6. Эффект Зеебека

различных температурах (рис. 6.6). Цепь, которая состоит только из двух различных проводников, называется термоэлементом или термопарой. Эффект Зеебека также иногда называют просто термоэлектрическим эффектом.

Величина возникающей термо-ЭДС зависит только от материала проводников и температур горячего (T_1) и холодного (T_2) контактов.

В небольшом интервале температур термо-ЭДС E можно считать пропорциональной разности температур: $E = \alpha_{12}(T_2 - T_1)$, где α_{12} – термоэлектрическая способность пары (или коэффициент термо-ЭДС).

В простейшем случае коэффициент термо-ЭДС определяется только материалами проводников, однако строго говоря, он зависит и от температуры, и в некоторых случаях с изменением температуры α_{12} меняет знак. Более корректное выражение для термо-ЭДС:

$$E = \int_{T_2}^{T_1} \alpha_{12}(T) dT. \quad (6.26)$$

Величина термо-ЭДС составляет милливольты при разности температур в 100 градусов и температуре холодного спая в 0 °С (например, пара «медь – константан» дает 4,25 мВ, «платина – платинородий» – 0,643 мВ, «нихром – никель» – 4,1 мВ).

Возникновение эффекта Зеебека обусловлено тремя составляющими:

1. Различная зависимость *средней энергии электронов* от температуры в разных веществах. Если вдоль проводника существует градиент температур, то электроны на горячем конце приобретают более высокие энергии и скорости, чем на холодном; в полупроводниках в дополнение к этому концентрация электронов проводимости растет с температурой. В результате возникает поток электронов от горячего конца к холодному и на холодном конце накапливается отрицательный заряд, а на горячем остается нескомпенсированный положительный заряд. Процесс накопления заряда продолжается до тех пор, пока возникшая разность потенциалов не вызовет поток электронов в обратном направлении, равный первичному, благодаря чему установится равновесие. ЭДС, возникновение которой описывается данным механизмом, называется *объемной ЭДС*.

2. Различная зависимость от температуры *контактной разности потенциалов*. Контактная разность потенциалов вызвана отличием энергии Ферми у контактирующих проводников:

$$V_k = \frac{E_{2F} - E_{1F}}{e}, \quad (6.27)$$

где E_F – энергия Ферми, e – заряд электрона.

На контакте существует электрическое поле, локализованное в тонком приконтактном слое. Если составить замкнутую цепь из двух металлов, то разность потенциалов возникает на обоих контактах. Электрическое поле будет направлено одинаковым образом в обоих контактах – от большего E_F к меньшему. Это значит, что если совершить обход по замкнутому контуру, то в одном контакте обход будет происходить по полю, а в другом – против поля. В результате циркуляция вектора \mathbf{E} будет равна нулю.

Если температура одного из контактов изменится на dT , то, поскольку энергия Ферми зависит от температуры, U также изменится. Но если изменилась внутренняя контактная разность потенциалов, то изменилось

электрическое поле в одном из контактов, и поэтому циркуляция вектора E будет отлична от нуля, т. е. появляется ЭДС в замкнутой цепи. Данная ЭДС называется *контактной ЭДС*.

3. *Фононное увлечение*. Если в твердом теле существует градиент температуры, то число фононов, движущихся от горячего конца к холодному, будет больше, чем в обратном направлении. В результате столкновений с электронами фононы могут увлекать за собой последние, и на холодном конце образца будет накапливаться отрицательный заряд (на горячем – положительный) до тех пор, пока возникшая разность потенциалов не уравновесит эффект увлечения.

Эта разность потенциалов и представляет собой третью составляющую термо-ЭДС, которая при низких температурах может быть в десятки и сотни раз больше рассмотренных выше. В магнетиках наблюдается дополнительная составляющая термо-ЭДС, обусловленная эффектом увлечения электронов магнонами.

Эффект Зеебека используется при создании термодатчиков (например, в компьютерах). Такие датчики миниатюрны и очень точны.

Эффект Пельтье – термоэлектрическое явление, при котором происходит выделение или поглощение тепла при прохождении электрического тока в месте контакта (спая) двух разнородных проводников (рис. 6.7).

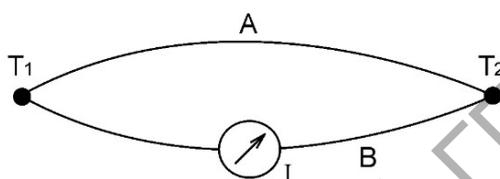


Рис. 6.7. Эффект Пельтье

Величина выделяемого тепла и его знак зависят от вида контактирующих веществ, направления и силы протекающего электрического тока:

$$Q = P_{AB}I = (P_B - P_A)I, \quad (6.28)$$

где Q – количество выделенного или поглощенного тепла; I – сила тока; P – коэффициент Пельтье, который связан с коэффициентом термо-ЭДС α соотношением Томсона $P = \alpha T$, где T – абсолютная температура в К.

Эффект был открыт французским физиком Ж. Пельтье в 1834 г., суть явления исследовал несколькими годами позже. В 1838 г. российский физик Э. Х. Ленц провел эксперимент, поместив каплю воды в углубление на стыке двух стержней из висмута и сурьмы. При пропускании электрического тока в одном направлении капля превращалась в лед, при смене направления тока лед таял, что позволило установить, что в зависимости от направления тока в эксперименте, помимо джоулева тепла, выделяется или поглощается дополнительное тепло, которое получило название тепла Пельтье. Эффект Пельтье «обратен» эффекту Зеебека. Эффект Пельтье более заметен у полупроводников, это свойство используется в элементах Пельтье.

Причина возникновения явления Пельтье заключается в следующем. На контактах двух веществ имеется контактная разность потенциалов, которая создает внутреннее контактное поле. Если через контакт протекает электрический ток, то это поле будет либо способствовать прохождению тока, либо препятствовать. Если ток идет против контактного поля, то внешний источник должен затратить дополнительную энергию, которая выделяется в контакте, что приводит к его нагреву. Если же ток идет по направлению контактного поля, то он может поддерживаться этим полем, которое и совершает работу по перемещению зарядов. Необходимая для этого энергия отбирается у вещества, что приводит к охлаждению его в месте контакта.

Эффект Томсона (1854 г.) относится к термоэлектрическим эффектам и заключается в следующем: при пропускании электрического тока через полупроводник (или проводник), вдоль которого существует градиент температуры, в нем, помимо джоулева тепла, в зависимости от направления тока будет выделяться или поглощаться дополнительное количество тепла (теплота Томсона).

Неравномерное нагревание первоначально однородного образца меняет его свойства, делая вещество неоднородным. Поэтому явление Томсона – это, в сущности, своеобразное явление Пельтье с той разницей, что неоднородность вызвана не различием химического состава образца, а неодинаковостью температуры.

Опыт и теоретические расчеты показывают, что явление Томсона подчиняется следующему закону:

$$Q_{\tau} = \tau j \frac{dT}{dx}, \quad (6.29)$$

где Q_{τ} – тепло Томсона, выделяющееся (или поглощающееся) за единицу времени в единице объема полупроводника (удельная тепловая мощность);

j – плотность тока; $\frac{dT}{dx}$ – градиент температуры вдоль образца; τ – коэффициент Томсона, зависящий от природы полупроводника и его температуры.

Приведенная выше формула (так называемая *дифференциальная форма закона Томсона*) может быть применена к отрезку образца Δx , вдоль которого течет ток I и имеется некоторый перепад температур:

$$\Delta T = \frac{dT}{dx} \Delta x. \quad (6.30)$$

Закон Томсон в *интегральной форме* определяет полное количество тепла Томсона Q_{τ} , выделившегося (или поглотившегося) во всем

рассматриваемом объеме полупроводника ($\Delta V = S\Delta x$) за время t : $Q_T = Q_\tau \Delta V t$, или окончательно:

$$Q_T = Q_\tau S \Delta x t. \quad (6.31)$$

При этом коэффициент Томсона считается положительным, если электрический ток, текущий в направлении градиента температуры (dT/dx), вызывает нагревание полупроводника ($Q_\tau > 0$), и отрицательным, если при том же направлении тока происходит его охлаждение ($Q_\tau < 0$).

Объяснение эффекта Томсона для полупроводников с одним типом носителей (электроны или дырки) аналогично случаю металлических проводников. Во-первых, необходимо учесть изменение средней энергии носителей заряда вдоль образца из-за его неравномерного нагрева. В более нагретой части полупроводника средняя энергия электронов (или дырок) больше, чем в менее нагретой. Поэтому, если направление тока в полупроводнике соответствует движению носителей тока от горячего конца к холодному, то они будут передавать свою избыточную энергию кристаллической решетке, в результате чего происходит выделение теплоты Томсона ($Q_T > 0$).

При обратном направлении тока носители заряда, двигаясь от холодного конца к нагретому, будут пополнять свою энергию за счет решетки, т. е. происходит поглощение соответствующего количества теплоты ($Q_T < 0$).

В полупроводниках со смешанной проводимостью при наличии тока электроны и дырки движутся навстречу друг другу, и переносимые ими тепловые потоки будут компенсироваться. Так, на рис. 6.8 дырки движутся от горячего конца к холодному,

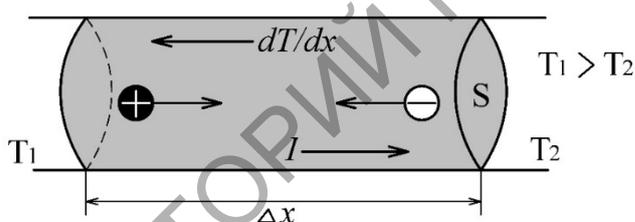


Рис. 6.8. Эффект Томсона в полупроводнике со смешанной проводимостью

что при отсутствии электронной проводимости должно приводить к выделению тепла. Однако с движением электронов (от холодного конца к горячему) связано поглощение тепла. В результате при равенстве концентраций и подвижностей электронов и дырок тепло Томсона не выделяется

($Q_T = 0$).

Второй фактор, который необходимо учесть, связан с электрическим полем термо-ЭДС, возникающим в условиях неоднородности температуры.

Рассмотрим полупроводник с электронной проводимостью. Пусть $T_1 > T_2$, т. е. градиент температуры направлен от точки 2 к точке 1 (рис. 6.9). Диффузия электронов от горячего конца к холодному приводит к разделению зарядов, в результате возникает электрическое поле термо-ЭДС E_T , направленное от 1 к 2, т. е. против градиента температуры. Если ток течет

в направлении градиента температуры (электроны движутся в направлении поля E_T), то поле E_T будет замедлять электроны, а участок полупроводника 1–2 станет охлаждаться ($Q_T < 0$). Если ток течет в обратном направлении, то произойдет нагревание участка 1–2 ($Q_T > 0$).

В дырочном полупроводнике соотношения будут обратными (рис. 6.10). Явление выглядит так, как если бы на обычный поток тепла, вызванный теплопроводностью, накладывался дополнительный поток тепла, связанный с прохождением электрического тока. В дырочных полупроводниках дополнительный поток тепла направлен в ту же сторону, куда течет электрический ток. В электронных полупроводниках направления тока и тепла противоположны. Рассмотренные факты действуют в противоположных направлениях, определяя не только величину, но и знак τ и Q_T .

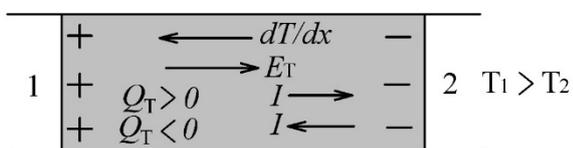


Рис. 6.9. Выделение и поглощение тепла Томсона в электронном полупроводнике (*n*-типа)

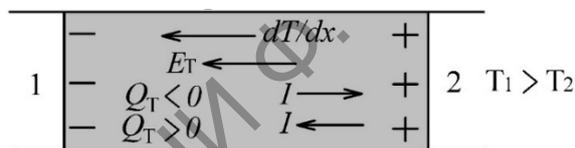


Рис. 6.10. Выделение и поглощение тепла Томсона в дырочном полупроводнике (*p*-типа)

Коэффициент Томсона связан с коэффициентами Пельтье Π и термо-ЭДС α соотношением Томсона:

$$\tau = -\frac{d\Pi}{dT} + \alpha = -T \frac{d\alpha}{dT}. \quad (6.32)$$

Для цепи, составленной из двух разнородных материалов, имеем:

$$\alpha_{12} = -\frac{d\Pi_{12}}{dT} + \tau_1 - \tau_2. \quad (6.33)$$

Учитывая эти соотношения, можно получить величину зависимости τ от температуры, концентрации носителей и др.

Из измерения коэффициента Томсона можно определить коэффициент термо-ЭДС одного материала, а не разность коэффициентов двух материалов, как при непосредственном измерении α и Π . Это позволяет, измерив τ и определив из него α в одном из материалов, получить абсолютную термо-электрическую шкалу.

Эффект Томсона не имеет технического применения, однако его необходимо учитывать в точных расчетах термоэлектрических устройств.

6.4. Поглощение света полупроводниками. Фотопроводимость

Согласно представлениям классической электронной теории, переменное электромагнитное поле световой волны, распространяющейся в конденсированной среде, вызывает вынужденные колебания связанных зарядов (электронов и ионов), входящих в состав молекул среды. Поэтому каждую молекулу среды можно рассматривать как систему осцилляторов с различными частотами собственных колебаний. Ионы значительно массивнее электронов и совершают заметные колебания только под действием низкочастотного (инфракрасного) излучения. В области частот видимого и ультрафиолетового излучения определяющую роль играют вынужденные колебания внешних, наиболее слабо связанных электронов атомов и молекул (их называют *оптическими электронами*).

Под действием падающего на вещество света электроны совершают вынужденные колебания с частотой ν падающего света, и дипольные электрические моменты молекул периодически изменяются. Следовательно, молекулы излучают вторичные электромагнитные волны, частота которых также равна ν . Вторичные волны, излучаемые большим числом соседних молекул среды, когерентны между собой и с первичной волной. При наложении они интерферируют. В однородном изотропном веществе в результате интерференции образуется проходящая волна, направление которой совпадает с направлением первичной волны. В оптически неоднородной среде в результате наложения первичной и вторичной волн возникает рассеяние света. Кроме того, при падении света на границу раздела двух сред в результате интерференции возникает не только проходящая, но и отраженная волна.

По мере распространения световой волны в веществе ее интенсивность уменьшается. Явление уменьшения интенсивности света, проходящего через среду вследствие взаимодействия его с частицами среды, называют **поглощением света**. В результате поглощения световая энергия переходит в другие виды энергии или в оптическое излучение другого спектрального состава. Основным законом поглощения света, связывающим интенсивность I световой волны, прошедшей слой среды толщиной l , с интенсивностью I_0 падающей волны, является закон Бугера – Ламберта:

$$I = I_0 \exp(-\alpha_\lambda l). \quad (6.34)$$

Французский физик П. Бугер (P. Vauger) открыл этот закон в 1729 г., а немецкий ученый, создатель фотометрии И. Ламберт (I. Lambert) уточнил его в 1760 г. Не зависящий от интенсивности света, но различный для

разных длин волн λ коэффициент α_λ назван *показателем (коэффициентом) поглощения* среды. Зависимость α_λ от длины волны света называют *спектром поглощения* вещества. В отличие от спектров поглощения атомов и молекул спектр поглощения конденсированных тел характеризуется очень широкими областями длин волн (сотни и тысячи нанометров) с большими значениями α_λ . Это объясняется тем, что в конденсированных средах сильное взаимодействие между частицами обуславливает быструю передачу всему коллективу частиц энергии, отданной светом одной из них.

Коэффициент поглощения имеет размерность обратной длины (см^{-1}) и может быть определен из уравнения (6.34) как

$$\alpha_\lambda = \frac{1}{l} \ln \frac{I_0}{I}. \quad (6.35)$$

Отсюда следует, что коэффициент поглощения пропорционален величине $D = \ln(I_0/I)$, которую принято называть *оптической плотностью* поглощения. Она имеет тот же физический смысл, что и коэффициент поглощения, но относится к толщине l слоя поглощающей среды: $D = d_\lambda l$. Часто пользуются понятием *прозрачности* (или пропускания) поглощающего слоя, определяя ее как отношение интенсивностей проходящей и падающей волн. Оптическая плотность поглощения D и коэффициент пропускания T связаны между собой соотношением: $D = \ln(1/T) = -\ln T$. Коэффициент пропускания зависит от размера, формы и состояния поверхности конденсированного тела, а также от угла падения на него потока излучения, спектрального состава тела (рис. 6.11) и поляризации излучения.

При поглощении света твердыми телами энергия фотонов превращается в другие виды энергии. Поэтому могут изменяться энергетическое состояние свободных или связанных с атомами электронов, а также колебательная энергия атомов. Рассмотрим механизмы поглощения света, характерные для полупроводников.

Собственное поглощение в полупроводниковых кристаллах связано с переходом электронов из валентной зоны в зону проводимости. В результате в зоне проводимости появляется свободный электрон, а в валентной

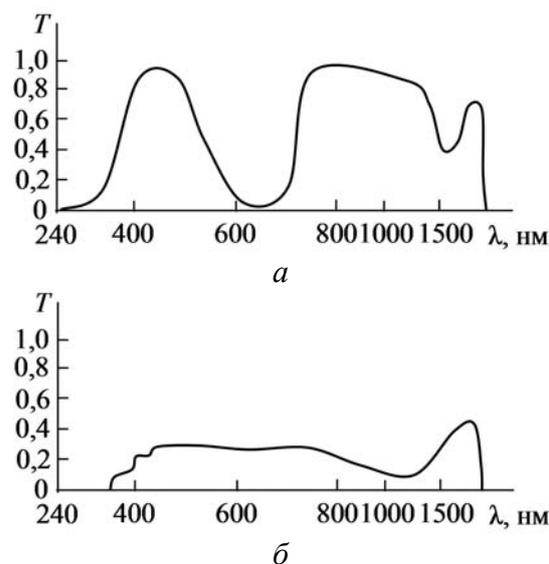


Рис. 6.11. Зависимость коэффициента пропускания от длины волны света: а – сине-зеленый; б – нейтральный светофильтры

зоне – дырка. Для фотонов с энергией, меньшей ширины запрещенной зоны ($h\nu < E_g$), кристалл прозрачен, т. е. фотоны не поглощаются. В области малых длин волн, т. е. при больших значениях $h\nu$, наблюдается сплошной спектр интенсивного поглощения, ограниченный так называемым краем поглощения при $h\nu \approx E_g$. У большинства полупроводников этот край находится в инфракрасной области спектра (рис. 6.11).

Если к кристаллу, облученному светом, приложить электрическое поле, то образовавшиеся в результате поглощения света свободные носители заряда приходят в движение, т. е. возникает проводимость. Изменение удельной электропроводности γ однородного полупроводника под действием излучения определяется концентрацией и подвижностью неравновесных носителей заряда:

$$\Delta\gamma = e(\mu_n\Delta n + \mu_p\Delta p), \quad (6.36)$$

где Δn , Δp – изменение концентрации электронов проводимости (n) и дырок (p), μ_n и μ_p – их подвижности.

Экситонное поглощение – механизм поглощения света, при котором электрон валентной зоны переводится в возбужденное состояние, но остается связанным с образовавшейся дыркой. Образуется экситон, перемещение которого не вызывает фотопроводимость кристалла, так как электрон и дырка движутся вместе. Экситоны могут возникать в диэлектриках, так как в них кулоновское притяжение электрона и дырки значительно. В полупроводниках это притяжение мало, и поэтому энергия связи экситона также мала.

Поглощение света свободными носителями заряда связано с переходом электронов (или дырок) с одного энергетического уровня на другой в пределах одной и той же разрешенной зоны. Такое поглощение характерно для полупроводников при достаточно больших концентрациях носителей зарядов. Оно плавно возрастает с ростом длины волны света, а коэффициент поглощения для каждой длины волны пропорционален концентрации основных носителей заряда в веществе.

Примесное поглощение наблюдается в полупроводниках, содержащих примесные атомы. Поглощение света обусловлено возбуждением примесных центров или их ионизацией. Например, в полупроводнике n -типа электроны с донорных уровней могут быть возбуждены в зону проводимости. Полосы примесного поглощения лежат за краем собственного поглощения полупроводника, поскольку энергия ионизации примесных уровней меньше, чем энергия, требуемая для перевода электронов из валентной зоны в зону проводимости.

Увеличение электропроводности полупроводников под действием электромагнитного излучения называют *фотопроводимостью полупроводников*. Фотопроводимость полупроводников может быть связана со свойствами как основного вещества, так и содержащихся в нем примесей. В первом случае при поглощении фотонов, энергия которых равна или больше ширины запрещенной зоны полупроводника ($h\nu \geq \Delta E_g$), могут совершаться перебросы электронов из валентной зоны в зону проводимости (рис. 6.12, а), что приведет к появлению добавочных (неравновесных) электронов (в зоне проводимости) и дырок (в валентной зоне). В результате возникает *собственная фотопроводимость*, обусловленная неравновесными электронами и дырками.

Если полупроводник содержит примеси, то фотопроводимость может возникать и при $h\nu < \Delta E$: для полупроводников с донорной примесью фотон должен обладать энергией $h\nu \geq \Delta E_d$, а для полупроводников с акцепторной примесью $h\nu \geq \Delta E_a$. При поглощении света примесными центрами происходит переход электронов с донорных уровней в зону проводимости в случае полупроводника *n*-типа или из валентной зоны на акцепторные уровни в случае полупроводника *p*-типа (рис. 6.12, б). В результате возникает *примесная фотопроводимость*, являющаяся чисто электронной для полупроводников *n*-типа и чисто дырочной для полупроводников *p*-типа.

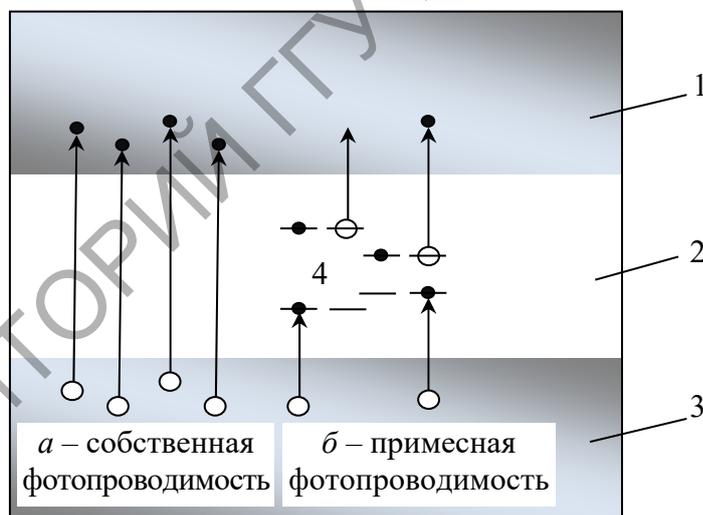


Рис. 6.12. Собственная (а) и примесная (б) фотопроводимость: 1 – зона проводимости, 2 – запрещенная зона, 3 – валентная зона, 4 – примесные уровни

Из условия $h\nu = hc/\lambda$ можно определить *красную границу фотопроводимости* – максимальную длину волны, при которой еще фотопроводимость возбуждается:

- для собственных полупроводников $\lambda_0 = hc/\Delta E_g$;

• для примесных полупроводников $\lambda_0 = hc/\Delta E_{\text{п}}$, где $\Delta E_{\text{п}}$ – в общем случае энергия активации примесных атомов.

Учитывая значения ΔE_g и $\Delta E_{\text{п}}$ для конкретных полупроводников, можно показать, что красная граница фотопроводимости для собственных полупроводников приходится на видимую область спектра, для примесных же полупроводников – на инфракрасную.

Тепловое или электромагнитное возбуждение электронов и дырок может и не сопровождаться увеличением электропроводности. Одним из таких механизмов может быть механизм возникновения экситонов. *Экситоны* представляют собой квазичастицы – электрически нейтральные связанные состояния электрона и дырки, образующиеся в случае возбуждения с энергией, меньшей ширины запрещенной зоны. Уровни энергии экситонов располагаются у дна зоны проводимости. Так как экситоны электрически нейтральны, то их возникновение в полупроводнике не приводит к появлению дополнительных носителей тока, вследствие чего экситонное поглощение света не сопровождается увеличением фотопроводимости.

Вопросы

1. Какие явления в полупроводниках относятся к термоэлектрическим/гальваномагнитным?
2. Какие сведения можно получить о примесном полупроводнике, измерив его удельную электропроводность и определив постоянную Холла?
3. Какой параметр электрона не учитывается при элементарном выводе постоянной Холла?
4. В чем заключается магниторезистивный эффект?
5. Какими механизмами обусловлено возникновение эффекта Зеебека?
6. В каких полупроводниках при наличии тока тепло Томсона не выделяется?
7. В какой области спектра находится край собственного поглощения у большинства полупроводниковых кристаллов?
8. В каких полупроводниках фотопроводимость возникает при меньших энергиях фотонов: в собственных или примесных, и почему?

Глава 7

ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ В СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ ПОЛЯХ

Концентрация и подвижность носителей заряда не зависят от напряженности электрического поля до некоторой ее величины, следовательно, и удельная электропроводность полупроводника γ не зависит от напряженности электрического поля. Электрические поля, которые практически не меняют подвижность и концентрацию носителей заряда, называются **слабыми**. Минимальная напряженность поля, при которой начинается заметная зависимость подвижности и концентрации носителей заряда от напряженности электрического поля, называется **критической**. Критическая напряженность $E_{кр}$ электрического поля зависит от природы полупроводника, температуры и концентрации примесей. Электрические поля, для которых подвижность или концентрация носителей заряда зависит от напряженности электрического поля, называются **сильными**. При напряженности поля выше критической линейность закона Ома уже не выполняется, т. е. величина плотности тока j не будет прямо пропорциональна напряженности поля, так как γ начинает зависеть от напряженности поля. Для значительного числа полупроводников величина $E_{кр}$ колеблется вблизи 10^6 В/м, для селена $E_{кр} \approx 10^3$ В/м. Напряженность $E_{кр}$ определяется тем условием, что дополнительная дрейфовая скорость, приобретаемая носителем заряда в поле, становится сравнимой с тепловой скоростью. При уменьшении температуры напряженность $E_{кр}$ уменьшается, так как она зависит от подвижности носителей заряда, а чем ниже температура, тем больше подвижность μ .

Нарушение закона Ома в сильных электрических полях может происходить по двум причинам. Во-первых, в сильных полях скорость дрейфа носителей заряда становится соизмеримой с их тепловой скоростью. Это влияет на процессы рассеяния, что проявляется в изменении времени релаксации τ , а следовательно, и подвижности μ . Сюда можно отнести явление «разогрева» электронно-дырочного газа и эффект Ганна. Во-вторых, сильные электрические поля могут вызывать изменение концентрации носителей заряда, что может быть обусловлено рядом эффектов. Такими эффектами являются ударная ионизация, туннельный эффект (эффект Зинера) и электростатическая ионизация.

7.1. «Разогрев» электронного газа

В слабых электрических полях электропроводность полупроводников подчиняется закону Ома, т. е. плотность тока пропорциональна напряженности электрического поля:

$$\mathbf{J} = en\mu_0\mathcal{E} = \gamma_0\mathcal{E}, \quad (7.1)$$

где μ_0 , γ_0 – подвижность и удельная проводимость при слабом электрическом поле.

При столкновении с акустическим фононом электрон может либо передать энергию фонону, либо получить энергию от фонона. Пусть при этом изменение энергии электрона будет ΔE . Поэтому относительная потеря энергии электрона при столкновении составляет

$$\delta = \frac{\Delta E}{kT}. \quad (7.2)$$

Следовательно, изменение энергии электрона при столкновении с фононом равно $\Delta E = kT\delta$. Соударение будет упругим в том случае, если энергия, набираемая электроном в электрическом поле на длине свободного пробега $E = e\mathcal{E}l$, будет значительно меньше энергии ΔE , теряемой в среднем при одном столкновении с решеткой, т. е.

$$\frac{e\mathcal{E}l}{\Delta E} \ll 1. \quad (7.3)$$

Это означает, что закон Ома выполняется в том случае, когда

$$e\mathcal{E}l \ll kT\delta, \text{ или } \mathcal{E} \ll \delta \frac{kT}{el}. \quad (7.4)$$

При этом условии распределение скоростей носителей заряда остается почти максвелловским с температурой, равной температуре решетки. В данном случае на хаотическое движение носителей заряда накладывается их медленный дрейф, обусловленный действием слабого внешнего электрического поля. При этом скорость дрейфа носителей заряда v значительно меньше их тепловой скорости v_0 , т. е.

$$v \ll v_0. \quad (7.5)$$

Рассмотрим электропроводность полупроводника в сильном электрическом поле. В сильных электрических полях, для которых не выполняется условие (6.4), носители заряда при столкновении с акустическими фононами не успевают отдавать кристаллической решетке набираемую ими в поле энергию. В результате средняя энергия носителей заряда повышается

и электронно-дырочный газ «разогревается». Но с ростом энергии носителей заряда увеличивается их скорость, а значит, уменьшается время свободного пробега и следовательно, увеличивается вероятность рассеяния носителей заряда на тепловых колебаниях решетки. В силу этого рост энергии носителей заряда с возрастанием напряженности поля будет продолжаться до тех пор, пока электронно-фононное взаимодействие не усилится настолько, что установится стационарное состояние, при котором энергия, приобретаемая носителем заряда в электрическом поле за единицу времени $e\varepsilon v$, сравняется с энергией, теряемой этим носителем заряда за одну секунду в результате столкновения $\Delta E/\tau = kT\delta/\tau$, т. е. $e\varepsilon v \ll \delta kT/\tau$.

Критическая напряженность поля, при которой начинается разогрев электронно-дырочного газа, равна:

$$\varepsilon_{\text{кр}} \ll \frac{kT}{el} (2\delta)^{1/2}, \quad (7.6)$$

т. е. напряженность критического поля уменьшается при снижении температуры. Это согласуется с экспериментальными данными.

Разогрев электронно-дырочного газа будет происходить в том случае, когда скорость, приобретенная носителем заряда во внешнем электрическом поле, станет сравнимой по величине с его тепловой скоростью:

$$v \approx v_0. \quad (7.7)$$

Таким образом, в сильных электрических полях, когда напряженность электрического поля больше критического, т. е.

$$\varepsilon > \varepsilon_{\text{кр}}, \text{ или } \varepsilon > \frac{kT}{el} (2\delta)^{1/2}, \quad (7.8)$$

дрейфовая скорость движения носителей заряда будет увеличиваться пропорционально корню квадратному из величины напряженности поля:

$$v \sim \varepsilon^{1/2}, \quad (7.9)$$

а время релаксации соответственно будет уменьшаться:

$$\tau = \frac{l}{v_0 + v} \sim \varepsilon^{-1/2}. \quad (7.10)$$

Отсюда следует, что подвижность

$$\mu_n = \frac{e\tau}{m^*} = \mu_0 \varepsilon^{-1/2}, \quad (7.11)$$

а удельная электропроводность и плотность тока:

$$\gamma = en\mu_n = \gamma_0 \varepsilon^{-1/2}; J = \gamma \varepsilon = \gamma_0 \varepsilon^{-1/2}. \quad (7.12)$$

Следовательно, при напряженности внешнего электрического поля $\mathcal{E} > \mathcal{E}_{\text{кр}}$ закон Ома не выполняется. Например, в германии отклонение от омической проводимости при комнатной температуре наступает при напряженности поля около 700 В/см.

В области еще более сильных электрических полей энергия носителей заряда увеличивается настолько, что возможно рассеяние их на оптических колебаниях. В этом случае при столкновении с оптическим фоном энергия носителя изменяется на значение, равное $\hbar\omega_0$, и скорость потерь энергии сильно возрастает. В результате устанавливается новое стационарное состояние, при котором энергия $e\mathcal{E}\nu$, набираемая носителем заряда в поле на единицу времени, равна энергии $\Delta E/\tau$, теряемой этим носителем заряда за то же время, т. е.

$$e\mathcal{E}\nu = \frac{\Delta E}{\tau}. \quad (7.13)$$

В пределе плотность тока, текущего через образец, стремится к насыщению и равна

$$J_s = env_s = en(\hbar\omega_0/m^*)^{1/2}, \quad (7.14)$$

где индекс s означает насыщение (от англ. *saturation*). Это значит, что

$$\gamma = \frac{J_s}{\mathcal{E}} = \gamma_0 \mathcal{E}^{-1}. \quad (7.15)$$

При увеличении напряженности электрического поля до $\mathcal{E}'_{\text{кр}}$ (рис. 7.1) в ковалентных кристаллах наступает насыщение тока, а следовательно, и насыщение дрейфовой скорости.

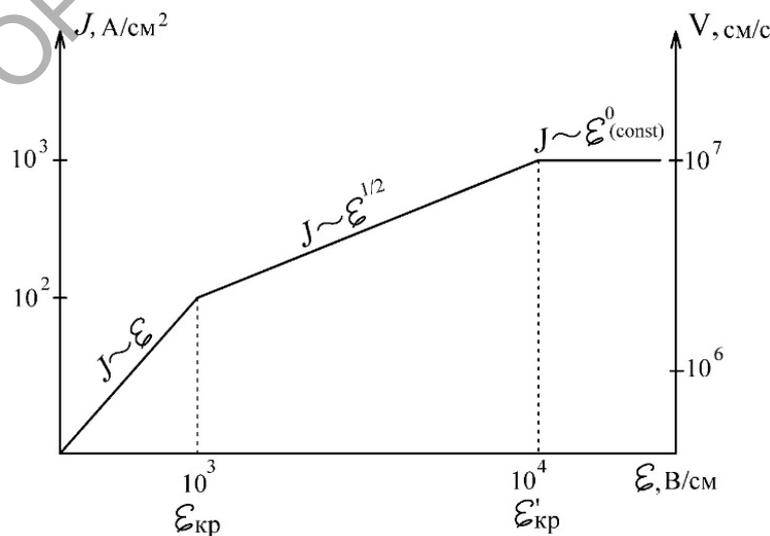


Рис. 7.1. Плотность тока и скорость носителей заряда в сильных электрических полях

При $\mathcal{E} > \mathcal{E}'_{кр}$ ток и дрейфовая скорость не зависят от напряженности внешнего электрического поля. Оценки показывают, что насыщение тока и дрейфовой скорости наступает в полях порядка 10^4 В/см, а дрейфовая скорость достигает $v \approx 10^7$ см/с. Это согласуется с экспериментальными данными.

7.2. Ударная ионизация

В сильном электрическом поле, если оно возрастает примерно до 10^5 В/см, электроны (или дырки) приобретают энергию, достаточную для ионизации атомов вещества. В результате этого возникают электронно-дырочные пары (рис. 7.2), которые в свою очередь ускоряются полем и генерируют дополнительные носители заряда. Этот процесс создания носителей заряда называется *ударной ионизацией*.

Наблюдать, однако, это явление возможно лишь в *p-n*-переходах, так как создать поля 10^5 – 10^6 В/см в однородном полупроводнике чрезвычайно трудно при разумных толщинах последнего. Исключением является ударная ионизация примесных атомов, которая требует небольших полей и наблюдается при низких температурах, когда атомы примеси не ионизированы. Ударная ионизация доноров и акцепторов – элементов V и III групп – в германии происходит при полях, приблизительно равных 5–10 В/см. Например, из рис. 7.3, на котором представлена зависимость плотности тока от напряженности электрического поля \mathcal{E} в электронном германии при 4,2 К, следует что при $\mathcal{E} \approx 5$ В/см происходит резкое скачкообразное возрастание плотности тока, вызванное ударной ионизацией атомов примеси.

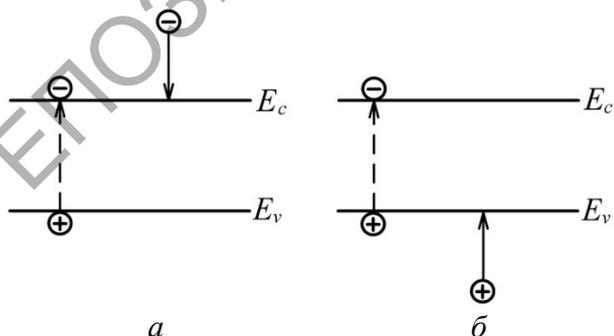


Рис. 7.2. Два возможных варианта ударной ионизации

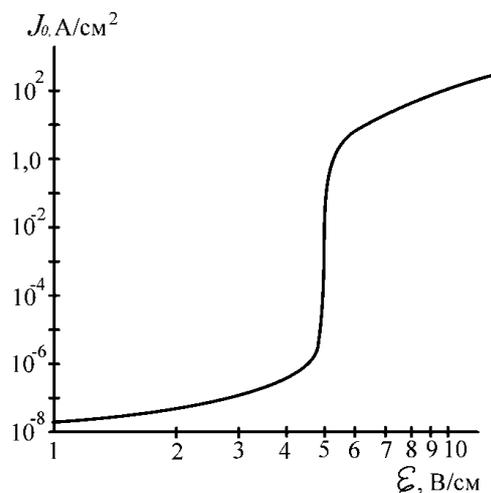


Рис. 7.3. Ударная ионизация в электронном германии при $T = 4,2$ К

Скорость генерации электронно-дырочных пар G вследствие ударной ионизации дается в виде

$$G = \alpha_n n \mu_n + \alpha_p p \mu_p, \quad (7.16)$$

где α_n и α_p – коэффициент ударной ионизации электронов и дырок соответственно.

Он определяется как число электронно-дырочных пар, генерируемых электроном (дыркой) на единице пути перемещения. Коэффициент ударной ионизации как электронов α_n , так и дырок α_p сильно зависит от значения напряженности электрического поля и определяется соотношением:

$$\alpha(E) = \frac{1}{v} \int_0^\infty W_{и}(E) f(E) N(E) dE, \quad (7.17)$$

где $W_{и}(E)$ – вероятность ионизации, рассчитываемая квантовомеханическим путем; $f(E)$ – неравновесная функция распределения; $N(E)$ – плотность квантовых состояний; v – скорость дрейфа.

Для полупроводников с большой диэлектрической проницаемостью (например, для Ge и Si) вероятность ионизации:

$$W_{и}(E) \approx A(E - E_{и})^2, \quad (7.18)$$

где A – коэффициент пропорциональности; $E_{и} = 3/2(E_g)$ – пороговая энергия ионизации; E_g – ширина запрещенной зоны.

Ускорение электронов до порога ионизации $E_{и}$ зависит от соотношения двух факторов – ускорения во внешнем электрическом поле и рассеяния энергии при столкновении с фононами. Наибольший вклад в релаксацию по энергиям вносят оптические фононы. При столкновениях с оптическим фононом энергия, теряемая электроном, равна $(\hbar\omega_0) = \text{const}$, а средняя длина свободного пробега не зависит от энергии, т. е. $l = \text{const}$.

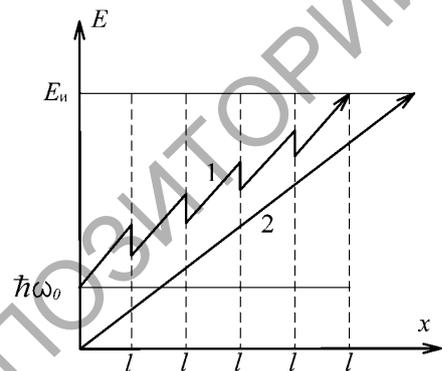


Рис. 7.4. Условная диаграмма ударной ионизации

Ускорение до порога ионизации $E_{и}$ можно проиллюстрировать условной диаграммой, изображенной на рис. 7.4, на котором по вертикали отложена энергия, приобретаемая электроном во внешнем электрическом поле, а по горизонтали – расстояние, проходимое электроном в среднем между двумя ионизирующими столкновениями.

Как следует из рис. 7.4, существует две возможности ускорения электронов до порога ионизации. В первом случае в очень сильных электрических полях, когда выполняется условие $eEl \gg \hbar\omega_0$, электроны могут набрать энергию в результате многократных столкновений с фононами, так

как относительная потеря энергии в каждом таком столкновении мала (линия 1). В этом случае, как показывает теория,

$$\alpha(\mathcal{E}) \sim e^{-E_n/\mathcal{E}^2}. \quad (7.19)$$

Здесь можно провести аналогию с диффузионным движением частицы, поэтому указанное приближение для $\alpha(\mathcal{E})$ называют *диффузионным*.

Во втором случае, когда поля относительно слабые ($e\mathcal{E}l \ll \hbar\omega_0$), носители заряда набирают энергию на расстоянии, равном многим длинам свободного пробега (линия 2 на рис. 7.4). В такой ситуации существует только одна возможность набрать энергию E_n , когда носители заряда не сталкиваются с фононами до порога ионизации. В связи с этим функция распределения, а следовательно, и коэффициент ударной ионизации пропорциональны

$$\alpha(\mathcal{E}) \sim e^{-E_n/e\mathcal{E}l}. \quad (7.20)$$

Функция распределения будет сильно вытянута вдоль направления электрического поля, т. е. в направлении дрейфа носителей, поэтому это приближение для $\alpha(\mathcal{E})$ называют *дрейфовым*.

Как показали эксперименты, в Ge и Si $\alpha(\mathcal{E})$ определяется в основном законом вида (7.20). Однако в узких *p-n*-переходах (особенно четко это проявляется в Si) наблюдается зависимость вида (7.19). Это имеет место при напряженности поля $\mathcal{E} \geq 5 \cdot 10^5$ В/см.

7.3. Туннельный эффект и электростатическая ионизация

В очень сильных электрических полях становится возможным еще один механизм образования свободных носителей заряда – *туннельный эффект*, или эффект Зинера. Как известно, у полупроводника, находящегося во внешнем электрическом поле, имеет место наклон энергетических зон. Наклон зон тем больше, чем больше напряженность электрического поля \mathcal{E} . На рис. 7.5 представлена зонная структура собственного полупроводника, находящегося в сильном электрическом поле. В этом случае возможен переход электрона из валентной зоны в зону проводимости благодаря туннельному эффекту. Вероятность туннельного перехода зависит от высоты и ширины потенциального барьера.

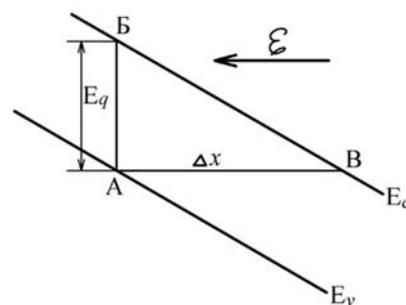


Рис. 7.5. Зонная структура собственного полупроводника при наличии сильного электрического поля

В данном случае высота потенциального барьера AB представляет собой ширину запрещенной зоны E_g . Эффективная ширина барьера $AB = \Delta x$ может быть определена из разности потенциальной энергии электрона в зоне проводимости в точке B и в валентной зоне – в точке A . Так как потенциальная энергия с точностью до аддитивной постоянной есть

$$U(B) - U(A) = -e\mathcal{E}\Delta x = -E_g,$$

то эффективная ширина барьера

$$\Delta x = \frac{E_g}{e\mathcal{E}}. \quad (7.21)$$

Следовательно, ширина потенциального барьера зависит от напряженности электрического поля.

Переход электрона из точки A в точку B связан с переходом сквозь треугольный потенциальный барьер ABB . Но, как известно из квантовой механики, вероятность перехода для барьера треугольной формы имеет вид:

$$D = D_0 e^{-\frac{8\pi\sqrt{2m_n^*}E_g^{3/2}}{3e\hbar\mathcal{E}}}. \quad (7.22)$$

Из этого выражения следует, что *вероятность туннельного перехода из зоны в зону растет экспоненциально с ростом напряженности электрического поля \mathcal{E}* . Кроме того, *вероятность туннельного перехода будет больше у полупроводников с малыми значениями ширины запрещенной зоны E_g и эффективной массой m^** .

Вероятность туннельного перехода одна и та же для переходов как из валентной зоны в зону проводимости, так и из зоны проводимости в валентную зону. Но поскольку в валентной зоне электронов больше, чем в зоне проводимости, то будет иметь место преимущественный переход электронов из валентной зоны в зону проводимости. Поэтому туннельный эффект может приводить к значительному увеличению концентрации свободных носителей заряда. Он наблюдается в полях 10^6 В/см.

Увеличение концентрации свободных носителей заряда может быть также обусловлено *электростатической ионизацией*, т. е. уменьшением энергии ионизации атомов донорной примеси в сильном электрическом поле. На рис. 7.6 представлена потенциальная энергия электрона в атоме донорной примеси без поля (пунктирная кривая) и с внешним электрическим полем (сплошная кривая). При отсутствии электрического поля электрон локализован в потенциальной яме, образованной положительно заряженным донором, а энергия, необходимая для его ионизации, равна E_H .

Внешнее электрическое поле понижает потенциальный барьер в направлении, противоположном направлению электрического поля. Потенциальная

энергия электрона с учетом действия внешнего электрического поля, направленного вдоль вектора \mathbf{r} , может быть записана в виде:

$$U(\mathbf{r}) = -e\mathcal{E}r - \frac{e^2}{4\pi\epsilon_r\epsilon_0 r}. \quad (7.23)$$

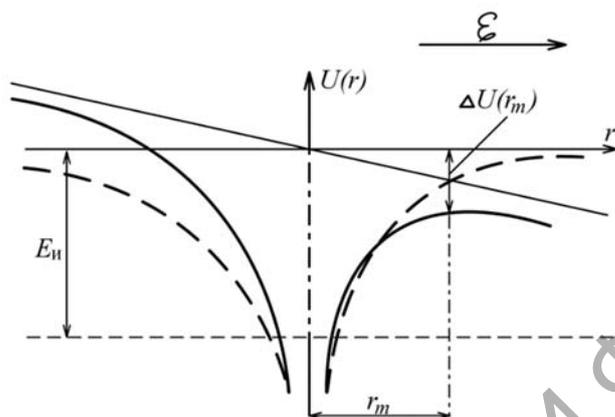


Рис 7.6. Понижение потенциального барьера под действием внешнего электрического поля

Первое слагаемое в этом равенстве отражает изменение потенциальной энергии во внешнем электрическом поле напряженностью \mathcal{E} , а второе – в кулоновском поле донора. Как следует из (7.23), энергия $U(\mathbf{r})$ в некоторой точке \mathbf{r}_m принимает экстремальное значение, а потенциальный барьер в этой точке снижается на

$$\Delta U(\mathbf{r}_m) = -e\sqrt{e\mathcal{E} / \pi\epsilon_r\epsilon_0}, \quad (7.24)$$

а следовательно, на такое же значение уменьшается и энергия ионизации донора.

Напряженность электрического поля, при которой имеет место заметное возрастание концентрации электронов, обычно составляет 10^4 – 10^5 В/см. Этот механизм образования свободных носителей заряда проявляется только в том случае, когда концентрация неионизированной донорной примеси значительно превышает концентрацию свободных электронов.

7.4. Эффект Ганна

Эффект Ганна – явление возникновения осцилляций тока ($\sim 10^9$ – 10^{10} Гц) в однородном многодолинном полупроводнике при приложении к нему сильного электрического поля. Впервые этот эффект наблюдал британский физик Джон Ганн в 1963 г. на арсениде галлия, затем явление осцилляции

тока было обнаружено в фосфиде индия, фосфиде галлия и ряде других полупроводниковых соединений.

Эффект Ганна может возникнуть в полупроводнике, в котором в зоне Бриллюэна имеется более одного минимума энергии. На рис. 7.7 показаны основной минимум, которым определяется ширина запрещенной зоны, и побочный минимум, смещенный на конечный волновой вектор от нуля зоны, имеющий большее расстояние до потолка валентной зоны, чем основной минимум, как например в GaAs, InAs. В полупроводниках, зона проводимости которых имеет более одного минимума энергии, электрон с волновым вектором \mathbf{k} , соответствующим одному из минимумов, при рассеянии может оказаться в состоянии с волновым вектором \mathbf{k}' , принадлежащим другому минимуму. В результате такого рассеяния будет иметь место переброс электронов из одного минимума зоны проводимости в другой. Такой вид рассеяния получил название *междолинного*.

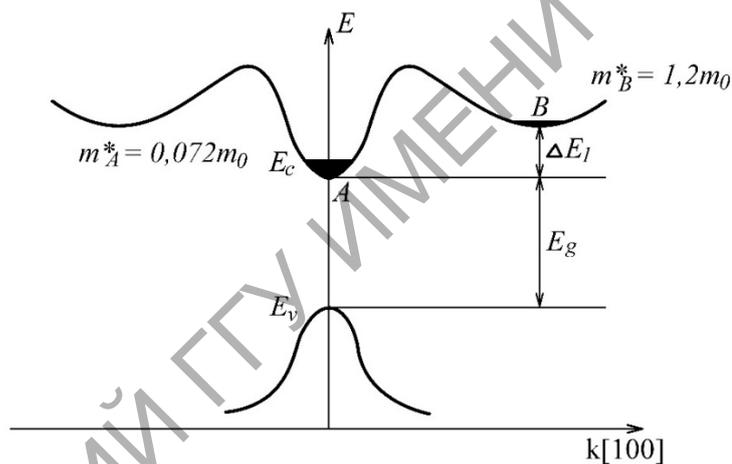


Рис. 7.7. Зонная структура GaAs в направлении [100]

Рассмотрим энергетическую структуру GaAs n -типа в направлении [100]. Возможны переходы из минимума A с состоянием \mathbf{k} в минимум B с состоянием \mathbf{k}' . Минимумы A и B разделены энергетическим интервалом ΔE_1 . Вблизи минимумов закон дисперсии можно представить в виде параболического с разной кривизной для долин A и B . Эффективные массы электронов в них также различны и равны $m_A^* = 0,072m_0$ и $m_B^* = 1,2m_0$ соответственно. Подвижность легких электронов выше, чем подвижность тяжелых электронов $\mu_A \gg \mu_B$. Плотность состояний в верхней долине примерно в 70 раз выше, чем в нижней.

При слабых внешних полях электроны находятся в термодинамическом равновесии с решеткой и, поскольку при обычных температурах $k_0T \ll E_g$, электроны в основном занимают энергетические состояния вблизи минимума A . Плотность тока $J = en_A\mu_A\mathcal{E}$ определяется концентрацией легких

электронов и их подвижностью. В этом случае концентрация электронов $n_0 = n_A$, $n_B = 0$. Плотность тока будет линейно возрастать с ростом напряженности поля до некоторого критического значения \mathcal{E}_1 (участок 1 на рис. 7.8).

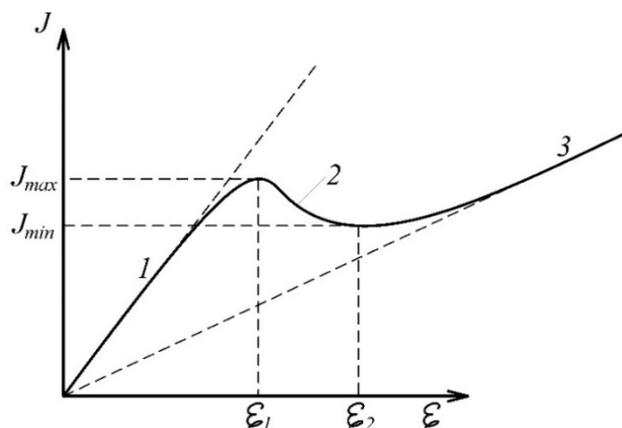


Рис. 7.8. N-образная вольт-амперная характеристика: \mathcal{E} – электрическое поле, создаваемое приложенной разностью потенциалов; J – плотность тока

По мере возрастания \mathcal{E} средняя энергия и скорость электронов повышается, и при $\mathcal{E} > E_g$ становится возможным переход электронов в долину B . Тогда суммарная концентрация электронов будет $n_0 = n_A + n_B$. Таким образом, с ростом напряженности от \mathcal{E}_1 до некоторого значения \mathcal{E}_2 будет иметь место уменьшение подвижности электронов, а следовательно, уменьшение плотности тока J , и на вольт-амперной характеристике появится падающий участок 2. При дальнейшем росте напряженности поля ($\mathcal{E} > \mathcal{E}_2$) все электроны перейдут в минимум B , и снова установится линейная ВАХ (участок 3), при этом суммарная концентрация электронов будет $n_0 = n_B$, а $n_A = 0$.

Таким образом, эффект Ганна – это генерация ВЧ колебаний электрического тока в полупроводнике с N-образной вольт-амперной характеристикой (рис. 7.8).

Генерация возникает, если постоянное напряжение U , приложенное к образцу длиной l , таково, что среднее значение электрического поля \mathcal{E} в образце равно: $\mathcal{E} = U/l$, что соответствует падающему участку вольт-амперной характеристики на участке \mathcal{E}_1 – \mathcal{E}_2 , на котором дифференциальное сопротивление отрицательно (рис. 7.8). Колебания тока имеют вид периодических последовательностей импульсов (рис. 7.9), частота их повторения обратно пропорциональна напряженности поля \mathcal{E} .

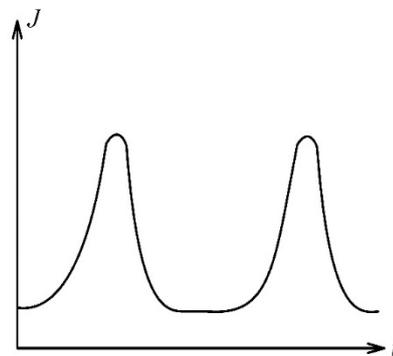


Рис. 7.9. Форма колебаний тока при эффекте Ганна

Причиной эффекта Ганна является то, что в образце периодически появляется, перемещается по нему и исчезает область сильного электрического поля, называемая *доменом Ганна*. Домен возникает в результате того, что однородное распределение электрического поля при объемном отрицательном дифференциальном сопротивлении неустойчиво. Действительно, если в полупроводнике случайно возникает неоднородное рас-

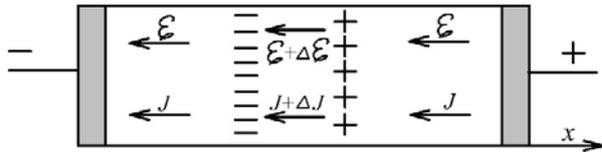


Рис. 7.10. Развитие электрического домена

пределение концентраций электронов в виде дипольного слоя, то между заряженными областями создается дополнительное поле ΔE (рис. 7.10).

Если область повышенной концентрации электронов находится ближе к катоду, то ΔE добавляется к внешнему полю, так что поле внутри дипольного слоя становится больше, чем вне его. Если при этом дифференциальное сопротивление образца положительно, т. е. ток растет с ростом поля, то ток и внутри слоя больше, чем вне его ($\Delta J > 0$). Поэтому из области с повышенной плотностью электронов они вытекают в большем количестве, чем втекают, в результате чего неоднородность рассасывается. Если дифференциальное сопротивление отрицательно (ток уменьшается с ростом поля), то ток меньше там, где E больше, т. е. внутри слоя, и неоднородность не рассасывается, а нарастает. Увеличивается и падение напряжения на дипольном слое, а вне его уменьшается (так как полное напряжение на образце задано). В результате образуется электрический домен (рис. 7.10). Вне домена $E < E_1$, благодаря чему новые домены не образуются. Устойчивое состояние образца – состояние с одним доменом.

Поскольку домен образован электронами проводимости, он движется в направлении их дрейфа со скоростью v , близкой к дрейфовой скорости носителей вне домена. Обычно домен возникает вблизи катода и, дойдя до анода, исчезает. По мере его исчезновения падение напряжения на домене уменьшается, а на остальной части образца соответственно растет. Одновременно возрастает ток в образце, так как увеличивается поле вне домена. По мере приближения напряженности поля E к E_1 плотность тока j приближается к j_{\max} . Когда вне домена $E > E_1$, у катода начинает формироваться новый домен, ток уменьшается и процесс повторяется. Частота колебаний тока $f = v/l$.

В GaAs с электронной проводимостью при комнатной температуре $E_1 \approx 3 \cdot 10^3$ В/см, $v \approx 10^7$ см/с и при $l = 50\text{--}300$ мкм, $f = 0,3\text{--}2$ ГГц. Размер домена $\approx 10\text{--}20$ мкм. Помимо GaAs и InP эффект Ганна наблюдается в электронных полупроводниках CdTe, ZnS, InSb, InAs и др., а также в Ge с дырочной проводимостью. Созданные на основе эффекта Ганна источники

СВЧ-колебаний называют генераторами на диодах Ганна, или просто *генераторами Ганна*.

Вопросы

1. По какому критерию разделяют слабое и сильное электрическое поле при рассмотрении электропроводности полупроводников?
2. Какие процессы в полупроводниках приводят к отклонению электропроводности от закона Ома?
3. При каком условии начинается разогрев электронно-дырочного газа? При каких напряженностях внешнего поля в полупроводниках происходит насыщение тока и дрейфовой скорости?
4. Какой процесс создания носителей заряда называется ударной ионизацией?
5. От чего зависит вероятность туннельного перехода электронов в полупроводниках?
6. В каких полупроводниках может возникнуть эффект Ганна?

Глава 8

КОНТАКТНЫЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

Контактные явления в полупроводниках – это неравновесные электронные процессы, возникающие при прохождении электрического тока: 1) через контакт полупроводника с металлом или другим полупроводником (гетеропереход) либо 2) через границу двух областей одного и того же полупроводника, отличающихся типом носителей заряда (p-n-переход) или разной их концентрацией. Контакты полупроводника с металлом или другим полупроводником обладают выпрямляющим свойством, т. е. значительно эффективнее пропускают электрический ток в одном направлении, чем в обратном. Это свойство обусловило их широкое применение в технике.

8.1. Электронно-дырочный переход

Электронно-дырочный, или p-n-переход – переходная область между двумя соприкасающимися одинаковыми полупроводниками, один из которых имеет электронную, а другой – дырочную проводимость. Концентрация дырок в p-области гораздо выше, чем в n-области, поэтому дырки из p-области диффундируют в n-область, а электроны из n-области – в p-область. В результате в p-области остаются отрицательно заряженные акцепторные атомы, а в n-области – положительно заряженные донорные атомы. Так как и те и другие неподвижны, в области p-n-перехода образуется двойной слой пространственного заряда: отрицательные заряды в p-области и положительные – в n-области (рис. 8.1, а). Возникающее поле E_k двойного электрического слоя препятствует дальнейшей диффузии основных носителей тока. Поле E_k обуславливает контактную разность потенциалов:

$$V_k = \frac{kT}{e} \ln \left(\frac{np}{n_0^2} \right), \quad (8.1)$$

где e – заряд электрона, T – температура полупроводника, n и p – концентрации электронов и дырок в n- и p-областях, n_0 – концентрация электронов в собственном полупроводнике.

При определенной толщине двойного электрического слоя наступает равновесное состояние контакта, характеризуемое выравниванием уровня Ферми E_F для обоих полупроводников (рис. 8.1, в). В области p-n-перехода энергетические зоны искривляются, в результате чего возникают

потенциальные барьеры eV_k как для электронов, так и для дырок. Толщина d слоя p - n -перехода в полупроводниках составляет 10^{-6} – 10^{-7} м, а контактная разность потенциалов V_k – десятые доли вольт. При обычных температурах носители тока не способны преодолеть потенциальный барьер eV_k , поэтому контактный слой является *запирающим*, т. е. характеризуется повышенным сопротивлением.

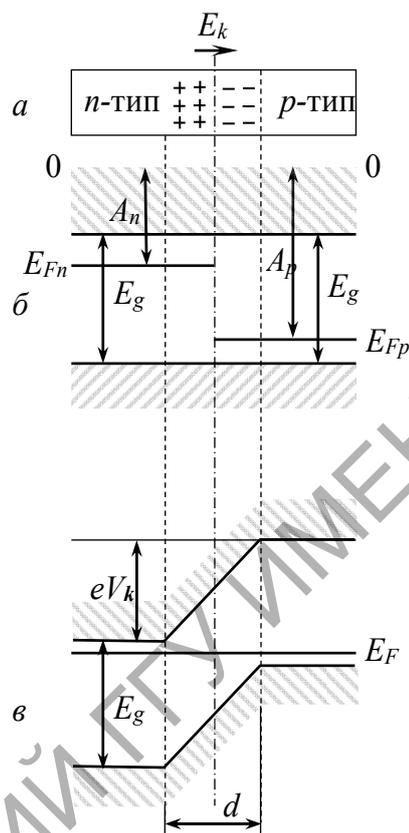


Рис. 8.1. Схема электронно-дырочного перехода: ***a*** – образование двойного электрического слоя в контакте полупроводников с n - и p -типами проводимости; ***б*** и ***в*** – энергетические зоны в изолированных друг от друга полупроводниках и после образования p - n -перехода. A_n и A_p – работы выхода электронного и дырочного полупроводников; E_g – ширина запрещенной зоны; E_{Fn} и E_{Fp} – уровни Ферми; eV_k – высота потенциального барьера; d – толщина слоя p - n -перехода

Примечание: работа выхода – разность значений энергий уровня вакуума и уровня Ферми, т. е. минимальная энергия, которую необходимо сообщить электрону для его удаления из объема твердого тела – металла или полупроводника: $A = E_{vac} - E_F$.

Внешнее электрическое поле изменяет высоту потенциального барьера eV_k и нарушает равновесие потоков носителей тока через барьер. Внешнее поле, направление которого совпадает с полем E_k контактного слоя (рис. 8.2, *a*), вызывает движение электронов в n -полупроводнике и дырок в p -полупроводнике от границы p - n -перехода в противоположные стороны. В результате запирающий слой расширится и его сопротивление возрастет.

Такое направление внешнего поля называют *запирающим (обратным)*. В запирающем направлении электрический ток через $p-n$ -переход практически не проходит.

Внешнее электрическое поле, приложенное к $p-n$ -переходу, противоположно полю контактного слоя (рис. 8.2, б) и вызывает движение электронов и дырок к границе $p-n$ -перехода навстречу друг другу.

В области перехода носители рекомбинируют, толщина и сопротивление контактного слоя уменьшаются. При таком направлении внешнего поля электрический ток проходит сквозь $p-n$ -переход от p - к n -полупроводнику. В результате $p-n$ -переход приобретает *одностороннюю проводимость*, а направление приложенного поля называют *пропускным, или прямым*.

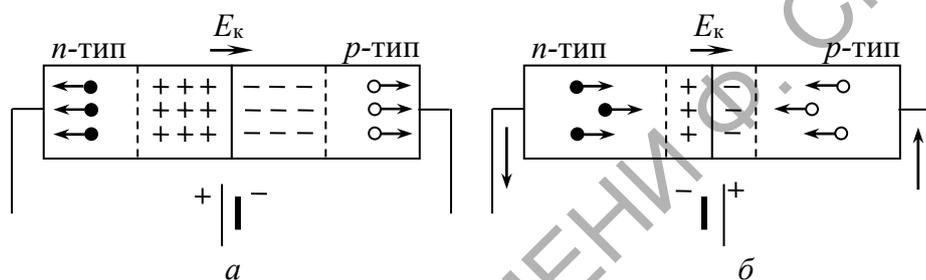


Рис. 8.2. Электронно-дырочный переход во внешнем электрическом поле: а – запирающее направление; б – пропускное направление

Зависимость тока, протекающего через $p-n$ -переход, от приложенного напряжения (вольт-амперная характеристика) обладает резко выраженной нелинейностью (рис. 8.3). При изменении знака U значение тока через $p-n$ -переход изменяется в 10^5 – 10^6 раз. Благодаря этому электронно-дырочный переход является вентильным элементом, используемым в полупроводниковых диодах для выпрямления переменного тока.

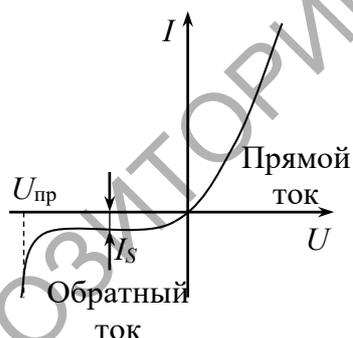


Рис. 8.3. Вольт-амперная характеристика $p-n$ -перехода: I_S – ток насыщения; $U_{пр}$ – напряжение пробоя

Гетеропереход – контакт двух разных по химическому составу полупроводников. На границе гетероперехода происходит изменение свойств полупроводникового материала: структуры энергетических зон, ширины запрещенной зоны, эффективных масс носителей заряда, их подвижности и т. д. Гетеропереход называют *изотипным*, если он образован полупроводниками с одинаковым типом проводимости, и *анизотипным*, если механизмы проводимости контактирующих полупроводников разные.

Наибольшее практическое значение имеют гетеропереходы, образованные монокристаллами. Для получения идеальных монокристаллических гетеропереходов (без дефектов кристаллической решетки и поверхностных состояний на границе раздела полупроводников) необходимо, чтобы у полупроводников совпадали коэффициенты термического расширения, типы и периоды кристаллических решеток. Для получения гетеропереходов, близких к идеальным, периоды решеток полупроводников должны совпадать с точностью до $\sim 0,1\%$. В зависимости от способа получения гетероперехода толщина d переходной области между двумя однородными полупроводниками может колебаться в широких пределах. В наиболее «резких» гетеропереходах $d \approx 20 \text{ \AA}$, т. е. равна толщине 4–5 атомных слоев. Комбинации различных гетеропереходов образуют *гетероструктуры*.

Механизм протекания тока в гетеропереходах обусловлен различием в высотах потенциальных барьеров для электронов и дырок. При так называемом прямом смещении на анизотипном гетеропереходе не одинаковы как потоки носителей между полупроводниками с разной шириной запрещенной зоны, так и токи инжектированных электронов и дырок. Поэтому в гетеропереходах обычно происходит односторонняя инжекция носителей из широкозонного полупроводника (*эмиттера*) в узкозонный.

В электронной технике широко применяют гетеропереходы между полупроводниковыми монокристаллами типа $A^{III}B^V$, например на основе арсенида галлия. Гетероструктуры получают также на основе многокомпонентных твердых растворов, период решетки которых не изменяется при изменении состава раствора в широких пределах. Изготавливают гетероструктуры методами эпитаксиального наращивания полупроводниковых кристаллов. Гетеропереходы используются в полупроводниковых приборах: транзисторах, тиристорах, гетеролазерах, светодиодах, фотодетекторах и др.

8.2. Электрические процессы в p - n -переходе в отсутствие внешнего поля

Обе области полупроводника, изображенные на рис. 8.4, *a*, электрически нейтральны, поскольку как сам материал полупроводника, так и примеси электрически нейтральны. Отличия этих областей в том, что левая из них содержит свободно перемещающиеся дырки, а правая – свободно перемещающиеся электроны.

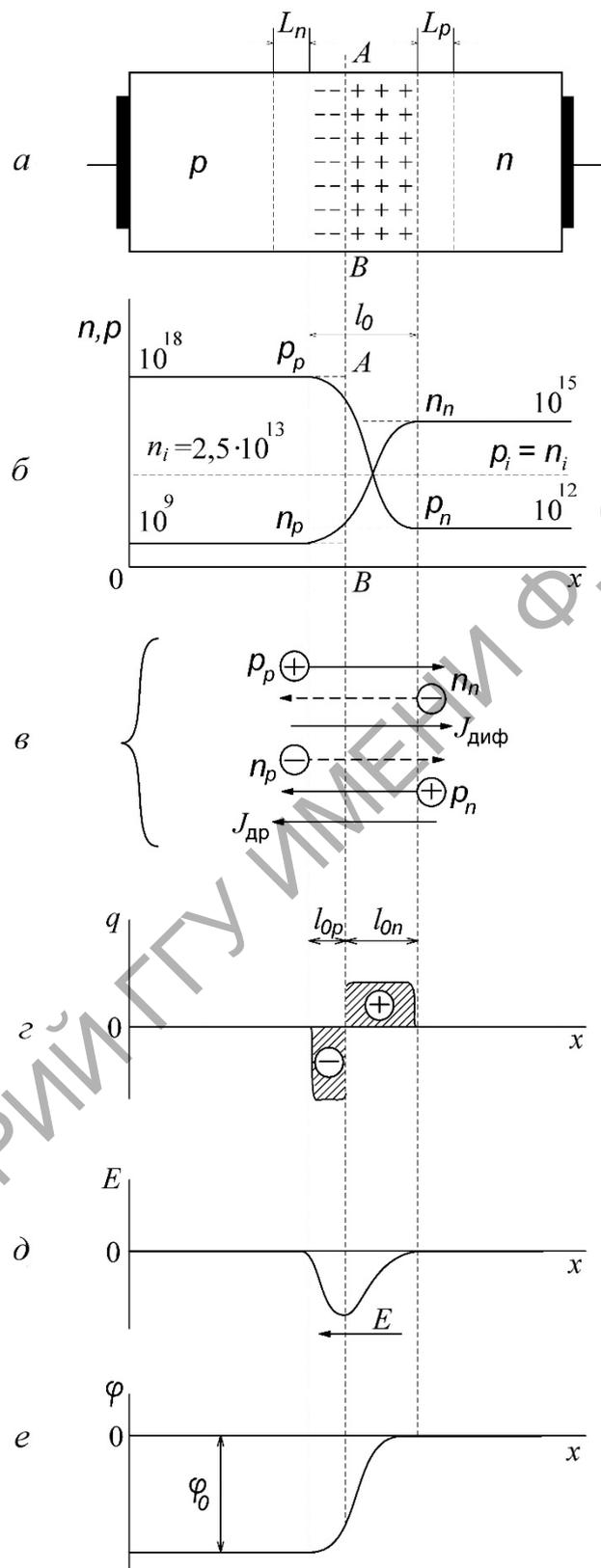


Рис. 8.4. Образование p - n -перехода в p - n -структуре полупроводника: a – схема p - n -перехода; $б$ – распределение концентраций носителей заряда; $в$ – направление диффузионного и дрейфового токов; $г$ – объемные заряды; $д$ – внутреннее электрическое поле; $е$ – потенциальный барьер

В германиевых и кремниевых диодах двуслойная p - n -структура создается введением в один из слоев монокристалла акцепторной примеси, а в другой – донорной примеси. При комнатной температуре атомы акцепторов и доноров можно считать полностью ионизированными, т. е. практически все акцепторные атомы присоединяют к себе электроны, создавая при этом дырки, а донорные атомы отдают свои электроны, которые становятся свободными. Кроме основных носителей заряда в каждом из слоев имеются неосновные носители заряда, создаваемые путем перехода электрона основного материала из валентной зоны в зону свободных уровней.

На практике наибольшее распространение получили p - n -структуры с неодинаковой концентрацией внесенных акцепторной N_a и донорной N_d примесей, т. е. с неодинаковой концентрацией основных носителей заряда в слоях $p_p \approx N_a$ и $n_n \approx N_d$. Типичными являются структуры с $N_a \ll N_d$ ($p_p \gg n_n$). Распределение концентраций носителей заряда для таких структур показано на рис. 8.4, б на примере германия, где приняты $p_p = 10^{18} \text{ см}^{-3}$, $n_n = 10^{15} \text{ см}^{-3}$. Концентрация свободных носителей заряда в германии при комнатной температуре $n_i = 2,5 \cdot 10^{13} \text{ см}^{-3}$. Концентрации неосновных носителей заряда существенно меньше концентраций основных носителей заряда, составляют для рассматриваемой структуры в соответствии с (2.2) $n_p = 10^9 \text{ см}^{-3}$.

Характерной особенностью полупроводников рассматриваемых типов является то, что произведение концентраций основных и неосновных носителей заряда при данной температуре является постоянной величиной и определяется соотношением

$$p_p n_p = n_n p_n = p_i n_i = A^2 e^{-\frac{E_g}{kT}}, \quad (8.2)$$

где $n_i = p_i$ – собственные концентрации носителей заряда в чистом полупроводнике.

В соответствии с выражением (8.2) концентрация неосновных носителей заряда в примесном полупроводнике меньше концентрации собственных носителей заряда в чистом полупроводнике. Это связано с тем, что с увеличением концентрации основных носителей заряда возрастает роль рекомбинаций, вследствие чего концентрация неосновных носителей заряда уменьшается. Равновесие достигается, когда при данной температуре произведение концентрации носителей заряда в примесном полупроводнике становится равным произведению концентрации носителей заряда в чистом полупроводнике.

В p - n -структуре на границе раздела слоев AB возникает разность концентраций одноименных носителей заряда: в одном слое они являются

основными, а в другом – неосновными. В приграничной области под действием разности концентраций возникает диффузионное движение основных носителей заряда во встречном направлении через границу раздела. Дырки из p -области диффундируют в n -область, электроны из n -области – в p -область. Дырки, вошедшие в n -область, рекомбинируют с электронами этой области, а электроны, вошедшие в p -область, – с дырками p -области. Вследствие двух факторов (ухода основных носителей заряда из приграничных областей и их рекомбинации с носителями заряда противоположного знака) концентрации основных носителей заряда (p_p и n_n) в обеих приграничных областях, суммарная ширина которых l_0 , снижаются (рис. 8.4, б). Кроме того, в соответствии с выражением (8.2) снижение концентрации носителей заряда одного знака сопровождается повышением концентрации носителей заряда другого знака. Вследствие этого в приграничной p -области повышается концентрация электронов, а в приграничной n -области – концентрация дырок. Таким образом, становится понятным характер распределения концентрации носителей заряда в p - n -переходе, показанной на рис. 8.2, б сплошными линиями.

Важнейшим следствием диффузионного движения носителей заряда через границу раздела слоев является появление в приграничных областях *объемных зарядов*, создаваемых ионами атомов примесей. Так, при уходе дырок из p -слоя в нем создается нескомпенсированный отрицательный объемный заряд за счет оставшихся отрицательных ионов акцепторных атомов примеси. Электроны же, ушедшие из n -слоя, оставляют здесь нескомпенсированный положительный объемный заряд, создаваемый положительными ионами донорных атомов примеси. Наличие объемного заряда является главной особенностью p - n -перехода. Кривая распределения объемного заряда в p - n -переходе показана на рис. 8.4, г.

Ввиду наличия объемного заряда в p - n -переходе создаются электрическое поле и разность потенциалов. Кривые $E(x)$ и $\varphi(x)$ показаны на рис. 8.2, д, е (за нулевой принят потенциал n -слоя). Отметим, что рассмотренный процесс формирования p - n -перехода происходит уже на этапе введения в монокристалл акцепторной и донорной примесей.

Толщина слоя объемного заряда l_0 составляет доли микрометров и зависит от концентрации примеси (основных носителей заряда) в p - и n -областях (от удельного сопротивления слоев). Объемные заряды по обе стороны границы раздела равны и создаются, как известно, неподвижными ионами примеси. Если бы концентрации акцепторной N_a и донорной N_d примесей были равны (симметричный p - n -переход), то концентрации отрицательных ионов слева от границы раздела и положительных ионов справа были бы также равны и p - n -переход имел бы одинаковые толщины слоев l_{0p} и l_{0n} .

В рассматриваемом случае несимметричного p - n -перехода ($N_a > N_d$) концентрация неподвижных отрицательных ионов слева от границы раздела AB будет выше концентрации неподвижных положительных ионов справа (рис. 8.4, a), в связи с чем равенству объемных зарядов обоих знаков (рис. 8.4, z) здесь будет отвечать условие $l_{0n} \gg l_{0p}$. Иными словами, p - n -переход толщиной l_0 будет преимущественно сосредоточен в n -области, как в более высокоомной.

Внутреннее электрическое поле, созданное объемными зарядами, является фактором, под действием которого обеспечивается равенство потоков носителей заряда через переход в обоих направлениях, т. е. равенство нулю суммарного тока в отсутствие внешнего электрического поля. Это обусловлено тем, что внутреннее электрическое поле с потенциальным барьером ϕ_0 (рис. 8.4, e) создает тормозящее действие для основных и ускоряющее – для неосновных носителей заряда. Таким образом, внутреннее электрическое поле приводит к уменьшению плотности диффузионного тока $J_{\text{диф}}$ через переход и появлению встречного ему дрейфового тока плотностью $J_{\text{др}}$.

Диффузионный ток плотностью $J_{\text{диф}}$, обусловленный основными носителями заряда (рис. 8.4, b), направлен вдоль оси x и состоит из потока дырок, перемещающихся под действием диффузии из p -области в n -область, и потока электронов, диффундирующих из n -области в p -область.

Плотность дрейфового тока $J_{\text{др}}$ (рис. 8.4, b) создается неосновными носителями заряда прилегающих к p - n -переходу слоев с толщиной, равной диффузионной длине: L_n – для электронов p -слоя и L_p – для дырок n -слоя (рис. 8.4, a). Неосновные носители заряда, совершая тепловое движение в этих слоях, успевают за время своей жизни попасть в область действия электрического поля, увлекаются этим полем и перебрасываются через переход. Таким образом, плотность дрейфового тока определяется потоками подходящих неосновных носителей из прилегающих к p - n -переходу слоев. Она зависит от концентрации неосновных носителей заряда в слоях и диффузионной длины. Дрейфовый ток имеет направление, противоположное направлению диффузионного тока.

Равенство нулю тока через переход в отсутствие внешнего напряжения соответствует уменьшению диффузионной составляющей тока до величины его дрейфовой составляющей. Равенство составляющих тока $J_{\text{диф}} = J_{\text{др}}$ создается установлением соответствующей величины потенциального барьера ϕ_0 в p - n -переходе. Величина потенциального барьера ϕ_0 (называемого также *контактной разностью потенциалов*) зависит от соотношения концентраций носителей заряда одного знака по обе стороны перехода и определяется соотношением

$$\varphi_0 = \varphi_T \ln \frac{p_p}{p_n} = \varphi_T \ln \frac{n_n}{n_p}, \quad (8.3)$$

где $\varphi_T = \frac{kT}{e}$ – тепловой потенциал (при $T = 300$ К $\varphi_T \approx 0,026$ В).

Высота потенциального барьера зависит от температуры ввиду зависимости от нее теплового потенциала и концентрации неосновных носителей заряда в слоях полупроводниковой структуры. Более сильное влияние температуры на концентрацию неосновных носителей заряда, чем влияние на величину φ_T , приводит к тому, что с ростом температуры высота потенциального барьера уменьшается. При комнатной температуре для германия $\varphi_0 = 0,3-0,5$ В, а для кремния $\varphi_0 = 0,6-0,8$ В. Различие в значениях φ_0 объясняется большей величиной E_g в кремнии ($E_g = 1,12$ эВ), чем в германии ($E_g = 0,72$ эВ), и, следовательно, меньшей концентрацией неосновных носителей заряда (при одинаковой температуре и одинаковых концентрациях внесенных примесей).

Уход неосновных носителей заряда через p - n -переход из прилегающих к нему слоев, казалось бы, должен привести к уменьшению их концентрации с приближением к границе p - n -перехода. Вместе с тем концентрации неосновных носителей заряда в прилегающих к p - n -переходу слоях сохраняются на уровнях p_n и n_p (рис. 8.4, б), так как в условиях динамического равновесия уменьшение концентрации неосновных носителей заряда за счет их ухода через p - n -переход будет постоянно восполняться носителями того же знака за счет их диффузии из противоположных слоев.

8.3. Электрические процессы в p - n -переходе при наличии внешнего напряжения

Подключение к p - n -структуре внешнего напряжения (напряжения смещения) приводит к изменению условий переноса заряда через p - n -переход. Существенную роль при этом играет полярность внешнего напряжения, с которой оно прикладывается к p - n -переходу.

Рассмотрим случай, когда внешнее напряжение подключено к p - n -структуре **в прямом направлении**, т. е. плюсом источника к выводу p -области, а минусом источника – к выводу n -области (рис. 8.5, а). При таком подключении источника создаваемое им электрическое поле направлено противоположно внутреннему полю в переходе, что приводит к уменьшению результирующего поля в p - n -переходе. Слой, осуществляющий эмиссию дырок через p - n -переход (p -слой), называется **эмиттером**. Поскольку основой

при получении p - n -структуры диода обычно служит полупроводниковый материал n -типа, n -слой называют *базой*.

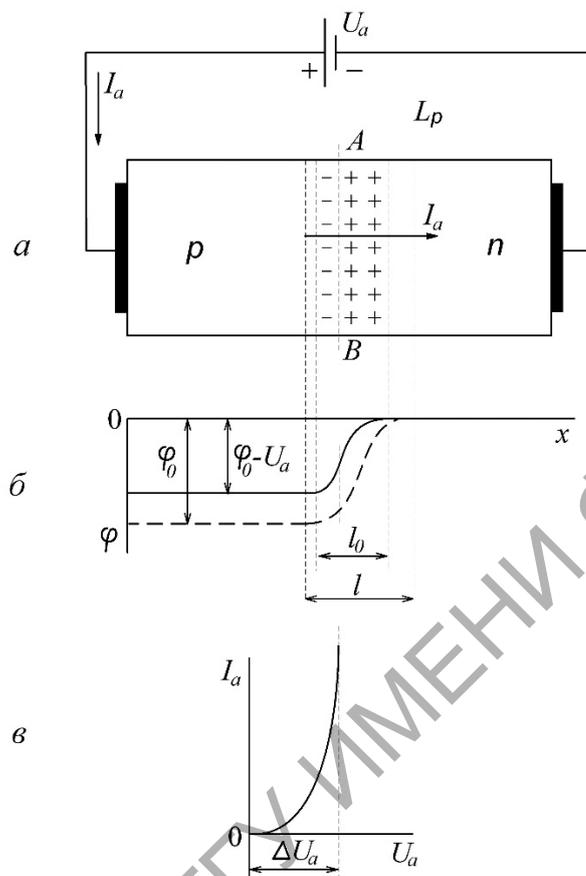


Рис. 8.5. Полупроводниковый диод при подключении внешнего напряжения в прямом направлении: a – схема включения; b – потенциальный барьер при прямом напряжении; c – прямая ветвь вольт-амперной характеристики

Объемный заряд обоих знаков, сосредоточенный в переходе по разные стороны границы раздела, будет определяться не только величиной φ_0 , обусловленной, как было показано выше, диффузионным движением носителей заряда под действием разности их концентраций в приграничных слоях, но и внешним напряжением U_a . Если пренебречь падением напряжения в слоях p - и n -структуры, то объемному заряду в переходе будет отвечать напряжение $\varphi_0 - U_a$, меньшее, чем в отсутствие внешнего источника. Следовательно, уменьшится и обусловленный напряжением объемный заряд в p - n -переходе.

Величина $\varphi_0 - U_a$ определяет высоту потенциального барьера в p - n -переходе при включении внешнего напряжения в прямом направлении (рис. 8.5, b). Уменьшение объемного заряда (потенциального барьера) проявляется в сужении p - n -перехода, которое происходит в основном за счет n -слоя как более высокоомного.

Уменьшение потенциального барьера облегчает переход основных носителей заряда под действием диффузии через границу раздела в соседние области, что приводит к увеличению диффузионного тока через p - n -переход (рис. 8.5, в). Указанное явление называют *инжекцией носителей заряда* через p - n -переход.

Вместе с тем дрейфовый ток через p - n -переход, создаваемый потоками неосновных носителей заряда, подходящих из приграничных слоев толщиной L к p - n -переходу, остается без изменения. Разность диффузионного и дрейфового токов определяет результирующий прямой ток через p - n -переход (прямой ток диода). Плотность прямого тока

$$J_a = J_{\text{диф}} - J_{\text{др}}. \quad (8.4)$$

С повышением приложенного внешнего напряжения диффузионный ток увеличивается (так как уменьшившийся потенциальный барьер способен преодолеть основные носители заряда, обладающие меньшей энергией), в связи с чем возрастает прямой ток через p - n -переход. Примерный вид прямой ветви вольт-амперной характеристики p - n -перехода показан на рис. 8.5, в (ток I_a на рис. 8.5, в равен произведению плотности тока J_a через p - n -переход на площадь его сечения S).

В кремниевых диодах величина ϕ_0 выше, чем в германиевых. Одинаковая величина внешнего напряжения U_a здесь создает меньшее относительное снижение потенциального барьера, чем в германиевых диодах, и обуславливает меньший прямой ток при одинаковой площади p - n -перехода. Большая величина ϕ_0 является одной из причин большего падения напряжения ΔU_a в кремниевых переходах (0,8–1,2 В) по сравнению с германиевыми переходами (0,3–0,6 В) при протекании тока в прямом направлении. Таким образом, падение напряжения ΔU_a не превышает 1,2 В, что выгодно отличает их от переходов других типов, в частности электровакуумных и газоразрядных (ионных).

При подключении к диоду источника внешнего напряжения **в обратном направлении** (рис. 8.6, а) потенциальный барьер возрастает на величину U_b и становится равным $(\phi_0 + U_b)$ (рис. 8.6, б).

При этом увеличивается объемный заряд в p - n -переходе и его ширина. Возросший потенциальный барьер затрудняет прохождение через p - n -переход основных носителей заряда, вследствие чего диффузионный ток, создаваемый этими носителями, уменьшается. Дрейфовый же ток, обусловленный концентрациями неосновных носителей заряда по обе стороны перехода ($J_{\text{др}} = J_{\text{др}p} + J_{\text{др}n}$), можно считать неизменным. Однако теперь он будет превышать диффузионный ток. Через диод будет протекать ток в обратном направлении:

$$J_b = J_{\text{др}} - J_{\text{диф}}. \quad (8.5)$$

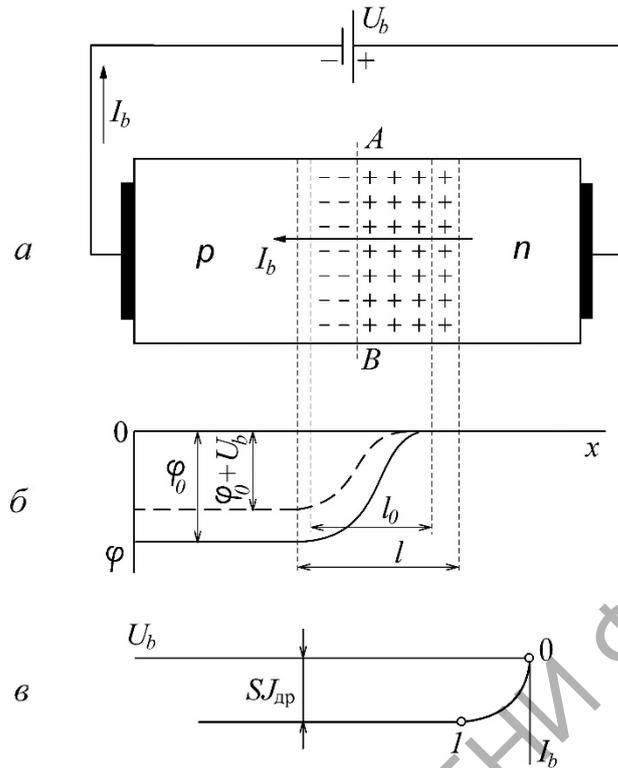


Рис. 8.6. Полупроводниковый диод при подключении внешнего напряжения в обратном направлении: *a* – схема включения; *б* – потенциальный барьер при обратном напряжении; *в* – обратная ветвь вольт-амперной характеристики

Обратная ветвь вольт-амперной характеристики перехода показана на рис. 8.6, *в*. При небольших обратных напряжениях (участок 0–1) увеличение обратного тока наблюдается за счет уменьшения диффузионной составляющей. При обратном напряжении, соответствующем точке 1 и большем, основные носители заряда не способны преодолеть потенциальный барьер, в связи с чем диффузионный ток равен нулю. Этим объясняется отсутствие роста обратного тока при увеличении обратного напряжения (участок характеристики левее точки 1).

Полная вольт-амперная характеристика полупроводникового перехода приведена на рис. 8.7. От характеристики идеального диода она отличается наличием некоторого падения напряжения на приборе при пропускании прямого тока и обратного тока в случае приложения обратного напряжения.

Как известно, прямой ток перехода создается основными, а обратный – неосновными носителями заряда. Концентрация основных носителей заряда на несколько порядков

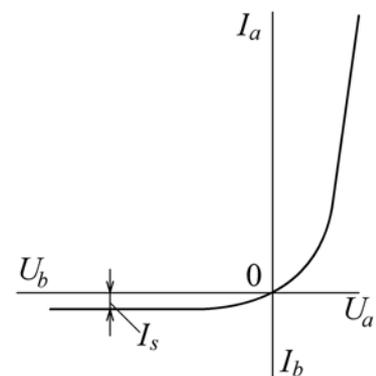


Рис. 8.7. Идеализированная вольт-амперная характеристика диода

превышает концентрацию неосновных носителей. Этим и обуславливаются вентильные свойства p - n -перехода.

Проведенному техническому анализу вольт-амперной характеристики перехода соответствует ее запись в аналитической форме:

$$I_a = I_s (e^{U/\varphi_T} - 1), \quad (8.6)$$

где $I_s = S J_{др}$ – ток насыщения (тепловой ток), создаваемый неосновными носителями заряда: φ_T – тепловой потенциал.

При $U = 0$ согласно соотношению (8.6) $I_a = 0$. В случае приложения прямого напряжения ($U = U_a > 0$) в (8.6) единицей можно пренебречь и зависимость $I_a(U_a)$ будет иметь экспоненциальный характер. В случае обратного напряжения ($U = U_b < 0$) можно не учитывать достаточно малую величину e^{-U/φ_T} и тогда $I_a = I_b = -I_s$.

Рассмотрим **дополнительные факторы**, влияющие на вольт-амперную характеристику перехода. В проведенном анализе, позволяющем главным образом объяснить принцип действия p - n -перехода, не учитывались некоторые факторы, отражающиеся на его реальной вольт-амперной характеристике.

На прямую ветвь вольт-амперной характеристики перехода оказывает влияние объемное сопротивление слоев p - n -структуры (особенно при больших токах), увеличивающее падение напряжения ΔU_a на диоде. В кремниевых диодах это влияние более значительно, чем в германиевых, так как из-за меньшей подвижности носителей заряда удельное сопротивление кремния выше. С учетом падения напряжения в слоях в кремниевых диодах при протекании прямого тока $\Delta U_a = 0,8-1,2$ В, а в германиевых $\Delta U_a = 0,3-0,6$ В.

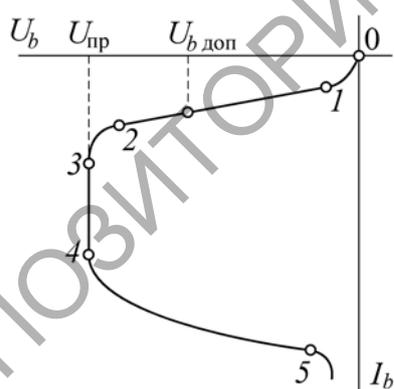


Рис. 8.8. Обратная ветвь вольт-амперной характеристики реальных переходов

На обратную ветвь вольт-амперной характеристики перехода оказывают влияние ток утечки через поверхность p - n -перехода и генерация носителей заряда, которая является причиной возможного пробоя p - n -перехода. Оба фактора приводят к тому, что обратная ветвь вольт-амперной характеристики диода принимает вид, показанный на рис. 8.8.

Ток утечки связан линейной зависимостью с напряжением U_b . Он создается различными загрязнениями на внешней поверхности p - n -структуры, что повышает поверхностную электрическую проводимость p - n -перехода и обратный ток через переход. Эта составляющая обратного тока обуславливает появление наклонного участка 1–2 на характеристике диода (рис. 8.8).

Влияние генерации носителей заряда в p - n -переходе обычно сказывается при повышенных обратных напряжениях. Оно проявляется в нарушении линейной зависимости изменения обратного тока от напряжения U_b (участок 2–3), а затем в резком возрастании обратного тока (участок 3–4–5), характеризующем пробой p - n -перехода.

В зависимости от причин, вызывающих появление дополнительных носителей заряда в p - n -переходе, различают *электрический* пробой и *тепловой* пробой. Электрический пробой в свою очередь может быть лавинным или туннельным. Рассмотрим эти виды пробоя.

Лавинный пробой обусловлен лавинным размножением носителей в p - n -переходе в результате ударной ионизации атомов быстрыми носителями заряда (см. п. 7.2). Он происходит следующим образом. Неосновные носители заряда, поступающие в p - n -переход при действии обратного напряжения, ускоряются полем и при движении в нем сталкиваются с атомами кристаллической решетки. При соответствующей напряженности электрического поля носители заряда приобретают энергию, достаточную для отрыва валентных электронов. При этом образуются дополнительные пары носителей заряда – электроны и дырки, которые, ускоряясь полем, при столкновении с атомами также создают дополнительные носители заряда. Описанный процесс носит лавинный характер.

Лавинный пробой возникает в широких p - n -переходах, где при движении под действием электрического поля носители заряда, встречаясь с большим количеством атомов кристалла, в промежутке между столкновениями приобретают достаточную энергию для их ионизации.

В основе *туннельного пробоя* лежит непосредственный отрыв валентных электронов от атомов кристаллической решетки под действием сильного электрического поля. Образующиеся при этом дополнительные носители заряда (электроны и дырки) увеличивают обратный ток через p - n -переход. Туннельный пробой развивается в достаточно узких p - n -переходах, где при сравнительно небольшом обратном напряжении имеется высокая напряженность поля.

Лавинный и туннельный пробой сопровождаются появлением почти вертикального участка 3–4 на обратной ветви вольт-амперной характеристики (рис. 8.8). Причина заключается в том, что небольшое повышение напряжения на p - n -переходе вызывает более интенсивную генерацию в нем носителей заряда при лавинном или туннельном пробое.

Оба эти вида пробоя являются обратимыми процессами. Это означает, что они не приводят к повреждению перехода и при снижении напряжения его свойства сохраняются.

Тепловой пробой возникает за счет интенсивной термогенерации носителей в p - n -переходе при недопустимом повышении температуры. Процесс развивается лавинообразно и ввиду неоднородности p - n -перехода обычно носит локальный характер. Лавинообразное развитие теплового пробоя обуславливается тем, что увеличение числа носителей заряда за счет повышения температуры вызывает увеличение обратного тока и, следовательно, еще больший разогрев участка p - n -перехода. Процесс заканчивается расплавлением этого участка и выходом прибора из строя.

Тепловой пробой может произойти в результате перегрева отдельного участка p - n -перехода вследствие протекания большого обратного тока при лавинном или туннельном пробое (участок 4–5 на рис. 8.8). Тепловой пробой здесь является следствием недопустимого повышения обратного напряжения (перенапряжения). Велика вероятность наступления теплового пробоя при общем перегреве p - n -перехода ввиду, например, ухудшения условий теплоотвода. В этом случае он может произойти при меньших напряжениях U_b , минуя стадии лавинного или туннельного пробоя.

Возможность теплового пробоя p - n -перехода учитывается указанием в паспорте на прибор допустимого обратного напряжения $U_{b \text{ доп}}$ и температурного диапазона работы. Величина допустимого обратного напряжения устанавливается с учетом исключения возможности электрического пробоя и составляет $(0,5–0,8) U_{пр}$.

8.4. Емкости p - n -перехода

Емкость p - n -перехода равна сумме так называемых барьерной и диффузионной емкостей.

Барьерная (или *зарядная*) *емкость* характеризуется сосредоточением по обе стороны границы раздела p - и n -слоев объемных зарядов, создаваемых ионами примесей. Физическим аналогом барьерной емкости приближенно может служить емкость плоского конденсатора.

Наличие барьерной емкости проявляется протеканием тока через p - n -переход вследствие изменения объемных зарядов (а следовательно, ширины p - n -перехода) при изменении напряжения на переходе и определяется соотношением $C_{\delta} = dQ/dU$. Зарядная емкость возрастает с уменьшением толщины p - n -перехода, т. е. при снижении обратного напряжения. Она выше при прямых напряжениях, чем при обратных. Величина барьерной емкости зависит от площади p - n -перехода и может составлять десятки и сотни пикофарад. Зависимость барьерной емкости p - n -перехода от обратного напряжения используется в варикапах (параметрических диодах), применяемых в качестве конденсаторов переменной емкости, управляемых напряжением.

В отличие от барьерной емкости, определяемой шириной области объемного заряда p - n -перехода, *диффузионная емкость* обусловлена изменением суммарных зарядов неравновесных электронов и дырок соответственно слева и справа от p - n -перехода в результате изменения напряжения на нем. Так как эти заряды создаются за счет диффузии (инжекции) носителей через p - n -переход, диффузионную емкость следует учитывать при прямом напряжении смещения. В несимметричных p - n -переходах, для которых $p_p \gg n_n$, диффузионная емкость определяется преимущественно суммарным зарядом неравновесных дырок в n -слое, величина которого изменяется при изменении прямого напряжения.

Величина диффузионной емкости зависит от протекающего через p - n -переход прямого тока и может составлять сотни и тысячи пикофарад, т. е. она существенно больше барьерной емкости. Таким образом, при малых напряжениях смещения емкость p - n -перехода определяется в основном диффузионной емкостью, а при обратных напряжениях, когда диффузионная емкость равна нулю, – барьерной емкостью.

Вопросы

1. При каких условиях в полупроводниках возникает электронно-дырочный переход?
2. Чем отличается гетеропереход от p - n -перехода двух полупроводников?
3. Какими видами носителей заряда обусловлены диффузионный и дрейфовый токи в p - n -переходе?
4. Какое явление называют инжекцией носителей заряда через p - n -переход.
5. Какие факторы оказывают влияние на вид обратной ветви вольт-амперной характеристики реального p - n -перехода?
6. Какие виды электрического пробоя p - n -перехода существуют и чем они обусловлены?
7. В чем отличие барьерной и диффузионной емкости p - n -перехода?

ЗАКЛЮЧЕНИЕ

Как было отмечено во введении, вступление в третье тысячелетие электроника отметила зарождением нового направления – наноэлектроники. В настоящее время нанотехнология в электронике считается одним из наиболее перспективных направлений в совершенствовании свойств материалов и создании материалов с заданным комплексом свойств.

Нанотехнология имеет дело с процессами, протекающими на так называемом наноуровне. При этом в зависимости от позиции исследователя нанотехнология может рассматривать в качестве своего объекта как сами нанобъекты (т. е. материалы, состоящие большей частью из низкоразмерных систем с характерными размерами примерно от 100 до 1 нм), так и материалы на их основе. Чтобы подчеркнуть тот факт, что тот или иной материал обладает определенным комплексом свойств именно благодаря наличию в нем нанобъектов, такие материалы зачастую называют *нано-структурированными*, или *нанокомпозитами*. Собственно говоря, в ряде случаев это вполне оправдано, поскольку основные физико-химические характеристики целого ряда материалов действительно определяются свойствами содержащихся в них нанобъектов. Изменяя размеры, форму или взаимное расположение низкоразмерных систем как компонентов наноматериала, можно управлять его свойствами (например, температурой плавления, растворимостью, прозрачностью) без изменения химического состава.

Нанокомпозиты, содержащие металлы или полупроводники, привлекают внимание, прежде всего, уникальными свойствами входящих в их состав кластеров, образованных разным количеством атомов металла или полупроводника – от десяти до нескольких тысяч. Типичные размеры такого агрегата – от 1 до 10 нм, что обуславливает огромную удельную поверхность. Подобные наночастицы отличаются по многим физическим свойствам (ширина полосы поглощения, спектральные характеристики, плотность электронных состояний) как от блочного материала, так и от индивидуального атома или молекулы.

Особенно сильно по свойствам отличаются полупроводниковые наночастицы, даже если их размер достигает сотен нанометров. Энергетическая зона металла, независимо от размера его частиц, заполнена не вся, поэтому электроны могут переходить на более высокие уровни. У полупроводника же валентная зона заполнена целиком и отделена от зоны проводимости на 2–3 эВ. Из-за малых размеров полупроводниковых нанокристаллов эти зоны расщепляются, что приводит к эффективному увеличению ширины запрещенной зоны. Так, при переходе от макрокристалла CdS к нанокристаллу ширина запрещенной зоны увеличивается с 2,5 до 4,5 эВ, время жизни

на нижнем возбужденном уровне уменьшается с нескольких наносекунд до пикосекунд, температура плавления снижается с 1600 до 400 °С. Нелинейные оптические свойства нанокластеров позволяют создавать на их основе управляемые квантовые светодиоды для применения в микроэлектронике и телекоммуникации.

Особое место в разработке нанокompозитных материалов занимает *графен* (от англ. *graphene*) – двумерная аллотропная модификация углерода, образованная слоем атомов углерода толщиной в один атом, находящихся в sp^2 -гибридизации и соединенных посредством σ - и π -связей в гексагональную двумерную кристаллическую решетку. Его можно представить как одну плоскость графита, отделенную от объемного кристалла. Графен относится к бесщелевым полупроводникам ($E_g = 0$). Толщина графена принимается равной 0,34 нм, т. е. расстоянию между соседними атомными плоскостями в трехмерном кристаллическом графите. Графен площадью 1 см² содержит $3,8 \cdot 10^{15}$ атомов углерода. Каждый атом углерода имеет четыре валентных электрона: три из них локализованы (участвуют в образовании ковалентных С–С-связей и определяют механические свойства графена), а один делокализован (обуславливает оптическое поглощение и электрическую проводимость). Пропускание света оптического диапазона графеном не зависит от длины волны и составляет примерно 97,7 %. Присоединение к каждому атому углерода одного атома водорода превращает графен в *графан* – кристаллический диэлектрик в виде квазиплоского слоя. При термической обработке графан восстанавливается до исходного полупроводникового графена. Нанокompозит, содержащий графен и олово, способен заметно увеличить емкость литий-ионных аккумуляторов и уменьшить их вес.

Для исследования, создания и применения низкоразмерных систем необходимо развивать их теорию. Обусловлено это тем, что основой для описания электрических, магнитных, оптических, акустических и тепловых явлений в трехмерных (объемных) кристаллических диэлектриках, полупроводниках и металлах с размерами, большими или примерно равными 0,1 мкм, является зонная теория. Однако для описания низкоразмерных систем она неприменима из-за нарушений точечной и трансляционной симметрий, а также условий применимости адиабатического и одноэлектронного приближений. Не может быть использована для адекватного описания состояний и процессов в системах размером порядка 0,1 мкм и квантовая теория одиночных атомов, являющаяся базой для описания небольших молекул (0,1–1 нм). Поэтому необходимо развить теорию, рассматривающую состояния и процессы в отдельных молекулах и атомах при учете их агломерации в низкоразмерную систему [Dresselhaus, M. S. What's next for low-dimensional materials? / M. S. Dresselhaus // Mater. Res. Lett. – 2014. – Vol. 2, № 1. – P. 1–9].

По мнению профессора Н. А. Поклонского совершенствование методов создания и диагностики, а также расширение областей применения полупроводниковых систем нового поколения в электронике, фотонике, спинтронике и акустике будет происходить по следующим основным перспективным направлениям исследований [Поклонский, Н. А. Физика и техника низкоразмерных систем из углерода / Н. А. Поклонский // Наноструктурные материалы: Беларусь – Россия – Украина. НАНО-2014: сб. пленар. докл. IV Междунар. науч. конф., Минск, 7–10 окт. 2014 г. / редкол.: П. А. Витязь [и др.]. – Минск, 2015. – С. 121–136]:

- разработка методов математического моделирования и физико-химических принципов молекулярного зодчества низкоразмерных систем и приборных структур на их основе; в итоге это позволит целенаправленно и экономно реализовать «малотоннажные» технологии;

- развитие квантовой теории ионизационного равновесия и миграции электронов, дырок и ионов в низкоразмерных полупроводниковых системах для целей водородной и солнечной энергетики; решение этой задачи позволит создать новые материалы для фотоэлектрических преобразователей и «неотравляемые» электроды для фотолиза воды;

- разработка физико-технологических способов формирования ассоциатов из атомов примеси или собственных атомных дефектов структуры в кристаллических полупроводниках при создании твердотельных аналогов катушек индуктивности для силовой электроники;

- интегрирование магнетизма в полупроводниковую микро- и наноэлектронику; развитие инженерии магнитных низкоразмерных систем в кремниевых пластинах обеспечит возможность распространить их использование в рамках планарной технологии на спинтронику;

- установление зависимости механической прочности широкозонных полупроводников (например, алмаза и нитрида бора, содержащих точечные, линейные и плоскостные дефекты кристаллической структуры) от положения уровня Ферми в запрещенной зоне; это позволит предсказывать (и предотвращать) процессы разрушения этих материалов в устройствах высокотемпературной электроники, оптики и механики;

- изучение прыжковой миграции электронов по многозарядным точечным дефектам кристаллической матрицы частично разупорядоченных полупроводников для разработки выпрямителя прыжкового электрического тока и элемента Пельтье.

Таким образом, перспективы развития физики и техники полупроводников во многом обусловлены исследованием, разработкой и внедрением в производство новых многофункциональных нанокompозитных материалов.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная литература

1. *Бонч-Бруевич, В. Л.* Физика полупроводников / В. Л. Бонч-Бруевич, С. Г. Калашников. – М.: Наука, 1977. – 672 с.
2. *Гаркуша, Ж. М.* Основы физики полупроводников / Ж. М. Гаркуша. – М.: Высш. шк., 1981. – 295 с.
3. *Грундман, М.* Основы физики полупроводников. Нанопизика и технические приложения / М. Грундман. – М.: Физматлит, 2012. – 772 с.
4. *Лебедев, А. И.* Физика полупроводниковых приборов. / А. И. Лебедев. – М.: Физматлит, 2012. – 488 с.
5. *Ощепков, А. Ю.* Физика полупроводников: учебник / А. Ю. Ощепков. – 4-е изд., стер. – СПб.: Лань П, 2016. – 400 с.
6. *Смит, Р. М.* Полупроводники / Р. М. Смит. – М.: Мир, 1982. – 560 с.
7. *Старосельский, В. И.* Физика полупроводниковых приборов микроэлектроники: учеб. для академического бакалавриата / В. И. Старосельский. – Люберцы: Юрайт, 2016. – 463 с.
8. *Фистуль, В. И.* Введение в физику полупроводников / В. И. Фистуль. – М.: Высш. шк., 1984. – 352 с.
9. *Шалимова, К. В.* Физика полупроводников / К. В. Шалимова. – М.: Энергоатомиздат, 1985. – 392 с.
10. *Шалимова, К. В.* Физика полупроводников: учебник / К. В. Шалимова. – 4-е изд., стер. – СПб.: Лань, 2010. – 400 с.
11. *Ю, Питер.* Основы физики полупроводников / Питер Ю, М. Кардона. – М.: Физматлит, 2002.

Дополнительная литература

1. *Анималу, А. М.* Квантовая теория кристаллических твердых тел / А. М. Анималу. – М.: Мир, 1981. – 576 с.
2. *Гольдаде, В. А.* Физика конденсированного состояния / В. А. Гольдаде, Л. С. Пинчук. – Минск: Беларус. навука, 2009. – 657 с.
3. *Елифанов, Г. И.* Физика твердого тела / Г. И. Елифанов. – М.: Высш. шк., 1977. – 276 с.
4. *Займан, Дж.* Принципы теории твердого тела / Дж. Займан. – М.: Мир, 1974. – 379 с.

5. *Зи, С.* Физика полупроводниковых приборов / С. Зи. – М.: Мир, 1984. – 456 с.

6. *Поклонский, Н. А.* Физика и техника низкоразмерных систем из углерода / Н. А. Поклонский // Наноструктурные материалы: Беларусь – Россия – Украина. НАНО-2014: сб. пленар. докл. IV Междунар. науч. конф., Минск, 7–10 окт. 2014 г. / редкол.: П. А. Витязь [и др.]. – Минск, 2015. – С. 121–136

7. *Ридли, Б.* Квантовые процессы в полупроводниках / Б. Ридли. – М.: Мир, 1986. – 304 с.

8. *Тилл, У.* Интегральные схемы. Материалы, приборы, изготовление / У. Тилл, Дж. Лаксон. – М.: Мир, 1985. – 501 с.

ОГЛАВЛЕНИЕ

СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	3
ВВЕДЕНИЕ	4
Глава 1. МОДЕЛЬНЫЕ ПРЕДСТАВЛЕНИЯ О СТРУКТУРЕ ПОЛУПРОВОДНИКОВ.....	8
1.1. Типы химических связей.....	8
1.2. Кристаллическая решетка	16
1.3. Кристаллографические обозначения.....	20
1.4. Обратная решетка.....	23
1.5. Кристаллические структуры твердых тел	24
1.6. Дефекты в кристаллах	27
Вопросы.....	34
Глава 2. ОСНОВЫ ЗОННОЙ ТЕОРИИ ТВЕРДЫХ ТЕЛ.....	35
2.1. Квантово-механические предпосылки зонной теории.....	36
2.1.1. Корпускулярно-волновой дуализм.....	36
2.1.2. Волновые функции и уравнение Шрёдингера	38
2.1.3. Волна (функция) Блоха.....	43
2.2. Зоны Бриллюэна	45
2.3. Поверхность и энергия Ферми.....	48
2.4. Энергетический спектр электронов в кристалле	50
2.4.1. Одномерная модель электронных зон	50
2.4.2. Заполнение зон электронами в идеальном кристалле.....	52
2.4.3. Эффективная масса электрона.....	56
2.4.4. Примесные уровни и поверхностные состояния	59
2.5. Энергетический спектр некристаллических твердых тел	62
2.5.1. Плотность состояний	62
2.5.2. Границы применимости зонной теории	65
Вопросы.....	67
Глава 3. ОСНОВНЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ.....	69
3.1. Зонная структура полупроводников	69
3.2. Дрейфовая и диффузионная электропроводность	70
3.3. Собственная и примесная проводимость	74

3.4. Основные электрофизические свойства полупроводников	80
3.5. Получение и легирование полупроводников.....	83
Вопросы	88

Глава 4. РАССЕЯНИЕ НОСИТЕЛЕЙ ЗАРЯДА

В ПОЛУПРОВОДНИКАХ	90
4.1. Механизмы рассеяния электронов и дырок.....	91
4.2. Рассеяние на ионах примеси	93
4.3. Рассеяние на атомах примеси и дислокациях.....	95
4.4. Рассеяние на тепловых колебаниях решетки (фононах)	96
Вопросы	101

Глава 5. ГЕНЕРАЦИЯ И РЕКОМБИНАЦИЯ

ЭЛЕКТРОНОВ И ДЫРОК	102
5.1. Равновесные и неравновесные носители заряда	102
5.2. Биполярная оптическая генерация носителей заряда.....	104
5.3. Монополярная оптическая генерация носителей заряда. Максвелловское время релаксации.....	106
5.4. Механизмы рекомбинации	107
5.4.1. Межзонная излучательная рекомбинация	108
5.4.2. Межзонная ударная рекомбинация	109
5.4.3. Рекомбинация носителей заряда через ловушки	110
Вопросы	115

Глава 6. КИНЕТИЧЕСКИЕ ЯВЛЕНИЯ

В ПОЛУПРОВОДНИКАХ	116
6.1. Эффект Холла	116
6.2. Магниторезистивный эффект.....	121
6.3. Термоэлектрические явления	124
6.4. Поглощение света полупроводниками. Фотопроводимость	130
Вопросы	134

Глава 7. ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

В СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ ПОЛЯХ	135
7.1. «Разогрев» электронного газа	136
7.2. Ударная ионизация	139
7.3. Туннельный эффект и электростатическая ионизация.....	141
7.4. Эффект Ганна.....	143
Вопросы	147

Глава 8. КОНТАКТНЫЕ ЯВЛЕНИЯ	
В ПОЛУПРОВОДНИКАХ	148
8.1. Электронно-дырочный переход.....	148
8.2. Электрические процессы в <i>p-n</i> -переходе в отсутствие внешнего поля	151
8.3. Электрические процессы в <i>p-n</i> -переходе при наличии внешнего напряжения	156
8.4. Емкости <i>p-n</i> -перехода	162
Вопросы	163
ЗАКЛЮЧЕНИЕ	164
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА	167

Учебное издание

Гольдаде Виктор Антонович

ВВЕДЕНИЕ В ФИЗИКУ ПОЛУПРОВОДНИКОВ

Учебное пособие

Редактор *А. И. Кизик*

Компьютерная верстка *Т. В. Лукашенок*

Корректор *И. М. Подоматько*

Подписано в печать 26.10.2022. Формат 60×84/16.

Бумага офсетная. Ризография.

Усл. печ. л. 10,0. Уч.-изд. л. 11,6. Тираж 300 экз. Заказ 92.

Издатель и полиграфическое исполнение:

государственное учреждение образования

«Республиканский институт высшей школы».

Свидетельство о государственной регистрации издателя,

изготовителя, распространителя печатных изданий

№ 1/174 от 12.02.2014.

Ул. Московская, 15, 220007, г. Минск.