

Московский физико-технический институт
Кафедра общей физики

Лекция 6

ОБЪЁМНЫЕ ПОЛУПРОВОДНИКИ

заметки к лекциям по общей физике

В.Н.Глазков

Москва

2021

В данном пособии представлены материалы к лекции по теме «Объёмные полупроводники» из курса «Квантовая макрофизика», преподаваемого на кафедре общей физики МФТИ.

Пособие не претендует на полноту изложения материала и в основном является авторскими заметками к лекциям, оно содержит основные сведения по этой теме курса. Для подробного изучения тем студентам рекомендуется обратиться к классическому курсу Ч.Киттеля «Введение в физику твёрдого тела» [1] и другим источникам.

Основной материал по этой теме содержится в главе 11 книги Киттеля [1] и в методичке А.И.Морозова «Физика твёрдого тела» [2]. Обсуждение ряда вопросов в привязке к измерению характеристик полупроводника имеется в лабораторном практикуме [3]. В качестве подробного классического учебника по физике полупроводников можно рекомендовать учебник Бонч-Бруевича [4].

Символом † отмечены дополнительные разделы.

Основные понятия этой лекции:

- полупроводник, валентная зона, зона проводимости, запрещённая зона;
- положение химпотенциала при $T=0$ в чистом полупроводнике и в полупроводнике с донорными или акцепторными примесями;
- энергетическая диаграмма чистого полупроводника, полупроводника n-типа и p-типа;
- водородоподобные состояния, экситоны;

Оглавление

Общие представления о полупроводниках.....	4
Зонная структура полупроводникового кристалла.....	4
Полупроводники с нулевой шириной запрещенной зоны†.....	5
Прямозонные и непрямозонные полупроводники.....	7
Электроны и дырки в полупроводнике.....	9
Зонная структура реального полупроводника.	12
Свойства чистого полупроводника.....	16
Концентрации носителей заряда в чистом полупроводнике.....	16
Химпотенциал чистого полупроводника.....	18
Примесные слаболегированные полупроводники.....	18
Донорные и акцепторные примеси в полупроводнике.....	18
Концентрация носителей в легированных полупроводниках. Закон действующих масс (правило рычага).....	21
Энергетические диаграммы для полупроводника.....	25
Электропроводность полупроводников.....	26
Влияние концентрации носителей на электропроводность.....	26
Подвижность зарядов в полупроводнике†.....	27
Связанные состояния и рекомбинация носителей в полупроводниках.....	29
Экранирование ионизованной примеси в полупроводнике.....	29
Водородоподобные состояния вблизи примеси.....	30
Экситоны Ваннье-Мотта.....	31
Экситон-фононное взаимодействие†.....	33
Рекомбинация носителей†.....	35

Список литературы

- 1: Ч.Киттель, Введение в физику твёрдого тела., ,
- 2: А.И.Морозов, Физика твёрдого тела, МИРЭА (есть на сайте кафедры), 2008
- 3: Ф.Ф.Игошин, Ю.А.Самарский, Ю.М.Ципенюк, Лабораторный практикум по физике: Квантовая физика., Физматкнига, 2012
- 4: В.Л.Бонч-Бруевич и С.Г.Калашников, Физика полупроводников, М.: Наука, 1990
- 5: A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov and A. K. Geim, The electronic properties of graphene, Reviews of Modern Physics, 81, 109(2009)
- 6: Brian Ray, II-IV Compounds: International Series of Monographs in The Science of The Solid State, vol.2, Pergamon Press, 1969
- 7: Charles P. Poole, Jr., Encyclopedic Dictionary of Condensed Matter Physics, vol.1, Academic Press, 2004
- 8: W. R. Frensley and N. G. Einspruch editors, Heterostructures and Quantum Devices, Academic Press, 1994
- 9: Semiconductors on NSM , <http://www.ioffe.ru/SVA/NSM/Semicond/>
- 10: под ред. В.А.Овчинкина, Сборник задач по общему курсу физики, часть 3 (раздел "Строение вещества"), Физматкнига, 2009
- 11: Bart Van Zeghbroeck, Principles of Semiconductor Devices , , 2011
- 12: T. Kazimierzczuk, D. Fröhlich, S. Scheel, H. Stolz and M. Bayer, Giant Rydberg excitons in the copper oxide Cu₂O, Nature, 514, 343 (arxiv 1407.0691)(2014)
- 13: С.Готовко, Бакалаврская дипломная работа МФТИ, "Нелинейные явления в экситон-фононных спектрах в заиси меди", под. рук. А.В.Горбунова, ИФТТ РАН, 2016

Общие представления о полупроводниках.

Зонная структура полупроводникового кристалла.

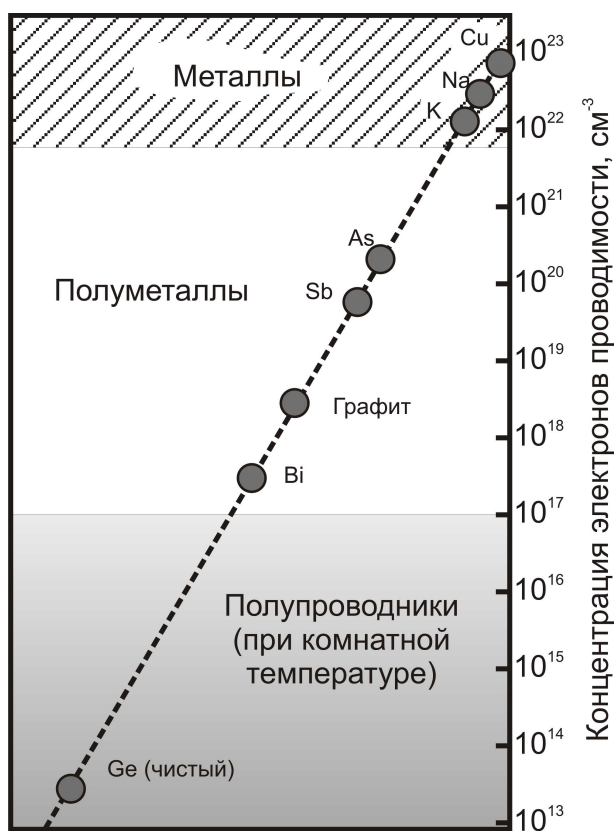


Рисунок 1: Концентрации электронов проводимости в металлах и полупроводниках. На основе рисунка из книги [1].

Главным свойством металлов, в каком-то смысле определением того, что такое металл, является наличие электронов проводимости (наличие частично заполненной энергетической зоны) при температуре $T=0$. Все остальные вещества являются в строгом смысле этого слова диэлектриками: в них при $T=0$ есть только полностью заполненные или совершенно свободные энергетические зоны.

При конечной температуре в диэлектриках возникает ненулевая вероятность перехода электрона из заполненной зоны в свободную, что будет приводить к возникновению проводимости. Однако часто энергетическое расстояние между заполненной при $T=0$ зоной (называемой валентной) и свободной (называемой зоной проводимости) оказывается настолько велико, что вероятность такого переброса, определяемая бoльцмановской экспонентой $e^{-\Delta/T}$, оказывается ничтожна — с практической точки зрения именно эти вещества являются изоляторами (диэлектриками).

Полупроводниками *по определению* называют вещества, в которых энергетическое расстояние между зоной проводимости и валентной зоной (щель между дном зоны проводимости и потолком валентной зоны, ширина запрещенной зоны) оказывается «не слишком велико», так что при интересующих нас температурах оказывается достаточно много электронов в зоне проводимости. Это определение весьма условно, обычно к полупроводникам относят соединения с шириной запрещенной зоны менее 3 эВ, хотя иногда и системы с несколько большей шириной запрещенной зоны называют широкозонными.

полупроводниками. В качестве некоторой численной характеристики можно использовать концентрацию свободных носителей заряда при комнатной температуре. В хороших металлах она составляет $10^{22} \dots 10^{23} \text{ 1/см}^3$, в полуметаллах она выше 10^{17} 1/см^3 , в «хороших» диэлектриках концентрация носителей ниже 10^{10} 1/см^3 , а в полупроводниках она обычно составляет от 10^{13} до 10^{17} 1/см^3 (рисунок 1).

Полупроводники с нулевой шириной запрещенной зоны[†]

Промежуточным звеном между металлами и полупроводниками оказываются так называемые полупроводники с нулевой шириной запрещенной зоны. В таких системах в силу случайных¹ или симметричных причин валентная зона и зона проводимости соприкасаются. При $T=0$ в таких системах оказывается полностью заполнена валентная зона и полностью свободна зона проводимости. Однако из-за отсутствия запрещенной зоны часть электронов перейдет в зону проводимости при низкой температуре — то есть, как в металлах, электронная система оказывается восприимчива к слабому воздействию. В то же время, число электронов, откликающихся на это воздействие, оказывается мало и их количество зависит от температуры иначе, чем в металлах.

Широко известным примером такого типа спектра является графен² (рисунок 2) [5], однако этот эффект известен с 60-х годов 20 века для полупроводников на основе HgTe и HgSe [6], [7]. Полупроводниками с нулевой шириной запрещенной зоны являются также некоторые соединения из семейств $\text{Bi}_x\text{Sb}_{1-x}$, $\text{Pb}_{1-x}\text{Sn}_x\text{Te}$ и $\text{Pb}_{1-x}\text{Sn}_x\text{Se}$ (при некоторых концентрациях x), серое олово ($\alpha\text{-Sn}$) [7]. Чистый HgTe является полуметаллом с небольшим перекрытием зон, при замещении более 20% ртути кадмием он становится полупроводником, а при пороговом уровне примеси кадмия ширина запрещенной зоны обращается в ноль (рисунок 3).

-
- 1 Напомним, что такое закрытие щели между некоторыми зонами происходит при специальных значениях параметров в рассмотренной на лекции зонной структуре модели Кронига-Пенни.
 - 2 Пример графена невозможно обойти стороной в силу того, что он находится «на слуху». В нашем курсе мы не будем систематически рассматривать свойства графена. Отметим, что многие из широко известных свойств графена являются скорее результатами теоретического расчёта в идеализированной модели (идеальная двумерность, отсутствие примесей и дефектов), а на свойства реальных графеновых образцов эти «несовершенства» реального мира могут оказывать определяющее значение. Так в образцах графена наблюдают небольшую проводимость при $T=0$, которая по всей видимости связана с неконтролируемыми дефектами на поверхности.

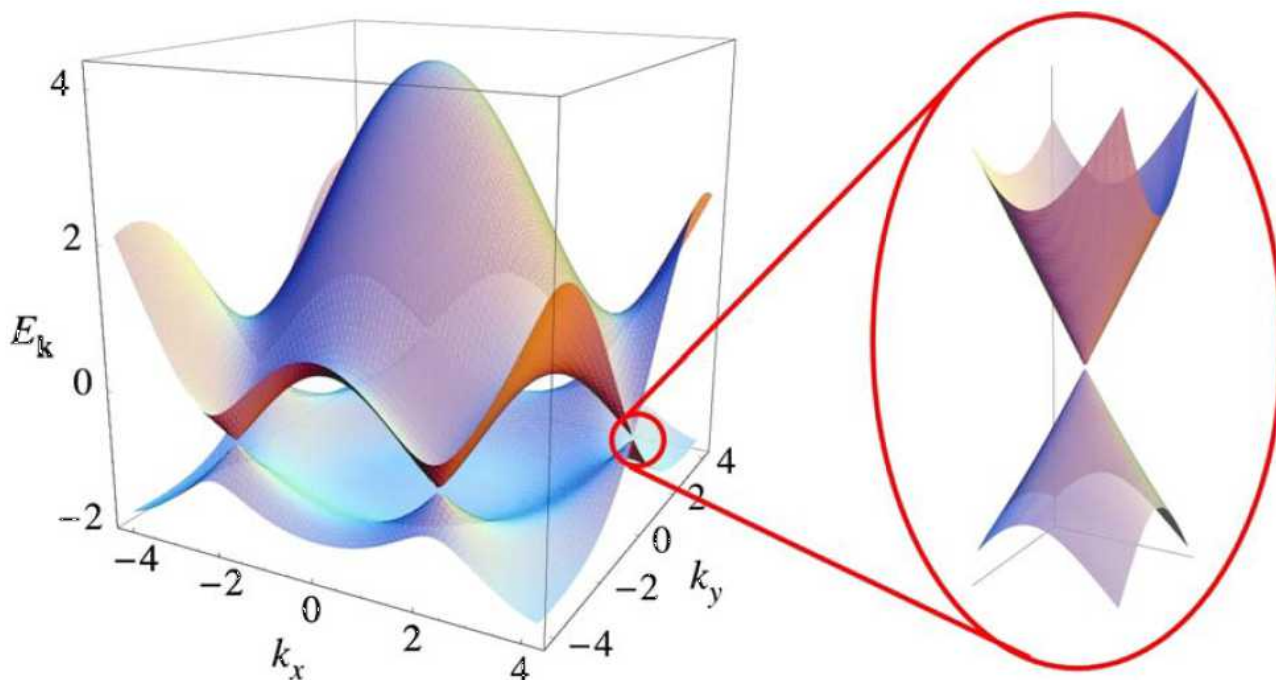


Рисунок 2: Слева: Электронный спектр графена. Справа: увеличенный фрагмент спектра в одной из дираковских точек. Из обзора [5].

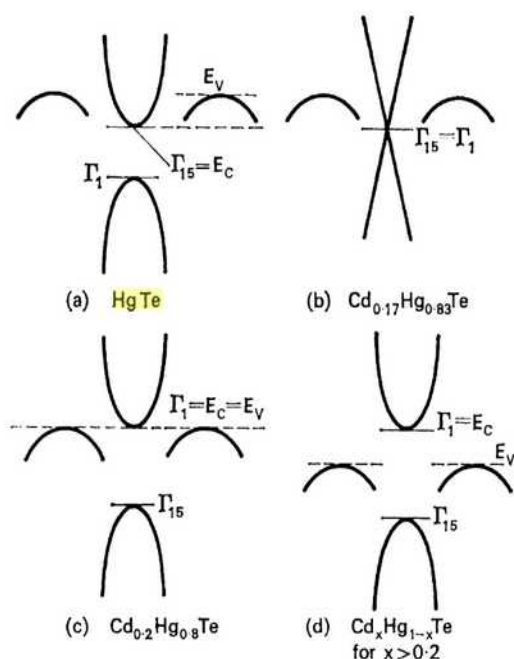


Рисунок 3 Спектры электронов (зонная структура) соединений Cd-Hg-Te для разной концентрации кадмия. (a) чистый HgTe, полуметалл с перекрытием зон (b) 17% Cd, полуметалл с линейным спектром электронов на некоторых ветвях спектра (c) 20% Cd, перекрытие зон пропало — полупроводник с нулевой запрещённой зоной (d) более высокие концентрации кадмия, обычный полупроводник. Из книги [6].

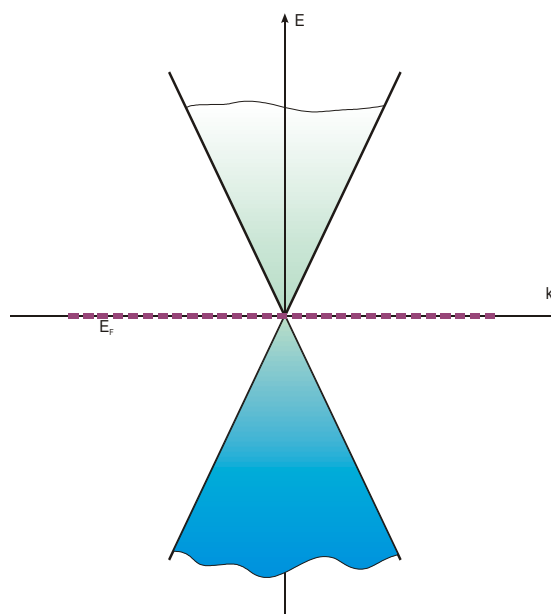


Рисунок 4: Фрагмент спектра электронов в полупроводнике с нулевой шириной запрещённой зоны и линейным спектром в точке касания зон (случай графена): если уровень энергии Ферми проходит точно через вершины "конусов", то плотность состояний на уровне Ферми нулевая.

Зависимость энергии электрона от волнового вектора в точке соприкосновения зон зависит от типа полупроводника: в соединениях на основе HgTe и HgSe спектр в точке касания квадратичен (рисунок 3), в графене — линейен (рисунки 2 и 4). Отметим, что для линейного в окрестности точки касания зон спектра (случай графена) можно записать для энергии $\varepsilon = \hbar v_F |\vec{k}|$, где характерную скорость v_F называют скоростью Ферми. Этот спектр напоминает спектр безмассовых (нерелятивистских) частиц и поэтому иногда его называют дираковским, а точки в k -пространстве, вблизи которых спектр представим в таком виде — точками Дирака.

При $T=0$ валентная зона заполнена полностью и уровень Ферми³ для полупроводника с нулевой шириной запрещённой зоны проходит через точку пересечения валентной зоны и зоны проводимости. Площадь поверхности Ферми при этом оказывается равна нулю, что приводит к занулению плотности состояний на поверхности Ферми. Для изотропного спектра электронов это утверждение доказывается прямолинейно (вычисления далее для трёхмерного случая): $D(E) = \frac{dN}{dE} = \frac{dN}{dk} \frac{1}{dE/dk} = \frac{1}{(2\pi)^3} \frac{dV_k}{dk} \frac{1}{\hbar V_{gp}} = \frac{1}{(2\pi)^3 \hbar} \left[\frac{S_k}{V_{gp}} \right]_E$, где

S_k — площадь (в k -пространстве) изоэнергетической поверхности с заданной энергией, а V_{gp} — групповая скорость электронов с этой энергией. Поэтому, несмотря на отсутствие запрещённой зоны, проводимости при $T=0$ в такой системе (в идеальном случае) нет. Мы не будем систематически рассматривать полупроводники с нулевой шириной запрещённой зоны в нашем курсе.

Прямозонные и непрямозонные полупроводники

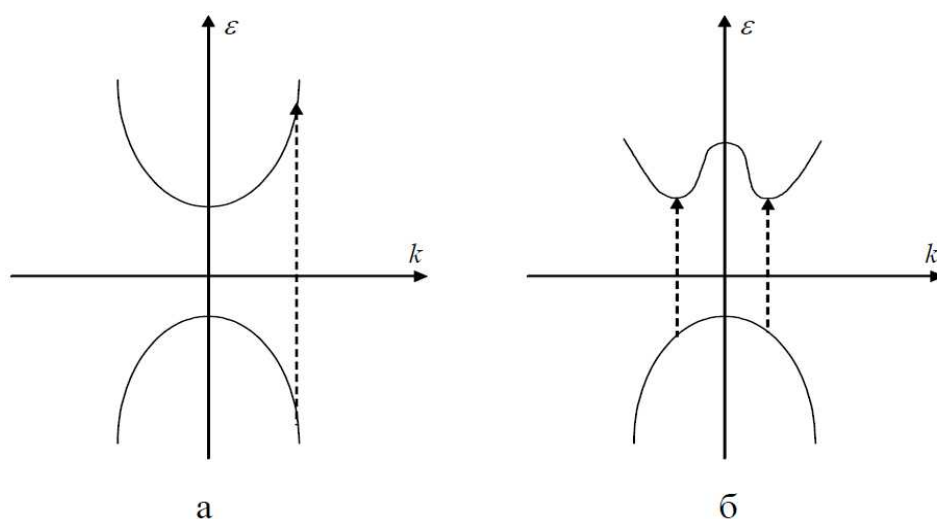


Рисунок 5: Схематическое изображение полупроводника с прямой и непрямыми запрещёнными зонами. Пунктирные линии — возможные оптические переходы из валентной зоны в зону проводимости. Из книги [2].

Ширина запрещённой зоны (E_g , иногда также называется энергетической щелью) есть

³ Напомним, что для металла уровень Ферми — это максимальная энергия заполненного электронного состояния. Для полупроводников с нулевой шириной запрещённой зоны энергетический спектр непрерывен, поэтому представление об уровне Ферми может быть перенесено на них непосредственно. Для полупроводников с ненулевой шириной запрещённой зоны вопрос о положении уровня химического потенциала при $T=0$ (иногда также называемого уровнем Ферми) оказывается более сложен и мы подробно рассмотрим его позднее (стр.19).

расстояние между минимальной энергией электронного состояния в зоне проводимости (*дном* зоны проводимости) и максимальной энергией электронного состояния в валентной зоне (*потолком* валентной зоны). Мы уже знаем, что зонная структура электронного спектра в реальном трёхмерном кристалле может быть устроена весьма сложным образом. Однако при малом числе термоактивированных электронов в зоне проводимости можно всегда считать, что освобождаются только состояния вблизи максимумов спектра в валентной зоне и оказываются занятыми только состояния вблизи минимумов спектра в зоне проводимости.

Таким образом, качественно важными оказываются следующие особенности спектра:

- совпадают ли в k -пространстве положения минимумов в зоне проводимости и максимумов в валентной зоне,
- количество таких максимумов и минимумов,
- а также эффективные массы (вообще говоря, анизотропные), описывающие эти минимумы и максимумы спектра.

Если минимум энергии в зоне проводимости совпадает в k -пространстве с максимумом энергии в валентной зоне, то запрещённую зону называют *прямой*, а если не совпадает, то *непрямой*, а полупроводник: *прямозонным* и *непрямозонным*, соответственно. Схематически это различие изображено на рисунке 5. Оно оказывается принципиально, например, в процессах оптического возбуждения электронов и релаксации возбуждённых состояний. Так как волновой вектор фотона с энергией порядка электронвольт много меньше бриллюэновского, межзонные процессы с участием фотонов происходят практически без изменения волнового вектора электрона. Поэтому процесс перехода электрона из минимума зоны проводимости в максимум валентной зоны с излучением фотона (и обратный процесс с поглощением фотона минимальной энергии) оказывается существенно подавлен (запрещён) для непрямой запрещённой зоны. В результате время излучательной рекомбинации (процесса с отдачей избытка энергии в форме излучения светового кванта) электрона в прямозонном полупроводнике InSb составляет около 1 мксек, а аналогичные времена в непрямозонных полупроводниках Si или Ge составляют около 3 часов и 0.43 сек [2].

Параметры различных полупроводников представлены в таблице 1.

Таблица 1: Ширина и тип запрещённой зоны (d - прямая, i - непрямая) в некоторых полупроводниковых кристаллах. Из книги [1].

Кристалл	Тип запрещ. зоны	E_g , эВ		Кристалл	Тип запрещ. зоны	E_g , эВ	
		0 °K	300 °K			0 °K	300 °K
алмаз	i	5.4		PbS	d	0.29	0.34...0.37
Si	i	1.17	1.14	PbSe	d	0.17	0.27
Ge	i	0.74	0.67	PbTe	d	0.19	0.30
α -Sn	d , нулевая ширина	0	0	CdS	d	2.58	2.42
InSb	d	0.23	0.18	CdSe	d	1.84	1.74
InAs	d	0.36	0.35	ZnO		3.44	3.2
InP	d	1.29	1.35	ZnS		3.91	3.6
GaP	i	2.32	2.26	SnTe	d	0.3	0.18
GaAs	d	1.52	1.43	AgCl			3.2
GaSb	d	0.81	0.78	AgI			2.8
AlSb	i	1.65	1.52	Cu ₂ O		2.17	
SiC (hex)		3.0		TiO ₂		3.03	
Te	d	0.33		ZnSb		0.53	0.56

Электроны и дырки в полупроводнике.

Количество электронов в зоне проводимости и вакансий в валентной зоне в случае полупроводника обычно невелико. Поэтому оказывается удобно рассматривать свойства полупроводника на языке квазичастиц, а не на языке «настоящих» электронов. Это позволит вместо рассмотрения большого числа «настоящих» электронов⁴, заполняющих валентную зону со сложным спектром $\epsilon(\vec{k})$, рассматривать малое число квазичастиц с квадратичной дисперсией вблизи квадратичных экстремумов зон.

Процесс перехода электрона из валентной зоны в зону проводимости удобно для этого рассматривать как рождение пары «частица-античастица» на фоне «вакуума» полностью заполненной валентной зоны и полностью свободной зоны проводимости. При этом «частица» рождается в зоне проводимости и полностью соответствует «настоящему» электрону, а античастица описывает поведение валентной зоны *без одного электрона*, как мы увидим чуть позже эта античастица по своим свойствам не совсем тождественна вакансии в валентной зоне, образовавшейся после ухода электрона.

Такие частицы традиционно называют электронами⁵, античастицы — дырками. Отметим, что эти обозначения похожи на введенные нами при обсуждении металлов электронные и дырочные ферми-поверхности, но есть некоторые отличия: в металлах при $T=0$ имеется определённая поверхность Ферми (поверхность в k -пространстве на которой скачком меняется электронная плотность), а в полупроводниках никакой поверхности Ферми нет, при практически важных температурах газ квазичастиц в полупроводнике в большинстве случаев

4 Подчеркнём, что эти электроны движутся в кристалле, то есть уже имеют сложный спектр и закон дисперсии и не являются свободными электронами, то есть уже являются квазичастицами. Может быть не совсем удачный термин «настоящие» подчёркивает, что речь идёт о многочисленных электронных состояниях в кристалле, число которых равно числу электронов.

5 Мы надеемся, что в ходе изложения удастся избежать путаницы между электронами в смысле электронов, попавших в зону проводимости и «настоящими» электронами, заполняющими в том числе и валентную зону.

оказывается невырожденным, при $T=0$ полупроводник становится диэлектриком. Поэтому в металле для описания его свойств при не слишком высоких (и реально — при всех практически значимых) температурах нас интересовала лишь плотность состояний на поверхности Ферми, для вычисления которой можно было перейти к понятию дырки, как просто отсутствия электрона в недозаполненной зоне. В полупроводнике же образовавшаяся в валентной зоне вакансия существует среди других электронов.

Квазиимпульс электронов и дырок принято отсчитывать от минимума или максимума соответствующей зоны. При необходимости⁶ он легко пересчитывается в квазиимпульс в координатах первой зоны Бриллюэна по известному положению этих экстремумов. Начало отсчёта энергии произвольно, есть два стандартных выбора: отсчитывать энергию от потолка валентной зоны или от дна зоны проводимости. В рамках нашего изложения мы будем отсчитывать энергию от потолка валентной зоны. Обратите внимание, что от этого выбора будет зависеть запись некоторых результатов.

Тогда, считая эффективную массу изотропной, закон дисперсии электронов можно представить в виде $\varepsilon_e(\vec{q}) = E_g + \frac{\hbar^2 \vec{q}^2}{2 m_e}$, где m_e - эффективная масса электрона. Заряд электрона конечно же отрицательный.

Античастица (дырка) имеет положительный заряд по определению (это просто требование закона сохранения заряда), а вот её масса, волновой вектор и закон дисперсии заслуживают более детального разговора. Для простоты будем полагать, что имеем дело с единственным максимумом в валентной зоне. Спектр «настоящих» электронов вблизи квадратичного максимума $\varepsilon(\vec{q}) \approx -\frac{\hbar^2 \vec{q}^2}{2 M}$ формально соответствует *отрицательной* эффективной массе

$(-M) < 0$. Уравнение динамики для вакансии (например, при включении электрического поля) будет таким же как и для электрона $\hbar \frac{d\vec{q}}{dt} = -e \vec{E}$, это связано с тем, что вакансия движется не независимо, а вместе с электронами: движение вакансии в электрическом поле можно рассматривать как перемещение свободного места в цепочке электронных состояний — если все электроны сдвинулись в k -пространстве⁷ на δk , то и вакансия сместилась ровно на столько же (рисунок 6).

6 Точное положение этих экстремумов может быть важно для некоторых правил отбора, например в оптических процессах, однако для термодинамических и транспортных свойств, о которых мы будем в основном говорить, этот перенос начала отсчёта не принципиален.

7 Напомним, что мы говорим о заполненной валентной зоне. Мы рассматриваем поведение вблизи потолка зоны, а вблизи края зоны Бриллюэна при смещении в импульсном пространстве будут происходить процессы переброса, переставляющие электроны с левого на правый (по рисунку 6) края зоны Бриллюэна и таким образом сохраняя зону заполненной (за исключением рассматриваемой вакансии).

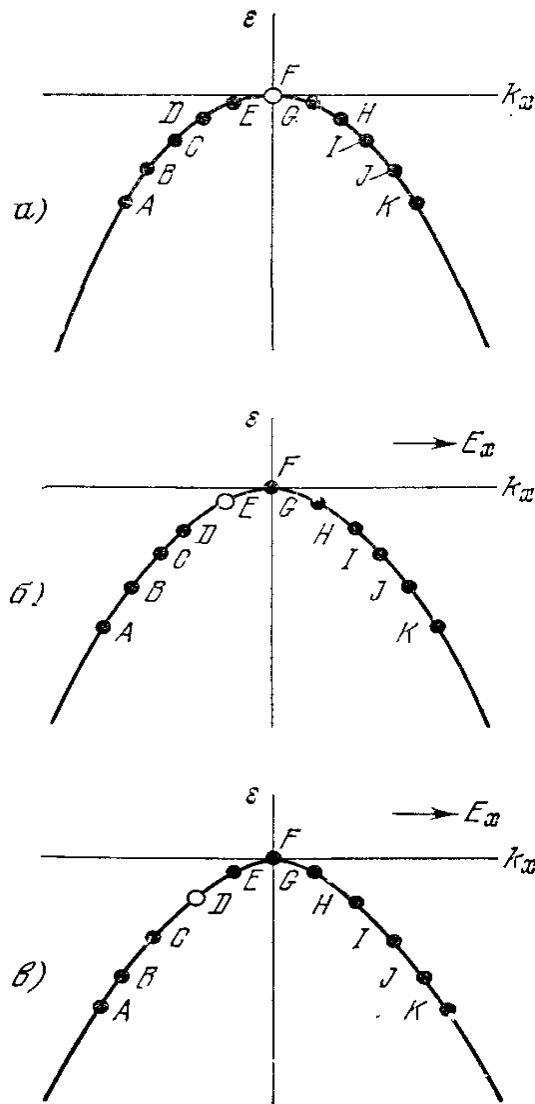


Рисунок 6: Движение вакансии в заполненной валентной зоне в присутствии электрического поля. Буквы A, B, C.... нумеруют электронные состояния. В начальный момент (а) вакансия образована в состоянии F. При смещении всех электронов на одну позицию влево по рисунку (б) вакансия смещается в положение E левее, и так далее (в). Из книги [1].

Суммарный квазиимпульс полностью заполненной зоны равен нулю. Поэтому, если забрать электрон из состояния с волновым вектором \vec{k}_e , то квазиимпульс «зоны с вакансией» будет равен $-\vec{k}_e$. Таким образом, если описывать свойства заполненной зоны с вакансией как появление некоторой квазичастицы-дырки, то этой квазичастице должен быть приписан квазиимпульс $\vec{k}_h = -\vec{k}_e$. Подчеркнём, волновой вектор квазичастицы-дырки не совпадает (противоположен) с волновым вектором вакансии в зоне «настоящих» электронов.

Уравнение динамики для дырки тогда также поменяет знак: $\hbar \frac{d\vec{q}_h}{dt} = e\vec{E}$, то есть формально будет, действительно, соответствовать положительному заряду. Энергия дырок, как и для вводившихся нами при рассмотрении возбуждений ферми-газа античастиц, должна быть положительна: энергия электронов валентной зоны может быть понижена, если самый энергичный электрон (с «потолка» валентной зоны) переместить на вакантное место. Соответственно, закон дисперсии для квазичастиц должен быть «отражён» в положительную полуплоскость. Действительно, если оставить его неизменным, то в геометрии рисунка 6

групповая скорость дырки в состоянии с $k_h = -k_e > 0$ окажется отрицательной, то есть изменится направление электрического тока (ток потечёт против оси X), в то время как в представлении «настоящих» электронов именно избыточные электроны правее от вакансии (положения G на панели (б) и H на панели (в)) создают нескомпенсированный электрический ток и отрицательность их групповой скорости в совокупности с отрицательностью заряда электрона даст электрический ток, текущий вдоль оси X.

Таким образом, если вблизи потолка валентной зоны закон дисперсии «настоящих» электронов имеет вид $\epsilon(\vec{q}) \approx -\frac{\hbar^2 \vec{q}^2}{2|M|} = -\frac{\hbar^2 \vec{q}^2}{2m_h}$, то для квазичастиц дырок закон дисперсии

$$\epsilon_h(\vec{q}) \approx \frac{\hbar^2 \vec{q}^2}{2m_h}, \text{ где } m_h > 0 \text{ — эффективная масса дырки.}$$

Это отражение закона дисперсии требует некоторой аккуратности при применении фермиевской функции распределения к дыркам. Дырке с энергией $\epsilon > 0$ соответствует вакансия в валентной зоне с энергией $\epsilon_{vac} = -\epsilon$. Концентрации вакансий и дырок равны по построению. Равновесное распределение для вакансий можно найти как отличие от 1 фермиевской функции распределения для «настоящих» электронов с энергией, равной энергии вакансии ϵ_{vac} :

$$\begin{aligned} n_h(\epsilon) &= n_{vac}(\epsilon_{vac}) = 1 - n_e(\epsilon_{vac}) = 1 - \frac{1}{\exp((\epsilon_{vac} - \mu)/T) + 1} = \frac{1}{\exp(-(\epsilon_{vac} - \mu)/T) + 1} = \\ &= \frac{1}{\exp((\epsilon + \mu)/T) + 1} \end{aligned}$$

То есть, функция распределения для дырок имеет тот же фермиевский вид, но с изменённым знаком химпотенциала. В качестве мнемонического правила, это можно представить как изменение направления отсчёта энергии для дырок.⁸ Другой мнемонической формулировкой этого правила является правило: «электроны — тонут, а дырки — всплывают»: минимуму энергии электрона соответствует положение на дне зоны проводимости, а минимуму энергии дырки — положение под потолком валентной зоны.

8 Обратите внимание, что применение этого мнемонического правила требует внимания при несовпадении нуля отсчёта энергии с потолком валентной зоны. Если энергия отсчитывается от произвольного уровня, то для правильного вычисления функции распределения для дырок (см. уравнения выше) мы должны приписать дыркам энергию $\epsilon_h = -\epsilon_{vac}$ и сменить знак химпотенциала. Например, при отсчёте энергии от дна зоны проводимости спектр электронов $\epsilon_e = \frac{\hbar^2 q^2}{2m_e}$, спектр вакансий $\epsilon_h = -E_g - \frac{\hbar^2 q^2}{2m_h}$, а спектр дырок $\epsilon_h = -\epsilon_{vac} = E_g + \frac{\hbar^2 q^2}{2m_h}$.

Зонная структура реального полупроводника.

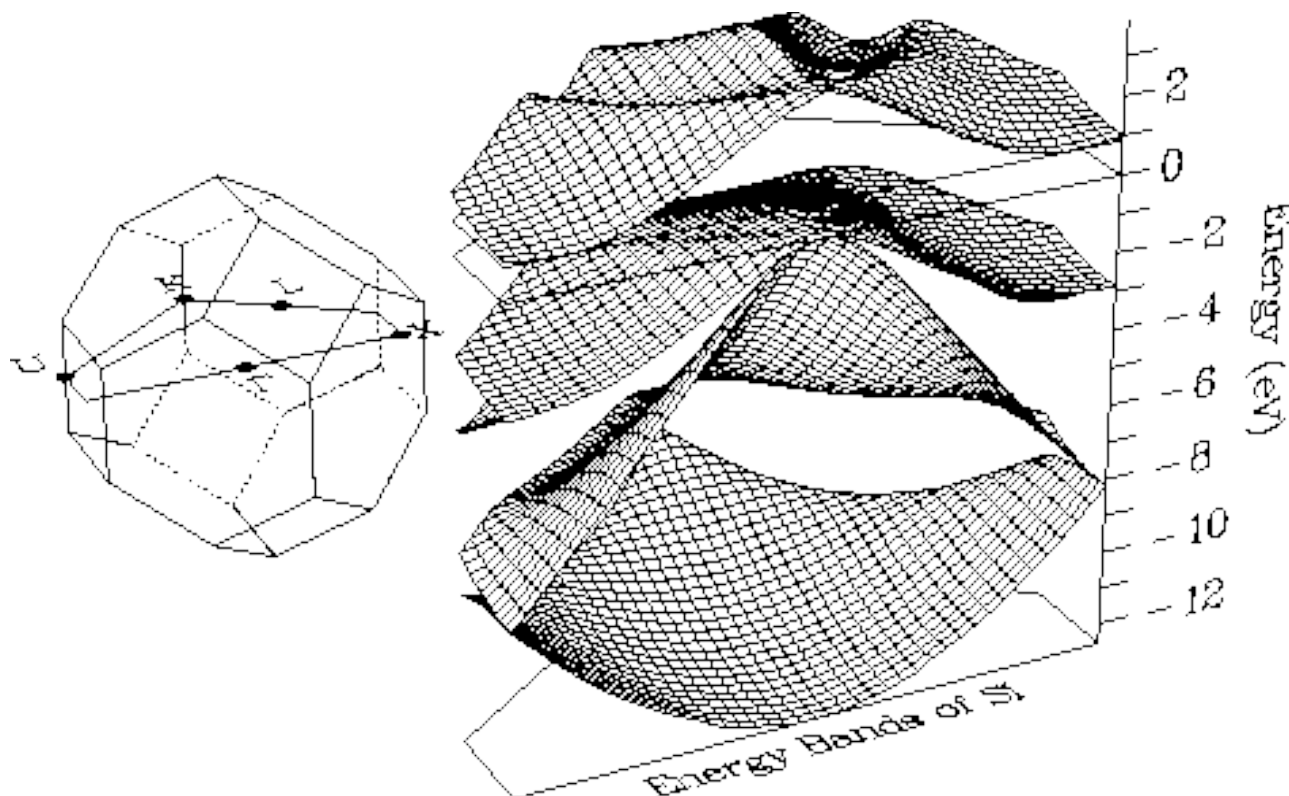


Рисунок 7: Зонная структура кремния. Слева: первая зона Бриллюэна и сечение, в котором построены профили энергии. Справа: вычисленная энергия для различных зон, показана только нижняя зона проводимости. Ноль отсчёта энергии соответствует потолку валентной зоны. Из книги [8].

Большинство решаемых нами задач будет рассматриваться в простой модели, когда имеются единственные минимумы и максимумы энергии электронов в зоне проводимости и в валентной зоне, и когда эффективные массы электронов и дырок изотропны. Этого достаточно для качественного понимания многих простых задач физики полупроводников, однако не всегда соответствует реальным системам. В качестве иллюстративного примера рассмотрим широко используемые в полупроводниковых технологиях материалы — кремний и германий, и широко используемый в приготовлении низкоразмерных полупроводниковых систем арсенид галлия (GaAs).

Примеры вычисленных зонных структур для кремния и арсенида галлия [8] показаны на рисунках 7, 10. Сечения дисперсионных поверхностей для кремния, германия и арсенида галлия с указанием характерных энергий показаны на рисунках 8, 11, 12 [9].

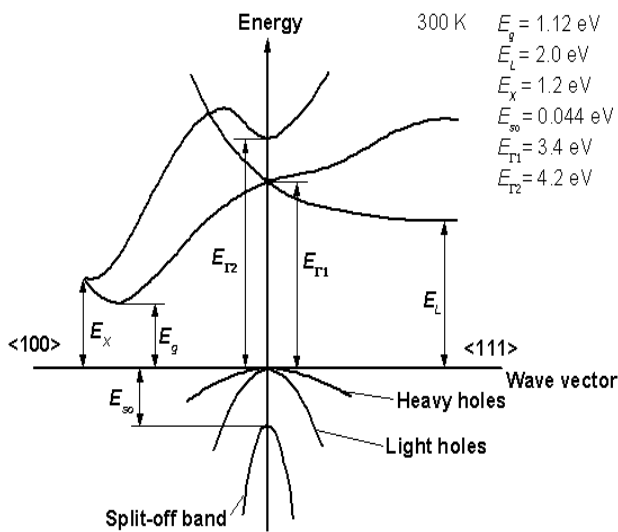


Рисунок 8: Сечение дисперсионных поверхностей кремния в направлениях $[100]$ и $[111]$. С сайта [9].

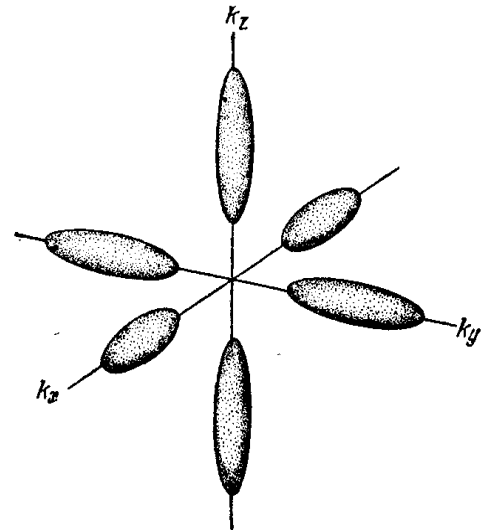


Рисунок 9: Эллипсоиды постоянной энергии для электронов в зоне проводимости кремния. Из книги [1].

Отметим вкратце происхождение зонной структуры в кремнии, которая оказывается сложнее простых модельных случаев рассмотренных нами ранее. Кремний кристаллизуется в структуру алмаза с двумя атомами в примитивной ячейке. Электронная конфигурация атома кремния $3s^2 3p^2$, но s -орбиталь в окружении подходящей симметрии гибридизуется с тремя p -орбиталями (sp^3 -гибридизация). В изолированном атоме sp^3 орбиталь четырёхкратно вырождена по энергии и способствует образованию четырёх эквивалентных химических связей в направлении на вершины тетраэдра. Для двух атомов в примитивной элементарной ячейке происходит расщепление вырожденных уровней на две группы по 4 уровня. При дальнейшем объединении в кристалл эти группы уровней sp^3 орбиталей расщепляются в зоны, описываемые поверхностями $\varepsilon(\vec{k})$ в первой зоне Бриллюэна. При этом нижняя зона из верхней группы не пересекается с зонами, образующимися из нижней группы.⁹ В результате имеющиеся 8 электронов на примитивную ячейку полностью заполняют четыре нижних зоны и образуют валентную зону. Одна из этих зон оказывается сильно ниже остальных (интервал от -12 до -8 эВ на рисунке 7) и всегда остаётся полностью заполнена. Для трёх остальных ветвей спектра в валентной зоне поверхности $\varepsilon(\vec{k})$ имеют максимум в центре зоны Бриллюэна, причём у двух из них максимумы совпадают с потолком валентной зоны. Это означает, что в кремнии имеется два типа различных дырок с различными эффективными массами (так как кривизна поверхности в максимуме разная для разных зон), а при достаточно высокой температуре может сказываться и наличие третьего типа дырок. Два основных типа дырок имеют существенно отличные эффективные массы: более пологая дисперсионная поверхность соответствует «тяжёлым» дыркам с эффективной массой $0.49 m_0$, а поверхность с более острым максимумом — «лёгким» дыркам с эффективной массой $0.16 m_0$ (m_0 — масса свободного электрона). Эффективная масса дырок в

9 Это является результатом точных расчётов для зонной структуры и в рамках нашего курса к такому свойству зон можно относиться как к случайному (но очень важному для приложений) совпадению из-за удачного сочетания параметров взаимодействия электрона с решёткой и параметров перекрытия электронных волновых функций в кристалле кремния.

отщеплённой зоне $0.24 m_0$.

В зоне проводимости также есть несколько ветвей спектра, образующиеся при расщеплении в зоны второй группы sp^3 -орбиталей, минимум нижней ветви расположен вблизи от границы зоны Бриллюэна (в направлении $[100]$ и эквивалентных), поэтому кремний является непрямозонным полупроводником. Энергия в минимуме равна 1.12 эВ. При этом в силу кубической симметрии кристалла имеется 6 эквивалентных минимумов. Эти минимумы оказываются сильно анизотропны, эллипсоиды постоянной энергии (рисунок 9) сильно вытянуты от центра зоны: продольная (соответствующая сечению дисперсионной поверхности от центра зоны Бриллюэна) эффективная масса равна $0.98 m_0$, а поперечная — $0.19 m_0$ [9]. В принципе, в зоне проводимости есть и другие локальные минимумы, попадание в которые может быть важно при достаточно высоких температурах¹⁰ (на границе зоны имеется минимум одной из ветвей спектра на высоте 1.2 эВ, на 0.08 эВ выше основного).

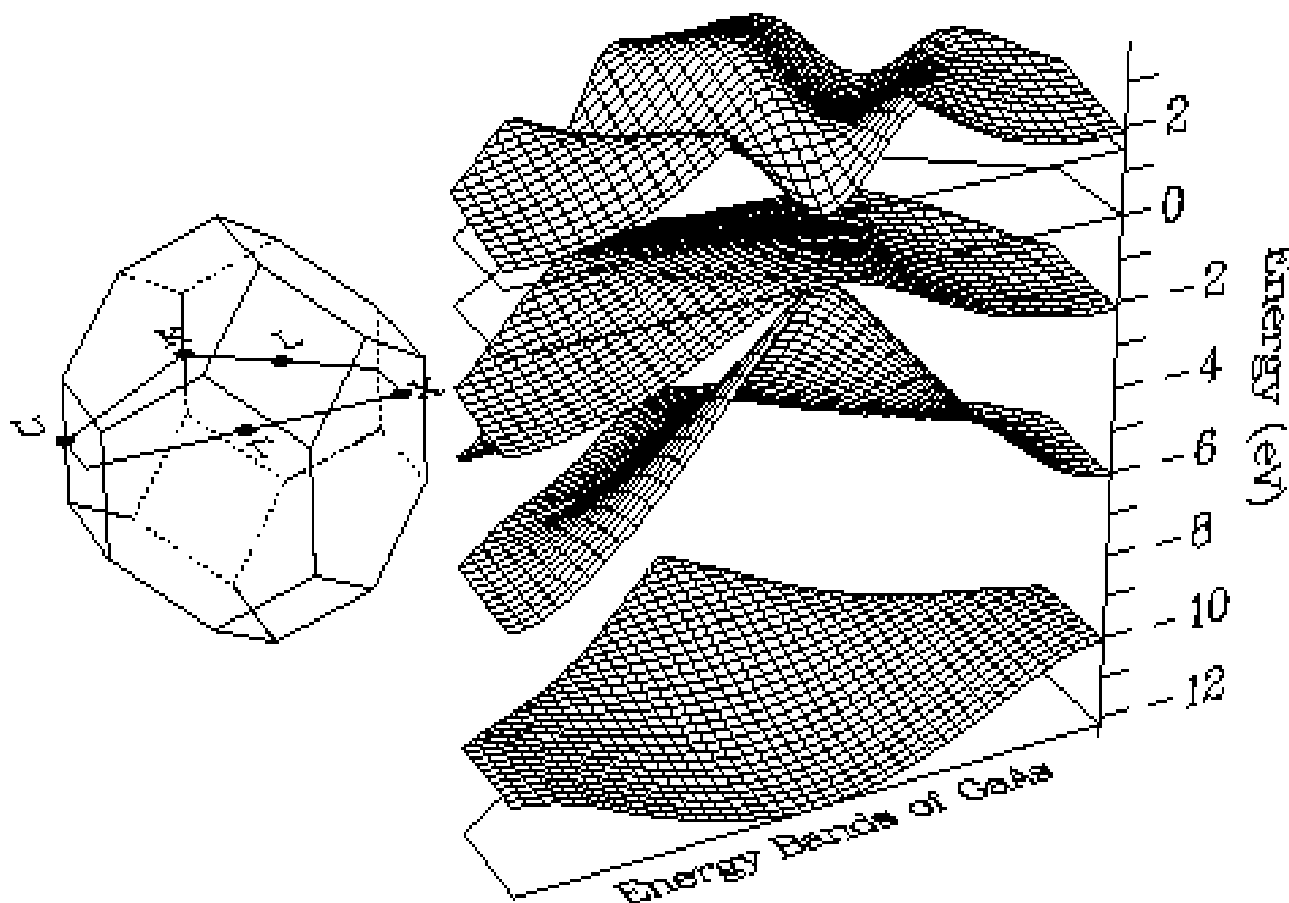


Рисунок 10: Зонная структура GaAs. Слева: первая зона Бриллюэна и рассматриваемое сечение первой зоны. Справа: вычисленная в выбранном сечении энергия, показана только нижняя из зон проводимости. Ноль отсчёта энергии соответствует потолку валентной зоны. Из книги [8].

В GaAs также имеется четыре поверхности $\varepsilon(\vec{k})$ в валентной зоне (рисунок 10), три из них имеют максимум при $k=0$, у двух этот максимум совпадает с потолком зоны. Главный минимум зоны проводимости оказывается также в центре зоны Бриллюэна, поэтому

¹⁰ Вблизи квадратичного экстремума спектра $\varepsilon(\vec{k})$ всегда есть максимум плотности состояний, поэтому при сложной форме спектра основной вклад в различные усреднённые физические величины будут давать состояния вблизи экстремумов.

GaAs – прямозонный полупроводник. Масштаб отщепления третьей валентной зоны и локальных минимумов зоны проводимости заметно больше, чем в кремнии (см. рисунок 11). Эффективные массы дырок равны $0.51 m_0$ и $0.082 m_0$, эффективная масса электронов в основном минимуме $0.063 m_0$ [9].

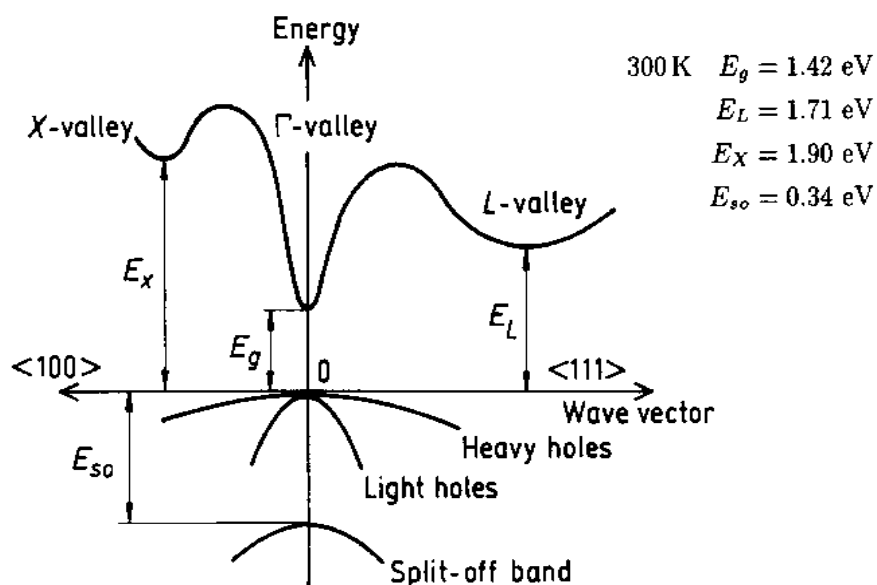


Рисунок 11: Сечение дисперсионных поверхностей арсенида галлия в направлениях $[100]$ и $[111]$. С сайта [9].

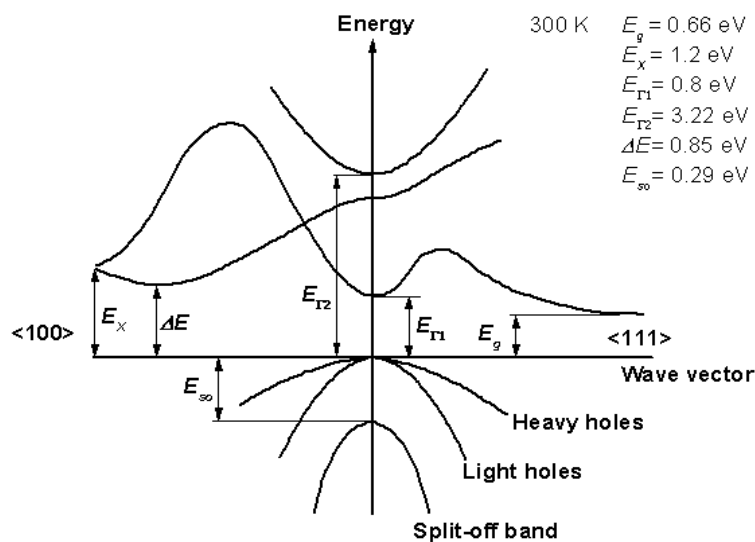


Рисунок 12: Сечение дисперсионных поверхностей германия в направлениях $[100]$ и $[111]$. С сайта [9].

Германий (рисунок 12), как и кремний, является непрямозонным полупроводником. В нём также имеется два основных типа дырок с массами $0.33 m_0$ и $0.043 m_0$. Главные

минимумы зоны проводимости лежат на границе зоны Бриллюэна в направлениях типа [111]. Таких направлений всего 8 (это направления на вершину куба), однако так как только половина из каждой области попадает в первую зону Бриллюэна (точка минимума лежит в центре шестиугольных граней первой зоны, показанной на рисунках 7, 10), то всего имеется 4 симметрично эквивалентных минимума. Энергия минимума 0.66 эВ, имеются также близкий минимум в центре зоны Бриллюэна с энергией 0.8 эВ и минимум в направлении [100] с энергией 0.85 эВ (который шестикратно размножится операциями симметрии). Электронные состояния в главных минимумах зоны проводимости, как и в кремнии, характеризуются сильно анизотропной эффективной массой: продольная эффективная масса равна $1.59m_0$, а поперечная $0.0815m_0$ [9].

Таким образом, в реальных полупроводниках электронный спектр бывает устроен сложно и для точного описания может оказываться необходимым учитывать наличие нескольких участков дисперсионной кривой, соответствующих электронам или дыркам (причём число электронных и дырочных участков дисперсионных поверхностей может быть разным), а также возможную анизотропию эффективной массы носителя. В задачах нашего курса, как правило, подразумевается упрощённая ситуация, когда имеется единственный максимум валентной зоны (один тип дырок) и единственный минимум зоны проводимости (один тип электронов), причём эффективные массы изотропны. Описание реальных систем на языке такой упрощённой модели приводит к перенормировке параметров (таких как эффективная масса), поэтому параметры полупроводников, приводимые в задачах, могут несколько отличаться от приведённых здесь.

Свойства чистого полупроводника.

Концентрации носителей заряда в чистом полупроводнике.

Рассмотрим идеальный полупроводник с единственным максимумом в валентной зоне и единственным минимумом в зоне проводимости. Будем также считать, что эффективные массы, описывающие эти экстремумы, изотропны.

При $T=0$ в зоне проводимости электронов нет, в валентной зоне все состояния заняты и наш кристалл будет демонстрировать чисто диэлектрические свойства.

При конечной температуре возникнет некоторое количество электронов и дырок, причём в силу общей электронейтральности их концентрации совпадают $n_e = n_h$. Эти термоактивированные носители заряда называют также *собственными* носителями заряда, а о таком полупроводнике говорят как о полупроводнике с собственной проводимостью.

Электроны и дырки являются фермионами, для них применимо распределение Ферми-

Дирака. Поэтому концентрации электронов и дырок равны $n_{e,h} = 2 \int n(\epsilon_{e,h}) \frac{d^3 \vec{q}_{e,h}}{(2\pi)^3}$, здесь

множитель 2 — спиновый, функция распределения $n(\epsilon_{e,h}) = \frac{1}{\exp((\epsilon_{e,h} \mp \mu)/T) + 1}$, где μ — химический потенциал для «настоящих» электронов¹¹, знак химпотенциала для дырок изменяется в соответствии со сформулированными выше соглашениями. Как правило, для полупроводников химпотенциал оказывается достаточно далеко от краёв запрещённой зоны (точнее, как правило, расстояние от уровня химпотенциала до дна зоны проводимости и

¹¹ Иногда в полупроводнике уровень химического потенциала называют уровнем Ферми (например [1]). Строго говоря, термин химический потенциал является более корректным, так как поверхности Ферми в полупроводнике нет.

потолка валентной зоны оказывается много больше температуры), поэтому газ электронов и дырок оказывается невырожден и можно пренебречь единицей в распределении Ферми-Дирака.

Тогда для концентрации электронов:

$$n_e = 2 \times \frac{1}{2\pi^2} \int_0^\infty \exp\left(-\left(E_g + \frac{\hbar^2 q^2}{2m_e} - \mu\right)/T\right) q^2 dq = \frac{1}{\pi^2} e^{-(E_g - \mu)/T} \left(\frac{\sqrt{2m_e T}}{\hbar}\right)^3 \int_0^\infty e^{-z^2} z^2 dz,$$

здесь множитель 2 связан со спином, интеграл $\int_0^\infty e^{-z^2} z^2 dz = \frac{1}{2} \int_0^\infty e^{-x} \sqrt{x} dx = \frac{\sqrt{\pi}}{4}$, откуда

$$n_e = 2 \left(\frac{m_e T}{2\pi \hbar^2}\right)^{3/2} e^{-(E_g - \mu)/T}.$$

Для дырок совершенно аналогичными вычислениями с учётом сделанного ранее замечания о фермиевской функции распределения для дырок получаем:

$$n_h = 2 \left(\frac{m_h T}{2\pi \hbar^2}\right)^{3/2} e^{-\mu/T}.$$

Множители $Q_{e,h} = 2 \left(\frac{m_{e,h} T}{2\pi \hbar^2}\right)^{3/2}$ называют статистическими факторами зон, они имеют смысл эффективного числа энергетических уровней в зоне проводимости и в валентной зоне. Подстановкой численных значений можно получить удобное для вычислений выражение

$$Q = 2.51 \cdot 10^{19} \left(\frac{m}{m_0} \times \frac{T[K]}{300 K}\right)^{3/2} 1/\text{см}^3, \text{ где } m_0 \text{ — масса свободного электрона.}$$

Отметим для последующего изложения, что мы пока никак не использовали равенство концентраций электронов и дырок в чистом полупроводнике, поэтому эти вычисления концентрации *подвижных* электронов и дырок можно будет применять и для примесных полупроводников – от наличия примесного уровня должен будет как-то меняться химпотенциал полупроводника. Также отметим, что при наличии нескольких электронных или дырочных карманов в зонной структуре полупроводника нужно просуммировать такие вклады от каждого кармана. Причём видно, что при наличии электронов или дырок разной массы, при одинаковой энергии активации наиболее многочисленными окажутся более тяжёлые носители (экспоненциальные множители одинаковые, а $Q \propto m^{3/2}$).

Вернёмся теперь к нашей простой модели. В силу электронейтральности $n_e = n_h$ (электроны и дырки рождаются парами), поэтому для равновесной концентрации $n_0 = n_e = n_h$ имеем, перемножив полученные равенства и возвращая в запись постоянную

Больцмана для совместимости с литературой, $n_0^2 = 4 (m_e m_h)^{3/2} \left(\frac{k_B T}{2\pi \hbar^2}\right)^3 e^{-E_g/(k_B T)}$, то есть

главным множителем в температурной зависимости концентрации оказывается $n_0 \propto e^{-E_g/(2k_B T)}$. Обратите внимание на множитель $1/2$ в показателе экспоненты, который может казаться несколько контринтуитивным: в термоактивационном законе оказывается не щель между валентной зоной и зоной проводимости, а её половина. Это, однако совершенно логично: в показателе больцмановской экспоненты стоит не разница между дном и потолком соседних зон, а разница между энергией частицы и уровнем химпотенциала.

Если бы имелось несколько эквивалентных дырочных или электронных экстремумов в

законе дисперсии, то мы бы получили аналогичный результат с точностью до некоторого численного множителя (вычисленные нами статфакторы зон нужно было бы умножить на количество экстремумов в соответствующей зоне), а зависимость от температуры $n \propto T^{3/2} e^{-E_g/(2T)}$ не изменилась бы.

Химпотенциал чистого полупроводника.

Приравняв явно концентрации электронов и дырок, получаем:

$$m_e^{3/2} e^{-(E_g - \mu)/T} = m_h^{3/2} e^{-\mu/T}$$

$$\left(\frac{m_e}{m_h}\right)^{3/2} = e^{(E_g - 2\mu)/T},$$

$$\mu = \frac{E_g}{2} - \frac{3}{4} k_B T \ln \frac{m_e}{m_h}$$

в последнем уравнении мы вернули в запись постоянную Больцмана для совместимости с литературой.

Таким образом, в чистом полупроводнике при $T=0$ уровень химического потенциала находится точно посередине запрещённой зоны, а при ненулевых температурах он смещается к той зоне, в которой масса носителей заряда меньше.¹² В чистом полупроводнике химпотенциал может отличаться от $E_g/2$ на величину порядка $k_B T$. С типичной шириной запрещённой зоны полупроводника в доли электронвольта это приводит при комнатной температуре к изменению уровня химического потенциала в пределах 10%. Это оправдывает сделанное выше предположение о возможности пренебречь единицей в фермиевском распределении. Однако это не является общим правилом — в узкозонном полупроводнике с сильно отличающимися массами электронов и дырок при конечной температуре уровень химпотенциала может даже попасть в валентную зону или зону проводимости.

Примесные слаболегированные полупроводники.

Донорные и акцепторные примеси в полупроводнике.

Практически важным вопросом является изменение свойств полупроводника при введении в него небольшого (критерий малости рассмотрим далее) количества примесных атомов, имеющих большее (донорные примеси) или меньшее (акцепторные примеси) количество валентных электронов. Для определённости будем говорить про кремний, в котором каждый атом связан со своими соседями четырьмя ковалентными связями (структура алмаза). Характерными концентрациями примесей, при которых применимы наши рассуждения

¹² Если бы имелось несколько эквивалентных экстремумов в зонной структуре полупроводника, то, как легко заметить, при $T=0$ ответ не изменится, а при конечных температурах возникнет дополнительная

линейная по температуре поправка $-\frac{1}{2} k_B T \ln \frac{N_e}{N_h}$, где N_e и N_h — число электронных и

дырочных экстремумов. При наличии носителей с несколькими массами (но с одинаковой энергией в экстремуме $\varepsilon(\vec{k})$) также ответ качественно не изменится, линейная по температуре поправка будет иметь

вид $-\frac{1}{2} k_B T \ln \left\{ \frac{\sum N_e^{(i)} (m_e^{(i)})^{3/2}}{\sum N_h^{(i)} (m_h^{(i)})^{3/2}} \right\}$.

являются концентрации порядка $10^{14} \dots 10^{17} \text{ 1/см}^3$. Снизу этот диапазон ограничен технологией (концентрация примесей 10^{14} 1/см^3 соответствует чистоте материала на уровне 10^{-10}), а при концентрациях, больших чем примерно 10^{19} 1/см^3 , образуются так называемые примесные зоны: электроны примесных атомов начинают туннелировать между примесями. Если такая примесная зона сольётся с валентной зоной или зоной проводимости, то, строго говоря, полупроводник становится полуметаллом [3]. В нашем изложении мы будем предполагать, что образования примесной зоны не произошло.

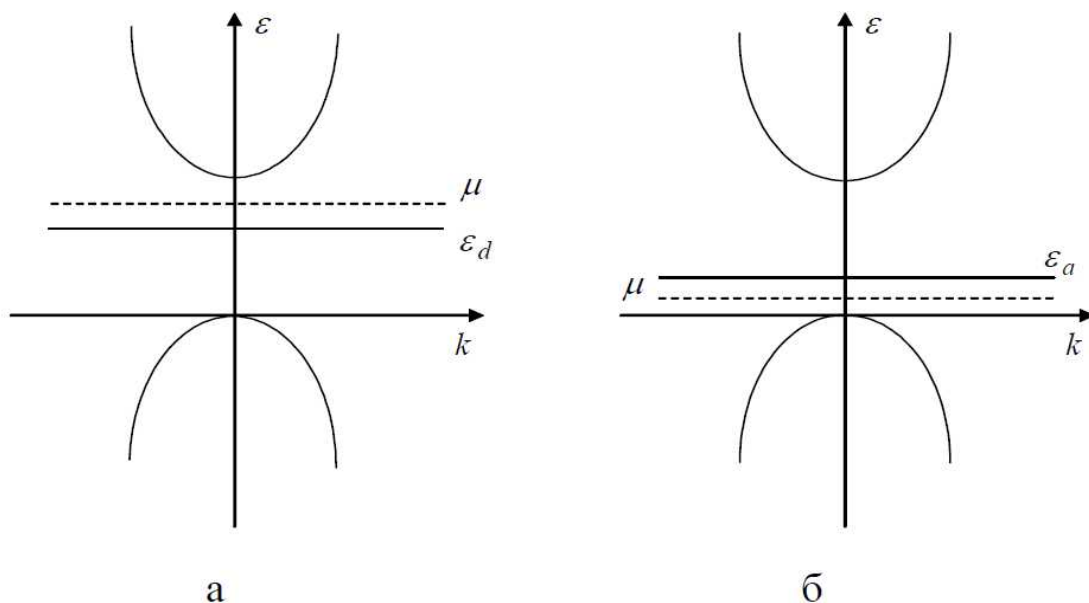


Рисунок 13: Типичное положение примесного уровня (сплошная прямая, $\epsilon_{d,a}$) и уровня химпотенциала при $T=0$ (пунктирная линия, μ) в слаболегированном полупроводнике для донорной (а) и акцепторной (б) примеси. Из книги [2].

Если один из атомов кремния заместить пятивалентным мышьяком, то четыре валентных электрона образуют ковалентные связи с соседними атомами кремния, а один избыточный электрон останется при $T=0$ на атоме примеси. Если энергия ионизации этого электрона окажется достаточно низкой, то этот избыточный электрон может перейти в зону проводимости (валентная зона занята полностью), оставив за собой полностью ионизованный атом примеси с избыточным положительным зарядом.

Если один из атомов кремния заменить трёхвалентным индием, то, наоборот, электронов примесного атома не хватает на устройство полного набора ковалентных связей в структуре алмаза. У одного из соседних атомов кремния на химической связи в направлении атома индия окажется неспаренный электрон, то есть на этой связи появится вакансия для электрона. На это вакантное место может попасть один из электронов из основной матрицы. При этом в матрице оказывается нехватка одного электрона — образуется вакансия, которая может перемещаться по кристаллу. Это можно описать как ионизацию связанного состояния «атом индия — дырка». Соответственно, после отрыва дырки от примесного атома остаётся отрицательно заряженный ион примеси.

Если концентрация примесных атомов мала, то в качестве простейшей модели можно охарактеризовать примесные атомы положением одного единственного уровня энергии ионизации примеси (примесного уровня) относительно валентной зоны и зоны проводимости.¹³ Это положение есть функция материала.

¹³ Необходимо подчеркнуть, что это упрощённое описание. Как мы увидим далее, электрон или дырка вблизи примесного иона могут иметь целую структуру водородоподобных уровней. Представление об

Для теоретического определения положения примесного уровня необходимо решать уравнение Шредингера с учётом взаимодействия электронов примесного атома и с «родительским» атомом, и с атомами матрицы полупроводника. Ответ будет зависеть от деталей связи электронов с родительским атомом и от свойств полупроводника. Строгое решение этой задачи выходит за рамки курса, один из частных случаев для мелких примесных уровней будет рассмотрен далее.

Рассмотрим феноменологически последствия возможного попадания примесного уровня в различные области зонной структуры полупроводника:

- Если примесной донорный уровень окажется выше дна зоны проводимости, то избыточным электронам с примеси будет энергетически выгодно перейти в зону проводимости основной матрицы, все примесные ионы ионизируются уже при $T=0$, и мы получим металлическое поведение нашей системы с маленькой (в меру концентрации примесей) концентрацией свободных носителей электронного типа. Аналогично, если примесной акцепторный уровень окажется ниже потолка валентной зоны (если образование связи «примесь-атом матрицы» окажется выгоднее образования связи в матрице), то части электронов матрицы оказывается выгодно уже при $T=0$ перейти на примесь, часть состояний в валентной зоне оказывается вакантна и, опять, получим металл с маленькой концентрацией свободных носителей заряда дырочного типа. В обоих этих случаях при нулевой температуре химпотенциал совпадёт с энергией Ферми образовавшихся свободных носителей заряда.
- Если примесной донорный уровень окажется ниже потолка валентной зоны или примесной акцепторный уровень окажется выше дна зоны проводимости, то ничего существенного по сравнению с чистым полупроводником не произойдёт: для термодинамических свойств такой системы по-прежнему главным (наиболее вероятным) будет переход электронов между потолком валентной зоны и дном зоны проводимости, так как разница энергий между этими состояниями будет минимальна.
- Наиболее интересным (в том числе и с точки зрения приложений) является случай, когда примесной уровень попадает в запрещённую зону. При этом стараются подобрать примеси так, чтобы донорный уровень оказался близко к дну зоны проводимости¹⁴, а акцепторный — к потолку валентной зоны (рисунок 13). Такие *неглубокие* примесные уровни не являются единственно возможными, но, как мы увидим далее (стр. 31), у них есть наглядная физическая интерпретация в виде водородоподобного состояния электрона или дырки в окрестности ионизованной примеси.

Мы можем, опираясь на предыдущие результаты, сразу указать положение химпотенциала в легированном полупроводнике с единственным неглубоким примесным уровнем при $T=0$ и в высокотемпературном пределе: при $T=0$ уровень химпотенциала будет ровно посередине между примесным уровнем и дном зоны проводимости для донорной примеси и ровно посередине между примесным уровнем и потолком валентной зоны для акцепторной примеси, а при высоких температурах уровень химпотенциала сместится к середине запрещённой зоны к положению, характерному для чистого полупроводника.¹⁵ Характерной

единственным уровне позволяет однако существенно упростить вычисления и получить качественное описание многих явлений.

14 В кремнии расстояние от примесного уровня донорной примеси (P, As, Sb) до дна зоны проводимости равно 0.04...0.05 эВ при ширине запрещённой зоны 1.12 эВ, в германии для тех же донорных примесей около 0.01 эВ при ширине запрещённой зоны 0.66 эВ.[1]

15 Это рассуждение не совсем общее — оно неявно предполагает, что имеется по одному экстремуму в валентной зоне и зоне проводимости и эффективные массы электронов и дырок равны. При таких условиях, действительно, высокотемпературный химпотенциал окажется посередине запрещённой зоны. В случае

температурой перехода от одного режима к другому является температура, при которой количество собственных носителей сравняется с количеством носителей, внесённых легированными примесями. Из этого рассуждения в частности следует, что для номинально чистого полупроводника (в котором в реальности всегда есть некоторое количество структурных или примесных дефектов) при $T=0$ химпотенциал окажется не посередине запрещённой зоны, а направление смещения химпотенциала в номинально чистом полупроводнике будет при низких температурах определяться не массами носителей, а типом случайно имеющихся дефектов. По этой, чисто технологической причине, кристаллы номинально чистого полупроводника оказываются слегка легированы центрами n или p-типа, для отработанных технологий тип этого «невольного» легирования известен.

Отметим также, что в одном и том же полупроводнике могут одновременно присутствовать и несколько видов примеси: доноры и акцепторы, доноры разной валентности или с разной глубиной уровней. Мы будем для простоты рассматривать случай наличия единственного типа примесей. Более сложные случаи рассматриваются аналогично: заселённости всех примесных уровней и зон определяются единственной характеристикой — химпотенциалом, в сочетании с условием электронейтральности это позволяет построить полную систему уравнений.

Концентрация носителей в легированных полупроводниках. Закон действующих масс (правило рычага).

Как легко заметить, ионизация донорных примесей приводит к появлению в полупроводнике избытка делокализованных электронов над дырками (такие полупроводники называют полупроводниками n-типа), а захват электронов акцепторами к появлению избытка подвижных дырок над электронами¹⁶ (такие полупроводники называют полупроводниками p-типа). Доминирующие носители называют *основными* носителями заряда, а второй тип носителей — *неосновными*.

Условие электронейтральности для легированного полупроводника превращается в $n_e + N_a^{(ion)} = n_h + N_d^{(ion)}$, где $N_{a,d}^{(ion)}$ — концентрация ионизованных акцепторных и донорных примесей. Ионизованные примеси не участвуют в переносе заряда, поэтому для практики важно определение концентрации электронов и дырок в зоне проводимости и в валентной зоне.

При получении концентрации собственных носителей заряда ранее мы получили выражение, в которое не входит химпотенциал. Следовательно, это выражение можно применять и для легированного полупроводника, просто теперь нельзя записать $n_0 = n_e = n_h$ и необходимо прямо выписывать произведение концентраций:

$$n_e n_h = 4 \left(m_e m_h \right)^{3/2} \left(\frac{k_B T}{2 \pi \hbar^2} \right)^3 e^{-E_g / (k_B T)} \quad (\text{напомним, что мы вычислили это в модели с}$$

единственными экстремумами в валентной зоне и зоне проводимости). Это выражение и есть *закон действующих масс*, иногда называемый правилом рычага.

Отметим, что произведение концентраций электронов и дырок в легированном полупроводнике в точности равно этому произведению в собственном полупроводнике. Это

сильно отличающихся эффективных масс (см. например задачу 4.40 из задачника [10]) при высоких температурах химпотенциал может сильно сместиться и даже попасть в одну из зон. Аналогично, переход от низкотемпературного к высокотемпературному пределу определяется свойствами конкретного полупроводника и может оказаться нетривиальным.

¹⁶ Захваченный акцепторной примесью электрон локализован, а оставшаяся в валентной зоне вакансия — делокализована.

означает, что, например, введение донорной примеси не только увеличивает количество электронов в зоне проводимости, но и *во столько же раз* уменьшает количество дырок. Утверждение может показаться контринтуитивным, но на самом деле оно совершенно логично — при появлении вакансий в валентной зоне (т.е. при появлении дырок) электронам с донорного уровня энергетически выгодно перейти в валентную зону, заполнив вакантное место (то есть, уменьшив число дырок), а не только занимать состояния в зоне проводимости с более высокой энергией.

Найдём равновесные концентрации носителей и положение химпотенциала μ , для определённости, полупроводнике с примесями донорного типа. Пусть N_d — концентрация донорных примесей, а ϵ_d — положение примесного уровня. Вероятность того, что электронное состояние на донорном уровне занято есть $\frac{1}{e^{(\epsilon_d - \mu)/T} + 1}$. Соответственно

концентрация ионизованных примесей $N_d^{(ion)} = N_d \left(1 - \frac{1}{e^{(\epsilon_d - \mu)/T} + 1} \right) = \frac{N_d}{1 + e^{-(\epsilon_d - \mu)/T}}$.

Выраженные через химпотенциал концентрации электронов и дырок остаются теми же (мы считаем газ электронов и дырок невырожденным). Тогда, подстановкой в уравнение электронейтральности получаем, используя для компактности понятие статфактора зон, уравнение на химпотенциал:

$$Q_e e^{-(E_g - \mu)/T} = \frac{N_d}{1 + e^{-(\epsilon_d - \mu)/T}} + Q_h e^{-\mu/T}.$$

В этом уравнении от температуры зависят химпотенциал μ и статфакторы зон $Q_{e,h}$, его точное решение невозможно. Поэтому мы рассмотрим предельные случаи высоких и низких температур.

При низких температурах $T \rightarrow 0$ можно пренебречь в этом равенстве концентрацией дырок (так как при $T=0$ $\mu = (E_g + \epsilon_d)/2$, то $\mu \approx E_g \gg (E_g - \mu)$) и единицей в выражении для концентрации ионизованных примесей ($(\mu - \epsilon_d)/T > 0$ и велико в этом пределе). Фактически, мы пользуемся тем, что при низкой температуре основная часть электронов в зону проводимости придёт с ионизованных донорных примесей, так как энергетический уровень примесей гораздо ближе к дну зоны проводимости. Тогда получаем

$$Q_e e^{-(E_g - \mu)/T} = N_d e^{-(\mu - \epsilon_d)/T}$$

и логарифмированием получаем, аналогично рассмотренной задаче о чистом полупроводнике

$$\mu = \frac{E_g + \epsilon_d}{2} - \frac{T}{2} \ln \frac{Q_e}{N_d}.$$

Для концентрации электронов получаем

$$n_e = Q_e e^{-(E_g - \mu)/(2T)} \sqrt{\frac{N_d}{Q_e}} = \sqrt{Q_e N_d} e^{-(E_g - \epsilon_d)/(2T)}.$$

Обратите внимание, что учёт температурной зависимости химпотенциала привёл к существенной перенормировке предэкспоненциального множителя. При 300K и для массы электрона равной массе свободного электрона $Q_e = 2.5 \cdot 10^{19} \text{ 1/см}^3 \gg N_d \approx 10^{15} \dots 10^{16} \text{ 1/см}^3$, то есть предэкспонента оказывается в десятки раз меньше, чем для чистого полупроводника. Это однако с лихвой компенсируется показателем экспоненты, в котором теперь вместо «большой» энергетической щели стоит «маленькое» расстояние от примесного уровня до ближайшей зоны.

Для концентрации неосновных носителей заряда (дырок в этом примере), пользуясь тем, что $n_e n_h = Q_e Q_h e^{-E_g/T}$, получаем

$$n_h = Q_h \sqrt{\frac{Q_e}{N_d}} e^{-(E_g + \epsilon_d)/(2T)}.$$

С учётом близости донорного уровня к дну зоны проводимости $E_g \approx \epsilon_d$ и показатель экспоненты по модулю оказывается почти вдвое больше, чем для чистого полупроводника. В результате, несмотря на заметное изменение предэкспоненты по сравнению с чистым полупроводником, концентрация дырок (неосновных носителей заряда) окажется меньше в полупроводнике n-типа.

Завершая этот анализ, найдём отношение концентраций $\frac{n_e}{n_h} = \frac{N_d}{Q_h} e^{\epsilon_d/T}$.

Аналогично может быть получен результат для химпотенциала и концентрации дырок в полупроводнике p-типа [2], приведём здесь только ответы:

$\mu = \frac{\epsilon_a}{2} + \frac{k_B T}{2} \ln \frac{Q_h}{N_a}$ и $n_h = \sqrt{Q_h N_a} e^{-\epsilon_a/(2T)}$, где N_a и ϵ_a — концентрация атомов-акцепторов и энергия акцепторного уровня.

Рассмотрим теперь высокотемпературный предел, когда доминируют «собственные» электроны полупроводника, перешедшие из валентной зоны в результате термоактивации. Будем считать, что химпотенциал при этом окажется близок к середине запрещённой зоны (но температура всё ещё много меньше ширины запрещённой зоны), что предполагает отсутствие сильного различия статфакторов валентной зоны и зоны проводимости. Тогда в уравнении электронейтральности для донорного уровня экспонента в знаменателе оказывается маленькой и все примеси будут полностью ионизованы. Для концентрации носителей тогда $n_e = n_h + N_d$ и поправка к концентрации носителей может быть легко найдена при помощи правила рычага. С линейной точностью¹⁷ концентрация электронов возрастёт на $N_d/2$, а число дырок — уменьшится на $N_d/2$ по сравнению с чистым полупроводником.

Для химпотенциала вычисляем по уравнению электронейтральности:

$$Q_e e^{-(E_g - \mu)/T} = N_d + Q_h e^{-\mu/T} = Q_h e^{-\mu/T} \left(1 + \frac{N_d}{Q_h e^{-\mu/T}} \right).$$

Логарифмированием (и используя то, что термоактивированных электронов и дырок уже много больше, чем донорных примесей $N_d \ll Q_h e^{-\mu/T}$) получаем:

$$\ln Q_e - \frac{E_g}{T} + \frac{\mu}{T} \approx \ln Q_h - \frac{\mu}{T} + \frac{N_d}{Q_h e^{-\mu/T}}$$

$$\mu \approx \left[\frac{E_g}{2} + \frac{T}{2} \ln \frac{Q_h}{Q_e} \right] + \frac{T}{2} \frac{N_d}{Q_h} e^{E_g/(2T)} = \mu_0(T) + \frac{T}{2} \frac{N_d}{Q_h} e^{E_g/(2T)}, \text{ где } \mu_0(T) \text{ — температурная}$$

зависимость химпотенциала в чистом полупроводнике. Напомним, что статфактор зоны также зависит от температуры $Q = 2.51 \cdot 10^{19} \left(\frac{m}{m_0} \times \frac{T[K]}{300 K} \right)^{3/2} 1/\text{см}^3$. То есть, действительно,

¹⁷ Считаем, что концентрация электронов $n_e = n_0 + N_d/2$, дырок $n_h = n_0 - N_d/2$. Тогда $n_e n_h = n_0^2 + Q(N_d^2)$ и уравнение электронейтральности также выполнено.

при высокой температуре химпотенциал приближается к своему положению в середине запрещённой зоны, характерному для чистого полупроводника.

Для перехода между этими пределами отметим, что низкотемпературный предел характеризуется условием (для полупроводника n-типа) $\frac{n_e}{n_h} \gg 1$, то есть $N_d \gg Q_h e^{-\epsilon_d/T}$.

В высокотемпературном пределе мы считаем концентрацию примесей малой по сравнению с концентрацией термоактивированных электронов из валентной зоны, то есть (не учитывая отличие химпотенциала от середины зоны в этом пределе) $N_d \ll \sqrt{Q_e Q_h} e^{-E_g/(2T)}$. Так как

$\epsilon_d \approx E_g$, эти условия не противоречивы, но есть некоторая область температур, не подпадающая ни под один предел. Более подробный разбор вопроса о концентрации носителей в легированном полупроводнике, включающий этот случай «средних» температур проведён в задаче 4.10 задачника [10]. Для оценки, зависимости характерных для этих условий концентраций дефектов от температуры (для простоты положено $m_e = m_h = m_0$,

$Q = Q_e = Q_h$, взяты параметры ширины запрещённой зоны для кремния) построены на рисунке 14. Видно, что практически для любой разумной концентрации примесей при комнатной температуре мы окажемся в пределе низких температур, когда свойства носителей заряда определяются легированием — это позволяет применять кремниевую полупроводниковую электронику при комнатных температурах. В то же время, видно, что при не слишком высокой температуре можно выйти из этого предела — это, в частности, объясняет, почему характеристики полупроводниковых приборов так чувствительны к температуре.

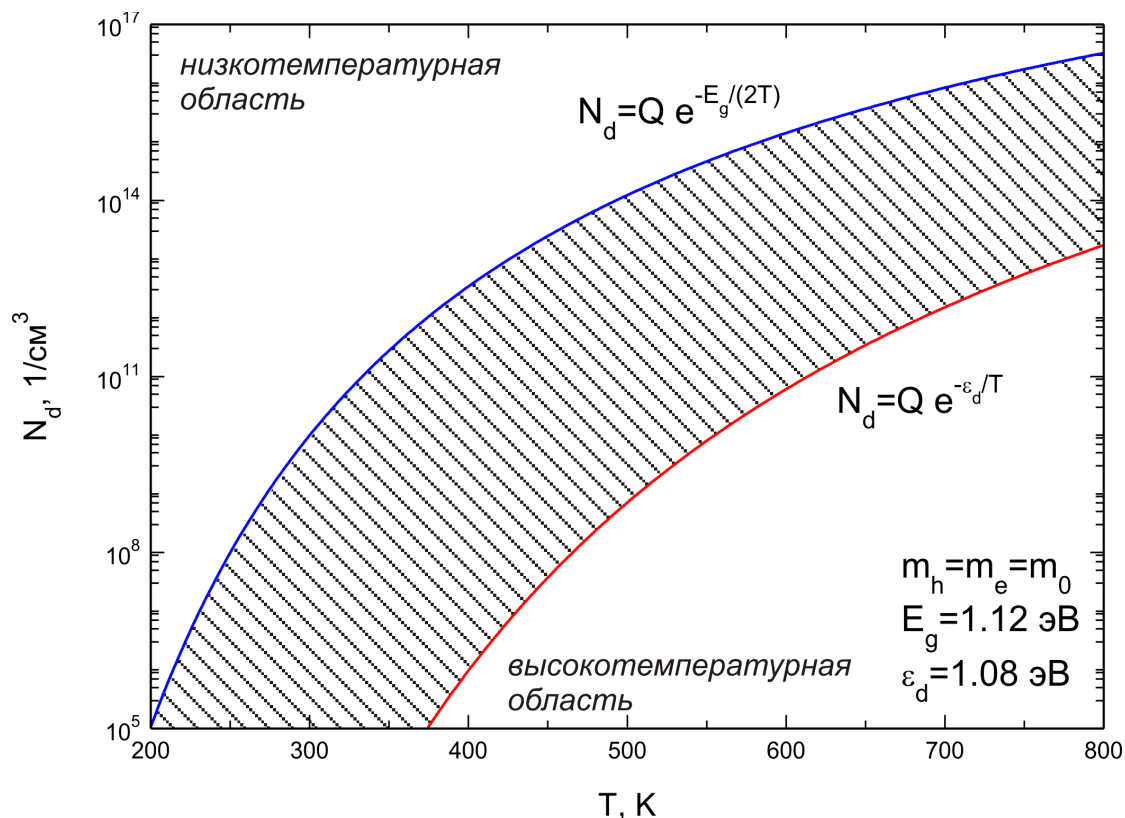


Рисунок 14: Оценка границ применимости низкотемпературного (красная кривая) и высокотемпературного (синяя кривая) приближений для легированного полупроводника. Вычисления сделаны для параметров ширины запрещённой зоны кремния. Заштрихована область «средних» температур, в которой ни один из пределов не применим.

Энергетические диаграммы для полупроводника.

Для многих практических задач и задач оценки оказываются не важны точные детали расположения зон в k -пространстве: прямозонный или не прямозонный полупроводник, количество электронных или дырочных «карманов» и т.д., а важна только энергетическая щель, положение в ней уровня химпотенциала и примесных уровней. Эту информацию удобно изображать в виде упрощённой энергетической диаграммы (рисунок 15).

В этом представлении показывается положение потолка валентной зоны и дна зоны проводимости (ширины этих зон в большинстве случаев показываются условно). Расстояние между этими зонами есть ширина запрещённой зоны, при необходимости указывается положение примесных уровней и химпотенциала. В некоторых задачах важно также положение минимального уровня энергии электрона в вакууме относительно указанных параметров зон. Расстояние от дна зоны проводимости до минимального уровня энергии электрона в вакууме называют *сродством к электрону* χ . Сродство к электрону есть работа, которую нужно затратить, чтобы вывести электрон со дна зоны проводимости в вакуум, то есть работа выхода. Отличие между минимальным уровнем энергии электрона в вакууме и уровнем химпотенциала в полупроводнике называют *термодинамической работой выхода*, которую необходимо отличать от сродства к электрону.

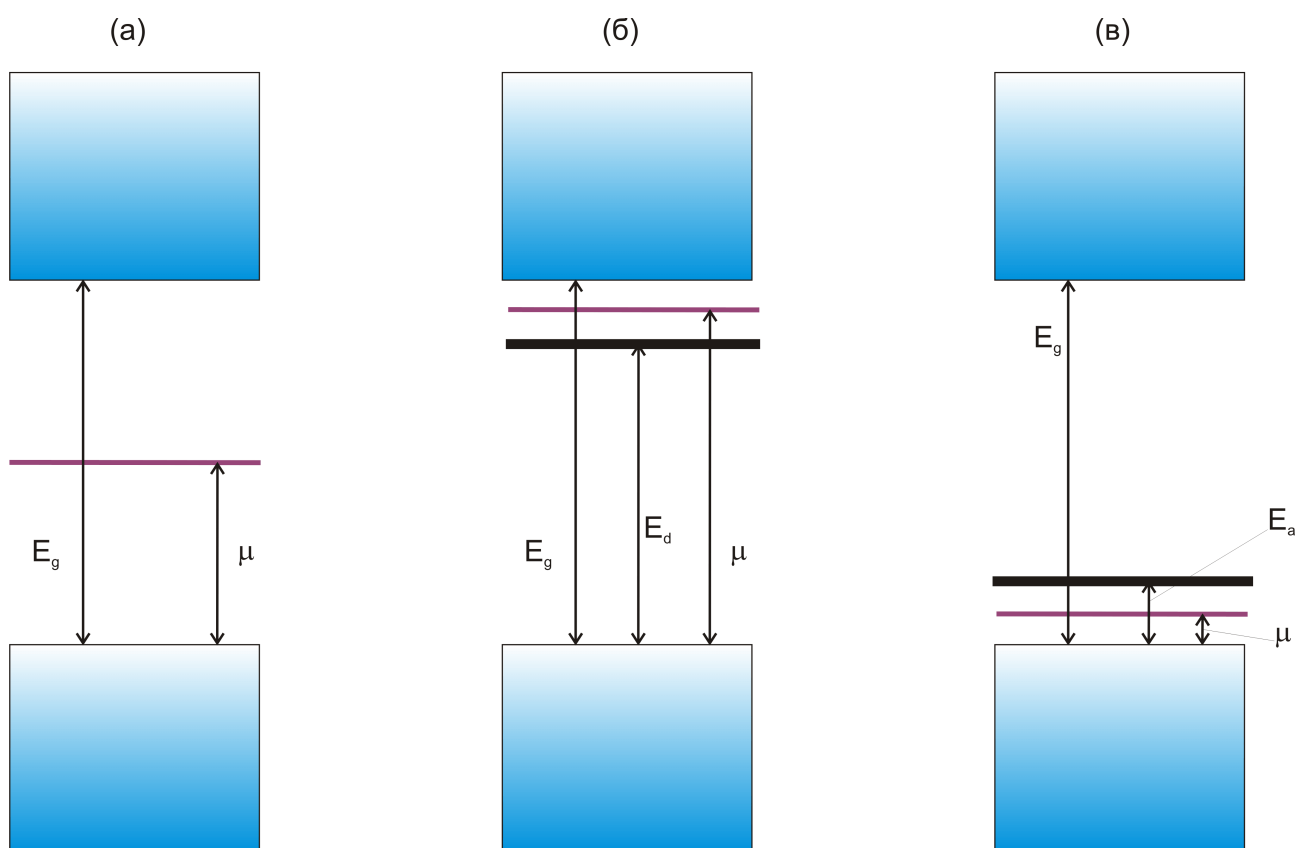


Рисунок 15: Упрощённое изображение зонной схемы полупроводника: (а) чистый полупроводник, (б) полупроводник с примесью донорного типа, (в) полупроводник с примесью акцепторного типа. E_g — ширина запрещённой зоны, μ — уровень химпотенциала, $E_{d,a}$ — уровни донорной и акцепторной примеси. Положение химпотенциала показано для случая $T=0$. Положение минимального уровня энергии электрона в вакууме не показано.

Электропроводность полупроводников.

Влияние концентрации носителей на электропроводность.

Электропроводность полупроводника имеет некоторую схожесть с электропроводностью металлов — движение электронов в зоне проводимости и дырок в валентной зоне описывается уравнениями типа второго закона Ньютона (для квазиимпульса), поэтому с учётом процессов рассеяния мы получим в общем схожую картину об установлении дрейфовой скорости движения носителей заряда. Так как свободные носители заряда в полупроводнике сосредоточены вблизи экстремумов зон, они все описываются одной эффективной массой и поэтому здесь могут быть прямо применены рассуждения простой классической модели Друде-Лоренца.

Некоторая сложность связана с тем, что в полупроводнике имеются одновременно различные носители заряда (электроны и дырки) и, иногда, носители с несколькими массами. В принципе, это может приводить к возникновению дополнительных механизмов релаксации, связанных с взаимодействием этих заряженных квазичастиц (рекомбинации электрона и дырки, превращения лёгкой дырки в тяжёлую и наоборот). Мы будем пренебрегать такими процессами, что допустимо, если концентрация носителей мала.

Тогда фактически имеются несколько независимых систем носителей и для проводимости мы можем обобщить формулу Друде-Лоренца: $\sigma = \sum_i \frac{n_i e^2 \tau_i}{m_i}$, где суммирование идёт по разным видам носителей. Обратим внимание, что так как заряд носителя в проводимость входит в квадрате, то вклады электронов и дырок в проводимость суммируются.

Видно, что основной вклад в температурную зависимость проводимости даст концентрация носителей, экспоненциально нарастающая с ростом температуры. Для чистого полупроводника основной зависимостью будет $\sigma \propto e^{-E_g/(2T)}$, для легированного полупроводника при низких температурах $\sigma \propto e^{-\delta E/(2T)}$, где δE — расстояние от примесного уровня до ближайшей зоны ($\delta E = E_g - \epsilon_d$ для донорной примеси и $\delta E = \epsilon_a$ для акцепторной). Важно иметь в виду, что номинально чистый полупроводник является некоторой абстракцией — даже в отсутствие целенаправленного легирования в реальном кристалле будут иметься дефекты и случайные примеси.

Пример зависимости сопротивления полупроводникового термометра сопротивления на основе германия от температуры показан на рисунке 16. В зависимости $R(T)$ есть два участка, на которых $R(T) \propto e^{\Delta/T}$, но энергия активации на этих участках отличается на порядок. Причём обе этих области связаны с какими-то примесными уровнями: ширина запрещённой зоны для германия $E_g = 0.66 \text{ эВ} \approx 7600 \text{ К}$.

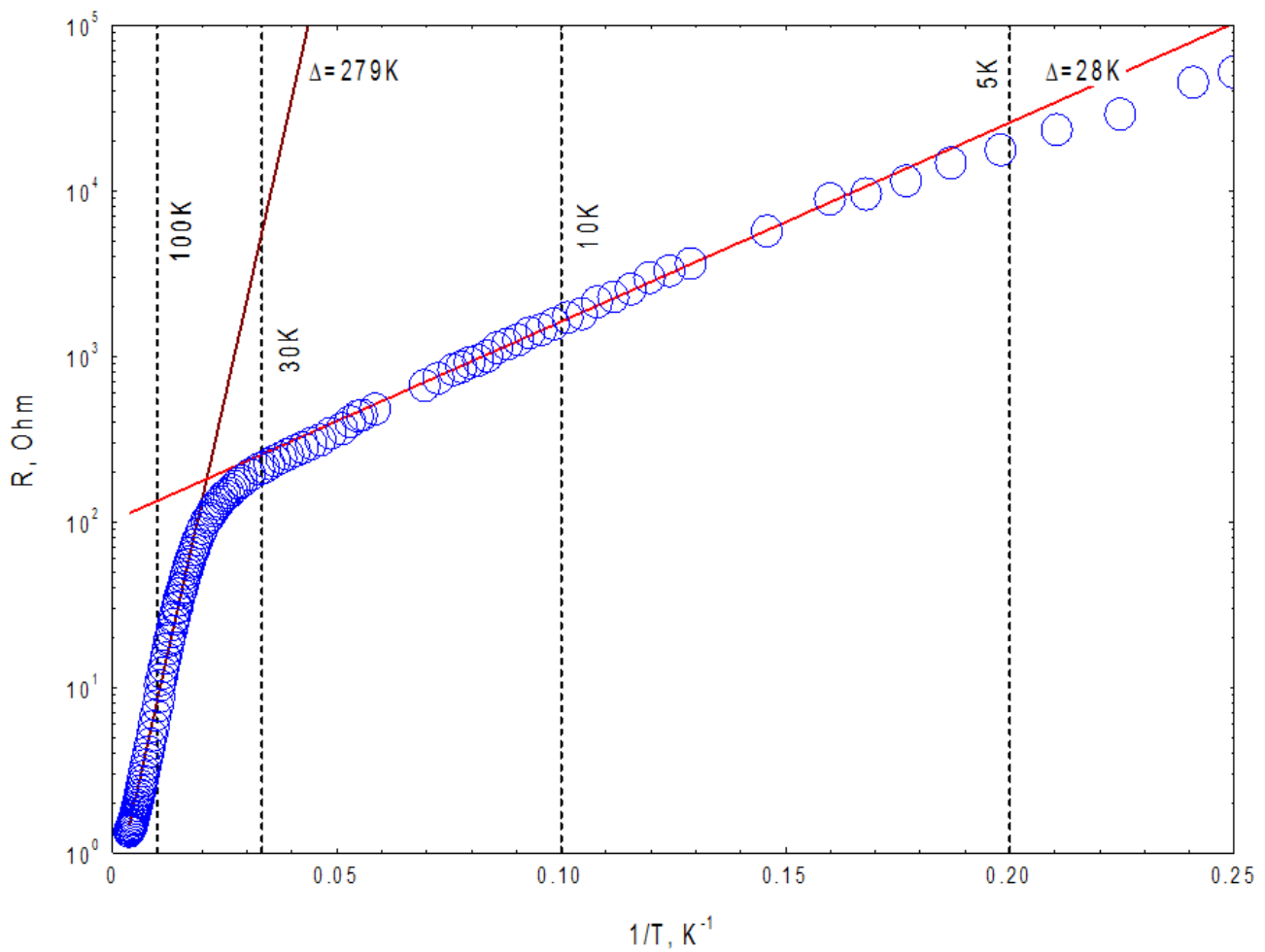


Рисунок 16 Зависимость сопротивления полупроводникового термометра сопротивления от температуры от комнатной температуры до 4.2K. Данные автора.

Подвижность зарядов в полупроводнике†

Из-за существенно изменяющейся концентрации носителей, в полупроводнике, в отличие от металла, проводимость является не самой удобной характеристикой для описания динамики движения заряда. Для этого обычно пользуются подвижностью, которую определяют (для изотропного случая) как отношение модуля средней дрейфовой скорости носителя заряда к напряжённости электрического поля: $|\vec{V}_{др}| = \mu |\vec{E}|$. По историческим причинам это обозначение несколько отличается от понятия подвижности в других разделах физики, где оно определяется как отношение дрейфовой скорости к силе. Иногда [3] при определении подвижности рассматривают не модули, а проекции скорости и напряжённости, тогда для электронов подвижность будет отрицательна, для дырок — положительна.

Подвижность связана с временем пробега (сравните с формулой Друде-Лоренца) $\mu = \frac{|e|\tau}{m^*}$.

Для малых концентраций носителей подвижность определяется рассеянием на фононах и рассеянием на примесях. Рассеяние носителей друг на друге является маловероятным в силу из малой концентрации. Подвижность является функцией температуры, характерные величины подвижности электронов и дырок в кремнии при комнатной температуре для разных уровней легирования (из [11]) показаны на рисунке 17.

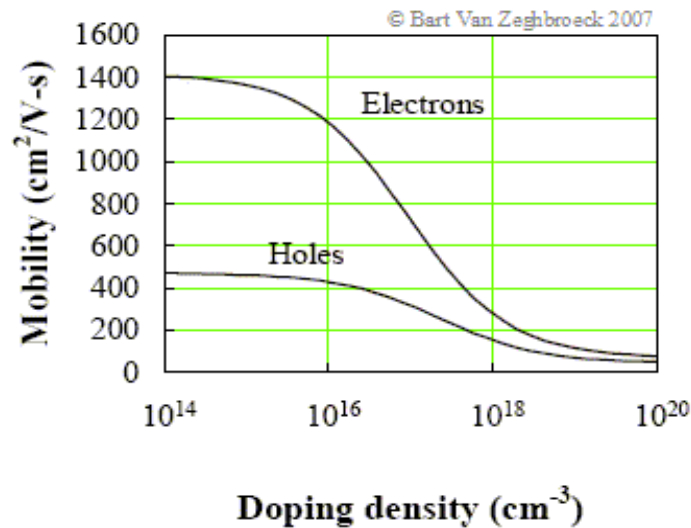


Рисунок 17: Подвижности электронов и дырок в кремнии при комнатной температуре при разных концентрациях примесей. Из книги [11].

Таблица 2: Подвижность носителей в различных полупроводниках при комнатной температуре. По книге [1].

Кристалл	Подвижность, $\text{см}^2/(\text{В}\cdot\text{сек})$		Кристалл	Подвижность, $\text{см}^2/(\text{В}\cdot\text{сек})$	
	электроны	дырки		электроны	дырки
алмаз	1800	1200	GaSb	4000	1400
Si	1300	500	PbS	550	600
Ge	4500	3500	PbSe	1020	930
InSb	77000	750	PbTe	1620	750
InAs	33000	460	AgCl	50	нет данных
InP	4600	150	KBr (100 K)	100	нет данных

Экспериментальные данные по подвижности в некоторых кристаллах полупроводников при комнатной температуре показаны в таблице 2. Для сравнения, подвижность электронов в меди при комнатной температуре $35 \text{ см}^2/(\text{В}\cdot\text{сек})$, для высококачественных образцов хороших металлов отношение RRR может достигать нескольких тысяч, что позволяет достичь в металлах подвижности $\sim 10^4 \text{ см}^2/(\text{В}\cdot\text{сек})$. В кристаллах PbTe при $T=4 \text{ K}$ подвижность достигает величины $5 \cdot 10^6 \text{ см}^2/(\text{В}\cdot\text{сек})$ [1]. Мы не будем в нашем курсе подробно обсуждать вопросы температурной зависимости подвижности в полупроводнике, некоторые сведения по этому вопросу могут быть найдены в [2].

Связанные состояния и рекомбинация носителей в полупроводниках.

Экранирование ионизованной примеси в полупроводнике.

В металле из-за наличия большого числа свободных электронов электрическое поле примесного иона экранируется на расстоянии порядка межатомного. При рассмотрении состояния носителя заряда вблизи ионизованного центра в полупроводнике также необходимо знать, как зависит электрическое поле кулоновского центра от расстояния. В отличие от металла, для полупроводников мы имеем дело с невырожденным газом квазичастиц. Эта задача по существу является известной задачей о дебаевском экранировании в плазме. Приведём здесь решение в упрощённом виде, считая, что мы имеем дело с примесной (для определённости — донорной) проводимостью полупроводника.

Уравнение для потенциала электрического поля $\varepsilon \Delta \varphi = 4\pi e(n(\vec{r}) - n_0)$, где n_0 — равновесная концентрация (обеспечивающая электронейтральность), а ε — диэлектрическая проницаемость среды (число порядка 10 для типичных полупроводников), константу e считаем положительной¹⁸. В присутствии медленно меняющегося электрического поля (мы считаем, что на электрон действует локальное электрическое поле, что требует малости его дебройлевской длины волны¹⁹ по сравнению с характерными длинами изменения электрического поля) условием термодинамического равновесия является постоянство электрохимического потенциала $\mu(\vec{r}) - e\varphi(\vec{r}) = \text{const}$, поэтому для химпотенциала в окрестности примеси имеем $\mu(\vec{r}) = \mu_0 + e\varphi(\vec{r})$, где μ_0 — положение химпотенциала вдали от примеси. Тогда концентрация невырожденного электронного газа в присутствии электрического поля определяется уравнением:

$$n(\vec{r}) = \exp\left(-\frac{(E - \mu(r))}{T}\right) = \exp\left(-\frac{(E - \mu_0 - e\varphi(\vec{r}))}{T}\right) = n_0 \exp\left(\frac{e\varphi(\vec{r})}{T}\right) \approx n_0 + \frac{en_0}{T} \varphi(\vec{r}), \text{ где } E$$

— кинетическая энергия электронов, константу e считаем положительной, что определяет знак потенциальной энергии $-e\varphi(\vec{r})$. Концентрация электронов увеличивается в области положительного потенциала, куда они притягиваются.

Тогда уравнение Пуассона для потенциала превращается в $\Delta \varphi = \frac{4\pi e^2 n_0}{\varepsilon T} \varphi$. Решением

этого уравнения будет, экранированный кулоновский потенциал $\varphi = \frac{Ze}{\varepsilon r} e^{-\lambda r}$, где

$$\lambda = \sqrt{\frac{4\pi n_0 e^2}{\varepsilon k_B T}} \quad \text{— обратная длина экранирования.}^{20} \text{ Радиус экранирования (дебаевский}$$

радиус) $r_D = \frac{1}{\lambda}$ в полупроводнике связан с концентрацией носителей и при их низких

18 В общем случае $\text{div } \vec{D} = 4\pi \rho$ и $-\varepsilon \Delta \varphi = 4\pi \rho$

19 Характерный волновой вектор, оцениваемый из $k_B T \simeq \frac{\hbar^2 k_0^2}{2m}$, при 300K для $m = m_0$ составляет

$\sim 10^7 \text{ 1/см}$, то есть дебройлевская длина волны при низких температурах заведомо будет много больше межатомного расстояния.

20 Напомним, что для *вырожденного* электронного газа в металле длина экранирования не зависела от

температуры: $\frac{1}{\lambda} = \sqrt{\frac{\varepsilon_F}{6\pi n_0 e^2}} = \frac{(3\pi^2)^{1/3}}{2(3\pi)^{1/2}} \frac{\hbar}{e \sqrt{m n_0^{1/3}}} \approx \frac{1}{2} \sqrt{\frac{a_0}{n_0^{1/3}}}$.

концентрациях растёт (в отсутствие носителей экранировка пропадает).

Поэтому (в особенности при низких температурах) можно считать дебаевский радиус большим, по сравнению с межатомным расстоянием.²¹ Это означает, что в ближайшей окрестности примесного иона экранирование можно не учитывать. Эти выводы останутся в силе и при рассмотрении газа из квазичастиц нескольких типов: дебаевский радиус при

условии общей электронейтральности равен $r_D^{-1} = \sqrt{\sum_i \frac{4\pi q_i^2 n_i}{k_B T}}$ (суммирование идёт по типам носителей заряда).

Водородоподобные состояния вблизи примеси.

Рассмотрим поведение носителя заряда в полупроводнике вблизи от противоположно заряженной примеси. При низких температурах можно считать дебаевский радиус экранирования сколь угодно большим. Тогда взаимодействие носителя заряда с примесью можно описать, как взаимодействие с кулоновским центром.

Принципиальным для дальнейшего решения оказывается вопрос о соотношении между дебройлевской длиной волны рассматриваемого электрона и межатомным расстоянием в кристалле. Для строгого ответа необходимо решать уравнение Шредингера с учётом свойств примесного атома и взаимодействия электрона с кристаллом. Мы воспользуемся тем, что из опыта известно, что у некоторых примесей энергии связи электрона оказываются маленькими и, соответственно, размер локализации электрона оказывается большим.

В этом случае можно считать, что электрон движется в сплошной среде, и нужно учесть диэлектрическую проницаемость среды, заменив заряд центра Ze на $\frac{Ze}{\epsilon}$. В этом же пределе динамика движения электрона в кристалле на большом удалении от примеси должна описываться эффективной массой.

Отсюда, используя результат для атома водорода, сразу получаем для уровней энергии:

$$E_n = -\frac{Z^2}{\epsilon^2} \frac{m^*}{m_0} R_y \frac{1}{n^2}, \text{ где уровни отсчитываются от энергии делокализованного носителя (то}$$

есть от края зоны проводимости для электронов и от потолка валентной зоны для дырок) вглубь запрещённой зоны. Видно, что для типичного для полупроводника значения $\epsilon \simeq 10$ и $m^* \simeq 0.1 m_0$, масштаб энергий оказывается $\sim 0.01 \text{ эВ} \sim 100 \text{ К}$. Это как раз масштаб энергий для тепловой ионизации при комнатной температуре. Имеются такие примеси с «почти свободным» носителем заряда (который уже не привязан жёстко к примесному атому, но ещё не может далеко уйти) позволяют создавать обычные полупроводниковые приборы, работающие при комнатной температуре.

Характерное расстояние до дефекта (радиус боровской орбиты) аналогично пересчитывается из известного результата: $r_0 = \frac{\epsilon}{Z} \frac{m_0}{m^*} a_0$, и для тех же параметров получим $r_0 \simeq 50 \text{ \AA}$.

Отметим, что если расстояние между примесными центрами окажется меньше r_0 , то эти связанные с примесями носители заряда делокализуются, аналогично делокализации частиц в задаче о сильной связи. В том случае образуется примесная зона и полупроводник приобретает металлические свойства.

²¹ Вычисленная для концентрации 10^{15} 1/см^3 , $\epsilon = 10$ и $T = 300 \text{ К}$ обратная длина экранирования $\lambda \sim 10^5 \text{ 1/см}$.

Экситоны Ванье-Мотта.

Связанное состояние могут образовывать и разноимённые квазичастицы. Это связанное состояние аналогично, например, связанному состоянию электрона и позитрона — позитронию, о котором упоминалось в курсе атомной физики. Эти состояния не привязаны к дефектам, поэтому могут распространяться по кристаллу. Их появление можно описать как появление в полупроводнике новых *нейтральных* квазичастиц — экситонов.

Если характерное расстояние между электроном и дыркой в экситоне много больше межатомного, то такой экситон называют экситоном Ванье-Мотта. Если расстояние между электроном и дыркой в экситоне оказывается порядка межатомного, то такую систему называют экситоном Френкеля. Мы ограничимся только рассмотрением экситонов Ванье-Мотта. Для экситона Ванье-Мотта при рассмотрении энергии взаимодействия электрона и дырки можно использовать приближение сплошной среды и зонные эффективные массы. Энергия связи экситона тогда легко определяется перенормировкой формул для атома

водорода: $E_n = -\frac{1}{\epsilon^2} \frac{m_e m_h}{m_0(m_e + m_h)} Ry \frac{1}{n^2}$, где $Ry = 13.6 \text{ эВ}$.

Отсчитанная от потолка валентной зоны энергия экситона:

$$E = E_g - \frac{1}{\epsilon^2} \frac{m_e m_h}{m_0(m_e + m_h)} Ry \frac{1}{n^2} + \frac{\hbar^2}{2(m_e + m_h)} k^2, \text{ где } k \text{ — волновой вектор экситона. Важно}$$

отметить, что для образования экситона требуется несколько меньшая энергия, чем для образования свободной электронно-дырочной пары. Поэтому образование экситонов можно наблюдать в оптическом эксперименте как появление узких линий оптического поглощения при энергии фотона чуть меньшей (на величину энергии связи), чем ширина запрещённой зоны. Соответствующие частоты поглощения определяются очевидным условием

$$\hbar \omega = E_g - \frac{1}{\epsilon^2} \frac{m_e m_h}{m_0(m_e + m_h)} Ry \frac{1}{n^2} : \text{ так как для оптических фотонов длина волны много}$$

больше межатомного расстояния, их волновой вектор мал по сравнению с бриллюэновским и можно считать, что при поглощении фотона $\Delta k = 0$ и возбуждаются только покоящиеся экситоны.

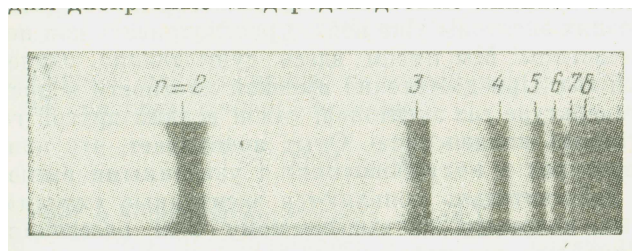


Рисунок 18: Спектр пропускания полупроводника Cu_2O на границе жёлтого края поглощения. Температура 1.3 К. По книге [4].

Экситоны были впервые наблюдаены в 50-х годах 20 века в полупроводниковой закиси меди Cu_2O советским физиком Е.Ф.Гроссом и его сотрудниками (см., например, учебник [4]). Пример фотографии спектра пропускания из оригинальной работы показан на рисунке 18. Пока энергия светового кванта меньше ширины запрещённой зоны поглощения с передачей энергии электрону быть не может. Однако вблизи границы поглощения (когда энергия кванта сравнивается с шириной запрещённой зоны) наблюдался ряд полос поглощения, связанных с возбуждением экситонных состояний.²²

²² Строго говоря, в случае Cu_2O из-за особенностей зонной структуры (обсуждаемых, например, в приложениях к статье [12]) возбуждаются экситоны в состоянии с орбитальным моментом $l=1$ (Р-экситоны). Это связано с дополнительными правилами отбора по чётности, возникающими при анализе

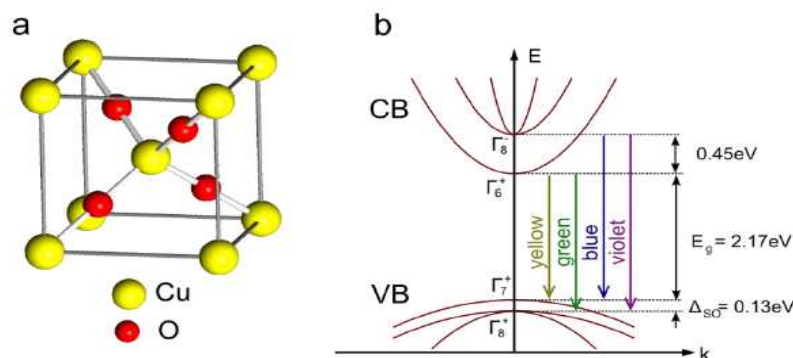


Рисунок 19: Структура и зонная схема закиси меди. Из статьи [12].

Этот эксперимент был недавно повторён с использованием современного оборудования [12]. С использованием перестраиваемого лазера в качестве источника света удалось наблюдать переходы в экситонные состояния с $n \leq 25$ (рисунок 20). Радиус экситона в состоянии с $n=25$ составляет около 1 мкм.

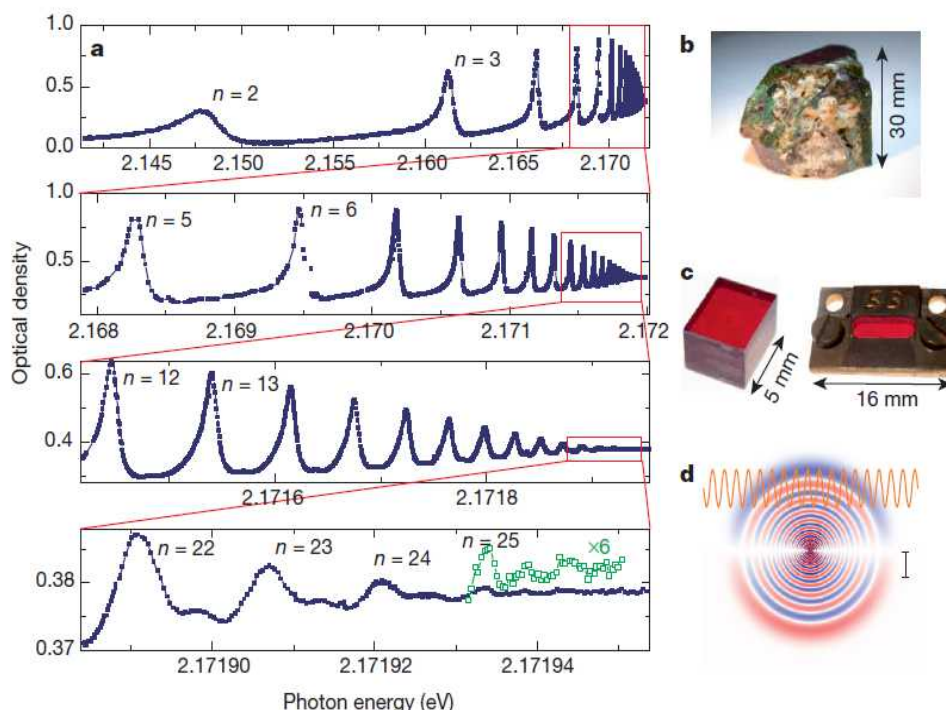


Рисунок 20: (a) Поглощение (оптическая плотность) в Cu_2O как функция энергии фотона. Температура 1.2 К. (b) Природный кристалл Cu_2O , из которого вырезались образцы для исследования. (c) смонтированные на приборные столики пластинки Cu_2O . Толщина образцов около 30 мкм. (d) Расчёт волновой функции экситона с $n=25$, для масштаба показана синусоида с длиной волны равной длине волны используемого для возбуждения света. Из статьи [12].

Зависимость энергии экситонного состояния (отсчитанной от дна зоны проводимости) от квантового числа n следует предсказанию простой водородоподобной модели $\propto \frac{1}{n^2}$ (рисунок 21).

поглощения света. Подробное рассмотрение выходит за рамки нашего курса. По этой причине оптический переход в состояние с $n=1$ не наблюдается (для $n=1$ есть только s-состояния).

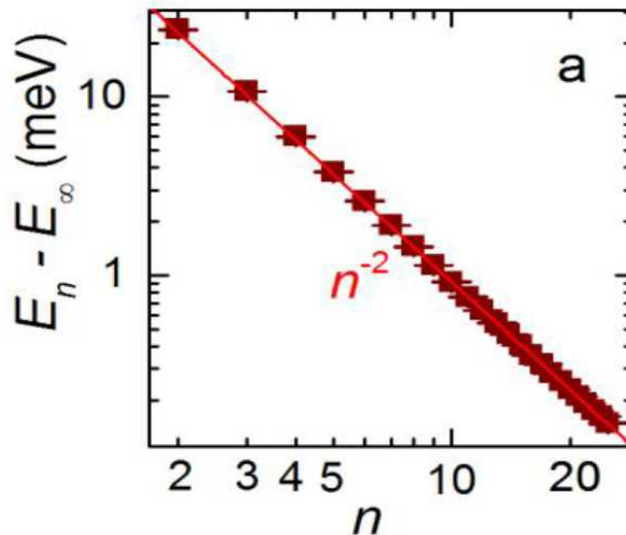


Рисунок 21: Зависимость энергии экситона от главного квантового числа. Из статьи [12].

Экситон-фононное взаимодействие[†]

Рассмотренный выше процесс рождения экситонов при поглощении кванта света наблюдался при низких температурах, то есть в отсутствие фононов. Как и в металле, электроны и дырки в полупроводнике могут взаимодействовать с фононами: рождать и уничтожать фононы.²³ Это изменяет «правила отбора» при рассмотрении процессов рождения экситонов.

В безфононном процессе, рассмотренном выше, рождается экситон с нулевым импульсом (так как волновой вектор фотона оптического диапазона много меньше бриллюэновского),

для чего требуется квант света с энергией $\hbar\omega = E_g - \frac{1}{\epsilon^2} \frac{m_e m_h}{m_0(m_e + m_h)} Ry \frac{1}{n^2}$. В процессе с

рождением или поглощением фонона энергия кванта света может отличаться от точной резонансной энергии, но и экситон теперь будет рождаться с ненулевым импульсом, равным импульсу фонона $\hbar\vec{K}$. Если (как будет в рассматриваемом далее примере) рождается или поглощается оптический фонон²⁴ в центре зоны с почти постоянной частотой Ω , то правило отбора по энергии принимает вид

$$E = E_g - \frac{1}{\epsilon^2} \frac{m_e m_h}{m_0(m_e + m_h)} Ry \frac{1}{n^2} \pm \hbar\Omega + \frac{\hbar^2}{2(m_e + m_h)} K^2 \quad (\text{рисунок 22}).$$

²³ Физическим механизмом взаимодействия остаётся, как и в металле, взаимодействие электрона с поляризацией решётки при распространении упругой волны. Однако, в отличие от находящегося в тепловом равновесии электронного газа в металле, рождение экситона при низких температурах есть процесс неравновесный (в электронную систему «закачивается» избыток энергии), поэтому процессы с рождением фонона возможны и при низких температурах. Процессы с поглощением фонона требуют наличия в кристалле «подходящих» фононов, поэтому они происходят только при достаточно высоких температурах.

²⁴ Напомним также, что именно в оптических фононных модах возникает наиболее сильная поляризация кристалла на малых волновых векторах, так как в этих модах разные ионы колеблются в противофазе.

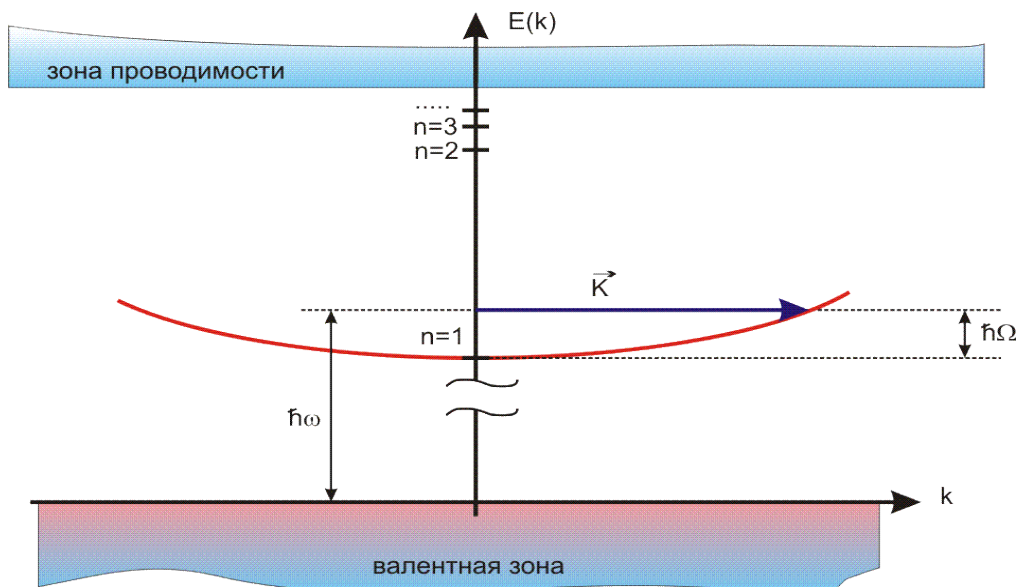


Рисунок 22: Схема процесса рождения экситона с рождением фонона.

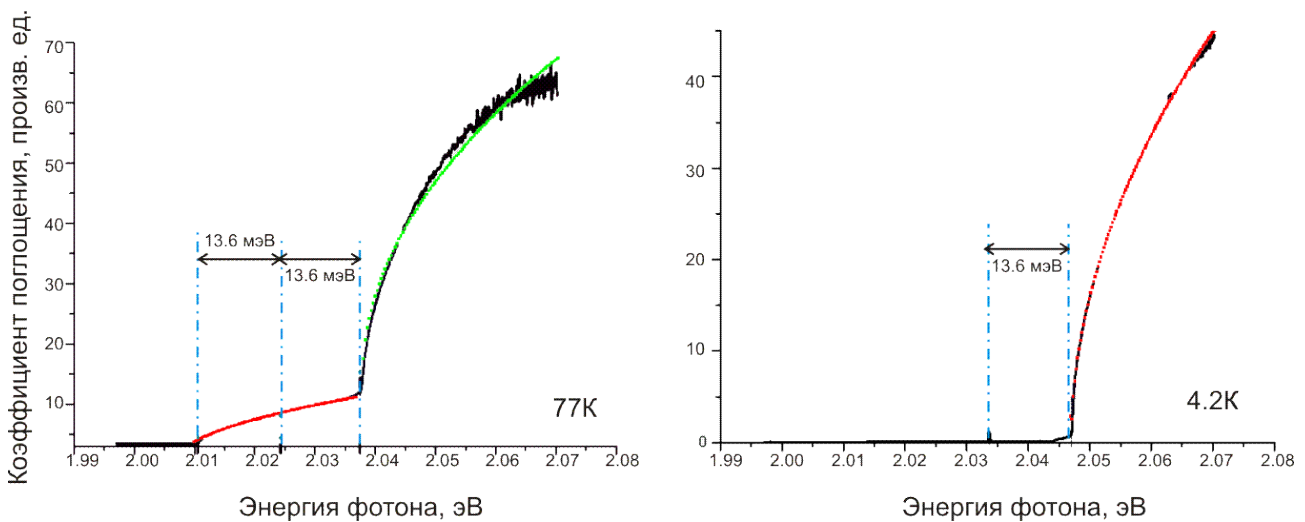


Рисунок 23: Коэффициент поглощения света в полупроводнике Cu_2O вблизи первой ($n=1$) экситонной линии поглощения. Слева: при температуре 77K. Справа: при температуре 4.2K. На основе рисунка из работы [13].

Рассмотрим пример экспериментального наблюдения таких процессов в полупроводнике Cu_2O [13]. Ширина запрещённой зоны Cu_2O (см. схему зон на рисунке 19) составляет 2.17 эВ, первый экситонный уровень находится на 150 мэВ ниже дна зоны проводимости, его энергия, отсчитанная от потолка валентной зоны приблизительно равна 2.02 эВ. Энергия оптической фононной моды равна 13.6 мэВ. Отметим также, что в силу особенностей формирования зонной структуры в Cu_2O рождение экситонов с $n=1$ в безфононном процессе запрещено правилами отбора (см. сноску на стр.32). В процессах с рождением и уничтожением фононов это правило отбора не действует — недостающий или избыточный момент импульса может быть передан всему кристаллу.

Пример экспериментальных данных по поглощению света в кристалле Cu_2O показан на рисунке 23. При энергии кванта света 2.033 эВ наблюдается слабый резонансный пик, связанный с безфононным процессом рождения экситона в состоянии с $n=1$. Этот процесс вероятно оказывается слабо разрешён какими-то дефектами кристалла. Вблизи этой энергии фотона (но всегда энергия фотона остаётся меньше ширины запрещённой зоны и меньше энергии второго экситонного уровня) наблюдаются непрерывные полосы поглощения в виде двух ступеней при $T=77\text{K}$ и одной ступени при $T=4.2\text{K}$. При этом

одна полоса начинается в точности на энергию оптического фонона *до* энергии безфононного процесса, а вторая — в точности на энергию оптического фонона *после* безфононного процесса. В соответствии с описанным выше, поглощение света при 77К в интервале энергий от 2.01 мэВ до 2.037 мэВ связано с процессами с поглощением фонона, и это поглощение предсказуемо пропадает при охлаждении до 4.2К, температуры много меньше энергии оптических фононов. Поглощение света при энергии кванта выше 2.037 мэВ связано в основном с процессами с излучением фонона, этот процесс остаётся возможным и при низкой температуре.

Рекомбинация носителей[†]

Как мы уже упомянули, в полупроводниках возможно поглощение фотонов с образованием свободных электронно-дырочных пар. Пороговая частота для такого процесса поглощения определяется шириной запрещённой зоны. При этом, в принципе, мы можем создать неравновесное распределение электронов и дырок (создать избыточное количество электронов и дырок, по сравнению с тепловым равновесием, и создать электронный и дырочный газы с неравновесным распределением по энергии). Простейшим способом будет облучение светом с частотой заметно выше пороговой — тогда будут появляться электроны и дырки в состояниях, которые в равновесии при этой температуре свободны от квазичастиц.

Во многих случаях можно считать, что процесс восстановления равновесного числа носителей медленнее, чем процесс восстановления равновесного распределения носителей по энергии. При комнатных температурах время установления равновесия в газе квазичастиц порядка 10^{-13} сек, а времена восстановления равновесного числа частиц меняются от 10^{-7} сек в прямозонных полупроводниках до секунд и даже часов в непрямозонных [2]. Таким образом, можно считать, что через небольшое время мы, независимо от способа воздействия, получим газы электронов и дырок с тепловыми энергиями (соответствующие квазичастицы соберутся вблизи минимумов своих энергетических спектров), но концентрация этих квазичастиц будет избыточна по сравнению с «настоящим равновесием» и пойдёт более медленный релаксационный процесс восстановления равновесной концентрации квазичастиц.

Процесс восстановления равновесного числа частиц заключается в уничтожении «лишних» электронно-дырочных пар. Этот процесс называют рекомбинацией, он похож на процесс аннигиляции частицы и античастицы в ядерной физике (но в отличие от электрон-позитронной аннигиляции в кристалле возможна рекомбинация электрона и дырки с испусканием единственного фотона, недостающий импульс можно передать кристаллу). При рекомбинации должны выполняться законы сохранения энергии и квазиимпульса. Рекомбинация может быть излучательной, когда вся энергия передаётся кванту света, и безызлучательной, когда энергия, например, передаётся другому носителю заряда (рекомбинация Оже). Закон сохранения квазиимпульса тривиально выполняется для прямозонных полупроводников (напомним, что для оптических фотонов с энергией ~ 1 эВ волновой вектор много меньше бриллюэновского, поэтому оптические переходы происходят почти без изменения волнового вектора состояния). Для непрямозонных полупроводников недостающий импульс передаётся фонону, этот процесс требует участия дефекта решётки (так называемый процесс Холла-Шокли-Рида).

Отметим также, что высвечивание фотонов при рекомбинации происходит с частотой $\omega = \omega_c \approx E_g / \hbar$, а генерация неравновесных носителей заряда производится путём облучения светом другой частоты $\omega > \omega_c$. Таким образом, мы имеем дело с фотолюминисценцией: несовпадением частот облучающего и переизлучаемого света.