



SEMICONDUCTORS

R. A. Smith

CBE, FRS, FRSE

**Principal Emeritus and Honorary Professor
Heriot-Watt University, Edinburgh**

Second Edition

**CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON — NEW YORK — MELBOURNE**

Р. СМИТ

ПОЛУПРОВОДНИКИ

Второе издание,
переработанное и дополненное

*Перевод с английского
под редакцией
д-ра физ.-мат. наук
Н. А. Пенина*

МОСКВА «МИР» 1982

ББК 22.379

С 50

УДК 539.2+537.33

Смит Р.

С 50 Полупроводники: Пер. с англ.—М.: Мир, 1982.—560 с., ил.

Учебное пособие известного английского физика Р. Смита, представляющее собой второе, переработанное и дополненное издание курса физики полупроводников (М.: ИЛ, 1962). С большим педагогическим мастерством и на высоком научном уровне излагаются основы современной теории электрических, гальваномагнитных, тепловых, оптических и контактных явлений в полупроводниках.

Для аспирантов и студентов университетов и технических вузов, а также для ученых и инженеров, занимающихся исследованием полупроводников и разработкой полупроводниковых приборов.

С $\frac{20403-064}{041(01)-82}$ 064-82, ч. 1 1704060000

ББК 22.379

С 50

Редакция литературы по физике

© Cambridge University Press, 1959, 1978

© Перевод на русский язык, «Мир», 1982

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Книга известного английского физика Р. Смита «Полупроводники», впервые изданная в русском переводе в 1962 г., приобрела в СССР широкую известность и долгое время была одним из лучших пособий по физике полупроводников. Сейчас она стала библиографической редкостью. В Англии книга выдержала четыре стереотипных издания. В 1978 г. вышло новое английское издание, которое выгодно отличается от прежнего. Книга подверглась существенной переработке. В ней были учтены современные достижения физики полупроводников.

При подготовке этого издания из книги были исключены две главы, посвященные применению полупроводников и методам измерения их параметров, поскольку по этим вопросам к тому времени уже были выпущены отдельные книги. Существенно сокращены и переработаны главы, в которых описываются свойства различных полупроводников: вместо двух глав этому вопросу в новом издании посвящена одна — гл. 13. В ней приведены самые общие сведения о наиболее изученных полупроводниках, в частности данные о структуре энергетических зон, об эффективных массах носителей заряда и их подвижностях.

Добавлены три новые главы: глава, в которой дан краткий обзор теоретических методов, используемых в зонной теории твердых тел (гл. 11), глава об аморфных полупроводниках (гл. 15) и глава, посвященная обзору новых явлений, открытых и изученных после выхода первого английского издания книги (гл. 14). В последней обсуждаются: явление конденсации экситонов и образования электронно-дырочных капель в полупроводниках, туннельный и лазерный эффекты в $p-n$ -переходах, лазерный эффект, в основе которого лежит рассеяние излучения с переворотом спина электронов в полупроводниках в магнитном поле.

В гл. 12 приведены сведения о влиянии сильных электрических и магнитных полей на процессы переноса носителей заряда и оптические свойства полупроводников.

Глава, посвященная оптическим и высокочастотным явлениям в полупроводниках (гл. 10), подверглась наибольшей переработке.

В нее включены новые сведения о таких явлениях, как двухфотонное, экситонное и решеточное поглощение, эффект увлечения носителей заряда фотонами, внутризонная фотопроводимость (μ -фотопроводимость), эффекты электропоглощения и электроотражения.

Разделы книги, посвященные процессам рассеяния и рекомбинации носителей заряда, в новом издании выделены в самостоятельные главы (соответственно гл. 8 и 9). Это отвечает тому значению, которое имеют процессы рассеяния и рекомбинации в физике полупроводников.

Основная часть текста книги существенным изменениям не подверглась. Это относится к таким наиболее устоявшимся разделам физики полупроводников, как представления об энергетических зонах и локальных уровнях в полупроводниках (гл. 1—3), статистика электронов и дырок (гл. 4), явления электронного переноса (гл. 5), тепловые явления (гл. 6), диффузия носителей заряда (гл. 7), процессы рассеяния и рекомбинации. Контактные явления рассматриваются в гл. 7. Всего в книге 15 глав. Объем ее несколько увеличен по сравнению с первым английским изданием.

Уровень изложения материала в книге остался прежним. Он довольно близок к уровню изложения физики полупроводников в высших учебных заведениях нашей страны.

Книга снабжена большим числом ссылок на оригинальные работы, что помогает в значительной мере восполнить краткость изложения, в особенности — новых вопросов физики полупроводников. Автор довольно часто дает ссылки на собственную книгу «Волновая механика кристаллических тел»^{*}). Это облегчило ему изложение материала в книге «Полупроводники». Поскольку, однако, «Волновая механика кристаллических тел» не переведена на русский язык, то следовать этим ссылкам затруднительно. Конечно, можно обратиться к другим руководствам, но обычно детальное изложение тех же вопросов в других книгах оказывается иным.

При переводе и редактировании были исправлены замеченные опечатки и погрешности в формулах и тексте. Там, где это казалось необходимым, сделаны подстрочные примечания. Библиографический список дополнен при переводе работами, имеющими в основном характер учебных пособий или монографий по отдельным разделам физики полупроводников.

Мы надеемся, что перевод второго издания книги Р. Смита на русский язык будет принят с удовлетворением как теми, кто уже изучал физику полупроводников (может быть, по старой книге Р. Смита), так и теми, кто ее изучает сейчас. Она может служить

^{*}) R. A. Smith. The Wave Mechanics of Crystalline Solids, 2nd Edition.— Chapman and Hall, 1969.

хорошим учебным пособием для студентов высших учебных заведений и аспирантов, специализирующихся по физике и технике полупроводников, а также полезным справочником для научных работников и инженеров.

Перевод переработанного и дополненного издания выполнен М. М. Бонч-Осмоловским (гл. 1—6 и 13—15) и В. А. Курбатовым (гл. 7—12).

Н. А. Пенин

ПРЕДИСЛОВИЕ

К моменту выхода в свет первого издания книги уже имелась обширная литература, посвященная полупроводникам, однако в основном это были статьи в научных и технических журналах. Книги же, которая давала бы обстоятельное изложение физических основ теории полупроводников, а также показывала роль полупроводников в технике, в то время явно не хватало. С тех пор объем литературы значительно возрос, причем не только в результате появления оригинальных статей в журналах, но также обзоров и книг, рассматривающих конкретные вопросы данной темы. Характерно, что именно в последнее время появилось несколько таких книг (как в форме монографий, так и в форме сборников статей), авторы которых являются ведущими специалистами в разных областях физики и техники полупроводников.

Может вызвать удивление тот факт, что книга «Полупроводники» так долго оставалась основным учебником для студентов старших курсов и аспирантов. Чтобы эта книга могла играть ту же роль и далее, ее, несомненно, следовало существенно переработать. Физические основы науки о полупроводниках изменений фактически не претерпели, тем не менее в последнее десятилетие в этой области появилось много нового. Например, замечательные исследования в области магнитооптики, связанные с магнитным квантованием, к моменту написания первого издания еще и не начинались (за исклю-

чением исследований циклотронного резонанса). Горячие электроны были известны и изучались, однако важность электронных переходных явлений осознана еще не была. Исследования, основанные на использовании очень сильных магнитных полей и очень высоких давлений, только начинались, а новые явления, основанные на взаимодействии света с различными возбуждениями, еще предстояло открыть. Не было еще и лазера, который играет такую важную роль в современной оптике, а важность исследования излучения света полупроводниками только начинала осознаваться. Хотя отдельные исследования аморфных полупроводников уже были выполнены, интерес к этой теме значительно вырос лишь в последние несколько лет.

Основную часть физических представлений, изложенных в первом издании, необходимо было сохранить; в то же время хотелось включить значительное количество нового материала. Однако новое издание по объему не должно было существенно превышать старое — таково было ограничение, поставленное издателями. Это очень затруднило отбор материала; только давление и уговоры со стороны коллег и друзей заставили меня взяться за это дело. Главным их аргументом было то, что такая книга, как «Полупроводники», все еще нужна, несмотря на изобилие литературы, посвященной данному вопросу вообще, и наличие нескольких научных трудов широкого профиля, сделанных на достаточно высоком уровне. Понятно, что изложение некоторых вопросов пришлось опустить. В основном это относится к главам, посвященным применениям, методам измерений и подробному перечислению физических параметров многих полупроводников; однако сейчас имеется много книг по электронике и оптическим применениям полупроводников, а также таблиц физических констант материалов, так что все эти сведения можно найти в соответствующей литературе. Кроме того, технические применения обычно не поспевают за научными открытиями. С опубликованием моей книги «Wave Mechanics of Crystalline Solids», написанной отчасти в качестве теоретического дополнения к «Полупроводникам», появилась возможность исключить из нового издания некоторые теоретические построения, дав ссылки на указанную книгу в соответствующих местах текста.

При подготовке нового издания мне очень помогли труды серии конференций по физике полупроводников. Конференции проводятся регулярно, каждые два года. Эти конференции можно назвать «золотым дном» научной информации. Они дают живую картину исторического развития физики полупроводников. Появились и периодические издания, такие, как «Progress in Semiconductors» (которое, к сожалению, прекратилось в 1967 г.), «Semiconductors and Semimetals» и «Solid State Physics». Значение их трудно переоценить, и мне хотелось бы выразить благодарность авторам обзорных работ. Большую помощь оказали мне переводы некоторых советских журналов, специально посвященных полупроводникам.

При столь обширной литературе возникла проблема библиографических ссылок. Полная библиография заняла бы слишком большую часть книги. Поэтому я давал ссылки только на некоторые ранние ключевые работы, а также на важнейшие более поздние работы и выбирал их так, чтобы читатель мог при желании почерпнуть из них значительно более полную библиографию по любой интересующей его теме. Я включил также краткую и очень неполную библиографию, указывающую на некоторые, на мой взгляд, наиболее полезные публикации.

В книге затрагивается большинство основных достижений в физике и технике полупроводников. Некоторые вопросы рассмотрены не так подробно, как мне хотелось бы, однако недостаток места не позволил дать более полное их изложение. Хотя почти во всей книге речь идет о кристаллических полупроводниках, последняя глава посвящена аморфным полупроводникам. Скорость появления публикаций по этой теме в настоящее время, может быть, даже превышает скорость появления публикаций, посвященных кристаллическим материалам. Поэтому, ничуть не надеясь дать адекватное изложение вопроса, я лишь попытался сопоставить основные идеи, характерные для этой быстро развивающейся области, с представлениями, которые позволили успешно описать кристаллическую форму полупроводников.

Я хочу выразить искреннюю благодарность за оказанную мне помощь по уточнению текста студентам, которым я преподавал, многим другим людям, которые мне писали, а также коллегам из

Массачусетского технологического института и позже из университета Хериот-Уотта, с которыми у меня состоялось много полезных обсуждений как по существу, так и по оформлению материала.

Наконец, мне хотелось бы поблагодарить сотрудников издательства Кембриджского университета за их постоянную помощь, в частности за ряд полезных предложений по уточнению текста.

Р. А. Смит

Эдинбург 1977

ИЗ ПРЕДИСЛОВИЯ АВТОРА К ПЕРВОМУ ИЗДАНИЮ

Поразительный рост за последние десять лет масштабов исследовательской работы, посвященной изучению полупроводников, привел к появлению огромного числа публикаций. Как это обычно бывает в быстроразвивающейся области науки, основную массу литературы составляют оригинальные статьи, опубликованные в научных и технических журналах; их дополняют обзорные статьи, в которых более детально рассмотрены отдельные аспекты проблемы. Значительное число прекрасных обзорных работ появилось за последние пять лет. Некоторые из них опубликованы в журналах, другие — в серии ежегодных сборников обзорных статей по более широкому кругу вопросов физики твердого тела, таких, как «Solid State Physics»¹⁾. Как минимум один том ежегодного издания «Progress in Semiconductors»²⁾ целиком посвящен обзорным статьям по полупроводникам.

В основу настоящей книги положен курс лекций по физике полупроводников, прочитанный автором весной 1955 г. на инженерном факультете Эдинбургского университета, куда он был приглашен в качестве внештатного лектора. Материал лекционного курса значительно расширен и дополнен современными данными, в связи с чем изменилось и его назначение. Если лекции в основном были рассчитаны на подготовку инженеров, то перед книгой ставится более широкая задача — она предназначается главным образом для физиков, поэтому автор стремился осветить основные физические свойства полупроводников. Автор надеется, однако, что его книга окажется полезной и для большого числа инженеров, занятых раз-

¹⁾ Под редакцией F. Seitz, D. Turnbull (Academic Press, New York, начиная с 1955 г.).

²⁾ Под редакцией A. Gibson, P. Aigrain, R. Burgess (Heywood, New York, начиная с 1956 г.).

работкой электронных устройств на базе полупроводников, так как она поможет им понять основные физические свойства используемых полупроводниковых материалов. Без такого понимания невозможно успешно применять эти вещества на практике, поскольку полупроводники, как правило, отнюдь не простые объекты и обладают рядом весьма тонких свойств.

В книге широко используется так называемый метод эффективной массы, тогда как более строгая квантовая теория движения электрона в кристалле изложена бегло и в элементарной форме. Различные виды энергетических спектров электронов в кристаллических телах также рассмотрены кратко и без привлечения строгой квантовомеханической трактовки вопроса. В настоящее время разработан целый ряд общих методов построения квантовой теории твердых тел, и автор предполагает в будущем опубликовать в качестве приложения к данному тому несколько более элементарную теорию энергетических спектров электронов в твердых телах¹⁾ В тексте по мере необходимости используется математический аппарат, однако всегда в наиболее простом виде.

Раздел книги, посвященный методам измерений, должен в основном ознакомить читателя с принципиальной стороной этого вопроса. Подробные сведения о деталях измерительной аппаратуры можно, как правило, почерпнуть из оригинальных статей, на которые в соответствующих местах книги даются ссылки. Автор не стремился составить исчерпывающий перечень таких статей, это было бы слишком трудной задачей. К счастью, обширная библиография имеется в обзорных работах, указанных в данной книге; кроме того, в ежегодном издании реферативного журнала *Semiconductor Abstracts*²⁾ приводятся аннотации статей, посвященных исследованию полупроводников. Представление о количестве этих работ можно составить на основании того факта, что выпуск журнала за 1957 г. содержит аннотации 1258 статей, вышедших в 1955 г. Наша цель заключалась лишь в том, чтобы обратить внимание читателя на одну-две наиболее важные статьи, в которых впервые освещено то или иное исследование, а также на несколько более поздних ключе-

¹⁾ Автор имеет в виду книгу *Wave Mechanics of Crystalline Solids* (Chapman and Hall, John Wiley and Sons, 1961). — *Прим. ред.*

²⁾ *Semiconductor Abstracts* (John Wiley and Sons, New York, начиная с 1955 г.).

вых работ. С помощью этих данных читатель, как правило, сможет составить себе более полный список работ по интересующему его вопросу.

Я благодарю моих коллег из Королевского института радиолокации за большую помощь при подготовке книги. В частности, много ценной информации по различным техническим вопросам и по форме их изложения я вынес из многочисленных бесед с д-рами А. Ф. Гибсоном, Е. Дж. С. Пэйджем, К. А. Хогартом, Е. Х. Патли. Я глубоко благодарен м-ру В. Е. Дж. Фарвису за то, что он подал мне мысль о написании этой книги и организовал чтение лекций, из которых в дальнейшем и выросла книга; д-ру Д. Шенбергу за полезную критику и за советы по способу изложения материала. Многие авторы любезно позволили использовать рисунки из их оригинальных статей; благодарности им приводятся в тексте. Я хотел бы выразить свою признательность также авторам обзорных статей, в особенности тех, которые перечислены в библиографии — без этих статей задача по написанию книги была бы неизмеримо сложнее.

Малверн, 23 июля 1958 г.

ПРОСТЕЙШИЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

1.1. Ранние исследования, посвященные полупроводникам

История изучения свойств веществ, которые в настоящее время относят к классу полупроводников, насчитывает более ста лет. Только на современном этапе развития науки о полупроводниках, при том понимании объекта исследования, которое достигнуто ныне, можно в полной мере оценить огромные трудности, стоявшие на пути многих ранних исследований. Чистота материалов, доступных в то время, была чрезвычайно низкой. Сейчас хорошо известно, что для получения надежных результатов при изучении свойств полупроводников необходимо использовать образцы из чрезвычайно чистого вещества. И если, несмотря ни на что, полупроводники все же были четко выделены в самостоятельный класс веществ и удалось определить их главные отличительные особенности задолго до того, как возникла теория, объясняющая эти свойства, то этим мы обязаны только высокому мастерству первых экспериментаторов. Неудивительно, что в таких условиях допускались и ошибки. Некоторые вещества, причисленные в то время к разряду полупроводников, оказались металлами, другие элементы, считавшиеся ранее металлами, в чистом виде оказались типичными полупроводниками.

Первым характерным признаком, отличающим проводники электричества этого класса от других плохих проводников и от металлов, является отрицательный температурный коэффициент сопротивления; иными словами, в отличие от металлов сопротивление полупроводников, как правило, падает с ростом температуры.

Фарадей [1] был, по-видимому, первым, кто отметил это обстоятельство при исследовании сернистого серебра. В настоящее время хорошо известна ограниченность этого критерия: в определенном температурном интервале сопротивление полупроводника может увеличиваться с ростом температуры, особенно при наличии в нем недостаточного количества примесей. Однако при дальнейшем повышении температуры всегда можно достичь такой точки, выше ко-

торой сопротивление с ростом температуры быстро падает. С другой стороны, некоторые металлические пленки, а иногда даже поликристаллические образцы металлов обнаруживают отрицательный температурный коэффициент электросопротивления. Теперь известно, что эти эффекты связаны с наличием оксидных пленок или существованием трещин между отдельными кристаллами, но в свое время они привели к тому, что металлы титан и цирконий были отнесены к числу полупроводников. Хотя возможности такого рода не исключаются, положение о том, что в чистых полупроводниках должен наблюдаться отрицательный температурный коэффициент сопротивления, все же остается справедливым. Наряду с этим имеется еще целый ряд других отличительных признаков полупроводников, которые были использованы для идентификации последних; их мы рассмотрим ниже. Необходимо, однако, признать, что строгий критерий для распознавания полупроводников удалось установить лишь после того, как квантовая теория твердого тела дала объяснение физическим причинам различных свойств, наблюдаемых на опыте.

За исключением области температур вблизи точки плавления, сопротивление полупроводников значительно больше сопротивления хороших металлических проводников тока и одновременно значительно меньше сопротивления хороших диэлектриков. Пределы изменения сопротивления в твердых телах поистине огромны. Хороший металлический проводник обладает при комнатной температуре удельным сопротивлением порядка 10^{-6} Ом·см, тогда как сопротивление полупроводников при комнатной температуре находится, как правило, в пределах от 10^{-3} до 10^6 Ом·см. Кроме того, имеется большое количество веществ, обладающих примерно такими же значениями сопротивления, но не являющихся полупроводниками в современном понимании этого слова. С другой стороны, хорошие диэлектрики обладают сопротивлением порядка 10^{12} Ом·см.

Успехи, достигнутые за последующие сорок лет после появления работы Фарадея, были незначительными. Удалось лишь выяснить, что некоторые вещества, принадлежащие к классу «плохих» проводников, например теллур, обладают исключительно большими значениями термо-э. д. с. Два важных открытия были сделаны лишь в 1873 и 1874 гг. Браун [2], исследуя такие вещества, как сернистый свинец и пирит, обнаружил явление выпрямления переменного тока; Смит [3] наблюдал фотопроводимость в селене. Вскоре и в других веществах были обнаружены некоторые из этих свойств, но не обязательно все сразу. В большинстве случаев этими веществами были сульфиды и окислы металлов, а также кремний.

Последовавшие за этим работы позволили выделить целый класс веществ, обладающих указанными свойствами,— класс полупроводников (по-немецки — Halbleiter). Обзор этих ранних работ, включающий также весьма обширную библиографию из 350 наиме-

нований, был составлен Ларк-Горовицем [4]. Этот период широко освещен также в появившихся еще раньше обзорах Гуддена [5], где подробно обсуждены проблемы идентификации полупроводников. Прекрасный обзор ранних работ русских авторов был сделан Иоффе [6].

К тому времени удалось установить следующие основные свойства полупроводниковых веществ:

- а) отрицательный температурный коэффициент сопротивления;
- б) значения удельного сопротивления, лежащие в пределах от 10^6 до 10^{-3} Ом·см;
- в) как правило, большие значения термо-э. д. с. (как положительные, так и отрицательные) относительно данного металла;
- г) эффект выпрямления переменного тока или, по крайней мере, неомическое поведение в контактах;
- д) чувствительность к свету (появление фото-э. д. с. или изменения сопротивления при освещении).

Незадолго до этого была выяснена важная роль, которую играют примеси в полупроводниках. Было отмечено, что некоторые характеристики различных образцов одного и того же вещества, в частности отрицательный температурный коэффициент сопротивления при высоких температурах, практически неизменны, тогда как другие характеристики сильно меняются от образца к образцу. Первые были названы «собственными» или «истинными» характеристиками. Этот факт весьма затруднял изучение «собственных» свойств полупроводников, так как никогда нельзя было быть полностью уверенным в том, что тот или иной наблюдаемый эффект не обусловлен наличием примесей в образце. С сомнениями такого рода ныне покончено, так как появилась возможность точно контролировать содержание примесей во многих полупроводниках.

Весьма важным событием, как оказалось впоследствии, явилось открытие в 1879 г. эффекта Холла [7], т. е. возникновение поперечной э. д. с. при прохождении электрического тока через проводник, помещенный в магнитное поле. Хотя к тому времени электрон еще не был открыт и поэтому не могла возникнуть мысль о числе носителей заряда, этот эффект оказался впоследствии тем ключом, который позволил раскрыть сущность процесса проводимости в полупроводниках и провести грань между полупроводниками и другими типами плохо проводящих веществ. Это обусловлено тем, что, как мы увидим позже, измерение э. д. с. Холла дает возможность непосредственно определить концентрацию и знак носителей заряда. Эти измерения позволяют также легко отделить случай ионной проводимости от случая электронной проводимости — весьма важное разграничение, так как электропроводность в случае ионной проводимости также быстро растет с ростом температуры, что может привести к ложным заключениям.

Различные методы разделения проводимости на электронную и ионную составляющие, а также трудности интерпретации в случаях, когда такое разделение невозможно, подробно обсуждены Ларк-Горовицем [4]. До использования эффекта Холла наличие или отсутствие ионной составляющей электрического тока устанавливалось в зависимости от того, имеется ли при этом электролитический перенос вещества или нет. Как мы увидим в дальнейшем, эффект Холла в случае ионной проводимости пренебрежимо мал по сравнению с эффектом, связанным с электронной проводимостью. Поэтому наличие эффекта Холла свидетельствует об электронном характере проводимости.

Два фактора обуславливают величину электропроводности вещества: концентрация имеющихся носителей заряда и легкость, с которой они движутся по кристаллу под действием приложенного электрического поля (подвижность носителей заряда). Более точное количественное определение подвижности будет дано в гл. 5 [разд. 1, формула (5.14)]. Здесь отметим лишь, что величина подвижности, как правило, убывает с ростом температуры, особенно при более высоких температурах. Этим объясняется падение электропроводности металлов с ростом температуры. Действительно, как показывают измерения эффекта Холла на образцах из металлической фольги, число носителей заряда в этом случае не зависит от температуры. В случае же полупроводника число носителей быстро увеличивается с ростом температуры, особенно в области более высоких температур, приводя к быстрому падению сопротивления. В этом и состоит основное отличие металла от *чистого* полупроводника. Мы подчеркиваем слово «чистого», так как слишком большое количество примесей в полупроводнике может замаскировать такое отличие. Металл — это проводник с практически постоянным числом носителей заряда, чистый полупроводник — проводник, в котором число носителей заряда быстро увеличивается с ростом температуры.

Первым, кто для изучения полупроводников систематически использовал эффект Холла, был, по-видимому, Бедкер [8], исследовавший соединение CuI . Подробные исследования большого числа веществ выполнены Кёнигсбергером [9], который воспользовался для этой цели как эффектом Холла, так и другими явлениями, упомянутыми выше. Было обнаружено, что число носителей заряда в полупроводниках намного меньше, чем в металлах, но подвижность их в полупроводниках обычно несколько больше. Одним из важных результатов этих работ было установление полупроводникового характера кремния, селена и теллура. Значительно позже к этому классу веществ был добавлен германий [10]. Кёнигсбергеру в полной мере удалось ощутить отмеченное выше различие между металлами и полупроводниками.

Весьма интересное наблюдение было сделано относительно зна-

ка носителей заряда. В некоторых случаях результаты опыта указывали на отрицательный знак носителей, как и следовало ожидать для случая электронов, но нередко он оказывался и положительным. Более того, при возрастании температуры знак носителей заряда в одном и том же веществе иногда менялся с положительного на отрицательный. Положительные значения коэффициента Холла (т. е. положительный знак носителей заряда) были обнаружены также у некоторых металлов. Все это привело в свое время к значительным затруднениям в области теории, которые были ликвидированы лишь с появлением квантовой механики.

1.1.1. «ИЗБЫТОЧНЫЕ» (ЭЛЕКТРОННЫЕ) И «ДЕФЕКТНЫЕ» (ДЫРОЧНЫЕ) ПОЛУПРОВОДНИКИ

Большая работа по исследованию множества веществ, отнесенных к разряду полупроводников, была выполнена между 1910 и 1930 гг. Однако при этом не было достигнуто сколько-нибудь значительных, фундаментальных результатов. К 1930 г. интерес к этим веществам сильно возрос, в основном благодаря применению полупроводников в технике. Для их изучения, как правило, использовались эффект Холла, электропроводность и термо-э. д. с. и было показано, что знак эффекта Холла при низких температурах почти всегда совпадает со знаком термо-э. д. с. Исследуя химические свойства ряда полупроводниковых соединений, Вагнер [11] обнаружил, что существуют два отличных друг от друга типа полупроводников — «избыточные» и «дефектные». Речь шла в основном об оксидах и сульфидах металлов, причем «дефектными» были образцы, содержащие меньшее количество металла по сравнению со стехиометрическим составом (например, окисленные образцы в случае оксидных соединений). В таких образцах знак термо-э. д. с. и эффекта Холла при низких температурах был положительным. «Избыточными» полупроводниками в случае окислов металлов были частично восстановленные образцы, обладающие некоторым избытком металла. Как правило, знак эффекта Холла в этом случае был отрицательным во всей области температур. Иногда в «дефектных» образцах знак эффекта Холла при высоких температурах становился отрицательным. В настоящее время «избыточные» и «дефектные» полупроводники называют обычно полупроводниками *n*-типа (от слова *negative*) и *p*-типа (от слова *positive*) соответственно. В результате этих работ было выяснено решающее влияние небольших отклонений от стехиометрического состава полупроводниковых соединений на их электрические свойства.

1.1.2. ГАЛОГЕНИДЫ ЩЕЛОЧНЫХ МЕТАЛЛОВ

Несмотря на то что эти вещества относятся скорее к классу диэлектриков, а не полупроводников, необходимо все же отметить многочисленные работы ¹⁾, выполненные Полем [12] и его сотрудниками на галогенидах щелочных металлов, поскольку эти исследования в значительной степени способствовали выяснению многих свойств полупроводников. Одна из главных причин, почему этот класс веществ привлек внимание исследователей, состояла в том, что такие соединения можно получать в виде больших монокристаллов высокой степени чистоты. Многие из работ, посвященных оксидам и сульфидам металлов, были выполнены в свое время на образцах из спрессованных порошков или на пленках, полученных путем испарения в вакууме или с помощью метода химического осаждения. В настоящее время хорошо известно, что на таких образцах часто получаются весьма обманчивые результаты. Необходимо отметить, что электропроводность кристаллов галогенидов щелочных металлов носит скорее ионный, нежели электронный характер. Однако освещением ультрафиолетовым светом можно возбудить в них электропроводность электронного характера. Именно исследование фотопроводности этих соединений и проложило путь к пониманию аналогичных явлений, происходящих под действием видимого или инфракрасного излучения в полупроводниках.

1.1.3. ПОВЕРХНОСТНЫЕ И ОБЪЕМНЫЕ ЯВЛЕНИЯ

Ненадежность многих результатов, полученных в ранних работах по исследованию полупроводников, связана с тем, что в то время еще не умели отделять эффекты, происходящие в объеме исследуемого образца, от явлений, связанных со свойствами его поверхности или границы раздела веществ различной природы. Широкое использование образцов из спрессованных порошков увеличило влияние поверхностных эффектов. Раньше считалось, что отрицательный температурный коэффициент электросопротивления всегда является объемной характеристикой; сейчас хорошо известно, что это отнюдь не так. Явление выпрямления переменного тока с самого начала было совершенно правильно отнесено к числу поверхностных или пограничных явлений, однако много недоразумений вызвали фотогальванический эффект и фотопроводимость.

Небольшое число полупроводников удалось исследовать на образцах, изготовленных из естественных кристаллов, например сернистый свинец (галенит), однако степень чистоты таких образцов была известна недостаточно точно. Новую фазу в исследовании это-

¹⁾ Эти работы начались около 1920 г., некоторые из них описаны Полем в работе [12]; см. также книгу Мотта и Герни [28].

го класса веществ открыла возможность выращивать искусственные монокристаллы некоторых полупроводников. В данной книге в основном освещаются работы именно этого периода. Исследование монокристаллов позволило не только отделить объемные свойства от поверхностных, но и значительно более подробно изучить поверхность полупроводников, а также переходную область между двумя различными полупроводниками или между полупроводником и металлом. Эти исследования убедительно показали, сколько ошибочных результатов можно получить, если не принимать во внимание свойства поверхности полупроводников, например, при изучении поликристаллических образцов.

Позже исследование аморфных полупроводников привело к более полному пониманию таких свойств, как высокая подвижность носителей, которые решающим образом зависят от качества изучаемых кристаллов, и других, более фундаментальных свойств, которые не зависят от дальнего порядка.

1.2. Применение полупроводников

Первым важным техническим применением полупроводников было изготовление выпрямителей переменного тока низкой частоты. Такие выпрямители на основе селена были изготовлены Фриттсом [13] еще в 1886 г., хотя широкое применение в электротехнике и электронике они получили значительно позже. Медно-закисные (купроксные) выпрямители были введены Грондалем и Гейгером [14] в 1927 г. и получили в то время широкое применение в качестве маломощных выпрямителей для зарядки батарей аккумуляторов, питания схем безпроводочной связи и т. д. Примерно тогда же была начата промышленная разработка селеновых выпрямителей. В настоящее время последние уже в значительной мере вытеснили медно-закисные выпрямители из обычного электронного оборудования. Развитие выпрямительной техники привело к появлению большого количества работ, посвященных изучению селена и закиси меди. Несмотря на это, наши знания об основных физических свойствах этих полупроводников весьма скудны по сравнению с теми сведениями, которыми мы располагаем в настоящее время о таких полупроводниках, как германий и кремний. В случае селена это объясняется сложной кристаллической структурой последнего, наличием у него различных кристаллических модификаций, а также трудностями получения вещества в чистом состоянии. Фотопроводимость селена и закиси меди была использована для изготовления фотоэкспонетров, а также фотоэлементов, которые широко применяются в кинопромышленности для превращения записи на звуковой дорожке киноленты в электрические колебания, которые затем усиливаются и воспроизводятся громкоговорящими. Фотоэле-

менты применяются также в таких автоматических устройствах, как счетчики изделий на конвейерах, фотоэлектрические предохранители и т. д. Выпрямители устройств электроснабжения, работающие с очень большими мощностями, равно как и диоды, выпрямляющие высокочастотные сигналы небольших мощностей для таких приборов, как радиоприемники, в настоящее время обычно изготавливаются на основе элементарного полупроводника — кремния.

Хотя радиоволны были открыты Герцем еще в 1888 г., только в 1904 г. выяснилось [15], что полупроводники могут служить прекрасными детекторами высокочастотных токов, возбуждаемых этими волнами в электрических цепях. В течение нескольких последующих лет с этой целью был испробован ряд разнообразнейших веществ. Открытие того факта, что контакт тонкой металлической проволоочки с кристаллом полупроводникового вещества представляет собой превосходный выпрямитель высокочастотных токов, позволило значительно увеличить чувствительность радиоприемников. Такого рода устройства (детекторные радиоприемники) получили широкое распространение на заре современной радиотехники. Для этой цели преимущественно использовались кремний и сернистый свинец — последний в виде естественных кристаллов галенита. Интересный спор возник по поводу природы (электрической или тепловой) механизма, лежащего в основе эффекта выпрямления. Этот спор был решен в обстоятельной работе Пирса [16], убедительно доказавшего, что явление выпрямления переменного тока имеет электрическую природу.

Вскоре кристаллические детекторы стали вытесняться электронными лампами и к 1939 г. были почти полностью забыты. Однако с развитием радиолокации, где применяются весьма короткие длины волн (порядка 10 см и меньше), пришлось опять вернуться к кристаллическим детекторам, так как оказалось, что при таких высоких частотах никакие другие приборы не могут эффективно использоваться в качестве выпрямителей или преобразователей частоты. Выяснилось, что наиболее подходящим материалом для этих применений является кремний. В обзоре Блини и др. [17] освещены работы английских ученых в этой области; в книге Торри и Уитмера [18] дан обзор английских и американских работ. Технические разработки с использованием кремния стимулировали появление большого количества фундаментальных работ, посвященных изучению этого вещества, особенно в лабораториях фирмы «Белл телефон» и в Университете Пардю в США. В этих работах всесторонне исследовался и германий — следующий за кремнием элемент IV группы периодической таблицы. В связи с тем что температура плавления германия намного ниже температуры плавления кремния и он легче поддается очистке, германий оказался более подходящим объектом для проведения фундаментальных исследований в области полупроводников и вскоре стал наиболее изученным среди полупро-

водниковых веществ. Хотя для изготовления детекторов для радиолокационных устройств германий поначалу казался неэффективным материалом, он нашел другое, весьма важное применение при разработке небольших компактных выпрямителей для работы в области низких частот — так называемых германиевых выпрямителей с высоким запирающим напряжением. Исследование полупроводниковых свойств германия оказалось впоследствии весьма плодотворным, так как привело к открытию Бардином и Браттейном [19] эффекта усиления в полупроводниках и к изобретению транзистора в лабораториях фирмы «Белл телефон» [20]¹⁾.

Подробное изучение эффекта усиления и контролируемой инжекции носителей заряда в полупроводниках, проведенное Шокли, Пирсоном и Хейнсом [21], резко расширило круг работ, посвященных исследованиям свойств германия, а несколько позже и кремния, так как стало ясно, что транзисторы сыграют революционную роль в развитии электроники и заменят большую часть множества электронно-вакуумных ламп, используемых в промышленности. Концентрация научно-исследовательской работы вокруг проблемы разработки транзисторов привела к значительно лучшему пониманию основных физических свойств полупроводников. Эффект усиления сам по себе является мощным методом изучения этих свойств, так что взаимосвязь чисто научных и прикладных исследований оказалась весьма плодотворной.

Другая группа полупроводников, нашедшая важное применение, — это так называемые фотопроводники для инфракрасной области спектра, особенно сульфид, селенид и теллурид свинца. Они были использованы для создания чувствительных детекторов инфракрасного излучения, что в свою очередь привело к появлению большого количества работ по изучению свойств этих соединений. Эти работы были начаты в Германии в 1940—1945 гг., а затем получили широкое развитие в целом ряде лабораторий. Обзор современных исследований, посвященных этим соединениям, а также разработке детекторов инфракрасного излучения, составлен автором настоящей книги [22]. Свойства этих детекторов и их использование в инфракрасной спектроскопии подробно описаны автором совместно с Джонсом и Чесмером [23].

В последние годы полупроводники нашли также множество других применений, так что электротехника и электроника, базирующиеся на использовании этих материалов, стали теперь одними из наиболее крупных и быстро развивающихся областей промышленности. Например, в технике больших мощностей полупроводники используются в выпрямителях, коммутирующих устройствах, регуляторах мощности, а также в большом количестве контрольно-

¹⁾ Обзор работ, которые привели к изобретению транзистора, сделали Шокли [20], Пирсон и Браттейн [29].

измерительного оборудования. В телевизионной технике большинство механических переключателей систем к настоящему времени уступило место твердотельным устройствам; они применяются и в автоматической телефонной связи. На явлении фотоэффекта в полупроводниках основано использование солнечных батарей для энергообеспечения спутников. В настоящее время активно изучается пригодность солнечных батарей для нужд земной энергетики.

Следующей ступенью развития транзистора явилось создание интегральной схемы, в которой множество транзисторов и связанных с ними элементов, таких, как резисторы, конденсаторы и т. д., изготавливаются методом контролируемой диффузии примесей в единой кремниевой матрице небольшого размера («чип»). Эти приборы уже через два десятилетия произвели новый переворот в электронной технике и сделали возможным появление таких необычайно сложных систем, как современные электронные цифровые вычислительные машины.

Следует особо упомянуть одно применение полупроводников, основанное на их оптических свойствах. Это твердотельные лазеры, предоставившие спектроскопистам перестраиваемый источник инфракрасного излучения с очень малой шириной спектра, что привело к новым достижениям в области спектроскопии. Такой полупроводниковый лазер, вероятно, будет играть важную роль в линиях оптической связи, так как он легко поддается модуляции и его можно состыковать с оптическим световодом. Менее сложный, но родственный ему прибор, светоизлучающий диод, находит широкое применение при разработке дисплеев для калькуляторов, компьютеров и сигнальных систем; он также может сыграть ключевую роль в технике оптической связи. По причинам, которые мы поясним позже, для этих оптических применений больше всего подходят сложные полупроводниковые соединения. В настоящее время наиболее широко используются с этой целью такие вещества, как GaAs и сплавы типа $Ga_xAl_{1-x}As$.

1.3. Элементарная теория полупроводников

Как мы уже указывали, прежде чем появилась теория, способная удовлетворительно объяснить свойства полупроводников, была проведена огромная работа по их экспериментальному изучению. Это неудивительно, так как для объяснения даже самых элементарных свойств полупроводников требуется привлечение аппарата квантовой механики. Лишь после применения его к исследованию движения электрона в кристаллах можно было приступить к построению более или менее удовлетворительной теории полупроводников. Это было сделано Вильсоном в 1931 г. [24]. До появления теории Вильсона нельзя было даже дать удовлетворительного оп-

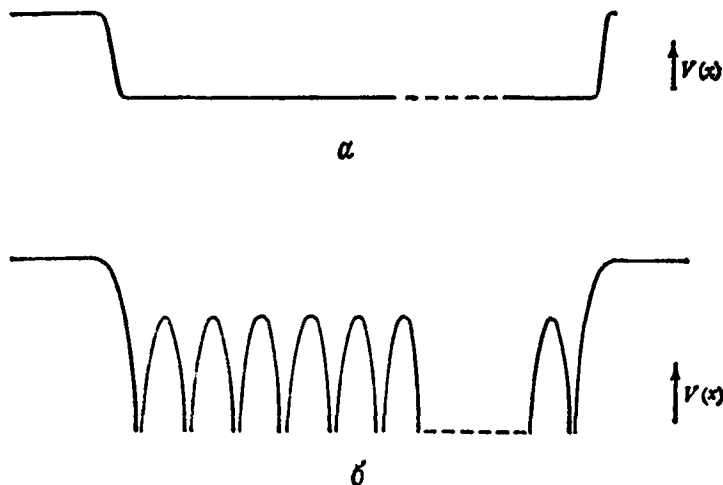


Рис. 1.1. Потенциальная энергия электрона в кристалле.
 а — модель Зоммерфельда; б — периодический потенциал атомных остатков.

ределения понятию полупроводник. Интересно отметить, что в первоначальных работах Вильсона выражались сомнения по поводу того, является ли кремний полупроводником или металлом. Это показательно для той степени неопределенности наших знаний о полупроводниках, которая существовала даже в начале 30-х годов и ликвидации которой в значительной мере способствовала сама теория Вильсона.

Первым применением квантовой механики к рассмотрению движения электронов в твердых телах явилась теория электропроводности металлов Зоммерфельда [25]. Единственное отличие этой работы от предшествовавших ей теоретических исследований, основанных на классической теории электронов, состояло в том, что энергетические уровни, которые могут занимать различные электроны в твердом теле, определялись в ней с помощью квантовой механики. Предполагалось, что в металле электроны внешних атомных оболочек, иначе говоря, валентные электроны (поскольку только они принимают участие в образовании химических связей), не принадлежат определенным атомам металла, а могут свободно перемещаться по всему кристаллу. Считалось также, что каждый электрон в металле движется в пространстве, лишенном поля, так как силовые поля атомных остатков и всех остальных электронов, кроме данного, в среднем взаимно компенсируются везде, за исключением поверхности (см. рис. 1.1, а). У поверхности металла по предположению существует поле, удерживающее электроны в объеме и затрудняющее выход электронов за его пределы. В простей-

шем варианте теории предполагалось, что силовые поля на границе твердого тела образуют абсолютно непроницаемый для электронов потенциальный барьер, удерживающий их внутри кристалла. Дальнейшее усовершенствование модели позволяет рассматривать также и такие явления, как холодную и термоэлектронную эмиссию. Для этого требуется только предположить, что потенциальный барьер у поверхности имеет конечную высоту. Теория Зоммерфельда мало изменила существовавшие в то время представления об электропроводности металлов, однако она сумела объяснить малый вклад электронов проводимости в удельную теплоемкость металлов. Ясно, что использованное в теории Зоммерфельда представление об электронах, которые свободно движутся по кристаллу и не испытывают частых столкновений с атомами, весьма грубо. Более того, эта теория не дает никакого объяснения существенному различию свойств металлов, полупроводников и диэлектриков.

Обычно считалось, что в металлах валентные электроны легко отщепляются от атомов, тогда как в диэлектриках они сильно связаны и не могут быть отщеплены; в этом и видели первопричину различного поведения металлов и диэлектриков. Однако эта точка зрения не согласуется с тем фактом, что энергии ионизации изолированных атомов мало отличаются друг от друга и иногда могут даже оказаться большими для атомов металла, чем для атомов неметалла (например, 9,20 эВ для Au и 8,09 эВ для Ge). Вскоре стало ясно, что для построения более общей теории требуется значительно менее грубое приближение, нежели приближение свободных электронов Зоммерфельда. Следующий шаг был сделан в направлении учета взаимодействия валентных электронов с атомными остатками; считалось, что последние находятся в узлах кристаллической решетки. По-прежнему предполагалось, что электроны в кристаллах движутся независимо друг от друга, однако постоянный потенциал, использованный в модели Зоммерфельда, заменялся потенциалом более общего вида. Характер изменения такого потенциала для определенного направления в кристалле вдоль линии, соединяющей ряд атомов решетки, показан на рис. 1.1, б. Отличительная особенность потенциала такого вида состоит в том, что он периодичен с периодом кристаллической решетки. Движение электронов в периодическом потенциальном поле было исследовано Блохом [26] с помощью аппарата квантовой механики. Блох получил при этом фундаментальный результат, который полностью изменил сам подход к проблеме движения электронов в кристаллических твердых телах. Этот результат не зависит от конкретного вида периодического потенциала и состоит в том, что в идеальной периодической решетке электрон может двигаться совершенно свободно, не испытывая «рассеяния» со стороны отдельных атомов кристалла. Рассеивать электроны могут только *нарушения* идеальной периодичности кристаллической решетки. Таким образом, с этой точки зрения модель Зоммерфельда

оказывается гораздо ближе к истине, чем можно было предполагать. Это верно по крайней мере для металлов. Имеются, однако, фундаментальные отличия между моделями Блоха и Зоммерфельда. Длина пути электрона, на которой его можно считать «свободным», определяется в модели Блоха несовершенством кристаллической решетки и в случае идеального кристалла стремится к бесконечности. Как правило, эта длина пути значительно больше постоянной решетки и электрон в кристалле свободно пролетает много межатомных расстояний, прежде чем испытает отклонение в результате «столкновения».

Модель Блоха является весьма общей и применима к описанию движения в идеальном периодическом поле любых, не обязательно валентных электронов. Как же в теории Блоха рассматриваются сильно связанные электроны, расположенные на внутренних оболочках атомов, из которых построен кристалл? Ответ, который следует из этой теории, состоит в том, что даже индивидуальные электроны нельзя считать принадлежащими какому-либо определенному атому в кристалле, наоборот, они могут свободно меняться местами с электронами других атомов кристалла, идентичных данному. Что же тогда произойдет, если приложить к кристаллу электрическое поле? Примут ли все эти электроны участие в электропроводности? Именно рассмотрение этой проблемы и привело Вильсона [24] к созданию теории полупроводников.

Согласно теории Зоммерфельда разрешенные уровни валентных электронов в кристалле макроскопических размеров расположены очень близко друг к другу и образуют систему уровней, простирающуюся почти от дна потенциальной ямы, в которой движутся электроны, до бесконечно больших значений энергии. Уровни энергии всех остальных электронов считаются невозмущенными и совпадают с атомными энергетическими уровнями. Эти уровни одни и те же для каждого из атомов кристалла, поэтому они вырождены по крайней мере N -кратно, по числу N атомов в кристалле (рис. 1.2). В теории Блоха вследствие введения периодического потенциала энергетические уровни группируются в определенные полосы (зоны) разрешенных энергий, разделенные областями, в которых нет разрешенных уровней энергии (запрещенные зоны). Для электронов, расположенных на внутренних оболочках атомов, эти разрешенные зоны чрезвычайно узки и соответствуют атомным энергетическим уровням. Зоны энергий валентных электронов оказываются достаточно широкими. Расположение уровней показано на рис. 1.3. Каждая зона состоит из множества близко расположенных уровней и во многих приложениях может рассматриваться как континуум состояний. Однако при расчете распределения электронов по уровням в зоне необходимо учитывать дискретность энергетического спектра и конечность общего числа уровней в зоне. Аналогично тому как заполнены электронами все внутренние состояния тяжелых

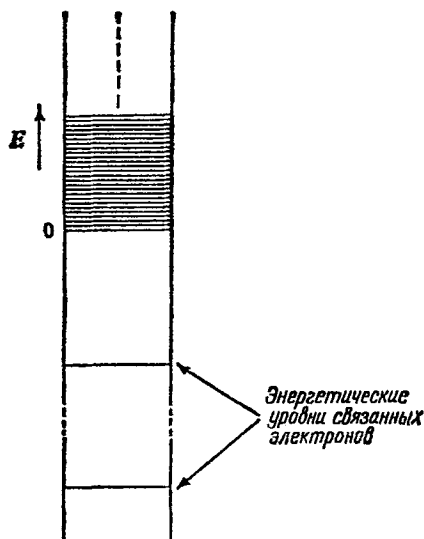


Рис. 1.2. Разрешенные уровни энергии электронов в металле в модели Зоммерфельда.

В качестве нулевого уровня энергии выбрана энергия «свободного» валентного электрона, не обладающего кинетической энергией.

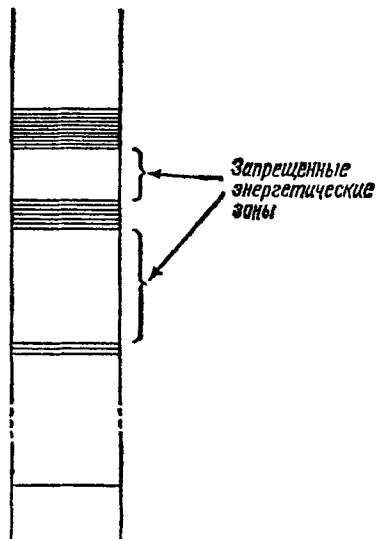


Рис. 1.3. Разрешенные уровни энергии в периодической решетке.

атомов, также заполнены и все уровни в нижних энергетических зонах. Полностью или частично пустыми могут быть лишь состояния в вышележащих зонах.

К такой картине энергетических зон в твердом теле можно прийти и другим путем. Рассмотрим два одинаковых атома, расположенных на большом расстоянии друг от друга, и предположим, что каждый из них обладает одним невырожденным уровнем энергии для электрона. Спектр разрешенных энергий системы двух удаленных атомов сводится к одному дважды вырожденному уровню, так как оба электрона в системе обладают в точности одной и той же энергией. Будем теперь сближать атомы. Вследствие возрастающего взаимодействия между ними вырожденный уровень энергии расщепляется на два различных уровня. Это весьма общий, известный из механики, результат, соответствующий расщеплению частот двух тождественных, взаимодействующих друг с другом осцилляторов. По мере приближения атомов величина расщепления растет (рис. 1.4). Аналогичные рассуждения можно применить и к N атомам кристалла. Вначале предположим, что кристалл равномерно растянут так, что межатомные расстояния в нем очень велики. Тогда разрешенные уровни для электронов совпадут с атомными уровнями, которые для простоты будем предполагать невырожденными. В кристалле каждый атомный уровень становится N -кратно вырож-

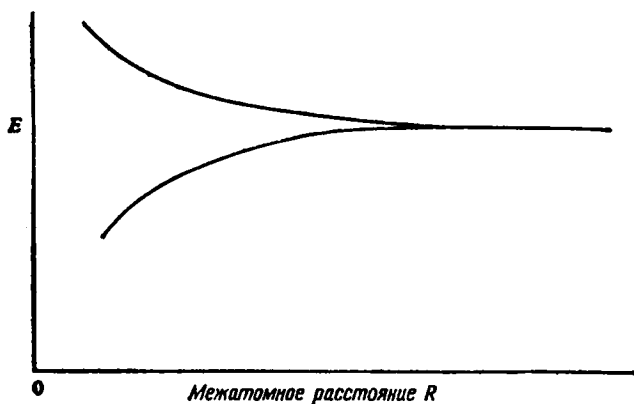


Рис. 1.4. Расщепление энергетического уровня в результате взаимодействия между двумя атомами.

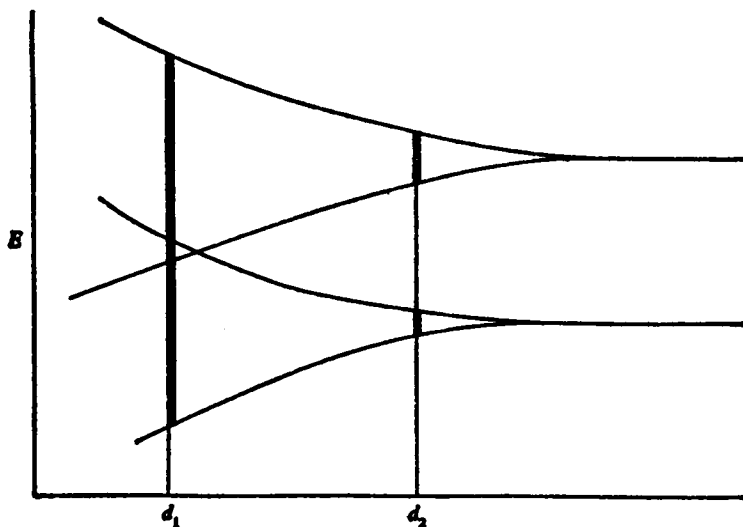


Рис. 1.5. Энергетические зоны кристалла в зависимости от межатомного расстояния.

денным. Если затем начать постепенно уменьшать межатомные расстояния в кристалле, то вследствие растущего взаимодействия атомов друг с другом каждый уровень расщепится на серию N различных уровней. Вместо каждого N -кратно вырожденного уровня получаем зону энергий, содержащую N плотно размещенных уровней. На рис. 1.5 показаны крайние значения энергии для двух зон в зависимости от величины межатомных расстояний в решетке. В слу-

чае глуболежащих атомных уровней возмущение оказывается слабым по сравнению с силами притяжения между электроном и ядром; соответственно и расщепление этих уровней будет мало. Для валентных электронов расщепление может быть довольно большим и приводит иногда даже к перекрытию соседних энергетических зон (см. рис. 1.5).

Рассмотрим сначала случай, когда зоны не перекрываются. Чтобы решить вопрос о распределении электронов по состояниям в зонах, необходимо привлечь принцип Паули, согласно которому на каждом уровне в зоне может разместиться не более двух электронов, т. е. по одному для каждого направления спина. Начнем с глуболежащих атомных уровней. В любом атоме каждый такой уровень энергии должен быть занят двумя электронами, так что в кристалле окажется $2N$ электронов, которые при больших межатомных расстояниях обладали бы энергией, соответствующей этому уровню. Эти $2N$ электронов необходимо разместить по состояниям в зоне. Но, как мы видели, зона, содержащая N уровней, может, согласно принципу Паули, вместить в точности $2N$ электронов. Таким образом, зоны, соответствующие атомным уровням, на которых имелись спаренные электроны со взаимно противоположными спинами, оказываются *полностью заполненными*. Различным полностью заполненным (замкнутым) оболочкам атомов можно опять-таки сопоставить полностью заполненные зоны в кристалле. Частично занятыми могут быть лишь зоны, соответствующие внешним, или валентным, электронам. Это различие между полностью и частично заполненными зонами имеет очень важное значение. Оно иллюстрируется рис. 1.6. Аналогично тому как в атомах существуют уровни возбужденных состояний, так и в спектре энергий кристаллов над валентной зоной должна располагаться серия более высоких, как правило, пустых энергетических зон, соответствующих возбужденным состояниям электронов.

Теперь легко понять, почему электроны в низших зонах, соответствующих «атомным остовам», не могут принимать участия в электропроводности, несмотря на то что, как показал Блох, в не-

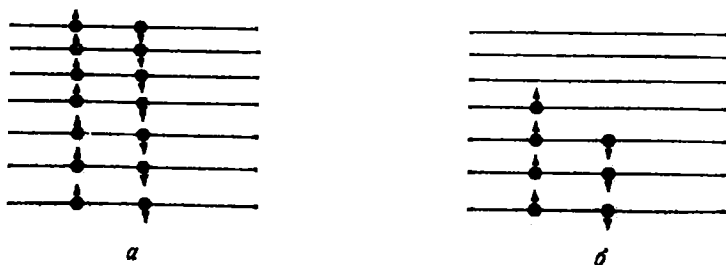


Рис. 1.6. Полностью заполненная (а) и частично заполненная (б) зоны.

котором смысле они могут беспрепятственно перемещаться по идеальному кристаллу. Чтобы участвовать в процессе электропроводности, электрону необходимо приобрести дополнительную энергию за счет электрического поля, т. е. поле должно оказывать на него ускоряющее действие. На языке квантовой механики это означает, что электрон должен перейти на более высокий энергетический уровень. Если же *все* уровни в зоне уже заняты, то электрон не может ускоряться до тех пор, пока в результате возбуждения он не перейдет в вышележащую зону, в которой имеются свободные уровни. Процесс возбуждения, однако, очень маловероятен, так как для перевода электрона в другую зону ему необходимо сообщить довольно большую энергию. Таким образом, чтобы электроны, занимающие определенную группу уровней в зоне, могли внести значительный вклад в электропроводность, необходимо наличие *близлежащих* пустых энергетических уровней, на которые рассматриваемые электроны могли бы перейти под действием ускоряющего поля, полностью заполненные зоны не вносят какого-либо вклада в процесс электропроводности. Это обстоятельство не только объяснило, почему внутренние электроны не участвуют в электропроводности, но и дало Вильсону ключ к пониманию существенного различия между металлами, с одной стороны, и диэлектриками и полупроводниками — с другой.

Рассмотрим теперь наивысшую зону, в которой обычно еще имеются занятые электронами уровни энергии. В щелочном металле, например в металлическом натрии, каждый атом содержит по одному валентному электрону; все остальные электроны образуют замкнутые оболочки. Соответствующая валентная зона в металле состоит из N уровней, и имеется N электронов, которые могут в ней разместиться. В невозбужденном состоянии эти электроны займут $N/2$ наиболее глубоких уровней в валентной зоне, т. е. по два электрона на уровень. Зона будет заполнена только наполовину, и в ней окажется множество доступных, т. е. близлежащих свободных уровней, которые, согласно сказанному, необходимы для осуществления процесса электропроводности. Таким образом, при наличии *одного лишь* валентного электрона на атом ситуация действительно во многом напоминает модель металла, принятую в теории Зоммерфельда.

Рассмотрим теперь кристалл, построенный из атомов, содержащих в свободном состоянии по два электрона с противоположными спинами на валентном уровне. В этом случае валентная зона кристалла, т. е. наивысшая из всех зон, состояния которых заняты электронами, оказывается полностью заполненной. Если к тому же между валентной зоной и следующей за ней верхней пустой зоной имеется энергетический зазор (запрещенная зона или щель), то на основании вышесказанного такой кристалл не должен обладать электропроводностью и соответствующее вещество может быть диэлектриком. Диэлектрик характеризуется тем, что наивысшая зона, содержа-

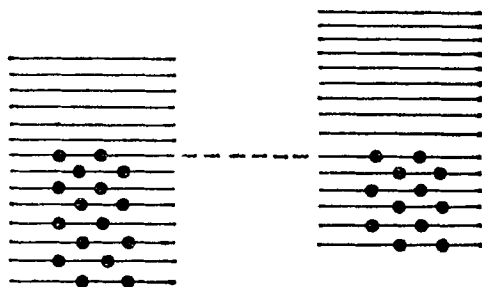


Рис. 1.7. Частичное заполнение двух перекрывающихся энергетических зон.

щая электроны, полностью заполнена и отделена широкой полосой запрещенных энергий от следующей пустой зоны. На основании этой простой картины следовало бы ожидать, что двухвалентные щелочные металлы типа магния окажутся диэлектриками. Тот факт, что на самом деле это не так, означает, что в данном случае обе зоны перекрываются, как это показано на рис. 1.5 для межатомного расстояния d_i . В этом случае каждая из указанных зон заполнена лишь частично (рис. 1.7) и поэтому возможность электропроводности сохраняется.

Ситуация, которую мы обрисовали для диэлектриков, действительно имеет место в некоторых твердых телах. В кристаллах четырехвалентных элементов, например в алмазе (модификация углерода), кремнии, германии, наивысшая зона, содержащая электроны, полностью заполнена и отделена широким энергетическим зазором от ближайшей пустой зоны. В соединениях ситуация сложнее, однако и здесь при четном числе валентных электронов, приходящихся на одну молекулу, валентная зона может быть полностью заполненной. Необходимо подчеркнуть, что это возможно *только* при четном числе валентных электронов.

Приведенные выше соображения применимы, строго говоря, лишь к идеальным кристаллическим структурам. В дальнейшем речь будет идти о кристаллических телах, представляющих собой не только монокристаллы больших размеров, но и поликристаллические структуры, т. е. агрегаты мелких кристалликов, случайным образом ориентированных по отношению друг к другу. В виде исключения образцы могут быть приготовлены в аморфном виде (например, слой аморфного кремния), интерес к свойствам таких структур недавно увеличился. Ниже мы кратко рассмотрим аморфные полупроводники (гл. 15), однако в основном речь будет идти здесь о материалах, представляющих собой совершенные монокристаллы.

Имеется еще и другой аспект квантовой теории, который проливает свет на высказанные выше идеи. Вследствие волновых свойств электрона последний не может долго оставаться в нег-

лубокой потенциальной яме, а переходит посредством туннельного эффекта в ближайшую соседнюю яму, если только потенциальный барьер достаточно тонок и не слишком высок. Атомы в реальном твердом теле расположены настолько тесно, что электрон, принадлежавший ранее какому-либо атому, — в особенности это относится к валентному электрону — может легко перейти из потенциальной ямы, удерживающей его вблизи данного атома, в потенциальную яму соседнего атома. Таким образом, электрон нельзя более считать принадлежащим определенному атому в кристалле. Это обстоятельство было выяснено вскоре после возникновения квантовой механики, и поэтому было трудно понять, каким образом можно считать электроны в диэлектриках «сильно связанными». Теория заполненных зон показывает, что электроны могут не быть связанными в обычном смысле и в то же время не участвовать в процессе электропроводности.

На рассмотренных принципах и основана теория Вильсона. Если ширина запрещенной зоны ΔE между наивысшей заполненной и следующей за ней пустой зонами велика, то электропроводность отсутствует, так как не существует близлежащих пустых уровней, обеспечивающих возможность ускорения электронов. Однако если ширина запрещенной зоны ΔE невелика, то имеется возможность теплового возбуждения электронов, приводящего к забросу последних в пустую зону, после чего возбужденные электроны уже могут участвовать в процессе электропроводности. Число возбужденных электронов будет увеличиваться с ростом температуры по закону, характерному для процессов, имеющих энергию активации порядка ΔE . Поэтому следует ожидать, что проводимость в этом случае будет быстро расти с температурой. Такие вещества, согласно теории Вильсона, и являются полупроводниками. Это определение оказывается слишком строгим, так как оно построено на основе зонной теории, применимой только к кристаллическим твердым телам. Вместе с тем желательно, чтобы определение было применимо также и к аморфным твердым телам. Как мы увидим ниже (разд. 3.2), такое определение можно дать на языке химических связей. Обычно полупроводник — это материал, у которого в химически чистом состоянии свободным электронам требуется энергия активации ΔE небольшой величины (скажем, $0 < \Delta E < \sim 2$ эВ), чтобы они могли участвовать в процессе электропроводности. Как мы увидим ниже, введение очень малого количества примесей может заметно повлиять на энергию активации этого процесса, так что материалы, обладающие величинами ΔE , значительно превышающими 2 эВ, могут вести себя как полупроводники, если они содержат определенную «активную» примесь.

1.3.1. ПРОЦЕССЫ ЭЛЕКТРОПРОВОДНОСТИ

Вскоре выяснилось, что процесс электропроводности в действительности значительно сложнее, и мы рассмотрим его более подробно. Когда один электрон переходит из полностью заполненной зоны в более высокую пустую зону, имеет место весьма интересное явление: в ранее заполненной зоне освобождается одно состояние (уровень) и остальные электроны получают возможность перейти на освободившийся уровень энергии. Поведение такой «вакансии» аналогично поведению пузырька воздуха в жидкости в поле сил тяготения. Во внешнем электрическом поле такая вакансия движется в направлении, противоположном направлению движения электрона, как если бы она обладала положительным зарядом. Она ведет себя и во многих других отношениях как положительно заряженная частица и поэтому ее обычно называют «положительной» дыркой или просто «дыркой». Подробнее это явление будет рассмотрено в разд. 2.4. Таким образом, электропроводность в идеальном полупроводнике обусловлена движением под действием электрического поля как электронов в почти пустой зоне, так и положительных дырок в зоне, почти заполненной электронами.

Этот процесс лежит в основе интерпретации свойств полупроводников при высоких температурах, в частности экспоненциального закона роста электропроводности с температурой, наблюдаемого в полупроводниках. Этот закон попросту является следствием существования некоторой энергии активации (в химическом смысле) порядка ΔE , необходимой для тепловой генерации электронно-дырочных пар. Одновременное образование свободного электрона (e), т. е. электрона в пустой зоне, или зоне «проводимости», и дырки (h) в заполненной зоне, или «валентной» зоне, можно рассматривать как реакцию



в которой нормальное невозбужденное состояние обозначено через (eh) . Если концентрации электронов n и дырок p малы по сравнению с полным числом атомов в кристалле N , то число невозбужденных состояний (eh) приближенно совпадает с N . Тогда из термодинамической теории химических реакций следует, что

$$\frac{np}{N^2} = K(T), \quad (1.2)$$

где функция $K(T)$ зависит только от температуры [27]. Известно также, что $K(T)$ можно записать в виде

$$K(T) = [F(T)]^2 \exp \left[-\frac{\Delta E}{kT} \right], \quad (1.3)$$

где ΔE — энергия активации данной реакции, k — постоянная Больцмана, T — абсолютная температура, $F(T)$ — функция, сравнительно слабо зависящая от T , если $\Delta E \gg kT$. Так как в рассмат-

риваемом случае $n=p$, то мы имеем

$$n = p = A \exp \left[-\frac{\Delta E}{2kT} \right], \quad (1.4)$$

где в первом приближении A можно считать постоянной. Если предположить, что температурная зависимость подвижности электронов и дырок в электрическом поле слаба по сравнению с соответствующей зависимостью экспоненциального множителя в (1.4), то удельная проводимость σ , которая при сделанных упрощающих предположениях просто пропорциональна числу носителей заряда, приобретает следующий вид:

$$\sigma = \sigma_0 \exp \left[-\frac{\Delta E}{2kT} \right]. \quad (1.5)$$

Такая экспоненциальная зависимость электропроводности уже давно наблюдалась в полупроводниках при высоких температурах. Однако при более низких температурах рост электропроводности с температурой значительно замедляется и одновременно появляется огромный разброс значений электропроводности при переходе от образца к образцу. Вильсон предположил, что поведение полупроводников в области низких температур обусловлено несовершенством кристалла, связанным либо с механическими дефектами, либо с химическими примесями. Выше предполагалось, что в запрещенной зоне отсутствуют разрешенные уровни энергии, однако это утверждение относится к случаю идеального кристалла. Вильсон показал, что наличие примесей в кристалле может приводить к появлению в запрещенной зоне локальных энергетических уровней в непосредственной близости к краям зоны проводимости или валентной зоны.

Если примесные уровни близки к краю пустой зоны, то в результате возбуждения электроны легко переходят с примесных уровней в зону, где они могут участвовать в электропроводности. Такие примеси называют донорными, или примесями n -типа. С другой стороны, если примесные уровни лежат вблизи края валентной зоны и могут захватывать электроны, то опять-таки вследствие возбуждения электроны легко покидают валентную зону, в которой остаются дырки, способные принять участие в электропроводности. Такие примеси называются акцепторными, или примесями p -типа. Таким образом, следует различать три процесса электропроводности в полупроводниках:

1) собственная проводимость, в которой участвует одинаковое число электронов и дырок, причем оба рода носителей создаются путем перехода электронов из валентной зоны прямо в зону проводимости;

2) электронная проводимость, или проводимость n -типа; участвуют лишь электроны, переходящие в зону проводимости с расположенных вблизи ее края донорных уровней;

3) дырочная проводимость, или проводимость p -типа; участвуют только дырки, образовавшиеся при переходах электронов из валентной зоны на близлежащие акцепторные уровни.

Обычно все три процесса происходят одновременно, но часто главную роль играет лишь один из них, и тогда в зависимости от природы преобладающего механизма электропроводности полупроводник называют соответственно либо собственным полупроводником, либо полупроводником n - или p -типа. Полупроводники n - и p -типа соответствуют описанным Вагнером «избыточным» и «дефектным» полупроводникам (см. разд. 1.1).

Во многих случаях оказалось, что энергия, необходимая для перевода электрона с донорного уровня в зону проводимости, настолько мала, что при комнатных температурах все электроны с примесных уровней находятся уже в зоне проводимости. Если к тому же концентрация доноров намного превышает концентрацию собственных электронов, то количество носителей почти перестает зависеть от температуры и изменение электропроводности с температурой в этих условиях определяется в основном лишь температурной зависимостью подвижности. В этой области температур подвижность, как правило, убывает с ростом температуры. Этим, кстати, объясняется рост сопротивления металлов с температурой, так как в металлах число носителей также постоянно. Таким образом, сопротивление примесного проводника n -типа будет увеличиваться с ростом температуры, пока концентрация собственных носителей (электронов) не начнет преобладать над концентрацией примеси. Начиная с этих температур, сопротивление будет падать по экспоненциальному закону, который считается характерным для полупроводников. Аналогичные соображения применимы и к случаю полупроводника p -типа. Рост сопротивления примесных полупроводников с температурой в области невысоких температур привел в свое время к немалой путанице в вопросе о классификации веществ, которая была по-настоящему преодолена лишь после широких исследований на германии, начатых в 1940 г. Ларк-Горовицем [4] и его сотрудниками в Университете в Пардю. Эти работы показали, что рост сопротивления с температурой еще не может считаться доказательством того, что данное вещество не является полупроводником. А ведь именно такая температурная зависимость сопротивления заставила многих сомневаться в том, является ли кремний полупроводником. В кремнии ширина запрещенной зоны настолько велика (порядка 1,2 эВ), что для достижения заметной собственной проводимости необходимы довольно высокие температуры (обычно выше 500 °С).

Если ширина запрещенной зоны еще больше, например несколько электрон-вольт, то температура, при которой могла бы наблюдаться заметная собственная проводимость, настолько высока, что

она заведомо лежит выше точки плавления кристалла. Такие вещества следует считать диэлектриками. Однако, чтобы вещество могло считаться хорошим диэлектриком, требуется еще и отсутствие примесных уровней, расположенных вблизи краев валентной зоны или зоны проводимости.

1.4. Способы управления концентрацией носителей

Наиболее характерное различие между металлами и полупроводниками состоит в том, что в металлах концентрация носителей заряда велика и неизменна, тогда как в полупроводниках она меньше и ее величина изменяема. Изменчивость этой характеристики в случае полупроводников наводит на мысль о возможности управлять числом носителей заряда, а значит, и электропроводностью. Одним из способов осуществления такого управления является дозировка содержания примеси в полупроводнике. Однако концентрацию носителей заряда в данном материале можно по желанию варьировать и при фиксированном содержании примесей. Одним из замечательных свойств полупроводников является наличие эффекта фотопроводимости. Это явление легко понять с помощью зонной модели. При поглощении кванта света, энергия которого достаточна для перевода электрона из состояния у потолка валентной зоны в зону проводимости, в обеих зонах появляются добавочные носители заряда и проводимость кристалла растет (см. разд. 10.9). Может иметь место также фотопроводимость, связанная с фотоионизацией примесных атомов, однако этот эффект менее значителен сравнительно с упомянутым выше, в связи с тем, что концентрация примесей обычно во много раз меньше концентрации валентных электронов и, кроме того, примесные уровни заполнены лишь при низких температурах (см., однако, разд. 10.7). Вместе с тем оказалось, что концентрации электронов и дырок в полупроводниках можно изменять и иными способами. Во-первых, путем инъекции носителей заряда в кристалл, например с помощью металлических электродов, и, во-вторых, при определенных условиях путем экстракции (удаления) носителей из кристалла электрическим полем.

Именно на этой возможности управления концентрацией носителей заряда основаны весьма важные технические применения полупроводников.

Как правило, в полупроводниках, используемых для практических целей, всегда преобладает тот или иной тип примесных уровней. Когда преобладает донорная примесь и ее концентрация значительно больше собственной концентрации носителей, тогда $n \gg p$. В обычных условиях (см. разд. 4.3) концентрации электронов

и дырок, даже при наличии примесей, по-прежнему связаны уравнением типа (1.2), причем произведение pn не зависит от концентрации примесей. Поэтому

$$pn = n_i^2, \quad (1.6)$$

где n_i — собственная концентрация носителей заряда при данной температуре, т. е. концентрация носителей в идеально чистом веществе. Равенство (1.6) следует из (1.2), если для собственного полупроводника положить $n=p=n_i$. Итак, при $n \gg n_i$ всегда $p \ll n_i$. В случае $n > p$ электроны называют основными носителями, а дырки — неосновными; в случае обратного неравенства основными носителями считают дырки, а неосновными — электроны. Таким образом, за исключением случая почти собственного полупроводника, концентрация основных носителей заряда всегда значительно превосходит концентрацию неосновных. Тем не менее во многих электронных процессах, весьма важных с точки зрения полупроводниковой техники, решающую роль играют именно неосновные носители, концентрацией которых значительно легче управлять из-за ее малости.

В заключение приведем численный пример для иллюстрации относительного вклада различных механизмов электропроводности. В чистом германии собственная концентрация n_i электронов или дырок при комнатной температуре представляет собой величину порядка $2,5 \cdot 10^{13} \text{ см}^{-3}$. Заметим, что в результате успехов, достигнутых в области технологии полупроводниковых материалов, получение германия, обладающего характеристиками собственного полупроводника при комнатной температуре, не представляет в настоящее время особых трудностей. Число атомов германия в единице объема равно $4,5 \cdot 10^{22} \text{ см}^{-3}$. Следовательно, если в германии присутствует донорная примесь, которая при комнатных температурах полностью «ионизирована», то при наличии одного атома примеси на 10^6 атомов германия мы получили бы $n = 4,5 \cdot 10^{16} \text{ см}^{-3}$ электронов. Отсюда $p = n_i^2/n = 1,4 \cdot 10^{10} \text{ см}^{-3}$, так что, действительно, $n \gg p$. Если температуру поднять до 300°C , то $n_i = 1,5 \cdot 10^{17} \text{ см}^{-3}$ и преобладающей становится уже собственная концентрация носителей заряда.

Из этой оценки видно, что для того чтобы германий сохранял свойства собственного полупроводника при комнатной температуре, примеси, образующие уровни вблизи зоны проводимости, могут присутствовать в кристалле в количестве не более одного атома примеси на 10^9 атомов германия. Поэтому неудивительно, что до сих пор лишь совсем небольшое число полупроводников удалось очистить до такой степени, чтобы они обладали собственной проводимостью при комнатной температуре. Почти все ранние работы, посвященные исследованию полупроводников, за исключением измерений в области высоких температур, были выполнены в условиях, когда свойства вещества в основном определялись примесью.

Что касается области высоких температур, то и здесь есть свои трудности; особенно это относится к полупроводниковым соединениям, которые имеют тенденцию разлагаться при высоких температурах. Неудивительно поэтому, что столь часто получались противоречивые или даже ошибочные результаты. Из всего сказанного следует также, что именно благодаря развитой в последнее время совершенной технике предельной очистки многих веществ удалось достичь таких больших успехов в изучении свойств полупроводников.

ЛИТЕРАТУРА

1. *Faraday M.*, Experimental Researches in Electricity, ser. IV, 433 (1833); Beibl. Ann. Phys., 31, 25 (1834).
2. *Braun F.*, Ann. Phys. Chem., 153, 556 (1874).
3. *Smith W.*, Journ. Soc. Electr. Eng., 2, 31 (1873).
4. *Lark-Horowitz K.*, The New Electronics, в сб.: The Present State of Physics, American Assn. for the Advancement of Science, Washington, 1954.
5. *Gudden B.*, Ergebn. exakt. Naturwiss., 3, 143 (1924); 13, 223 (1934).
6. *Иоффе А. Ф.*, Физика полупроводников, 2-е изд., М.—Л., 1957.
7. *Hall E. H.*, Amer. Journ. Math., 8, 287 (1879).
8. *Baedecker K.*, Phys. Zs., 29, 506 (1909).
9. *Königsberger J.*, Jahrb. d. Radioakt., 4, 158 (1907); 11, 84 (1914).
10. *Merrit E.*, Proc. Nat. Acad. Sci., Washington, 11, 743 (1925).
11. *Wagner C.*, Zs. Chem. Phys., B11, 163 (1930); 22, 195 (1933).
12. *Pohl R. W.*, Proc. Phys. Soc., 43, 3 (1937).
13. *Fritts C. E.*, Amer. Journ. Sci., 26, 465 (1883).
14. *Grondahl L. O.*, *Geiger P. N.*, Trans. Amer. Inst. Electr. Eng., 46, 357 (1927).
15. *Bose J. C.*, Патент США, 755840, 1904.
16. *Pierce G. W.*, Phys. Rev., 25, 31 (1907); 28, 153 (1909); 29, 478 (1909).
17. *Bleaney B.*, *Ryde J. W.*, *Kinsman T. H.*, Journ. Inst. Electr. Eng., IIIA, 93, 847 (1946).
18. *Torrey H. C.*, *Whitmer C. A.*, Crystal Rectifiers (Radiation Laboratory Series), New York, 1948. [Имеется перевод: А. Торрей, К. Уитмер. Кристаллические детекторы.— Сов. радио, 1950.]
19. *Bardeen J.*, *Brattain W. H.*, Phys. Rev., 74, 1208 (1949).
20. *Shockley W.*, Electrons and Holes in Semiconductors, New York, 1950. [Имеется перевод: Шокли В. Теория электронных полупроводников.— М.: ИЛ, 1953.]
21. *Shockley W.*, *Pearson G. L.*, *Haynes J. R.*, Bell Syst. Tech. Journ., 28, 344 (1949).
22. *Smith R. A.*, Adv. Phys., 2, 321 (1953); Sci. Mon., 82, 3 (1956).
23. *Smith R. A.*, *Jones F. E.*, *Chasmar R. P.*, The Detection and Measurement of Infra-red Radiation, Oxford, 1968. [Имеется перевод 1-го издания: Смит Р., Джонс Ф., Чесмер Р. Обнаружение и измерение инфракрасного излучения.— М.: ИЛ, 1959.]
24. *Wilson A. H.*, Proc. Roy. Soc., A133, 458 (1931); 134, 277 (1931).
25. *Sommerfeld A.*, Zs. Phys., 47, 1 (1928).
26. *Bloch F.*, Zs. Phys., 52, 555 (1928).
27. *Smith R. A.*, The Physical Principles of Thermodynamics, New York, 1952, p. 219.
28. *Mott N. F.*, *Gurney R. W.*, Electronic Processes in Ionic Crystals, Oxford University Press, Oxford, 1940. [Имеется перевод: Мотт Н., Гурни Р. Электронные процессы в ионных кристаллах.— М.: ИЛ, 1950.]
29. *Pearson G. L.*, *Brattain W. H.*, Proc. IRE, 43, 1794 (1955).

УРОВНИ ЭНЕРГИИ В КРИСТАЛЛИЧЕСКИХ ТВЕРДЫХ ТЕЛАХ

2.1. Волновая механика свободных электронов

В предыдущей главе мы ввели квантовые представления об уровнях энергии электронов в кристаллических твердых телах. Теперь необходимо подробнее изучить форму энергетических зон, обусловленных полем кристаллической решетки, а также энергетические уровни примесей, не вдаваясь в детали соответствующего теоретического вывода. Этот вопрос широко освещен во многих известных руководствах по квантовой теории твердых тел¹⁾. Здесь будут приведены лишь некоторые из наиболее фундаментальных результатов этой теории с указанием на то, как применяются эти общие положения к случаю полупроводников. Мы сможем убедиться в том, что теория позволяет описывать движение электронов в твердых телах, оперируя небольшим количеством величин, которым можно дать строгое определение, и что во многих случаях детальную структуру твердого тела можно не рассматривать. Теория, естественно, позволяет вычислить эти величины, но их значения можно также получить (причем, как правило, с гораздо большей точностью) из опытных данных и в этом случае их можно трактовать просто как параметры феноменологической теории.

Для начала рассмотрим движение электрона в идеальной кристаллической решетке в отсутствие примесей и дефектов. Предполагается, что электроны движутся независимо друг от друга, причем каждый из них подчиняется уравнению Шредингера для движения электрона в силовом поле, в котором он обладает потенциальной энергией $U^2)$,

$$\nabla^2 \Psi + \frac{8\pi^2 m_0}{h^2} [E - U] \Psi = 0, \quad (2.1)$$

¹⁾ См., например, [1—3]. Рассмотрение для случая полупроводников выполнено автором [4].

²⁾ В дальнейшем для краткости вместо термина «потенциальная энергия» будет использоваться термин «потенциал», — Прим. ред.

где Ψ — волновая функция электрона, E — его энергия. Вероятность обнаружения электрона в объеме dV равна $|\Psi|^2 dV$, а соответствующая плотность заряда — $e|\Psi|^2$, где e — заряд электрона. В случае пространства, свободного от полей, потенциал U постоянен и волновая функция определяется как

$$\Psi_{\mathbf{k}} = \exp[i\mathbf{k} \cdot \mathbf{r}], \quad (2.2)$$

что соответствует плоской волне. Тогда энергия E дается выражением

$$E = \frac{\hbar^2 k^2}{2m_0}, \quad (2.3)$$

где

$$k^2 = |\mathbf{k}^2| = k_x^2 + k_y^2 + k_z^2. \quad (2.4)$$

E — представляет собой кинетическую энергию электрона. Волновой вектор \mathbf{k} связан с импульсом электрона \mathbf{p} и скоростью \mathbf{v} соотношением

$$\mathbf{p} = \hbar \mathbf{k} = m_0 \mathbf{v}, \quad (2.5)$$

причем m_0 — масса электрона в свободном пространстве (см. [4], § 1.2.1).

Если $U(\mathbf{r})$ — периодический потенциал, имеющий периодичность кристаллической решетки, то уравнение Шредингера имеет формально сходное решение. Как будет видно в дальнейшем, мы можем ввести величины, соответствующие скорости и импульсу электрона, движущегося в кристалле, однако эти величины будут усреднены по большому количеству ячеек кристаллической решетки.

2.2. Движение в пространстве с периодическим потенциалом

В случае одномерного периодического потенциала с периодом d потенциальная энергия $U(x)$ удовлетворяет условию

$$U(x) = U(x + rd), \quad (2.6)$$

где r — любое целое число. Можно показать, что в этом случае решения соответствующего уравнения Шредингера должны иметь вид [5]

$$\psi_k(x) = u_k(x) \exp[ikx], \quad (2.7)$$

где $u(x)$ — периодическая функция с периодом d , k — постоянная, которая может служить «квантовым числом» для характеристики состояния, описываемого волновыми функциями (2.7). Эти волновые функции по своей природе аналогичны волновым функциям типа (2.2) для случая свободного электрона. Однако физический смысл величины k в данном случае не столь нагляден. При движении

электрона в периодическом поле нарушается пропорциональность между волновым числом k и импульсом, который не является константой при движении в пространстве с периодическим потенциалом, т. е. не является интегралом движения. Величину P , определяемую равенством

$$P = \hbar k, \quad (2.8)$$

обычно называют «квазиимпульсом», чтобы отличить ее от настоящего импульса электрона, с которым она совпадает, лишь когда $U(x) = \text{const}$. Квазиимпульс P , очевидно, является интегралом движения и может служить вместо k в качестве «квантового числа». Ниже мы увидим, что P обладает многими свойствами, присущими импульсу.

Из функций типа (2.7) можно построить волновые пакеты для описания движения электронов. В этом случае можно показать, что скорость электрона равна групповой скорости $d\omega/dk$, где величина ω выражается через энергию с помощью обычного квантового соотношения $E = \hbar\omega$. Теперь, однако, вид зависимости E от k значительно усложняется и определяется потенциальной функцией U . Скорость электрона v описывается равенством (см. [4], §§ 5.1, 5.2)

$$v = \frac{d\omega}{dk} = \hbar^{-1} \frac{dE}{dk} = \quad (2.9)$$

$$= \frac{dE}{dP}. \quad (2.10)$$

Лишь в том случае, когда $E = P^2/2m$, имеем $v = P/m$, как и для свободного электрона. Необходимо отметить одно весьма важное свойство волнового числа k , а также соответствующего ему квазиимпульса P . Величина k неоднозначно определяется волновой функцией выражения (2.7). Действительно, это выражение можно записать в виде

$$\psi(x) = \left\{ u_k(x) \exp \left[-\frac{2\pi i r x}{d} \right] \right\} \exp \left[i \left(k + \frac{2\pi r}{d} \right) x \right] = u_{k'}(x) \exp [i k' x], \quad (2.11)$$

где

$$k' = k + \frac{2\pi r}{d}. \quad (2.12)$$

Волновая функция, описываемая формулой (2.11), является решением уравнения Шредингера для того же значения энергии, что и волновая функция (2.7), причем $u_{k'}(x)$ тоже является периодической функцией с периодом d , так что k и k' полностью эквивалентны. Итак, энергия E является периодической функцией k с периодом $2\pi/d$. Это позволяет ограничить область изменения k интервалом $2\pi/d$, обычно от $-\pi/d$ до π/d , именуемым зоной Бриллюэна.

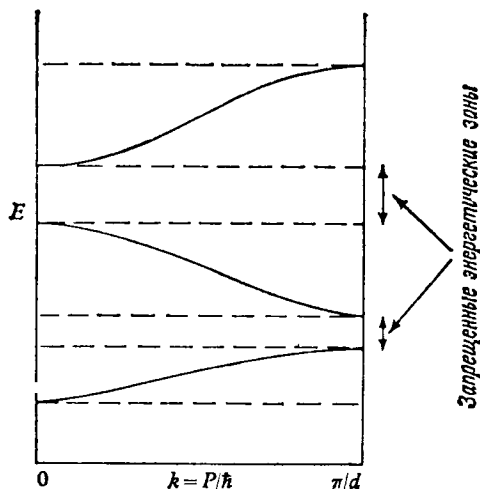


Рис. 2.1. Приведенное представление зависимости E от k .

Более того, если $U(x)$ — четная функция, как это и должно быть из соображений симметрии, то при изменении знака x энергия не изменится. Если мы теперь изменим и знак k , то мы вновь получим волновую функцию с той же энергией, что и ранее; следовательно, $E(k)$ — четная функция k . Итак, определяя энергию $E(k)$ как функцию k , мы можем еще больше ограничить область изменения k : $0 \leq k \leq \pi/d$. Такое представление носит название приведенного представления и иллюстрируется рис. 2.1¹⁾. При использовании приведенного представления энергия $E(k)$ оказывается многозначной функцией k , как это показано на рис. 2.1, причем каждая ветвь на этом рисунке соответствует определенной разрешенной зоне. Зоны, соответствующие одномерным задачам, не перекрываются и почти всегда отделены друг от друга запрещенными энергетическими зазорами. Именно вследствие такой особенности возможно качественное рассмотрение свойств полупроводников в рамках этой упрощенной теории, использованной нами в предыдущей главе.

Эти результаты можно легко обобщить на более реальный случай трехмерного кристалла. Для простоты предположим, что в кристалле можно выбрать три взаимно ортогональные оси (x , y , z) так, чтобы кристаллическая решетка была периодична в направлении этих осей. Это всегда можно сделать в случае кристаллов простой формы. Пусть (d_1, d_2, d_3) — периоды решетки вдоль осей

¹⁾ Соответственно, область изменения k , определяемая интервалами $-\pi/d \leq k \leq \pi/d$ или $0 \leq k \leq \pi/d$, называют приведенной зоной. — Прим. ред.

кристалла; иначе говоря, элементарная ячейка представляет собой параллелепипед со сторонами d_1 , d_2 и d_3 . Периодический потенциал в этом случае обладает свойством

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{d}), \quad (2.13)$$

где \mathbf{d} — любой вектор вида

$$\mathbf{d} = n_1 d_1 \mathbf{i} + n_2 d_2 \mathbf{j} + n_3 d_3 \mathbf{q},$$

n_1, n_2, n_3 — целые числа, $\mathbf{i}, \mathbf{j}, \mathbf{q}$ — единичные векторы вдоль осей x, y, z соответственно. Решение уравнения Шредингера можно записать в виде (см. [4], § 4.1.1, 4.3, 4.9)

$$\psi(x, y, z) = u_{\mathbf{k}}(x, y, z) \exp[i(\mathbf{k} \cdot \mathbf{r})], \quad (2.14)$$

где \mathbf{k} — постоянный вектор с компонентами (k_x, k_y, k_z) , $u_{\mathbf{k}}(x, y, z)$ — периодическая функция координат с периодом решетки. Энергия E зависит от k_x, k_y, k_z и по-прежнему является периодической функцией с периодами $2\pi/d_1, 2\pi/d_2$ и $2\pi/d_3$ по осям k_x, k_y и k_z соответственно. Поэтому достаточно ограничить область изменения \mathbf{k} в k -пространстве пределами одной зоны. Эта зона обычно выбирается в виде наименьшего объема в k -пространстве, централизованного в начале координат, который еще содержит все неэквивалентные значения \mathbf{k} . Форма этой зоны зависит от кристаллической структуры; эту зону обычно называют первой зоной Бриллюэна.

Вектор квазимпульса \mathbf{P} можно определить с помощью равенства (см. [4], § 5.1)

$$\mathbf{P} = \mathbf{k} \hbar; \quad (2.15)$$

тогда вектор скорости \mathbf{v} определяется формулой

$$\mathbf{v} = \hbar^{-1} \nabla_{\mathbf{k}} E(\mathbf{k}) = \nabla_{\mathbf{P}}(\mathbf{P}). \quad (2.16)$$

Значения \mathbf{P} отличаются от соответствующих значений \mathbf{k} лишь множителем \hbar , поэтому любая из этих величин может служить «квантовым числом», описывающим состояние электрона в кристалле.

Как правило, энергию нельзя уже представить в виде функции от одной переменной, за исключением специальных случаев, например, когда E зависит лишь от k , т. е. от величины вектора \mathbf{k} . Обычно зависимость E от \mathbf{k} изображают графически для некоторых главных направлений в кристалле.

При заданном направлении в кристалле соседние энергетические зоны, вообще говоря, всегда разделены запрещенными участками, однако минимум энергии в одной зоне может лежать ниже максимума энергии нижней зоны для другого направления в кристалле. Поэтому перекрытие соседних зон в трехмерных задачах — явление обычное. Кристалл будет обладать свойствами полупроводника лишь в том случае, если наивысшая заполненная зона и

наинишая пустая зона отделены друг от друга конечной зоной запрещенных значений энергии.

Можно показать, что в кристалле кубической формы с длиной ребра L , так же как и для случая свободных электронов, число разрешенных значений k_x в интервале $-\pi/d_1 < k_x \leq \pi/d_1$ равно L/d_1 , и аналогично для k_y, k_z . Полное число разрешенных значений \mathbf{k} поэтому равно $V/d_1 d_2 d_3$, где V — объем кристалла, что совпадает с числом элементарных ячеек кристалла. Результат верен для кристалла любой формы (см. [4], § 5.5). Если на каждую элементарную ячейку приходится по одному атому, то N равно числу атомов в кристалле. Таким образом, в кристаллах, состоящих из одновалентных атомов, электроны заполняют лишь половину валентной зоны, т. е. $N/2$ состояний, поскольку из-за наличия спина на каждом энергетическом уровне может находиться по два электрона. При наличии двух электронов на атом зона полностью заполняется. В общем случае при четном числе электронов на атом состоянии кристалла с минимально возможной энергией соответствуют полностью заполненные зоны. Вообще говоря, зависимость $E(\mathbf{k})$ имеет достаточно сложный вид. Однако в полупроводниках в отличие от металлов мы всегда будем иметь дело с почти пустой или почти заполненной зоной, так что для нас будет иметь значение вид зависимости $E(\mathbf{k})$ только вблизи краев зон, т. е. у дна зоны проводимости или у потолка валентной зоны.

2.3. Форма энергетических зон

Теоретический расчет функции $E(\mathbf{k})$ даже для простейших кристаллов представляет собой сложнейшую проблему теоретической физики. Сколько-нибудь подробные вычисления производились лишь для очень небольшого числа полупроводников. К этому вопросу мы вернемся в дальнейшем при рассмотрении конкретных полупроводников (см. гл. 13). Покажем теперь, что в одномерном случае на краях и в центре зон групповая скорость обращается в нуль. Рассмотрим точки, соответствующие $k=\pi/d$ и $k=-\pi/d$. Они разнесены на величину $2\pi/d$ и поэтому эквивалентны, так что волновая функция с $k=\pi/d$ должна описывать *стоячую* волну, так как в противном случае волновые функции с $k=\pi/d$ и $k=-\pi/d$ представляли бы собой две бегущие в противоположных направлениях волны и тогда соответствующие значения k не могли бы быть эквивалентными. Поскольку в стоячей волне нет переноса энергии, значения $k=\pi/d$ и $k=-\pi/d$ соответствуют состояниям электрона со скоростью, равной нулю; иными словами, групповая скорость $\partial\omega/\partial k=0$, следовательно, $\partial E/\partial k=0$. Аналогичные рассуждения позволяют доказать, что в симметричных точках на поверхности зоны Бриллюэна производная от энергии по нормали к поверхности

$\partial E/\partial k_n$ равна нулю. Итак, если графически изобразить зависимость E от k для любого из главных направлений в кристалле, которые всегда перпендикулярны поверхности зоны, то зависимость $E(k)$ на границе будет иметь вид, показанный на рис. 2.1. Исследуя уравнение Шредингера, можно показать, что $E(\mathbf{k})$ является непрерывной функцией \mathbf{k} вместе со своими первой и второй производными при условии, что E — однозначная функция \mathbf{k} , т. е. когда различные ветви не соприкасаются. Поскольку E — четная функция k , то отсюда следует, что вблизи точки $k=0$ она может содержать лишь квадратичные члены разложения по k . Следовательно, $\partial E/\partial k$ должно обратиться в нуль и при $k=0$, если только в этой точке нет вырождения. Тогда зависимость E от k вблизи точки $k=0$ должна иметь вид, показанный на рис. 2.1. Таким образом, в точке $k=0$ величина $E(k)$ должна иметь экстремум: максимум или минимум. В практически важных случаях экстремумы могут встретиться и в других точках. Если эти дополнительные экстремумы лежат в области промежуточных значений энергии, то они оказываются относительно несущественными. Если же они представляют собой абсолютный экстремум, т. е. наивысший максимум или наименьший минимум, то они приобретают первостепенное значение. При этом могут иметь место различные ситуации, которые будут рассмотрены ниже.

Самый простой вид зонной структуры имеет место тогда, когда наименьшая незаполненная зона имеет минимум в центре зоны, а вырождение отсутствует. Вблизи минимума энергии можно разложить $E(k_x, k_y, k_z)$ в ряд по степеням k_x, k_y, k_z в виде

$$E(k_x, k_y, k_z) = E_0 + Ak_x^2 + Bk_y^2 + Ck_z^2 + \text{члены более высокого порядка.} \quad (2.17)$$

где A, B, C — положительные числа. В частности, в кубическом кристалле с достаточным основанием можно принять $A=B=C$. Если при этом энергию отсчитывать от дна зоны, то для E получим следующее выражение, справедливое при малых значениях k :

$$E = \frac{\hbar^2}{2m_e} (k_x^2 + k_y^2 + k_z^2), \quad (2.18)$$

где m_e — постоянная, имеющая размерность массы. Значение m_e совпадает с массой свободного электрона лишь в том случае, когда $U(x) = \text{const}$. В терминах квазиимпульса (2.18) принимает следующий вид:

$$E = \frac{1}{2m_e} (P_x^2 + P_y^2 + P_z^2). \quad (2.19)$$

Равенства (2.18) и (2.19) отличаются от соответствующих зависимостей для свободных электронов лишь заменой m_0 на m_e . Таким образом, электроны ведут себя подобно свободным электронам, но с

другим значением массы. Величину m_e обычно называют эффективной массой. Такие зоны энергии обычно называют «сферическими» зонами, так как поверхности равной энергии в k -пространстве имеют вид сфер. Раньше считалось, что такой вид зон широко распространен, однако современные исследования показали, что он осуществляется довольно редко. Примером является зона проводимости в интерметаллическом соединении InSb . Прежде чем перейти к обсуждению других возможных вариантов, рассмотрим форму энергетических зон, когда в точке $k=0$ реализуется максимум. В этом случае при соответствующем выборе осей имеем

$$E = E'_0 - A'k_x^2 - B'k_y^2 - C'k_z^2. \quad (2.20)$$

При $A'=B'=C'$ это равенство можно записать еще и в следующем виде:

$$E = E'_0 - \frac{\hbar^2}{2m_h} (k_x^2 + k_y^2 + k_z^2). \quad (2.21)$$

Следовательно, электроны вблизи максимума энергии в зоне ведут себя так, как если бы это были свободные частицы с отрицательной эффективной массой. Ниже мы рассмотрим вытекающие отсюда важные следствия.

Вид энергетических зон при рассмотренных выше условиях показан на рис. 2.2. Энергия изображена как функция k_x , однако эта зависимость сохраняется и для любого другого направления в кристалле. В случае, когда верхняя зона представляет собой зону проводимости, а нижняя — валентную зону, имеет место простейшая из возможных зонных структур полупроводника. Ширина запрещенной зоны, которую мы обозначим через ΔE , равна $E_0 - E'_0$.

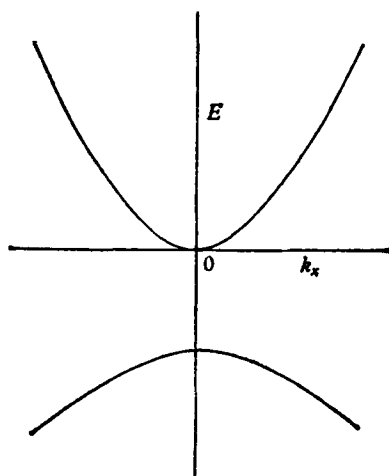


Рис. 2.2. «Сферические» энергетические зоны в центре зоны Бриллюэна.

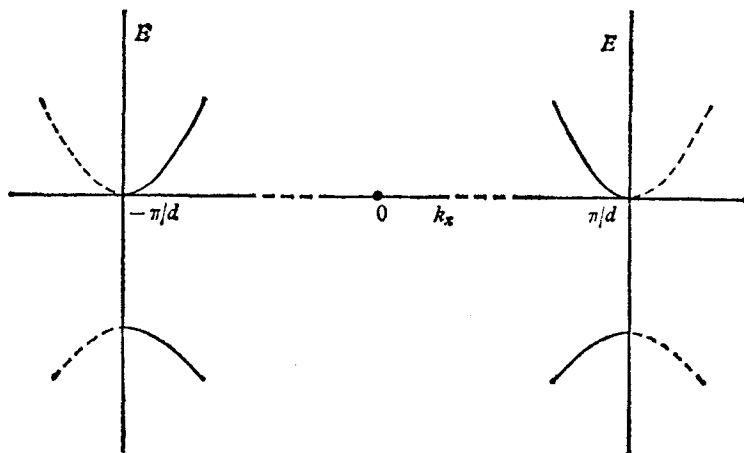


Рис. 2.3. Максимум или минимум энергии на границе зоны.

причем обычно $E'_0 < E_0$. К сожалению, до сих пор нельзя указать конкретный полупроводник, которому соответствовало бы такое строение энергетических зон.

Аналогичная ситуация имеет место тогда, когда точки экстремумов оказываются на поверхности зоны Бриллюэна. На рис. 2.3 показан пример, когда минимум энергии в зоне проводимости и максимум энергии в валентной зоне находятся на оси k_x у края зоны Бриллюэна. В этом случае изоэнергетические поверхности не обязательно должны быть сферически-симметричными, они должны лишь обладать цилиндрической симметрией относительно оси k_x , если речь идет о кубическом кристалле. Принимая в качестве начала отсчета энергию в минимуме, можно получить для энергии в окрестности минимума выражение

$$E = \frac{\hbar^2}{2} \left[\frac{\left(\frac{\pi}{d} - k_x \right)^2}{m_1} + \frac{(k_y^2 + k_z^2)}{m_2} \right]. \quad (2.22)$$

В данном случае имеются две константы m_1 и m_2 . Таким образом, эффективные массы электронов, движущихся по различным направлениям, различны. Вид сечений поверхностей равной энергии в плоскости (k_x, k_y) показан на рис. 2.4. Они имеют форму эллипсов. Пунктирные участки представляют собой продолжение эллипсоидов для значений k , лежащих вне зоны. Они эквивалентны эллипсоидам с центрами в точках $k_x = \pm \pi/d$, где d — период решетки. Если кристалл имеет кубическую симметрию, то должны существовать еще и другие эквивалентные минимумы, например вдоль осей k_y и k_z . Тогда в зоне проводимости имеются в общей сложности три

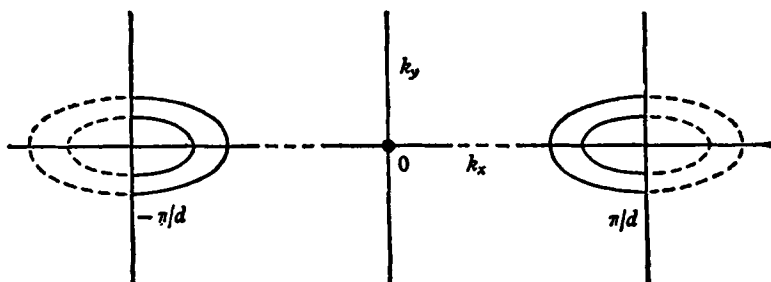


Рис. 2.4. Эллипсоидальные изоэнергетические поверхности для минимума энергии на границе зоны.

эквивалентных минимума в точках $(\pi/d, 0, 0)$, $(0, \pi/d, 0)$, $(0, 0, \pi/d)$ в k -пространстве. (Заметим, что минимум в точке $(-\pi/d, 0, 0)$ совпадает с минимумом в точке $(\pi/d, 0, 0)$ и т. д.) В кристаллах с кубической симметрией минимумы могут лежать и на других линиях симметрии, например в направлении $\langle 1, 1, 1 \rangle$, для которых

$$k_x = \pm k_y = \pm k_z.$$

Если к тому же минимумы располагаются на граничной поверхности зоны, то всегда должно быть четыре эквивалентных минимума. Такая структура зоны проводимости осуществляется в германии (см. разд. 13.3). Изоэнергетические поверхности в таких случаях многосвязны. В кристаллах с кубической симметрией для области изменения энергии вблизи минимумов они состоят из ряда эллипсоидов вращения. Центры эллипсоидов лежат при этом в симметричных точках, соответствующих минимальному значению энергии.

Предположим теперь, что абсолютные минимумы энергии расположены не на границе зоны, а в точке $(k_0, 0, 0)$ и других эквивалентных точках внутри зоны Бриллюэна, т. е. в области $0 < k_0 < \pi/d$. Это означает, что наряду с минимумом энергии в точке $(k_0, 0, 0)$ существует еще пять других, эквивалентных, но не сводящихся друг к другу минимумов в точках $(-k_0, 0, 0)$, $(0, \pm k_0, 0)$, $(0, 0, \pm k_0)$ — итого шесть минимумов. Вблизи $(k_0, 0, 0)$ энергию можно приближенно представить в виде

$$E = \frac{\hbar^2}{2} \left[\frac{(k_0 - k_x)^2}{m_1} + \frac{k_y^2 + k_z^2}{m_2} \right]. \quad (2.23)$$

В состоянии, соответствующем $(k_0, 0, 0)$ квазиимпульс отличен от нуля (за исключением случая $k_0=0$) и равен $k_0\hbar$. Однако скорость электрона в этом состоянии равна нулю, поскольку из уравнения

(2.16) следуют соотношения:

$$\left. \begin{aligned} v_x &= \frac{\hbar}{m_1} (k_x - k_0), \\ v_y &= \frac{\hbar}{m_2} k_y, \\ v_z &= \frac{\hbar}{m_2} k_z, \end{aligned} \right\} \quad (2.24)$$

в которых необходимо положить $k_x = k_0$, $k_y = 0$, $k_z = 0$.

Абсолютные минимумы энергии могут размещаться внутри зоны и на других осях симметрии. Так, в кремнии минимумы энергии для зоны проводимости расположены внутри зоны на осях $\langle 1, 0, 0 \rangle$. Таким образом, всего имеется шесть эквивалентных минимумов. Изоэнергетические поверхности в плоскости (k_x, k_y) для этого случая показаны на рис. 2.5, в. В дальнейшем при обсуждении конкретных полупроводников будут рассмотрены и другие возможные конфигурации зон. Некоторые примеры возможного строения энергетических зон показаны на рис. 2.5. Из формул (2.24) можно получить, что для значения квазиимпульса, соответствующего экстремуму энергии, скорость v равна нулю. Можно также показать, что при определенных условиях симметрии скорость v равна нулю и на краях зоны Бриллюэна [4], стр. 174).

Положение осложняется при наличии вырождения, т. е. когда в некоторой точке k -пространства энергетические зоны соприкасаются. Например, в валентной зоне Si и Ge имеются две почти параболические зоны, которые соприкасаются в центре зоны Бриллюэна, как показано на рис. 2.6, изображающем зависимость энергии от k_x . Каждому значению k_x соответствуют два значения энергии, причем при $k_x = 0$ они совпадают между собой. В этом случае простое разложение энергии E по четным степеням k_x, k_y, k_z невозможно. Выражение, описывающее зависимость E от k при малых k , имеет более сложный вид. В результате изоэнергетические поверхности приобретают вид искаженных сфер. Однако во многих случаях «вырожденные» зоны можно представить в виде двух «сферических» зон с эффективными массами дырок m_{h1} и m_{h2} . Обычно оказывается, что одна из эффективных масс намного меньше другой, и о них говорят как о «легких» и «тяжелых» дырках; ранние исследования их свойств в основном проводились на Si и Ge, которые имеют подобную зонную структуру (см. разд. 13.3).

Кристаллы с кубической симметрией, такие, как Si и Ge, обладающие кристаллической структурой алмаза или цинковой обманки, в действительности имеют зоны, которые *трехкратно* вырождены при $\mathbf{k} = 0$, если, конечно, пренебречь взаимодействием спина электрона с его орбитальным движением. Однако если это взаимодействие принять во внимание, то третья зона оказывается отщепленной от первых двух и становится по форме почти сферической

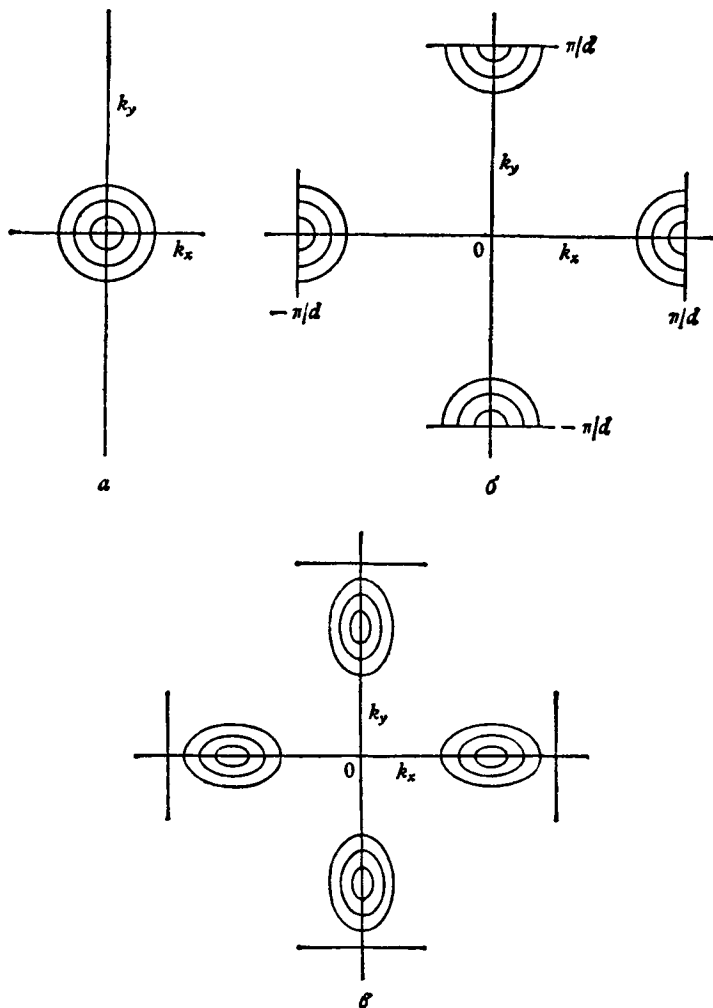


Рис. 2.5. Сечения изоэнергетических поверхностей плоскостью (k_x, k_y) .
 а — сферические поверхности при $k=0$; б — сферические поверхности при $k=\pi/d$ и т. д.; в — эллипсоидальные поверхности при $k=k_0 (< \pi/d)$ и т. д.

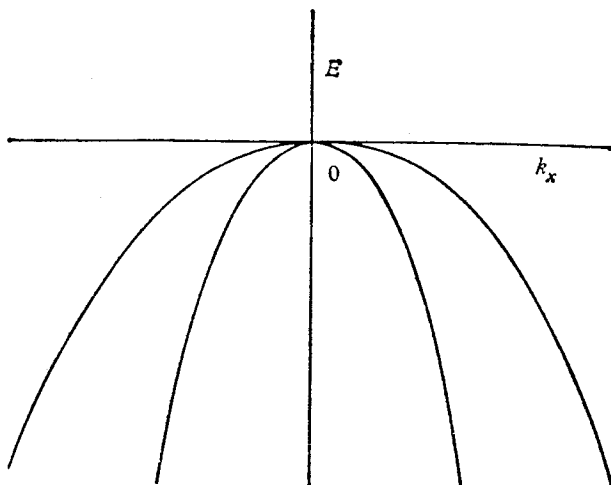


Рис. 2.6. Условие вырождения при $k=0$ для направления k_x .

с максимумом, лежащим на ΔE_s , ниже положения общего максимума двух других зон (см. [4], § 7.3.5). Эффективная масса дырки в этой зоне будет иметь еще одно значение $m_{\text{нз}}$. Энергию ΔE_s называют энергией спин-орбитального расщепления. Обычно ΔE_s заметно меньше, чем ширина запрещенной зоны ΔE . Однако в некоторых полупроводниках ΔE и ΔE_s имеют величины одного порядка, что, как мы увидим позже, придает таким материалам некоторые интересные свойства.

Из-за взаимодействия между вырожденными зонами изоэнергетические поверхности теряют сферическую форму, и для некоторых целей приближение на основе двух скалярных эффективных масс оказывается слишком грубым. Можно показать (см. [4], § 7.3.5), что вблизи $\mathbf{k}=0$ энергию можно представить в виде

$$E(k) = -\Delta E + \frac{\hbar^2}{2m_0} [Ak^2 \pm \{B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)\}^{1/2}], \quad (2.25)$$

где, как и обычно,

$$k^2 = k_x^2 + k_y^2 + k_z^2.$$

Если пренебречь членом, содержащим константу C , то можно заметить, что (2.25) сводится к двум отдельным сферическим энергетическим зонам со скалярными эффективными массами $m_0/(A+B)$ и $m_0/(A-B)$. Несколько лучшим приближением являются урав-

нения

$$\begin{aligned} \frac{m_0}{m_{h1}} &= A - \left(B^2 + \frac{C^2}{6} \right)^{1/2}, \\ \frac{m_0}{m_{h2}} &= A + \left(B^2 + \frac{C^2}{6} \right)^{1/2}. \end{aligned} \quad (2.26)$$

Это достаточно хорошее приближение для m_{h2} , при этом зона легких дырок оказывается почти совсем сферической, тогда как зона тяжелых дырок искажена ¹⁾ значительно сильнее [6].

В отсутствие спин-орбитального расщепления (см. разд. 12.6) валентные зоны Si и Ge, как отмечалось выше, были бы трехкратно вырождены при $k=0$; влияние спин-орбитального взаимодействия заключается в уменьшении энергии одной из зон и, следовательно, в отделении этой зоны от двух других зон, рассмотренных выше. Так как взаимодействие между этой зоной и двумя другими мало, то ее изоэнергетические поверхности оказываются почти сферическими. Для этой зоны получим

$$E(k) = -\Delta E - \Delta E_s - \frac{\hbar^2 k^2}{2m_0} (Ak^2), \quad (2.27)$$

где A та же самая константа, что и в (2.25); отсюда следует, что эффективная масса третьей дырки имеет величину

$$\frac{m_0}{m_{h3}} = A. \quad (2.26a)$$

Иной вид зонной структуры реализуется в халькогенидах свинца PbS, PbSe, PbTe, у которых минимум зоны проводимости и максимум валентной зоны расположены при одной и той же величине k , однако, отличной от нуля. В этих соединениях максимумы и минимумы лежат на краю зоны Бриллюэна в направлениях $\langle 111 \rangle$, так что существует четыре максимума и четыре минимума (см. рис. 13.7).

Упомянем еще один вид зонной структуры, хотя, строго говоря, к обычным полупроводникам он отношения не имеет. Может оказаться так, что вершина валентной зоны в точности совпадает с дном зоны проводимости и ширина запрещенной зоны в точности равна нулю. Такая структура осуществляется, например, в сером олове (α -Sn) и в HgTe. Структура (α -Sn) показана на рис. 13.5. В этом случае материал называют полуметаллом.

В отсутствие вырождения зон всегда можно выбрать оси координат так, чтобы выражение для E свелось к одной из форм, содержащих только квадраты компонент k . Однако если такого выбора системы координат не делать, то, разлагая функцию E вблизи экстре-

¹⁾ Относительно сферической зоны.— *Прим. ред.*

му, можно, вообще говоря, получить выражение вида

$$E = E_0 + \frac{\hbar^2}{2} \sum_{xyz} \frac{k_x k_y}{m_{xy}}. \quad (2.28)$$

В этой формуле члены первого порядка по \mathbf{k} отсутствуют, так как разложение производится в точке экстремума. Тензор $1/m_{xy}$ носит название тензора обратной эффективной массы. Он может быть записан в виде

$$\frac{1}{m_{xy}} = \hbar^{-2} \frac{\partial^2 E}{\partial k_x \partial k_y} \text{ и т. д.} \quad (2.28a)$$

2.4. Положительные дырки

В идеальном полупроводнике, когда все электроны находятся в наинизших энергетических состояниях, в зоне проводимости нет электронов. Однако такое положение теоретически возможно лишь при абсолютном нуле температуры. При обычных температурах в зоне проводимости всегда найдется некоторое количество электронов, заброшенных туда из заполненной зоны вследствие теплового возбуждения. Мгновенная плотность электрического тока \mathbf{I} , обусловленная движением какого-либо электрона, пропорциональна его скорости \mathbf{v} и совпадает с ней по направлению. Величину \mathbf{I} можно вычислить следующим образом. Пусть имеется N_v электронов (с вектором скорости, равным \mathbf{v}) в объеме V . Число электронов, пересекающих за единицу времени единичную площадку, перпендикулярную направлению скорости \mathbf{v} , определяется как $N_v |\mathbf{v}|/V$. Поэтому плотность тока равна $-e N_v |\mathbf{v}|/V$, а плотность тока \mathbf{I} , обусловленного движением одного электрона, может быть выражена в виде

$$\mathbf{I} = -e \frac{\mathbf{v}}{V}. \quad (2.29)$$

Предположим теперь, что речь идет об электроне, состояние которого характеризуется волновым вектором \mathbf{k} . Соответствующая скорость \mathbf{v} определяется равенством (2.16). Отметим, что скорость электрона в состоянии с волновым вектором $-\mathbf{k}$ равна $-\mathbf{v}$, поскольку энергия E является четной функцией \mathbf{k} . Полная плотность электрического тока для всей системы электронов равна

$$\mathbf{I} = - \left(\frac{e}{V} \right) \sum_{\mathbf{s}} \mathbf{v}_{\mathbf{s}}, \quad (2.30)$$

где векторы скорости $\mathbf{v}_{\mathbf{s}}$ соответствуют состояниям, занятым электронами. В частности, если электроны полностью заполняют какую-либо разрешенную зону, то соответствующий ток $\mathbf{I} = 0$, так как каждому значению $\mathbf{k}_{\mathbf{s}}$ в зоне соответствует волновой вектор $-\mathbf{k}_{\mathbf{s}}$,

порождающий равный по величине и обратный по направлению ток.

Предположим теперь, что вся зона полностью заполнена электронами, за исключением одного единственного состояния, характеризуемого волновым вектором \mathbf{k}_i , которому соответствует скорость \mathbf{v}_i . Суммарный ток всех электронов в зоне записывается в виде

$$\mathbf{I} = - \left(\frac{e}{V} \right) \sum_{s \neq i} \mathbf{v}_s = - \left(\frac{e}{V} \right) \sum_{\substack{\text{по всей} \\ \text{зоне}}} \mathbf{v}_s + \left(\frac{e}{V} \right) \mathbf{v}_i. \quad (2.31)$$

Первый член в правой части (2.31), как уже было сказано, равен нулю. Поэтому

$$\mathbf{I} = \left(\frac{e}{V} \right) \mathbf{v}_i. \quad (2.32)$$

Таким образом, суммарный ток всех электронов в зоне эквивалентен току одного электрона, если поместить последний в вакантное состояние и приписать ему *положительный* заряд $+e$. Такое вакантное состояние называется положительной дыркой или просто дыркой. Позже мы увидим, что положительной дырке можно приписать эффективную массу, равную по абсолютной величине эффективной массе того электрона, который занял бы это вакантное состояние. Ясно, что при наличии некоторого количества положительных дырок плотность соответствующего им электрического тока равна

$$\mathbf{I} = \left(\frac{e}{V} \right) \sum_i \mathbf{v}_i. \quad (2.33)$$

При наличии электронов и дырок, как это имеет место в собственном полупроводнике, когда взамен каждого электрона, перешедшего в зону проводимости, остается одна положительная дырка в валентной зоне, мгновенное значение плотности электрического тока определяется выражением

$$\mathbf{I} = - \left(\frac{e}{V} \right) \left\{ \sum_j \mathbf{v}_j - \sum_i \mathbf{v}_i \right\}, \quad (2.34)$$

где сумма по j относится к электронам, а сумма по i — к дыркам.

2.5. Движение электронов и дырок в кристалле под действием внешних силовых полей

Электрон, движущийся в идеальном кристалле, сохраняет свою скорость, когда он находится в состоянии, описываемом волновой функцией типа (2.7). В действительности скорость его будет испыты-

вать флуктуации, обусловленные полем кристалла, однако *наблюдаемая* скорость будет постоянной. Выясним теперь, как влияет наличие внешних полей на величину средней скорости. Рассмотрим сначала влияние электрического поля \mathcal{E} . Работа, совершаемая полем над электроном в единицу времени, равна

$$\frac{dE}{dt} = -e(\mathcal{E} \cdot \mathbf{v}). \quad (2.35)$$

Используя равенство (2.16) для \mathbf{v} , уравнение (2.35) можно записать в виде

$$\left(\nabla_{\mathbf{k}} E \cdot \frac{d\mathbf{k}}{dt} \right) = -e(\mathcal{E} \cdot \nabla_{\mathbf{k}} E) \hbar^{-1},$$

откуда имеем

$$\hbar \frac{d\mathbf{k}}{dt} = \frac{d\mathbf{P}}{dt} = -e\mathcal{E}. \quad (2.36)$$

Это одно из самых основных соотношений теории. Оно показывает, что под действием внешнего электрического поля квазиимпульс растет пропорционально напряженности приложенного поля. Однако вывод, приведенный выше, нельзя применить к случаю внешних полей, действующих на электрон с силой, перпендикулярной направлению скорости электрона \mathbf{v} , как, например, сила Лоренца в случае магнитного поля, поэтому здесь потребуется более сложное рассмотрение. Оказывается, что уравнение (2.36) имеет довольно общий характер и может быть записано в виде (см. [4], § 5.4)

$$\hbar \frac{d\mathbf{k}}{dt} = \frac{d\mathbf{P}}{dt} = \mathbf{F}, \quad (2.37)$$

где \mathbf{F} — сила действующая на электрон. Это объясняет, почему величина \mathbf{P} называется квазиимпульсом электрона в кристалле. Если наряду с электрическим полем имеется также и магнитное поле с индукцией \mathbf{B} , то

$$\mathbf{F} = -e\mathcal{E} - e(\mathbf{v} \times \mathbf{B}). \quad (2.38)$$

Дифференцируя равенство (2.16) для \mathbf{v} по времени, получаем

$$\frac{d\mathbf{v}}{dt} = \left(\nabla_{\mathbf{k}} \mathbf{v} \cdot \frac{d\mathbf{k}}{dt} \right) = \hbar^{-2} \nabla_{\mathbf{k}} (\nabla_{\mathbf{k}} E \cdot \mathbf{F}). \quad (2.39)$$

Тензор $\hbar^{-2} \nabla_{\mathbf{k}} \nabla_{\mathbf{k}} E$, который можно записать как $1/m_{\eta\zeta}$, где η, ζ обозначают попеременно x, y, z , называют тензором обратной эффективной массы. Компоненты этого тензора по-прежнему определяются формулами (2.28а), несмотря на то что они были введены для частных случаев сравнительно простой зависимости $E(\mathbf{k})$.

Векторные компоненты уравнения (2.39) можно записать в виде

$$\frac{dv_x}{dt} = \hbar^{-2} \left(\frac{\partial^2 E}{\partial k_x^2} F_x + \frac{\partial^2 E}{\partial k_x \partial k_y} F_y + \frac{\partial^2 E}{\partial k_x \partial k_z} F_z \right) \text{ и т. д.} \quad (2.40)$$

Если недиагональные компоненты тензора обратной эффективной массы равны нулю, то, обозначив его диагональные компоненты соответственно через $1/m_1$, $1/m_2$ и $1/m_3$, можно свести (2.40) к простому виду

$$\left. \begin{aligned} m_1 \frac{dv_x}{dt} &= F_x, \\ m_2 \frac{dv_y}{dt} &= F_y, \\ m_3 \frac{dv_z}{dt} &= F_z. \end{aligned} \right\} \quad (2.41)$$

При $m_1=m_2=m_3=m_e$ запись еще упрощается и мы получаем

$$m_e \frac{dv}{dt} = F. \quad (2.42)$$

В последнем случае имеется лишь одно значение эффективной массы, и электрон движется в силовом поле так, как если бы речь шла об обычной частице с массой m_e . В общем случае в зависимости от направления движения необходимо использовать соответствующее значение эффективной массы.

Исследуем теперь движение положительной дырки. Для начала рассмотрим движение какого-нибудь электрона в заполненной зоне. Если в момент времени t его скорость была v_i , а к моменту времени $t+\Delta t$ она стала равной $v_i+\Delta v_i$, то

$$\Delta v_i = \frac{dv_i}{dt} \Delta t, \quad (2.43)$$

где dv_i/dt определяется уравнением (2.39). Поэтому если бы рассматриваемое состояние было вакантным, то можно было бы считать, что в момент времени t отсутствует электрон, обладающий скоростью v_i , а к моменту времени $t+\Delta t$ отсутствует уже электрон, который имеет скорость $v_i+\Delta v_i$, т. е. имеется дырка со скоростью $v_i+\Delta v_i$. Таким образом, дырка движется с той же скоростью, с которой бы двигался электрон, занимающий это вакантное состояние. Однако вблизи потолка заполненной зоны диагональные компоненты тензора обратной эффективной массы электрона отрицательны, так что можно написать

$$-m_{\eta\zeta}^{-1} = \hbar^{-2} \nabla_{\zeta} \nabla_{\eta} E. \quad (2.44)$$

Назовем $1/m_{\eta\zeta}'$ тензором обратной эффективной массы дырки, теперь диагональные компоненты этого тензора положительны. Таким образом, в случае диагонального тензора для дырки имеем

$$v_x = - \frac{\partial E}{\partial P_x} = m_{\hbar}^{-1} P_x$$

и

$$m_h \frac{dv_x}{dt} = -F_x \text{ и т. д.} \quad (2.45)$$

где F_x является x -компонентой вектора силы, действующей на электрон. В случае электрических сил дырка ведет себя так, как если бы она обладала положительным зарядом $+e$ и эффективной массой, равной по величине и обратной по знаку эффективной массе отсутствующего электрона. При этом необходимо учесть, что эффективная масса электрона у потолка заполненной зоны отрицательна. Таким образом, дырка ведет себя как частица с положительной массой и положительным зарядом.

Резюмируя полученные выше результаты, можно сказать, что в идеальном кристалле электроны и дырки движутся под действием приложенных внешних полей соответственно как частицы с отрицательным и положительным зарядами и со свойственными им величинами эффективных масс. Вообще говоря, эффективные массы электронов и дырок отличаются друг от друга, поскольку они относятся к различным зонам. В сферически-симметричном случае, обозначив эффективные массы электронов и дырок соответственно через m_e и m_h , имеем

$$\frac{1}{m_e} = \hbar^{-2} \frac{\partial^2 E_c}{\partial k^2}, \quad \frac{1}{m_h} = -\hbar^2 \frac{\partial^2 E_v}{\partial k^2}, \quad (2.46)$$

где E_c и E_v — энергии электронов в зоне проводимости и в валентной зоне соответственно. Величины m_e и m_h обратно пропорциональны кривизне изоэнергетических поверхностей в \mathbf{k} -пространстве. Поэтому во многих случаях нет необходимости знать подробности кристаллической структуры; достаточно воспользоваться соответствующими величинами эффективных масс, которые можно определить, как мы увидим позже, из опыта.

Различные методы определения компонент тензора обратной эффективной массы будут в дальнейшем обсуждены более подробно, так как эффективная масса носителей заряда является одной из наиболее важных характеристик полупроводника. Если конкретный вид тензора обратной эффективной массы определен, то отсюда можно сделать много общих заключений относительно строения энергетических зон.

В качестве простого примера движения электрона в кристалле под действием внешнего силового поля рассмотрим вкратце самый прямой метод определения эффективной массы m_e для зоны со «сферической» симметрией. Предположим, что вдоль оси z создано постоянное магнитное поле с индукцией B . Уравнения движения элект-

рона на основании (2.42) имеют вид

$$\left. \begin{aligned} m_e \dot{v}_x &= -eBv_y, \\ m_e \dot{v}_y &= eBv_x, \\ \dot{v}_z &= 0. \end{aligned} \right\} \quad (2.47)$$

Решение этой системы уравнений хорошо известно. Если начальная составляющая скорости $v_z=0$, то частица описывает в плоскости (x, y) окружность. Обозначив через r радиус этой окружности, а через v — скорость движения по ней, имеем

$$m_e \frac{v^2}{r} = Bev,$$

или

$$v = B \frac{er}{m_e}. \quad (2.48)$$

Электрон обращается по окружности с частотой ν_c , которая дается формулой

$$\nu_c = \frac{Be}{2\pi m_e}. \quad (2.49)$$

Эта частота известна под названием циклотронной частоты. Она не зависит от радиуса описываемой окружности. Если, кроме магнитного поля, приложить также слабое радиочастотное электрическое поле, то при приближении частоты поля к ν_c электрон начнет поглощать энергию поля, увеличивая одновременно радиус своей круговой орбиты. Определение этой резонансной частоты позволяет измерить ν_c , а значит, и m_e . Такой метод непосредственного измерения m_e был впервые предложен Дорфманом [7], затем обсужден теоретически Динглем [8] и Шокли [9] и с успехом применен на опыте Дрессельхаузом и др. [10]. Эти опыты были проведены на германии. Было показано, что эффективные массы как электронов, так и дырок нельзя считать простыми скалярными величинами m_e . Точное определение тензора обратной эффективной массы для германия будет обсуждено подробнее в разд. 13.3.

2.6. Схема энергетических уровней

Часто бывает удобно представить энергетический спектр полупроводника с помощью схемы энергетических уровней, показанной на рис. 2.7. На этой схеме энергия электронов возрастает по вертикали вверх. На схеме указаны уровни в зоне проводимости. Ту же схему можно использовать и для описания энергии положительных дырок с той только разницей, что в этом случае энергию следует считать возрастающей по вертикали вниз. Это обстоятельство лег-

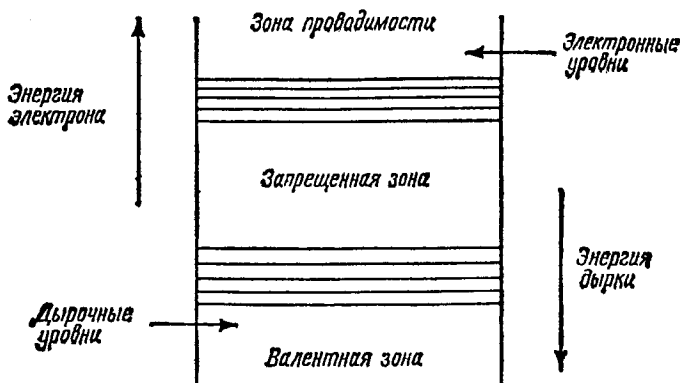


Рис. 2.7. Схема энергетических уровней в полупроводнике для электронов в зоне проводимости и дырок в валентной зоне.

че всего уяснить себе следующим образом. Рассмотрим в качестве примера вид энергетических зон, показанный на рис. 2.2. Пусть E_0 обозначает энергию электронов в полностью заполненной зоне, в которой зависимость E от k определяется соотношением $E = E_v(k)$, причем за начало отсчета энергии взята вершина заполненной зоны. При малых значениях k функцию $E_v(k)$ можно записать в виде:

$$E_v(k) = -\frac{\hbar^2}{2} \left(\frac{k_x^2}{m_1^2} + \frac{k_y^2}{m_2^2} + \frac{k_z^2}{m_3^2} \right). \quad (2.50)$$

Зависимость E от k в следующей зоне, т. е. в зоне проводимости, можно представить в виде $E = \Delta E + E_c(k)$. У дна зоны проводимости имеем

$$E_c(k) = \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_1^2} + \frac{k_y^2}{m_2^2} + \frac{k_z^2}{m_3^2} \right). \quad (2.51)$$

Предположим теперь, что один электрон изъят из валентной зоны, где он пребывал в состоянии с волновым вектором k' и переведен в зону проводимости в состояние, характеризуемое волновым вектором k . При этом будет создана электронно-дырочная пара. Такой процесс может быть осуществлен в результате поглощения электроном энергии колебаний решетки либо кванта излучения. Тогда энергия системы станет равной

$$E = E_0 + \Delta E + E_c(k) - E_v(k'), \quad (2.52)$$

где E_0 — энергия системы до образования пары.

Второй и третий члены в правой части (2.52), очевидно, представляют собой энергию электрона, последний член — энергию дырки. Необходимо отметить, что $-E_v(k')$ — положительная величина, так

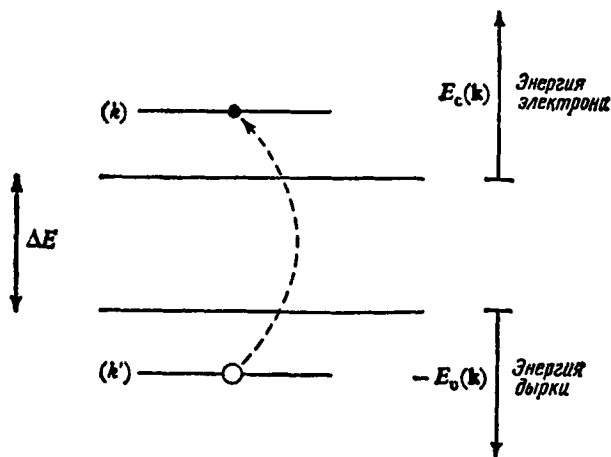


Рис. 2.8. Иллюстрация процесса образования электронно-дырочной пары.

как начало отсчета для $E_v(\mathbf{k})$ взято в максимуме валентной зоны. Таким образом, можно использовать одну и ту же схему энергетических уровней как для электронов, так и для дырок при условии, что энергия дырок будет отсчитываться в направлении, противоположном направлению отсчета энергии электронов (рис. 2.8). Воспользовавшись приближенными выражениями $E(\mathbf{k})$, можно представить изменение энергии кристалла в виде

$$E - E_0 = \Delta E + \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_1} + \frac{k_y^2}{m_2} + \frac{k_z^2}{m_3} \right) + \frac{\hbar^2}{2} \left(\frac{k_x'^2}{m_1'} + \frac{k_y'^2}{m_2'} + \frac{k_z'^2}{m_3'} \right). \quad (2.53)$$

Здесь ΔE — минимальная энергия, необходимая для перевода электрона из валентной зоны в зону проводимости. Как будет видно из дальнейшего, эту величину можно измерить разными способами, причем простейшим из них в принципе является метод измерения длинноволнового края поглощения света, который еще вызывает электронные переходы между зонами. Энергия ΔE — самый важный из всех параметров, характеризующих полупроводник, поскольку величина ΔE в большой мере определяет электропроводность полупроводников и ее температурную зависимость. Второй и третий члены в (2.53) могут, очевидно, рассматриваться как кинетические энергии электрона и дырки. Это равенство можно записать также и через квазиимпульс \mathbf{P} :

$$E - E_0 = \Delta E + \frac{1}{2} \left(\frac{P_x^2}{m_1} + \frac{P_y^2}{m_2} + \frac{P_z^2}{m_3} \right) + \frac{1}{2} \left(\frac{P_x'^2}{m_1'} + \frac{P_y'^2}{m_2'} + \frac{P_z'^2}{m_3'} \right). \quad (2.54)$$

Если $m_1=m_2=m_3=m_e$ и $m'_1=m'_2=m'_3=m_h$,

$$E - E_0 = \Delta E + \frac{1}{2} \frac{P^2}{m_e} + \frac{1}{2} \frac{P'^2}{m_h}, \quad (2.55)$$

где P и P' — абсолютные величины векторов квазиимпульса для электрона и дырки. Сходство этого равенства с соответствующим классическим выражением очевидно.

2.7. Влияние нарушений периодичности кристаллической решетки на движение электронов и дырок в кристалле

До сих пор мы пренебрегали влиянием примесей и дефектов в кристаллах, а также тепловых колебаний решетки, приводящих к смещению атомов кристалла из идеально правильных положений в узлах решетки, на движение электронов и дырок. Все эти факторы нарушают стационарность движения электронов и дырок как квазисвободных частиц, т. е. вызывают рассеяние. Эффект рассеяния можно описывать либо в терминах изменения во времени волнового вектора \mathbf{k} или скорости \mathbf{v} частиц, либо на языке теории столкновений. К этому кругу вопросов мы еще вернемся после того, как будет обсуждена более детально природа примесей и дефектов в кристаллах.

Как мы увидим в дальнейшем, выводы этой главы дают нам возможность обсуждать процесс движения электронов и дырок в почти идеальных кристаллах. В большинстве практических применений полупроводников в современной электронике используемые материалы являются кристаллами высокого качества. В таких полупроводниках электроны и дырки обладают высокой подвижностью и большим временем жизни в возбужденном состоянии, так что при воздействии на кристалл электрического и магнитного полей, так же как светового потока или температурных градиентов, они могут передвигаться на значительные расстояния. Однако такое движение не является необходимым для *всех* применений полупроводников: аморфные и поликристаллические материалы также нашли себе важные применения, к тому же они значительно дешевле. Мы убедемся далее, что хотя некоторые из описанных выше понятий останутся в силе для таких материалов, тем не менее для объяснения явлений электронного переноса в них придется ввести некоторые новые представления, например прыжковый характер движения электрона в кристалле.

ЛИТЕРАТУРА

1. *Wilson A. H.*, The Theory of Metals, Cambridge University Press, 1953.
2. *Mott N. F., Jones H.*, The Theory of the Properties of Metals and Alloys, Oxford University Press, 1936.
3. *Scits F.*, Modern Theory of Solids, McGraw-Hill, 1940.
4. *Smith R. A.*, The Wave Mechanics of Crystalline Solids, 2nd Ed., Chapman and Hall, 1969.
5. *Bloch F.*, Zs. Phys., **52**, 555 (1928).
6. *Herman F.*, Proc. IRE, **40**, 1703 (1955).
7. Дорфман Я. Г., ДАН СССР, **81**, 765 (1951).
8. *Dingle R. B.*, Proc. Roy. Soc., **A212**, 38 (1952).
9. *Shockley W.*, Phys. Rev., **90**, 491 (1953).
10. *Dresselhaus G., Kip A. F., Kittel C.*, Phys. Rev., **92**, 827 (1953).

ПРИМЕСИ И НЕСОВЕРШЕНСТВА В КРИСТАЛЛАХ

3.1. Типы несовершенств решетки

В предыдущей главе мы исходили из того, что атомы в кристаллах занимают положения, соответствующие идеальной периодической решетке. Однако в реальных кристаллах это условие, вообще говоря, не выполняется. Мы отложим на некоторое время обсуждение тепловых колебаний атомов в кристаллической решетке около своих равновесных положений, вызывающих, вполне очевидно, нарушения *идеальной* периодичности в кристалле, и рассмотрим сначала другие типы нарушений — дефекты в кристалле.

3.1.1. ПРИМЕСИ

Наиболее очевидный тип дефектов в кристаллах обусловлен присутствием чужеродных атомов — примесей. Несмотря на крупные успехи, достигнутые в области технологии очистки материалов, главным образом полупроводников, полностью очистить какое бы то ни было вещество от всех содержащихся в нем примесей невозможно. Даже предельно очищенные кристаллы содержат еще до-

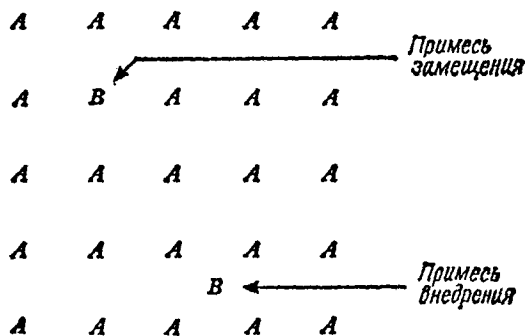


Рис. 3.1. Примеси в одноатомном кристалле,

статочны много чужеродных атомов; так, содержание примесей в отношении $1 : 10^9$ означает наличие около 10^{14} примесных атомов в 1 см^3 кристалла, так как в большинстве кристаллов содержится от 10^{22} до 10^{23} атомов на 1 см^3 . Основную часть атомов, из которых состоит кристалл, будем называть «основными атомами кристалла». Чужеродные атомы обычно называют примесями, и, как мы увидим ниже, они играют первостепенную роль, поскольку многие свойства полупроводников обусловлены именно наличием в них примесей. Существует два рода примесей: примесь замещения, когда атомы примеси заменяют основные атомы кристалла в узлах кристаллической решетки, и примесь внедрения, когда примесные атомы занимают положения между узлами. Оба рода примесей проиллюстрированы на рис. 3.1, изображающем кристалл, построенный из одинаковых атомов типа *A* и содержащий небольшое количество примесных атомов типа *B*.

3.1.2. МЕЖДУУЗЕЛЬНЫЕ АТОМЫ И ВАКАНСИИ

Если один из основных атомов кристалла перемещается из узла кристаллической решетки в промежутки между узлами, то получается новый тип дефекта (рис. 3.2), который носит название «междуузельного атома» (чтобы отличить его от случая примеси внедрения). При этом один из узлов решетки может остаться незаполненным, что представляет собой также определенный тип несовершенства решетки, называемый «вакансией» (рис. 3.2). Естественно, могут возникнуть и более сложные конфигурации дефектов, например, две или более вакансии могут разместиться в соседних узлах кристаллической решетки.

Атомы, покинувшие свои места в глубине кристалла, могут переместиться по направлению к его поверхности и занять положения, соответствующие узлам решетки на поверхности. Это приведет к увеличению объема кристалла. В этом случае в кристалле

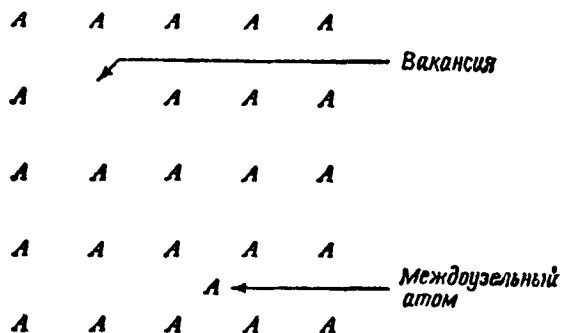


Рис. 3.2. Междуузельные атомы и вакансии в одноатомном кристалле.

остаются вакантные узлы без соответствующего числа междоузельных атомов. Такие несовершенства известны как дефекты по Шоттки. С другой стороны, вакансии могут образовывать пары с междоузельными атомами. Такие несовершенства носят название дефектов по Френкелю. Из сказанного следует, что число вакансий в кристалле не обязательно совпадает с числом атомов в междоузлиях. Все перечисленные выше типы дефектов могут отличаться по величине и знаку электрического заряда, т. е. при известных условиях они могут захватить один или большее число электронов или положительных дырок и могут существовать в различных состояниях.

При более высоких температурах все эти дефекты становятся достаточно подвижными, в результате чего между различными типами дефектов легко устанавливается термодинамическое равновесие. Чем меньше энергия образования дефекта, тем больше число дефектов данного типа. Например, если энергия образования дефекта по Шоттки много меньше энергии образования дефекта по Френкелю, то первый тип дефектов будет преобладающим и число атомов в междоузлиях окажется сравнительно небольшим. В кристаллах с плотной упаковкой атомов появление междоузельного атома, как правило, затруднено. Примесные атомы малых размеров сравнительно легко внедряются в междоузлия, тогда как более крупные атомы могут образовывать дефекты такого рода лишь с трудом, и, как правило, являются примесями замещения. Обычно один из типов дефектов по концентрации преобладает над всеми остальными.

Как уже было сказано, при высоких температурах, близких к точке плавления, между различными типами дефектов устанавливается термодинамическое равновесие. Однако время установления равновесия между дефектами сильно возрастает при понижении температуры. В этих условиях возможно «вмораживание» дефектов и тогда количество и род дефектов, имеющих в кристалле, будет зависеть не только от температуры, но и от предыдущей тепловой обработки кристалла. Выдержав кристалл некоторое время при промежуточной температуре, можно в значительной мере уменьшить количество «вмороженных» дефектов. Этот процесс называют отжигом.

В условиях теплового равновесия число дефектов определенного типа можно вычислить с помощью методов статистической физики, если известна энергия образования этих дефектов. С другой стороны, энергию образования дефектов легко определить, если тем или иным путем найдена температурная зависимость концентрации дефектов в кристалле. В гл. 2 книги Мотта и Герни [1] приведена термодинамическая теория образования дефектов по Шоттки и по Френкелю с привлечением понятия свободной энергии образования дефектов. Из теории следует, например, что когда дефекты по Шоттки преобладают над другими видами дефектов, количество вакансий

N_v можно вычислить по формуле

$$N_v = NF \exp \left[- \frac{W_v}{kT} \right], \quad (3.1)$$

где N — число атомов в кристалле, W_v — энергия образования вакансий, F — медленно меняющаяся с температурой функция, численные значения которой по порядку величины лежат в пределах от 10^3 до 10^4 . Величина W_v обычно имеет значения порядка $1 \div 2$ эВ, так что при обычных температурах в условиях равновесия N_v ничтожно мало, но быстро растет с повышением температуры.

3.1.3. ДИСЛОКАЦИИ

Перечисленные выше несовершенства решетки в большинстве случаев представляют собой точечные дефекты. Небольшое количество близко расположенных дефектов может, правда, группироваться и образовывать небольшие скопления (кластеры) или короткие цепочки дефектов, однако обычно эти дефекты четко локализованы. Рассмотрим теперь линейные (часто прямолинейные) дефекты, которые пронизывают кристалл. Такие дефекты называют дислокациями. В кристалле, подвергнутом механическим напряжениям, вообще говоря, имеют место относительные сдвиги кристаллических плоскостей. Кристаллическая плоскость не сдвигается как целое относительно других плоскостей. Сдвиг происходит внутри некоторого участка, который постепенно распространяется по плоскости. Линию раздела сдвинутой и несдвинутой плоскостей называют дислокацией. Если направление скольжения составляет пря-

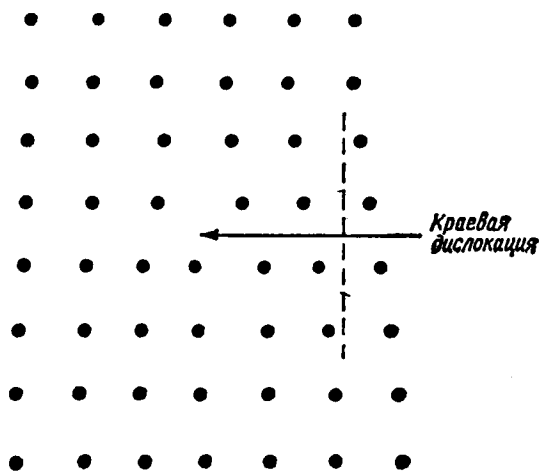


Рис. 3.3. Краевая дислокация.

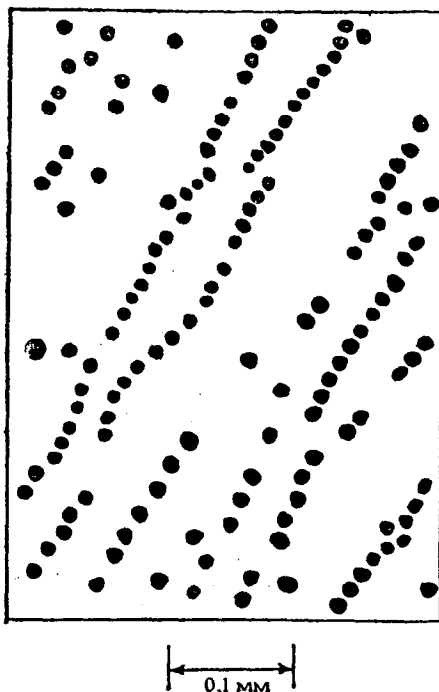


Рис. 3.4. Ямки травления у выхода краевых дислокаций на поверхность кристалла. (С микрофотографии, полученной в работе [7]).

мой угол с осью дислокации, то такую дислокацию называют краевой. Можно показать, что краевая дислокация эквивалентна введению лишней полуплоскости атомов (рис. 3.3). Если направление скольжения совпадает с направлением дислокации, то дислокация носит название винтовой. Рассмотрение весьма интересных вопросов геометрической теории дислокаций выходит за рамки настоящей книги. Теория дислокаций занимает центральное место во всех современных работах по сопротивлению материалов. Она подробно освещается в ряде учебников [2—4]. Дислокации сказываются и на физических свойствах полупроводников; эти вопросы будут обсуждены далее в разд. 8.11.

У места выхода линии дислокации на поверхность кристалла правильное расположение атомов на поверхности нарушается. Эти поверхностные дефекты можно подвергнуть избирательному действию некоторых травителей. При соответствующих условиях в местах выхода дислокаций на поверхность кристалла после травления образуются небольшие углубления, так называемые ямки травления, которые легко наблюдаются в микроскоп даже при не-

большом увеличении. На рис. 3.4 показаны такие ямки травления, полученные от краевых дислокаций в InSb. Число дислокаций, пересекающих единицу поверхности кристалла, может служить мерой его совершенства (см. разд. 8.11).

Полное описание свойств дислокаций в кристаллах со структурой алмаза, в особенности в Si и в Ge, сделали Александер и Хаасен [5]. Полупроводники типа InSb, имеющие кристаллическую структуру цинковой обманки, являются более сложными; типы дислокаций, которые оказывают существенное влияние на электрические свойства таких полупроводников, рассмотрели Гатос и др. [6].

3.1.4. УПОРЯДОЧЕНИЕ ДИСЛОКАЦИЙ

Распределение дислокаций в объеме кристалла может быть, вообще говоря, полностью хаотичным. Было найдено, однако, что в кристаллах, подвергавшихся отжигу, дислокации очень часто стремятся сгруппироваться, выстраиваясь в виде дислокационных «стенок». Иногда такие дислокационные стенки в свою очередь образуют трехмерную сетку, охватывающую весь кристалл. Сечение этой структуры плоскостью представляет собой сетку многоугольников, как было наглядно показано на примере бромида серебра в работе Митчелла [8], использовавшего для этой цели фотолитическое осаждение серебра на дислокациях. Дислокационные стенки можно видеть на рис. 3.4. Весьма интересная разновидность параллельных краевых дислокаций возникает на границах слабо разориентированных относительно друг друга зерен кристалла. Она является обычной формой несколько более крупных несовершенств существующих в кристаллах. В этом случае атомные плоскости двух соседних зерен кристалла, вместо того чтобы быть параллельными, как в идеальном кристалле, образуют между собой небольшой угол θ , другими словами, малоугловую границу (рис. 3.5). Это эквивалентно введению серии дополнительных атомных полуплоскостей, что в конечном итоге эквивалентно появлению ряда параллельно расположенных краевых дислокаций (рис. 3.6). Расстояния d между дислокациями при малых углах θ равны a/θ , где a — межатомное расстояние в направлении скольжения. Тогда число дислокаций, приходится на единицу длины границы между зернами, равно θ/a .

Если прямой кристаллический стержень прямоугольного сечения изогнуть по окружности с большим радиусом кривизны, то деформированный кристалл может рассматриваться как состоящий из большого числа слегка разориентированных относительно друг друга кристаллов. Таким образом, деформация изгиба порождает множество краевых дислокаций, параллельных оси изгиба. В этом случае дислокации должны распределяться равномерно по кристал-

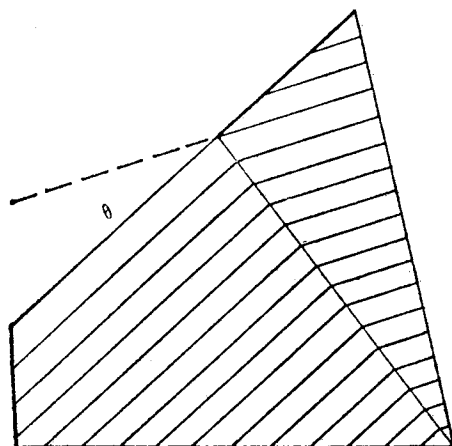


Рис. 3.5. «Малуюгловая» граница зерен.

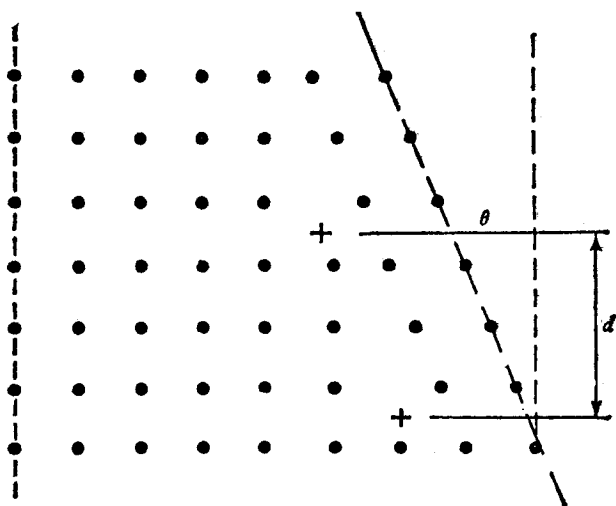


Рис. 3.6. Дислокация у границ зерен.

ду. Их количество может быть сосчитано, что позволяет проверить надежность метода травления, обычно используемого для подсчета числа дислокаций. Деформация изгиба может служить также для изготовления образцов с заданным числом дислокаций, параллельных одной и той же оси в кристалле (см. разд. 8.11). Монокристаллы часто состоят из большого числа блоков, весьма слабо разориентированных относительно друг друга. Если же углы между эквива-

лентными атомными плоскостями на границе блоков не очень малы, то кристалл перестает быть монокристаллом и его следует рассматривать скорее как поликристаллическое тело, состоящее из отдельных кристаллических зерен. В предельном случае взаимная ориентация зерен в поликристалле может быть совершенно случайной. Между отдельными зернами могут даже существовать микротрещины. Многие из ранних работ, посвященных полупроводникам, были выполнены именно с такими несовершенными материалами. В современных работах стремятся по возможности исследовать монокристаллы без заметных границ между блоками.

3.2. Химическая связь в полупроводниках

Прежде чем приступить к более подробному исследованию примесей и дефектов в полупроводниках, полезно рассмотреть различные виды химической связи между атомами, удерживающей их в кристаллах. Теория химической связи представляет собой один из возможных путей трактовки проблемы электронной структуры твердых тел. Однако мы ограничимся лишь некоторыми элементарными соображениями, которые проливают свет на довольно характерные свойства полупроводников.

3.2.1. ИОННАЯ СВЯЗЬ

Образование замкнутых электронных оболочек играет главную роль как в теории строения атома, так и в теории химической связи в твердых телах. Важнейшим примером замкнутой электронной оболочки являются внешние оболочки благородных газов, таких, как Ne, Ar, Kr и т. д., состоящие из восьми электронов: двух электронов в *s*-состоянии и шести электронов в *p*-состоянии. В таких оболочках все разрешенные *s*- и *p*-уровни заполнены, и на каждом уровне содержится по два электрона с противоположными спинами. Такие восьмиэлектронные замкнутые оболочки образуют очень устойчивые конфигурации с весьма низкой энергией. Поэтому неудивительно, что даже при сближении атомов у них наблюдается тенденция к сохранению таких замкнутых оболочек. Эти общие соображения позволяют понять простейший вид химической связи — ионную связь, которая проявляется, например, в кристаллах галогенидов щелочных металлов. Рассмотрим для примера такое вещество, как NaCl. Внешняя оболочка каждого атома Na содержит всего один электрон, тогда как в каждом атоме Cl число таких электронов равно семи. При сближении атомов Na и Cl у единственного валентного электрона Na появляется резко выраженная тенденция к переходу на атом Cl, после чего у последнего образуется замкнутая оболочка из восьми электронов. В результате такого перехода

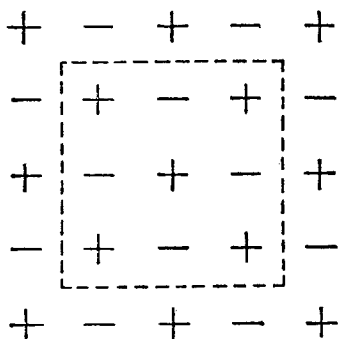


Рис. 3.7. Расположение ионов в ионном кристалле в плоскости {100}.

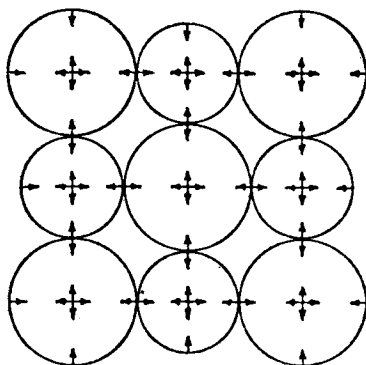


Рис. 3.8. Силы взаимного притяжения и отталкивания между ионами в ионном кристалле.

возникают положительные ионы Na^+ и отрицательные ионы Cl^- , между которыми действуют силы притяжения обычной электростатической природы. Можно показать, что электростатическими силами обусловлена большая часть энергии связи в кристалле. Большинство ионных соединений кристаллизуется либо в гранецентрированной решетке (как, например, NaCl), либо в объемноцентрированной решетке (как, например, CsCl). Зная кристаллическую структуру вещества, можно вычислить соответствующую ей энергию связи [9]. Расположение атомов в одной из главных плоскостей типичного ионного кристалла показано на рис. 3.7.

Вполне естественно, что наряду с чисто электростатическими силами, которые стремятся удержать атомы в кристалле, имеются еще и силы отталкивания, уравновешивающие электростатические силы притяжения и обеспечивающие вместе с первыми условия устойчивого равновесия кристаллической решетки. Представление о замкнутых электронных оболочках позволяет объяснить также и природу сил отталкивания. При достаточном сближении ионов имеет место существенное взаимное проникновение соответствующих замкнутых электронных оболочек. Однако оболочки должны энергично противодействовать этому взаимному проникновению, так как в каждой из них все состояния заняты, и поэтому, согласно принципу Паули, в них нет места для добавочных электронов. Силы отталкивания гораздо быстрее убывают при увеличении расстояния между ионами, чем силы электростатического взаимодействия, поэтому с достаточной степенью точности ионы могут рассматриваться как непроницаемые шарики определенного радиуса. Таким образом, мы приходим к понятию ионных радиусов. Расстояния между центрами двух ближайших ионов в кристалле по определе-

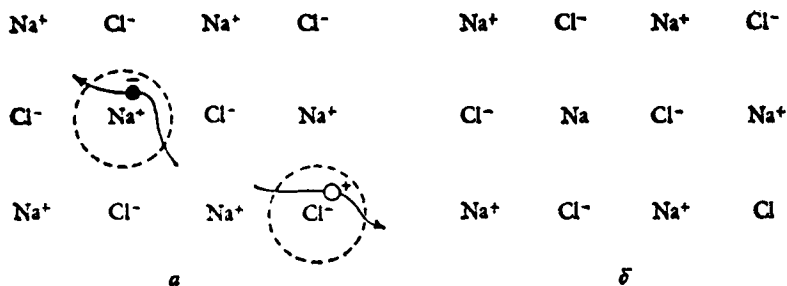


Рис. 3.9. Образование электронно-дырочной пары в ионном кристалле (NaCl).

нию равно сумме соответствующих ионных радиусов. Это проиллюстрировано на рис. 3.8.

Главная отличительная особенность ионных кристаллов состоит в том, что они, как правило, прозрачны в видимой области спектра. При высоких температурах они обладают небольшой электропроводностью, обусловленной переносом ионов, но с точки зрения электронной теории должны рассматриваться как диэлектрики, так как наивысшая заполненная зона отделена от следующей пустой зоны большим энергетическим зазором (валентная зона должна быть полностью заполненной, поскольку на один ион приходится восемь электронов). Кристалл начинает сильно поглощать свет лишь в ультрафиолетовой части спектра. Освещение кристаллов светом такой длины волны сопровождается появлением фотопроводимости. На языке химической связи процесс образования свободного электрона сводится к тому, чтобы оторвать электрон у отрицательного иона и перевести его на какой-либо далеко расположенный положительный ион. Образованные в результате такого перехода нейтральные атомы щелочного металла следует рассматривать как центры локализации свободных электронов. Свободный электрон без каких-либо дополнительных энергетических затрат может совершать переходы от одного иона щелочного металла к другому и, таким образом, может свободно перемещаться по всему кристаллу. Этот процесс описывается рассмотренными в предыдущей главе блоховскими волновыми функциями.

Образовавшийся вследствие ухода электрона нейтральный атом галогена может рассматриваться как положительная дырка, которая в принципе также может свободно перемещаться по кристаллу. Однако на опыте обнаружено, что дырки, как правило, обладают значительно меньшей способностью к перемещению, чем свободные электроны. На рис. 3.9 с помощью двух различных схем описания состояния кристалла показан процесс образования свободного электрона и дырки в NaCl.

3.2.2. ГОМЕОПОЛЯРНАЯ СВЯЗЬ

Образование полярных молекул типа HCl можно в известной мере объяснить теми же механизмами, которые приводят к образованию ионных кристаллов, однако существование неполярных молекул типа H_2 требует для своей интерпретации привлечения совершенно новых представлений. До возникновения квантовой механики это была одна из труднейших проблем структурной химии и ее разрешение представляло в свое время выдающийся успех теории. В своей знаменитой работе, посвященной молекуле водорода, Гайтлер и Лондон [10] показали, что перераспределение двух электронов между двумя одинаковыми атомами приводит к сильной связи, если спины электронов антипараллельны, и сильному отталкиванию, если спины электронов параллельны. Возникающие в первом случае силы связи в своей основе являются электростатическими, но имеют чисто квантовомеханическую природу, обусловленную неразличимостью тождественных частиц. Представление о двухэлектронной или гомеополярной связи имеет в настоящее время фундаментальное значение в органической структурной химии. Возможна также гомеополярная связь, обусловленная перераспределением плотности заряда одного электрона между двумя атомами, как, например, в молекулярном ионе H_2^+ . Однако, как правило, одноэлектронная связь слабее двухэлектронной.

Используя представления о двухэлектронной гомеополярной связи и воспользовавшись приведенными выше соображениями относительно тенденции к образованию восьмиелектронных замкнутых оболочек, легко понять природу химической связи в очень важной группе полупроводников, а именно в одноатомных полупроводниках: кремнии, германии, сером олове, а также в алмазе. Все эти элементы принадлежат к четвертой группе периодической системы Менделеева и в соответствии с этим в атомарном состоянии имеют по четыре валентных электрона: два электрона в s -состоянии и два в p -состоянии. При образовании твердого тела валентные электроны распределяются между соседними атомами кристалла. Если число ближайших соседей у каждого атома равно *четырем*, то с учетом перераспределения электронов каждый атом окружен группой из восьми электронов, так что образуется весьма стабильная структура. Схематически эта ситуация изображена на рис. 3.10. Каждая линия связи на этом рисунке представляет собой распределенный между двумя атомами электрон. На связях размещаются электроны с противоположными спинами, так что в каждой паре суммарный спин равен нулю. Такой тип связи иногда называют ковалентной, а рассматриваемую группу полупроводников — полупроводниками с ковалентной связью. Рис. 3.10 на самом деле дает лишь схематическое, двумерное представление о структуре связи. В реальных кристаллах связи образуют, естествен-

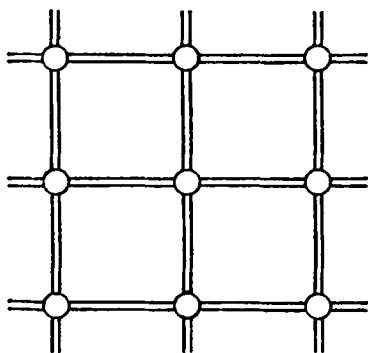


Рис. 3.10. Схема связей в кристаллах, состоящих из четырехвалентных атомов.

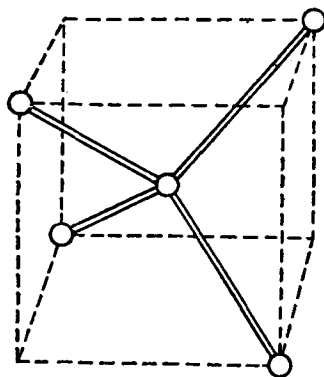


Рис. 3.11. Тетраэдрическое расположение связей.

но, не двумерную, а трехмерную сетку. Вместе с тем на рис. 3.10 совершенно правильно отражено то обстоятельство, что у каждого атома в кристалле имеется четыре симметрично расположенных ближайших соседа. Пространственная структура связей в этих кристаллах такова, что четыре ближайших соседа данного атома размещены по углам правильного тетраэдра, причем сам атом занимает центр тетраэдра, как показано на рис. 3.11. Таким путем можно построить правильную кристаллическую решетку, известную под названием решетки типа алмаза. Ее имеют, в частности, кристаллы Ge и Si (см. [11], § 3.8.1).

Теперь необходимо выяснить вопрос о связи этой наглядной картины сил сцепления в кристалле со структурой зоны проводимости и валентной зоны. Совершенно очевидно, что валентная зона соответствует состояниям валентных электронов, образующих связи между атомами в кристалле. Для примера рассмотрим кристалл Ge. Если удалить один электрон из какой-либо связи, то образовавшаяся электронная вакансия может свободно перемещаться по кристаллу, переходя от одного атома к другому без изменения энергии. Движение электронной вакансии равносильно перемещению по кристаллической решетке ионного состояния Ge^+ , т. е. положительного заряда $+e$, а это соответствует движению положительной дырки в валентной зоне. Электрон, вырванный из связи и помещенный на удаленный от нее атом Ge, эквивалентен свободному электрону в зоне проводимости. Его движение соответствует перемещению по кристаллу ионного состояния Ge^- . С точки зрения представлений о химической связи ширина запрещенной зоны это не что иное, как минимальная энергия, которую необходимо затратить для того, чтобы вырвать электрон из связи и перевести его на удаленный атом Ge, где он окажется сверхкомплектным от-

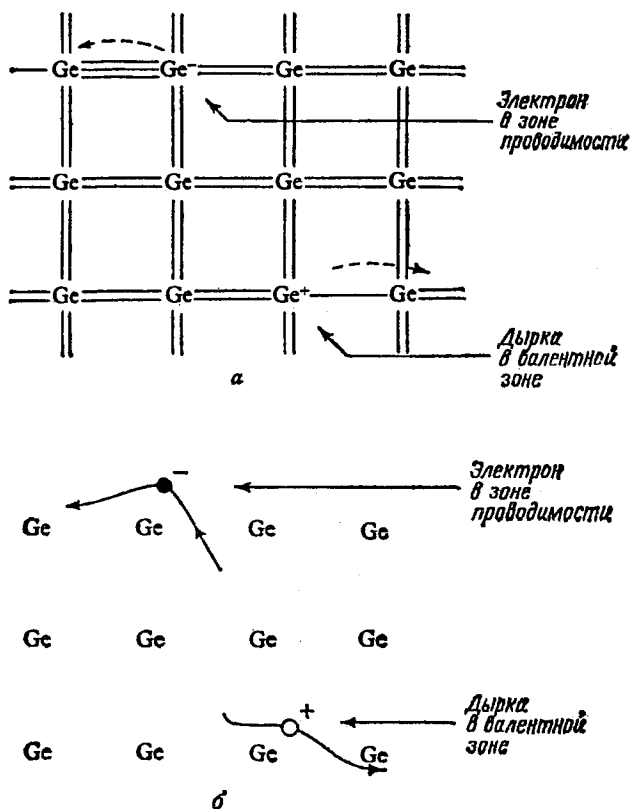


Рис. 3.12. Образование электронно-дырочных пар в Ge.

носителем заполненной оболочки. На рис. 3.12 двумя различными способами проиллюстрирован процесс образования электронно-дырочной пары в Ge.

Энергия образования электронно-дырочной пары в кристаллах элементов четвертой группы убывает по мере роста атомного номера соответствующего элемента. В алмазе она равна $\sim 5,6$ эВ, так что чистый алмаз при обычных температурах является типичным диэлектриком, так как его запрещенная зона слишком широка и поэтому число электронов, попадающих в зону проводимости вследствие теплового возбуждения, ничтожно мало.

В случае кремния ширина запрещенной зоны равна приблизительно 1,12 эВ, а для германия — около 0,67 эВ. Эти одноатомные полупроводники обычно называют полупроводниками IV группы.

3.2.3. СМЕШАННАЯ ХИМИЧЕСКАЯ СВЯЗЬ

Единственными полупроводниками, в которых связь носит чисто гомеоплярный характер, являются полупроводники IV группы. Во всех остальных случаях связь носит смешанный гомеоплярно-ионный характер. Такая смешанная связь встречается и в молекулах. Например, в молекуле HF электрон *не все время* находится на атоме фтора, оставляя протон совершенно оголенным; часть времени он движется вокруг протона по своего рода «размазанной» орбите. Вследствие этого химическая связь в молекуле HF осуществляется частично ионными, частично обменными силами. Главные характеристики смешанной связи зависят от того, какой вид связи преобладает, ионный или гомеоплярный. В полупроводниках, обозначаемых обычно как $A^{III}B^V$, преобладает гомеоплярный характер связи, однако в них явно присутствует и ионная компонента связи. Эти вещества представляют собой соединения элементов III и V групп периодической системы элементов. Преимущественно гомеоплярный характер связи в этих соединениях обычно следует из кристаллической структуры, весьма схожей со структурой полупроводников IV группы: каждый атом типа *A* окружен четырьмя ближайшими соседями типа *B*, находящимися в углах правильного тетраэдра. Такую структуру обычно называют структурой типа цинковой обманки. Она схематически изображена на рис. 3.13 для случая соединения InSb. Чтобы вокруг каждого узла этой решетки могли образоваться замкнутые восьмиелектронные оболочки, необходимо, чтобы с каждого атома элемента V группы один из электронов перешел на атом элемента III группы. Таким образом, в случае соединения InSb можно считать, что узлы кристаллической решетки заняты поочередно ионами In^- и Sb^+ . Поэтому наряду с чисто гомеоплярной связью здесь имеется ионная составляющая, обусловленная электростатическим притяжением между разноименными ионами.

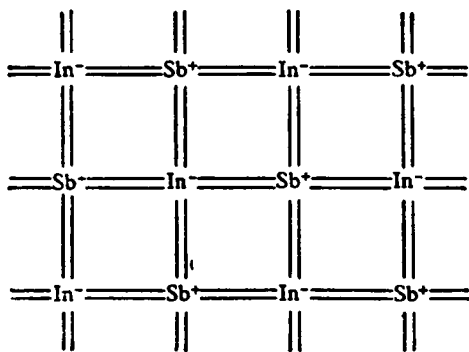


Рис. 3.13. Двумерная схема строения связей в InSb.

Примером полупроводников с преимущественно ионным типом связи может служить группа широко исследованных солей свинца: PbS, PbSe, PbTe. В первом приближении можно принять, что кристалл PbS построен из ионов Pb^{2+} и S^{2-} . Существует, однако, немало доказательств тому, что в этих кристаллах имеется известная доля гомеоплярного типа связи. Вещества со связью преимущественно гомеоплярного типа образуют, как правило, кристаллы со структурой цинковой обманки, или с какой-нибудь другой родственной структурой, тогда как соединения с преобладающим ионным характером связи образуют кристаллы кубической структуры, родственной структуре галогенидов щелочных металлов. За редкими исключениями это относится и к соединениям элементов II и VI групп периодической системы элементов (соединения типа $A^{II}B^{VI}$). Полупроводники этого типа обычно называются полярными полупроводниками.

В соединениях типа XY, где X обозначает двухвалентный атом, а Y — шестивалентный, имеются два пути образования замкнутых электронных оболочек. Во-первых, возможно образование упомянутого выше полярного соединения типа $X^{2+}Y^{2-}$. Этот тип соединения характеризуется электронной структурой валентных оболочек {0—8}. С другой стороны, возможно также образование соединений типа $X^{2-}Y^{2+}$ с электронной структурой валентных оболочек {4—4}, аналогичной структуре в соединениях типа $A^{II}B^{VI}$, причем в последнем случае связь будет главным образом гомеоплярной, но с некоторой примесью ионной связи. В то время как соли свинца PbS, PbSe, PbTe являются соединениями первого типа, соответствующие соли бериллия BeS, BeSe, BeTe относятся к соединениям второго типа. Как и следовало ожидать, последние обладают кристаллической структурой, аналогичной структуре соединений типа $A^{II}B^{VI}$, т. е. структурой цинковой обманки. В каждом отдельном случае трудно *заранее* предсказать, какая именно из указанных структур реализуется в действительности. Можно указать лишь на то, что если атом металла является сильно электроположительным, то более вероятно образование ионной структуры.

В твердых телах известную роль играют также иные типы связи, такие, как металлическая, вандерваальсова и др. Обзор различных типов связи, а также методов их идентификации, особенно в случае связи ионного и гомеоплярного типов, имеется в монографии Эванса [12].

3.3. Другое определение полупроводников

Введенная выше картина химических связей дает возможность сделать еще одно определение полупроводника. Определение, данное в разд. 1.1 и основанное на ширине запрещенной зоны, приме-

нимо только к полупроводникам с четко выраженной структурой энергетических зон, как это описано в предыдущей главе. С другой стороны, в аморфных материалах хотя и можно выявить зонную структуру, но она не так четко выражена, как в кристаллических материалах (см. разд. 15.2). В то же время представление о разрыве химической связи не зависит от дальнего порядка в расположении атомов. Поэтому материал можно считать полупроводником, если требуется небольшая энергия ΔE (скажем, $0 < \Delta E < 2$ эВ) для разрыва химической связи и, следовательно, освобождения электрона, который уже может участвовать в электропроводности. Естественно, при этом предполагается, что если нет разорванных связей, то нет и электронов, которые могли бы принять участие в электропроводности, т. е. химическая связь в материале построена на основе замкнутых электронных оболочек. Здесь ничего не говорится о том, насколько эффективно электрон, освободившийся благодаря тепловому или оптическому возбуждению, сможет участвовать в электропроводности. Если он окажется малоподвижным, то данный материал будет вести себя как диэлектрик. Как бы то ни было, в принципе величину ΔE можно получить, измеряя зависимость поглощения инфракрасного или оптического излучения от длины волны, и, таким образом, доказать наличие полупроводниковых свойств материала. По-видимому, это определение является более фундаментальным, чем сделанное ранее.

Более полно проблему химической связи в полупроводниках рассмотрели Моозер и Пирсон [13]. Взаимосвязь между подходами на основе химической связи и на основе структуры энергетических зон очень наглядно изложил Филлипс [14].

3.4. Примеси замещения в полупроводниках

В качестве простейшего и вместе с тем технологически чрезвычайно важного примера примеси замещения можно указать на элементы III или V групп в качестве примесей в полупроводниковых кристаллах группы IV. Рассмотрим сначала атом пятивалентного As в кристалле Ge. Четыре из пяти валентных электронов атома As вместе с четырьмя непарными электронами, принадлежащими четырем ближайшим соседним атомам Ge, образуют систему двух-электронных связей, описанную выше. Пятый электрон, будучи, так сказать, *лишним*, оказывается весьма слабо связанным с атомом As. Ниже мы произведем приближенную оценку величины энергии связи этого электрона с атомом, примеси. Она оказывается настолько малой, что при комнатной температуре все атомы As в Ge теряют свои лишние электроны. В этом случае говорят, что все атомы As ионизованы. С другой стороны, при очень низких температурах электроны связываются с атомами As. Сказанное илю-

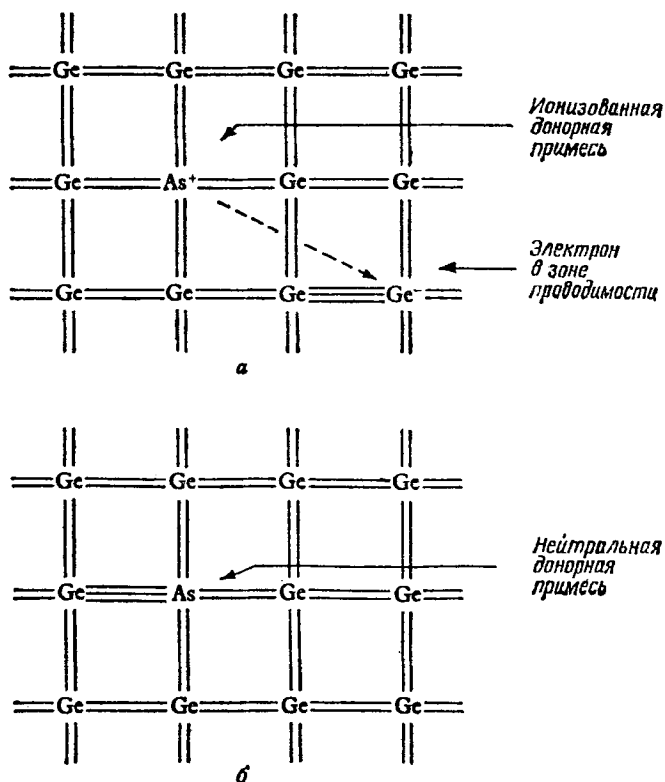


Рис. 3.14. Атом элемента V группы в решетке Ge или Si.

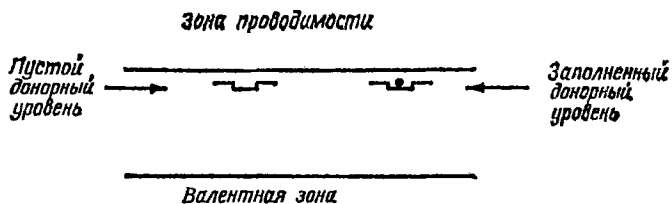


Рис. 3.15. Зонная схема для донорной примеси замещения (атом элемента V группы). Точкой обозначен электрон.

стрируется на рис. 3.14. На языке зонной теории это означает, что каждый атом вносит один, локальный уровень, лежащий на небольшой глубине под дном зоны проводимости. При очень низких температурах этот уровень занят электроном. При обычных температурах он пуст, причем на каждый незанятый уровень приходится

по одному электрону в зоне проводимости (рис. 3.15). Поэтому такой тип примеси носит название *донорной примеси*. Сказанное относится и к другим атомам элементов V группы, таким, как P и Sb. Тот факт, что атомы элементов V группы действительно образуют примесь замещения, следует из различных наблюдений, в частности из определения величины скорости диффузии этих примесей в кристалле; однако и сам факт, что они ведут себя в кристалле так, как это было описано выше, также является убедительным тому доказательством.

Интересно отметить, что положительный ион As^+ , образовавшийся после перехода лишнего электрона в зону проводимости, можно рассматривать как положительную дырку, связанную с атомом As. Поскольку ион As^+ неподвижен, то речь идет о локализованной, *неподвижной дырке*. Иногда такие представления оказываются весьма удобными, так как они согласуются с моделью образования электронно-дырочных пар. Однако часто достаточно рассматривать

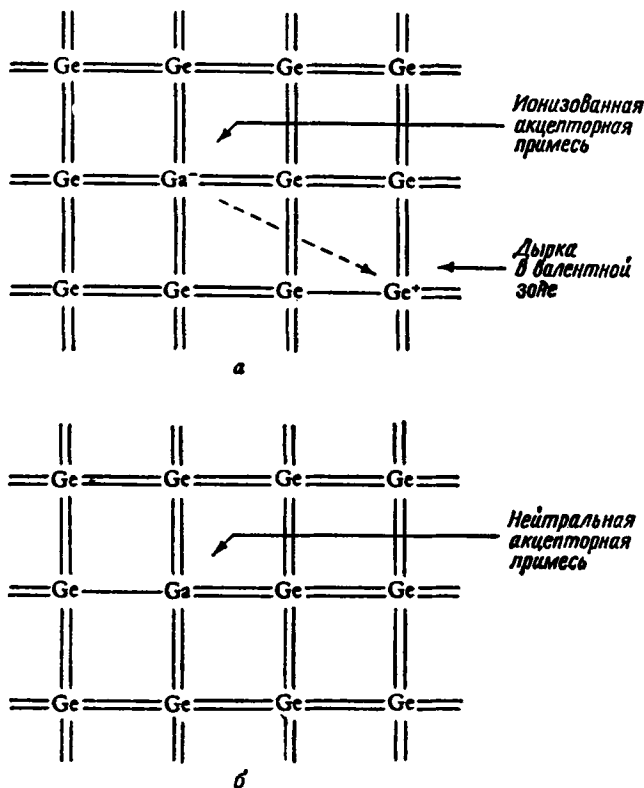


Рис. 3.16. Атом элемента III группы в решетке Ge.

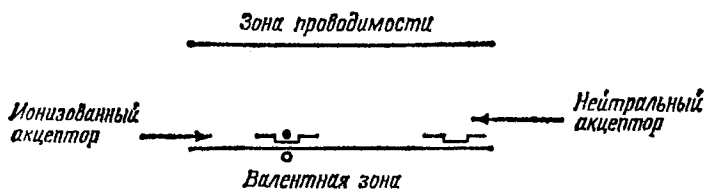


Рис. 3.17. Зонная схема для примеси замещения (атом элемента III группы) (в электронном представлении).

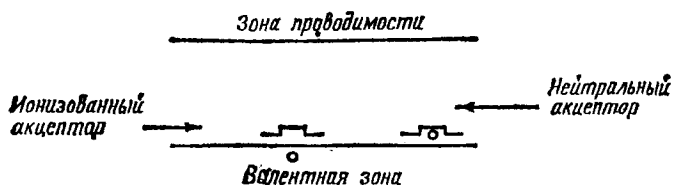


Рис. 3.18. Другой вариант зонной схемы, показанной на рис. 3.17, с использованием уровней энергии для дырок.

электроны в зоне проводимости и свободные дырки в валентной зоне.

Рассмотрим теперь случай трехвалентной примеси замещения, например Ga в кристаллах типа Si или Ge. Атомы элементов III группы имеют лишь три валентных электрона, так что, для того чтобы укомплектовать четыре двухэлектронные связи с соседними атомами Ge, необходимо снять один электрон еще с какого-нибудь атома IV группы. В результате в кристалле образуется положительная дырка (рис. 3.16). Чтобы перевести электрон со связи Ge=Ge на связь Ga=Ge, требуется совсем небольшая энергия, которую мы оценим ниже. Таким образом, при низких температурах атомы Ga, замещающие атомы Ge в узлах решетки, оказываются *нейтральными*. При обычных температурах атом Ga присоединяет к себе лишний электрон, т. е. превращается в ион Ga^- , и в кристалле появляется свободная положительная дырка.

С помощью зонных представлений эти явления можно описать двояким образом. При очень низких температурах примесный уровень Ga можно рассматривать как пустой *электронный* уровень, лежащий несколько выше потолка валентной зоны, на который может перейти один из электронов валентной зоны. С ростом температуры электроны в валентной зоне возбуждаются и начинают переходить на эти уровни, так что при обычных температурах практически все они оказываются занятыми. Это иллюстрируется рис. 3.17; такие примесные энергетические уровни носят название акцепторных уровней. С другой стороны, можно утверждать, что при низких температурах имеются дырки, связанные с ионами Ga^- , ко-

торые при более высоких температурах освобождаются и переходят в валентную зону. Такая схема описания процесса показана на рис. 3.18. Примесные уровни, приписываемые дыркам, изображены с помощью специального значка ¹⁾, указывающего на то, что энергия дырок растет *сверху вниз* и что поэтому на схеме дырки имеют тенденцию занять по возможности более высокое положение. Акцептор в виде отрицательного иона, оставшегося после освобождения связанной с ним дырки, по-прежнему считается ионизованным атомом примеси. Рис. 3.17 и 3.18 представляют собой два возможных описания одного и того же состояния. Сравнивая рис. 3.15 и 3.18, легко убедиться в аналогии между электронами и дырками. Ионизованный акцептор можно также представить себе как связанный с атомом примеси (локализованный) электрон плюс свободная дырка в валентной зоне.

3.4.1. ЭНЕРГЕТИЧЕСКИЕ УРОВНИ ТРЕХВАЛЕНТНЫХ И ПЯТИВАЛЕНТНЫХ АТОМОВ ПРИМЕСИ В ПОЛУПРОВОДНИКАХ IV ГРУППЫ

Перейдем теперь к приближенному расчету энергии отрыва электрона от атома пятивалентной донорной примеси в полупроводниковых кристаллах IV группы и соответственно энергии, необходимой для отрыва дырки от трехвалентного атома примеси. Используемый ниже метод расчета заимствован в основном из книги Мотта и Герни (см. [1], стр. 101). Рассмотрим однократно ионизованный пятивалентный атом примеси, замещающий один атом IV группы (см. рис. 3.14, *a*), например ион As^+ , замещающий атом Ge. Картина гомеоплярных связей в этом случае почти как такая же, как и в идеальном кристалле. Сила F , действующая на достаточно удаленный от иона As^+ свободный электрон в кристалле, приблизительно равна силе, обусловленной единичным положительным зарядом, локализованным на ионе As^+ . Она определяется формулой

$$F = - \frac{e^2}{4\pi\epsilon\epsilon_0 r^2}, \quad (3.2)$$

где r — расстояние электрона от иона As^+ , ϵ — диэлектрическая проницаемость кристалла, ϵ_0 — электрическая постоянная. Этот закон справедлив при $r > (2 \div 3)a$, где a — постоянная решетки. При r порядка расстояния между ближайшими соседями в кристалле лишний электрон оказывает уже значительное возмущающее действие на связанные электроны, и формула (3.2) перестает быть

¹⁾ «Потенциальной ямки» для электрона \sqcup или «потенциальной ямки» для дырок \sqcap . — *Прим. ред.*

справедливой. Однако ниже можно будет убедиться в том, что при больших значениях постоянной ϵ область малых значений r играет ничтожную роль. Силу F можно в данном случае рассматривать как «внешнюю» силу, действующую на электрон, движущийся в идеальном кристалле, и воспользоваться результатами теории, изложенной в разд. 2.5. Итак, рассмотрим энергию электрона с эффективной массой m_e , движущегося в поле силы F . По существу задача сводится к известной задаче об атоме водорода. Поэтому спектр энергетических уровней с соответствующими величинами энергии связи примесного электрона получается сразу в виде

$$W_n = \frac{W_H m_e}{\epsilon^2 m_0 n^2} = \frac{13,6}{\epsilon^2} \cdot \frac{m_e}{m_0} \cdot \frac{1}{n^2} \text{ эВ}, \quad (3.3)$$

где n — целое число.

В случае Ge и Si необходимо видоизменить эту простейшую модель, с тем чтобы учесть отклонение формы энергетических зон от сферической. В дальнейшем мы еще вернемся к обсуждению такого обобщения теории. Предположим, что электрону в германии можно приписать некоторую среднюю скалярную эффективную массу m_e , равную $0,2 m_0$ (такое значение средней эффективной массы m_e , как мы увидим ниже, дает хорошие результаты и при интерпретации других явлений). Полагая $\epsilon=16$, получаем $W_n \approx 0,01/n^2$ эВ, т. е. для основного состояния $W_1 \approx 0,01$ эВ. В принятом здесь приближении значение W_1 получается одинаковым для любых примесных атомов V группы. На самом деле опытные значения энергии в основном состоянии слегка изменяются в зависимости от разновидности донорных атомов (см. разд. 13.3). Однако этот разброс значений невелик и совпадение результатов столь простой теории с опытом следует считать неожиданно хорошим. В случае донора V группы в Si, принимая $\epsilon=12$ и $m_e=0,5 m_0$, получаем $W_n \approx 0,05/n^2$ эВ, тогда как экспериментальные значения W_1 лежат между $0,039$ эВ для Sb и $0,067$ эВ для Bi (см. разд. 13.3). Величина W_1 равна энергетическому зазору между донорным уровнем и дном зоны проводимости. Из предыдущего видно, что в Si и в Ge этот зазор мал по сравнению с шириной запрещенной зоны.

В точности такие же представления применимы и для свободной дырки, движущейся в электрическом поле акцепторного иона трехвалентного элемента (например, иона Ga в Ge). На достаточно больших расстояниях ион Ga^- создает кулоновское поле, в котором и движется дырка. Формула для примесных уровней энергии по виду совпадает с формулой (3.3), необходимо лишь вместо m_e подставить m_h — эффективную массу дырки. Однако, оказалось, что в кристаллах Ge и Si, где энергетический спектр примесей изучен с наибольшей полнотой, ситуация осложняется из-за вырождения валентной зоны. Обсуждение влияния структуры зон на спектр

уровней примеси мы пока отложим и отметим лишь, что, например, в Ge значения W_1 для акцепторов почти такие же, как и для доноров. Для Si, приняв $m_h=0,4 m_0$ из формулы (3.3) получаем $W_1 \approx 0,04$ эВ, тогда как опытные значения W_1 для акцепторов лежат в пределах от 0,045 эВ для В до 0,065 эВ для Ga (исключая In, для которого $W_1=0,16$ эВ; см. разд. 13.3). Здесь W_1 представляет собой энергетический зазор между основным примесным акцепторным уровнем и потолком валентной зоны. Так же как в случае донорной примеси, величина этого зазора оказывается гораздо меньше ширины запрещенной зоны.

Интересно оценить эффективные размеры области локализации волновых функций, которыми описываются электрон или дырка, связанные с соответствующими примесными центрами. Такой оценкой может служить величина радиуса «боровской орбиты» a_n , соответствующей водородоподобной волновой функции. Ее можно вычислить по формуле

$$a_n = a_0 e \frac{m_0}{m_e} n^2, \quad (3.4)$$

где a_0 — радиус первой «боровской орбиты» для атома водорода ($0,53 \cdot 10^{-8}$ см). В случае донорных пятивалентных атомов в германии $a_1 \approx 80a_0$, откуда следует, что волновая функция примесного электрона простирается на расстояния, во много раз превышающие величину постоянной кристаллической решетки. Как можно показать, в этом случае вклад потенциальной энергии при значениях $r \approx a_0$ не играет существенной роли. Это может служить оправданием для использования обычного кулоновского поля для приближенного вычисления W_n . В Si для донорной пятивалентной примеси получаем $a_1 \approx 30a_0$. Хотя волновая функция примесного электрона простирается на расстояния, все еще во много раз превышающие постоянную решетки, следует ожидать, что согласие упрощенной теории с опытом будет в кремнии несколько хуже, чем в германии. Для возбужденных уровней энергии согласие должно быть лучше из-за наличия множителя n^2 , что на самом деле и имеет место (см. разд. 13.3).

Если в зоне проводимости полупроводника имеется несколько эквивалентных минимумов энергии, то необходимо видоизменить использованный выше простой метод расчета, основанный на водородоподобной модели примесного атома. Более того, как мы уже видели, использованное нами приближение теряет силу, если a_n перестает быть большим по сравнению с постоянной решетки. Это имеет место для примесей, основные состояния которых лежат на большом расстоянии от зоны проводимости и от валентной зоны. В этом случае требуется значительно более сложный метод расчета. Обсуждение его мы отложим до разд. 11.5.1.

3.4.2. ЭНЕРГЕТИЧЕСКИЕ УРОВНИ ДРУГИХ ПРИМЕСЕЙ В ПОЛУПРОВОДНИКАХ IV ГРУППЫ

Для примесей с избытком (или недостатком) более чем одного валентного электрона, не существует *простого* способа оценки положения энергетических уровней. Экспериментальным путем установлено, что донорные примеси такого рода, как правило, имеют энергетические уровни, лежащие довольно глубоко под дном зоны проводимости, соответственно акцепторные примеси имеют уровни, лежащие сравнительно высоко над потолком валентной зоны. Некоторые примесные атомы этого типа порождают несколько уровней в соответствии с тем, сколькими электронами занят один и тот же примесный атом¹⁾. Например, Au в Ge может иметь акцепторный уровень, лежащий на 0,15 эВ выше валентной зоны, или донорный уровень, расположенный на 0,20 эВ ниже зоны проводимости²⁾.

Примесные уровни некоторых атомов лежат, по-видимому, в полосе разрешенных энергий, и по этой причине их трудно обнаружить экспериментально. Такая ситуация может иметь место тогда, когда энергия ионизации соответствующих примесных атомов замещения в кристалле сравнима с шириной запрещенной зоны, или даже больше ее. Поэтому такие примесные атомы не могут присоединить к себе лишний электрон и образовать устойчивый отрицательный ион. Имеются основания считать, что именно таким образом ведет себя примесь водорода в решетках Ge и Si. В качестве других примеров можно привести случай малых добавок атомов Si в кристалле Ge, или, наоборот, примесь Ge в кристалле Si. Это неудивительно, если учесть большое сходство структуры внешних электронных оболочек этих двух элементов.

Пустые узлы решетки, или вакансии, по-видимому, очень слабо влияют на концентрацию свободных электронов или дырок. Они могут играть роль ловушек для носителей заряда, энергетические уровни которых лежат далеко от краев зоны проводимости и валентной зоны. У вакансий имеется тенденция собираться на дислокациях. Дислокации, несомненно, имеют большое влияние на скорости диффузии примесей в полупроводниках типа Si и Ge.

¹⁾ Число электронов или дырок, связанных с одним и тем же примесным атомом принято называть зарядовым состоянием примесного атома.— *Прим. ред.*

²⁾ В первом случае атом золота находится в зарядовом состоянии Au^- , во втором — в зарядовом состоянии Au^{2-} . Оба этих состояния атома золота в германии являются акцепторными. Однако зарядовое состояние Au^{2-} проявляется в кристаллах с электронным типом проводимости и ведет себя подобно донорной примеси. Атомы Au в Ge могут иметь еще два уровня: акцепторный, расположенный на 0,05 эВ ниже дна зоны проводимости, и донорный, располагающийся на 0,04 эВ над потолком валентной зоны.— *Прим. ред.*

3.4.3. ПРИМЕСИ В ПОЛЯРНЫХ ПОЛУПРОВОДНИКАХ

Одно из главных отличий полупроводниковых соединений от одноатомных полупроводников связано с тем, что даже в отсутствие посторонних примесей они содержат более одного типа атомов. В связи с этим в них могут возникать дефекты, связанные с отклонением от точного стехиометрического состава. Например, идеальный кристалл PbS должен содержать в точности одинаковое число атомов Pb и S. Однако хорошо известно, что соединение PbS может существовать в виде гомогенной кристаллической фазы с составом $Pb_{1+\delta}S$, где δ — дробное число, принимающее любые значения от нуля вплоть до 10^{-3} . Оказывается, что каждый лишний атом Pb в кристалле PbS действует как донорная примесь, энергетический уровень которой лежит вблизи дна зоны проводимости. Значению $\delta=10^{-3}$ соответствует примерно 10^{19} донорных атомов на 1 см^3 , т. е. очень большое число. Поэтому проблема очистки полупроводниковых соединений значительно усложняется по сравнению с очисткой одноатомных полупроводников, где из кристалла требуется лишь удалить чужеродные атомы и не возникает проблемы точного соблюдения стехиометрического состава.

В двухатомном полярном полупроводнике стехиометрического состава вакантные анионные и катионные узлы должны встречать-

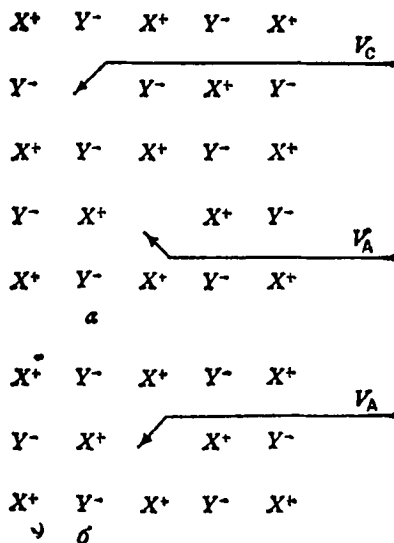


Рис. 3.19. Анионные и катионные вакансии в полярных кристаллах.

a — пара вакансий V_A , V_C ; b — вакансия, обусловленная избыточным атомом металла.

ся одинаково часто. Это — простое следствие условия нейтральности кристалла в целом. Вакантные анионные и катионные узлы обозначим соответственно через V_A и V_C . Если в узлах решетки имеется избыток катионов, то одновременно в кристалле должен иметь место равный избыток анионных вакансий (рис. 3.19). Если, однако, избыток одного из компонентов получается за счет внедрения в междоузлие, то равенство числа обоих видов вакансий, естественно, нарушается. Более того, значительный избыток катионов приводит к заполнению большинства катионных вакансий, так что большая концентрация вакансий анионов сопровождается малой концентрацией катионных вакансий. В условиях термодинамического равновесия концентрация анионных вакансий $[V_A]$ связана с концентрацией катионных вакансий $[V_C]$ законом, аналогичным «закону действующих масс»:

$$[V_A][V_C] = K(T), \quad (3.5)$$

где $K(T)$ — функция, зависящая только от температуры. В основе равенства (3.5) лежит предположение о том, что катионная и анионная вакансии могут выйти на поверхность кристалла и там аннигилировать. При этом, однако, не учитывается возможность образования двойных вакансий. На самом деле такие комплексные дефекты возможны, но соответствующая энергия образования настолько велика, что их число обычно очень мало ¹⁾. Равенство (3.5) можно вывести, если процесс образования и взаимной аннигиляции анионной и катионной вакансий промоделировать с помощью бимолекулярной химической реакции в виде



где S — пара вакансий, вышедших на поверхность кристалла, ΔW — энергетический эффект реакции. Закон действующих масс²⁾ имеет в этом случае вид

$$\frac{[V_A][V_C]}{[S]} = f(T). \quad (3.7)$$

Поскольку изменение $[S]$ мало при условии, что концентрации $[V_A]$ и $[V_C]$ малы по сравнению с концентрацией атомов, формулу (3.7) можно переписать в виде (3.5). Эту формулу можно также представить в виде

$$[V_A][V_C] = [V_s]^2, \quad (3.7a)$$

где $[V_s]$ — концентрация катионных или анионных вакансий в кристалле стехиометрического состава. Из (3.7a) следует, что если

¹⁾ Подробно о методе исследования вакансий на основе термодинамических соотношений для химического равновесия см. [15—17].

²⁾ Этот тип уравнений приведен во многих учебниках по термодинамике (см., например, [18]).

$[V_A] \gg [V_s]$, то должно быть $[V_C] \ll [V_s]$, и наоборот. Функцию $K(T)$ можно представить в виде (см. [18])

$$K(T) = [F(T)]^2 \exp \left[-\frac{\Delta W}{kT} \right], \quad (3.8)$$

где $F(T)$ слабо зависит от температуры по сравнению с $\exp(-\Delta W/2kT)$. Поэтому

$$V_s = F(T) \exp \left[-\frac{\Delta W}{2kT} \right]. \quad (3.9)$$

В кристаллах с большим значением энергии активации ΔW число вакансий невелико. Такая ситуация имеет место, по-видимому, в InSb (см. разд. 13.5). Вместе с тем в солях свинца PbS, PbSe, PbTe вакансии, по-видимому, образуются гораздо легче (см. разд. 13.6).

Рассмотрим теперь подробнее влияние избытка одного из компонентов на примере двухатомного ионного кристалла, построенного из одновалентных ионов. Такие кристаллы являются обычно ионными проводниками, например NaCl, который считается типичным представителем кристаллов этого класса. Предположим, что избыточные атомы Na занимают нормальные положения в узлах решетки. Эти узлы они занимают в виде ионов Na^+ , поэтому каждый избыточный атом Na отдает по одному избыточному электрону. С каждым избыточным атомом Na в кристалле можно сопоставить избыточную вакансию, занятую анионом Cl^- . Такая вакансия обладает единичным эффективным *положительным* зарядом, поэтому она притягивает электрон. Освобожденный избыточным атомом Na электрон захватывается этой вакансией и заряд ее нейтрализуется. Это сочетание электрона с анионной вакансией, обозначаемое обычно символом V_A^- , показано на рис. 3.20 на примере кристалла NaCl. В галогенидах щелочных металлов энергия связи электрона в анионной вакансии порядка 0,5 эВ. Такие образования, известные под названием *F-центров*, приводят к окрашиванию кристаллов галогенидов, содержащих избыток металла. В этих кристаллах *F-центры* могут захватить по второму электрону и превратиться в дважды заряженные вакантные центры V_A^{2-} или *F'-центры*. При избытке анионов, т. е. ионов Cl^- , каждый атом галогена должен освободить положительную дырку, которая захватывается вакансией катиона и образует V_C^+ -центр.

В полупроводниках мы обычно имеем дело с двухвалентными ионами, и энергия связи электрона, захваченного анионной вакансией, как правило, настолько мала, что вакансия легко ионизируется (если температура кристалла не очень низка) и электрон попадает в зону проводимости. Как известно, в PbS, например, каждый избыточный атом свинца отдает при обычных температурах по одному электрону в зону проводимости. Однако если допустить, что кристалл состоит из ионов Pb^{2+} и S^{2-} , то с каждого избыточного иона

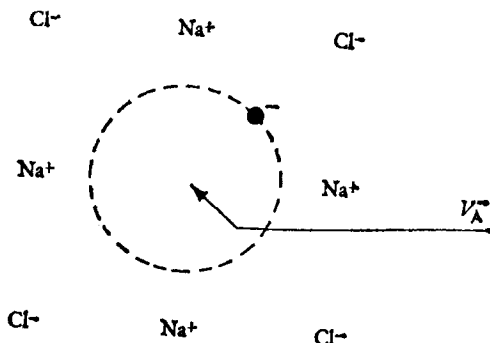


Рис. 3.20. F -центр (V_A^- -центр) в решетке NaCl.

Pb^{2+} освобождается по два электрона. Вакансия V_A теперь уже дважды заряжена по отношению к случаю идеального кристалла и может присоединить к себе два электрона. Один электрон захватывается вакансией на глубоко лежащий уровень, образуя V_A^- -центр. Такой центр обладает единичным отрицательным зарядом. Присоединяя к себе второй электрон, V_A^- -центр превращается в V_A^{2-} -центр. При обычных температурах V_A^{2-} -центры ионизируются с образованием V_A^- -центра и свободного электрона. Аналогичным образом можно предположить также существование V_C^{2+} -центров и V_C^+ -центров, которые могут играть важную роль при избытке анионов, например при избытке S в кристаллах PbS. В этом случае каждый избыточный атом серы должен был бы освободить положительную дырку. Однако на опыте это не всегда так. В CdS, например, при комнатной температуре избыточные атомы Cd поставляют свободные электроны, тогда как избыточные атомы серы не дают свободных дырок. Это означает, что акцепторный уровень, соответствующий V_C^{2+} -центру, расположен слишком далеко от потолка валентной зоны, чтобы он мог ионизоваться при комнатных температурах.

Чужеродные атомы с валентностью, отличающейся от валентности основных атомов, также порождают донорные или акцепторные уровни в энергетическом спектре полярных полупроводников. Предполагается, например, что атомы Bi образуют в PbS примесь замещения в виде ионов Bi^{3+} . При комнатных температурах каждый замещенный ион Pb^{2+} освобождает один электрон, который легко отщепляется от иона Bi^{2+} . При этом возникает примесный центр (Bi^{3+}) с эффективным зарядом $+e$ относительно случая идеального кристалла. Аналогично, Ag замещает Pb в PbS в виде иона Ag^+ , порождая одну свободную дырку на каждый замещенный ион Pb^{2+} и один примесный центр (Ag^+) с эффективным зарядом в кристалле $-e$. Возможно также замещение аниона чужеродным атомом.

Например, предполагается, что Cl замещает S в кристалле PbS в виде иона Cl^- , так что каждый замещенный хлором ион S^{2-} освобождает по одному электрону и представляет собой примесный центр (Cl^-) с эффективным зарядом $+e$ (см. [16], стр. 68). Действие примесей замещения в полярных полупроводниках подробнее обсуждено в разд. 13.6.

3.4.4. ПРИМЕСИ В ПОЛУПРОВОДНИКОВЫХ СОЕДИНЕНИЯХ ТИПА $A^{III}B^V$

Аналогичные соображения применимы и к рассмотрению таких соединений, как полупроводники типа $A^{III}B^V$, связь в которых в основном оказывается гомеополлярной, хотя ионный тип связи также играет в них некоторую роль (см. разд. 13.6). Замена пятивалентного атома атомом более высокой валентности, например замещение атома Sb в InSb атомом Te, приводит к появлению донорного уровня вблизи дна зоны проводимости, поскольку ион Sb^+ , вероятно, замещается ионом Te^{2+} (Te имеет шесть валентных электронов). При этом освобождается один электрон. С другой стороны, Zn в InSb порождает положительные дырки. В этом случае ион In^- замещается ионом Zn^{2-} (Zn имеет два валентных электрона) и возникает при этом одна положительная дырка. При введении в соединения типа $A^{III}B^V$ примеси четырехвалентных атомов заранее не очевидно, какой из основных ионов кристалла будет замещаться примесью. По-видимому, Ge в InSb порождает неглубокий акцепторный уровень, так что Ge предположительно замещает ион Sb^+ и создает один электрон. Во многих соединениях типа $A^{III}B^V$ энергия образования вакансий в узлах решетки, по-видимому, настолько велика, что имеется сильная тенденция к образованию соединений стехиометрического состава. Так обстоит дело, в частности, в InSb (см. разд. 13.5), и это обстоятельство оказалось весьма благоприятным, ибо оно значительно облегчает приготовление чистых кристаллов InSb стехиометрического состава.

3.4.5. ДРУГИЕ ТИПЫ НЕСОВЕРШЕНСТВ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКИ

Существует много других типов несовершенств кристаллической решетки твердых тел. Мы упомянули только несколько наиболее часто встречающихся типов, которые играют определяющую роль в формировании свойств полупроводников. Все рассмотренные выше виды несовершенств решетки могут быть введены в кристалл химическими методами. Например, вакансии можно создать, вводя в полярные кристаллы атомы другой валентности (см. разд. 3.4.3). Большие концентрации вакансий и междоузельных атомов можно

получить, облучая образцы тяжелыми частицами или электронами, либо γ -лучами, посредством рождаемых ими электронов с высокой энергией. Если при этом возникают нежелательно большие концентрации дефектов, то говорят, что имеют место «радиационные повреждения». Они могут привести к выходу из строя полупроводникового прибора.

В небольших дозах, однако, ионизирующее излучение оказалось наиболее удобным средством для изучения несовершенств кристаллической решетки. С его помощью можно создавать не только одиночные вакансии, но и двойные вакансии и даже макроскопические полости, состоящие из большого числа вакансий. В этих условиях вакансии стремятся сгруппироваться с образованием дислокаций, так что кристалл оказывается подвергнутым значительным локальным механическим напряжениям (см. разд. 3.1.3), которые могут возникнуть и в результате введения в кристалл больших концентраций примесей, скажем $> 10^{19}$ см $^{-3}$.

Облучение тяжелыми частицами может быть использовано для введения в кристалл химических примесей в виде ионов. Ионы ускоряются до высоких энергий, так что они проникают на некоторое расстояние вглубь кристалла. В этих условиях кристалл получает сильные радиационные повреждения, однако большая их часть может быть устранена путем тщательного отжига, при котором необходимые примеси остаются в узлах решетки или в междоузлиях. Обычно глубина проникновения имеет малую величину, достаточную, однако, для создания полупроводникового прибора. Такая методика¹⁾ широко применяется для «легирования» полупроводников. (Этот вопрос подробно рассмотрен в нескольких статьях и книгах [19—24].)

Позже мы обсудим, как различные типы несовершенств решетки влияют на оптические и электрические свойства полупроводников. Мы также покажем, как можно определить положение энергетических уровней более глубоких и более сложных примесей и несовершенств решетки и как их можно описать теоретически (см. разд. 11.5).

Обширная литература (см., например, [25, 26]) посвящена несовершенствам решетки, вызванным радиационным повреждением. Это неудивительно, так как они в значительной степени влияют на работу полупроводниковых приборов.

3.5. Экситоны

При исследовании движения электрона в идеальном кристалле мы раньше везде пренебрегали взаимным отталкиванием электронов. Более строгая теория приводит к результатам, оправдывающим

¹⁾ Она известна как «ионное внедрение». — *Прим. ред.*

это, по-видимому, довольно грубое предположение [27]. Точно так же ранее не учитывалось взаимное отталкивание между дырками и, кроме того, взаимное притяжение между электронами и дырками. Последнее утверждение, однако, не совсем верно, поскольку большая часть энергии образования электронно-дырочной пары определяется именно работой по преодолению кулоновского притяжения между разноименными носителями. Взаимодействие электронов между собой приводит лишь к взаимному рассеянию, которое может играть существенную роль только при больших концентрациях носителей заряда [28]. Вместе с тем учет притяжения между дыркой и электроном в условиях, когда концентрация носителей заряда не слишком велика, приводит к новым, чрезвычайно интересным результатам. Исходя из того, что дырки и электроны движутся в кристалле под действием электрического поля как свободные частицы, каждая со своей эффективной массой, легко убедиться в том, что электрон и дырка в кристалле могут при определенных условиях образовать связанную систему с дискретным энергетическим спектром, которая может двигаться по кристаллу как единое целое. Такая система носит название экситона [29]. В известном смысле экситон представляет собой возбужденное состояние атома в кристалле, передающееся от атома к атому посредством квантового механического резонанса [30].

Если для простоты предположить, что электрону и дырке можно приписать скалярные эффективные массы m_e и m_h , то задача об экситоне сводится к задаче о простой водородоподобной системе, т. е. о движении двух частиц под действием взаимного кулоновского притяжения. Если на время пренебречь движением центра масс этой системы частиц, то энергия W_{ex}^n (отсчитанная от состояния, когда электрон и дырка находятся на бесконечном расстоянии друг от друга) равна

$$W_{ex}^n = - \frac{W_n}{\epsilon^2} \cdot \frac{m_r}{m_0} \cdot \frac{1}{n^2}, \quad (3.10)$$

где m_r — «приведенная» масса системы двух частиц $m_e m_h / (m_e + m_h)$. Величины W_n , n и ϵ имеют тот же смысл, что и в формуле (3.3), которая аналогична формуле (3.10). Если положить $m_e = m_h$, то $m_r = 1/2 m_e$ и $W_{ex}^n = -1/2 W_n$. Таким образом, энергия связи экситона в основном состоянии в полупроводниках IV группы по порядку величины совпадает с глубиной примесного донорного уровня элементов V группы. Следует ожидать, что энергия связи экситона в Ge будет порядка 0,005 эВ, что очень близко к экспериментально найденному значению 0,0036 эВ (см. разд. 10.6).

Рассмотрим теперь подробнее физический смысл величины W_{ex} . Если исходить из системы энергетических уровней для одиночного электрона, представленной на рис. 3.15, то основной уровень экситона необходимо поместить на глубине W_{ex}^1 ниже дна зоны проводимости, а возбужденные состояния экситона — соответствен-

но на глубинах $W_{\text{ex}}^1/4$, $W_{\text{ex}}^1/9$ и т. д. Таким образом, мы получаем бесконечное число уровней, переходящих в сплошной спектр при $n \rightarrow \infty$. Сплошной спектр соответствует электрону и дырке в свободном, т. е. несвязанном, состоянии, иначе говоря, состоянию в зоне проводимости. Может показаться, что приведенные выше соображения противоречат положению, вытекающему из теоремы Блоха (см. разд. 2.2), согласно которому между зоной проводимости и валентной зоной в идеальном кристалле *не должно* быть никаких разрешенных уровней энергии. Объясняется это противоречие тем, что в теории Блоха мы имеем дело с *невоздействующими* электронами и дырками, тогда как введение экситона представляет собой приближение более высокого порядка.

Рассмотрим кристалл, в котором валентная зона отделена от зоны проводимости запрещенной зоной конечной ширины, и выясним, какова природа первого возбужденного состояния в таком кристалле. С точки зрения энергии электронов кристалл находится в «основном» состоянии, т. е. в состоянии с наименьшей возможной энергией, когда все уровни в валентной зоне заняты электронами, а зона проводимости совершенно пуста. Если отвлечься от возможности образования экситонов, то первому возбужденному состоянию кристалла соответствует один электрон на самом дне зоны проводимости и одна дырка у потолка валентной зоны, т. е. электрон и дырка, кинетические энергии которых равны нулю. Однако ясно, что такое состояние окажется неустойчивым, если мы учтем взаимодействие между электроном и дыркой, которое приводит эти частицы во взаимно связанное состояние с меньшей энергией. Но связанные электрон и дырка, которые перемещаются по кристаллу как единое целое, по определению представляют собой экситон. Отсюда следует, что если кристалл содержит один экситон в состоянии с наименьшей энергией, то кристалл находится в наименьшем энергетическом состоянии, не считая «основного». Иными словами, первый возбужденный уровень кристалла соответствует такому

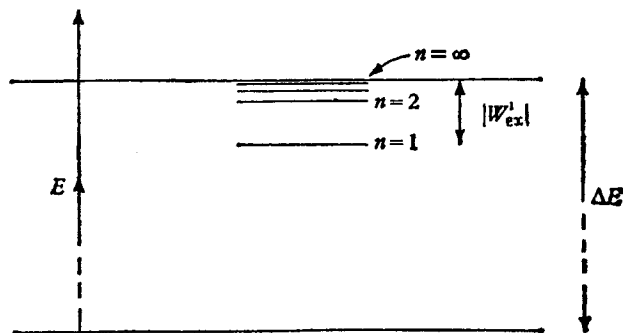


Рис. 3.21. Уровни энергии экситона.

состоянию, когда имеется один экситон в низшем энергетическом состоянии, причем первый возбужденный уровень энергии отделен от «основного» состояния кристалла зазором величиной $\Delta E - |W_{\text{ex}}^1|$.

Строго говоря, изображение экситонных уровней энергии на одноэлектронных диаграммах некорректно, поскольку речь идет о системе двух взаимодействующих между собой частиц — электрона и дырки. Но иногда все же бывает удобно пользоваться такой наглядной схемой (рис. 3.21).

До сих пор мы пренебрегали движением центра масс экситона. Это движение описывается волновой функцией вида

$$\psi(\mathbf{R}) = \exp[i(\mathbf{K} \cdot \mathbf{R})], \quad (3.11)$$

где \mathbf{R} — радиус-вектор центра масс экситона, а \mathbf{K} — волновой вектор, связанный с импульсом движения центра масс соотношением, аналогичным зависимости $\mathbf{p} = \hbar \mathbf{k}$ для электрона. При малых значениях \mathbf{K} кинетическая энергия экситона W_{ex}^0 равна

$$W_{\text{ex}}^0 = \hbar^2 \frac{|\mathbf{K}|^2}{2M}, \quad (3.12)$$

где $M = m_e + m_h$. Таким образом, при учете движения центра масс экситона линии энергетического спектра экситонов расширяются в *зоны*. По аналогии с электроном энергию экситона можно изображать как функцию \mathbf{K} , причем принимая во внимание периодичность кристалла, по-прежнему достаточно рассматривать лишь приведенные значения \mathbf{K} . Поэтому правильнее говорить об экситонных зонах, а не о дискретных уровнях энергии экситона. Однако при образовании экситонов в результате поглощения одиночных квантов излучения возникают экситоны с очень малыми значениями \mathbf{K} . Это следует из закона сохранения импульса, если учесть то обстоятельство, что импульс светового кванта чрезвычайно мал. Поэтому очень часто экситону можно приписать дискретную совокупность уровней W_{ex}^n . Как будет видно из дальнейшего, спектр экситона будет линейчатым, если только вершина валентной зоны и дно зоны проводимости соответствуют одному и тому же значению квазиимпульса, например $\mathbf{k} = 0$. В указанном выше простейшем случае спектр поглощения твердого тела в области экситонного поглощения состоит из серии дискретных линий. Частота ν первой линии определяется равенством $h\nu = \Delta E - |W_{\text{ex}}^1|$. Граница серии характеризуется предельной частотой ν , которую можно найти из условия $h\nu = \Delta E$ (рис. 3.22). При частотах, превышающих предельную частоту экситонного поглощения, начинается область сплошного поглощения, связанного с переходами электронов из валентной зоны в зону проводимости¹⁾. Колебания кристаллической решетки и раз-

¹⁾ То есть область собственного поглощения, или поглощения зона — зона. — Прим. ред.

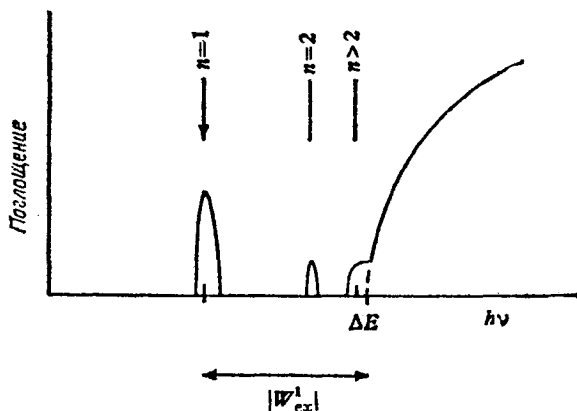


Рис. 3.22. Предсказываемый теорией вид спектра поглощения в твердом теле.

личные дефекты кристаллической структуры приводят к уширению линий экситонного поглощения, а в некоторых случаях даже к слиянию экситонного спектра со спектром собственного поглощения. Аналогичный вид спектра поглощения в твердых телах предсказал Пайерлс [31], исходя из более общей теории возбужденных состояний кристалла. Основным экспериментальным доказательством образования экситонов при низких температурах обычно служит наличие поглощения света кристаллом вблизи красной границы спектра собственного поглощения (см. рис. 3.22), которое не приводит к образованию свободных носителей заряда. При более высоких температурах экситоны термически диссоциируют на свободные электроны и дырки. Еще одним доказательством является форма спектра поглощения вблизи края фундаментального поглощения, а также наличие линейчатых спектров, которые принято связывать с образованием экситонов. Этот вопрос будет рассмотрен в разд. 10.6. Обзор выполненных до 1955 г. экспериментальных работ, посвященных экситонам (на английском языке), опубликован Гроссом [32].

Линейчатые спектры поглощения в виде водородоподобных серий наблюдались в ряде соединений с очень широкими запрещенными зонами [33], например в CdS , HgI_2 , PbI_2 , CdI_2 . Линейчатый спектр поглощения в Cu_2O впервые наблюдали Апфель и Хэдди [34], Гросс и Захарченя [35] и Никитин [36] и отождествили его со спектром экситонного поглощения. Одно время считалось, что в Ge и Si условия для наблюдения линейчатого экситонного спектра поглощения должны быть весьма благоприятными, поскольку эти вещества доступны в виде чрезвычайно чистых и совершенных монокристаллов. Однако выделение экситонного спектра поглощения в

Ge и Si оказалось трудным делом по причинам, которые удалось полностью понять лишь после выяснения зонной структуры этих веществ. Максимум энергии в валентной зоне и минимум энергии в зоне проводимости расположены для этих кристаллов в разных точках k -пространства, так что условия применимости сформулированной выше простейшей теории не выполняются. Можно показать, что экситонное поглощение в этом случае должно проявиться в виде серии полос поглощения. Такие полосы действительно наблюдались в Si и Ge Макфарланом и др. [37]. Свойства экситонных полос поглощения подробно рассмотрены в разд. 10.6.

Может показаться нелогичным обсуждение свойств экситонов в главе, посвященной примесям, поскольку экситоны характеризуют энергию кристалла в целом, и вопрос об их возникновении не связан с наличием примесей. Однако некоторые свойства экситонных уровней аналогичны свойствам примесных уровней, так что сравнение этих двух случаев не лишено интереса. В частности, как экситоны, так и примеси приводят к появлению энергетических уровней в «запрещенной» энергетической зоне.

ЛИТЕРАТУРА

1. Mott N. F., Gurney R. W., *Electronic Processes in Ionic Crystals*, Oxford University Press, Oxford, 1940. (Имеется перевод: Мотт Н., Герни Р. Электронные процессы в ионных кристаллах.— М.: ИЛ, 1950.)
2. Cottrell A. H., *Dislocations and Plastic Flow in Crystals*, Oxford University Press, Oxford, 1953.
3. Read W. T., *Dislocations in Crystals*, McGraw-Hill, New York, 1953.
4. Cottrell A. H., *Theory of Crystal Dislocations*, Gordon and Breach, New York, 1964.
5. Alexander H., Haasen P., в сб.: *Solid State Physics*, ed. Seitz F., Turnbull D., Academic Press, New York, 22, 27 (1968).
6. Gatos H. C., Finn M. C., Lavine M. C., *Journ. Appl. Phys.*, 32, 1174 (1961).
7. Bardsley W., Bell R. L., *Journ. Electron. Cont.*, 3, 103 (1957).
8. Mitchell J. W., *Phil. Mag. Ser. 7*, 44, 223 (1953).
9. Born M., Huang K., *The Dynamical Theory of Crystal Lattices*, Oxford University Press, Oxford, 1954. [Имеется перевод: Борн М., Хуан Кунь. Динамическая теория кристаллических решеток.— М.: ИЛ, 1958.]
10. Heitler W., London F., *Zs. Phys.*, 44, 455 (1927).
11. Smith R. A., *The Wave-Mechanics of Crystalline Solids*, 2nd ed. Chapman and Hall, and John Wiley and Sons, New York, 1969.
12. Evans R. C., *Crystal Chemistry*, 2nd ed., Cambridge University Press, Cambridge, 1964.
13. Mooser E., Pearson W. B., в сб.: *Progress in Semiconductors*, ed. Gibson A. F., Aigrain P., Burgess R. E., New York, 5, 103 (1960).
14. Phillips J. C., *Bonds and Bands in Semiconductors*, Academic Press, New York, 1973.
15. Rees A. L. G., *Chemistry of the Defect Solid State*, Methuen Monographs, London, 1954. [Имеется перевод: Риз А. Химия кристаллов с дефектами.— М.: ИЛ, 1956.]
16. Kröger F. A., Vink H. J., *Solid State Physics*, Academic Press, New York, 3, 307 (1956).

17. *Vink H. J.*, Rendiconti della Scuola Enrico Fermi, XXII Corso, Academic Press, New York, 1963, p. 68.
18. *Smith R. A.*, Physical Principles of Thermodynamics, Chapman and Hall, New York, 1952, p. 214.
19. *Mayer J. W.*, *Eriksson L.*, *Davies J. A.*, Ion Implantation in Semiconductors, Academic Press, New York, 1970.
20. *Dearnaley G. et al.*, Ion Implantation, North Holland, Amsterdam, 1973.
21. *Bassani F.*, *Iadonisi G.*, *Preziosi B.*, Rept. Prog. Phys., **37**, 1099 (1974).
22. Lattice Defects in Semiconductors, ed. Huntley F. A., Inst. Phys Conf. Series., 1974.
23. Point Defects in Solids, ed. Crawford J. N., Sliifkin L. M., Plenum Press, New York, 1975, vol. 2.
24. *Stoneham A. M.*, Theory of Defects in Solids, Oxford University Press, Oxford, 1975. [Имеется перевод: *Стоунхэм А. М.* Теория дефектов в твердых телах, — М.: Мир, 1978.]
25. Radiation Effects in Semiconductors, ed. Corbett J. W., Watkins G. D., Gordon and Breach, New York, 1971.
26. Radiation Damage and Defects in Semiconductors, ed. Whitehouse J. E., Inst. Phys. Conf. Series, 1973.
27. *Pines D.*, Solid State Physics, Academic Press, New York, **1** (1955).
28. *Fröhlich H.*, *Doniach S.*, Proc. Phys. Soc., **B69**, 961 (1956).
29. *Френкель Я.*, Phys. Rev., **37**, 17, 1276 (1931).
30. *Wannier G.*, Phys. Rev., **52**, 191 (1937).
31. *Peierls R.*, Ann. Phys., **13**, 905 (1932).
32. *Гросс Е. Ф.*, Nuovo Cimento (Suppl.), **3**, 672 (1956).
33. *Гросс Е. Ф.*, *Каплянский А.*, ЖТФ, **25**, 2061 (1955).
34. *Apfel J. H.*, *Hadley L. N.*, Phys. Rev., **100**, 1689, (1955).
35. *Гросс Е. Ф.*, *Захарченко В. Р.*, ДАН СССР, **90**, 745 (1953).
36. *Nikiline S.*, Journ. de Phys. et le Radium, **17**, 817 (1956).
37. *Macfarlane G. G.*, *McLean T. P.*, *Quarrington J. E.*, *Roberts V.*, Phys. Rev., **108**, 1377 (1957); **111**, 1245 (1958).

РАВНОВЕСНЫЕ КОНЦЕНТРАЦИИ НОСИТЕЛЕЙ ЗАРЯДА

4.1. Распределение электронов по энергетическим уровням

В двух предыдущих главах были рассмотрены различные системы уровней энергии для электронов в полупроводниковых кристаллах. При изучении многих физических явлений в полупроводниках можно не принимать во внимание энергетические уровни, расположенные глубже валентной зоны. Существуют, конечно, явления, например поглощение рентгеновских лучей кристаллом, которые нельзя описать без учета глубоких уровней, однако во всех проблемах, которые будут обсуждены в дальнейшем, эти уровни можно считать полностью заполненными и поэтому их можно вовсе исключить из рассмотрения. В действительности имеют существенное значение только уровни в верхней части валентной зоны, уровни в нижней части зоны проводимости, а также примесные уровни, занимающие промежуточные положения. Наша задача заключается в том, чтобы выяснить, как распределяются электроны по уровням энергии при заданной температуре T , полагая, что в состоянии с минимальной энергией, соответствующей абсолютному нулю температуры, они заполняют все состояния в валентной зоне и все примесные уровни в запрещенной зоне. Для этой цели необходимо воспользоваться фундаментальными соотношениями статистической физики, которые определяют вероятность заполнения электроном состояния с заданной энергией E . Пусть энергия E , соответствующая определенному уровню, измеряется относительно произвольного, но фиксированного начала отсчета. Тогда вероятность $P_e(E)$ заполнения энергетического уровня электроном определяется выражением

$$P_e(E) = \frac{1}{\exp \left[\frac{E - E_F}{kT} \right] + 1}, \quad (4.1)$$

где E_F — постоянная, которую находят из условия равенства ожидаемого полного числа электронов во всех состояниях фактическому числу электронов в системе. Функция, описываемая равенством

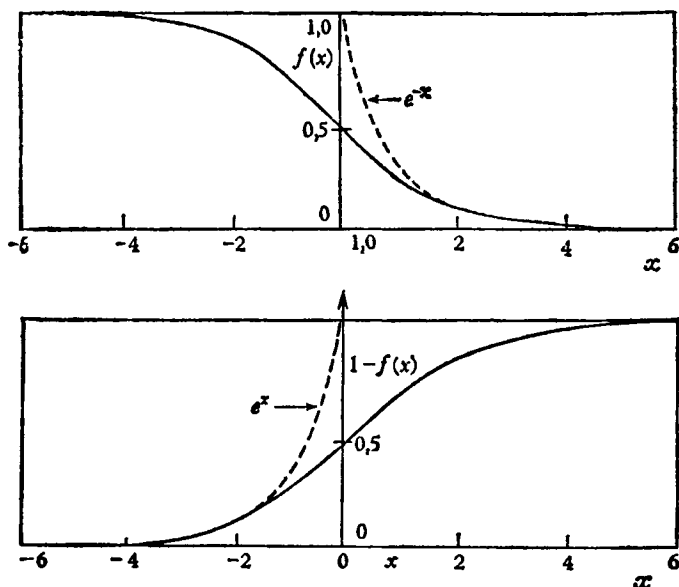


Рис. 4.1. Функция Ферми — Дирака $f(x)$ и функция $1-f(x)$. Пунктиром нанесены графики функций e^x и e^{-x} .

(4.1), носит название функции распределения Ферми — Дирака. Это распределение справедливо для любой совокупности свободных частиц, которые, аналогично электронам, подчиняются принципу Паули¹⁾. Равенство (4.1) можно записать в виде

$$P_e(E) = f \left[\frac{E - E_F}{kT} \right], \quad (4.1a)$$

где

$$f(x) = \frac{1}{e^x + 1}. \quad (4.2)$$

Функция $f(x)$ есть функция Ферми — Дирака. На рис. 4.1 представлена ее зависимость от x . При выводе формулы (4.1) предполагалось, что все уровни энергии невырождены. Если уровни энергии g -кратно вырождены и если вероятность заполнения любого уровня энергии электроном не зависит от степени заполнения этого же уровня другими электронами, то в этом случае мы имеем

$$P_e(E) = gf \left[\frac{E - E_F}{kT} \right]. \quad (4.3)$$

¹⁾ Вывод распределения Ферми — Дирака можно найти в любом элементарном учебнике по квантовой статистике. Простой и весьма изящный вывод этого распределения приведен в книге Шредингера [1], гл. VII.

Этот результат получается непосредственно, если каждый g -кратно вырожденный уровень заменить группой из g очень близких друг к другу невырожденных уровней энергии и воспользоваться формулой (4.1a). Разрешенные уровни энергии для свободных электронов в идеальном кристалле обладают вырождением именно такого типа, и каждый из них можно считать двукратно вырожденным в соответствии с двумя возможными направлениями спина. Заполнение такого уровня электроном с заданным направлением спина не мешает заполнению этого же уровня электроном с противоположным направлением спина. При рассмотрении примесных уровней мы столкнемся с вырожденными уровнями, для которых заполнение уровня одним электроном совершенно исключает возможность заполнения этого же уровня вторым электроном.

Можно показать, что функция $f(x)$ обладает следующими свойствами: она мала при больших положительных значениях x , поэтому вероятность заполнения уровня E электроном мала, если $(E - E_F) \gg \gg kT$. С другой стороны, $f(x) \simeq 1$ при больших отрицательных значениях x , откуда следует, что уровень будет почти наверняка занят, если $(E - E_F) \ll \ll kT$. Отметим также, что $f(0) = 1/2$, так что $P_e(E) = 1/2$ при $E = E_F$. Это значит, что если в спектре энергии существует уровень с $E = E_F$, то он с одинаковой вероятностью может оказаться занятым электроном или быть пустым. В любом случае все уровни с энергией $E > E_F$ оказываются преимущественно незанятыми, тогда как все уровни с энергией $E < E_F$ преимущественно заполнены. Уровень энергии, соответствующий E_F , носит название уровня Ферми, и указанное выше свойство этого уровня может использоваться для определения E_F . Величина E_F связана с термодинамическим потенциалом, она остается постоянной в «многофазной» равновесной системе, например в контакте двух различных полупроводников в состоянии термодинамического равновесия. При больших положительных x с большой точностью можно заменить функцию $f(x)$ на e^{-x} , так что при $E - E_F \gg \gg kT$ имеем

$$P_e(E) = \exp\left[-\frac{E - E_F}{kT}\right] = A \exp\left[-\frac{E}{kT}\right], \quad (4.4)$$

где A — нормирующая постоянная. Такой вид функции $f(x)$ в точности совпадает с известным классическим распределением $P(E)$ электронов по энергиям, при выводе которого принцип Паули не учитывается. Отсюда ясно, что распределение Ферми — Дирака переходит в классическое распределение для тех значений E , при которых $E - E_F \gg \gg kT$. В этих условиях система считается невырожденной.

Вероятность $P_n(E)$ того, что уровень с энергией E не будет

занят электроном, определяется, очевидно, выражением

$$P_h(E) = 1 - P_e(E) = \frac{1}{\exp\left[\frac{E_F - E}{kT}\right] + 1}. \quad (4.5)$$

Функцию $P_h(E)$ можно трактовать как вероятность заполнения уровня E положительной дыркой. Здесь снова мы видим аналогию между положительными дырками и электронами, с той только разницей, что для дырок направление отсчета энергии необходимо изменить на противоположное. При $(E - E_F) \ll kT$ мы получаем $P_h(E) = \exp[(E - E_F)/kT]$.

4.2. Собственные полупроводники

Воспользуемся полученными выше зависимостями для вывода распределения электронов по энергиям в собственном полупроводнике, в запрещенной зоне которого нет примесных уровней. Ширину запрещенной зоны обозначим через ΔE , а начало отсчета энергии поместим на дне зоны проводимости. Пусть $N_c(E)dE$ — число разрешенных энергетических уровней в единице объема в зоне проводимости в интервале энергий от E до $E + dE$. Аналогично можно определить соответствующую величину для валентной зоны. Тогда число электронов $n(E)dE$ в зоне проводимости с энергиями в интервале от E до $E + dE$ равно

$$n(E)dE = 2N_c(E)P_e(E)dE. \quad (4.6)$$

Множитель 2 в формуле (4.6) учитывает спиновое вырождение каждого уровня. Иными словами, каждый уровень может быть занят двумя электронами с противоположными спинами. Полное число n_i электронов в зоне проводимости, рассчитанное на единицу объема кристалла, определяется выражением

$$n_i = 2 \int_0^{E_t} N_c(E)P_e(E)dE, \quad (4.7)$$

где E_t — энергия, соответствующая потолку зоны проводимости. Аналогично число дырок $p(E)dE$ в валентной зоне, занимающих электронные уровни энергии в интервале от E до $E + dE$, равно

$$p(E)dE = 2N_v(E)P_h(E)dE. \quad (4.8)$$

Полное число дырок в валентной зоне в расчете на единицу объема равно

$$p_i = 2 \int_{E_b}^{-\Delta E} N_v(E)P_h(E)dE, \quad (4.9)$$

где E_b — энергия, соответствующая дну валентной зоны. В собственном полупроводнике полное число электронов в зоне проводимости должно совпадать с полным числом дырок в валентной зоне, поскольку электроны и дырки возникают и аннигилируют в кристалле парами, поэтому $n_1 = p_1$. Из этого равенства можно определить величину E_F , а следовательно, и вероятностные функции $P_e(E)$ и $P_h(E)$.

Чтобы выполнить такой расчет, необходимо знать конкретный вид функций $N_c(E)$ и $N_v(E)$. Рассмотрим сначала простейший тип полупроводника с одним сферически-симметричным минимумом энергии в зоне проводимости, которому соответствует скалярная эффективная масса m_e , а также с одним сферически-симметричным максимумом в валентной зоне со скалярной эффективной массой m_h . Такой случай уже обсуждался в разд. 2.3. Рассматривая граничные условия, накладываемые на волновой вектор k , можно с полной общностью показать, что независимо от конкретного вида $E(k)$ число разрешенных состояний в кристалле объема V в области dk в k -пространстве, равно

$$N(k) dk = V \frac{dk}{8\pi^3}. \quad (4.10)$$

Для электронов в свободном пространстве, для которых импульс дается соотношением $p = \hbar k$, из (4.10) следует, что существует один уровень на объем h^3 в пространстве импульсов и единичный объем в координатном пространстве, что является очень важным результатом (см. [2], § 5.5). Тогда число уровней, для которых величина волнового вектора k лежит между k и $k + dk$, дается выражением

$$N(k) dk = V k^2 \frac{dk}{2\pi^2}. \quad (4.11)$$

В полупроводнике со скалярной эффективной массой m_e в зоне проводимости имеем $E = \hbar^2 k^2 / 2m_e$, поэтому мы можем сразу получить число уровней в единичном объеме в интервале энергий от E до $E + dE$ (для одного направления спина):

$$N_c(E) dE = 2\pi (2m_e)^{3/2} h^{-3} E^{1/2} dE. \quad (4.12a)$$

Аналогично для валентной зоны со скалярной эффективной массой m_h :

$$N_v(E) dE = 2\pi (2m_h)^{3/2} h^{-3} (-\Delta E - E)^{1/2} dE. \quad (4.12b)$$

Из физических соображений ясно, что при $(\Delta E/kT) \gg 1$ вероятности заполнения энергетических уровней зоны проводимости электронами и уровней валентной зоны дырками будут очень малы. Поэтому при определении E_F можно предположить, что вырождение отсутствует как в валентной зоне, так и в зоне проводимости. Условия, при которых справедливо сделанное предположение, будут рассмотрены

ниже. Подставляя значение $N_c(E)$ из (4.12а) в (4.7), получаем

$$n_i = 4\pi (2m_e)^{3/2} h^{-3} \int_0^{\infty} \frac{E^{1/2} dE}{\exp\left(\frac{E-E_F}{kT}\right) + 1}. \quad (4.13)$$

При выводе этого равенства сделаны некоторые упрощения. Во-первых, мы предположили, что равенство (4.12а) выполняется при *любых* значениях E . Такое предположение оправдывается только в том случае, когда $(E-E_F) \ll kT$, и, следовательно, существенный вклад в интеграл в правой части (4.13), вносят лишь малые значения E . Во-вторых, по тем же соображениям верхний предел в интеграле (4.7) заменен на $+\infty$.

Аналогичным образом

$$p_i = 4\pi (2m_h)^{3/2} h^{-3} \int_{-\infty}^{-\Delta E} \frac{(-\Delta E - E)^{1/2} dE}{\exp\left(\frac{E_F - E}{kT}\right) + 1}. \quad (4.13a)$$

Поскольку, по предположению, в формуле (4.13) $(E-E_F)/kT \gg 1$, а в формуле (4.13а) $(E_F-E)/kT \gg 1$, то в подынтегральных выражениях в (4.13) и (4.13а) распределение Ферми — Дирака можно заменить простыми экспонентами. Введя в формулу (4.13) обозначение $x = E/kT$, имеем

$$n_i = 4\pi (2m_e)^{3/2} h^{-3} (kT)^{3/2} \exp\left(\frac{E_F}{kT}\right) \int_0^{\infty} x^{1/2} e^{-x} dx, \quad (4.14)$$

$$n_i = 2 \left(\frac{2\pi m_e kT}{h^2}\right)^{3/2} \exp\left(\frac{E_F}{kT}\right), \quad (4.15)$$

$$n_i = N_c \exp\left(\frac{E_F}{kT}\right). \quad (4.15a)$$

Аналогичным образом, обозначив в формуле (4.13а) величину $(-\Delta E - E)/kT$ через x , получим

$$p_i = 4\pi (2m_h)^{3/2} h^{-3} (kT)^{3/2} \exp\left[-\frac{E_F + \Delta E}{kT}\right] \int_0^{\infty} x^{1/2} e^{-x} dx = \\ = 2 \left(\frac{2\pi m_h kT}{h^2}\right)^{3/2} \exp\left[-\frac{E_F + \Delta E}{kT}\right], \quad (4.16)$$

$$p_i = N_v \exp\left[-\frac{E_F + \Delta E}{kT}\right]. \quad (4.16a)$$

Приравнявая p_i и n_i , получаем

$$m_e^{3/2} \exp\left(\frac{E_F}{kT}\right) = m_h^{3/2} \exp\left[-\frac{E_F + \Delta E}{kT}\right],$$

откуда

$$E_F = -\frac{1}{2}\Delta E + \frac{3}{4}kT \ln\left(\frac{m_h}{m_e}\right). \quad (4.17)$$

При $m_h = m_e$ уровень Ферми располагается в точности посередине между валентной зоной и зоной проводимости. В большинстве собственных полупроводников отклонение уровня Ферми от этого среднего положения при обычных температурах невелико. Однако в таких полупроводниках, как InSb, где $m_h/m_e \approx 20$ и $\Delta E \approx 0,2$ эВ, уровень Ферми при комнатных температурах ($kT \approx 0,025$ эВ) заметно сдвинут в сторону зоны проводимости¹⁾.

Подставив значение E_F из (4.17) в формулу (4.15), легко получить для n_i или p_i выражение

$$n_i = p_i = 2 \left(\frac{2\pi kT}{h^3}\right)^{3/2} (m_e m_h)^{3/2} \exp\left[-\frac{\Delta E}{2kT}\right]. \quad (4.18)$$

После подстановки численных значений физических констант h и k выражение (4.18) приобретает вид

$$n_i = p_i = 4,82 \cdot 10^{15} T^{3/2} \left(\frac{m_e m_h}{m_0^2}\right)^{3/2} \exp\left(-\frac{\Delta E}{2kT}\right) \text{ см}^{-3}. \quad (4.18a)$$

На рис. 4.4 показана зависимость n_i от $1/T$ (кривая 1) для германия, у которого при $T=300\text{К}$ $\Delta E \approx 0,665$ эВ (см. разд. 13.3). Из этого графика следует, что $\ln n_i$ почти линейно зависит от $1/T$. Едва заметное на рис. 4.4 отклонение от строгой пропорциональной зависимости обусловлено множителем $T^{3/2}$. В случае германия при комнатной температуре для состояний в зоне проводимости $(E - E_F) > 15kT$ и аналогично для состояний в валентной зоне $(E_F - E) > 15kT$, так что условия отсутствия вырождения выполняются с большой точностью.

Если использованное экспоненциальное приближение становится недопустимым, т. е. начинает сказываться вырождение, то выражение для n_i можно записать в виде

$$n_i = 4\pi \left(2m_e \frac{kT}{h^3}\right)^{3/2} F_{1/2}\left(\frac{E_F}{kT}\right), \quad (4.19)$$

где функция $F_n(\zeta)$ определяется интегралом

$$F_n(\zeta) = \int_0^{\infty} \frac{x^n dx}{\exp(x - \zeta) + 1}.$$

Таблицы функций $F_n(\zeta)$ для небольших полуцелочисленных значений n составлены МакДуглом и Стонером [3]. Для дырок при

¹⁾ В данном примере уровень Ферми сдвинут относительно середины запрещенной зоны к зоне проводимости на 0,056 эВ.— *Прим. ред.*

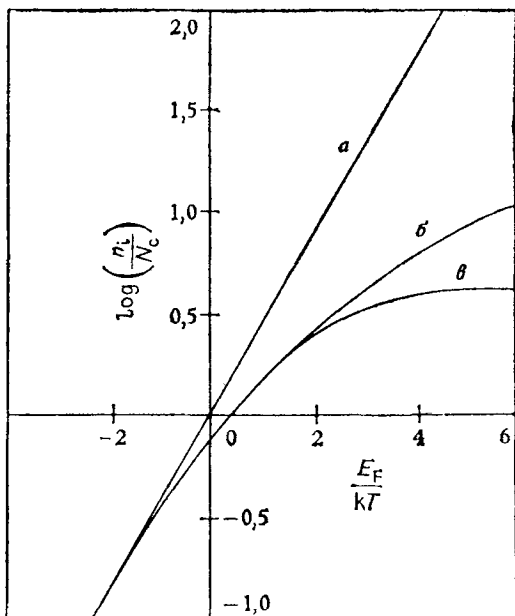


Рис. 4.2. Концентрация электронов в зоне проводимости как функция E_F/kT в различных приближениях.

a — классическое приближение; *б* — точное значение; *в* — приближение Эренберга,

наличии вырождения справедливо соотношение

$$p_i = 4\pi \left(2m_n \frac{kT}{h^2} \right)^{3/2} F_{1/2} \left[-\frac{E_F + \Delta E}{kT} \right]. \quad (4.20)$$

Энергия Ферми E_F определяется в этом случае из условия

$$m_e^{3/2} F_{1/2} \left[-\frac{E_F}{kT} \right] = m_n^{3/2} F_{1/2} \left[-\frac{E_F + \Delta E}{kT} \right]. \quad (4.21)$$

Вообще говоря, такое уравнение можно решить лишь численными методами. При $\zeta \gg 5$ функция $F_{1/2}(\zeta) \approx \frac{2}{3} \zeta^{3/2}$, и система считается при этом сильно вырожденной. Однако сильное вырождение имеет место в металлах, тогда как в обычных собственных полупроводниках оно отсутствует. При больших отрицательных значениях аргумента ζ функция $F_{1/2}(\zeta) \approx \frac{1}{2} \pi^{1/2} e^{\zeta}$, и система становится невырожденной. Эренберг [4] предложил следующую аппроксимацию для $F_{1/2}(\zeta)$:

$$F_{1/2}(\zeta) = 2\pi^{1/2} \frac{e^{\zeta}}{(4 + e^{\zeta})}. \quad (4.22)$$

На рис. 4.2 представлены графики зависимости n_1/N_c от E_F/kT , вычисленные в приближении простой экспоненты, в приближении Эренберга, а также по точной формуле (4.19). По этим графикам видно, что классическое приближение можно использовать с хорошей степенью точности уже при значениях $E_F/kT < -1$, т. е. оно справедливо до тех пор, пока уровень Ферми лежит по крайней мере на величину kT ниже дна зоны проводимости, чему соответствует $n < 0,4N_c$. Для более высоких значений n наступающее вырождение приводит уже к заметным отклонениям от точной формулы. Приближение Эренберга справедливо до $E_F/kT = 2$.

Для определения численных значений N_c и N_v [см. равенства (4.15а) и (4.16а)] необходимо знать величины эффективных масс m_e и m_h , которые известны далеко не всегда. Однако если энергия активации ΔE известна, то этого достаточно для того, чтобы оценить порядок величин n_1 и p_1 .

С другой стороны, зависимость n_1 от температуры можно использовать для определения ΔE . Несомненно, простейшим способом определения ΔE является изучение температурной зависимости собственной проводимости. Если значение ΔE достаточно велико, то изменение проводимости с температурой в основном будет обусловлено температурной зависимостью концентрации носителей заряда. Поэтому, предполагая, что удельная проводимость σ пропорциональна n_1 (так как $n_1 = p_1$), можно изобразить графически зависимость $\ln \sigma$ от $1/T$; при этом из равенства (4.18) следует, что она будет иметь вид почти прямой (если пренебречь членом $\ln T^{3/2}$) с наклоном $\Delta E/2k$. Такой метод определения ΔE весьма широко использовался, однако он может привести к существенным ошибкам, если применять его без разбора. Кроме того, для применения этого метода необходимо удостовериться в том, что измерения действительно относятся к области собственной проводимости. Позже мы подробнее обсудим различные методы определения ΔE . Когда имеется несколько эквивалентных минимумов энергии в зоне проводимости или несколько эквивалентных максимумов в валентной зоне (см. разд. 2.3), интегрирование по энергиям необходимо произвести в каждом экстремуме в отдельности. Однако поскольку вклад каждого из них одинаков, то выражение (4.15) для n_1 необходимо просто умножить на M_c — число эквивалентных минимумов в зоне проводимости. Аналогичным образом в (4.16) появляется M_v — число эквивалентных максимумов в валентной зоне. Для зонного спектра со многими экстремумами сферическая симметрия в k -пространстве в общем случае отсутствует. Если тензор обратной эффективной массы электронов в зоне проводимости можно привести к диагональному виду, так что изоэнергетические поверхности будут представлять собой эллипсоиды вращения с эффективными массами m_1 и m_2 , то для E можно воспользоваться

выражением (2.23). При помощи простого преобразования

$$k'_x = \left(\frac{m_0}{m_1}\right) k_x, \quad k'_y = \left(\frac{m_0}{m_2}\right) k_y, \quad k'_z = \left(\frac{m_0}{m_2}\right) k_z$$

мы можем снова привести выражение для E к «сферической» форме и далее, действуя как и ранее, получить $N(E)$, однако заменив в (4.12) m_e на «эффективную массу плотности состояний» $m_{де}$, где $m_{де}^3 = m_1 m_2^2$. Тогда для n_1 получаем

$$n_1 = 2m_1^{1/2} m_2 M_c h^{-3} (2\pi kT)^{3/2} \exp\left(\frac{E_F}{kT}\right). \quad (4.23)$$

Если валентная зона вырождена в точке $\mathbf{k}=0$ (как это имеет место в Ge и Si), то изоэнергетические поверхности имеют довольно сложный вид и точный расчет оказывается чрезвычайно громоздким¹⁾. Однако можно аппроксимировать имеющиеся поверхности равной энергии двумя сферически-симметричными поверхностями, которым соответствуют эффективные массы m_{h1} и m_{h2} . Тогда вычисления можно провести без особого труда, так как дело сводится к интегрированию по состояниям в каждой сферически-симметричной зоне, взятой в отдельности. В результате получаем

$$p_1 = 2(m_{h1}^{3/2} + m_{h2}^{3/2}) h^{-3} (2\pi kT)^{3/2} \exp\left(-\frac{\Delta E + E_F}{kT}\right). \quad (4.24)$$

«Эффективная масса плотности состояний» в этом случае определяется соотношением $m_{дн}^3 = m_{h1}^3 + m_{h2}^3$. В каждом из рассмотренных случаев энергию Ферми E_F можно получить, приравняв n_1 и p_1 в точности так, как это мы делали ранее.

Так как равенства (4.15) и (4.23) можно записать в виде [ср. с выражением (4.15а)]

$$n_1 = N_c \exp\left(\frac{E_F}{kT}\right), \quad (4.25)$$

то зону проводимости в этом приближении можно рассматривать как одиночный уровень энергии с фактором вырождения N_c , расположенный на дне зоны. Это следует также из формулы (4.3), если положить в ней $E=0$ и $-E_F \gg kT$. Действительно, среднее число электронов, занимающих g -кратно вырожденный уровень при $E=0$ и $-E_F \gg kT$, согласно (4.3), равно $g \exp(E_F/kT)$. Сравнивая с (4.25), можно убедиться в том, что эти формулы совпадают между собой при $g=N_c$. Аналогичным образом равенства (4.16) или (4.24) можно записать в виде

$$p_1 = N_v \exp\left(-\frac{\Delta E + E_F}{kT}\right). \quad (4.26)$$

¹⁾ Для Ge и Si такой расчет выполнен в работе [5].

Таким образом, валентную зону можно рассматривать как одиночный уровень энергии при $E = -\Delta E$, имеющий фактор вырождения N_v . Вследствие того что для собственного полупроводника $n_1 = p_1$, в отсутствие вырождения в зонах

$$n_i^2 = N_c N_v \exp\left(-\frac{\Delta E}{kT}\right). \quad (4.27)$$

Соотношение для энергии Ферми, выраженное через N_c и N_v , имеет вид

$$E_F = -\frac{1}{2}\Delta E + \frac{1}{2}kT \ln\left(\frac{N_v}{N_c}\right). \quad (4.28)$$

4.3. Примесные полупроводники

При рассмотрении распределения электронов по примесным уровням, лежащим между валентной зоной и зоной проводимости, мы сразу сталкиваемся со следующей трудностью. В выражении для числа электронов на примесных уровнях необходимо учесть то обстоятельство, что, с одной стороны, электрон может пребывать на атоме примеси в двух различных состояниях, соответствующих противоположным направлениям спина, с другой — примесный уровень не может считаться дважды вырожденным, поскольку наличие на нем одного электрона полностью исключает возможность заселения его любым другим электроном. При этом могут иметь место различные ситуации, некоторые из них мы рассмотрим подробнее. Необходимо отметить, что приближенные зависимости для определения числа электронов в зоне проводимости, полученные в разд. 4.2, применимы и в рассматриваемом здесь случае, если уровень Ферми лежит ниже дна зоны проводимости, по крайней мере на величину порядка $2kT$, так что электроны в зоне могут считаться невырожденными. Таким образом, концентрация n электронов в зоне проводимости по-прежнему подчиняется соотношению (4.15a), т. е.

$$n = N_c \exp\left(\frac{E_F}{kT}\right). \quad (4.29)$$

Однако в данном случае величина E_F уже не определяется формулой (4.28) или эквивалентной ей зависимостью, а является функцией концентрации примесей. Аналогичным образом если дырки в валентной зоне можно считать невырожденными, то для концентрации p дырок в валентной зоне получаем

$$p = N_v \exp\left(-\frac{E_F + \Delta E}{kT}\right). \quad (4.30)$$

Таким образом, в отсутствие вырождения по-прежнему остается справедливым выражение

$$np = N_c N_v \exp\left(-\frac{\Delta E}{kT}\right) = n_i^2 \quad (4.31)$$

[ср. равенство (4.27)]. Это очень важное соотношение, показывающее, что в отсутствие фермиевского вырождения, т. е. когда малы концентрации электронов в зоне проводимости и дырок в валентной зоне, произведение np не зависит от концентрации и распределения примесей в полупроводнике.

Зависимость типа (4.31) можно получить и другим путем, хотя и менее строгим, но весьма поучительным. Если рассматривать возникновение и рекомбинацию электрона (e) и дырки (h) как химическую реакцию, приводящую к образованию нормального атома или иона в кристалле (N), то такую реакцию можно описывать уравнением

$$(e) + (h) = (N) + \Delta E. \quad (4.32)$$

Из закона действующих масс (ср. (1.2) и (1.3)) следует

$$np = K(T), \quad (4.33)$$

где $K(T)$ имеет вид $[F(T)]^2 \exp(-\Delta E/kT)$, причем $F(T)$ — функция, медленно меняющаяся с температурой. Равенство (4.33) эквивалентно равенству (4.31), в котором приводится вид функции $F(T)$. Если нам известна концентрация электронов в зоне проводимости, то очень легко проверить степень вырождения в полупроводнике. Пусть для примера $T=300\text{K}$, тогда при $m_e \approx m_h \approx m_0$ величина $N_c \approx 2,5 \cdot 10^{19} \text{ см}^{-3}$. Поэтому в области концентраций $n < 10^{19} \text{ см}^{-3}$ для E_F получаем $E_F < -2kT$, т. е. электронный газ невырожден. С другой стороны, если $n \approx 2 \cdot 10^{19} \text{ см}^{-3}$, то $E_F > -2kT$, и исходное предположение перестает быть справедливым, т. е. наступает вырождение электронов в зоне проводимости. Если $m_e \ll m_0$, как это имеет место в InSb, то вырождение в зоне проводимости наступает при значительно меньших концентрациях электронов.

Прежде чем приступить к исследованию более сложных случаев, рассмотрим полупроводник, содержащий N_d донорных уровней в единице объема. Предположим, что донорные уровни расположены в непосредственной близости от дна зоны проводимости, так что энергия ионизации примесей ϵ_d очень мала по сравнению с шириной запрещенной зоны. Мы уже видели, что такие уровни довольно типичны, например в Ge, где $\epsilon_d \approx 0,01 \text{ эВ}$ при $\Delta E \approx 0,7 \text{ эВ}$. Предположим также, что уровень Ферми расположен существенно ниже дна зоны проводимости, т. е. $E_F \ll -kT$. Вследствие малости ϵ_d практически все атомы примеси будут ионизованными и их электроны окажутся в зоне проводимости. Выясним сперва, какова

предельная концентрация примесей, при которой эти условия перестают быть справедливыми. Для начала допустим, что $N_d \gg n_1$, так что можно пренебречь «собственными» электронами, т. е. теми, которые перебрасываются в зону проводимости из валентной зоны. В этом случае мы имеем простое соотношение $n = N_d$. В этих условиях число электронов в зоне проводимости будет слабо зависеть от температуры. Такое состояние полупроводника обычно называют состоянием «примесного истощения». Полупроводники, содержащие доноры в состоянии примесного истощения, называют несовершенными полупроводниками n -типа¹⁾.

Положение уровня Ферми в этих условиях определяется уравнением

$$n = N_d = N_c \exp\left(\frac{E_F}{kT}\right)$$

или

$$E_F = kT \ln\left(\frac{N_d}{N_c}\right). \quad (4.34)$$

Формула (4.34) справедлива лишь при условии $E_F < -kT$, т. е. при $N_d \leq N_c/2$. Для Ge при $T=300\text{K}$ это означает, что для применимости равенства (4.34) требуется, чтобы $N_d < 10^{19} \text{ см}^{-3}$; с другой стороны, однако, должно выполняться условие $N_d > 10^{14} \text{ см}^{-3}$, чтобы можно было пренебречь концентрацией собственных носителей заряда. Итак, в Ge при комнатных температурах существует довольно широкая область значений N_d , для которых выполняются условия примесного истощения.

Из формулы (4.34) следует, что с ростом N_d уровень Ферми перемещается вверх примерно с середины запрещенной зоны до значения энергии примерно на kT ниже дна зоны проводимости, когда $N_d \rightarrow N_c$. Если $N_d \geq N_c$, то система электронов в зоне проводимости становится вырожденной и для определения положения уровня Ферми необходимо воспользоваться уже равенством типа (4.19). В этом случае полупроводник будет вести себя скорее как металл.

Интересно выяснить, что происходит в более общем случае, когда N_d того же порядка величины что и n_1 , но при этом $N_d \ll N_c$. Ясно, что, если $\Delta E \gg kT$, уровень Ферми будет находиться достаточно далеко от дна зоны проводимости, на основании чего по-прежнему можно предполагать, что все атомы примеси ионизованы. Тогда между n и p имеет место простое соотношение

$$n - p = N_d, \quad (4.35)$$

¹⁾ Термин «несовершенный полупроводник», как правило, используется в более широком смысле. Несобственными полупроводниками обычно называют примесные полупроводники, у которых концентрация носителей заряда, обусловленная ионизацией атомов примеси, существенно превышает концентрацию «собственных» носителей заряда.— *Прим. ред.*

выражающее условие электрической нейтральности полупроводника. На основании равенств (4.31) и (4.35) можно выразить n и p через n_1 и N_d и получить

$$n = \frac{1}{2} N_d \left[1 + \left(1 + \frac{4n_1^2}{N_d^2} \right)^{1/2} \right], \quad (4.36)$$

$$p = \frac{2n_1^2}{N_d} \left[1 + \left(1 + \frac{4n_1^2}{N_d^2} \right)^{1/2} \right]^{-1}$$

Если $N_d \gg n_1$ (а также $N_d < N_c$), то приближенно имеем

$$n = N_d + \frac{n_1^2}{N_d}, \quad (4.37)$$

$$p = \frac{n_1^2}{N_d}.$$

Заметим, что концентрация дырок в этом случае намного меньше концентрации дырок в собственном веществе (в отношении n_i/N_d). Например, при $n_1 = 2,37 \cdot 10^{13} \text{ см}^{-3}$ (как для Ge) и $N_d \approx 10^{16} \text{ см}^{-3}$ мы получаем $n \approx 10^{16} \text{ см}^{-3}$ и $p \approx 5,6 \cdot 10^{10} \text{ см}^{-3}$. В этом случае электроны принято называть основными носителями заряда, а дырки — неосновными носителями заряда. Несмотря на малость концентрации неосновных носителей по сравнению с концентрацией основных, первые могут играть в известных условиях очень важную роль. Предположив, что имеет место обратное неравенство $N_d \ll n_1$, мы приближенно получаем

$$n = n_1 + \frac{1}{2} N_d, \quad (4.38)$$

$$p = n_1 - \frac{1}{2} N_d.$$

Предположим теперь, что вместо донорной примеси имеется N_a акцепторных атомов в единице объема, уровни энергии которых расположены непосредственно над потолком валентной зоны. Аналогично предыдущему будем считать, что уровень Ферми расположен выше потолка валентной зоны по крайней мере на несколько kT . Тогда почти все акцепторные уровни будут заняты электронами из валентной зоны, т. е. примесные атомы будут «ионизованными», причем в валентной зоне останутся около N_a свободных дырок, и мы получим

$$p = N_a = N_v \exp \left(-\frac{\Delta E + E_F}{kT} \right). \quad (4.39)$$

Эта формула справедлива лишь при условии, что отношение $N_a/N_v \ll 1/2$. Снова с ростом N_a от 0 до N_v уровень Ферми смещается примерно с середины запрещенной зоны до потолка валентной зоны.

Ясно, что вместо равенства (4.35) теперь имеем

$$p - n = N_a, \quad (4.40)$$

что вместе с формулой (4.31) дает

$$\begin{aligned} n &= \frac{2n_i^2}{N_a} \left[1 + \left(1 + \frac{4n_i^2}{N_a^2} \right)^{1/2} \right]^{-1}, \\ p &= \frac{N_a}{2} \left[1 + \left(1 + \frac{4n_i^2}{N_a^2} \right)^{1/2} \right]. \end{aligned} \quad (4.41)$$

При $N_a \gg n_i$ приближенно имеем

$$\begin{aligned} n &= \frac{n_i^2}{N_a}, \\ p &= N_a + \frac{n_i^2}{N_a}. \end{aligned} \quad (4.42)$$

Если же $n_i \gg N_a$, то приближенно

$$\begin{aligned} n &= n_i - \frac{1}{2} N_a, \\ p &= n_i + \frac{1}{2} N_a. \end{aligned} \quad (4.43)$$

Допустим теперь, что в кристалле одновременно присутствуют как донорные, так и акцепторные примеси. Предположим вначале, что $(N_d - N_a) \gg n_i$. При $N_a = 0$ уровень Ферми окажется существенно выше середины запрещенной зоны. При добавлении в кристалл акцепторов вероятность заполнения соответствующих им уровней, т. е. вероятность «ионизации» акцепторных атомов очень велика, поскольку эти акцепторные уровни лежат значительно ниже уровня Ферми. Электроны, заполняя акцепторные уровни (по одному электрону на уровень), переходят с донорных уровней или из зоны проводимости. Число свободных дырок чрезвычайно мало, так как уровень Ферми располагается достаточно высоко над краем валентной зоны. Поэтому эффективное число доноров равно $N_d - N_a$, и положение уровня Ферми определяется из соотношения

$$N_d - N_a = N_c \exp \left(\frac{E_F}{kT} \right). \quad (4.44)$$

При $N_a \rightarrow N_d$ акцепторы все в большей степени нейтрализуют действие доноров и уровень Ферми постепенно перемещается к середине запрещенной зоны. При $N_a > N_d$ эффективное число акцепторов становится равным $N_a - N_d$, все электроны с донорных центров переходят на акцепторные уровни, уровень Ферми опускается еще ниже и в пределе оказывается у самого края валентной зоны. Зависимость E_F от концентрации примесных атомов пока-

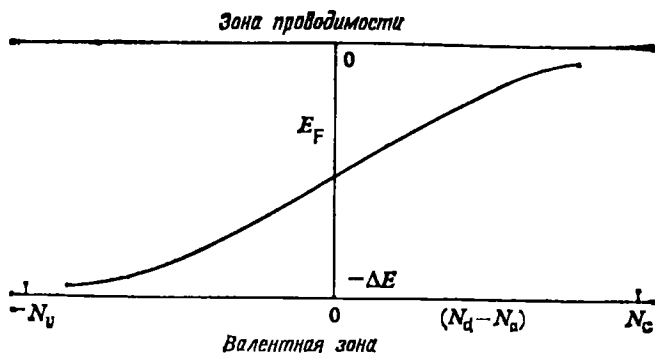


Рис. 4.3. Изменение положения уровня Ферми в зависимости от концентраций доноров и акцепторов,

зона на рис. 4.3 (фактический ход кривой зависит от температуры). Серия кривых, изображающих температурную зависимость n и p при различных значениях $N_d - N_a$ для Ge, представлена на рис. 4.4.

До сих пор предполагалось, что уровни доноров и акцепторов лежат очень близко к краям соответствующих зон. Однако на самом деле эти уровни отделены от краев зон энергетическими зазорами конечной величины ϵ_d и ϵ_a соответственно. Если ϵ_d и ϵ_a малы и $|N_d - N_a|$ больше N_c или N_v , то, за исключением области очень низких температур, действительно можно считать, что примеси полностью ионизованы. Рассмотрим теперь весьма важную область низких температур, когда ионизация примесей перестает быть полной, а также такие примесные уровни, которые отделены от краев зон достаточно большими энергетическими зазорами, соизмеримыми со значением ΔE . Для исследования таких более сложных ситуаций необходимо сначала подробнее остановиться на вопросе о вырождении примесных уровней.

Если бы каждый примесный уровень мог принять по два электрона (по одному для каждого направления спина), то вероятность заполнения примесного уровня электроном с произвольным направлением спина описывалась бы формулой типа (4.1), где вместо E необходимо подставить либо $-\epsilon_d$ в случае донорных уровней, либо $-\Delta E + \epsilon_a$ в случае акцепторных. Тогда среднее число электронов n_d на донорных уровнях равно

$$n_d = \frac{2N_d}{\exp\left(-\frac{\epsilon_d - E_F}{kT}\right) + 1}. \quad (4.45)$$

Однако такие условия заполнения примесных уровней доноров и акцепторов в полупроводниках почти никогда не реализуются.

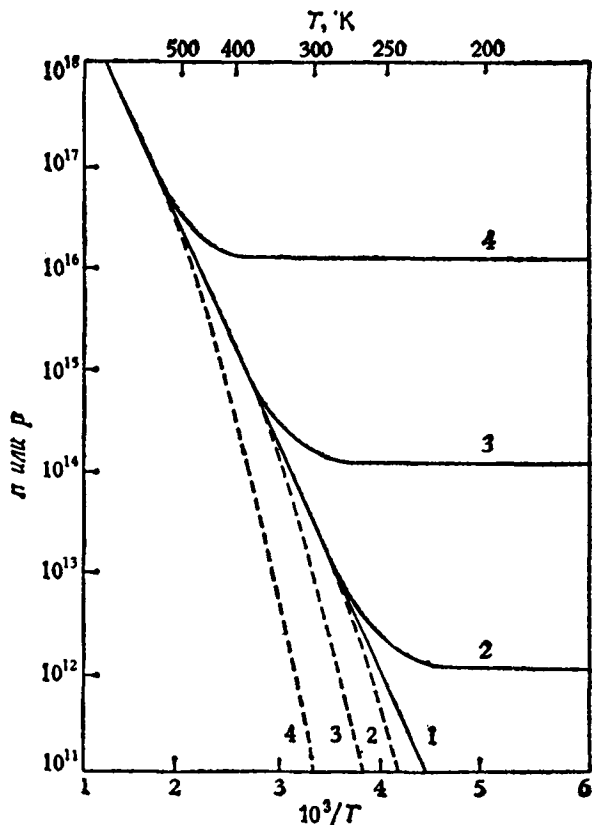


Рис. 4.4. Температурная зависимость n и p для Ge. Сплошные кривые для n , пунктирные — для p .

1) $N_d - N_a = 0$ (кривая для собственного полупроводника); 2) $N_d - N_a = 10^{12} \text{ см}^{-3}$; 3) $N_d - N_a = 10^{14} \text{ см}^{-3}$; 4) $N_d - N_a = 10^{16} \text{ см}^{-3}$.

Обычно заполнение примесных уровней определяется двумя условиями:

1) Примесный (донорный) уровень может либо захватить *один* электрон с произвольным направлением спина, либо оставаться незаполненным.

2) Примесный (акцепторный) уровень может содержать либо два электрона со спаренными спинами, либо один электрон с произвольным направлением спина. Используя соответствующую терминологию, второе условие можно сформулировать в той же форме, что и первое, но для дырок.

Можно показать, что при выполнении первого условия вероятность заполнения примесного уровня электроном σ *произвольным*

направлением спина дается выражением

$$P(E) = \frac{1}{1 + \frac{1}{2} \exp \frac{E - E_F}{kT}}. \quad (4.46)$$

Первое условие выполняется в полупроводниковых кристаллах типа Ge и Si, содержащих в качестве донорной примеси пентавалентные атомы элементов V группы. Четыре электрона атома примеси образуют с электронами кристалла систему ковалентных связей со спаренными спинами, тогда как пятый электрон может обладать спином любого направления. Если N_d — концентрация такой донорной примеси, то количество неионизованных доноров n_d определяется соотношением

$$n_d = \frac{N_d}{1 + \frac{1}{2} \exp \frac{E - E_F}{kT}}, \quad (4.47)$$

где $E = -\varepsilon_d$.

Можно показать, что в случае, когда выполняется второе условие, вероятность $P(E)$ заполнения примесного уровня E двумя электронами со спаренными спинами равна

$$P(E) = \frac{1}{1 + 2 \exp \frac{E_F - E}{kT}}. \quad (4.48)$$

В полупроводниках IV группы второму условию соответствуют примеси элементов III группы, атомы которых в условиях ионизации (т. е. после присоединения к себе одного добавочного электрона) обладают четырьмя электронами со спаренными спинами. Число неионизованных акцепторов определяется равенством

$$n_a = N_a [1 - P(E)] = \frac{N_a}{1 + \frac{1}{2} \exp \frac{E_F - E}{kT}}, \quad (4.49)$$

где $E = -\Delta E + \varepsilon_a$.

Здесь явно видна полная аналогия между формулами (4.47) и (4.49), в которых для доноров использована электронная терминология, а для акцепторов — дырочная. Для вывода равенств (4.46) и (4.48), а так же других эквивалентных им зависимостей были использованы различные методы. Однако эти формулы обладают тем же фундаментальным характером, что и равенство (4.1), и поэтому могут быть выведены непосредственно из основных положений статистики Ферми — Дирака¹⁾.

При выводе формул (4.47) и (4.49) из выражений для вероятностей заполнения $P(E)$ неявно предполагалось наличие только одного

¹⁾ См., например, [6], стр. 326—328.

уровня у каждой примеси. Однако мы уже видели, что примеси в кристалле наряду с основными уровнями могут обладать системой возбужденных уровней. Поэтому в равенстве типа (4.47) следует учесть и эти уровни, причем с соответствующими весовыми множителями. Тогда имеем

$$n_d = \frac{N_d}{1 + \left(\sum g_r \exp \frac{E_F - E_r}{kT} \right)^{-1}}, \quad (4.50)$$

где E_r — энергия r -го возбужденного состояния, g_r — числа, учитывающие спиновые и другие виды вырождения. Для доноров, принимающих на свои уровни только по одному электрону, $g_0 = 2$, $E_0 = -\varepsilon_d$. Аналогичную формулу можно написать для акцепторов. В двух наиболее важных с практической точки зрения случаях возбужденные состояния не играют существенной роли. Первый имеет место тогда, когда $(E_0 - E_F)/kT \gg 1$. Поэтому все члены суммы в правой части (4.50) малы, мало также отношение n_d/N_d , и почти все уровни примеси можно считать ионизованными. Второй случай имеет место при весьма низких температурах. Если $(E_1 - E_0)/kT \gg 1$, то первый член суммы в правой части (4.50) преобладает над всеми остальными и равенство (4.47) оказывается весьма хорошим приближением. Поэтому в дальнейшем мы отвлечемся от возбужденных состояний примеси. Влияние уровней возбуждения примеси подробно исследовали Шифрин [7], а также Бурштейн и др. [8].

Перейдем к вопросу об определении концентраций электронов и дырок n и p в полупроводнике, содержащем N_d донорных атомов с примесными уровнями на глубине ε_d под дном зоны проводимости и N_a акцепторных атомов с примесными уровнями на высоте ε_a над потолком валентной зоны. Уравнение, выражающее условие электрической нейтральности кристалла и определяющее положение уровня Ферми, можно записать в виде

$$n + n_d + N_a = p + N_d + n_a. \quad (4.51)$$

Все члены этого равенства выражаются через E_F с помощью формул (4.29), (4.30), (4.47), (4.49), и мы приходим к уравнению четвертой степени относительно $\exp(E_F/kT)$. Для получения простых аналитических решений необходимо сделать дальнейшие упрощения, справедливость которых ограничена конкретными условиями. Предположим для начала, что $|N_a - N_d| \ll N_c$ или N_v . Допустим также, что уровень Ферми расположен на глубине по крайней мере на несколько kT ниже края зоны проводимости, а также уровней примеси, но вместе с тем выше положения уровня Ферми в собственном полупроводнике. Легко убедиться в том, что при сделанных предположениях n_a пренебрежимо мало по сравнению с N_a и может быть опущено в (4.51). Пока уровень Ферми не достигнет значения $-\varepsilon_d$, значение n_d также будет мало по сравнению с N_d .

Это означает, что в данных условиях практически все доноры и акцепторы ионизованы. Тогда получаем

$$N_c \exp \frac{E_F}{kT} - N_v \exp \frac{-E_F + \Delta E}{kT} = N_d - N_a. \quad (4.52)$$

Предположение о том, что уровень Ферми сдвинут вверх по отношению к положению этого уровня в собственном полупроводнике, означает, что правая часть (4.52) положительна и, следовательно, $N_d > N_a$. Если бы сдвиг уровня Ферми имел другой знак, то опять получилось бы равенство (4.52), но уже с $N_a > N_d$. Формула (4.52) аналогична формуле (4.35), в которой N_d заменено на $N_d - N_a$ и приводит к уже обсужденным выше результатам. При высоких температурах члены, фигурирующие в левой части (4.52), становятся преобладающими, и положение уровня Ферми стремится к положению, характерному для собственного полупроводника. Однако с понижением температуры второй член в левой части уравнения, определяющий концентрацию дырок, становится мал по сравнению с первым, и мы получаем

$$E_F = kT \ln \left(\frac{N_d - N_a}{N_c} \right). \quad (4.53)$$

При достаточно низких температурах нельзя уже предполагать, что уровень Ферми располагается ниже уровней донорной примеси и полученные выше зависимости перестают быть справедливыми. С понижением температуры уровень Ферми перемещается к примесным донорным уровням и система доноров становится частично заполненной.

К счастью, в этом случае можно пренебречь вкладом свободных дырок в уравнении (4.51), если n_a мало. Не делая предположения о малости n_d по сравнению с N_d , мы получаем следующее выражение для определения уровня Ферми:

$$n = N_d - N_a - \frac{N_d}{1 + \frac{1}{2} \exp \left[-\frac{(\epsilon_d + E)}{kT} \right]}. \quad (4.54)$$

При $n = N_c \exp(E_F/kT)$ отсюда получаем следующее квадратное уравнение относительно n или $\exp(E_F/kT)$:

$$n^2 + n(N_a + n_1) - n_1(N_d - N_a) = 0, \quad (4.55)$$

где $n_1 = \frac{1}{2} N_c \exp(-\epsilon_d/kT)$. Соответствующее решение этого уравнения имеет вид

$$n = -\frac{1}{2}(N_a + n_1) + \frac{1}{2}[(N_a + n_1)^2 + 4n_1(N_d - N_a)]^{1/2}. \quad (4.56)$$

При $\epsilon_d/kT \gg 1$, n_1 — малое число и решение (4.56) можно заменить приближенным, но более простым выражением. При этом необхо-

дима, однако, некоторая осторожность, так как разложение в ряд корня в правой части равенства (4.56) возможно лишь при $n_1 \gg N_a$. Такое приближение справедливо лишь для частично скомпенсированных полупроводников. В этом случае приближенно имеем

$$n = \frac{n_1 (N_d - N_a)}{N_a} = \frac{N_d - N_a}{2N_a} N_c \exp\left(-\frac{\epsilon_d}{kT}\right). \quad (4.57)$$

Если N_a очень мало, так что $N_a \ll n_1 \ll N_d$, то оказывается справедливым следующее приближенное выражение

$$n = (n_1 N_d)^{1/2} = \frac{1}{\sqrt{2}} (N_d N_c)^{1/2} \exp\left(-\frac{\epsilon_d}{2kT}\right). \quad (4.58)$$

Эти формулы по существу тем же путем получены впервые де Буром и Ван Геелем [9]. Как показали Мотт и Герни (см. [10], с. 156—160), эти соотношения можно получить также из условия минимума изменения свободной энергии, связанного с переходами электронов с донорных уровней в зону проводимости. Заметим, что в указанных выше условиях график зависимости $\ln n$ от $1/T$ имеет вид прямых с наклонами ϵ_d/kT и $1/2(\epsilon_d/kT)$. Этим можно воспользоваться для определения значения ϵ_d для доноров в германии и кремнии. Однако ясно, что этот метод следует применять с осторожностью, лишь убедившись, что соответствующие условия выполняются. Аналогичные формулы легко получить и для случая полупроводников, в которых преобладает акцепторная примесь.

Формулы (4.57) и (4.58) указывают на то, что при понижении температуры должны исчезнуть все свободные носители заряда. Как правило, в области температур жидкого водорода и жидкого гелия концентрация носителей заряда действительно сильно убывает и обычно можно получить кривые, наклон которых позволяет определить значения ϵ_d или ϵ_a . Однако при дальнейшем понижении температуры оказалось, что свободные носители заряда вовсе не исчезают, как это предсказывает простая теория. Это явление, известное обычно под названием проводимости по примесным зонам, будет рассмотрено позже (см. разд. 5.4). Если отвлечься от подобных эффектов, то общий ход зависимости концентрации основных носителей заряда от $1/T$ имеет вид, показанный на рис. 4.5¹⁾.

¹⁾ При некоторых соотношениях между концентрациями донорной и акцепторной примесей на графике зависимости $\ln n$ от $1/T$ могут проявиться три разных наклона. Так, при $N_a < \frac{1}{2} N_d$ в области наименьших температур имеет место участок с наклоном ϵ_d/k , при более высоких температурах появляется участок с наклоном $\epsilon_d/2k$, и, наконец, при дальнейшем увеличении температуры наступает источник доноров, когда наклон кривой обращается в нуль. Переход от участка с наклоном ϵ_d/k к участку, имеющему наклон $\epsilon_d/2k$, соответствует равенству $n = N_a$. При $N_d > N_a \geq \frac{1}{2} N_d$ участок с наклоном $\epsilon_d/2k$ не возникает. — Прим. ред.

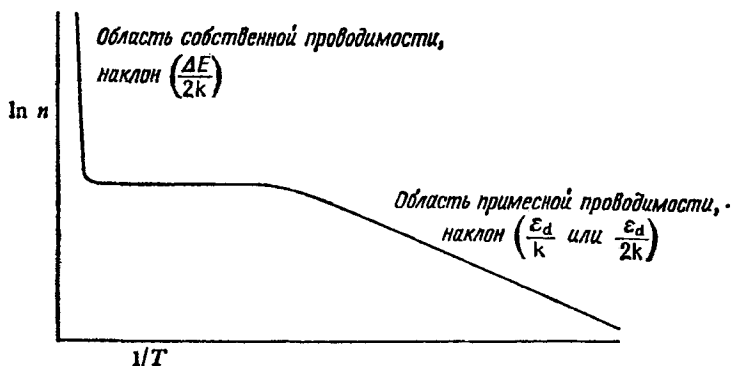


Рис. 4.5. Температурная зависимость концентрации носителей заряда в полупроводнике n -типа при низких температурах.

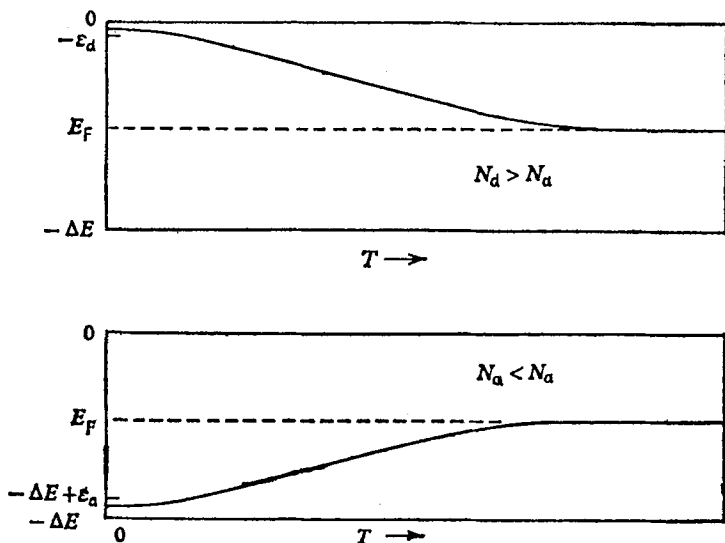


Рис. 4.6. Температурная зависимость положения уровня Ферми. Пунктиром обозначено положение уровня Ферми в области собственной проводимости.

Положение уровня Ферми в условиях применимости равенства (4.57) определяется по формуле

$$E_F = -\varepsilon_d + kT \ln \left(\frac{N_d - N_a}{2N_a} \right), \quad (4.59)$$

т. е. уровень Ферми расположен несколько выше уровня $E = -\varepsilon_d$. Если положение уровня Ферми определяется формулой (4.58),

например, при $N_a=0$, то

$$E_F = -\frac{1}{2} \epsilon_d + \frac{1}{2} kT \ln \left(\frac{N_d}{N_c} \right). \quad (4.60)$$

Поэтому в полупроводнике, содержащем лишь донорную примесь, уровень Ферми располагается примерно посередине между донорным уровнем и дном зоны проводимости. Аналогичные соотношения имеют место при $N_a > N_d$. В этом случае уровень Ферми располагается либо непосредственно под уровнями акцепторов, либо посередине между потолком валентной зоны и акцепторными уровнями. Общий ход изменения положения уровня Ферми с температурой показан на рис. 4.6. Во всех приведенных выше вычислениях предполагалось, что рассматриваемые полупроводники невырожденные. Для области температур, где наступает вырождение, вычисления усложняются¹⁾. При достаточно низких температурах вследствие исчезновения электронов из зоны проводимости вырождение снимается.

ЛИТЕРАТУРА

1. *Schrödinger E.*, Statistical Thermodynamics, Cambridge, 1946. [Имеется перевод: Шредингер Э. Статистическая термодинамика.— М.: ИЛ, 1948.]
2. *Smith R. A.*, Wave Mechanics of Crystalline Solids, 2nd ed., New York, 1969.
3. *McDougall J., Stoner E. C.*, Phil. Trans., A237, 67 (1938).
4. *Ehrenberg W.*, Proc. Phys. Soc., A63, 75 (1950).
5. *Lax B., Mavroides J. G.*, Phys. Rev., 100, 1650 (1955).
6. *Wilson A. H.*, The Theory of Metals, 2nd ed., Cambridge University Press, 1953.
7. *Шифрин К. С.*, ЖТФ, 14, 43 (1944).
8. *Burstein E. et al.*, Journ. Phys. Chem., 57, 849 (1953).
9. *de Boer J. H., van Geel W. C.*, Physica, 2, 186 (1935).
10. *Mott N. F., Gurney R. W.*, Electronic Processes in Ionic Crystals, Oxford University Press, Oxford, 1940. [Имеется перевод: Мотт Н., Герни Р. Электронные процессы в ионных кристаллах.— М.: ИЛ, 1950.]
11. *Blakemore J. S.*, Semiconductor Statistics, Pergamon Press, 1962. [Имеется перевод: Блекмор Дж. Статистика электронов и дырок в полупроводниках.— М.: Мир, 1964.]

¹⁾ См., например, [11].

ЯВЛЕНИЯ ЭЛЕКТРОННОГО ПЕРЕНОСА

5.1. Рассеяние электронов на дефектах в кристалле. Время релаксации

В этой главе будут рассмотрены различные кинетические явления, связанные с движением электронов и дырок в полупроводниках под действием электрического и магнитного полей. Излагаемая теория основана на понятии эффективной массы, введенном в разд. 2.5 с помощью основного уравнения

$$\dot{\mathbf{p}} = \hbar \dot{\mathbf{k}} = \mathbf{F}, \quad (5.1)$$

где \mathbf{F} — сила, действующая на электрон. Аналогичное уравнение движения справедливо и для дырок (см. разд. 2.5). Рассмотрим для простоты одномерный случай движения электрона в направлении оси x ¹⁾. Предположим, что минимум энергии в зоне проводимости расположен в точке $\mathbf{k} = 0$ и что на границе зоны Бриллюэна имеется единственный максимум при $k_x = k_d$. Вблизи $k_x = 0$ эффективная масса m_e положительна, и тогда

$$m_e \frac{dv_x}{dt} = F_x. \quad (5.1a)$$

При $F_x > 0$ электрон ускоряется в направлении оси x . Под действием силы F_x значение k_x электрона будет равномерно расти, пока не достигнет точки k_d . Затем значение k_x окажется за пределами первой зоны Бриллюэна, что равносильно вторичному появлению электрона в пределах первой приведенной зоны в точке $k_x = -k_d$. По мере приближения значения k_x к k_d эффективная масса электрона становится отрицательной (см. разд. 2.5), так что электрон при этом замедляется, и, наконец, при $k_x \rightarrow k_d$ он совсем останавливается, так как при $k_x = k_d$ величина $v_x = 0$. Затем электрон начнет ускоряться, но уже в обратном направлении, так как в этой области значений k_x эффективная масса оказывается отрицатель-

¹⁾ В данном случае предполагается, что при своем движении электрон не испытывает рассеяния (соударений). — *Прим. ред.*

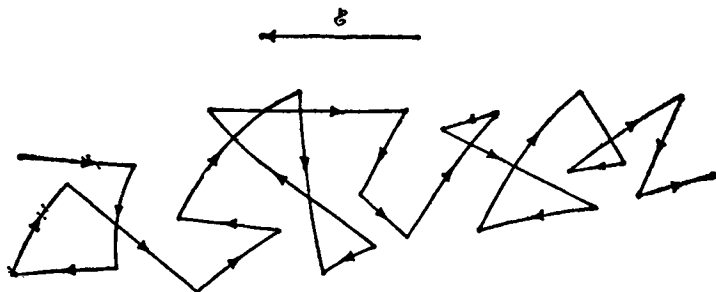


Рис. 5.1. Траектория электрона в кристалле под действием электрического поля \mathcal{E} .

ной, и снова придет в состояние покоя при $k_x=0$, где эффективная масса положительна. Это соответствует явлению брэгговского отражения от кристаллических плоскостей (см. [1], § 5.8). Таким образом, под влиянием постоянного внешнего поля электрон в отсутствие рассеяния должен совершать колебания по направлению оси x около начала координат. На самом деле, однако, в реальных кристаллах поведение электрона совсем иное. Это обусловлено большим количеством содержащихся в них дефектов (включая и колебания решетки). За время, необходимое для того, чтобы значение k_x заметно возросло под действием поля, электрон успевает много раз столкнуться с различными дефектами. Среднее для всех электронов значение k может измениться существенным образом лишь в очень сильных электрических полях. Вместе с тем в каждом отдельном акте столкновения значение k отдельно взятого электрона может претерпевать очень сильные изменения. Количество энергии, отдаваемой электроном при соударении с тяжелым ионом, весьма мало. Можно показать, что энергия, затрачиваемая электроном на возбуждение колебаний решетки, также невелика (см. разд. 12.2). В этом случае соударения электрона можно считать упругими. При таких соударениях вектор k электрона меняется лишь по направлению, оставаясь на одной и той же поверхности равной энергии в k -пространстве. Поэтому движение электрона в кристалле весьма напоминает движение молекулы в газе. Двигаясь в отсутствие внешних полей совершенно хаотично, электрон обладает при этом определенной длиной свободного пробега между двумя последовательными столкновениями. Под действием внешнего поля на это хаотическое движение накладывается направленный дрейф, подобно ветру в газовой среде. Действительная длина пути, пройденного электроном ¹⁾, значительно больше пути дрейфа по

¹⁾ Имеется в виду длина пути пакета электронных волн, так как согласно квантовой механике электрону нельзя приписать вполне определенной траектории.— *Прим. ред.*

направлению поля и истинная скорость движения значительно больше скорости дрейфа. Это показано на рис. 5.1.

Для описания процессов соударений удобно ввести ряд новых величин. В квантовой теории столкновений обычно вводят эффективное сечение рассеяния, являющееся мерой вероятности столкновения того или иного рода. Например, можно ввести величину $\sigma(\theta, \varphi)d\omega$, определяющую вероятность соударения, в результате которого электрон с исходным направлением движения $\theta=0$ будет двигаться после столкновения по направлению (θ, φ) , заключенному в небольшом телесном угле $d\omega$, где θ, φ — сферические угловые координаты. Исходя из этой величины, можно определить интегральное сечение рассеяния σ_t по формуле

$$\sigma_t = \int \sigma(\theta, \varphi) d\omega. \quad (5.2)$$

Другими полезными величинами являются среднее время свободного пробега τ и средняя длина свободного пробега $l = v\tau$, где v — скорость электрона. Пусть n электронов движется со скоростью v в заданном направлении. Количество соударений, которые испытывают электроны за небольшой промежуток времени dt , должно быть пропорционально как n , так и dt . Поэтому количество электронов $n(t)$, которые к моменту времени t еще не испытали ни одного соударения, должно удовлетворять дифференциальному уравнению

$$\frac{dn}{dt} = -\frac{n}{\tau(v)}, \quad (5.3)$$

где $\tau(v)$ — некоторая величина, не зависящая от t и совпадающая, как можно будет убедиться ниже, со средним временем между соударениями. Из (5.3) следует

$$n = n_0 \exp\left(-\frac{t}{\tau}\right), \quad (5.4)$$

где $n = n_0$ при $t=0$. Вероятность того, что электрон к моменту времени t еще не испытал столкновения, определяется отношением n/n_0 и поэтому равно $\exp(-t/\tau)$. Среднее время \bar{t} между двумя соударениями равно

$$\bar{t} = \tau^{-1} \int_0^{\infty} t \exp\left[-\frac{t}{\tau}\right] dt = \tau. \quad (5.5)$$

Таким образом, величина τ действительно играет роль среднего времени между соударениями, или среднего времени свободного пробега. Отсюда следует, что длину свободного пробега l можно найти с помощью равенства $l(v) = v\tau(v)$. Легко показать также, что $1/l = N\sigma_t$, где N — число центров рассеяния электронов в единице объема, так что $1/\tau = N\sigma_t v$.

Простейшим из возможных вариантов рассеяния является случай изотропного рассеяния, характеризующийся тем, что после столкновения все направления движения электрона равновероятны. В этом случае скорость дрейфа, приобретенная электроном под действием поля на длине свободного пробега, полностью теряется в процессе столкновения. Как будет видно из дальнейшего (см. разд. 8.5), такое положение имеет место в случае рассеяния на тепловых колебаниях решетки. При других механизмах рассеяния, в частности для рассеяния кулоновским полем ионизованных примесных центров, преобладает рассеяние под малыми углами. В этом случае интегральное сечение рассеяния σ_t , определяемое равенством (5.2), удобнее заменить другим интегральным сечением рассеяния σ_c , которое получается в результате усреднения $\sigma(\theta, \varphi)$ с помощью надлежащей весовой функции, учитывающей изменение скорости дрейфа при соударении.

Рассмотрим электрон, движущийся до соударения параллельно оси z ($\theta=0$) со скоростью v_{z0} , а после соударения — в направлении (θ, φ) . Так как угол между начальным и конечным направлениями скорости равен θ , то изменение δv_z составляющей скорости электрона по оси z равно

$$\delta v_z = v_{z0} (1 - \cos \theta). \quad (5.6)$$

Если $\sigma(\theta, \varphi)$ зависит лишь от θ , то среднее изменение v_z в расчете на одно столкновение дается формулой

$$\overline{\delta v_z} = v_{z0} \frac{\int_0^\pi (1 - \cos \theta) \sigma(\theta) \sin \theta d\theta}{\int_0^\pi \sigma(\theta) \sin \theta d\theta}. \quad (5.7)$$

Определим сечение рассеяния σ_c с помощью равенства

$$\sigma_c(\theta) = 2\pi \int_0^\pi (1 - \cos \theta) \sigma(\theta) \sin \theta d\theta. \quad (5.8)$$

Тогда

$$\overline{\delta v_z} = v_{z0} \frac{\sigma_c}{\sigma_t}. \quad (5.9)$$

Отсюда следует, что первоначальная скорость дрейфа v_{z0} полностью теряется в среднем после σ_t/σ_c столкновений. Вероятность одного столкновения электрона за время dt равна $Nv\sigma dt$, а вероятность столкновения, приводящего к потере первоначальной скорости, равна $Nv\sigma_c dt$. Поэтому сечение рассеяния σ_c является более уместной величиной для описания явлений переноса (см. [1], § 10.3.1). В связи с этим необходимо ввести еще понятие об эффек-

тивном времени релаксации τ_c с помощью равенства

$$\frac{1}{\tau_c} = \frac{1}{\tau} \overline{(1 - \cos \theta)}, \quad (5.10)$$

где

$$\overline{(1 - \cos \theta)} = \frac{\sigma_e}{\sigma_i}.$$

Ограничимся пока рассмотрением только таких соударений, в результате которых полностью теряется приобретенная ранее скорость дрейфа. Используя введенные величины σ_e и σ_i , можно будет учесть также и другие возможные случаи рассеяния. Пусть электрическое поле \mathcal{E} направлено по оси x . Рассмотрим электрон, испытавший столкновение в момент времени $t=0$. Составляющая его скорости по оси x к моменту времени t при условии, что электрон не успел столкнуться вторично, равна

$$v_x = v_{x0} - \frac{\mathcal{E}et}{m_e}, \quad (5.11)$$

где v_{x0} — значение v_x при $t=0$; m_e — соответствующая эффективная масса. Выражение (5.11) необходимо усреднить по всевозможным значениям времени t с помощью выражения $\exp(-t/\tau)dt/\tau$, определяющего вероятность соударения в интервале времени от t до $t+dt$. Тогда

$$\bar{v}_x = \bar{v}_{x0} - \frac{\mathcal{E}e}{\tau m_e} \int_0^{\infty} t \exp\left(-\frac{t}{\tau}\right) dt = \bar{v}_{x0} - \frac{\mathcal{E}e\tau}{m_e}. \quad (5.12)$$

Из предположения, что в среднем после каждого соударения полностью теряется приобретенная ранее дрейфовая скорость, следует, что $\bar{v}_{x0} = 0$. Поэтому

$$\bar{v}_x = -\frac{e\mathcal{E}\tau}{m_e}. \quad (5.13)$$

Таким образом, в этом приближении средняя скорость дрейфа \bar{v}_x пропорциональна напряженности электрического поля \mathcal{E} . Равенство (5.13) можно записать в виде

$$\bar{v}_x = -\mu_e \mathcal{E}, \quad (5.14)$$

где μ_e (положительная величина) — подвижность электрона, т. е. величина скорости дрейфа электрона, обусловленная полем с напряженностью, равной единице. Таким образом, для μ_e получим соотношение

$$\mu_e = \frac{e\tau}{m_e}. \quad (5.15)$$

Аналогично в случае положительных дырок

$$\bar{v}_x' = \frac{e\tau\mathcal{E}}{m_h}, \quad (5.16)$$

причем

$$\mu_n = \frac{e\tau}{m_n}. \quad (5.16a)$$

Оценим значения некоторых из введенных величин. В чистом германии при комнатной температуре $\mu_e = 3900 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$. (В системе СИ: $0,39 \text{ м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$). Если принять $m_e = 0,3m_0$, то $e/m_e \approx 6 \cdot 10^{11} \text{ Кл} \cdot \text{кг}^{-1}$, и мы получаем $\tau = 6 \cdot 10^{-13} \text{ с}$. Среднюю квадратичную скорость электрона \bar{v}^2 можно получить из соотношения

$$\frac{1}{2} m_e \bar{v}^2 = \frac{3}{2} kT.$$

Поскольку средняя скорость \bar{v} лишь незначительно отличается от $[\bar{v}^2]^{1/2}$, то

$$\bar{v} \approx \left(\frac{3kT}{e} \right)^{1/2} \left(\frac{e}{m_e} \right)^{1/2}. \quad (5.17)$$

При $T = 300 \text{ К}$ $kT/e \approx 1/40 \text{ В}$, откуда $\bar{v} \approx 2,5 \cdot 10^5 \text{ м} \cdot \text{с}^{-1}$, так что $l = 1,5 \cdot 10^{-2} \text{ м}$, $= 1,5 \cdot 10^{-2} \text{ см}$. Это означает, что длина свободного пробега превышает сотни межатомных расстояний. В среднем за 1 с электрон пробегает путь порядка $\bar{v} \cdot 1 \text{ с}$, т. е. $2,5 \cdot 10^7 \text{ см}$, тогда как в поле напряженностью $1 \text{ В} \cdot \text{см}^{-1}$ он дрейфует на расстояние $\mu_e \cdot 1 \text{ В} / \text{см} \cdot 1 \text{ с} = 3,9 \cdot 10^3 \text{ см}$, что примерно на четыре порядка меньше пути, пройденного при тепловом движении.

5.2. Время релаксации, не зависящее от энергии

Значения τ для электронов и дырок, как правило, отличаются друг от друга. Кроме того, следует помнить, что в общем случае время релаксации τ является функцией скорости, так что его необходимо, вообще говоря, усреднить по всевозможным скоростям v . Для начала, однако, рассмотрим более простой случай, когда время не зависит от скорости, т. е. когда равенства (5.13) и (5.16) относятся сразу ко всем свободным электронам в зоне проводимости и соответственно ко всем дыркам в валентной зоне. Как мы уже выяснили в подразд. 1.3.1, все прочие электроны не принимают участия в процессе электропроводности.

5.2.1. ЭЛЕКТРОПРОВОДНОСТЬ

Обозначим через n и p концентрации электронов в зоне проводимости и дырок в валентной зоне соответственно. Плотность j_x электрического тока вдоль оси x дается формулой

$$j_x = -en\bar{v}_x + ep\bar{v}_x = \quad (5.18)$$

$$= e(n\mu_e + p\mu_n) \mathcal{E}_x. \quad (5.18a)$$

Электропроводность σ определяется равенством

$$\mathbf{j} = \sigma \mathcal{E}. \quad (5.19)$$

Таким образом,

$$\sigma = e(n\mu_e + p\mu_h). \quad (5.20)$$

Это весьма важное соотношение, связывающее электропроводность с концентрациями и подвижностями электронов и дырок, будет широко использовано впоследствии.

В полупроводниках со сферическими изоэнергетическими поверхностями и скалярной эффективной массой приведенные выше соотношения остаются справедливыми для любого направления тока и поля в кристалле, иначе говоря, σ является скалярной величиной. Предположим, однако, что мы имеем дело с полупроводником n -типа ($n \gg p$), у которого имеется один минимум энергии в зоне проводимости при $\mathbf{k}=0$, а поверхности равной энергии не обладают сферической симметрией. Для малых значений k энергию можно представить в виде

$$E = \frac{\hbar^2}{8\pi^2} \left[\frac{k_x^2}{m_1} + \frac{k_y^2}{m_2} + \frac{k_z^2}{m_3} \right] = \quad (5.21)$$

$$= \frac{1}{2} [m_1 v_x^2 + m_2 v_y^2 + m_3 v_z^2] \quad (5.21a)$$

[ср. с формулой (2.18)].

В системе координат, связанной с главными осями тензора обратной эффективной массы, уравнения движения для электронов имеют вид

$$\begin{aligned} m_1 \dot{v}_x &= -e\mathcal{E}_x, \\ m_2 \dot{v}_y &= -e\mathcal{E}_y, \\ m_3 \dot{v}_z &= -e\mathcal{E}_z, \end{aligned} \quad (5.22)$$

где \mathcal{E}_x , \mathcal{E}_y , \mathcal{E}_z — компоненты поля вдоль соответствующих осей координат. Путем обычного обобщения приведенного выше расчета на трехмерный случай, получим

$$\begin{aligned} j_x &= ne\mu_1 \mathcal{E}_x, \\ j_y &= ne\mu_2 \mathcal{E}_y, \\ j_z &= ne\mu_3 \mathcal{E}_z, \end{aligned} \quad (5.23)$$

где $\mu_1 = e\tau/m_1$, $\mu_2 = e\tau/m_2$, $\mu_3 = e\tau/m_3$. Таким образом, в данном случае электропроводность является тензором, причем благодаря соответствующему выбору системы координат этот тензор имеет диагональный вид. Если закон Ома выписать в виде тензорного равенства

$$j_r = \sum_s \sigma_{rs} \mathcal{E}_s, \quad (5.24)$$

где $r, s = x, y, z$, то мы получим

$$\begin{aligned} \sigma_{xx} &= ne\mu_1, \\ \sigma_{yy} &= ne\mu_2, \quad \sigma_{xy} = \sigma_{yz} = \sigma_{zx} = 0. \\ \sigma_{zz} &= ne\mu_3, \end{aligned} \quad (5.25)$$

Такой проводник электрического тока, очевидно, анизотропен, и, за исключением направлений, параллельных главным осям кристалла, вектор плотности тока \mathbf{j} уже не будет коллинеарен вектору приложенного внешнего электрического поля. Анизотропные кристаллы часто обладают симметрией вращения относительно одной из осей кристалла, которую можно выбрать в качестве главной координатной оси. Остальные две оси координат можно совместить с любыми двумя взаимно перпендикулярными осями, лежащими в плоскости, ортогональной к главной оси симметрии. Направим ось z по направлению главной оси кристалла. Тогда получаем $m_1 = m_2$ и, следовательно, $\mu_1 = \mu_2$, $\sigma_{xx} = \sigma_{yy} = \sigma_T$ и $\sigma_{zz} = \sigma_L$. Пусть вектор тока \mathbf{j} образует угол θ с осью z . Направим ось y так, чтобы она была перпендикулярна к плоскости, содержащей вектор \mathbf{j} и ось z . Компоненты вектора напряженности электрического поля по осям координат в этой системе соответственно будут $\mathcal{E}_x = j \sin \theta / \sigma_T$, $\mathcal{E}_y = 0$, $\mathcal{E}_z = j \cos \theta / \sigma_L$. Отсюда следует, что проекция поля \mathcal{E} на направление электрического тока \mathbf{j} равна $j \sin^2 \theta / \sigma_T + j \cos^2 \theta / \sigma_L$. Поэтому

$$\frac{1}{\sigma} = \frac{\sin^2 \theta}{\sigma_T} + \frac{\cos^2 \theta}{\sigma_L}. \quad (5.26)$$

Примером полупроводника с таким типом проводимости может служить теллур, у которого $\sigma_T = 1,95 \sigma_L$.

Если зона проводимости содержит несколько эквивалентных минимумов (как, например, в полупроводнике со структурой зоны, показанной на рис. 2.5, б или в), то необходимо проводить суммирование по электронным состояниям, относящимся ко всем минимумам. Если число таких симметрично расположенных минимумов равно M , то в единице объема кристалла каждый из них содержит n/M электронов. Тогда плотность тока определяется симметричным выражением вида

$$\left. \begin{aligned} j_x \\ j_y \\ j_z \end{aligned} \right\} = \frac{1}{3} ne (\mu_1 + \mu_2 + \mu_3) \left\{ \begin{aligned} \mathcal{E}_x, \\ \mathcal{E}_y, \\ \mathcal{E}_z, \end{aligned} \right. \quad (5.27)$$

так как из-за свойств симметрии кристалла в случае, когда электрическое поле направлено вдоль оси x , компоненты вектора электрического тока j_y и j_z обращаются в нуль. Электропроводность опять вырождается в скаляр, и мы получаем $\mathbf{j} = \sigma \mathcal{E}$, где

$$\sigma = \frac{1}{3} ne (\mu_1 + \mu_2 + \mu_3). \quad (5.28)$$

Если записать электропроводность σ в виде $\sigma = ne\mu_c$, то μ_c можно назвать «дрейфовой» подвижностью¹⁾ и определить соотношением

$$\mu_c = \frac{1}{3} (\mu_1 + \mu_2 + \mu_3). \quad (5.29)$$

С другой стороны, если μ_c записать в виде

$$\mu_c = \frac{e\tau}{m_c}, \quad (5.30)$$

то m_c можно назвать «дрейфовой» эффективной массой, так что

$$\frac{1}{m_c} = \frac{1}{3} \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right). \quad (5.31)$$

В кристаллах с кубической симметрией поверхности равной энергии вблизи минимумов имеют вид эллипсоидов вращения (см. разд. 2.3). Каждому такому эллипсоиду можно сопоставить диагональный тензор обратной эффективной массы, характеризуемый двумя совпадающими между собой компонентами $1/m_2$ и отличной от них третьей компонентой $1/m_1$. Величины m_1 и m_2 одни и те же для всех эллипсоидов, хотя и соответствуют в каждом из них различным направлениям в кристалле. Легко видеть, что при суммировании по всем эллипсоидам опять получается равенство (5.30), где

$$\frac{1}{m_c} = \frac{1}{3} \left(\frac{1}{m_1} + \frac{2}{m_2} \right). \quad (5.32)$$

«Дрейфовая» эффективная масса m_c несколько отличается от «эффективной массы плотности состояний» m_d , определяемой с помощью равенства (4.23).

Когда эллипсоиды равной энергии расположены не на осях $\langle 100 \rangle$ (которые можно принять в качестве осей координат), а на других осях симметрии, таких, как, например, $\langle 111 \rangle$, то для каждого из таких эллипсоидов получается тензор проводимости с отличными от нуля недиагональными компонентами. Однако при усреднении по всем эквивалентным минимумам недиагональные компоненты тензора проводимости исчезают и все диагональные компоненты снова совпадают между собой, так что тензор электропроводности по-прежнему вырождается в скаляр. В дальнейшем мы увидим, что это заключение справедливо лишь в случае кристаллов с кубической симметрией в отсутствие внешних магнитных полей. В присутствии магнитного поля сопротивление кристалла становится анизотропным.

¹⁾ То есть подвижностью, проявляющейся при движении (дрейфе) электронов в полупроводнике под действием электрического поля.— *Прим. ред.*

5.2.2. ЭФФЕКТ ХОЛЛА

Если проводник, по которому течет ток, поместить в однородное магнитное поле, составляющее прямой угол с направлением электрического тока, то в проводнике возникает э. д. с., направленная перпендикулярно плоскости, содержащей векторы электрического тока и напряженности магнитного поля. Этот эффект, названный эффектом Холла по имени Холла [2], открывшего его в 1879 г. в образцах тонкой металлической фольги, стал одним из самых действенных методов исследования свойств носителей заряда в полупроводниках. В то время как электропроводность при высоких температурах наряду с электронной компонентой может содержать также и известную долю ионной составляющей электрического тока, эффект Холла, обусловленный ионной проводимостью, настолько мал, что им можно полностью пренебречь.

Для простоты рассмотрим полупроводник в виде полоски неограниченной длины, по которой в направлении оси x течет электрический ток плотностью j и которая находится в магнитном поле, направленном по оси z . Для начала речь будет идти о полупроводнике n -типа со сферическими поверхностями равной энергии.

Рассмотрим влияние магнитного поля на дрейфовую скорость электронов в электрическом поле в плоскости (x, y) . Легко убедиться в том, что в стационарных условиях должна возникнуть составляющая электрического поля, перпендикулярная направлению тока, для того чтобы скомпенсировать поперечную силу, обусловленную магнитным полем. Дрейфовая скорость v_x равна $-j/ne$. Средняя поперечная сила, действующая на электрон в направлении оси y , равна $+eBv_x$, тогда компенсирующее электрическое поле \mathcal{E}_y определяется из условия

$$e\mathcal{E}_y = eBv_x = -\frac{Bj}{n}, \quad (5.33)$$

где B — магнитная индукция. Так как $\vec{v}_x = -\mathcal{E}_x \mu_e$, то составляющая электрического поля \mathcal{E}_x , параллельная току, определяется из равенства

$$j = ne\mu_e \mathcal{E}_x. \quad (5.34)$$

Поэтому угол θ между током и результирующим электрическим полем можно найти по формуле ¹⁾

$$\operatorname{tg} \theta = \frac{\mathcal{E}_y}{\mathcal{E}_x} = -B\mu_e. \quad (5.35)$$

Угол θ (рис. 5.2) принято называть углом Холла.

¹⁾ Необходимо заметить, что произведение $B\mu_e$ — безразмерная величина. На первый взгляд это не очевидно. В системе СИ μ_e имеет размерность $\text{м}^2 \cdot \text{В}^{-1} \text{с}^{-1}$, а индукция B — $\text{Вб} \cdot \text{м}^{-2}$ (Тесла) = $\text{В} \cdot \text{с} \cdot \text{м}^{-2}$, т. е. размерность, обратную размерности μ_e .

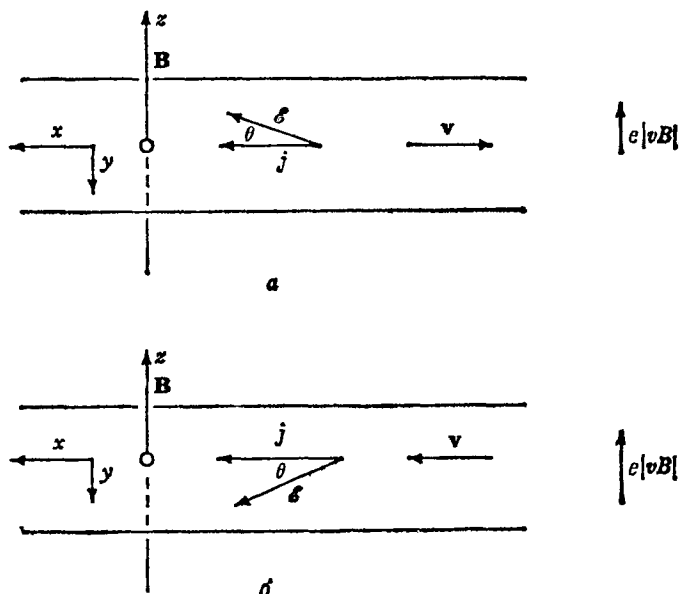


Рис. 5.2. Эффект Холла в слабом поле.
а — электроны, б — дырки.

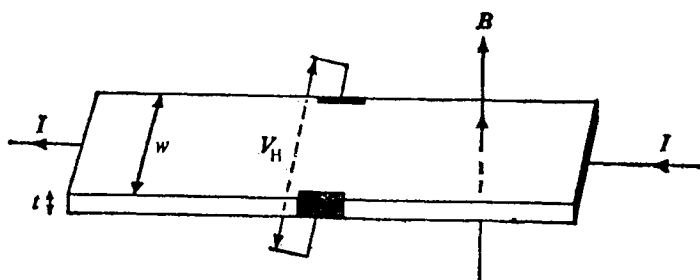


Рис. 5.3. Схема измерения эффекта Холла.

Для измерения напряжения, обусловленного эффектом Холла, обычно используют продолговатые плоские прямоугольные пластинки, толщина и ширина которых мала по сравнению с длиной (рис. 5.3). Направим ось x вдоль пластинки, а ось y — по направлению ее ширины. Наблюдение эффекта Холла сводится к измерению поперечной э. д. с., возникающей между узкими сторонами пластинки под действием магнитного поля. По сути это внешнее напряжение, необходимое для того, чтобы электрический ток был направлен точно по оси x . Эффект Холла описывается с помощью

коэффициента Холла R_H , определяемого равенством

$$\mathcal{E}_H = R_H j B, \quad (5.36)$$

где \mathcal{E}_H — напряженность поля Холла. Обозначим через t толщину, а через w ширину пластинки (рис. 5.3), тогда для напряжения Холла имеем

$$V_H = \frac{R_H I B}{t}, \quad (5.37)$$

или

$$R_H = \frac{t V_H}{I B}, \quad (5.38)$$

где I — полный электрический ток, протекающий по пластинке. Так как $j = ne\mu_e \mathcal{E}_x$ и $\mathcal{E}_H \equiv \mathcal{E}_y$, то из (5.35) и (5.36) имеем

$$\mathcal{E}_H = R_H n e \mu_e \mathcal{E}_x B = - B \mu_e \mathcal{E}_x,$$

так что

$$R_H = -1/ne. \quad (5.39)$$

Как будет видно из дальнейшего, формула (5.39) справедлива, строго говоря, лишь тогда, когда τ не зависит от скорости или энергии. В более общем случае необходимо еще включить сюда численный множитель r , который обычно изменяется в пределах от 1 до 2 в зависимости от преобладающего механизма рассеяния и, кроме того, зависит от степени вырождения электронов в зоне проводимости, т. е. следует записывать R_H в виде

$$R_H = -\frac{r}{ne}. \quad (5.40)$$

Таким образом, на основании измерения коэффициента Холла можно определить концентрацию электронов в зоне проводимости с неопределенностью, зависящей от множителя r , если только его значение неизвестно.

Аналогично, когда преобладает дырочная проводимость, имеем

$$\operatorname{tg} \theta = B \mu_h, \quad (5.35a)$$

$$R_H = \frac{1}{pe}, \quad (5.39a)$$

или точнее,

$$R_H = \frac{r'}{pe}, \quad (5.40a)$$

причем величина r' для дырок не обязательно совпадает с величиной r для электронов (хотя обе величины положительны).

Знак коэффициента Холла совпадает со знаком основных носителей заряда в полупроводниках. Коэффициент Холла R_H , как видно из (5.39), имеет размерность $L^3 Q^{-1}$, где Q — электрический заряд. В системе СИ это дает $\text{м}^3 \cdot \text{Кл}^{-1}$. Поскольку n чаще всего вы-

ражают в см^{-3} , то и для R_H обычно используют единицы $\text{см}^3 \cdot \text{Кл}^{-1}$. Естественно, что, используя формулу (5.37), необходимо позаботиться о том, чтобы все величины были выражены в одних и тех же единицах. Весьма удобно в качестве единицы длины использовать сантиметр, I и V выражать соответственно в амперах и вольтах, а в качестве единицы измерения B принять гаусс. Тогда, если мы хотим получить величину R_H в единицах $\text{см}^3 \cdot \text{Кл}^{-1}$, в правую часть равенства (5.38) необходимо ввести множитель 10^9 .

Так как электропроводность, обусловленная электронами или дырками, выражается одинаковыми формулами $\sigma = ne\mu_e$ или $\sigma = p\mu_h$, то вместо (5.39) и (5.39а) можно написать общее выражение

$$|R_H| \sigma = \mu. \quad (5.41)$$

Это равенство является точным, если $r=1$. В общем случае

$$\frac{|R_H| \sigma}{r} = \mu. \quad (5.42)$$

С помощью равенства

$$|R_H| \sigma = \mu_H \quad (5.43)$$

можно дать определение некоторой величины μ_H , имеющей размерность подвижности. Ее обычно называют холловской подвижностью. Подвижность, определяемую из электропроводности (дрейфовую подвижность), впредь будем обозначать через μ_c . При $\tau = \text{const}$ имеем $\mu_c = \mu_H$; в противном случае $\mu_H = r\mu_c$. Ниже будет проведено вычисление коэффициента r для некоторых частных случаев.

Произведем оценку величины э. д. с. Холла для типичного полупроводника. Пусть речь идет о пластинке германия n -типа шириной 1 см и толщиной 1 мм с концентрацией электронов $n = 5 \cdot 10^{14} \text{ см}^{-3}$. Из формулы (5.39) следует, что $-R_H \approx 10^4 \text{ см}^3 \cdot \text{Кл}^{-1}$. Для вычисления напряжения удобнее перейти к системе СИ. Тогда $R_H = 10^{-2} \text{ м}^3 \cdot \text{Кл}^{-1}$. Если $B = 10^3 \text{ Гс} = 10^{-1} \text{ Т}$, а ток равен 1 мА, то из уравнения (5.37) следует $V_H = 1 \text{ мВ}$.

На рис. 5.2 можно видеть, что магнитное поле заставляет электроны и дырки дрейфовать в одну и ту же сторону относительно направления тока, а именно в правую сторону по ходу тока, в то время как в электрическом поле их дрейф направлен в противоположные стороны. Если в образце имеются сравнимые концентрации электронов и дырок, то в отсутствие магнитного поля и при $j_x = 0$ дрейфовые скорости электронов и дырок должны быть равны нулю, т. е. не существует электронного и дырочного токов, которые могли бы скомпенсировать друг друга. Однако при наличии магнитного поля возможна компенсация электронного и дырочного токов так, что $j_y = 0$, но при этом скорости v_y как у электронов, так и у дырок по отдельности будут отличны от нуля. Это следует из того, что в магнитном поле электроны и дырки стремятся дрейфовать в одном и том же направлении, перпенди-

кулярном направлении тока. Следовательно, в условиях смешанной проводимости к вычислению величины коэффициента Холла следует подходить с большой осторожностью. Это наглядно поясняет существенное различие между металлами, которые обладают только одним типом носителей заряда, и полупроводниками.

При вычислении коэффициента Холла R_H для случая, когда преобладает один тип носителей заряда, электроны или дырки, нет необходимости заранее предполагать малость холловского угла θ . Однако, как мы увидим позже, для получения *простого* выражения для R_H , соответствующего уравнениям (5.39) и (5.39а), мы все-таки должны принять, что холловский угол θ мал одновременно для электронов и дырок; т. е. $V\mu_e \ll 1$ и $V\mu_h \ll 1$. Позже мы уберем это ограничение при вычислении R_H в условиях смешанной проводимости, однако поучительно показать, почему простой подход, пригодный для случая проводимости одного типа неприменим в случае, когда *и электроны, и дырки* присутствуют в сравнимых концентрациях.

Теперь при $j_y = 0$ мы не можем считать, что v_y для электронов или дырок равно нулю, так что нам придется обратиться к уравнениям движения. Для электронов

$$\begin{aligned} m_e \dot{v}_x &= -e\mathcal{E}_x - ev_y B, \\ m_e \dot{v}_y &= -e\mathcal{E}_y + ev_x B. \end{aligned} \quad (5.44)$$

Если $V\mu_e$ мало, то член Bev_y будет порядка $Be\mu_e\mathcal{E}_y = (V\mu_e)^2 e\mathcal{E}_x$, т. е. будет пренебрежимо мал. Тогда первое из уравнений (5.44) можно заменить более простым,

$$m_e \dot{v}_x = -e\mathcal{E}_x, \quad (5.45)$$

и, повторяя рассуждения, приведенные ранее, получить $\bar{v}_x = -\mu_e \mathcal{E}_x$. Подставляя этот результат в (5.44), получаем

$$m_e \dot{v}_y = -e\mathcal{E}_y - eV\mu_e \mathcal{E}_x. \quad (5.46)$$

Так как правая часть (5.46) теперь постоянна, то мы можем, как и ранее, записать

$$\bar{v}_{ye} = -\mu_e \mathcal{E}_y - \mu_e^2 B \mathcal{E}_x. \quad (5.47)$$

Аналогично для дырок

$$\bar{v}_{yh} = \mu_h \mathcal{E}_y - \mu_h^2 B \mathcal{E}_x. \quad (5.48)$$

Составляющая плотности тока j_y в направлении оси y теперь является суммой электронной и дырочной компонент,

$$\begin{aligned} j_y &= -ne\bar{v}_{ye} + p e\bar{v}_{yh} = \\ &= n_e \mu_e \mathcal{E}_y + ne\mu_e^2 B \mathcal{E}_x + p e\mu_h \mathcal{E}_y - p e\mu_h^2 B \mathcal{E}_x. \end{aligned} \quad (5.49)$$

Теперь уже мы можем положить $j_y=0$ и получить отношение полей $\mathcal{E}_y/\mathcal{E}_x$, т. е. записать

$$\operatorname{tg} \theta = \frac{\mathcal{E}_y}{\mathcal{E}_x} = \frac{\rho\mu_h^2 - n\mu_e^2}{n\mu_e + \rho\mu_h}. \quad (5.50)$$

В этом случае коэффициент Холла R_H дается соотношением $\mathcal{E}_y/jB = = \mathcal{E}_y/\mathcal{E}_x \sigma B$. Далее, используя выражение (5.20) для σ , получаем

$$R_H = \frac{\rho\mu_h^2 - n\mu_e^2}{e(n\mu_e + \rho\mu_h)^2}. \quad (5.51)$$

Равенства (5.50) и (5.51) переходят в (5.35) и (5.39) при $p=0$ и в (5.35a) и (5.39a) при $n=0$.

Особый интерес представляет собой значение коэффициента Холла R_H для собственного полупроводника. В этом случае $n=p=n_1$ и

$$R_{H1} = \frac{1}{en_1} \frac{\mu_h - \mu_e}{\mu_e + \mu_h}. \quad (5.52)$$

Если ввести отношение подвижностей $b = \mu_e/\mu_h$, то формулу (5.47) можно переписать в виде

$$R_H = \frac{1}{e} \frac{\rho - b^2 n}{(nb + \rho)^2}. \quad (5.53)$$

Таким образом, при $p = b^2 n$ имеем $R_H = 0$. Можно заметить, что при наличии смешанной проводимости произведение $R_H \sigma$ является функцией отношения p/n . Значение коэффициента Холла в собственном полупроводнике определяется формулой

$$R_{H1} = -\frac{1}{en_1} \frac{b-1}{b+1}. \quad (5.54)$$

При $b > 1$, как это часто и бывает, $R_{H1} < 0$, т. е. в данном случае преобладает электронная проводимость.

Изменение коэффициента Холла с температурой показано на рис. 5.4 для полупроводника n -типа при $b > 1$ и на рис. 5.5 для p -типа при $b > 1$. Следует обратить внимание на прохождение через нуль и смену знака коэффициента R_H в этом случае.

Поучительно вычислить средние скорости дрейфа \bar{v}_{ye} и \bar{v}_{yh} . Подставляя значение \mathcal{E}_y в (5.47) и (5.48), получаем

$$\bar{v}_{ye} = \frac{-\rho\mu_e\mu_h B \mathcal{E}_x (\mu_e + \mu_h)}{n\mu_e + \rho\mu_h}, \quad (5.55)$$

$$\bar{v}_{yh} = -n\mu_e\mu_h B \mathcal{E}_x \frac{\mu_e + \mu_h}{n\mu_e + \rho\mu_h}. \quad (5.56)$$

Опять мы видим, что и электроны, и дырки стремятся дрейфовать вправо по ходу тока, т. е. в отрицательном y -направлении. Мы видим также, что $\bar{v}_{ye} = 0$ при $j_y = 0$, если только $p = 0$, и $\bar{v}_{yh} = 0$

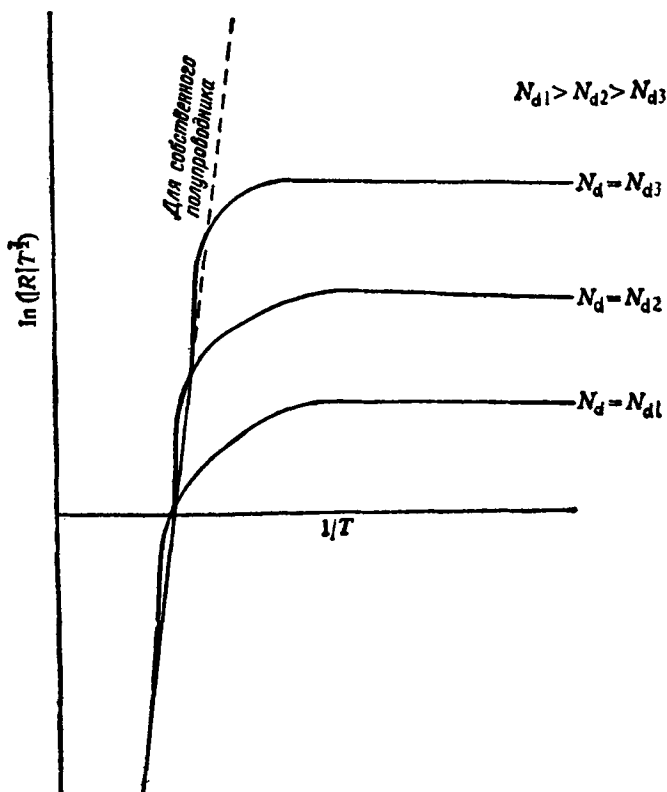


Рис. 5.4. Температурная зависимость коэффициента Холла для полупроводника n -типа.

при $j_y = 0$, если $n = 0$, как это предполагалось ранее для случая, когда преобладает один тип носителей заряда.

В выводе этих формул предполагалось, что холловский угол θ мал как для электронов, так и для дырок, иными словами, $B\mu_e$ и $B\mu_h$ одновременно малы. Это ограничение можно снять, решая соответствующие уравнения движения. Уравнения движения для электронов (5.44) можно записать в виде

$$\begin{aligned} \dot{v}_x &= -\frac{e}{m_e} \mathcal{E}_x - \omega v_y, \\ \dot{v}_y &= -\frac{e}{m_e} \mathcal{E}_y + \omega v_x, \end{aligned} \quad (5.57)$$

где $\omega = eB/m_e$, $\omega/2\pi$ — частота циклотронного резонанса, введенная в разд. 2.5. Это хорошо известные уравнения, описывающие движение электрона в скрещенных электрическом и магнитном

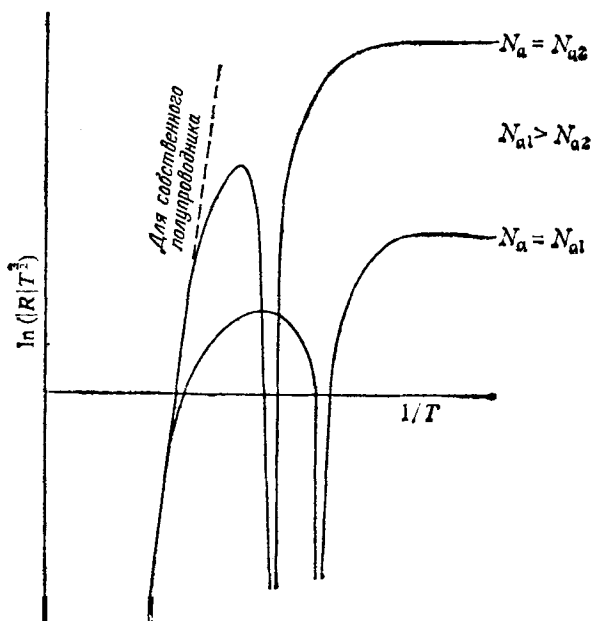


Рис. 5.5. Температурная зависимость коэффициента Холла для полупроводника p -типа.

полях. Используя комплексные переменные $Z = v_x + iv_y$ и $z = x + iy$, систему уравнений (5.57) можно записать в виде

$$\dot{Z} - i\omega Z = -\frac{e}{m_e} (\mathcal{E}_x + i\mathcal{E}_y). \quad (5.58)$$

Общее решение этого уравнения имеет вид

$$Z = Z_0 e^{i\omega t} + \frac{e}{m_e} (\mathcal{E}_x + i\mathcal{E}_y) \frac{1 - e^{-i\omega t}}{i\omega}, \quad (5.59)$$

где $Z_0 = v_{x0} + iv_{y0}$. При $t=0$ оно сводится, как это и должно быть, к значению $Z = Z_0$. Необходимо усреднить теперь Z по временам свободного пробега, приняв по-прежнему $\bar{Z}_0 = 0$. Тогда получаем

$$\bar{Z} = \bar{v}_x + i\bar{v}_y = \frac{1}{\tau} \int_0^{\infty} Z e^{-t/\tau} dt = -\frac{e\tau}{m_e} \frac{\mathcal{E}_x + i\mathcal{E}_y}{1 - i\omega\tau}. \quad (5.60)$$

Разделяя (5.54) на действительную и мнимую части и полагая $j_x = -en\bar{v}_x$ и $j_y = -en\bar{v}_y$, получаем для составляющих плотности

тока j_x и j_y :

$$j_x = \frac{ne^2}{m_e} \left[\frac{\tau \mathcal{E}_x}{1 + \omega^2 \tau^2} - \frac{\omega \tau^2 \mathcal{E}_y}{1 + \omega^2 \tau^2} \right], \quad (5.61)$$

$$j_y = \frac{ne^2}{m_e} \left[\frac{\tau \mathcal{E}_y}{1 + \omega^2 \tau^2} + \frac{\omega \tau^2 \mathcal{E}_x}{1 + \omega^2 \tau^2} \right]. \quad (5.62)$$

Условие $j_y = 0$ сводится к требованию, чтобы

$$\frac{\mathcal{E}_y}{\mathcal{E}_x} = -\omega\tau, \quad (5.63)$$

поэтому угол θ между током j_x и результирующим полем \mathcal{E} определяется выражением

$$\operatorname{tg} \theta = -\omega\tau = -\frac{Be\tau}{m_e} = -B\mu_e. \quad (5.64)$$

Это не что иное, как формула (5.35), что является подтверждением упрощенного вывода, с помощью которого ранее была получена формула (5.35). Подставляя \mathcal{E}_y из (5.63) в формулу (5.61), получаем

$$j_x = \frac{ne^2\tau \mathcal{E}_x}{m_e} = ne\mu_e \mathcal{E}_x = \sigma \mathcal{E}_x, \quad (5.65)$$

где $\mu_e = e\tau/m_e$. Этот результат в точности совпадает с результатом, полученным ранее. Заметим, что в указанных выше условиях, т. е. при постоянном τ и сферической симметрии изоэнергетических поверхностей, изменение сопротивления в поперечном магнитном поле равно нулю. Кроме того,

$$R_H = -\frac{1}{ne}. \quad (5.66)$$

Отсутствие зависимости коэффициента Холла от напряженности магнитного поля нарушается при наличии двух типов носителей заряда. Систему равенств (5.61) и (5.62) можно применить и к дыркам, предварительно изменив знак ω , тогда получаем $R_H = 1/pe$. При наличии дырок и электронов необходимо в выражениях для j_x и j_y учесть вклады каждого из типов носителей заряда. Кроме того, допустим, что τ_e и τ_h , т. е. времена свободного пробега для электронов и дырок различны. Систему равенств (5.61) и (5.62) можно записать в виде

$$\begin{aligned} j_x &= A\mathcal{E}_x - D\mathcal{E}_y, \\ j_y &= A\mathcal{E}_y + D\mathcal{E}_x, \end{aligned} \quad (5.67)$$

так что, полагая $j_y = 0$ и производя вычисления по той же схеме, что и ранее, получаем

$$\begin{aligned} -BR_H &= D \frac{1}{A^2 + D^2}, \\ \sigma &= \frac{A^2 + D^2}{A}. \end{aligned} \quad (5.68)$$

При наличии двух видов носителей заряда эта система может быть записана в виде

$$\begin{aligned} j_x &= (A_1 + A_2) \mathcal{E}_x - (D_1 + D_2) \mathcal{E}_y, \\ j_y &= (A_1 + A_2) \mathcal{E}_y + (D_1 + D_2) \mathcal{E}_x. \end{aligned} \quad (5.69)$$

При $j_y = 0$ имеем

$$\mathcal{E}_x = - (A_1 + A_2) \frac{\mathcal{E}_y}{D_1 + D_2}. \quad (5.70)$$

Подставляя (5.70) в первое равенство системы (5.69), получаем

$$BR_H = - \frac{D_1 + D_2}{(A_1 + A_2)^2 + (D_1 + D_2)^2}. \quad (5.71)$$

Коэффициенты A_1 , A_2 , D_1 и D_2 можно выразить через электропроводности σ_1 , σ_2 , и коэффициенты Холла R_{H1} , R_{H2} , относящиеся к случаю одного типа носителей заряда, — с помощью формул

$$\begin{aligned} A_1 &= \frac{\sigma_1}{1 + \sigma_1^2 B^2 R_{H1}^2}, \\ D_1 &= - \frac{\sigma_1^2 B R_{H1}}{1 + \sigma_1^2 B^2 R_{H1}^2} \text{ и т. д.} \end{aligned}$$

После несложных алгебраических преобразований формула (5.71) принимает вид

$$R_H = \frac{R_{H1} \sigma_1^2 (1 + \sigma_2^2 B^2 R_{H2}^2) + R_{H2} \sigma_2^2 (1 + \sigma_1^2 B^2 R_{H1}^2)}{(\sigma_1 + \sigma_2)^2 + \sigma_1^2 \sigma_2^2 B^2 (R_{H1} + R_{H2})^2}. \quad (5.72)$$

Формула (5.72) получена Чемберсом [3]¹⁾. Она справедлива и в более общем случае, когда τ зависит от v , и применима для любой совокупности двух типов носителей заряда (не обязательно электронов и дырок). Воспользовавшись равенством (5.65) и выражением (5.66), а также аналогичными формулами для дырок, можно записать

$$R_H = \frac{(p - nb^2) + b^2 \mu_h^2 B^2 (p - n)}{(bn + p)^2 + b^2 \mu_h^2 B^2 (p - n)^2} \frac{1}{e}. \quad (5.73)$$

Эта формула справедлива, строго говоря, лишь для случая $\tau = \text{const}$ (см. подразд. 5.3.3). Коэффициент Холла обращается в нуль при

$$p = \frac{nb^2 (1 + \mu_h^2 B^2)}{1 + b^2 \mu_h^2 B^2}. \quad (5.74)$$

Таким образом, значение отношения p/n , при котором $R_H = 0$, зависит от магнитного поля. Эта зависимость исчезает лишь при

¹⁾ Раньше Чемберса эффект Холла для системы двух типов носителей заряда подробно рассмотрели в общем виде Б. И. Давыдов и И. М. Шмушкевич, — УФН, 1940 г., т. 24, с. 21, — *Прим. ред.*

$\mu_h B \ll 1$. Для германия $b\mu_h = 0,39 \text{ м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$. При $B = 10^3 \text{ Гс} = 0,1 \text{ Т}$ значение $B\mu_h = 0,039$. Поэтому в случае германия изменение R_H с полем до 10^3 Гс мало. Однако уже при $B = 10^4 \text{ Гс}$ R_H заметно отличается от значения коэффициента Холла в слабом поле. Интересно отметить, что при больших B (и $n \neq p$)

$$R_H \rightarrow \frac{1}{e(p-n)}. \quad (5.75)$$

В проведенных выше вычислениях всюду предполагалось, что поперечная составляющая электрического тока равна нулю. Практически, однако, дырки и электроны могут рекомбинировать на поверхности образца, так что в стационарном режиме возможны условия, когда $j_y \neq 0$: Этот эффект исследовали Бэнбери и др. [5], а также Ландауер и Свансон [6]. Их исследования показали, что при соблюдении известных мер предосторожности поверхностная рекомбинация не сказывается на результатах измерения коэффициента Холла.

5.2.3. ПОПЕРЕЧНОЕ МАГНИТОСОПРОТИВЛЕНИЕ

Как уже отмечалось выше, если полупроводник со сферическими поверхностями постоянной энергии содержит лишь один тип носителей заряда и если $\tau = \text{const}$, то поперечное магнитосопротивление у такого полупроводника в слабом магнитном поле не возникает. Однако при наличии более одного типа носителей это утверждение перестает быть справедливым. Из равенств (5.69) и (5.70) следует, что

$$\sigma = \frac{j_x}{\mathcal{E}_x} = (A_1 + A_2) \frac{(D_1 + D_2)^2}{A_1 + A_2}. \quad (5.76)$$

Тогда из (5.71) и (5.76) имеем

$$R_H \sigma = - \frac{D_1 + D_2}{B(A_1 + A_2)} = \frac{R_{H1}\sigma_1^2(1 + \sigma_2^2 B^2 R_{H2}^2) + R_{H2}\sigma_2^2(1 + \sigma_1^2 B^2 R_{H1}^2)}{(\sigma_1 + \sigma_2) + B^2 \sigma_1 \sigma_2 (\sigma_1 R_{H1}^2 + \sigma_2 R_{H2}^2)} \quad (5.77)$$

и

$$\sigma = \frac{(\sigma_1 + \sigma_2)^2 + \sigma_1^2 \sigma_2^2 B^2 (R_{H1} + R_{H2})^2}{(\sigma_1 + \sigma_2) + B^2 \sigma_1 \sigma_2 (\sigma_1 R_{H1}^2 + \sigma_2 R_{H2}^2)} \quad (5.78)$$

С точностью до членов первого порядка по B^2 и учитывая, что σ_1 и σ_2 при постоянном τ не зависят от B , получаем

$$- \frac{\Delta \sigma}{\sigma_0 B^2} = \frac{\Delta \rho}{\rho_0 B^2} = \frac{-\sigma_1^2 \sigma_2^2 (R_{H1} + R_{H2})^2}{\sigma_0^2} + \frac{\sigma_1 \sigma_2 (\sigma_1 R_{H1}^2 + \sigma_2 R_{H2}^2)}{\sigma_0}. \quad (5.79)$$

Здесь $\Delta \sigma$ — изменение электропроводности в слабом магнитном поле, $\Delta \sigma = \sigma - \sigma_0$, где $\sigma_0 = \sigma_1 + \sigma_2$ — электропроводность при $B = 0$;

аналогично $\Delta\rho$ означает изменение удельного сопротивления в слабом магнитном поле, $\Delta\rho = \rho - \rho_0$, где ρ_0 — значение сопротивления при $B=0$. Подставляя вместо σ_1 , R_{H1} и σ_2 , R_{H2} соответствующие выражения для электронов и дырок, получаем

$$-\frac{\Delta\sigma}{\sigma_0 B^2} = \frac{\Delta\rho}{\rho_0 B^2} = \frac{n p \mu_e \mu_h (\mu_e + \mu_h)^2}{(n \mu_e + p \mu_h)^2} \quad (5.80)$$

или

$$-\frac{\Delta\sigma}{\sigma_0 B^2} = \frac{n p b \mu_h^2 (1+b)^2}{(n b + p)^2} \cdot \quad (5.81)$$

Используя равенство (5.53) для коэффициента Холла R_{H0} в слабом поле, можно переписать формулу (5.81) в виде

$$-\frac{\Delta\sigma}{\sigma_0} = \frac{\Delta\rho}{\rho_0} = \xi R_{H0}^2 \sigma_0^2 B^2, \quad (5.82)$$

где

$$\xi = \frac{n p b (1+b)^2}{(p - n b^2)^2} \cdot \quad (5.83)$$

Величину ξ обычно называют коэффициентом поперечного магнитосопротивления. При $n=0$ или $p=0$ этот коэффициент обращается в нуль. Отсюда видно, что даже те полупроводники, которые при $n=0$ или $p=0$ не обнаруживают поперечного магнитосопротивления, могут обладать этим эффектом в области смешанной проводимости. Однако поскольку j_z не зависит от B даже в области смешанной проводимости, то эффект продольного магнитосопротивления у таких полупроводников отсутствует.

5.3. Время релаксации, зависящее от энергии

Перейдем теперь к исследованию случая, когда время релаксации τ зависит от скорости v . Для начала рассмотрим несобственный полупроводник со сферическими изоэнергетическими поверхностями. Существуют два теоретических метода изучения кинетических явлений. Более фундаментальный метод — использование кинетического уравнения Больцмана, с помощью которого определяется отклонение функции распределения от ее равновесного значения (распределение Ферми — Дирака), вызванное внешними полями. Этот метод подробно излагается в известных руководствах по теории твердого тела (см., например, [7], гл. 8; [8]; [1], § 10.5, 10.6). С другой стороны, для этой цели можно использовать решения уравнений движения, т. е. равенства (5.61) и (5.62), и усреднить

их с помощью надлежащей весовой функции¹⁾. В качестве весовой функции для усреднения равенств (5.61) и (5.62) на первый взгляд кажется вполне естественным использовать обычную функцию распределения

$$N(E)f_0\left(\frac{E-E_F}{kT}\right) \equiv N(E)f_0(x)$$

определяющую число занятых состояний, приходящихся на единичный интервал энергии. Здесь $f_0(x)$ — функция Ферми, определенная в разд. 4.1. Это, однако, неверно. Такой способ усреднения не исключает переходов электронов в *занятые* состояния, которые на самом деле запрещены принципом Паули. Кроме того, даже в отсутствие вырождения электронного газа такой способ усреднения неверен, так как приписывает слишком малый вес наиболее энергичным электронам. Оказывается, что в отсутствие вырождения правильная весовая функция должна иметь вид $EN(E)f_0[(E-E_F)/kT]$, где $f_0(x)$ можно заменить показательной функцией.

5.3.1. ЭЛЕКТРОПРОВОДНОСТЬ

Для начала предположим, что имеется $\Delta n(E)$ электронов, энергия которых лежит в интервале E до $E+dE$. Вклад этих электронов в плотность тока j дается формулой

$$\Delta j = \frac{\Delta n(E) e^2 \tau(E) \mathcal{E}}{m_e}. \quad (5.84)$$

Предположим, что τ зависит только от величины скорости v и, следовательно, от энергии. Для того чтобы получить полную плотность тока, необходимо произвести суммирование по всем таким группам электронов, однако с учетом весового множителя $W(E)$. Таким путем получим

$$j = ne^2 \frac{\langle \tau \rangle \mathcal{E}}{m_e}. \quad (5.85)$$

В этом выражении для невырожденного полупроводника n -типа мы заменили $\sum \Delta n(E) \tau(E) W(E)$ на $n \langle \tau \rangle$ (см. [1], § 10.6), где

$$\langle \tau \rangle = \frac{\int_0^{\infty} EN(E) \exp(-E/kT) \tau(E) dE}{\int_0^{\infty} EN(E) \exp(-E/kT) dE}. \quad (5.86)$$

¹⁾ Такой метод используется в статье Брукса [9] (см. также [1], § 10.6).

В случае «сферических» зон энергии $N(E) \sim E^{1/2}$ (см. разд. 4.2)

$$\langle \tau \rangle = \frac{\int_0^{\infty} E^{3/2} \exp(-E/kT) \tau(E) dE}{\int_0^{\infty} E^{3/2} \exp(-E/kT) dE}. \quad (5.87)$$

Следовательно, мы можем использовать приведенное выше рассмотрение проводимости, лишь заменив постоянное время релаксации τ его усредненной величиной $\langle \tau \rangle$.

Если имеет место слабое вырождение, то в качестве весового множителя следует использовать значительно более сложное выражение (см. [1], § 10.6)

$$W(E) = E f_0 (1 - f_0). \quad (5.88)$$

Опустив несущественный множитель, это выражение можно представить в виде

$$W(E) = E \frac{\partial f_0}{\partial E}, \quad (5.89)$$

тогда для «сферических» зон

$$\langle \tau \rangle = \frac{\int_0^{\infty} E^{3/2} (\partial f_0 / \partial E) \tau(E) dE}{\int_0^{\infty} E^{3/2} (\partial f_0 / \partial E) dE}. \quad (5.90)$$

В условиях сильного вырождения весьма приближенно можно принять $f_0 = 1$ при $E < E_F$ и $f_0 = 0$ при $E \geq E_F$, так что интегралы, содержащие $\partial f_0 / \partial E$, дадут вклад только вблизи $E = E_F$, следовательно, остальные сомножители подынтегральной функции можно вынести за знак интеграла. Таким образом,

$$\int_0^{\infty} \tau(E) E^{3/2} \frac{\partial f_0}{\partial E} dE \approx -\tau(E_F) E_F^{3/2},$$

$$\int_0^{\infty} E^{3/2} \frac{\partial f_0}{\partial E} dE \approx \int_0^{E_F} E^{1/2} dE = -E_F^{3/2},$$

откуда приближенно

$$j_x \approx e^2 \mathcal{E} n \frac{\tau(E_F)}{m_e}. \quad (5.91)$$

Из этой формулы видно, что как будто бы все электроны в зоне участвуют в проводимости, однако с одним и тем же значением

времени релаксации, соответствующим энергии электрона вблизи уровня Ферми.

Если изоэнергетические поверхности не обладают сферической симметрией, то весовую функцию, используемую для усреднения, следует видоизменить. Рассмотрим случай, когда вырождение отсутствует, тогда функцию f_0 можно заменить экспонентой. Предположим, что изоэнергетические поверхности имеют форму эллипсоидов (так же как и в разд. 2.1). Пусть электрическое поле имеет отличные от нуля компоненты вдоль осей x , y , z . Можно легко показать (см. [1], § 10.6), что

$$\begin{aligned}\sigma_{xx} &= n \frac{e^2}{m_1} \langle \tau \rangle, \\ \sigma_{yy} &= n \frac{e^2}{m_2} \langle \tau \rangle, \\ \sigma_{zz} &= n \frac{e^2}{m_3} \langle \tau \rangle,\end{aligned}\tag{5.92}$$

где $\langle \tau \rangle$ та же самая усредненная величина, которая определяется выражением (5.87). Отсюда видно, что весовую функцию $E f_0$ можно использовать и для несферического случая.

Соотношения (5.92) совпадают с соотношениями (5.25), полученными для постоянного τ , если в них τ заменить на $\langle \tau \rangle$. Соответствующие подвижности даются формулами

$$\begin{aligned}\mu_{c1} &= \frac{e}{m_1} \langle \tau \rangle, \\ \mu_{c2} &= \frac{e}{m_2} \langle \tau \rangle, \\ \mu_{c3} &= \frac{e}{m_3} \langle \tau \rangle.\end{aligned}\tag{5.93}$$

Формула (5.31) для дрейфовой эффективной массы в кристаллах кубической симметрии также сохраняется, как это видно из вывода, приведенного в подразд. 5.2.1.

5.3.2. ЗАВИСИМОСТЬ τ ОТ ЭНЕРГИИ E

Можно показать, что в полупроводнике со сферическими поверхностями равной энергии длина свободного пробега l при рассеянии электрона на колебаниях решетки не зависит от его скорости (см. разд. 8.5). Поэтому $\tau_i \sim v^{-1}$ или $\tau_i \sim E^{-1/2}$. Можно показать также, что при рассеянии на ионизованных примесных центрах $\tau_1 \sim E^{1/2}$ (см. разд. 8.9). Таким образом, для двух указанных механизмов рассеяния получаются совершенно различные усреднения. В случае примесного рассеяния преобладает влияние весьма медленных электронов или дырок, тогда как при решеточном

рассеянии наибольший вклад вносят наиболее быстрые электроны или дырки. При одновременном действии обоих типов рассеяния выражение для общего времени релаксации получается из суммы обратных времен релаксации для отдельных механизмов¹⁾, т. е. из соотношения

$$\frac{1}{\tau} = \frac{1}{\tau_l} + \frac{1}{\tau_i}, \quad (5.94)$$

так как полное сечение рассеяния определяется суммой парциальных сечений (см. разд. 5.1). Оказалось, что для многих полупроводников в достаточно широкой области изменения энергии можно положить $\tau = aE^{-s}$, где s — некоторая постоянная, а величина a может зависеть от температуры. Значение s получают в общем случае из экспериментальной зависимости подвижности от температуры. В отсутствие вырождения при $\tau = aE^{-s}$ имеем

$$\langle \tau \rangle = a \frac{\int_0^{\infty} E^{3/2-s} \exp(-E/kT) dE}{\int_0^{\infty} E^{3/2} \exp(-E/kT) dE}. \quad (5.95)$$

Если s — дробная величина, то интеграл в числителе равенства (5.95) может быть выражен через Γ -функцию, определяемую соотношением

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Γ -функция обладает очевидным свойством: $\Gamma(x+1) = x\Gamma(x)$. Если $x = n$ — целое число, то $\Gamma(n+1) = n!$. При $x = 1/2$

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} t^{-1/2} e^{-t} dt = \int_0^{\infty} e^{-y^2} dy = \pi^{1/2}.$$

Итак, полагая $t \equiv E/kT$, имеем

$$\langle \tau \rangle = a (kT)^{-s} \frac{\Gamma\left(\frac{5}{2} - s\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (5.96)$$

Для рассеяния на колебаниях решетки $s = 1/2$. В этом случае длина свободного пробега l не зависит от скорости и может быть записана

¹⁾ Это не совсем правильно, так как τ_l и τ_i по-разному зависят от энергии. Для различных величин примесного рассеяния в предположении $\tau_l \sim E^{-1/2}$ усреднение было проведено Джонсом [10], Джонсом и Ларк-Горовицем [11]. Проблему совместного действия двух типов рассеяния в дальнейшем обсуждала Конуэлл [12].

в виде $l = \tau v = A/T$, где A — некоторая постоянная (см. формулу (8.24)). Это дает

$$\langle \tau \rangle = \frac{4m_e^{1/2} l}{3(2\pi kT)^{1/2}} \quad (5.97)$$

и

$$\mu_i = \frac{e \langle \tau \rangle}{m_e} = \frac{4el}{3(2\pi m_e kT)^{1/2}}. \quad (5.98)$$

Поскольку $l = A/T$, можно записать μ_i в виде

$$\mu_i = BT^{-3/2}, \quad (5.99)$$

где B — константа.

Для рассеяния на ионизованных примесях получаем (см. разд. 8.9)

$$\langle \tau \rangle = 8a \frac{(kT)^{3/2}}{\pi^{1/2}}. \quad (5.100)$$

В этом случае приближенно

$$a = \frac{C}{\ln(Td)},$$

где C и d — некоторые постоянные (см. разд. 8.9). Таким образом, при рассеянии на ионах примеси

$$\mu_i = \frac{DT^{3/2}}{\ln(Td)}, \quad (5.101)$$

где D и d — постоянные числа.

При наличии двух механизмов рассеяния можно определить полную подвижность из соотношения

$$\frac{1}{\mu_c} = \frac{1}{\mu_i} + \frac{1}{\mu_i}, \quad (5.102)$$

т. е. складывая обратные величины подвижностей так, как если бы речь шла о соответствующих временах релаксации¹⁾. Из (5.102) следует, что при высоких температурах преобладает решеточное рассеяние, однако при более низкой температуре, значение которой зависит от концентрации ионизованных примесных центров, начинает преобладать примесное рассеяние. Это проявляется в том, что температурная зависимость подвижности перестает укладываться в рамки простого степенного закона, который обычно наблюдается при более высоких температурах. Однако было показано, что формула (5.99), вообще говоря, несправедлива даже и при высоких температурах. Причины этого сводятся в основном к тому, что для применимости этой формулы необходимо, чтобы по-

¹⁾ См., однако, примечание к (5.94).

верхности равной энергии были сферическими. Мы вернемся к обсуждению этого вопроса в разд. 8.7.

Для чисто решеточного рассеяния при сильном вырождении имеем

$$\langle \tau \rangle = \tau(E_F) \sim l E_F^{-1/2}, \quad (5.103)$$

так что величина $\langle \tau \rangle / l$ очень слабо зависит от температуры и может считаться постоянной. Поскольку множитель l обратно пропорционален температуре, то

$$\mu_c = \mu_0 \frac{T_0}{T}. \quad (5.104)$$

Итак, в сильно примесном полупроводнике с большой и почти постоянной концентрацией электронов n

$$\sigma = n e \mu_0 \frac{T_0}{T} = \sigma_0 \frac{T_0}{T}. \quad (5.105)$$

Таким образом, удельное сопротивление оказывается прямо пропорциональным температуре T . Такой полупроводник ведет себя подобно металлу, поскольку у металлов в широкой температурной области, за исключением очень низких температур, обычно наблюдается такая же зависимость удельного сопротивления от температуры. Это указывает на ограниченность первоначального определения класса полупроводников, согласно которому электропроводность полупроводников должна возрастать при повышении температуры. В действительности такая зависимость электропроводности от температуры имеет место лишь в области почти собственной проводимости, когда увеличение концентрации свободных носителей (n или p) с температурой происходит значительно сильнее, чем уменьшение их подвижности. На основании формулы (4.18) можно прийти к заключению, что в случае решеточного рассеяния, когда $\mu \sim T^{-3/2}$, электропроводность собственного полупроводника подчиняется зависимости

$$\sigma = \sigma_0 \exp \left[-\frac{\Delta E}{2kT} \right], \quad (5.106)$$

где σ_0 не зависит от температуры.

Оказалось, однако, что даже в достаточно чистых полупроводниках, когда примесное рассеяние заведомо не играет никакой роли (например, в очень чистых образцах Ge или Si), подвижность электронов и дырок не подчиняется закону $T^{-3/2}$. В настоящее время полагают, что такое отклонение температурной зависимости подвижности электронов в Ge и Si от закона $T^{-3/2}$ обусловлено наличием нескольких минимумов или «долин» в зоне проводимости, вследствие чего возможно рассеяние электронов, которое сопровождается перебросами между состояниями, принадлежащими различным «долинам». Этот механизм рассеяния действительно

несколько видоизменяет закон $T^{-3/2}$ для подвижности. Простая степенная зависимость τ от энергии, по-видимому, не выполняется в широкой области изменения энергии, однако можно подобрать некоторую константу s так, чтобы закон $\tau = \tau_0 E^{-s}$ был справедлив в достаточно широком интервале изменения энергии E . Это позволяет вычислить различные величины, такие, как электропроводность, постоянную Холла и т. д. Обсуждение вопроса о «междолинном» механизме рассеяния электронов мы отложим до разд. 8.7.

Если поверхности равной энергии не обладают сферической симметрией, то для определения тензора электропроводности можно воспользоваться формулой (5.94). Как было показано в подразд. 5.3.1, среднее значение τ определяется при этом так же, как и раньше. После замены τ его средним значением можно воспользоваться всеми результатами, полученными в подразд. 5.2.1. В этом подразделе предполагалось, что τ зависит лишь от величины скорости. Как показали Херринг и Фохт [13], в случае зависимости τ от направления в кристалле теорию можно обобщить, считая τ тензорной величиной. В случае стационарных явлений переноса учет анизотропии τ сводится к усреднению каждой компоненты тензора обратной эффективной массы вместе с соответствующими компонентами тензора времени релаксации, т. е. к усреднению величин m_r/τ_r , которые входят в выражения для различных кинетических явлений.

При изучении электропроводности мы считали, что электрическое поле достаточно мало и не изменяет распределения электронов и дырок по энергиям. Позже (в разд. 12.1) мы рассмотрим, что происходит в случае более сильных полей, т. е. в ситуации, когда возникают «горячие электроны» — электроны с энергией, значительно превышающей термодинамически равновесную величину. Кроме того, мы пренебрегали квантовыми эффектами, которые возникают в очень сильных магнитных полях и которые также заметно изменяют равновесное распределение (см. подразд. 12.5.1 и разд. 12.8). Рассмотрение этих двух вопросов мы пока отложим.

5.3.3. ЭФФЕКТ ХОЛЛА В ПОЛУПРОВОДНИКАХ СО СФЕРИЧЕСКИМИ ПОВЕРХНОСТЯМИ ПОСТОЯННОЙ ЭНЕРГИИ

Исследование влияния магнитного поля на движение электронов и дырок в полупроводниках можно проводить с помощью двух различных методов. Первый из них сводится к тому, чтобы включить магнитное поле в выражение для внешних сил в уравнении Больцмана, и затем решать это уравнение с необходимой степенью точности. Такой путь, который представляется наиболее логичным и строгим, широко используется во многих современных исследо-

ваниях кинетических явлений как в металлах, так и полупроводниках (см., например, [1], § 10.8). Во втором методе исходят из решений уравнений движения для электрона, которые усредняют сначала по всевозможным временам пробега между столкновениями, а затем по энергиям, рассматривая среднее время пробега τ как функцию v или E , так же, как в разд. 5.3.

Вернемся к равенствам (5.61) и (5.62), определяющим значения плотностей токов j_x , j_y при наличии постоянного магнитного поля с индукцией B , направленного по оси z . Предположим для начала, что частота $\omega = eB/m_e$ настолько мала, что $\omega\tau \ll 1$ для всех существенных значений τ . Тогда в знаменателях равенств (5.61), (5.62) можно пренебречь членами с $\omega^2\tau^2$. Напомним, что эти равенства относятся к полупроводнику со сферическими поверхностями равной энергии, содержащему носителю заряда одного типа. После усреднения по E с помощью весовой функции из (5.87) или (5.88), получим

$$\begin{aligned} j_x &= \frac{ne^2}{m_e} (\langle\tau\rangle \mathcal{E}_x - \omega \langle\tau^2\rangle \mathcal{E}_y), \\ j_y &= \frac{ne^2}{m_e} (\langle\tau\rangle \mathcal{E}_y + \omega \langle\tau^2\rangle \mathcal{E}_x). \end{aligned} \quad (5.107)$$

Из условия $j_y = 0$ для угла Холла θ получаем выражение

$$\operatorname{tg} \theta = \frac{\mathcal{E}_y}{\mathcal{E}_x} = -\omega \frac{\langle\tau^2\rangle}{\langle\tau\rangle}. \quad (5.108)$$

Таким образом, пренебрегая, как и раньше, членами порядка ω^2 , имеем

$$\mathcal{E}_y = -\frac{B}{ne} \frac{\langle\tau^2\rangle j_x}{\langle\tau\rangle^2} = R_H B j_x, \quad (5.109)$$

откуда коэффициент Холла R_H равен

$$R_H = -\frac{r}{ne}, \quad (5.110)$$

где

$$r = \frac{\langle\tau^2\rangle}{\langle\tau\rangle^2}, \quad (5.111)$$

Формула (5.111) является обобщением формулы (5.39) на случай, когда τ зависит от E . Для дырок

$$R_H = \frac{r}{pe}, \quad (5.112)$$

где r по-прежнему определяется формулой (5.111), но под τ теперь следует понимать время релаксации дырок.

При рассеянии на колебаниях решетки (см. подразд. 5.3.2), для которого $\tau = aE^{-1/2}$, имеем

$$\langle \tau^2 \rangle = \frac{a^2 \int_0^{\infty} E^{1/2} \exp(-E/kT) dE}{\int_0^{\infty} E^{1/2} \exp(-E/kT) dE} = \frac{2a^2}{3kT}. \quad (5.113)$$

Воспользовавшись формулой (5.96), получаем $\langle \tau \rangle = 3\pi/8 = 1,18$. Итак,

$$R_H = -\frac{3\pi}{8ne} \quad \text{или} \quad \frac{3\pi}{8pe}. \quad (5.114)$$

В более общем случае при $\tau = aE^{-s}$ имеем

$$\langle \tau^2 \rangle = a^2 \frac{\int_0^{\infty} E^{\frac{3}{2}-2s} \exp(-E/kT) dE}{\int_0^{\infty} E^{1/2} \exp(-E/kT) dE} = a^2 (kT)^{-2s} \frac{\Gamma\left(\frac{5}{2}-2s\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (5.115)$$

Из (5.115) для $\langle \tau^2 \rangle$ и (5.96) для $\langle \tau \rangle$ получаем

$$r = \frac{\Gamma\left(\frac{5}{2}-2s\right) \Gamma\left(\frac{5}{2}\right)}{\left[\Gamma\left(\frac{5}{2}-s\right)\right]^2}. \quad (5.116)$$

В случае примесного рассеяния, для которого $s = -3/2$, получаем $r = 315\pi/512 = 1,93$. В вырожденных полупроводниках $\langle \tau^2 \rangle = [\tau(E_F)]^2$ и поэтому $r = 1$, т. е. получили тот же результат, что и для постоянного τ . При очень больших значениях B , как и раньше, $R \rightarrow -1/ne$ или $1/pe$, в чем можно убедиться, если в выражениях (5.107) перейти к большим значениям ω .

При наличии носителей заряда нескольких типов выражения (5.107), относящиеся к отдельным типам носителей, складываются. В частности, вклады электронов и дырок входят аддитивно в выражение для общего электрического тока. При условии одинаковой зависимости времен релаксации электронов и дырок от энергии E , поступая так же, как в подразд. 5.2.2, легко вывести следующую формулу для коэффициента Холла:

$$R_H = \frac{r}{e} \frac{p - b^2 n}{(p + bn)^2}, \quad (5.117)$$

где r определяется зависимостью (5.116).

Если поле таково, что членом $(\omega\tau)^2$ в знаменателях равенств (5.61) и (5.62) пренебречь уже нельзя, то для средних плотностей

токов можно написать формулы

$$\begin{aligned} j_x &= A \mathcal{E}_x - D \mathcal{E}_y, \\ j_y &= D \mathcal{E}_x + A \mathcal{E}_y, \end{aligned} \quad (5.118)$$

где

$$A = \frac{ne^2}{m_e} \left\langle \frac{\tau}{1 + \omega^2 \tau^2} \right\rangle, \quad (5.119)$$

$$D = \frac{ne^2 \omega}{m_e} \left\langle \frac{\tau^2}{1 + \omega^2 \tau^2} \right\rangle. \quad (5.120)$$

При $\tau = aE^{-s}$ получаем

$$A = \frac{4ane^2}{3\pi^{1/2} m_e (kT)^s} \int_0^\infty \frac{x^{2+s} e^{-x} dx}{x^{2s} + a^2 \omega^2 (kT)^{-2s}}, \quad (5.121)$$

$$D = \frac{4a^2 ne^2}{3\pi^{1/2} m_e (kT)^{2s}} \int_0^\infty \frac{x^{s/2} e^{-x} dx}{x^{2s} + a^2 \omega^2 (kT)^{-2s}}. \quad (5.122)$$

Коэффициент Холла, выраженный через коэффициенты A и D , имеет вид [ср. с формулой (5.68)]

$$R_H = -\frac{D}{B(A^2 + D^2)}. \quad (5.123)$$

Для случая рассеяния на колебаниях атомов решетки, для которого $s=1/2$, введя обозначение $\mu_{c0} = e \langle \tau \rangle / m_e$, получаем

$$A = ne \mu_{c0} I_2(\alpha), \quad (5.124)$$

$$D = \frac{3\pi^{1/2}}{4} ne \mu_{c0}^2 B I_{3/2}(\alpha), \quad (5.125)$$

где

$$\alpha = \frac{9\pi}{16} B^2 \mu_{c0}^2 \quad (5.126)$$

и

$$I_r(\alpha) = \int_0^\infty \frac{x^r e^{-x}}{x + \alpha} dx. \quad (5.127)$$

Для коэффициента Холла R_H в этом случае имеем

$$R_H = -\frac{3\pi^{1/2}}{4ne} \frac{I_{3/2}(\alpha)}{\left[I_2^2(\alpha) + \frac{9\pi}{16} \mu_{c0}^2 B^2 I_{3/2}^2(\alpha) \right]}. \quad (5.128)$$

При $B \rightarrow 0$, $I_2 \rightarrow 1$, а $I_{3/2} \rightarrow \pi^{1/2}/2$, так что, как и следовало ожидать,

$$R_H \rightarrow -\frac{3\pi}{8ne}.$$

При наличии смешанной проводимости используем формулу (5.71) со значениями коэффициентов A и D , соответствующими каждому типу носителей заряда в отдельности. Для дырок знак постоянной D необходимо изменить на обратный. В результате получаем

$$R_H = -\frac{3\pi^{1/2}}{4e} \frac{nb^2 I_{1/2}(\alpha_n) - p I_{1/2}(\alpha_p)}{\{nb I_{1/2}(\alpha_n) + p I_{1/2}(\alpha_p)\}^2 + \alpha_p \{nb^2 I_{1/2}(\alpha_n) - p I_{1/2}(\alpha_p)\}^2}, \quad (5.129)$$

где α_n и α_p — значения коэффициента α для электронов и дырок соответственно, b — отношение μ_{e0} для электронов к μ_{e0} для дырок. При очень больших значениях B и при $n \neq p$ имеем $R \rightarrow 1/e(p-n)$.

5.3.4. ЭФФЕКТ ХОЛЛА В ПОЛУПРОВОДНИКАХ С НЕСКОЛЬКИМИ МАКСИМУМАМИ ИЛИ МИНИМУМАМИ ЭНЕРГИИ

Приступим теперь к обобщению предыдущих результатов на случай полупроводника, обладающего несколькими эквивалентными минимумами в зоне проводимости или эквивалентными максимумами в валентной зоне. Для упрощения задачи допустим, что экстремумы лежат на трех взаимно перпендикулярных осях симметрии, как это имеет место, например, в зоне проводимости Si. Для начала рассмотрим простую эллипсоидальную поверхность постоянной энергии, причем будем считать, что магнитное поле направлено по оси z . Составляющие тензора обратной эффективной массы по осям x , y , z обозначим соответственно через $1/m_1$, $1/m_2$, $1/m_3$. Время релаксации τ предполагается изотропным и зависящим лишь от энергии E . В этом случае уравнения движения можно записать в виде

$$\begin{aligned} \dot{v}_x &= -\left(\frac{e}{m_1}\right) \mathcal{E}_x - \omega_1 v_y, \\ \dot{v}_y &= -\left(\frac{e}{m_2}\right) \mathcal{E}_y + \omega_2 v_x. \end{aligned} \quad (5.130)$$

Для решения этой системы умножим первое равенство на $\omega_2^{1/2}$, а второе на $\omega_1^{1/2}$ и введем обозначение $Z = \omega_1^{1/2} v_x + i \omega_2^{1/2} v_y$. Тогда система (5.130) будет эквивалентна уравнению

$$\dot{Z} = -e \left(\frac{\omega_2^{1/2}}{m_1} \mathcal{E}_x + i \frac{\omega_1^{1/2}}{m_2} \mathcal{E}_y \right) + i \omega_1^{1/2} \omega_2^{1/2} Z. \quad (5.131)$$

Соответствующее решение имеет вид

$$Z = Z_0 e^{i\omega t} - e \left\{ \frac{\omega_2^{1/2}}{m_1} \mathcal{E}_x + i \frac{\omega_1^{1/2}}{m_2} \mathcal{E}_y \right\} \frac{1 - e^{i\omega t}}{i\omega}, \quad (5.132)$$

где через ω^2 обозначено произведение $\omega_1\omega_2$. Изучение электропроводности показало, что при этих условиях для усреднения можно воспользоваться той же весовой функцией, что и раньше. Усредняя (5.132) по различным временам свободного пробега (см. подразд. 5.2.2), получаем

$$\bar{Z} = -\frac{\tau e}{1-i\omega\tau} \left\{ \frac{\omega_2^{1/2} \mathcal{E}_x}{m_1} + i \frac{\omega_1^{1/2} \mathcal{E}_y}{m_2} \right\}. \quad (5.133)$$

Разделив последнее выражение на действительную и мнимую части, приходим к системе равенств

$$-e\bar{v}_x = e^2 \left\{ \frac{\tau \mathcal{E}_x}{m_1(1+\omega^2\tau^2)} - \frac{\omega_1 \mathcal{E}_y \tau^2}{m_2(1+\omega^2\tau^2)} \right\}, \quad (5.134)$$

$$-e\bar{v}_y = e^2 \left\{ \frac{\tau \mathcal{E}_y}{m_2(1+\omega^2\tau^2)} + \frac{\omega_2 \mathcal{E}_x \tau^2}{m_1(1+\omega^2\tau^2)} \right\}. \quad (5.135)$$

При выводе формулы для коэффициента Холла в слабом поле можно отбросить члены, пропорциональные B^2 , и затем выполнить обычное усреднение τ по энергиям. Поскольку на всех трех осях (x , y , z) имеются эквивалентные эллипсоиды, необходимо произвести суммирование вкладов от всех эллипсоидов. Переставляя массы m_1 , m_2 , m_3 , получаем

$$-e \sum \bar{v}_x = \frac{ne^2}{3} \left\{ \langle \tau \rangle \mathcal{E}_x \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) - eB \langle \tau^2 \rangle \mathcal{E}_y \frac{m_1+m_2+m_3}{m_1m_2m_3} \right\}, \quad (5.136)$$

$$-e \sum \bar{v}_y = \frac{ne^2}{2} \left\{ \langle \tau \rangle \mathcal{E}_y \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) + eB \langle \tau^2 \rangle \mathcal{E}_x \frac{m_1+m_2+m_3}{m_1m_2m_3} \right\}. \quad (5.137)$$

Полагая $j_y = -e\sum \bar{v}_y = 0$ и пренебрегая членами, пропорциональными B^2 , получаем

$$j_x = -nem_1m_2m_3 \langle \tau \rangle^2 \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right)^2 \frac{\mathcal{E}_y}{3B \langle \tau^2 \rangle (m_1+m_2+m_3)}. \quad (5.138)$$

Таким образом, коэффициент Холла дается формулой

$$R_H = -\frac{\langle \tau^2 \rangle 3(m_1+m_2+m_3)}{ne \langle \tau \rangle^2 \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right)^2 m_1m_2m_3}. \quad (5.139)$$

Для эллипсоидов вращения $m_2=m_3=m_T$, $m_1=m_L$. Обозначая $K=m_L/m_T$, получаем

$$R_H = -\frac{r}{ne} \frac{3K(K+2)}{(2K+1)^2}, \quad (5.140)$$

где по-прежнему $r = \langle \tau^2 \rangle / \langle \tau \rangle^2$. Эта формула справедлива для любого кристалла с кубической симметрией, например для Ge (см.

формулу 5.219)¹⁾. Если, как и ранее, с помощью равенства $\sigma = ne\mu_c$ определить дрейфовую подвижность μ_c , а на основе равенства $R\sigma = \mu_H$ — холловскую подвижность, то отношение этих подвижностей запишется в виде

$$\frac{\mu_H}{\mu_c} = \frac{3rK(K+2)}{(2K+1)^2}. \quad (5.141)$$

Таким образом, значения коэффициента Холла в случае слабого поля для любой пары главных осей совпадают друг с другом. Отсюда следует, что измерение коэффициента Холла в слабом поле при любой ориентации тока j и магнитном поле, направленном перпендикулярно направлению тока, дает один и тот же результат. В этом можно убедиться следующим образом: отбросим члены второго порядка по полю и запишем выражения для плотностей тока в случае поля \mathbf{B} , направленного по оси z :

$$\left. \begin{aligned} j_x &= \sigma \mathcal{E}_x + R\sigma^2 \mathcal{E}_y B_z, \\ j_y &= \sigma \mathcal{E}_y - R\sigma^2 \mathcal{E}_x B_z, \\ j_z &= \sigma \mathcal{E}_z. \end{aligned} \right\} \quad (5.142)$$

Аналогичные системы равенств можно написать, когда магнитное поле направлено по оси x или по оси y . Они содержат те же коэффициенты σ и $R\sigma^2$, что и система (5.142), и отличаются от последней только перестановкой индексов. Далее, равенства, связывающие плотности тока с полями \mathcal{E} и \mathbf{B} в линейном приближении, можно в общем виде записать с учетом того, что в данном случае σ является скалярной величиной:

$$j_r = \sigma \mathcal{E}_r + \sum_{st} A_{rst} \mathcal{E}_s B_t. \quad (5.143)$$

В системе (5.142) и в двух других подобных системах полагаем по очереди $B_x = B_y = 0$, $B_x = B_z = 0$, $B_y = B_z = 0$. Тогда легко убедиться в том, что, за исключением случая $r \neq s \neq t$, коэффициент $A_{rst} = 0$; при $r \neq s \neq t$ имеем $|A_{rst}| = R\sigma^2$. Кроме того, $A_{rst} = -A_{rts}$ и т. д. В итоге равенство (5.143) можно переписать в векторной форме

$$\mathbf{j} = \sigma \mathcal{E} + R\sigma^2 [\mathcal{E} \times \mathbf{B}]. \quad (5.144)$$

Однако точно так должно выглядеть выражение для тока в изотропной среде. Таким образом, эффект Холла в слабом поле в рассматриваемом здесь случае изотропен. В сильных полях, когда нельзя пренебречь членами, содержащими B^2 , этот вывод, вообще говоря, может оказаться неверным. На основе формул (5.134) и

¹⁾ Общая теория изложена в работах [14, 15].

(5.135) легко показать, что при очень сильных полях $R_H \rightarrow -1/ne$ для электронов и $R_H \rightarrow 1/pe$ для дырок.

При смешанной проводимости достаточно подставить значения R_{H1} и R_{H2} для электронов и дырок в формулу (5.72). В приближении слабого поля имеем

$$R_H = \frac{R_1 \sigma_1^2 + R_2 \sigma_2^2}{(\sigma_1 + \sigma_2)^2}, \quad (5.145)$$

$$R_H = -\frac{1}{e} \frac{\mu_{He} \mu_{ce} n - \mu_{Hh} \mu_{ch} p}{(n \mu_{ce} + p \mu_{ch})^2}, \quad (5.146)$$

$$R_H = -\frac{1}{e} \frac{r_e b^2 n - r_h p}{(bn + p)^2}, \quad (5.147)$$

где $r_e = \mu_{He}/\mu_{ce}$, $r_h = \mu_{Hh}/\mu_{ch}$. Через μ_{He} , μ_{ce} обозначены соответственно холловская и дрейфовая подвижности электронов, а через μ_{Hh} , μ_{ch} — холловская и дрейфовая подвижности дырок. В случае когда зависимость τ от E можно представить в виде $\tau = aE^{-s}$ и s принимает одно и то же значение как для электронов, так и для дырок, формула (5.147) совпадает с выражением (5.117), причем $b = \mu_{ce}/\mu_{ch} = \mu_{He}/\mu_{Hh}$. Однако во многих полупроводниках на опыте обнаружено, что значения постоянной s для электронов и дырок несколько отличаются друг от друга. В этом случае необходимо пользоваться формулой более общего вида, а именно выражением (5.147). При очень сильных полях $R_H \rightarrow R_{H1}R_{H2}/(R_{H1} + R_{H2})$, что опять приводит к результату $R_H \rightarrow 1/e(p-n)$.

В наших рассуждениях мы пренебрегли магнитными квантовыми эффектами (магнитным квантованием), которые были упомянуты выше в связи с электропроводностью в очень сильных магнитных полях. Оказывается, что изменение равновесного распределения влияет на коэффициент Холла в меньшей степени, чем на электропроводность, так что результат $R_H \rightarrow 1/e(p-n)$ с хорошей степенью точности верен в условиях, когда $B\mu_e \gg 1$ или $B\mu_h \gg 1$ (см. разд. 12.8).

5.3.5. МАГНИТОПРОТИВЛЕНИЕ ПОЛУПРОВОДНИКА СО СФЕРИЧЕСКИМИ ПОВЕРХНОСТЯМИ ПОСТОЯННОЙ ЭНЕРГИИ

Рассмотрим изменение сопротивления полупроводника в магнитном поле, когда τ является функцией энергии E . Для этого необходимо вернуться к равенствам (5.61), (5.62). Поскольку мы рассматриваем лишь квадратичные по B эффекты, то в разложении выражения для j_x по степеням поля достаточно ограничиться членами, содержащими B^2 включительно. В случае полупроводника со сферическими поверхностями постоянной энергии такое

разложение имеет вид

$$j_x = \frac{ne^2}{m_e} (\tau \mathcal{E}_x - \omega \tau^2 \mathcal{E}_y - \omega^2 \tau^3 \mathcal{E}_x), \quad (5.148)$$

$$j_y = \frac{ne^2}{m_e} (\tau \mathcal{E}_y + \omega \tau^2 \mathcal{E}_x). \quad (5.149)$$

Эти выражения можно усреднить по энергиям, воспользовавшись уже знакомой нам весовой функцией. Полагая затем $j_y = 0$, получаем

$$j_x = \frac{ne}{m_e} \mathcal{E}_x \left[\langle \tau \rangle + \omega^2 \frac{\langle \tau^2 \rangle^2}{\langle \tau \rangle} - \omega^2 \langle \tau^3 \rangle \right]. \quad (5.150)$$

Введем обозначения

$$\sigma_0 = \frac{1}{\rho_0} = \frac{ne^2}{m_e} \langle \tau \rangle, \quad (5.151)$$

$$\sigma = \sigma_0 \left(1 - \frac{\Delta \sigma}{\sigma_0} \right), \quad (5.152)$$

$$\rho = \rho_0 \left(1 + \frac{\Delta \rho}{\rho_0} \right). \quad (5.153)$$

Тогда при $\Delta \sigma / \sigma_0 \ll 1$ имеем

$$-\frac{\Delta \sigma}{\sigma_0} = \frac{\Delta \rho}{\rho_0} = \frac{e^2 B^2 \langle \tau^3 \rangle \langle \tau \rangle - \langle \tau^2 \rangle^2}{m_e^2 \langle \tau \rangle^2}. \quad (5.154)$$

Обозначая через R_{Ho} значения коэффициента Холла в слабом поле [см. формулы (5.110), (5.111)], можно равенство (5.154) переписать в следующем виде:

$$\frac{\Delta \rho}{\rho_0} = -\frac{\Delta \sigma}{\sigma_0} = \xi R_{\text{Ho}}^2 \sigma_0^2 B^2, \quad (5.155)$$

где ξ — коэффициент магнитосопротивления, который определяется из соотношения

$$\xi + 1 = \frac{\langle \tau^3 \rangle \langle \tau \rangle}{\langle \tau^2 \rangle^2}. \quad (5.156)$$

В невырожденном полупроводнике в случае рассеяния только на колебаниях решетки, когда $\tau \sim E^{-1/2}$, получаем $\xi + 1 = 4/\pi$, так что $\xi = 0,275$. В этом легко убедиться, если соответствующие средние величины степеней τ выразить в виде интегралов типа (5.87), которые в свою очередь сводятся к обычным Γ -функциям. При рассеянии на ионизованной примеси $\tau \sim E^{1/2}$; в этом случае $1 + \xi = 32,768/6,615\pi$, так что $\xi = 0,577$. Можно показать, что численное значение произведения $\langle \tau^3 \rangle \langle \tau \rangle$ всегда больше или равно $\langle \tau^2 \rangle^2$, следовательно, $\xi \geq 0$. При $\tau \sim E^{-s}$

$$\xi + 1 = \frac{\Gamma\left(\frac{5}{2} - 3s\right) \Gamma\left(\frac{5}{2} - s\right)}{\left[\Gamma\left(\frac{5}{2} - 2s\right)\right]^2}. \quad (5.157)$$

Если $\tau = \text{const}$, то $\xi = 0$, в чем мы уже могли убедиться в подразд. 5.2.3. Это имеет место в сильно вырожденном полупроводнике, в котором вместо усреднения τ по E достаточно взять значение τ при энергии, равной энергии Ферми (см. подразд. 5.3.2). Согласно (5.155), в полупроводниках со сферическими поверхностями постоянной энергии в отсутствие вырождения может существовать поперечное магнитосопротивление, однако продольное магнитосопротивление при этом отсутствует, так как магнитное поле, направленное по оси z , не изменяет ток j_z .

Возвращаясь к равенствам (5.61), (5.62), легко видеть, что в очень сильных магнитных полях после усреднения по энергиям получаем

$$j_x = \frac{ne^2}{m_e} \left[-\frac{\mathcal{E}_y}{\omega} \right], \quad (5.158)$$

$$j_y = \frac{ne^2}{m_e} \left[\mathcal{E}_y \left\langle \frac{1}{\tau} \right\rangle + \omega \mathcal{E}_x \right]. \quad (5.159)$$

При $j_y = 0$ имеем $\mathcal{E}_x = -\mathcal{E}_y \frac{\langle 1/\tau \rangle}{\omega}$, так что

$$j_x = \frac{ne^2}{m_e} \left\langle \frac{1}{\tau} \right\rangle^{-1} \mathcal{E}_x. \quad (5.160)$$

Обозначив через σ_∞ и ρ_∞ соответственно значения электропроводности и электросопротивления в предельном случае сильного поля, получим

$$\frac{\sigma_0}{\sigma_\infty} = \left\langle \frac{1}{\tau} \right\rangle \langle \tau \rangle = \frac{\rho_\infty}{\rho_0}. \quad (5.161)$$

Итак, с ростом магнитного поля сопротивление стремится к предельному значению ρ_∞ . Для рассеяния по закону $\tau \sim E^{-s}$ имеем

$$\frac{\rho_\infty}{\rho_0} = \frac{\Gamma\left(\frac{5}{2} + s\right) \Gamma\left(\frac{5}{2} - s\right)}{\left[\Gamma\left(\frac{5}{2}\right)\right]^2}. \quad (5.162)$$

Для решеточного рассеяния, когда $s = 1/2$, получаем $\rho_\infty/\rho_0 = 32/9\pi$, тогда как при рассеянии на примесных ионах $s = -3/2$ и поэтому $\rho_\infty/\rho_0 = 32/3\pi$.

В случае смешанной проводимости можно определить зависимость σ от B с помощью формулы (5.78), куда необходимо подставить выражения для проводимостей σ_1 и σ_2 для каждого типа носителей заряда. При очень больших B , очевидно, имеем

$$\sigma_\infty = \frac{\sigma_1 \sigma_2 (R_1 + R_2)^2}{\sigma_1 R_1^2 + \sigma_2 R_2^2}, \quad (5.163)$$

где под σ_1 , σ_2 , R_1 и R_2 подразумеваются значения электропроводности и коэффициента Холла для случая сильного поля B . Если

индекс 1 у σ и R относится к электронам, а индекс 2 — к дыркам, так что $R_1 = -1/ne$ и $R_2 = 1/pe$, то, используя формулу (5.161), получаем

$$\frac{\sigma_\infty}{\sigma_0} = \frac{\rho_0}{\rho_\infty} = \frac{(n-p)^2}{(\sigma_{0e} + \sigma_{0h})(\sigma_1^{-1}\rho^2 + \sigma_2^{-1}n^2)} = \quad (5.164)$$

$$= \frac{(n-p)^2}{\left(n \frac{\langle \tau_e \rangle}{m_e} + p \frac{\langle \tau_h \rangle}{m_h}\right) \left(nm_e \left\langle \frac{1}{\tau_e} \right\rangle + pm_h \left\langle \frac{1}{\tau_h} \right\rangle\right)}. \quad (5.165)$$

Вводя обозначения

$$\sigma_{\infty e} = \frac{ne^2}{m_e} \left\langle \frac{1}{\tau_e} \right\rangle^{-1} = ne\mu_{\infty e} \quad (5.166)$$

и

$$\sigma_{\infty h} = \frac{pe^2}{m_h} \left\langle \frac{1}{\tau_h} \right\rangle^{-1} = pe\mu_{\infty h}, \quad (5.167)$$

формулу (5.165) можно переписать в виде

$$\frac{\sigma_\infty}{\sigma_0} = \frac{\rho_0}{\rho_\infty} = \frac{(n-p)^2}{(n\mu_e + p\mu_h) \left(\frac{n}{\mu_{\infty e}} + \frac{p}{\mu_{\infty h}}\right)}; \quad (5.168)$$

здесь $\mu_{\infty e}$ и $\mu_{\infty h}$ можно назвать подвижностями соответственно электронов и дырок в сильных полях. Так как они зависят от средних величин $\langle 1/\tau \rangle$, то при наличии более одного механизма рассеяния вклады различных механизмов просто складываются. В случае сильного поля не требуется отдельного определения холловской и дрейфовой подвижностей, так как для каждого типа носителей заряда $R_\infty \sigma_\infty = \mu_\infty$. В случае собственного полупроводника коэффициент Холла в сильном поле равен нулю, а предельное значение магнитосопротивления в сильном поле, согласно (5.168), стремится к бесконечности. Отметим, однако, что на опыте довольно трудно достичь области сильных магнитных полей, когда выполняется неравенство $\omega \langle \tau \rangle \gg 1$. Кроме того, уравнения, на основе которых выведены формулы для эффектов в сильном поле, строго говоря, неприменимы при $\omega \langle \tau \rangle \gg 1$, поскольку в них не учтено квантование энергии в магнитном поле. К этому вопросу мы еще вернемся в разд. 12.8. Если произведение $\omega \langle \tau \rangle$ выразить через магнитную индукцию и дрейфовую подвижность, то $\omega \langle \tau \rangle = B\mu_e$. При очень больших подвижностях, скажем порядка $10 \text{ м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ (такие подвижности можно получить при низких температурах в структурно совершенных и предельно очищенных монокристаллах некоторых веществ) и $B = 10^5 \text{ Гс}$ (10 Т), имеем $\omega \langle \tau \rangle = 100$, т. е. выполняется (классический) критерий сильного поля.

Магнитосопротивление полупроводника со смешанной проводимостью в слабом поле можно исследовать либо на основе формул вида (5.148) и (5.149) для каждого типа носителей заряда, либо с помощью формулы (5.78), которая справедлива в общем случае

как для постоянного, так и для зависящего от энергии времени релаксации τ . Заметим, что если в разложении σ удерживать члены с B^2 включительно, то необходимо учесть не только все члены равенства (5.79), но и члены, пропорциональные B^2 в разложениях σ_1 и σ_2 . Вводя обозначения $\Delta_1 = \Delta\sigma_1/\sigma_1 B^2$ и т. д., получаем

$$\frac{\Delta\sigma}{\sigma_0 B^2} = \frac{\sigma_{10}\Delta_1 + \sigma_{20}\Delta_2}{\sigma_{10} + \sigma_{20}} - \frac{\sigma_{10}\sigma_{20}(\sigma_{10}R_{10}^2 + \sigma_{20}R_{20}^2)}{\sigma_{10} + \sigma_{20}} + \frac{\sigma_{10}^2\sigma_{20}^2(R_{10} + R_{20})^2}{(\sigma_{10} + \sigma_{20})^2}. \quad (5.169)$$

Если под σ_1 , σ_2 понимать соответственно электронную и дырочную электропроводности, то последние два члена в правой части (5.169) сводятся к выражению

$$-\frac{n\rho\mu_{ce}\mu_{ch}(\mu_{He} + \mu_{Hh})^2}{(n\mu_{ce} + \rho\mu_{ch})^2}; \quad (5.170)$$

оно имеет тот же вид, что и правая часть в формуле (5.80) для постоянного τ , так как в этом случае $\mu_{ce} = \mu_{He} = \mu_e$, $\mu_{ch} = \mu_{Hh} = \mu_h$. Первый член можно записать в виде

$$-\frac{n\mu_{ce}\mu_{He}^2\xi_{He} + \rho\mu_{ch}\mu_{Hh}^2\xi_{Hh}}{n\mu_{ce} + \rho\mu_{ch}}, \quad (5.171)$$

где

$$1 + \xi_e = \frac{\langle\tau_e^3\rangle\langle\tau_e\rangle}{\langle\tau_e^2\rangle^2},$$

$$1 + \xi_h = \frac{\langle\tau_h^3\rangle\langle\tau_h\rangle}{\langle\tau_h^2\rangle^2}.$$

Поэтому полное выражение для $\Delta\sigma/\sigma_0$ имеет вид

$$\frac{\Delta\sigma}{\sigma_0} = -B^2\mu_{Hh}^2 \left\{ \frac{n\rho b(1+b')^2}{(nb+\rho)^2} + \frac{nb b'^2 \xi_e + \rho \xi_h}{nb+\rho} \right\}, \quad (5.172)$$

где $b = \mu_{ce}/\mu_{ch}$, $b' = \mu_{He}/\mu_{Hh}$. Оба члена в правой части (5.172) дают отрицательный вклад в $\Delta\sigma$. Первый из них обусловлен смешанным характером проводимости и исчезает при $n=0$ или $\rho=0$. Второй член описывает вклад в магнитосопротивление каждого типа носителей заряда в отдельности. Он исчезает лишь в том случае, если $\tau = \text{const}$ как для электронов, так и для дырок.

Коэффициент магнитосопротивления ξ в области смешанной проводимости определяется из равенства

$$\xi = -\frac{\Delta\sigma}{\sigma_0^3 R_{H0}^2 B^2}; \quad (5.173)$$

сюда необходимо подставить значение R_{H0} из формулы (5.146). Тогда на основе (5.172) и (5.173) можно получить

$$\xi + 1 = \frac{(nb+\rho)[nb b'^2(\xi_e+1) + \rho(\xi_h+1)]}{(b b' n - \rho)^2}. \quad (5.174)$$

Если воспользоваться известными выражениями для средних значений $\langle \tau^n \rangle$ для различных степеней n , то формулу (5.174) легко представить еще и в следующем виде:

$$\xi + 1 = \frac{\left(n \frac{\langle \tau_e \rangle}{m_e} + \rho \frac{\langle \tau_h \rangle}{m_h} \right) \left(n \frac{\langle \tau_e^3 \rangle}{m_e^3} + \rho \frac{\langle \tau_h^3 \rangle}{m_h^3} \right)}{\left(n \frac{\langle \tau_e^2 \rangle}{m_e^2} - \rho \frac{\langle \tau_h^2 \rangle}{m_h^2} \right)^2}. \quad (5.175)$$

5.3.6. МАГНИТОСОПРОТИВЛЕНИЕ ПОЛУПРОВОДНИКОВ С ЭЛЛИпсоИДАЛЬНЫМИ ПОВЕРХНОСТЯМИ ПОСТОЯННОЙ ЭНЕРГИИ

Строгая теория магнитосопротивления полупроводников с эллипсоидальными поверхностями постоянной энергии весьма сложна. Ее развили Эйблис и Мейбум [14], а также Шибуба [16]. Здесь мы рассмотрим лишь несколько сравнительно простых частных случаев. Для начала предположим, что имеется всего один минимум энергии при $k=0$ и что поверхности постоянной энергии имеют вид эллипсоидов, главные оси которых направлены по осям координат x , y и z . Магнитное поле предполагается направленным по оси z . Разлагая равенства (5.134) и (5.135) в ряды и ограничиваясь членами с B^2 , получаем после усреднения по τ

$$j_x = ne^2 \left[\frac{\langle \tau \rangle}{m_1} \mathcal{E}_x - \frac{eB \langle \tau^2 \rangle}{m_1 m_2} \mathcal{E}_y - \frac{e^2 B^2 \langle \tau^3 \rangle}{m_1^2 m_2} x \mathcal{E} \right], \quad (5.176)$$

$$j_y = ne^2 \left[\frac{\langle \tau \rangle}{m_2} \mathcal{E}_y + \frac{eB \langle \tau^2 \rangle}{m_1 m_2} \mathcal{E}_x \right].$$

Положив, как и ранее, $j_y=0$, получим следующее выражение для j_x :

$$j_x = ne^2 \left[\frac{\langle \tau \rangle}{m_1} + \frac{e^2 B^2 \langle \tau^2 \rangle^2}{m_1^2 m_2 \langle \tau \rangle} - \frac{e^2 B^2 \langle \tau^3 \rangle}{m_1^2 m_2} \right] \mathcal{E}_x, \quad (5.177)$$

так что

$$\frac{\Delta \sigma}{\sigma_0} = - \frac{e^2 B^2}{m_1 m_2} \xi \frac{\langle \tau^2 \rangle^2}{\langle \tau \rangle^2}, \quad (5.178)$$

где коэффициент ξ определяется формулой (5.156).

В случае если имеется несколько эллипсоидов постоянной энергии, оси вращения которых направлены вдоль осей координат, то к уравнениям (5.176) следует добавить вклад от каждого такого эллипсоида. Введем обозначения: $m_1 = m_L$, $m_2 = m_T = m_T$, $K = m_L/m_T$. Такая ситуация осуществляется в Si. Второй член в (5.177) всегда

можно записать в виде $B^2 R^2 \sigma_0^3$. При учете вкладов от других эллипсоидов это соотношение сохраняет силу, но с подстановкой соответствующей величины R_{H_0} . В результате выражение для полного тока приобретает вид

$$j_x = \mathcal{E}_x \left[\sigma_0 + R_{H_0}^2 \sigma_0^3 B^2 - ne^4 B^2 \langle \tau^3 \rangle \left(\frac{1}{m_L m_T^2} + \frac{1}{m_T m_L^2} + \frac{1}{m_T^3} \right) \right], \quad (5.179)$$

где

$$\sigma_0 = \frac{ne^2 \langle \tau \rangle}{3} \left(\frac{2}{m_T} + \frac{1}{m_L} \right).$$

Используя для R_{H_0} величину R_H [см. выражение (5.139)], перепишем формулу (5.179) в виде

$$j_x = \sigma_0 \mathcal{E}_x \left[1 + R_{H_0}^2 \sigma_0^3 B^2 \left(1 - \frac{\langle \tau^3 \rangle \langle \tau \rangle (K^2 + K + 1)(2K + 1)}{\langle \tau^2 \rangle^2 K(2 + K)^2} \right) \right]. \quad (5.180)$$

Далее находим, что

$$-\frac{\Delta \sigma}{\sigma_0} = R_{H_0}^2 \sigma_0^3 B^2 \{ (1 + \xi) F(K) - 1 \}, \quad (5.181)$$

где ξ определяется формулой (5.156), а

$$F(K) = \frac{(K^2 + K + 1)(2K + 1)}{K(K + 2)^2}. \quad (5.182)$$

При $m_L = m_T$ имеем $K = 1$ и $F(K) = 1$, тогда формула (5.181) непосредственно сводится к формуле (5.155).

Коэффициент магнитосопротивления, который в этом случае обозначают обычно, как ξ_{100}^{001} , где нижние индексы указывают направление электрического тока, а верхние — направление магнитного поля, определяется соотношением

$$\xi_{100}^{001} = (1 + \xi) F(K) - 1. \quad (5.183)$$

Поскольку при этих условиях

$$j_z = \sigma_0 \mathcal{E}_z, \quad (5.184)$$

то

$$\xi_{100}^{001} = 0, \quad (5.185)$$

т. е. продольное магнитосопротивление в направлении $\langle 100 \rangle$ равно нулю. Как будет видно из дальнейшего, в кристаллах с кубической симметрией имеются три независимых коэффициента магнитосопротивления, так что для расчета магнитосопротивления при произвольной ориентации тока и магнитного поля необходимо знать еще и остальные коэффициенты ξ . Расчет коэффициента магнитосопротивления, когда ток и поле направлены вдоль оси $\langle 110 \rangle$, т. е. когда $B_y = B_x$, $B_z = 0$, $j_x = j_y$, $j_z = 0$, может быть выполнен путем решения соответствующих уравнений движения и последующего их

усреднения. Эти вычисления весьма громоздки и мы их опускаем. Окончательный результат имеет вид

$$\xi_{110}^{\text{II}} = (1 + \xi) \frac{(K-1)^2 (2K+1)}{2K(K+2)^2}. \quad (5.186)$$

Общее выражение для электрического тока \mathbf{j} , рассматриваемого как функция векторов напряженности электрического поля \mathcal{E} и магнитной индукции \mathbf{B} в линейном по \mathcal{E} и квадратичном по \mathbf{B} приближении, можно записать следующим образом:

$$\mathbf{j} = \sigma_0 \mathcal{E} + R_0 \sigma_0^2 [\mathcal{E} \times \mathbf{B}] + \mathbf{M} \mathcal{E} \mathbf{B} \mathbf{B}, \quad (5.187)$$

где \mathbf{M} — тензор четвертого ранга. Первое и второе слагаемые следуют из формулы (5.144), третье представляет собой общий вид записи членов, линейных по \mathcal{E} и одновременно квадратичных по \mathbf{B} . Формула (5.187), записанная в компонентах векторов по осям координат (x, y, z) , обозначенных нижними индексами $r, s, t=1, 2, 3$, имеет вид

$$j_r = \sigma_0 \mathcal{E}_r + R_0 \sigma_0^2 (\mathcal{E}_s B_t - \mathcal{E}_t B_s) + M_{rstw} \mathcal{E}_w B_s B_w. \quad (5.188)$$

В этой формуле, как обычно, предполагается суммирование по дважды повторяющимся индексам. В кристаллах с кубической симметрией значительное число компонент тензора \mathbf{M} равно нулю. В этом можно убедиться, произведя замену направлений одного из векторов в ряде симметричных конфигураций. В общем случае отличны от нуля лишь следующие компоненты тензора:

$$M_{rrrr}, M_{rrss} (r \neq s), M_{rsrs} (r \neq s).$$

Эти компоненты не меняются при взаимных перестановках индексов r, s, t ; M_{rssr} и M_{rsrs} представляют собой одну и ту же составляющую. Итак, в квадратичном по B члене в (5.188) имеются три независимые постоянные. Если поверхности постоянной энергии представляют собой эллипсоиды вращения вокруг соответствующих осей x, y, z , то $M_{rrrr}=0$, поскольку в направлении этих осей коэффициент продольного магнитосопротивления равен нулю. Если же оси эллипсоидов вращения направлены по осям $\langle 111 \rangle$, то, как мы увидим ниже, коэффициент продольного магнитосопротивления отличен от нуля для любого направления в кристалле, поэтому $M_{rrrr} \neq 0$.

Обозначим $M_{rrrr}=S$, $M_{rrss}=P$, $M_{rsrs}=Q$ и перепишем формулу (5.188) в виде

$$j_r = \sigma_0 \mathcal{E}_r + R_0 \sigma_0^2 (\mathcal{E}_s B_t - \mathcal{E}_t B_s) + S \mathcal{E}_r B_s^2 + P \mathcal{E}_r (B_s^2 + B_t^2) + Q (\mathcal{E}_s B_r B_s + \mathcal{E}_t B_r B_t). \quad (5.189)$$

Легко видеть, что она полностью эквивалентна следующему выражению, которое вывел Зейтц [17] для кубических кристаллов, ис-

ходя из общих соображений симметрии:

$$\mathbf{j} = \sigma_0 \mathbf{E} + \alpha [\mathbf{E} \times \mathbf{V}] + \beta \mathbf{E} B^2 + \gamma \mathbf{V} (\mathbf{E} \cdot \mathbf{V}) + \delta \mathbf{T} \mathbf{E}, \quad (5.190)$$

где α , β , γ , δ — постоянные числа, \mathbf{T} — диагональный тензор с компонентами B_r^2 , B_s^2 , B_z^2 . Из сравнения двух формул (5.190) и (5.189) находим

$$\alpha = R_0 \sigma_0^2, \quad \beta = P, \quad \gamma = Q, \quad \delta = S - P - Q.$$

При экспериментальном исследовании эффекта магнитосопротивления обычно измеряют зависимость падения напряжения на образце от напряженности внешнего магнитного поля при неизменной величине протекающего по образцу электрического тока. Формулу (5.190) можно переписать таким образом, чтобы выразить напряженность электрического поля \mathbf{E} как функцию плотности тока \mathbf{j} . Очевидно, в линейном по \mathbf{V} приближении

$$\mathbf{E} = \frac{\mathbf{j}}{\sigma_0} - R_0 [\mathbf{j} \times \mathbf{V}]. \quad (5.191)$$

Подставляя это выражение для \mathbf{E} в линейный по \mathbf{V} член выражения (5.190) и заменяя \mathbf{E} в квадратичном по \mathbf{V} члене выражения \mathbf{j}/σ_0 , получаем

$$\sigma_0 \mathbf{E} = \mathbf{j} - \sigma_0 R_0 [\mathbf{j} \times \mathbf{V}] + b \mathbf{j} B^2 + c \mathbf{V} (\mathbf{j} \cdot \mathbf{V}) + d \mathbf{T} \mathbf{j};$$

при этом использовано соотношение

$$[\mathbf{V} \times [\mathbf{j} \times \mathbf{V}]] = \mathbf{j} B^2 - \mathbf{V} (\mathbf{j} \cdot \mathbf{V}).$$

Сопротивление ρ определяется выражением $(\mathbf{E} \cdot \mathbf{j})/j^2$, поэтому имеем

$$\frac{\Delta \rho}{\rho_0 B^2} = \frac{(\mathbf{E} \cdot \mathbf{j}) - \rho_0 j^2}{B^2 (\mathbf{E}_0 \cdot \mathbf{j})} = b + c \frac{(\mathbf{j} \cdot \mathbf{V})^2}{j^2 B^2} + d \frac{(j_x^2 B_x^2 + j_y^2 B_y^2 + j_z^2 B_z^2)}{j^2 B^2}, \quad (5.192)$$

где

$$b = -R_0^2 \sigma_0^2 - \frac{\beta}{\sigma_0} = -R_0^2 \sigma_0^2 - \frac{P}{\sigma_0},$$

$$c = R_0^2 \sigma_0^2 - \frac{\gamma}{\sigma_0} = R_0^2 \sigma_0^2 - \frac{Q}{\sigma_0},$$

$$d = -\frac{\delta}{\sigma_0} = \frac{P + Q - S}{\sigma_0};$$

S , P , Q — введенные выше обозначения для компонент M_{rrrr} , M_{rrss} , M_{rsrs} тензора \mathbf{M} . Обозначив через i , j , k направляющие косинусы вектора плотности электрического тока, а через l , m , n соответственно направляющие косинусы вектора напряженности

магнитного поля, получим

$$\frac{\Delta\rho}{\rho_0 B^2} = b + c(il + jm + kn)^2 + d(i^2l^2 + j^2m^2 + k^2n^2). \quad (5.193)$$

Это равенство, предложенное Бардином, приведено в работе Пирсона и Суля [18], которые воспользовались им для анализа результатов своих измерений магнитосопротивления в Si и Ge. Оно весьма удобно для определения угловой зависимости магнитосопротивления, возникающей при изменении ориентации полей относительно осей кристалла.

Если поверхности постоянной энергии представляют собой эллипсоиды вращения вокруг осей координат x, y, z , то $S=0$, а величины P и Q могут быть выражены через коэффициенты магнитосопротивления ζ . Например, при B , направленном по оси z , а j — по оси x , получим простое соотношение $\Delta\sigma/\sigma_0 B^2 = -b$, так что

$$R_0^2 \sigma_0^3 + P = \frac{\Delta\sigma}{B^2}, \quad (5.194)$$

или

$$\frac{\Delta\sigma}{R_0^2 \sigma_0^3 B^2} = 1 + \frac{P}{R_0^2 \sigma_0^3} = -\zeta_{100}^{001}. \quad (5.195)$$

Таким образом,

$$M_{rrss} = P = -R_0^2 \sigma_0^3 (\zeta_{100}^{001} + 1) = \quad (5.196)$$

$$= -R_0^2 \sigma_0^3 (1 + \xi) \frac{(K^2 + K + 1)(2K + 1)}{K(K + 2)^2}. \quad (5.197)$$

С другой стороны, полагая $B_x = B_y, B_z = 0$ и $j_x = j_y, j_z = 0$, имеем

$$\frac{\Delta\sigma}{\sigma_0 B^2} = -\left(b + c + \frac{d}{2}\right) = \frac{P + Q}{2\sigma_0}. \quad (5.198)$$

Но

$$\frac{1}{2}(P + Q) = -R_0^2 \sigma_0^3 \zeta_{110}^{110}, \quad (5.199)$$

поэтому

$$M_{rrss} = Q = R_0^2 \sigma_0^3 (\zeta_{110}^{001} - 2\zeta_{110}^{110} + 1) = \quad (5.200)$$

$$= 3R_0^2 \sigma_0^3 (1 + \xi) \frac{2K + 1}{K(K + 2)^2}. \quad (5.201)$$

Итак, получаем

$$\frac{M_{rrss}}{M_{rrss}} = \frac{Q}{P} = -\frac{3K}{K^2 + K + 1}. \quad (5.202)$$

Коэффициенты b, c, d выражаются следующим образом через величины ζ :

$$\left. \begin{aligned} b &= R_0^2 \sigma_0^3 \zeta_{100}^{001}, \\ c &= R_0^2 \sigma_0^3 (2\zeta_{110}^{110} - \zeta_{100}^{001}), \\ d &= -2R_0^2 \sigma_0^3 \zeta_{110}^{110}. \end{aligned} \right\} \quad (5.203)$$

В кристаллах с кубической симметрией возможно еще и другое симметричное расположение поверхностей постоянной энергии, а именно: эквивалентные эллипсоиды могут располагаться вдоль осей $\langle 111 \rangle$ куба. Примером подобной структуры является зона проводимости в Ge. Для рассмотрения такой более сложной ситуации необходимо обобщить теоретический метод, использованный в подразд. 5.3.4, на более общий случай недиагонального тензора обратной эффективной массы. Направим координатные оси (x, y, z) вдоль направлений $\langle 100 \rangle$ в кристалле. Рассмотрим вначале изоэнергетическую поверхность в виде эллипсоида вращения, ось которого направлена вдоль диагонали $\langle 111 \rangle$ и который характеризуется поперечной эффективной массой m_T и продольной эффективной массой m_L . Введем новую систему координат (x', y', z') с таким расчетом, чтобы направляющие косинусы оси x' в прежней системе координат равнялись $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, иными словами, так, чтобы ось x' была направлена по диагонали $\langle 111 \rangle$. Пусть, далее, ось y' образует прямой угол с осью x' и характеризуется направляющими косинусами $(1/\sqrt{2}, 0, -1/\sqrt{2})$, а ось z' направлена перпендикулярно к двум первым осям и имеет направляющие косинусы $(-1/\sqrt{6}, 2/\sqrt{6}, -1/\sqrt{6})$. В новой системе координат уравнение эллипсоида, соответствующего энергии E , имеет вид

$$E = \frac{\hbar^2}{8\pi^2} \left[\frac{k_x'^2}{m_L} + \frac{k_y'^2 + k_z'^2}{m_T} \right]. \quad (5.204)$$

Используя матрицу преобразования

$$\begin{array}{ccc} & x & y & z \\ x' & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ y' & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ z' & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{array},$$

перейдем к системе координат (x, y, z) . Тогда уравнение эллипсоида постоянной энергии относительно прежней системы отсчета приобретает вид

$$E = \frac{\hbar^2}{8\pi^2} \left[\frac{1}{3} \left(\frac{2}{m_T} + \frac{1}{m_L} \right) (k_x^2 + k_y^2 + k_z^2) - \frac{2}{3} \left(\frac{1}{m_T} - \frac{1}{m_L} \right) (k_x k_y + k_y k_z + k_z k_x) \right]. \quad (5.205)$$

Введем обозначения

$$\frac{1}{m_c} = \frac{1}{3} \left(\frac{2}{m_T} + \frac{1}{m_L} \right), \quad (5.206)$$

$$\frac{1}{m_d} = \frac{1}{3} \left(\frac{1}{m_T} - \frac{1}{m_L} \right), \quad (5.207)$$

причем m_c совпадает с использованной ранее так называемой дрейфовой эффективной массой. Таким образом, тензор обратной эффективной массы обладает следующими отличными от нуля компонентами: $m_{xx}^{-1} = m_{yy}^{-1} = m_{zz}^{-1} = m_c^{-1}$, $m_{xy}^{-1} = m_{xz}^{-1} = m_{yz}^{-1} = -m_d^{-1}$. Если магнитное поле направлено по оси z , то уравнения движения электронов имеют вид [см. формулу (2.39)]

$$\begin{aligned} \dot{v}_x &= -\frac{e}{m_c} (\mathcal{E}_x + Bv_y) + \frac{e}{m_d} (\mathcal{E}_y - Bv_x) + \frac{e}{m_d} \mathcal{E}_z, \\ \dot{v}_y &= -\frac{e}{m_c} (\mathcal{E}_y - Bv_x) + \frac{e}{m_d} (\mathcal{E}_x + Bv_y) + \frac{e}{m_d} \mathcal{E}_z, \\ \dot{v}_z &= -\frac{e}{m_c} \mathcal{E}_z + \frac{e}{m_d} (\mathcal{E}_x + Bv_y) + \frac{e}{m_d} (\mathcal{E}_y - Bv_x). \end{aligned} \quad (5.208)$$

С помощью обозначений $\omega_c = eB/m_c$ и $\omega_d = eB/m_d$ первые два уравнения можно переписать в виде

$$\begin{aligned} \dot{v}_x + \dot{v}_y &= -e \left(\frac{1}{m_c} - \frac{1}{m_d} \right) \mathcal{E}_x - \\ &- e \left(\frac{1}{m_c} - \frac{1}{m_d} \right) \mathcal{E}_y + \frac{2e}{m_d} \mathcal{E}_z + (\omega_c - \omega_d) (v_x - v_y), \\ \dot{v}_x - \dot{v}_y &= -e \left(\frac{1}{m_c} + \frac{1}{m_d} \right) \mathcal{E}_x + \\ &+ e \left(\frac{1}{m_c} + \frac{1}{m_d} \right) \mathcal{E}_y + (\omega_c + \omega_d) (v_x + v_y). \end{aligned} \quad (5.209)$$

Аналогично тому, как мы это делали в подразд. 5.3.4., эту систему уравнений можно решить относительно $(v_x + v_y)$ и $(v_x - v_y)$ и затем усреднить по времени релаксации τ . Действительно, равенства (5.209) можно представить в виде

$$\begin{aligned} \dot{v}_x + \dot{v}_y &= F + \omega_1 (v_x - v_y), \\ \dot{v}_x - \dot{v}_y &= G - \omega_2 (v_x + v_y), \end{aligned} \quad (5.210)$$

где $\omega_1 = \omega_c - \omega_d$ и $\omega_2 = \omega_c + \omega_d$. Введем обозначение $\Omega^2 = \omega_1 \omega_2 = \omega_c^2 - \omega_d^2$, тогда сравнение с (5.134), (5.135) позволяет по аналогии сразу написать решение

$$\begin{aligned} \bar{v}_x + \bar{v}_y &= F \left\langle \frac{\tau}{1 + \Omega^2 \tau^2} \right\rangle - \omega_1 G \left\langle \frac{\tau^2}{1 + \Omega^2 \tau^2} \right\rangle, \\ \bar{v}_x - \bar{v}_y &= G \left\langle \frac{\tau}{1 + \Omega^2 \tau^2} \right\rangle + \omega_2 F \left\langle \frac{\tau^2}{1 + \Omega^2 \tau^2} \right\rangle, \end{aligned} \quad (5.211)$$

откуда

$$\bar{v}_x = \frac{1}{2} (F + G) \left\langle \frac{\tau}{1 + \Omega^2 \tau^2} \right\rangle + \frac{1}{2} (\omega_2 F - \omega_1 G) \left\langle \frac{\tau^2}{1 + \Omega^2 \tau^2} \right\rangle, \quad (5.212)$$

$$\bar{v}_y = \frac{1}{2} (F - G) \left\langle \frac{\tau}{1 + \Omega^2 \tau^2} \right\rangle - \frac{1}{2} (\omega_2 F + \omega_1 G) \left\langle \frac{\tau^2}{1 + \Omega^2 \tau^2} \right\rangle.$$

Поскольку ω_1 , ω_2 и Ω пропорциональны магнитной индукции B , то выражения для \bar{v}_x и \bar{v}_y можно разложить в ряды по степеням B , после чего получим

$$\begin{aligned} \bar{v}_x = \frac{1}{2} (F + G) \langle \tau \rangle - \frac{1}{2} (\omega_2 F - \omega_1 G) \langle \tau^2 \rangle - \\ - \frac{1}{2} (F + G) \Omega^2 \langle \tau^3 \rangle, \end{aligned} \quad (5.213)$$

$$\bar{v}_y = \frac{1}{2} (F - G) \langle \tau \rangle - \frac{1}{2} (\omega_2 F + \omega_1 G) \langle \tau^2 \rangle. \quad (5.214)$$

Теперь необходимо произвести усреднение по различным эквивалентным эллипсоидам. Это осуществляется заменой эффективных масс m_{yz} , m_{zx} , m_{xy} на $\pm m_d$ путем последовательных перестановок. В результате находим

$$\bar{F} = -\frac{e}{m_c} (\mathcal{E}_x + \mathcal{E}_y),$$

$$\bar{G} = -\frac{e}{m_c} (\mathcal{E}_x - \mathcal{E}_y),$$

$$\overline{\omega_2 F} = -e^2 B \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) (\mathcal{E}_x + \mathcal{E}_y),$$

$$\overline{\omega_1 G} = -e^2 B \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) (\mathcal{E}_x - \mathcal{E}_y),$$

$$\overline{\Omega^2 F} = \frac{e^3 B^2}{m_c} \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) (\mathcal{E}_x + \mathcal{E}_y),$$

$$\overline{\Omega^2 G} = \frac{e^3 B^2}{m_c} \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) (\mathcal{E}_x - \mathcal{E}_y).$$

По определению $j_x = -ne\bar{v}_x$ и т. д. Поэтому

$$\begin{aligned} j_x = ne^2 \left\{ \frac{\langle \tau \rangle}{m_c} \mathcal{E}_x - eB \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) \langle \tau^2 \rangle \mathcal{E}_y - \right. \\ \left. - \frac{e^2 B^2}{m_c} \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) \langle \tau^3 \rangle \mathcal{E}_x \right\}, \end{aligned} \quad (5.215)$$

$$j_y = ne^2 \left\{ \frac{\langle \tau \rangle}{m_c} \mathcal{E}_y + eB \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right) \langle \tau^2 \rangle \mathcal{E}_x \right\}. \quad (5.216)$$

Отсюда сразу получим

$$\sigma_0 = ne^2 \frac{\langle \tau \rangle}{m_c}, \quad (5.217)$$

$$R_{H0} = -\frac{1}{ne} \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} m_c^2 \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right). \quad (5.218)$$

Используя обозначение $K = M_L / M_T$, имеем $m_c = 3M_L / (2K + 1)$, $m_d = 3M_L (K - 1)$ и тогда формула (5.218) приобретает вид

$$R_{H0} = -\frac{1}{ne} \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} \frac{3K(K+2)}{(2K+1)^2}. \quad (5.219)$$

Выражения для R_{H0} и σ_0 получились такими же, как и для полупроводников, у которых оси вращения эллипсоидов постоянной энергии ориентированы вдоль направлений $\langle 100 \rangle$.

Из (5.215) получаем для магнитосопротивления

$$\frac{\Delta\sigma}{\sigma_0} = R_{H0}^2 \sigma_0^2 B^2 - \frac{ne^4 B^2}{m_c \sigma_0} \langle \tau^3 \rangle \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right), \quad (5.220)$$

откуда

$$-\frac{\Delta\sigma}{R_{H0}^2 \sigma_0^2 B^2} = \frac{\langle \tau^3 \rangle \langle \tau \rangle}{\langle \tau^2 \rangle^2} \frac{(2K+1)^2}{3K(K+2)} - 1. \quad (5.221)$$

Таким образом,

$$1 + \zeta_{100}^{001} = \frac{(2K+1)^2 (1+\xi)}{3K(K+2)}. \quad (5.222)$$

Из выражений (5.212) находим также предельное значение σ при $B \rightarrow \infty$, а именно:

$$\frac{\sigma_0}{\sigma_\infty} = \left\langle \frac{1}{\tau} \right\rangle \langle \tau \rangle m_c^2 \left(\frac{1}{m_c^2} - \frac{1}{m_d^2} \right), \quad (5.223)$$

$$\frac{\sigma_0}{\sigma_\infty} = \left\langle \frac{1}{\tau} \right\rangle \langle \tau \rangle \frac{(2K+1)^2}{3K(K+2)}, \quad (5.224)$$

и, как раньше, $R_\infty = -1/ne$.

Вычисление значений ξ и σ_∞ для других направлений является более громоздким. Окончательный результат следующий [9]:

$$\zeta_{100}^{100} = 2(\xi + 1)(K-1)^2 \frac{2K+1}{3K(K+2)^2}, \quad (5.225)$$

$$\zeta_{110}^{110} = \frac{1}{2} \zeta_{100}^{100}. \quad (5.226)$$

Кроме того,

$$\left(\frac{\sigma_0}{\sigma_\infty} \right)_{100}^{100} = \frac{(2K+1)(K+2)}{9K}; \quad (5.227)$$

$$\left(\frac{\sigma_0}{\sigma_\infty} \right)_{110}^{110} = \frac{(2K+1)^2}{3K(K+2)}. \quad (5.228)$$

Заметим, что в этом случае коэффициент продольного магнитосопротивления ζ_{100}^{100} отличен от нуля, так что сопротивление в магнитном поле меняется даже тогда, когда ток параллелен направлению магнитного поля и оба вектора ориентированы вдоль одного из направлений $\langle 100 \rangle$.

Зная величины ζ_{100}^{001} , ζ_{100}^{100} , ζ_{110}^{110} , можно вычислить также коэффициенты M_{ruvw} формулы (5.188). По-прежнему [см. формулу (5.196)] имеем

$$M_{rrss} = P = -R_{\text{H}0}^2 \sigma_0^3 (\zeta_{100}^{001} + 1), \quad (5.229)$$

$$M_{rrss} = -R_{\text{H}0}^2 \sigma_0^3 (1 + \xi) \frac{(2K+1)^2}{3K(K+2)}. \quad (5.230)$$

Для определения ζ_{110}^{110} используем соотношение

$$\frac{\Delta\sigma}{\sigma_0 B^2} = -\left(b + c + \frac{1}{2}d\right) = \frac{1}{2}(P + Q + S), \quad (5.231)$$

откуда

$$P + Q + S = -2R_{\text{H}0}^2 \sigma_0^3 \zeta_{110}^{110} = -R_{\text{H}0}^2 \sigma_0^3 \zeta_{100}^{100}. \quad (5.232)$$

Коэффициент ζ_{100}^{100} можно определить из соотношения

$$\frac{\Delta\sigma}{\sigma_0 B^2} = -(b + c + d) = \frac{S}{\sigma_0}, \quad (5.233)$$

так что

$$M_{rrrr} = S = -R_{\text{H}0}^2 \sigma_0^3 \zeta_{100}^{100}, \quad (5.234)$$

$$M_{rrrr} = -2R_{\text{H}0}^2 \sigma_0^3 (1 + \xi) \frac{(K-1)^2 (2K+1)}{3K(K+2)^2}. \quad (5.235)$$

Отсюда следует также, что $P = -Q$ и

$$M_{rrss} = -M_{rsrs}. \quad (5.236)$$

Наконец, через величины ξ можно выразить постоянные b , c , d . Из предыдущих соотношений следует

$$\left. \begin{aligned} b &= R_{\text{H}0}^2 \sigma_0^2 \zeta_{100}^{001}, \\ c &= -R_{\text{H}0}^2 \sigma_0^2 \zeta_{100}^{001}, \\ d &= R_{\text{H}0}^2 \sigma_0^2 \zeta_{100}^{100}. \end{aligned} \right\} \quad (5.237)$$

Эффект магнитосопротивления в полупроводниках, имеющих поверхности постоянной энергии в виде эллипсоидов, а также имеющих несколько минимумов (так же, как это имеет место для зон проводимости Si и Ge), обстоятельно описан Гликсманом [19]. Обсуждение еще более сложной проблемы в случае полупроводников с гофрированными поверхностями постоянной энергии (как в случае валентных зон Si и Ge) было проведено Донованом и Гербертом [20]. Обзор этой проблемы сделал Бир [21].

Магнитное квантование, упомянутое выше в связи с измерениями проводимости и коэффициента Холла при очень сильных магнитных полях, также будет оказывать определенное влияние на величи-

ны коэффициентов магнитосопротивления. Однако обычно измерения проводятся при достаточно малых магнитных полях, при которых эффекты магнитного квантования несущественны (см. разд. 12.8).

5.4. Электропроводность при очень низких температурах

На первый взгляд может показаться, что при температурах столь низких, что все свободные носители «заморожены» на донорных или акцепторных центрах, проводимость полупроводника должна быть очень мала. Это будет иметь место при условии, что kT значительно меньше энергии ионизации примеси ϵ_d . Для примесных центров с мелкими уровнями, например с $\epsilon_d \approx 0,01$ эВ (таких, как элементы V группы в Ge), это означает, что температура T должна быть существенно ниже 100 К. Например, следует ожидать, что при 10К или ниже такие доноры в большинстве своем будут неионизованными, т. е. не будут поставлять электроны в зону проводимости. При больших концентрациях доноров, скажем порядка 10^{19} см $^{-3}$, волновые функции электронов соседних доноров перекрываются и образуется так называемая примесная зона, т. е. квазинепрерывная узкая полоса разрешенных уровней, лежащих внутри запрещенной зоны (см. рис. 5.6). Если на каждый донор приходится по одному электрону, то эта зона будет заполнена только наполовину и процесс электропроводности будет проходить так же, как и в металле. Из-за узости этой зоны подвижность носителей заряда в ней оказывается небольшой, однако достаточной для создания легко наблюдаемой проводимости.

Оказалось, однако, что проводимость имеет место и при концентрациях доноров или акцепторов намного меньших, чем 10^{19} см $^{-3}$, вплоть до 10^{12} — 10^{16} см $^{-3}$. В этих случаях проводимость по примесной зоне не может объяснить величины электропроводности. Найдено, что электропроводность при низких температурах оказывается наибольшей, если присутствуют *одновременно как доноры, так и акцепторы*. В таких случаях полупроводник называют «компенсированным». В полупроводнике p -типа с $N_a > N_d$ степень компенсации K равна отношению N_d/N_a . Рассмотрим, что будет происходить в полупроводнике p -типа с $K \neq 0$. При низких температурах, когда $kT \ll \epsilon_d$, все N_d уровней доноров будут пустыми, так как их электроны перейдут на N_d акцепторов, которые в результате окажутся заряженными отрицательно, т. е. ионизованными. Таким образом мы получим N_d ионизованных и $N_a - N_d$ неионизованных акцепторов (см. рис. 5.7). Ионизованные акцепторы будут распределены по материалу, так что дырки с неионизованных акцепторов смогут при помощи туннельного процесса перескакивать с одного вакантного центра на другой.

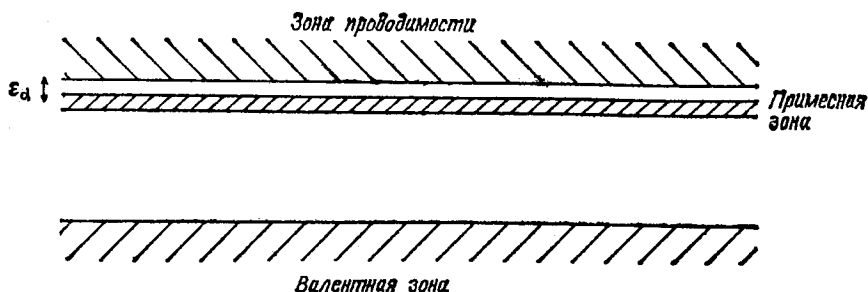


Рис. 5.6. Возникновение примесной зоны вследствие высокой концентрации доноров на глубине ϵ_d (в пределе малых концентраций) под дном зоны проводимости.

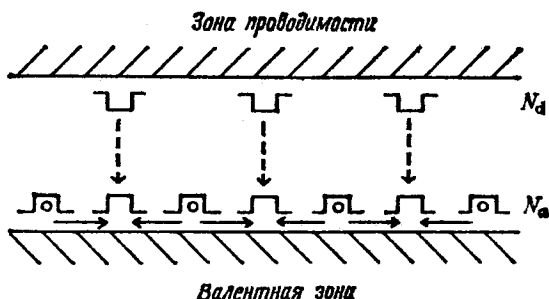


Рис. 5.7. Частичная компенсация N_a акцепторов N_d донорами ($N_a > N_d$) при низких температурах, в результате которой становится возможной прыжковая проводимость дырок.

Эту форму электропроводности обстоятельно исследовал на постоянном токе Фрицше [22], на переменном токе низкой частоты ее изучали Поллак и Геболл [23] и на частотах микроволнового диапазона — Танака и Фэн [24]. Сопоставление электропроводности с поглощением в далекой инфракрасной области позволило обнаружить электропроводность такого типа на частотах, соответствующих длинам волн порядка 100 мкм (см. подразд. 10.2.1).

Вообще говоря, электропроводность σ можно выразить в виде

$$\sigma = A_1 \exp \left[-\frac{\epsilon_1}{kT} \right] + A_2 \exp \left[-\frac{\epsilon_2}{kT} \right] + A_3 \exp \left[-\frac{\epsilon_3}{kT} \right]. \quad (5.238)$$

Первый член обусловлен тепловой ионизацией примесей, так что при самых низких температурах он пренебрежимо мал. Второй член возникает в результате проводимости по примесной зоне и оказывается несущественным, если либо N_d , либо N_a меньше 10^{18} см⁻³. Третий член обусловлен прыжковым характером проводимости и обычно наиболее важен в области очень низких температур. Фрицше [22] обнаружил, что на постоянном токе A_3 зависит глав-

ным образом от концентрации основных носителей заряда, в то время как ε_3 зависит от степени компенсации $K = N_a/N_d$ (для n -типа или $K = N_d/N_a$ (для p -типа). При $K \rightarrow 0$ ε_3 возрастает. ε_3 зависит также и от частоты переменного тока, уменьшаясь при возрастании последней. Действительно, для частот, соответствующих длинам волн между 200 и 800 мкм, σ очень слабо зависит от температуры; это указывает на то, что для таких частот ε_3 мало (см. подразд. 10.2.1). Аналогичные результаты получил Блекмор [25], который изучал вопросы, связанные с одновременным определением N_a и N_d из измерений постоянной Холла в широком диапазоне температур. Эти эксперименты проводились на Ge. Вудбери и Блекмор [26] провели измерения на GaAs, легированном Mn, и обнаружили существование всех трех типов проводимости.

Проводимость по примесной зоне и прыжковую проводимость систематически исследовали Мотт и Туз [27], которые получили довольно хорошее согласие с экспериментом, если принять во внимание значительную неопределенность степени компенсации ввиду сложности измерений ее величины.

Считается, что энергия активации ε_3 обусловлена электрическим полем, притягивающим электрон или дырку обратно к центру, с которого носитель перескакивает. Используя частоты порядка 9 ГГц при 10 К, Учинокура и Танака [28] исследовали влияние поля гомеоплярных примесных пар на прыжковую проводимость и тоже получили хорошее согласие с теорией.

С теоретической точки зрения эти процессы проводимости очень похожи на процессы, происходящие при электропроводности в некристаллических твердых телах, и в этой связи обсуждались Моттом и Дэвисом [29].

ЛИТЕРАТУРА

1. *Smith R. A.*, Wave Mechanics of Crystalline Solids, Chapman and Hall and John Wiley and Sons, New York, 2 nd ed., 1969.
2. *Hall E. H.*, Amer. Journ. Maths., 2, 287 (1879); Amer. Journ. Sci., 19, 200 (1880).
3. *Chambers R. G.*, Proc. Phys. Soc., A65, 903 (1952).
- 4*. *Давыдов Б. И., Шмушкевич И. М.*, УФН, 24, 21 (1940).
5. *Banbury P. C., Henisch H. K., Many A.*, Proc. Phys. Soc., A66, 753 (1953).
6. *Landauer R., Swanson J.*, Phys. Rev., 91, 555 (1953).
7. *Wilson A. H.*, The Theory of Metals, Cambridge University Press, Cambridge, 2 nd ed., 1953. [Имеется перевод 1-го изд.: *Вильсон А.* Квантовая теория металлов.— М.-Л., 1941.]
8. *Ziman J. M.*, Electrons and Phonons, Oxford University Press, Oxford, 1960. [Имеется перевод: *Займан Дж.* Электроны и фононы.— М.: ИЛ, 1962.]
9. *Brooks H.*, Advances in Electronics and Electron Physics, Academic Press, New York, 7, 85 (1955).
10. *Jones H.*, Phys. Rev., 81, 149 (1951).
11. *Johnson V. A., Lark-Horowitz K.*, Phys. Rev., 82, 977 (1951).
12. *Conwell E. M.*, Proc. IRE, 11, 1327 (1952).
13. *Herring C., Vogt E.*, Phys. Rev., 101, 944 (1956).

14. *Abeles B., Meiboom S.*, Phys. Rev., **95**, 31 (1954).
15. *Herring C.*, Bell Syst. Techn. Journ., **34**, 237 (1955).
16. *Shibuya M.*, Phys. Rev., **95**, 1388 (1954).
17. *Seitz F.*, Phys. Rev., **79**, 372 (1950).
18. *Pearson G. L., Suhl H.*, Phys. Rev., **83**, 768 (1951).
19. *Gliksman M.*, Progress in Semicond., ed. Gibson A. F., Aigrain P., Burgess R. E. Heywood, New York, **3**, 1 (1958).
20. *Donovan B.*, в сб.: Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, New York, 1969, p. 79.
21. *Beer A. C.*, в сб.: Solid State Physics. ed. Seitz F., Turnbull D., Academic Press, New York, 1963, Suppl. 4.
22. *Fritzsche H.*, Phys. Rev., **99**, 406 (1955).
23. *Pollak M., Geballe T. H.*, Phys. Rev., **122**, 1472 (1961).
24. *Tanaka S., Fan H. Y.*, Phys. Rev. **132**, A564 (1964).
25. *Blakemore J. S.*, Phil. Mag., **4**, 560 (1959).
26. *Woodbury D. A., Blakemore J. S.*, Phys. Rev. **B8**, 3803 (1973).
27. *Mott N. F., Twose W. D.*, Adv. in Phys., **10**, 107 (1961).
28. *Uchinokura K., Tanaka S.*, Phys. Rev., **153**, 828 (1967).
29. *Mott N. F., Davis A. E.*, Electronic Processes in Non-crystalline Solids, Oxford University Press, Oxford, 1971.

ЭЛЕКТРОННЫЕ ТЕПЛОВЫЕ ЯВЛЕНИЯ

6.1. Теплопроводность

В предыдущей главе мы рассмотрели явления переноса электрического заряда в полупроводниках, обусловленные электрическими и магнитными полями. Теперь мы остановимся на некоторых эффектах, возникающих при наличии температурных градиентов. Если имеется градиент температуры, то функция $f(\mathbf{k})$, описывающая распределение электронов по разрешенным энергетическим уровням в зависимости от величины волнового вектора \mathbf{k} электрона, отличается от равновесной. Это вызывает перенос энергии в форме потока тепла и перенос электрического заряда. Отсюда ясно, что электрические и тепловые эффекты должны быть взаимно связаны. В простейшем случае, в отсутствие внешних полей, вдоль направления градиента температуры возникает перенос кинетической энергии, обусловленный перераспределением электронов по энергиям, т. е. явление электронной теплопроводности:

В температурно-неоднородной среде уже нельзя использовать простые уравнения движения, усредненные по равновесному распределению. Для того чтобы получить выражения для плотности электрического тока и потока тепла следует воспользоваться изменением функции распределения $f(\mathbf{k})$, которое определяется уравнением Больцмана. Соответствующий вывод описан в известных учебниках по кинетическим явлениям в твердых телах и газах, и мы не будем приводить подробного вывода нового выражения для плотности тока (см. [1], § 10.5—10.7). Формулу (5.85) следует заменить выражением

$$j_x = \frac{ne}{m_e} \left[e\mathcal{E}_x + T \frac{\partial}{\partial T} \left(\frac{E_F}{T} \right) \frac{dT}{dx} \right] \langle \tau \rangle + \frac{ne}{Tm_e} \frac{dT}{dx} \langle \tau E \rangle, \quad (6.1)$$

где, как и ранее, E_F —энергия Ферми. Для того чтобы найти тепловой поток W_x , прежде всего заменим $-e$ на E перед прямыми скобками и во втором члене и затем произведем усреднение, тогда полу-

чим

$$W_x = -\frac{n}{m_e} \left[e\mathcal{E}_x + T \frac{\partial}{\partial T} \left(\frac{E_F}{T} \right) \frac{dT}{dx} \right] \langle \tau E \rangle - \frac{n}{T m_e} \frac{dT}{dx} \langle \tau E^2 \rangle. \quad (6.2)$$

В отсутствие электрического тока можно исключить \mathcal{E}_x и получить выражение для W_x (заметим, что $\mathcal{E}_x \neq 0$, если $dT/dx \neq 0$); таким образом, получим выражение

$$W_x = -\frac{n}{m_e T} \frac{dT}{dx} \frac{\langle \tau \rangle \langle E^2 \tau \rangle - \langle E \tau \rangle^2}{\langle \tau \rangle}. \quad (6.3)$$

Ниже мы увидим, что наряду со средней величиной $\langle \tau \rangle$ мы будем иметь дело со средними $\langle \tau E \rangle$ и $\langle \tau E^2 \rangle$; весовая функция используется прежняя.

В полупроводнике теплопроводность имеет две составляющие: одна обусловлена колебаниями атомов около своих положений равновесия (в кристалле это колебания решетки), а другая составляющая обусловлена подвижными носителями заряда, электронами и дырками. Теплопроводность первого типа мы будем обозначать через κ_L . Электронную часть теплопроводности обозначим через κ_e для полупроводника n -типа и через κ_h для полупроводника p -типа. В случае смешанной проводимости обозначим ее через κ_m . Электронная составляющая теплопроводности определяется выражением

$$\kappa_e = -\frac{W_x}{dT/dx} = \frac{n [\langle \tau \rangle \langle E^2 \tau \rangle - \langle E \tau \rangle^2]}{T m_e \langle \tau \rangle}. \quad (6.4)$$

Поскольку в нашем случае электропроводность σ равна $ne^2 \langle \tau \rangle / m_e$, то

$$\kappa_e = L \sigma T = \mathcal{L} \left(\frac{k^2}{e^2} \right) \sigma T, \quad (6.5)$$

где

$$\mathcal{L} = \frac{\langle \tau \rangle \langle E^2 \tau \rangle - \langle E \tau \rangle^2}{k^2 T^2 \langle \tau \rangle^2} \quad (6.6)$$

— простой численный множитель. Обычно $L = \mathcal{L} (k^2/e^2)$ называют числом Лоренца. В невырожденном полупроводнике при $\tau = aE^{-s}$

$$\left. \begin{aligned} \langle \tau \rangle &= \frac{a \Gamma \left(\frac{5}{2} - s \right)}{\Gamma \left(\frac{5}{2} \right)}, \\ \langle \tau E \rangle &= akT \frac{\Gamma \left(\frac{7}{2} - s \right)}{\Gamma \left(\frac{5}{2} \right)}, \\ \langle \tau E^2 \rangle &= ak^2 T^2 \frac{\Gamma \left(\frac{9}{2} - s \right)}{\Gamma \left(\frac{5}{2} \right)}. \end{aligned} \right\} \quad (6.7)$$

Воспользовавшись соотношением $\Gamma(1+x) = x\Gamma(x)$, получаем $\mathcal{L} = \frac{5}{2} - s$. При рассеянии на акустических колебаниях, когда $s = \frac{1}{2}$, имеем $\mathcal{L} = 2$ (классическая электронная теория дает значение $\mathcal{L} = 3$).

В случае сильного вырождения при вычислении средних величин типа $\langle \tau \rangle$ нельзя ограничиться первым приближением. В первом приближении, т. е. с точностью до членов первого порядка по kT/E_F , число Лоренца $\mathcal{L} = 0$. Действительно [ср. с формулой (5.91)], в этом приближении

$$\langle E\tau \rangle = E_F \tau(E_F), \quad \langle E^2\tau \rangle = E_F^2 \tau(E_F).$$

Для вычисления $\langle \tau \rangle$, $\langle E\tau \rangle$ и $\langle E^2\tau \rangle$, определяемых выражениями вида

$$\langle \tau \rangle = -E_F^{-3/2} \int_0^{\infty} \tau E^{3/2} \frac{\partial f_0}{\partial E} dE \quad (6.8)$$

[ср. с формулой 5.90)], необходимо вычислить интеграл

$$-\int_0^{\infty} \varphi(E) \frac{\partial f_0}{\partial E} dE$$

с более высоким приближением. Ранее в случае сильного вырождения такой интеграл полагался равным просто $\varphi(E_F)$. Теперь необходимо вычислить его с большей степенью точности. Разлагая $\varphi(E)$ в ряд по степеням $(E - E_F)$ вблизи $E = E_F$, получаем

$$\varphi(E) = \varphi(E_F) + (E - E_F) \varphi'(E_F) + \frac{1}{2} (E - E_F)^2 \varphi''(E_F) + \dots \quad (6.9)$$

Подстановка в исходный интеграл и почленное интегрирование дают

$$\begin{aligned} \int_0^{\infty} \frac{\partial f_0}{\partial E} dE &= -1, \\ \int_0^{\infty} (E - E_F) \frac{\partial f_0}{\partial E} dE &= 0, \\ \int_0^{\infty} (E - E_F)^2 \frac{\partial f_0}{\partial E} dE &= -\frac{\pi^2}{3} k^2 T^2. \end{aligned}$$

Таким образом,

$$-\int_0^{\infty} \varphi(E) \frac{\partial f_0}{\partial E} dE = \varphi(E_F) + \frac{\pi^2}{6} (kT)^2 \varphi''(E_F) + \dots \quad (6.10)$$

Формула (6.10) позволяет вычислить величины $\langle \tau \rangle$, $\langle E\tau \rangle$, $\langle E^2\tau \rangle$ во втором приближении по kT/E_F . Проводя затем необходимые преобразования, получим, что

$$\langle \tau \rangle \langle E^2\tau \rangle - \langle E\tau \rangle^2 = \frac{1}{3} \pi^2 k^2 T^2 \langle \tau \rangle^2, \quad (6.11)$$

откуда $\mathcal{L} = \pi^2/3$. Для промежуточной степени вырождения интегралы, определяющие $\langle \tau \rangle$, $\langle E\tau \rangle$ и т. д., необходимо оценить численно. Если $\tau = aE^{-s} (s < 3/2)$, то

$$\frac{\langle E^n \tau \rangle}{\langle \tau \rangle} = \frac{\left(\frac{3}{2} + n - s\right) F_{\frac{1}{2} + n - s}}{\left(\frac{3}{2} - s\right) F_{\frac{1}{2} - s}}, \quad (6.12)$$

где

$$F_r(E_F) = \int_0^\infty E^r f_0 dE. \quad (6.13)$$

Таблицы функций $F_r(E_F)$ составили Макдугл и Стонер [2] для некоторых полуцелочисленных значений r и Родс [3] для целочисленных значений r . Зависимость \mathcal{L} от E_F/kT при $s = 1/2$ изображена на рис. 6.1.

В случае полупроводников p -типа удобнее пользоваться функцией распределения $f'_0(E)$ для дырок

$$f'_0(E) = 1 - f_0(E). \quad (6.14)$$

Если ввести энергию Ферми для дырок $E'_F = -\Delta E - E_F$, где ΔE — ширина запрещенной зоны, а E' — кинетическая энергия дырки, и учесть, что энергию любого вакантного электронного уровня (т. е. занятого дыркой) можно представить в виде $E = -\Delta E - E'$, то

$$f'_0(E', E'_F) = f_0(E, E_F). \quad (6.14a)$$

Предполагая, что $\tau_h = a'E^{-s'}$, можно для плотности электрического тока получить следующее выражение:

$$j_x = p\mu_h \left\{ e\mathcal{E}_x - \left[T \frac{\partial}{\partial T} \left(\frac{E'_F}{T} \right) + k \left(\frac{5}{2} - s' \right) \right] \frac{dT}{dx} \right\}. \quad (6.15)$$

Для вычисления потока энергии необходимо заменить e на $E' + \Delta E$. Тогда

$$\begin{aligned} W_x = & \frac{p\mu_h \left(\frac{5}{2} - s' \right) kT}{e} \left\{ e\mathcal{E}_x - \left[T \frac{\partial}{\partial T} \left(\frac{E'_F}{T} \right) + k \left(\frac{7}{2} - s' \right) \right] \frac{dT}{dx} \right\} + \\ & + \frac{p\mu_h \Delta E}{e} \left\{ e\mathcal{E}_x - \left[T \frac{\partial}{\partial T} \left(\frac{E'_F}{T} \right) + k \left(\frac{5}{2} - s' \right) \right] \frac{dT}{dx} \right\}. \quad (6.16) \end{aligned}$$

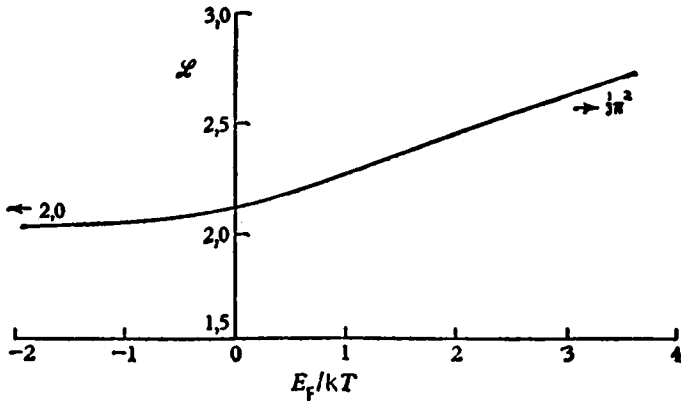


Рис. 6.1. Множитель \mathcal{L} в числе Лоренца $L = \mathcal{L}(k/e)^2$ как функция E_F/kT при $\tau = aE^{-1/2}$

Второй член (пропорциональный ΔE) в правой части (6.16) представляет собой как бы перенос рекомбинационной энергии дырок и электронов (см. ниже); он исчезает одновременно с током дырок. Полагая $j_x = 0$ и исключая $e\mathcal{E}_x$, получаем

$$\kappa_n = -\frac{\bar{W}_x}{dT/dx} = p\mu_n \left(\frac{5}{2} - s' \right) \frac{k^2 T}{e}. \quad (6.17)$$

Таким образом,

$$\kappa_n = \left(\frac{5}{2} - s' \right) \left(\frac{k^2}{e^2} \right) \sigma T, \quad (6.18)$$

что отличается от κ_e лишь заменой s на s' .

Весьма интересное явление имеет место в области смешанной проводимости, когда электронный и дырочный токи не исчезают, несмотря на то что плотность суммарного тока равна нулю. Второй член в формуле (6.16) может дать значительный вклад в теплопроводность лишь при $\Delta E \gg kT$. Выражения для суммарного электрического и теплового потоков в случае смешанной проводимости можно получить, сложив попарно выражения для соответствующих потоков, определяемых выражениями (6.1), (6.15) и (6.2), (6.16). Если из полученной таким путем системы уравнений исключить \mathcal{E}_x и пренебречь небольшой зависимостью ΔE от T , то для теплопроводности, обусловленной носителями заряда, в области смешанной проводимости получим выражение

$$\frac{\kappa_m}{\sigma T} = \frac{k^2 \left(\frac{5}{2} - s \right) n\mu_e + k^2 \left(\frac{5}{2} - s' \right) p\mu_n}{e^2 (n\mu_e + p\mu_n)} + \frac{k^2 \left[5 - s - s' + \frac{\Delta E}{kT} \right]^2 n\mu_e \mu_n}{e^2 (n\mu_e + p\mu_n)^2}. \quad (6.19)$$

К полученному выражению для κ_m — теплопроводности носителей заряда следует прибавить κ_L — теплопроводность решетки. Полагая $s=s'=1/2$, получаем для полной теплопроводности кристалла κ выражение

$$\frac{\kappa}{\sigma T} = \frac{\kappa_L}{\sigma T} + \frac{2k^2}{e^2} + \left(4 + \frac{\Delta E}{kT}\right)^2 \frac{k^2 n p \mu_e \mu_h}{e^2 (n \mu_e + p \mu_h)^2}, \quad (6.20)$$

которое совпадает с результатом, приведенным в работе Патли [4]¹⁾. Первый член в правой части (6.20), как было сказано выше, описывает теплопроводность решетки, третий связан с наличием смешанной электропроводности. При $n \approx p$ и $\Delta E \gg kT$ вклад последнего механизма в теплопроводность может быть значительным. Рекомбинационный механизм не может служить объяснением значительного роста теплопроводности при низких температурах, обнаруженного в некоторых полупроводниках. Последний эффект, исследованный экспериментально в InSb Бушем и Шнейдером [5], несомненно, обусловлен дополнительным переносом тепла фононами (см. разд. 8.4). Теорию этого вопроса разработал Тер-Хаар и Нивс [6].

Относительный вклад теплопроводности, обусловленной свободными электронами и дырками, в полную теплопроводность полупроводниковых кристаллов в общем довольно мал. Например, если $\mathcal{L}=2$, то $\kappa_e = 2(k^2/e^2)\sigma T = 4,4 \cdot 10^{-5} \sigma$ при $T=300\text{K}$ (если σ выразить в единицах $\text{Ом}^{-1} \cdot \text{м}^{-1}$, а κ_e — в единицах $\text{Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$). Рассмотрим два почти предельных случая. В хороших монокристаллах германия

$$\kappa = 0,15 \text{ кал} \cdot \text{с}^{-1} \cdot \text{см}^{-1} \cdot \text{К}^{-1} \approx 60 \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}.$$

Таким образом, в германии при сопротивлении $1 \text{ Ом} \cdot \text{см}$ ($\sigma = 100 \text{ Ом}^{-1} \cdot \text{м}^{-1}$) имеем $\kappa_e = 4,4 \cdot 10^{-4} \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$, так что $(\kappa_e/\kappa) = 10^{-5}$. В случае же $\text{Вi}_2\text{Тe}_3$, когда $\kappa = 6 \cdot 10^{-3} \text{ кал} \cdot \text{с}^{-1} \cdot \text{см}^{-1} \cdot \text{К}^{-1} = 2,4 \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$ и $\sigma = 10^3 \text{ Ом}^{-1} \cdot \text{м}^{-1}$, получаем $\kappa_e = 4,5 \cdot 10^{-1} \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$, так что $(\kappa_e/\kappa) = 0,2$.

Обзоры работ по исследованию теплопроводности полупроводников составлены А. В. Иоффе и А. Ф. Иоффе [7], А. Ф. Иоффе [17], Аппелем [8] (см. также монографию Драббла и Голдсмита [9]).

6.2. Термоэлектродвижущая сила

Выше мы видели, что если в полупроводнике существует градиент температуры, то в нем возникает э. д. с. даже в отсутствие электрического тока. Для полупроводника n -типа, полагая $j_x=0$

¹⁾ Необходимо отметить, что эквивалентная зависимость для κ получена Давыдовым и Шмушкевичем еще в 1940 г. [28*]. Тогда же этими авторами предсказан эффект добавочной теплопроводности, обусловленной электронно-дырочной рекомбинацией. — *Прим. ред.*

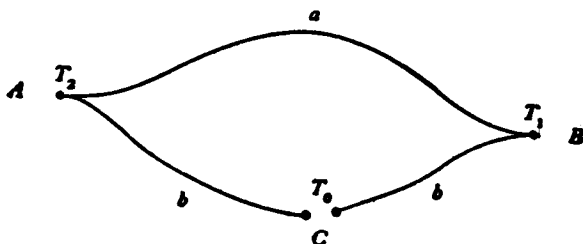


Рис. 6.2.

в уравнении (6.1), найдем, что электрическое поле равно

$$\mathcal{E}_x = -\frac{1}{e} \left[T \frac{d}{dT} \left(\frac{E_F}{T} \right) + \frac{1}{T} \frac{\langle E\tau_e \rangle}{\langle \tau_e \rangle} \right] \frac{dT}{dx}. \quad (6.21)$$

Скорость обратимого нагревания, обусловленного протеканием электрического тока j_x при наличии градиента температуры dT/dx , в однородном материале выражается уравнением

$$\frac{dW_x}{dt} = -\mathcal{S} j_x \frac{dT}{dx}, \quad (6.22)$$

где \mathcal{S} — коэффициент пропорциональности, называемый коэффициентом Томсона. С другой стороны, скорость нагревания единицы объема материала может быть представлена уравнением

$$\frac{dW_x}{dt} = j_x \mathcal{E}_x - \frac{\partial W_x}{\partial x} = j_x \mathcal{E}_x - \frac{\partial W_x}{\partial T} \frac{dT}{dx}. \quad (6.23)$$

Решая уравнение (6.1) относительно \mathcal{E}_x и подставляя \mathcal{E}_x в (6.23), получаем, оставляя только члены, *линейные* по j_x ,

$$\frac{dW_x}{dt} = -\frac{j_x}{e} \left[T \frac{d}{dT} \left(\frac{E_F}{T} \right) - T \frac{d}{dT} \left(\frac{\langle E\tau_e \rangle}{T \langle \tau_e \rangle} \right) \right] \frac{dT}{dx}. \quad (6.24)$$

Сравнивая уравнения (6.24) и (6.22), видим, что

$$\mathcal{S} = -T \frac{d}{dT} \left[\frac{\langle E\tau_e \rangle - E_F \langle \tau_e \rangle}{eT \langle \tau_e \rangle} \right]. \quad (6.25)$$

Если же $\frac{dT}{dx} = 0$, то $W_x \neq 0$ и оказывается пропорциональным j_x .

Этот эффект известен под названием томсоновского нагрева. Он тесно связан с термоэлектрическим эффектом. Для описания этих эффектов обычно используют относительную термо-э. д. с. \mathcal{P}_{ab} (см., например, [10]). Эта величина относится к двум веществам a и b . Пусть цепь (рис. 6.2) содержит два контакта A и B при температурах T_2 и T_1 соответственно. Пусть далее материал b разорван в точке C , причем температура в точке C равна T_0 . Тогда напряжение \mathcal{V} , между разомкнутыми концами цепи равно [10]

$$\mathcal{V}_0 = \int_{T_1}^{T_2} \mathcal{P}_{ab}(T) dT. \quad (6.26)$$

Если $T_2 = T_1 + \Delta T$, то для малых ΔT

$$\mathcal{V}_0 = \mathcal{P}_{ab} \Delta T. \quad (6.27)$$

Если $\mathcal{F}_a, \mathcal{F}_b$ — коэффициенты Томсона для данных материалов [10],

$$\mathcal{F}_a - \mathcal{F}_b = T \frac{d\mathcal{P}_{ab}}{dT}. \quad (6.28)$$

Если для материала b значение $\mathcal{F}_b = 0$ (например, для свинца), то тогда для термо-э. д. с. относительно этого материала можно написать: $\mathcal{P}_{ab} = \mathcal{P}_a$. Иногда \mathcal{P}_a называют абсолютной термо-э. д. с. В этом случае

$$\mathcal{F}_a = T \frac{d\mathcal{P}_{ab}}{dT}. \quad (6.29)$$

Сравнивая (6.29) и (6.25), можно видеть, что

$$\mathcal{P} = \frac{E_F \langle \tau_e \rangle - \langle \tau_e E \rangle}{kT \langle \tau_e \rangle} \left(\frac{k}{e} \right). \quad (6.30)$$

Для полупроводника со скалярной эффективной массой m_e и временем релаксации τ , зависящим от энергии по закону $\tau = aE^{-s}$,

$$\mathcal{P} = - \left\{ \left(\frac{5}{2} - s \right) - \frac{E_F}{kT} \right\} \left(\frac{k}{e} \right). \quad (6.31)$$

Так как при сделанных допущениях $E_F/kT < 0$, то для полупроводника n -типа \mathcal{P} отрицательно, поскольку обычно $s < \frac{5}{2}$ (при $s > \frac{5}{2}$ интеграл для $\langle \tau \rangle$ расходится). Тогда для любых двух веществ a, b

$$\mathcal{P}_{ab} = \mathcal{P}_a - \mathcal{P}_b. \quad (6.32)$$

Если оба вещества a и b — полупроводники n -типа, отличающиеся между собой только концентрациями носителей, то, как известно [см. формулу (4.29)],

$$E_F^{(a)} = -kT \ln \left(\frac{N_c}{n_a} \right), \quad E_F^{(b)} = -kT \ln \left(\frac{N_c}{n_b} \right),$$

так что

$$\mathcal{P}_{ab} = - \frac{k}{e} \ln \left(\frac{n_b}{n_a} \right). \quad (6.33)$$

Формулу (6.31) можно также выразить через электронную концентрацию n в виде ¹⁾

$$\mathcal{P}_e = - \frac{k}{e} \left[\left(\frac{5}{2} - s \right) + \ln \left(\frac{N_c}{n} \right) \right], \quad (6.34)$$

¹⁾ Формула (6.34) впервые выведена Н. Л. Писаренко (см. [28*]). — *Прим.*

где $N_c = 2(2\pi m_e \cdot kT/h^2)^{3/2}$. Поскольку $(k/e) \simeq 100$ мкВ·К⁻¹, то величины термо-э. д. с. в полупроводниках, как правило, порядка 1 мВ·К⁻¹, тогда как в металлах термо-э. д. с. \mathcal{P}_{ab} по порядку величины равна нескольким единицам мкВ·К⁻¹. Поэтому термо-э. д. с. полупроводника по отношению к какому-либо металлу \mathcal{P}_{ab} можно считать равной абсолютной термо-э. д. с. \mathcal{P}_a для данного полупроводника.

В невырожденном полупроводнике p -типа по аналогии с (6.30) имеем

$$\mathcal{P}_h = \frac{\langle E\tau_h \rangle - E'_F \langle \tau_h \rangle}{eT \langle \tau_h \rangle}.$$

Отсюда при $\tau_h = a'E^{-s'}$ получаем

$$\mathcal{P}_h = \frac{k}{e} \left[\left(\frac{5}{2} - s' \right) - \frac{E'_F}{kT} \right], \quad (6.35)$$

$$\mathcal{P}_h = \frac{k}{e} \left[\left(\frac{5}{2} - s' \right) + \frac{E_F}{kT} + \frac{\Delta E}{kT} \right], \quad (6.36)$$

$$\mathcal{P}_h = \frac{k}{e} \left[\left(\frac{5}{2} - s' \right) + \ln \left(\frac{N_v}{\rho} \right) \right], \quad (6.37)$$

где

$$N_v = 2 \left(\frac{2\pi m_h kT}{h^2} \right)^{3/2}.$$

Так как при сделанных допущениях $E'_F/kT < 0$ и поскольку обычно $s' < 5/2$, то \mathcal{P}_h положительно. Таким образом, измерение термо-э. д. с. может служить простым средством определения знака основных носителей заряда в полупроводнике.

В случае смешанной проводимости, объединяя равенства (6.1) и (6.15), можно получить для \mathcal{G}_x выражение

$$\mathcal{G}_x = T \frac{d}{dT} \left[\frac{\frac{n \langle E\tau_e \rangle}{m_e} - \frac{n E'_F \langle \tau_e \rangle}{m_e} + \frac{\rho E'_F \langle \tau_h \rangle}{m_h} - \frac{\rho \langle E\tau_h \rangle}{m_h}}{eT \left(\frac{n \langle \tau_e \rangle}{m_e} + \frac{\rho \langle \tau_h \rangle}{m_h} \right)} \right], \quad (6.38)$$

так что при $\tau_e = aE^{-s}$ и $\tau_h = a'E^{-s'}$

$$\mathcal{P} = \frac{\left(\frac{n}{m_e} \right) (E_F \langle \tau_e \rangle - \langle E\tau_e \rangle) - \left(\frac{\rho}{m_h} \right) (E'_F \langle \tau_h \rangle - \langle E\tau_h \rangle)}{eT \left(\frac{n \langle \tau_e \rangle}{m_e} + \frac{\rho \langle \tau_h \rangle}{m_h} \right)}, \quad (6.39)$$

или

$$\mathcal{P} = -\frac{k}{e} \left[\frac{n\mu_e \left(\frac{5}{2} - s - \frac{E_F}{kT} \right) - \rho\mu_h \left(\frac{5}{2} - s' - \frac{E'_F}{kT} \right)}{n\mu_e + \rho\mu_h} \right]. \quad (6.40)$$

Если обозначить через \mathcal{P}_e и σ_e термо-э. д. с. и электропроводность, обусловленные электронами в отсутствие дырок, а через \mathcal{P}_h и σ_h

соответственно термо-э. д. с. и электропроводность, обусловленные дырками в отсутствие электронов, то в смешанном случае термо-э. д. с. равна

$$\mathcal{P} = \frac{\mathcal{P}_e \sigma_e + \mathcal{P}_h \sigma_h}{\sigma_e + \sigma_h}. \quad (6.41)$$

Таким образом, вклады электронов и дырок в общем выражении для термо-э. д. с. складываются так, как складывались бы напряжения от двух параллельно подключенных генераторов, полные проводимости которых равны проводимости в соответствующих зонах.

Воспользовавшись зависимостями E_F и E'_F соответственно от n и p , можно получить

$$\begin{aligned} \mathcal{P} = & \\ = -\frac{k}{e} & \left[\frac{n\mu_e \left[\left(\frac{5}{2} - s \right) + \ln \left(\frac{N_c}{n} \right) \right] - p\mu_h \left[\left(\frac{5}{2} - s' \right) - \ln \left(\frac{N_v}{p} \right) \right]}{n\mu_e + p\mu_h} \right]. \quad (6.42) \end{aligned}$$

При наличии в зоне более одного типа носителей заряда (как, например, в германии p -типа) соответствующие вклады различных типов носителей сочетаются в общем выражении для \mathcal{P} по указанному выше правилу с учетом знака каждого типа носителей.

В области сильного вырождения на основании (6.10) имеем

$$\begin{aligned} \langle \tau \rangle = \tau + \frac{1}{6} (\pi kT)^2 (\tau E^{3/2})'' E_F E_F^{-3/2} + \dots = \\ = \tau \left[1 + A \left(\frac{3}{2} - s \right) \left(\frac{1}{2} - s \right) + \dots \right], \quad (6.43) \end{aligned}$$

$$\langle E\tau \rangle = \tau E_F \left[1 + A \left(\frac{5}{2} - s \right) \left(\frac{3}{2} - s \right) + \dots \right], \quad (6.44)$$

$$\langle E^2\tau \rangle = \tau E_F^2 \left[1 + A \left(\frac{7}{2} - s \right) \left(\frac{5}{2} - s \right) + \dots \right], \quad (6.45)$$

где

$$A = \frac{\pi^2}{6} \left(\frac{kT}{E_F} \right)^2.$$

Значение τ в правых частях приведенных выше равенств берется при $E = E_F$. В результате получаем

$$\frac{\langle E\tau \rangle - E_F \langle \tau \rangle}{\langle \tau \rangle} = \frac{\pi^2}{3} \left(\frac{3}{2} - s \right) \frac{k^2 T^2}{E_F}. \quad (6.46)$$

Итак, в случае полупроводника n -типа

$$\mathcal{P}_e = -\frac{\pi^2 k T}{e E_F} \left(\frac{1}{2} - \frac{1}{3} s \right). \quad (6.47)$$

Так как для принятых условий $E_F/kT > 0$, то \mathcal{P}_e , как и ранее, остается отрицательным.

Для полупроводника p -типа

$$\mathcal{P}_h = \frac{\pi^2 k T}{e E_F'} \left(\frac{1}{2} - \frac{1}{3} s' \right) \quad (6.48)$$

и, следовательно, $\mathcal{P}_h > 0$, так как, согласно исходному предположению, $E_F'/kT > 0$. В области смешанной проводимости можно по-прежнему пользоваться зависимостью (6.41). Легко видеть, что при наступлении вырождения, когда выполняются условия $E_F'/kT \gg 1$ или $E_F'/kT \gg 1$, численные значения \mathcal{P}_e и \mathcal{P}_h сильно уменьшаются по абсолютной величине. В области частичного вырождения величины $\langle \tau \rangle$, $\langle E\tau \rangle$ и т. д. необходимо вычислять или оценивать на основе точного значения функции f_0 . На основании (6.12) имеем

$$\frac{\langle E\tau_e \rangle}{\langle \tau_e \rangle} = \frac{\left(\frac{5}{2} - s \right) F_{3/2-s}}{\left(\frac{3}{2} - s \right) F_{1/2-s}}, \quad (6.49)$$

$$\frac{\langle E\tau_h \rangle}{\tau_h} = \frac{\left(\frac{5}{2} - s' \right) F_{3/2-s'}}{\left(\frac{3}{2} - s' \right) F_{1/2-s'}}$$

Отсюда для случая электронной проводимости

$$\mathcal{P}_e = \frac{k}{e} \left[\frac{E_F}{kT} - \frac{\left(\frac{5}{2} - s \right) F_{3/2-s}}{kT \left(\frac{3}{2} - s \right) F_{1/2-s}} \right], \quad (6.50)$$

а для дырочной проводимости

$$\mathcal{P}_h = \frac{k}{e} \left[-\frac{E_F'}{kT} + \frac{\left(\frac{5}{2} - s' \right) F_{3/2-s'}}{kT \left(\frac{3}{2} - s' \right) F_{1/2-s'}} \right], \quad (6.51)$$

где интегралы $F_r(E_F)$ определяются формулой (6.13). Эти выражения эквивалентны тем, которые получил Тауц [11]. В области смешанной проводимости по-прежнему справедливо соотношение (6.41), где \mathcal{P}_e и \mathcal{P}_h определяются формулами (6.50) и (6.51).

В области температур, при которых необходимо учитывать одновременно рассеяние на колебаниях атомов решетки и рассеяние на ионах примеси, зависимость времени релаксации τ от энергии нельзя уже представить в виде $\tau = aE^{-s}$, так как она приобретает более сложный вид $\tau = aE^{-1/2} / (1 + cE^{-2})$ (см. подразд. 5.3.2). Тогда интегралы, определяющие $\langle \tau \rangle$ и $\langle E\tau \rangle$, необходимо оценить численно. Значения термо-э. д. с. \mathcal{P} при различных численных значениях s (отношение рассеивающих способностей решетки и ионов примеси) вычислены в работе Джонсон и Ларк-Горвица [12].

Рассмотрим теперь, как влияет на рассмотренные выше эффекты замена скалярной эффективной массы тензором обратной эффективной массы. Выражение (6.1) для плотности тока j_x зависит от эффективной массы через множитель ne/m_e и через энергию Ферми E_F . В первом случае тензор обратной эффективной массы входит так же, как m в электропроводность, которая уже рассматривалась выше, т. е. это приводит к той же комбинации составляющих тензора обратной эффективной массы, которая типична для подвижности. Если в проводимости участвуют носители заряда одного типа, подвижность выпадает из выражения для термо-э. д. с. и не оказывает на нее никакого влияния. В области смешанной проводимости тензорный характер обратной эффективной массы сказывается на термо-э. д. с. через подвижности обоих типов носителей, для которых можно воспользоваться уже установленными ранее зависимостями. Рассмотрим теперь влияние эффективной массы, входящей в определение E_F . В разд. 4.2 было установлено, что в выражении для E_F содержится эффективная масса «плотности состояний» и что при этом необходимо еще учесть общее количество эквивалентных экстремумов (максимумов или минимумов) в зоне. В полупроводнике n -типа с поверхностями постоянной энергии в виде эллипсоидов вращения в случае кубической симметрии имеем

$$N_c = 2m_l^3 m_T M_c h^{-3} (2\pi kT)^{3/2}, \quad (6.52)$$

где m_l и m_T — соответственно продольная и поперечная эффективные массы; M_c — число эквивалентных минимумов в зоне. При учете формулы (6.52) равенство (6.34) остается в силе.

Несмотря на то что приведенные выше зависимости хорошо согласуются с результатами измерений при обычных температурах, они не могут объяснить большого возрастания термо-э. д. с. (по абсолютной величине), которое обнаруживается иногда в чистом веществе при очень низких температурах. Такой низкотемпературный рост $|\mathcal{S}|$ в чистом германии наблюдал Фредерикс [13], а также Джебболл и Халл [14]. Теория этого эффекта разработана Фредериксом [13] и рядом других авторов [15, 16, 6]; они показали, что эффект обусловлен особым видом электрон-фононного взаимодействия, о котором уже упоминалось выше при обсуждении явления теплопроводности. Это явление обычно называют эффектом «увлечения» электронов фононами¹⁾. При наличии градиента температуры число фононов, переносящих энергию со стороны горячего конца образца по направлению к холодному, превышает число фононов, переносящих энергию в обратном направлении. Поэтому рассеяние электронов на фононах перестает быть изотропным и добавочная часть элект-

¹⁾ Еще в 1946 г. Гуревич [22] впервые предсказал эффект «увлечения» электронов фононами и развил теорию этого явления для электронов в металлах.—
Прим. ред.

тронов как бы «увлекается» фононами в сторону холодного конца, порождая добавочное термоэлектрическое поле в дополнение к тому, которое возникает на основе обычного, исследованного выше механизма электронной термо-э. д. с. Эффект «увлечения» приводит к значительному росту (по абсолютной величине) термо-э. д. с. при низких температурах. При более высоких температурах рассеяние фононов на фононах восстанавливает изотропное распределение фононов, и эффект увлечения исчезает. Херринг [15] подверг тщательному анализу современные опытные данные, относящиеся к эффекту увлечения в германии и показал, что они полностью согласуются с теорией. Имеются обзоры [18, 19] современных работ, посвященных исследованию термо-э. д. с. в полупроводниках, содержащие также подробное сравнение теории с опытными данными. Экспериментальные данные по термоэлектрическим свойствам полупроводников типа $A^{III}B_V$ подробно описаны Уре [20].

6.3. Термомагнитные явления

Теперь необходимо перейти к рассмотрению группы явлений, обусловленных совместным действием электрического поля и градиента температуры при наличии внешнего магнитного поля. Предполагается, что магнитное поле направлено по оси z и рассматриваются лишь случаи, когда векторы электрического поля и градиента температуры расположены в плоскости (x, y) . В этих условиях можно наблюдать целую серию эффектов, среди которых выделяется очень важная группа так называемых поперечных эффектов. Например, если электрический ток направлен по оси x , то вдоль оси y наряду с поперечным электрическим полем (полем Холла) может возникнуть градиент температуры. До сих пор при рассмотрении эффекта Холла поперечный температурный эффект нами не учитывался. Иными словами, выводились соотношения для напряжения Холла, которые справедливы при выполнении изотермического условия $dT/dy=0$. Такой эффект обычно называют *изотермическим эффектом Холла*. В адиабатических условиях появляется градиент температуры $dT/dy \neq 0$, и к напряжению Холла добавляется соответствующее термоэлектрическое напряжение. Возникающее в этом случае поперечное напряжение является мерой *адиабатического эффекта Холла*. Ложные термо-э. д. с. при измерении эффекта Холла обычно исключаются путем одновременного изменения направления электрического тока и магнитного поля. Однако таким путем нельзя исключить из адиабатического эффекта Холла термо-э. д. с., связанную с указанным выше эффектом (его обычно называют эффектом Эттингсгаузена). Вместе с тем, как будет видно из дальнейшего, поперечный градиент температуры, возникающий вследствие эффекта Эттингсгаузена, обратно пропорционален пол-

ной теплопроводности. Отношение полной теплопроводности (теплопроводность носителей заряда плюс теплопроводность решетки) к электропроводности значительно больше в полупроводниках, чем в металлах. С другой стороны, величина термо-э. д. с. в полупроводниках также больше, чем термо-э. д. с. в металлах. Несмотря на это, при измерениях адиабатического эффекта Холла в полупроводниках термо-э. д. с., обусловленные эффектом Эттингсгаузена, составляют обычно менее 1% от измеряемого напряжения Холла, так что в полупроводниках отличие изотермического и адиабатического эффектов Холла, как правило, очень мало. При измерениях на переменном токе с частотой, большей 50 Гц, поперечный градиент температуры не успевает устанавливаться и поэтому измерения дают непосредственно изотермическое напряжение Холла.

Можно дать следующую простую физическую интерпретацию рассматриваемого эффекта. Средняя сила, с которой магнитное поле действует на электрон, пропорциональна дрейфовой скорости последнего, так что наиболее быстрые электроны отклоняются магнитным полем сильнее. Поэтому в образце *n*-типа, например, та сторона, где вследствие эффекта Холла имеет место скопление зарядов и которая заряжается отрицательно, должна одновременно быть несколько более горячей. Поскольку знак термо-э. д. с. в веществе *n*-типа отрицательный (так же, как и знак постоянной Холла), то ясно, что напряжение, вызванное эффектом Эттингсгаузена, складывается с напряжением Холла. Очевидно, что аналогичная ситуация имеет место и в полупроводниках *p*-типа.

Если в образце в отсутствие электрического тока существует тепловой поток в направлении оси *x*, то магнитное поле приводит к возникновению поперечной э. д. с., направленной по оси *y*. Это явление представляет собой тепловой аналог эффекта Холла и носит название эффекта Нернста. Природа его такова. При наличии потока тепла по оси *x* в отсутствие электрического тока существует результирующий поток быстрых электронов от горячего конца образца к холодному. Так как отклонение электронов в магнитном поле пропорционально их скорости, то в результате в направлении оси *y* должны возникнуть одновременно поперечная э. д. с. и поперечный градиент температуры. Этот последний эффект (эффект Риги — Ледюка) является тепловым аналогом эффекта Эттингсгаузена.

Для исследования всех перечисленных выше термомагнитных явлений необходимо обобщить уравнение Больцмана путем включения в него магнитной силы Лоренца. Предположим, что вектор индукции имеет компоненты по осям координат (0, 0, *B*). Вместо \mathcal{E} необходимо подставить $\mathcal{E} + [\mathbf{v} \times \mathbf{B}]$ и затем, считая *B* заданным, найти решение соответствующего кинетического уравнения (см., например, книгу Вильсона [21], стр. 210). Имея это решение, можно вывести выражения для электрического и теплового токов аналогично

тому, как мы это делали и ранее. Можно получить решение кинетического уравнения и другим путем, если заметить, что в первом приближении уравнение Больцмана для термически неоднородных систем формально совпадает с уравнением для изотермического случая, если произвести замену $\mathcal{E} \rightarrow \mathcal{E}'$ (в случае электронов) или $\mathcal{E} \rightarrow \mathcal{E}''$ (в случае дырок), где

$$\mathcal{E}' = \mathcal{E} + \left[\frac{T}{e} \frac{\partial}{\partial T} \left(\frac{E_F}{T} \right) + \frac{E}{eT} \right] \frac{dT}{dx}, \quad (6.53)$$

$$\mathcal{E}'' = \mathcal{E} \left[\frac{T}{e} \frac{\partial}{\partial T} \left(\frac{E'_F}{T} \right) + \left(\frac{E'}{eT} \right) \right] \frac{dT}{dx}, \quad (6.54)$$

а E'_F и E' определены в разд. 6.1. Итак, можно воспользоваться найденными ранее решениями уравнений движения для проводимости в присутствии магнитного поля, а также выражениями (5.107) для плотностей электрических токов j_x и j_y . Ограничиваясь рассмотрением лишь линейных по магнитному полю эффектов, мы всюду отбросим члены порядка $\omega_e^2 \tau^2$, где по-прежнему $\omega_e = eB/m_e$. Итак, для электронов получаем

$$j_x = \frac{ne^2}{m_e} [\tau_e \mathcal{E}'_x - \omega_e \tau_e^2 \mathcal{E}'_y], \quad (6.55)$$

$$j_y = \frac{ne^2}{m_e} [\tau_e \mathcal{E}'_y + \omega_e \tau_e^2 \mathcal{E}'_x]. \quad (6.56)$$

Уравнения для плотностей потоков энергии в случае электронов получаются непосредственно, если в (6.55) и (6.56) формально заменить заряд $-e$ энергией E . Тогда получим

$$W_x = -\frac{ne}{m_e} [E \tau_e \mathcal{E}'_x - \omega_e E \tau_e^2 \mathcal{E}'_y], \quad (6.57)$$

$$W_y = -\frac{ne}{m_e} [E \tau_e \mathcal{E}'_y + \omega_e E \tau_e^2 \mathcal{E}'_x]. \quad (6.58)$$

Аналогичные системы уравнений для тока дырок отличаются от формул (6.55), (6.56) только тем, что \mathcal{E}' , n и ω_e заменяются соответственно на \mathcal{E}'' , p и $-\omega_h$, где $\omega_h = eB/m_h$. Кроме того, система для потоков энергии дырок отличается от системы уравнений для дырочного тока лишь заменой e на $E' + \Delta E$. Таким образом, получим

$$j_x = \frac{pe^2}{m_h} [\tau_h \mathcal{E}''_x + \omega_h \tau_h^2 \mathcal{E}''_y], \quad (6.59)$$

$$j_y = \frac{pe^2}{m_h} [\tau_h \mathcal{E}''_y - \omega_h \tau_h^2 \mathcal{E}''_x], \quad (6.60)$$

$$W_x = \frac{pe}{m_h} [E' \tau_h \mathcal{E}''_x + \omega_h E' \tau_h^2 \mathcal{E}''_y] + \frac{pe \Delta E}{m_h} [\tau_h \mathcal{E}''_x + \omega_h \tau_h^2 \mathcal{E}''_y], \quad (6.61)$$

$$W_y = \frac{pe}{m_h} [E' \tau_h \mathcal{E}''_y - \omega_h E' \tau_h^2 \mathcal{E}''_x] + \frac{pe \Delta E}{m_h} [\tau_h \mathcal{E}''_y - \omega_h \tau_h^2 \mathcal{E}''_x]. \quad (6.62)$$

Во всех приведенных выше равенствах необходимо еще провести обычное усреднение по энергиям. При этом дело сводится к вычислению средних величин типа $\langle E^n \tau \rangle$, с которыми мы уже встречались в разд. 6.1 и 6.2, а также величин типа $\langle E^n \tau^2 \rangle$. Все коэффициенты в системах уравнений (6.59) — (6.62) после усреднения можно выразить через указанные величины. Тогда каждую систему равенств можно записать в виде

$$j_x = \sigma_{11} \mathcal{E}_x + \sigma_{12} \mathcal{E}_y + \alpha \frac{\partial T}{\partial x} + \beta \frac{\partial T}{\partial y}, \quad (6.63)$$

$$j_y = -\sigma_{12} \mathcal{E}_x + \sigma_{11} \mathcal{E}_y - \beta \frac{\partial T}{\partial x} + \alpha \frac{\partial T}{\partial y}, \quad (6.64)$$

$$W_x = \gamma \mathcal{E}_x + \delta \mathcal{E}_y + \kappa_{11} \frac{\partial T}{\partial x} + \kappa_{12} \frac{\partial T}{\partial y}, \quad (6.65)$$

$$W_y = -\delta \mathcal{E}_x + \gamma \mathcal{E}_y - \kappa_{12} \frac{\partial T}{\partial x} + \kappa_{11} \frac{\partial T}{\partial y}. \quad (6.66)$$

Некоторые коэффициенты в этих равенствах, как будет видно далее, совпадают между собой (иногда с точностью до обратного знака). Это следует из самой формы записи исходных уравнений. Из сопоставления систем уравнений (6.63) — (6.66) и (6.59) — (6.62) следует, что в случае чисто электронной проводимости

$$\sigma_{11} = \frac{e^2 n \langle \tau_e \rangle}{m_e} = ne\mu_e = \sigma_e, \quad (6.67)$$

$$\sigma_{12} = -\frac{\omega_e e^2 n \langle \tau_e^2 \rangle}{m_e} = -\frac{B (ne\mu_e)^2 r}{en} = \sigma_e^2 R_{He} B, \quad (6.68)$$

$$\alpha = \frac{en}{T m_e} \langle E \tau_e \rangle + \frac{T en}{m_e} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_e \rangle, \quad (6.69)$$

$$\beta = -\frac{en \omega_e}{m_e T} \langle E \tau_e^2 \rangle - \frac{en \omega_e T}{m_e} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_e^2 \rangle, \quad (6.70)$$

$$\gamma = -\frac{en}{m_e} \langle E \tau_e \rangle, \quad (6.71)$$

$$\delta = \frac{en \omega_e}{m_e} \langle E \tau_e^2 \rangle, \quad (6.72)$$

$$\kappa_{11} = -\frac{n}{m_e T} \langle E^2 \tau_e \rangle - \frac{n T}{m_e} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle E \tau_e \rangle - \kappa_L, \quad (6.73)$$

$$\kappa_{12} = \frac{n \omega_e}{m_e T} \langle E^2 \tau_e^2 \rangle + \frac{m \omega_e T}{m_e} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle E \tau_e^2 \rangle. \quad (6.74)$$

Заметим, что здесь учтена также теплопроводность решетки κ_L . Величина r в равенстве для σ_{12} равна $\langle \tau^2 \rangle / \langle \tau \rangle^2$ и совпадает со множителем r в формулах для коэффициента Холла R_H . Если $\tau = aE^{-s}$, то значение r определяется формулой (5.116). При такой зависимости времени релаксации τ от энергии на основании равенств (6.7)

имеем

$$\langle E\tau_e \rangle = \left(\frac{5}{2} - s \right) kT \langle \tau_e \rangle, \quad (6.75)$$

$$\langle E^2\tau_e \rangle = \left(\frac{7}{2} - s \right) \left(\frac{5}{2} - s \right) k^2 T^2 \langle \tau_e \rangle.$$

Аналогично

$$\langle E\tau_e^2 \rangle = \left(\frac{5}{2} - 2s \right) kT \langle \tau_e^2 \rangle = \left(\frac{5}{2} - 2s \right) kT r \langle \tau_e \rangle^2, \quad (6.76)$$

$$\begin{aligned} \langle E^2\tau_e^2 \rangle &= \left(\frac{7}{2} - 2s \right) \left(\frac{5}{2} - 2s \right) k^2 T^2 \langle \tau_e^2 \rangle = \\ &= \left(\frac{7}{2} - 2s \right) \left(\frac{5}{2} - 2s \right) k^2 T^2 r \langle \tau_e \rangle^2. \end{aligned}$$

В случае чисто дырочной проводимости имеем

$$\sigma_{11} = -\frac{e^2 p \langle \tau_h \rangle}{m_h} = p e \mu_h = \sigma_h, \quad (6.77)$$

$$\sigma_{12} = \frac{\omega_h e^2 p \langle \tau_h^2 \rangle}{m_h} = \sigma_h^2 R_{Hh} B, \quad (6.78)$$

$$\alpha = -\frac{ep}{Tm_h} \langle E\tau_h \rangle - \frac{ep \Delta E}{Tm_h} \langle \tau_h \rangle + \frac{Tep}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_h \rangle, \quad (6.79)$$

$$\beta = -\frac{ep\omega_h}{m_h T} \langle E\tau_h^2 \rangle - \frac{ep\omega_h \Delta E}{m_h T} \langle \tau_h^2 \rangle + \frac{ep\omega_h T}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_h^2 \rangle, \quad (6.80)$$

$$\psi = \frac{ep}{m_h} \langle E\tau_h \rangle + \frac{ep \Delta E}{m_h} \langle \tau_h \rangle, \quad (6.81)$$

$$\delta = \frac{ep\omega_h}{m_h} \langle E\tau_h^2 \rangle + \frac{ep\omega_h \Delta E}{m_h} \langle \tau_h^2 \rangle, \quad (6.82)$$

$$\begin{aligned} \kappa_{11} &= -\frac{p}{m_h T} \langle E^2\tau_h \rangle - \frac{2p \Delta E}{m_h T} \langle E\tau_h \rangle - \frac{p \Delta E^2}{m_h T} \langle \tau_h \rangle + \\ &+ \frac{pT}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle E\tau_h \rangle + \frac{pT \Delta E}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_h \rangle - \kappa_L. \end{aligned} \quad (6.83)$$

$$\begin{aligned} \kappa_{12} &= -\frac{p\omega_h}{m_h T} \langle E^2\tau_h^2 \rangle - \frac{2p\omega_h \Delta E}{m_h T} \langle E\tau_h^2 \rangle - \frac{p\omega_h \Delta E^2}{m_h T} \langle \tau_h^2 \rangle + \\ &+ \frac{p\omega_h T}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle E\tau_h^2 \rangle + \frac{p\omega_h T \Delta E}{m_h} \frac{d}{dT} \left(\frac{E_F}{T} \right) \langle \tau_h^2 \rangle. \end{aligned} \quad (6.84)$$

Выписывая эту систему равенств, мы заменили E_F на E_F , согласно соотношению $E_F = -E_F - \Delta E$, и пренебрегли слабой зависимостью ΔE от T . При вычислении средних величин типа $\langle E^2\tau \rangle$ знак штриха у E опущен. Величины $\langle E\tau_h \rangle$ считаются положительными. В случае смешанной проводимости необходимо сложить попарно равенства из систем (6.67) — (6.74) и (6.77) — (6.84), за исключением, конечно, теплопроводности решетки κ_L , которую необходимо учесть *лишь один раз*. Заметим, что теплопроводность κ определяется зависимостью

$$\kappa = -\kappa_{11} + \frac{\alpha\psi}{\sigma_{11}}. \quad (6.85)$$

6.3.1. ЭФФЕКТ ЭТТИНГСГАУЗЕНА

Эффект Эттингсгаузена проявляется как поперечный температурный эффект, возникающий в отсутствие градиента температуры по оси x при наличии поперечного магнитного поля. Итак, полагаем $\partial T/\partial x=0$, а также $j_y=W_y=0$. Исключая \mathcal{E}_y из равенств (6.64) и (6.66), получаем уравнение для $\partial T/\partial y$:

$$\frac{\partial T}{\partial y} = \frac{\left(\delta - \sigma_{12} \frac{\gamma}{\sigma_{11}}\right) \mathcal{E}_x}{\kappa_{11} - \alpha \frac{\gamma}{\sigma_{11}}}. \quad (6.86)$$

Но в первом приближении по B имеем $\mathcal{E}_x = j_x/\sigma_{11}$. Используя также формулы (6.67) — (6.74), получаем

$$B^{-1} \frac{\partial T}{\partial y} = -\frac{\mu_e j_x}{\kappa_e} \left[\frac{\langle E\tau_e^2 \rangle}{\langle \tau_e \rangle^2} - \frac{\langle \tau_e^2 \rangle \langle E\tau_e \rangle}{\langle \tau_e \rangle^3} \right]. \quad (6.87)$$

Коэффициент Эттингсгаузена P определяется из уравнения

$$\frac{\partial T}{\partial y} = -PBj_x, \quad (6.88)$$

так что в случае проводимости n -типа

$$P_e = \frac{\mu_e}{e\kappa} \left[\frac{\langle E\tau_e^2 \rangle}{\langle \tau_e \rangle^2} - \frac{\langle \tau_e^2 \rangle \langle E\tau_e \rangle}{\langle \tau_e \rangle^3} \right]. \quad (6.89)$$

Если $\tau = aE^{-s}$, то отсюда получаем ($r = \langle \tau_e^2 \rangle / \langle \tau_e \rangle^2$)

$$P_e = -\frac{rskT\mu_e}{e\kappa}. \quad (6.90)$$

При рассеянии на колебаниях решетки, когда $s=1/2$, получаем $r=3\pi/8$, так что

$$P_e = -\frac{3\pi kT\mu_e}{16e\kappa}. \quad (6.91)$$

Заметим, что коэффициент P от концентрации n не зависит, если только подвижность μ_e можно считать не зависящей от n . Аналогичным образом можно показать, что в случае проводимости p -типа

$$P_h = \frac{\mu_h}{e\kappa} \left[\frac{\langle E\tau_h^2 \rangle}{\langle \tau_h \rangle^2} - \frac{\langle \tau_h^2 \rangle \langle E\tau_h \rangle}{\langle \tau_h \rangle^3} \right]. \quad (6.92)$$

Необходимо отметить, что в отличие от коэффициента Холла или термо-э. д. с. знак коэффициента Эттингсгаузена не зависит от знака носителей заряда. При $\tau_h = a'E^{-s'}$ из (6.92) следует, что

$$P_h = -\frac{r's'kT\mu_h}{e\kappa}, \quad (6.93)$$

где $r' = \langle \tau_h^2 \rangle / \langle \tau_h \rangle^2$.

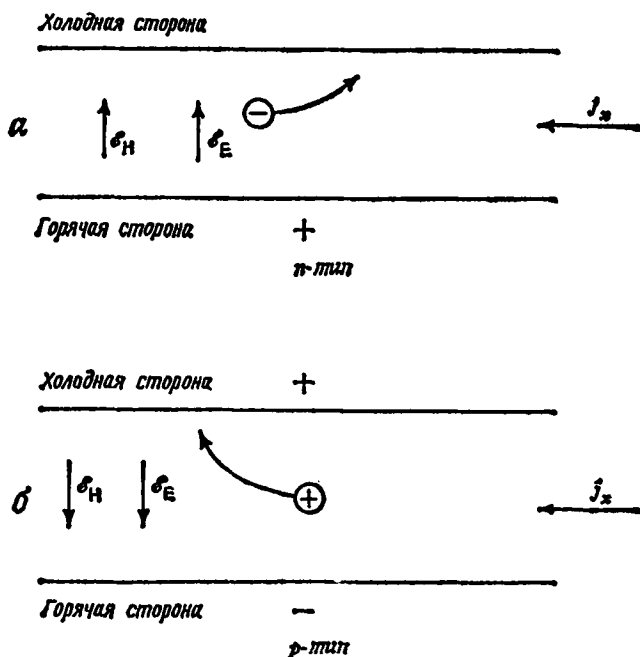


Рис. 6.3. Разности потенциалов, возникающие благодаря изотермическому эффекту Холла и эффекту Эттингсгаузена.

a — в полупроводнике *n*-типа; *b* — в полупроводнике *p*-типа.

Поэтому для рассеяния на колебаниях решетки, когда $s' = \frac{1}{2}$,

$$P_h = - \frac{3\pi k T \mu_n}{16 e \kappa} . \quad (6.94)$$

Возникновение э. д. с. Холла и Эттингсгаузена для электронов и дырок иллюстрируется рис. 6.3, из которого видно, что термо-э. д. с., возникающая благодаря эффекту Эттингсгаузена, совпадает по знаку с э. д. с. Холла, если $s > 0$. В некоторых работах, посвященных эффекту Эттингсгаузена, в знаменателях формул (6.91) и (6.94) под κ понимают только электронную часть теплопроводности. Это ошибочное заключение. Во всех формулах (6.89) — (6.94) под κ необходимо понимать полную теплопроводность образца. Это обстоятельство отметил Патли [4] в связи с измерениями термомагнитных явлений, проведенными им на селениде свинца.

Если $\tau = aE^{-s}$, то, как видно из формулы (6.90), коэффициент Эттингсгаузена пропорционален s . Таким образом, знак коэффициента зависит от природы механизма рассеяния: так, в случае рассеяния на акустических колебаниях решетки ($s = 1/2$) коэффици-

ент отрицателен, а при рассеянии на ионизованных примесях ($s = -3/2$) — положителен. В области смешанного рассеяния знак эффекта зависит от соотношения рассеивающих способностей решетки и примесей. При низких температурах, когда преобладает примесное рассеяние, знак P положителен, тогда как при более высоких температурах основную роль начинает играть рассеяние на колебаниях решетки, и знак P меняется на обратный. Рассмотрим, каков будет знак эффекта при смешанной проводимости.

В этом случае коэффициент P получается путем попарного сложения выражений типа (6.63) и (6.66) для электронов с соответствующими выражениями для дырок и подстановки коэффициентов σ_{ii} , α и т. д. из формул (6.67) — (6.74) и (6.77) — (6.84). Величину поперечного градиента температуры $\partial T/\partial y$ по-прежнему необходимо определить из условий $W_y = j_y = 0$. В общем виде выражение для P имеет довольно сложный вид, и мы его здесь приводить не будем. Для случая когда $s = s' = 1/2$, Патли [4] приводит выражение

$$P = -\frac{3\pi}{16} \frac{kT}{e} \frac{(n^2 \mu_e^3 + p^2 \mu_h^3) - n p \mu_e \mu_h (\mu_e + \mu_h) \left(7 + 2 \frac{\Delta E}{kT}\right)}{\kappa (n \mu_e + p \mu_h)^2}. \quad (6.95)$$

В полупроводниках со смешанной проводимостью в связи с тем, что обычно $\Delta E/kT \gg 1$, в собственной области, когда $n \approx p$, знак P оказывается положительным. Этот вывод согласуется с наблюдениями Патли [4].

6.3.2. ЭФФЕКТ НЕРНСТА

Изотермический эффект Нернста наблюдается в условиях, когда $\partial T/\partial x \neq 0$, $W_x \neq 0$, но $j_x = \partial T/\partial y = j_y = 0$. Из равенства (6.63) и (6.64) можно исключить \mathcal{E}_x , тогда для \mathcal{E}_y получаем

$$\mathcal{E}_y = -\frac{\beta \sigma_{ii} - \alpha \sigma_{12}}{\sigma_{11}^2 + \sigma_{12}^2} \frac{dT}{dx}, \quad (6.96)$$

при этом можно опустить член $\sigma_{12}^2 \sim B^2$. Коэффициент Нернста Q определяется из соотношения

$$\mathcal{E}_y = -QB \frac{dT}{dx}. \quad (6.97)$$

Подставляя соответствующие выражения для σ_{ii} , σ_{12} , α и β в (6.96), получаем

$$Q_c = \frac{\mu_e}{Te} \left[\frac{\langle E \tau_e^2 \rangle}{\langle \tau_e \rangle^2} - \frac{\langle \tau_e^2 \rangle \langle E \tau_e \rangle}{\langle \tau_e \rangle^3} \right]. \quad (6.98)$$

Если $\tau = aE^{-s}$, то

$$Q_c = -\frac{k \mu_e r s}{e} = -\frac{k \mu_{He} s}{e}, \quad (6.99)$$

откуда при $s=1/2$

$$Q_e = -\frac{3\pi k \mu_e}{16e}. \quad (6.99a)$$

В полупроводниках с дырочной проводимостью аналогичным образом можно показать, что

$$Q_h = \frac{\mu_e}{Te} \left[\frac{\langle E\tau_h^2 \rangle}{\langle \tau_h \rangle^2} - \frac{\langle \tau_h^2 \rangle \langle E\tau_h \rangle}{\langle \tau_h \rangle^3} \right]. \quad (6.100)$$

Если $\tau_h = a' E^{-s'}$, то

$$Q_h = -\frac{k\mu_h r' s'}{e} = -\frac{ks' \mu_{Hh}}{e}, \quad (6.101)$$

откуда при $s'=1/2$

$$Q_h = -\frac{3\pi k \mu_h}{16e}. \quad (6.102)$$

На основании формул (6.89) и (6.98) легко показать, что в использованном здесь приближении справедливо равенство

$$Q = \frac{\kappa P}{T}. \quad (6.103)$$

Это соотношение впервые получено Бриджменом [23] на основе классической термодинамики. В качестве термодинамического соотношения оно справедливо в довольно общем смысле. Поэтому значения коэффициента Нернста Q для полупроводников со смешанной проводимостью можно получить на основе соответствующей зависимости для коэффициента Эттингсгаузена.

Итак, коэффициент Нернста по аналогии с коэффициентом Эттингсгаузена должен изменить свой знак при переходе от механизма решеточного рассеяния к механизму рассеяния на ионах примеси. Он может также менять знак благодаря смешанной проводимости. На основании формулы (6.95) легко убедиться, что при $n \approx p$ такое изменение знака Q может произойти, если

$$\frac{b^3 + 1}{b(b+1)} < 7 + \frac{2\Delta E}{kT}, \quad (6.104)$$

где $b = \mu_e / \mu_h$. Неравенство (6.104), как правило, выполняется всегда. Исключение представляют довольно редкие случаи, когда отношение b очень велико.

Явление «увлечения фононов», упомянутое выше в связи с рассмотрением теплопроводности и термо-э. д. с. полупроводников при низких температурах, сказывается и на эффекте Нернста. Теоретически этот вопрос рассмотрен Парротом [24], который показал, что при выполнении определенных условий явление «увлечения» может привести даже к изменению знака коэффициента Нернста, особенно при низких температурах. При этом изменение знака Q

вследствие «увлечения» легко отличить от изменения знака, обусловленного смешанной проводимостью, которое обычно имеет место при более высоких температурах.

Подобно эффекту Холла, эффекты Эттингсгаузена и Нернста можно также наблюдать либо в изотермических, либо в адиабатических условиях. Впрочем, в полупроводниках различие между изотермическими и адиабатическими эффектами невелико. В экспериментальных работах обычно пользуются изотермическим определением коэффициентов Нернста и Эттингсгаузена, хотя на опыте бывает трудно строго соблюсти условия изотермичности.

6.3.3. ЭФФЕКТ РИГИ — ЛЕДЮКА

Эффект Риги — Ледюка возникает в условиях, когда $j_x = j_y = W_y = 0$. При этом поперечный градиент температуры $\partial T / \partial y$ связан с продольным градиентом $\partial T / \partial x$ соотношением

$$\frac{\partial T}{\partial y} = SB \frac{\partial T}{\partial x}, \quad (6.105)$$

где S — коэффициент Риги — Ледюка. Для полупроводника n -типа, исключая \mathcal{E}_y из равенств (6.64) и (6.66) и полагая в линейном приближении по B , что $\mathcal{E}_x = -\alpha (\partial T / \partial x) / \sigma_{11}$, получаем соотношение

$$\kappa \frac{\partial T}{\partial y} = \frac{\partial T}{\partial x} \left\{ -\kappa_{12} + \frac{\beta\gamma + \alpha\delta}{\sigma_{11}} - \frac{\alpha\gamma\sigma_{12}}{\sigma_{11}^2} \right\}. \quad (6.106)$$

Если подставить сюда соответствующие значения для α , β , γ и т. д., то для S_e можно записать

$$S_e = -\frac{n\mu_e^2}{Tex} \left[\frac{\langle E^2\tau_e^2 \rangle}{\langle \tau_e \rangle^2} + \frac{\langle E\tau_e \rangle^2 \langle \tau_e^2 \rangle}{\langle \tau_e \rangle^4} - \frac{2 \langle E\tau_e \rangle \langle E\tau_e \rangle}{\langle \tau_e \rangle^3} \right]. \quad (6.107)$$

При $\tau_e = aE^{-s}$ равенство (6.107) сводится к

$$S_e = -nk^2\mu_e^2 Tr \frac{(5-4s+2s^2)}{2ex}. \quad (6.108)$$

Отсюда в случае рассеяния на колебаниях атомов решетки, когда $s=1/2$, получаем

$$S_e = -21\pi k^2 T \frac{n\mu_e^2}{32ex}, \quad (6.109)$$

где $r=3\pi/8$. Таким же путем можно показать для дырочного полупроводника, что

$$S_h = \frac{p\mu_h^2}{Tex} \left[\frac{\langle E^2\tau_h^2 \rangle}{\langle \tau_h \rangle^2} + \frac{\langle E\tau_h \rangle^2 \langle \tau_h^2 \rangle}{\langle \tau_h \rangle^4} - \frac{2 \langle E\tau_h \rangle \langle E\tau_h \rangle}{\langle \tau_h \rangle^3} \right]. \quad (6.110)$$

При $\tau_h = a'E^{-s'}$

$$S_h = pk^2\mu_h^2 Tr' \frac{(5-4s'+2s'^2)}{2ex}, \quad (6.111)$$

и для $s' = 1/2$, имеем

$$S_h = 21\pi k^2 T \frac{\rho\mu_h}{32e\kappa}, \quad (6.112)$$

где $r' = 3\pi/8$. Знак коэффициента Риги — Ледюка зависит от знака носителей заряда, он отрицателен для электронов и положителен для дырок. Заметим, что значения R_h (коэффициент Холла) и S (коэффициент Риги — Ледюка) отличны, вообще говоря, от нуля при $\tau = \text{const}$, т. е. когда $s=0$, тогда как в тех же условиях коэффициенты P и Q обращаются в нуль.

6.4. Термомагнитные явления в условиях сильного вырождения

В условиях сильного вырождения значения коэффициентов P , Q , S можно получить путем разложения величин $\langle E^n \tau \rangle$, $\langle E^n \tau^2 \rangle$ в ряды с помощью формулы (6.12). Поскольку коэффициенты P , Q в нулевом приближении исчезают, то в соответствующих рядах необходимо сохранить по крайней мере величины порядка $(kT/E_F)^2$. В этом приближении имеем

$$\begin{aligned} \langle \tau^2 \rangle &= \tau^2 \left[1 + A \left(\frac{3}{2} - 2s \right) \left(\frac{1}{2} - 2s \right) \right], \\ \langle \tau^2 E \rangle &= \tau^2 E_F \left[1 + A \left(\frac{5}{2} - 2s \right) \left(\frac{3}{2} - 2s \right) \right], \\ \langle \tau^2 E^2 \rangle &= \tau^2 E_F^2 \left[1 + A \left(\frac{7}{2} - 2s \right) \left(\frac{5}{2} - 2s \right) \right], \end{aligned}$$

где $A = \frac{1}{6} \pi^2 (kT/E_F)^2$. Используя полученные соотношения, для коэффициента Эттингсгаузена P_e получаем выражение

$$P_e = - \frac{\pi^2 s k^2 T^2 \mu_e}{3E_F e \kappa}. \quad (6.113)$$

Аналогично для P_h имеем

$$P_h = - \frac{\pi^2 s' k^2 T^2 \mu_h}{3E_F e \kappa}. \quad (6.114)$$

Соотношение Бриджмена (6.103), связывающее между собой коэффициенты Эттингсгаузена P и Нернста Q , остается, естественно, справедливым и в случае сильного вырождения. Поэтому

$$Q_e = - \frac{\pi^2 s k^2 T \mu_e}{3eE_F}, \quad (6.115)$$

$$Q_h = - \frac{\pi^2 s' k^2 T \mu_h}{3eE_F}. \quad (6.116)$$

Как можно видеть, величины P и Q в случае сильного вырождения меньше значений тех же коэффициентов в отсутствие вырождения примерно в kT/E_F или kT/E_F' раз. Аналогичным образом получаем выражения для коэффициентов Риги — Ледюка

$$S_e = - \frac{\pi^2 \mu_e^2 k^2 T}{3\epsilon k}, \quad (6.117)$$

$$S_h = \frac{\pi^2 \mu_h^2 k^2 T}{3\epsilon k}. \quad (6.118)$$

В литературе иногда приводятся формулы для этих коэффициентов, совпадающие по виду с (6.117) и (6.118), однако вместо κ в знаменателях стоят коэффициенты κ_e или κ_h . Такая замена до некоторой степени оправдана для металлов, где $\kappa \approx \kappa_e$, однако в случае полупроводников это неверно. Здесь должна фигурировать полная теплопроводность κ , включающая наряду с κ_e (или κ_h) также и κ_L . Формулы (6.113) — (6.118) применимы лишь в условиях сильного вырождения, когда $E_F/kT \gg 1$ или $E_F'/kT \gg 1$. В промежуточной области вырождения величины $\langle E^n \tau \rangle$ и $\langle E^n \tau^2 \rangle$ необходимо оценивать, используя соответствующие интегралы, которые определяются равенствами типа (6.49). Подобное описание этих эффектов сделал Патли [25].

6.5. Случай сильных магнитных полей

В случае сильного магнитного поля, когда величиной $\omega\tau$ нельзя пренебречь по сравнению с единицей, следует заменить равенства (6.55) — (6.58) более общими выражениями типа (5.61) и (5.62). Это равносильно умножению каждого члена в правой части формул (6.55) — (6.58) на $1/(1+\omega^2\tau^2)$, причем этот множитель должен быть включен в выражения, которые подлежат усреднению по энергиям. Вместо средних величин типа $\langle E^n \tau \rangle$ появляются средние величины типа $\langle E^n \tau / (1+\omega^2\tau^2) \rangle$ и т. д. Простые соотношения между величинами $\langle E\tau \rangle$, $\langle \tau \rangle$ и т. д., имевшие место, когда $\tau = aE^{-s}$, теперь уже не выполняются, и каждая из этих величин должна рассчитываться в отдельности. В отсутствие вырождения это сводится к вычислению интегралов вида

$$K_{rl} = \int_0^\infty \frac{r^l E^r \exp(-E/kT) dE}{1 + \omega^2 \tau^2}. \quad (6.119)$$

Для каждого частного значения s для оценки величины этих интегралов необходимо воспользоваться численными методами. Необходимо также обратить внимание на то, чтобы члены порядка B^2 , которые мы ранее опускали при выводе полученных выше соотношений для P , Q , S , были теперь включены в соответствующие ра-

венства. Значения коэффициентов для различных эффектов находятся из прежней системы уравнений (6.63) — (6.66), но с более общими выражениями для коэффициентов α , β , γ , δ и т. д. Получающиеся при этом формулы достаточно сложны (см., например, [26, 27]), и мы их здесь воспроизводить не будем.

6.6. Относительные величины гальваномагнитных и термомагнитных эффектов

Если проводимость обусловлена носителями заряда в основном одного сорта, то отношение напряжения V_E , обусловленного эффектом Эттингсгаузена, к изотермическому напряжению Холла V_H можно выразить в виде

$$\frac{V_E}{V_H} = \mathcal{P} \frac{P}{R_H}. \quad (6.120)$$

Если положить $P = qk/e$, то это отношение можно записать в виде

$$\frac{V_E}{V_H} = Aq \frac{\kappa_e}{\kappa}, \quad (6.121)$$

где A — численный коэффициент (при $s=1/2$, $A=4$). Коэффициент Риги — Ледюка также можно представить в виде

$$S = C\mu \frac{\kappa_e}{\kappa}, \quad (6.122)$$

где C — число порядка единицы (при $s=1/2$, $C=21\pi/64$). Вычислим величины различных эффектов для полупроводникового образца, размеры и характеристики которого приводятся ниже.

Длина l	1 см = 10^{-2} м
Ширина w	1 мм = 10^{-3} м
Толщина t	0,1 мм = 10^{-4} м
Ток I	10^{-3} А
Плотность тока j_x	10^4 А·м $^{-2}$
Теплопроводность κ	50 Вт·м $^{-1}$ ·К $^{-1}$
Магнитная индукция B_z	10^3 Гс = 0,1 Т
Электропроводность σ	1 Ом $^{-1}$ ·см $^{-1}$ = 100 Ом $^{-1}$ ·м $^{-1}$
Подвижность электронов	2000 см 2 ·В $^{-1}$ ·с $^{-1}$ = = 0,2 м 2 ·В $^{-1}$ ·с $^{-1}$
$\mu_e \approx R\sigma$	
Термо-э. д. с.	—500 мкВ·К $^{-1}$
Градиент температуры $\partial T/\partial x$	10^3 К·м $^{-1}$

На основе этих данных вычислены приближенные значения следующих величин:

Коэффициент Холла R_H	$2 \cdot 10^{-3} \text{ м}^3 \cdot \text{Кл}^{-1}$
Напряжение Холла V_H	2 мВ
Коэффициент Нернста Q	$-10^{-5} \text{ м}^2 \cdot \text{с}^{-1} \cdot \text{К}^{-1}$
Напряжение Нернста V_N	1 мкВ
Электронная теплопроводность κ_e	$5 \cdot 10^{-4} \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$
κ_e/κ	10^{-5}
Коэффициент Эттингсгаузена P	$-6 \cdot 10^{-5} \text{ К} \cdot \text{м}^3 \cdot \text{Дж}^{-1}$
dT_E/dy	$6 \cdot 10^{-2} \text{ К} \cdot \text{м}^{-1}$
T_E — разность температур в образце, обусловленная эффектом Эттингсгаузена	$6 \cdot 10^{-5} \text{ К}$
Напряжение Эттингсгаузена V_E	0,03 мкВ
V_E/V_H	10^{-5}
Коэффициент Риги — Ледюка S	$2 \cdot 10^{-6} \text{ м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$
$\frac{\partial T_{R-L}}{\partial y} / \frac{\partial T}{\partial x} = SB$	$2 \cdot 10^{-7}$
Разность температур ΔT_{R-L} , обусловленная эффектом Риги — Ледюка	$2 \cdot 10^{-7} \text{ К}$
Напряжение Риги — Ледюка V_{R-L}	10^{-10} В

Как видно, при этих условиях эффекты Эттингсгаузена и Риги — Ледюка дают чрезвычайно малые поперечные э. д. с. Они приблизительно пропорциональны электропроводности σ , и для полупроводника с электропроводностью $\sigma = 10^5 \text{ Ом}^{-1} \cdot \text{м}^{-1}$ и теплопроводностью $\kappa = 5 \text{ Вт} \cdot \text{м}^{-1} \cdot \text{К}^{-1}$ V_E может уже составить примерно 10% от V_H . Из формул (6.121) и (6.122) ясно, почему эти эффекты могут играть значительно более важную роль в металлах, где $\kappa_e \approx \kappa$.

ЛИТЕРАТУРА

1. *Smith R. A.*, Wave Mechanics of Crystalline Solids, 2 nd ed. Chapman and Hall and John Wiley and Sons, New York, 1969, § 10.5, 10.6, 10.7.
2. *McDougall J.*, *Stoner E. C.*, Phil. Trans., A237, 67 (1938).
3. *Rhodes P.*, Proc. Roy. Soc., A204, 396 (1950).
4. *Putley E. H.*, Proc. Phys. Soc., B58, 35 (1955).
5. *Busch G.*, *Schneider M.*, Physica, 20, 1084 (1954).
6. *ter Haar D.*, *Neaves A.*, Advances in Physics (Phil. Mag. Suppl.) 5, 241 (1956).
7. *Иоффе А. В.*, *Иоффе А. Ф.*, ДАН СССР, 98, 757 (1954).
8. *Appel J.*, в сб.: Progress in Semiconductors, ed. Gibson A. F., Aigrain P., Burgess R. E., Heywood, New York, 5, 141 (1960).
9. *Drabble J. R.*, *Goldsmid H. J.*, Thermal Conduction in Semiconductors, Pergamon Press, 1961.
10. *Smith R. A.*, The Physical Principles of Thermodynamics, Chapman and Hall, 1952, p. 84.
11. *Tauc J.*, Phys. Rev., 95, 1394 (1954).

12. *Johnson V. A., Lark-Horowitz K.*, Phys. Rev., **92**, 226 (1953).
13. *Frederikse H. P. R.*, Phys. Rev., **91**, 491 (1953); **92**, 248 (1953).
14. *Geballe T. H., Hull G. W.*, Phys. Rev., **94**, 1134 (1954).
15. *Herring C.*, Phys. Rev., **92**, 857 (1953); **96**, 1163 (1954).
16. *Sondheimer E. H.*, Proc. Roy. Soc., **A234**, 391 (1956).
17. *Иоффе А. Ф.*, Canad. Journ. Phys., **34**, 1342 (1956).
18. *Johnson V. A.*, Photo- and Thermoelectric Effects in Semiconductors, Pergamon Press, 1962.
19. Сборник трудов I и II совещаний по термоэлектричеству / Под ред. Кутасова В. А.— М.-Л., 1963 г.
20. *Ure R. W.*, Semiconductors and Semimetals, Academic Press, **8**, 67 (1972).
21. *Wilson A. H.*, The Theory of Metals, Cambridge University Press, 1954, p. 210.
- 22*. *Гуревич Л. Э.*, ЖЭТФ, **16**, 193 (1946).
23. *Bridgman P. W.*, Thermodynamics of Electrical Phenomena in Metals, Macmillan, New York, 1934.
24. *Parrott J. E.*, Proc. Phys. Soc., **71**, 82 (1958).
25. *Putley E. H.*, The Hall Effect and Related Phenomena, Butterworths, 1960.
26. *Beer A. C., Armstrong J. A., Greenberg I. N.*, Phys. Rev., **107**, 1506 (1957).
27. *Beer A. C.*, в сб.: Solid State Phys. ed. Seitz F., Turnbull D., Suppl. 4, Academic Press, 1963.
- 28*. *Давыдов А. С., Шмушкевич И. М.*, УФН, **24**, 21 (1940).

ДИФФУЗИЯ ЭЛЕКТРОНОВ И ДЫРОК

7.1. Неоднородные полупроводники

В гл. 4 рассматривались однородные полупроводники в состоянии равновесия. Теперь необходимо распространить этот анализ на случай неоднородных образцов. Подобные образцы, главным образом в виде переходов между частями кристалла с разными концентрациями примесей, играют важную роль в техническом применении полупроводников. В неоднородном полупроводнике концентрации электронов n и дырок p меняются от точки к точке и, следовательно, являются функциями пространственных координат (x, y, z) . В условиях равновесия можно определить положение уровня Ферми, который будет постоянным во всем образце. Это вытекает из соотношения между энергией, соответствующей уровню Ферми, и термодинамическим потенциалом (см. разд. 4.1); условием равновесия при неизменном давлении является постоянство термодинамического потенциала.

Так как концентрации электронов и дырок меняются с координатами (x, y, z) , должен возникнуть диффузионный ток электронов и дырок. Число электронов n_x^d , пересекающих за единицу времени единичную площадку в направлении x , связано с градиентом концентрации dn/dx уравнением вида

$$n_x^d = -D_e \frac{dn}{dx}. \quad (7.1)$$

Величину D_e называют коэффициентом диффузии электронов. Следовательно, плотность электронного тока j_{ex}^d , обусловленного диффузией, определяется уравнением

$$j_{ex}^d = eD_e \frac{dn}{dx}. \quad (7.2)$$

Аналогичным образом можно определить и коэффициент диффузии дырок D_h ; число дырок p_x^d , пересекающих за единицу времени еди-

ничную площадку в направлении x , дается уравнением

$$p_x^d = -D_h \frac{dp}{dx}. \quad (7.3)$$

Таким образом, плотность диффузионного дырочного тока j_{hx}^d определяется уравнением

$$j_{hx}^d = -eD_h \frac{dp}{dx}. \quad (7.4)$$

В состоянии равновесия электронный и дырочный токи, очевидно, обращаются в нуль; следовательно, в неоднородном полупроводнике должно существовать статическое электрическое поле, препятствующее протеканию диффузионного тока. Расчет величины такого поля в общем случае оказывается весьма сложным, но для ряда частных условий его можно легко осуществить; этот вопрос будет рассмотрен позднее. Уравнения (7.1) — (7.4) можно представить в векторной форме, описывающей зависимость диффузионного тока от градиента концентрации в трехмерном пространстве. Чтобы найти плотность полного тока электронов j_e , необходимо учесть еще и ток, обусловленный электрическим полем. В результате получим

$$j_e = ne\mu_e \mathcal{E} + eD_e \nabla n \quad (7.5)$$

и аналогично для плотности дырочного тока

$$j_h = p\mu_h \mathcal{E} - eD_h \nabla p. \quad (7.6)$$

Таким образом, плотность полного тока j определяется уравнением

$$j = e(n\mu_e + p\mu_h) \mathcal{E} + e(D_e \nabla n - D_h \nabla p). \quad (7.7)$$

Теперь интересно выяснить, почему диффузионный ток существует в полупроводниках и несуществен в металлах. Причина состоит в том, что концентрации электронов и дырок в полупроводнике могут меняться в широких пределах при постоянной величине суммарной концентрации зарядов. Как будет видно из дальнейшего, любое сильное изменение плотности зарядов должно сопровождаться появлением сильных электрических полей — следствие из уравнения Пуассона. С другой стороны, в металлах концентрация электронов практически постоянна.

7.2. Соотношение Эйнштейна

В состоянии равновесия плотность электронного тока должна быть равна нулю. Направим ось x вдоль электрического поля. Из (7.5) получим

$$n\mu_e \mathcal{E}_x = -D_e \frac{dn}{dx}. \quad (7.8)$$

Если ψ — электростатический потенциал, то

$$\mathcal{E}_x = -\frac{d\psi}{dx}, \quad (7.9)$$

а потенциальная энергия электрона равна $-e\psi$. В отсутствие вырождения концентрация электронов n будет подчиняться классической статистике. Поэтому

$$n = \text{const} \cdot \exp\left(\frac{e\psi}{kT}\right), \quad (7.10)$$

так что

$$\frac{1}{n} \frac{dn}{dx} = \frac{e}{kT} \frac{d\psi}{dx}. \quad (7.11)$$

Подставляя в уравнение (7.8) \mathcal{E}_x и dn/dx , выраженные через ψ , получаем

$$D_e = \frac{kT}{e} \mu_e. \quad (7.12)$$

Это уравнение, связывающее коэффициент диффузии D_e с дрейфовой подвижностью μ_e , известно под названием соотношения Эйнштейна. Оно справедливо в самом общем случае для любых заряженных частиц, диффундирующих в электрическом поле. Для дырок получим соотношение

$$D_h = \frac{kT}{e} \mu_h. \quad (7.13)$$

Это следует по аналогии с предыдущим из условия, что в состоянии равновесия результирующий дырочный ток также должен быть равен нулю. Коэффициенты диффузии D_h и D_e имеют размерность $[\text{м}^2 \cdot \text{с}^{-1}]$, если величина μ выражена в $[\text{м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}]$, или $[\text{см}^2 \cdot \text{с}^{-1}]$, если μ выражена в $[\text{см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}]$. Например, если $\mu = 10^3 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, то $D = 25 \text{ см}^2 \cdot \text{с}^{-1}$. Уравнение (7.12) справедливо только в отсутствие вырождения, однако его можно легко обобщить и на случай вырождения.

Вид энергетических зон неоднородного полупроводника показан на рис. 7.1. Теперь уже нельзя за нуль энергии принимать энергию,

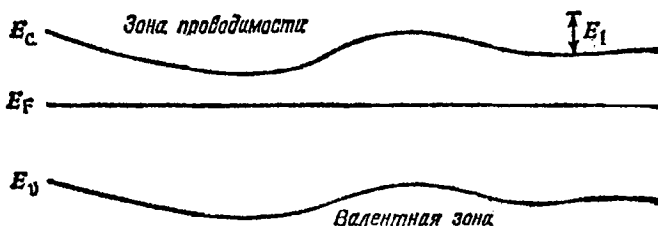


Рис. 7.1. Энергетические зоны неоднородного полупроводника.

соответствующую дну зоны проводимости, поскольку энергетическое расстояние от дна зоны до постоянного уровня Ферми меняется. Пусть E_c — самый низкий энергетический уровень в зоне проводимости, а E_v — самый высокий энергетический уровень в валентной зоне. Они будут зависеть от их места положения в неоднородном образце. Далее, имеем

$$E_c - E_v = \Delta E, \quad (7.14)$$

где ΔE — ширина запрещенной зоны, которую будем предполагать не зависящей от содержаний примесей, т. е. одинаковой для всего материала. Пусть E — энергия какого-либо электронного уровня, а E_1 — его энергия относительно дна зоны проводимости, т. е. $E = E_c + E_1$. Теперь функцию распределения f_0 можно представить в виде

$$f_0 = \frac{1}{\exp\left(\frac{E - E_F}{kT}\right) + 1} = \frac{1}{\exp\left(\frac{E_1 + E_c - E_F}{kT}\right) + 1}. \quad (7.15)$$

Тогда имеем

$$n = \int_{E_c}^{\infty} N(E - E_c) f_0 dE = \int_0^{\infty} N(E_1) f_0 dE_1. \quad (7.16)$$

Очевидно, что E_1 можно, как и раньше, рассматривать как кинетическую энергию электрона, потенциальная энергия которого может быть принята с точностью до произвольной постоянной, равной E_c . Таким образом, E_c связано с электростатическим потенциалом ψ уравнением

$$E_c = \text{const} - e\psi. \quad (7.17)$$

Следовательно, электрическое поле \mathcal{E}_x в направлении x определяется выражением

$$\mathcal{E}_x = -\frac{d\psi}{dx} = \frac{1}{e} \frac{dE_c}{dx}. \quad (7.18)$$

Далее имеем

$$\frac{\partial n}{\partial E_c} = \int_0^{\infty} N(E_1) \frac{\partial f_0}{\partial E_1} dE_1, \quad (7.19)$$

так что, используя уравнение (7.18), получаем

$$\frac{dn}{dx} = \frac{\partial n}{\partial E_c} \frac{dE_c}{dx} = e\mathcal{E}_x \int_0^{\infty} N(E_1) \frac{\partial f_0}{\partial E_1} dE_1. \quad (7.20)$$

Подставляя в выражение (7.8), получаем

$$eD_e = -\mu_e \frac{\int_0^{\infty} N(E_1) f_0 dE_1}{\int_0^{\infty} N(E_1) \frac{\partial f_0}{\partial E_1} dE_1}, \quad (7.21)$$

что, очевидно, приводит к соотношению Эйнштейна (7.12), если f_0 имеет вид $C \exp(-E_1/kT)$, справедливый для невырожденного случая. Раньше было показано, что довольно часто $N(E_1)$ можно представить в виде $N(E_1) = AE_1^{1/2}$. Интегрирование по частям знаменателя в уравнении (7.21) показывает, что это выражение можно привести к виду

$$eD_e = \frac{(2\mu_e kT) F_{1/2} [(E_F - E_c)/kT]}{F_{-1/2} [(E_F - E_c)/kT]}, \quad (7.22)$$

где $F_r(x)$ — функция, определенная для различных значений r в разд. 4.2.

7.3. Отклонения от теплового равновесия

Мы уже видели, что когда к полупроводнику приложено электрическое поле, тепловое равновесие нарушается, и для описания распределения электронов по различным энергетическим состояниям необходимо ввести другую функцию. В этом случае нельзя воспользоваться такой схемой энергетических уровней с постоянным уровнем Ферми, как на рис. 7.1. Если внешнее электрическое поле \mathcal{E}_x приложено к однородному полупроводнику, то изменения f_0 могут быть очень малыми. Можно показать, что в невырожденном случае [1]

$$f = f_0 \left(1 - \frac{ev_x \mathcal{E}_x \tau}{kT} \right), \quad (7.23)$$

и напряжение $v_x \tau \mathcal{E}_x$ при $T=300$ К оказывается малым по сравнению с kT/e , пока \mathcal{E}_x не станет порядка $10^3 \text{ В} \cdot \text{см}^{-1}$. После интегрирования по всем значениям v_x в любом случае окажется, что величина n не отличается от своего равновесного значения. Однако достаточно большие изменения n и p можно получить, например облучая полупроводник светом, длина волны которого настолько мала, что становятся возможными междузонные переходы (оптическая инжекция), или подавая напряжение соответствующего знака на инжектирующие электроды (см. подразд. 7.5.3). Пусть полная концентрация электронов n равна $n_0 + \Delta n$, где n_0 — равновесная концентрация электронов. Аналогично пусть $p = p_0 +$

$\pm \Delta p$, где p_0 — равновесная концентрация дырок. Если нарушается условие $\Delta n = \Delta p$, то возникает объемный заряд ρ , определяемый соотношением

$$\rho = e(\Delta p - \Delta n). \quad (7.24)$$

Один из наиболее важных результатов, справедливость которого сейчас будет доказана, состоит в том, что в отсутствие сильного электрического поля Δn должно быть близко по величине к Δp . Сильное поле может наблюдаться только вблизи поверхности или в области, где происходит очень резкое изменение концентрации примесей, приводящее к образованию большого градиента концентрации носителей заряда. Из уравнения Пуассона следует

$$\operatorname{div} \mathfrak{E} = \frac{\rho}{\epsilon \epsilon_0} = \frac{e(\Delta p - \Delta n)}{\epsilon \epsilon_0}, \quad (7.25)$$

где ϵ — диэлектрическая проницаемость. Например, пусть $\rho = 10^{12} \text{ см}^{-3} = 10^{18} \text{ м}^{-3}$, $(\Delta p - \Delta n) = 0,01 \rho$, $\epsilon = 10$ и $\epsilon_0 \approx 10^{-11} \text{ Ф} \cdot \text{м}^{-1}$. Если теперь при $x=0$ поле $\mathfrak{E}_x = 0$, то $\mathfrak{E}_x(x) = 10^{17} \cdot x \text{ В} \cdot \text{м}^{-1}$, и тогда на расстоянии в 1 см получим $\mathfrak{E}_x = 10^8 \text{ В} \cdot \text{м}^{-1}$.

Можно также рассмотреть изменения во времени ρ в объеме вещества (пренебрегая пока процессом диффузии). Уравнение непрерывности электрического тока дает

$$\operatorname{div} \mathbf{j} = \operatorname{div} \sigma \mathfrak{E} = -\frac{\partial \rho}{\partial t} = e \frac{\partial}{\partial t} (\Delta n - \Delta p). \quad (7.26)$$

Используя уравнение (7.25), получаем

$$\frac{\partial}{\partial t} (\Delta n - \Delta p) = -\frac{\sigma}{\epsilon \epsilon_0} (\Delta n - \Delta p). \quad (7.27)$$

Решением этого уравнения будет

$$(\Delta n - \Delta p) = (\Delta n - \Delta p)_{t=0} \exp\left(-\frac{t}{\tau_0}\right), \quad (7.28)$$

где $\tau_0 = \epsilon \epsilon_0 / \sigma$. Если $\sigma = 100 \text{ Ом}^{-1} \cdot \text{м}^{-1}$, то $\tau_0 \approx 10^{-12} \text{ с}$, и, следовательно, всякий объемный заряд, возникший в полупроводнике, очень скоро исчезнет. Это, казалось бы, означает, что *ни при каких условиях* объемный заряд существовать не может. Более точные вычисления, в которых принята во внимание диффузия носителей заряда и которые будут рассмотрены ниже, показывают, однако, что могут существовать *очень слабые* объемные заряды и что они могут приводить к возникновению довольно сильных внутренних электрических полей. Предыдущий анализ неоднородных полупроводников показывает, что подобные объемные заряды и поля существуют везде, где меняется концентрация доноров и акцепторов. Пока мы будем предполагать, что $\Delta n = \Delta p$.

7.3.1. КВАЗИУРОВНИ ФЕРМИ

Если концентрация электронов n и концентрация дырок p отклоняются от равновесных значений n_0 и p_0 , то n и p уже нельзя выразить через одну величину — энергию Ферми E_F . В разд. 4.3 было показано, что в равновесных условиях при температуре T и в отсутствие вырождения n и p можно записать в виде

$$n = N_c \exp\left(\frac{E_F}{kT}\right), \quad (7.29)$$

$$p = N_v \exp\left[-\frac{(\Delta E + E_F)}{kT}\right]. \quad (7.29a)$$

Мы можем, однако, ввести два новых энергетических параметра E_{F_e} и E_{F_h} , которые позволяют нам выразить n и p в том же виде, что и уравнения (7.29), (7.29a), а именно:

$$n = N_c \exp\left(\frac{E_{F_e}}{kT}\right), \quad (7.30)$$

$$p = N_v \exp\left[-\frac{(\Delta E + E_{F_h})}{kT}\right]. \quad (7.30a)$$

Величины E_{F_e} и E_{F_h} называют соответственно квазиуровнями Ферми для электронов и дырок. В состоянии теплового равновесия оба они совпадают с E_F , но в общем случае один из них расположен выше, а другой — ниже уровня Ферми. Это показано на рис. 7.2 для обычного случая, когда $n > n_0$, а $p > p_0$; при этом уровень E_{F_e} расположен выше E_F , а E_{F_h} — ниже E_F . Это легко видеть, если записать E_{F_e} и E_{F_h} в виде

$$E_{F_e} = kT \ln\left(\frac{n}{N_c}\right), \quad (7.31)$$

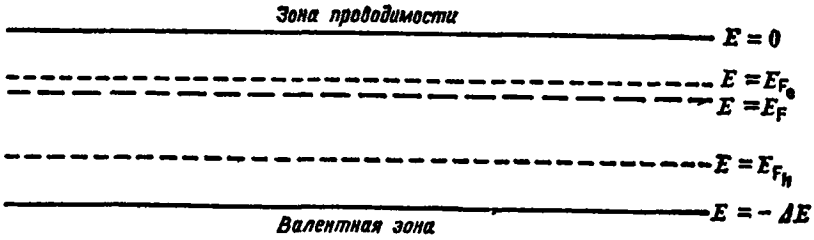
$$E_{F_h} = -\Delta E - kT \ln\left(\frac{p}{N_v}\right). \quad (7.31a)$$

Если $n_0 \gg p_0$, $\Delta n = \Delta p \ll n_0$, но в то же время $\Delta n = \Delta p \gg p_0$ (это довольно общий случай), то можно видеть, что E_{F_e} сдвигается относительно E_F очень немного, но E_{F_h} может сдвинуться весьма существенно.

Простое соотношение, подобное $n_0 p_0 = n_i^2$, теперь записать нельзя, но можно отметить, что

$$np = n_i^2 \exp\left[\frac{E_{F_e} - E_{F_h}}{kT}\right], \quad (7.32)$$

так что разность $E_{F_e} - E_{F_h}$ служит мерой отклонения произведения np от его равновесного значения.

Рис. 7.2. Квазиуровни Ферми, $\Delta n = \Delta p > 0$.

Как можно видеть из уравнений (7.31) и (7.31а), в случае обеднения носителями, когда $n < n_0$ и $p < p_0$, положения квазиуровней Ферми относительно уровня Ферми становятся обратными.

В случае, который нельзя считать невырожденным, мы можем все же ввести квазиуровни Ферми с помощью уравнений, подобных уравнениям (4.19) и (4.20). В этом случае, однако, они не так удобны для вычислений и использование их не столь наглядно. Квазиуровни Ферми ничего не говорят о распределении электронов по уровням энергии в зоне проводимости или дырок в валентной зоне и поэтому не очень полезны при сильном отклонении этого распределения от равновесного.

7.4. Электронно-дырочная рекомбинация

Теперь необходимо рассмотреть поведение полупроводников, в которых возникли отклонения от равновесных концентраций носителей заряда, равные Δn и Δp . Ясно, что если причина, порождающая эти отклонения, устранена, то концентрации через некоторое время вновь достигнут своих равновесных значений. Пусть Δn_0 и Δp_0 — величины Δn и Δp при $t=0$ ($\Delta n_0 = \Delta p_0$), и пусть \mathcal{R}_0 — темп (постоянный) генерации электронно-дырочных пар. Тогда темп изменения n и p можно выразить уравнением

$$\frac{dn}{dt} = \frac{d\Delta n}{dt} = \mathcal{R}_0 - \frac{n}{\tau_n}, \quad (7.33)$$

где τ_n — величина, которая может зависеть от p . В состоянии равновесия $d\Delta n/dt = 0$ и $n = n_0$, так что $\mathcal{R}_0 = n_0/\tau_{n_0}$, где τ_{n_0} — значение τ , когда $n = n_0$. Следовательно, уравнение (7.33) можно представить в виде

$$\frac{d\Delta n}{dt} = \frac{n_0}{\tau_{n_0}} - \frac{n}{\tau_n}. \quad (7.34)$$

Аналогично получим

$$\frac{d\Delta p}{dt} = \frac{p_0}{\tau_{p_0}} - \frac{p}{\tau_p}. \quad (7.34a)$$

Большой практический интерес представляет случай $n \gg p$. При этом условии p называют концентрацией неосновных носителей заряда. Изменения n будут малы, если $|\Delta p| \ll p$, так как $|\Delta n| = |\Delta p| \ll p$. В этом случае время жизни τ не будет зависеть существенным образом от концентрации n ; его можно положить равным константе, и тогда

$$\frac{d\Delta p}{dt} = -\frac{\Delta p}{\tau_p} \quad (7.35)$$

при $\mathcal{R}_0 = p_0/\tau_p$. Решением уравнения (7.35) будет

$$\Delta p = \Delta p_0 \exp\left(-\frac{t}{\tau_p}\right). \quad (7.36)$$

Мы видим, что концентрация избыточных дырок уменьшается в $e \approx 2,7$ раза за время τ_p , которое по этой причине называют временем жизни неосновных носителей заряда — дырок. Заметим, что наряду с (7.36) должно иметь место равенство

$$\Delta n = \Delta n_0 \exp\left(-\frac{t}{\tau_p}\right), \quad (7.36a)$$

поэтому одновременно τ_p является и средним временем жизни электронно-дырочных пар. В равновесии темп генерации электронно-дырочных пар \mathcal{R}_0 равен p_0/τ_p , но если $n \approx p$, то положение оказывается более сложным. Этот случай будет рассмотрен позднее. Значения τ_p меняются в пределах от 1—2 мс (для очень чистого Ge или Si) до величин, меньших 10^{-8} с. Аналогичным образом можно определить величину τ_n , если $p \gg n$. Необходимо хорошо усвоить смысл параметров τ_n и τ_p и времен релаксации τ_e и τ_h для электронов и дырок, а также смысл времени диэлектрической релаксации τ_0 , и не смешивать их друг с другом. Приведенные выше обозначения применены, чтобы уменьшить опасность такой путаницы. Порядки величин этих параметров весьма различны: при комнатной температуре τ_e , τ_h и τ_0 оказываются порядка 10^{-12} с, в то время как τ_n и τ_p обычно бывают значительно большими.

7.5. Диффузия и проводимость в примесных полупроводниках ($n \gg p$ или $p \gg n$)

Теперь мы можем рассмотреть движение электронов и дырок под действием электрического и магнитного полей при условии, что n и p сильно отличаются от своих равновесных величин. Прежде всего рассмотрим одномерное движение в направлении x , обусловленное электрическими полями и градиентами концентрации, направленными вдоль оси x . Предположим, что мы имеем сильно

примесный полупроводник n -типа, т. е. что $n \gg p$. Ход рассуждений для вещества p -типа будет точно таким же. Допустим также, что образец однороден и что объемного заряда нет, следовательно, $\Delta n = \Delta p$. Ниже будет показано, что условий $n \gg p$ или $p \gg n$ вполне достаточно, чтобы считать это приближение хорошим.

7.5.1. УРАВНЕНИЕ НЕПРЕРЫВНОСТИ

Уравнение непрерывности в применении к теории электрического тока выражает условие отсутствия где-либо накопления зарядов. Обычно оно записывается в форме

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} \mathbf{j}. \quad (7.37)$$

Теперь следует учесть тот факт, что нигде нет накопления электронов и дырок, так что для описания концентраций электронов и дырок можно пользоваться уравнениями, подобными (7.37). Необходимо также включить член, учитывающий генерацию и рекомбинацию электронно-дырочных пар. Для нашей цели достаточно одного лишь уравнения для дырок, поскольку если известно p , то при помощи равенства

$$n = n_0 + p - p_0, \quad (7.38)$$

которое следует из условия $\Delta n = \Delta p$, легко найти также и n . Общее уравнение непрерывности для дырок имеет вид

$$\frac{\partial p}{\partial t} = -\frac{p - p_0}{\tau_p} - \frac{1}{e} \operatorname{div} \mathbf{j}_h, \quad (7.39)$$

где \mathbf{j}_h — плотность дырочного тока, определяемая уравнением (7.6). Подставляя выражение для \mathbf{j}_h и ограничиваясь лишь однородным случаем, получаем

$$\frac{\partial p}{\partial t} = -\frac{p - p_0}{\tau_p} - p\mu_h \frac{\partial \mathcal{E}}{\partial x} - \mathcal{E}\mu_h \frac{\partial p}{\partial x} + D_h \frac{\partial^2 p}{\partial x^2}. \quad (7.40)$$

Аналогичное уравнение можно использовать для нахождения темпа изменения n , но следует проявить некоторую осторожность при написании члена, связанного с рекомбинацией. В этом случае имеем

$$\frac{\partial n}{\partial t} = \frac{p_0}{\tau_p} - \frac{n}{\tau_n} + n\mu_e \frac{\partial \mathcal{E}}{\partial x} + \mathcal{E}\mu_e \frac{\partial n}{\partial x} + D_e \frac{\partial^2 n}{\partial x^2}. \quad (7.40a)$$

Хотя в данной главе это уравнение использоваться не будет, заметим, однако, что, если подставить в него соотношение (7.38), оно может показаться несовместимым с уравнением (7.40). Причиной этого является сделанное здесь допущение $\Delta n = \Delta p$. При $n \gg p$ это допущение хорошо выполняется для уравнения (7.40) и не справедливо для уравнения (7.40a). Решим теперь уравнение (7.40) при некоторых ограничивающих условиях.

7.5.2. СЛАБОЕ ЭЛЕКТРИЧЕСКОЕ ПОЛЕ

Рассмотрим сначала случай столь слабого электрического поля, что током проводимости неосновных носителей заряда можно пренебречь; это соответствует приближению чисто диффузионного тока, так что уравнение (7.40) примет вид

$$\frac{\partial \Delta p}{\partial t} = -\frac{\Delta p}{\tau_p} + D_n \frac{\partial^2 \Delta p}{\partial x^2}. \quad (7.41)$$

Рассмотрим стационарный случай, когда при $x=0$ имеется источник дырок, благодаря которому в этой точке $\Delta p = \Delta p_0$. Тогда $\partial \Delta p / \partial t = 0$, и, следовательно,

$$D_n \frac{d^2 \Delta p}{dx^2} - \frac{1}{\tau_p} \Delta p = 0. \quad (7.42)$$

Предположим, что мы имеем полубесконечную пластинку полупроводникового материала, простирающуюся от $x=0$ до $x=\infty$. Решением в этом случае будет

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_p}\right), \quad (7.43)$$

где $L_p^2 = D_n \tau_p$; таким образом, равновесные условия достигаются на расстоянии нескольких L_p . Величина L_p называется диффузионной длиной неосновных носителей заряда. Оценим величину L_p для типичных условий, например для значений $\mu_n = 10^3 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, $D_n = 25 \text{ см}^2 \cdot \text{с}^{-1}$ и $\tau_p = 10^{-4} \text{ с}$ получим $L_p = 0,05 \text{ см}$.

Стационарное состояние может поддерживаться лишь благодаря потоку дырок, входящему в полупроводник при $x=0$; плотность дырочного тока j_{h_0} при $x=0$ определяется уравнением

$$j_{h_0} = -D_n e \left(\frac{d \Delta p}{dx} \right)_{x=0} = \frac{e D_n \Delta p_0}{L_p}, \quad (7.44)$$

в то время как плотность полного тока j подчиняется уравнению (7.37) и остается постоянной. Введем величину γ , называемую коэффициентом инжекции. Тогда дырочный ток при $x=0$ и γ связаны соотношением

$$j_{h_0} = \gamma j, \quad (7.45)$$

из которого следует, что электронный ток при $x=0$, j_{e_0} , дается формулой

$$j_{e_0} = (1 - \gamma) j. \quad (7.45a)$$

Если $\gamma=1$, то весь ток при $x=0$ переносится дырками, тогда как при $x \gg L_p$, очевидно, весь ток практически переносится электронами. Электронный ток при $x=0$ может обратиться в нуль лишь при наличии в этой точке электрического поля, но, как мы увидим далее,

оно может быть очень малым. Из уравнений (7.44) и (7.45) имеем

$$\Delta p_0 = \frac{\gamma j L_p}{e D_n}, \quad (7.46)$$

$$\Delta p_0 = \frac{\gamma j}{e} \left(\frac{\tau_p}{D_n} \right)^{1/2}. \quad (7.46a)$$

Отсюда видно, что если $j > 0$, то полный ток при $x=0$ направлен *внутрь* образца, т. е. $\Delta p_0 > 0$; следовательно, концентрация неосновных носителей заряда может возрасти, только если $\gamma > 0$ и если общий ток при $x=0$ имеет и электронную, и дырочную компоненты. Правда, даже при $\gamma=0$ будет протекать очень слабый дырочный ток проводимости, который здесь не учитывается. С другой стороны, если $j < 0$, то и $\Delta p_0 < 0$, и концентрация неосновных носителей заряда падает. Если коэффициент γ не мал, то дырочный ток при $x=0$ будет сравним по величине с электронным, однако на расстоянии нескольких L_p практически весь ток, как мы увидим ниже, будет создаваться электронами. Так как

$$j_n = \gamma j \exp\left(-\frac{x}{L_p}\right), \quad (7.47)$$

то

$$j_e = j - j_n = j \left[1 - \gamma \exp\left(-\frac{x}{L_p}\right) \right]. \quad (7.48)$$

Это показано на рис. 7.3. Теперь можно вычислить электрическое поле \mathcal{E} . Так как $\Delta n = \Delta p$, то диффузионный ток, обусловленный электронами, равен $e D_e d\Delta p/dx$. Далее, поскольку $D_e/D_n = \mu_e/\mu_n = b$, этот ток равен $-bj_n$. Электронный ток проводимости приблизительно равен $en_0\mu_e\mathcal{E}$, где \mathcal{E} — электрическое поле, ибо мы предполагаем, что $\Delta p \ll n$; вместе с тем этот ток приблизительно равен $\sigma\mathcal{E}$, так как $n \gg p$. Тогда полный электронный ток равен

$$j_e = \sigma\mathcal{E} - bj_n \exp\left(-\frac{x}{L_p}\right), \quad (7.49)$$

так что

$$j = \sigma\mathcal{E} - (b-1)\gamma j \exp\left(-\frac{x}{L_p}\right) \quad (7.50)$$

и

$$\mathcal{E} = \mathcal{E}_\infty \left[1 + (b-1)\gamma \exp\left(-\frac{x}{L_p}\right) \right], \quad (7.51)$$

где $\mathcal{E}_\infty = j/\sigma$ — величина электрического поля при $x \gg L_p$; для $x=0$ имеем $\mathcal{E} = \mathcal{E}_0$, где

$$\mathcal{E}_0 = \mathcal{E}_\infty [1 + (b-1)\gamma]. \quad (7.52)$$

Зависимость \mathcal{E} от x также показана на рис. 7.3. Если, как обычно, $b > 1$, то наибольшее электрическое поле будет при $x=0$; если же $\gamma=1$, то $\mathcal{E}_0/\mathcal{E}_\infty = b$, и если $\gamma=0$, то \mathcal{E} постоянно и равно \mathcal{E}_∞ .

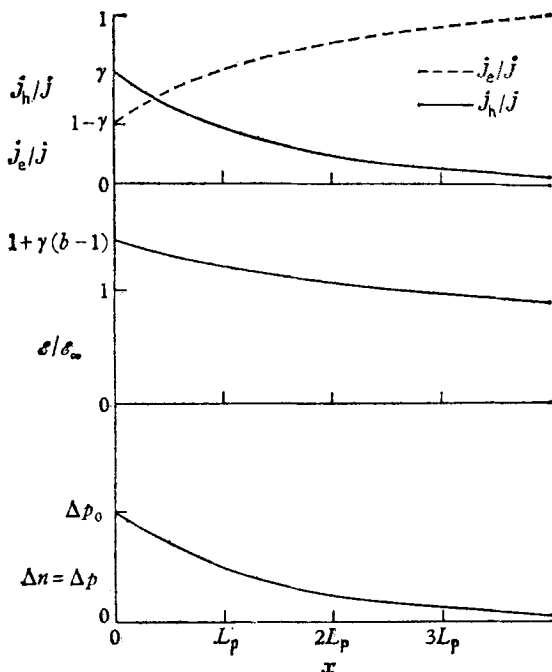


Рис. 7.3. Зависимость электронного j_e и дырочного j_h токов, электрического поля \mathcal{E} и неравновесных концентраций носителей заряда Δn и Δp от координаты (случай слабого поля).

Теперь можно выяснить, когда именно дырочный ток становится в основном диффузионным. При $x=0$, когда \mathcal{E} самое большое, дырочный ток проводимости, равный $e(p+\Delta p)\mu_h\mathcal{E}$, будет мал по сравнению с диффузионным в случае, когда

$$\mathcal{E}\mu_h \ll D_h/L_p, \quad (7.53)$$

т. е. если

$$e\mathcal{E}L_p \ll kT. \quad (7.53a)$$

Здесь $e\mathcal{E}L_p$ — энергия, приобретаемая электроном в поле при прохождении расстояния, равного диффузионной длине. Пусть критическое поле \mathcal{E}_c определяется уравнением

$$\mathcal{E}_c = \frac{kT}{eL_p}, \quad (7.54)$$

тогда дырочный ток будет в основном диффузионным при $\mathcal{E} \ll \mathcal{E}_c$. Это имеет место, если $\mathcal{E}_\infty = j/\sigma \ll \mathcal{E}_c$; например, при $L_p = 0,05 \text{ см} = 5 \cdot 10^{-4} \text{ м}$ критическое поле $\mathcal{E}_c = 50 \text{ В} \cdot \text{м}^{-1}$.

7.5.3. ИНЖЕКЦИЯ НОСИТЕЛЕЙ ЗАРЯДА (СЛАБОЕ ПОЛЕ)

На данном этапе важно оценить численное значение различных величин, используемых для описания диффузионных процессов. Пусть $n=3 \cdot 10^{20} \text{ м}^{-3}$ и $p=3 \cdot 10^{18} \text{ м}^{-3}$, так что $n=100 p$; при $\mu_n = 0,1 \text{ м}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ и $b=2$ получим $\sigma \approx 10 \text{ Ом}^{-1} \cdot \text{м}^{-1}$. Положим также $\mathcal{E}_\infty = 1 \text{ В} \cdot \text{м}^{-1}$ и $\mathcal{E}_c = 50 \text{ В} \cdot \text{м}^{-1}$, так что $j = 10 \text{ А} \cdot \text{м}^{-2} = 1 \text{ мА} \cdot \text{см}^{-2}$. Для Δp_0 имеем

$$\Delta p_0 = \frac{\gamma j L_p}{e D_h} = \frac{\gamma n_0 \mu_c L_p \mathcal{E}_\infty}{D_h} = \gamma b \left(\frac{\mathcal{E}_\infty}{\mathcal{E}_c} \right) n_0. \quad (7.55)$$

Так как мы предположили, что $\mathcal{E}_\infty \ll \mathcal{E}_c$, откуда $\mathcal{E}_0 \ll \mathcal{E}_c$, следовательно, $\Delta p_0 \ll \Delta n_0$; с другой стороны, соотношение между значениями Δp_0 и p_0 может быть произвольным. Положив $\gamma = 0,5$, получим $\Delta p_0 = 0,02 n_0 = 2 p_0$. Таким образом, мы видим, что уже при сравнительно малом положительном напряжении на электроде можно получить значительный рост концентрации неосновных носителей. Общее число добавочных носителей в рассмотренном выше примере составляет $2 \Delta p_0 = 0,04 n_0$. Это явление, известное под названием инжекции неосновных носителей, имеет место (для материала n -типа) при приложении к электроду с $\gamma \neq 0$ положительного потенциала. Чем ближе γ к единице, тем более эффективен процесс инжекции. Необходимо отметить, что пока γ не станет чрезмерно мало, всегда будет происходить *некоторая* инжекция дырок. На практике очень трудно получить коэффициент инжекции γ меньше $\sim 0,01$, хотя это иногда бывает необходимо. Как мы увидим позже, необходимо принять специальные меры, чтобы получить γ малой величины. Обычные же напаянные контакты металл — полупроводник в случае материала n -типа обладают коэффициентом инжекции γ , изменяющимся в пределах от 0,5 до 0,1.

В связи с процессом инжекции необходимо особо отметить два обстоятельства. Поскольку должно выполняться условие $\Delta n = \Delta p$, то на каждую инжектированную дырку приходится один инжектированный электрон. Изменение $\Delta \sigma$ проводимости σ дается, таким образом, выражением

$$\Delta \sigma = e \Delta p (1 + b) \mu_n. \quad (7.56)$$

Вторая особенность заключается в следующем: когда инжекция осуществляется в условиях такого слабого поля, что $\mathcal{E}_\infty \ll \mathcal{E}_c$, изменение проводимости $\Delta \sigma$ ограничивается областью вокруг инжектирующего контакта с радиусом порядка диффузионной длины L_p . Как мы увидим позже, это ограничение может быть в некоторой степени снято, если для инжекции использовать сильное поле. Явление инжекции неосновных носителей заряда лежит в основе работы транзисторов.

7.5.4. ЭКСТРАКЦИЯ НОСИТЕЛЕЙ ЗАРЯДА (СЛАБОЕ ПОЛЕ)

Теперь мы рассмотрим явление, при котором происходит уменьшение концентрации дырок p . В этом случае ток течет из электрода в образец, так что значения как j , так и Δp отрицательны. Это явление известно под названием экстракции носителей заряда. Сразу видно, что этот процесс существенно отличается от инжекции носителей заряда, поскольку в данном случае накладывается дополнительное ограничение $|\Delta p| \ll p_0$, которого раньше не было. Наибольший дырочный ток, который может возникнуть вследствие диффузии, равен $e p_0 D_n / L_p$. По достижении этого значения дальнейший рост тока сопровождается изменением величины γ . В приведенном выше анализе предполагалось, что γ не зависит от j , но это справедливо только для малых величин j . Ясно, однако, что в случае контактов с коэффициентом γ , близким к единице, можно получить заметное уменьшение числа дырок.

Для материала p -типа рассмотренные выше условия будут противоположными: контакт с $\gamma \approx 0$ обеспечивает хорошую инжекцию электронов, если на электрод подан малый отрицательный потенциал.

7.5.5. СИЛЬНОЕ ЭЛЕКТРИЧЕСКОЕ ПОЛЕ

Рассмотрим теперь противоположный случай, когда электрическое поле столь велико, что диффузионными токами можно пренебречь. Это будет справедливо, если $\mathcal{E} \gg \mathcal{E}_c$. Тогда уравнение (7.40) упрощается и в стационарном случае имеет вид

$$\mathcal{E} \mu_h \frac{d\Delta p}{dx} + \frac{\Delta p}{\tau_p} = 0 \quad (7.57)$$

(\mathcal{E} предполагается постоянным). Пусть сначала \mathcal{E} положительно. Тогда решением, удовлетворяющим рассматриваемому уравнению, будет

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_{d_0}}\right), \quad (7.58)$$

где

$$L_{d_0} = \mathcal{E} \mu_h \tau_p = \frac{\mathcal{E}}{\mathcal{E}_c} L_p. \quad (7.59)$$

Величину L_{d_0} называют дрейфовой длиной. Она пропорциональна электрическому полю \mathcal{E} и при условии $\mathcal{E} \gg \mathcal{E}_c$ намного превышает L_p . Дырки смещаются в среднем на расстояние L_{d_0} от точки инжекции. Величина L_{d_0} в точности совпадает с расстоянием, которое они проходили бы в электрическом поле за время τ_p . Дырочный диффузионный ток будет мал по сравнению с дырочным током про-

водимости, если $D_n/L_d \ll \mu_n \mathcal{E}$, т. е. если $\mathcal{E} \gg \mathcal{E}_c$. Диффузионный ток электронов также будет пренебрежимо мал по сравнению с током проводимости. Из формулы (7.45) следует, что в случае сильного тока (при $x=0$)

$$\gamma = \frac{\rho_0 + \Delta\rho_0}{b(n_0 + \Delta\rho_0) + \rho_0 + \Delta\rho_0}. \quad (7.60)$$

Так как $\rho_0 \ll n_0$, это выражение упрощается:

$$\gamma = \frac{\rho_0 + \Delta\rho_0}{bn_0 + (1+b)\Delta\rho_0}. \quad (7.61)$$

Таким образом, пока коэффициент γ достаточно велик (скажем, $\gamma \geq 0,1$), вблизи $x=0$ условие $n \gg p$ или $\mathcal{E} = \text{const}$ уже не будет выполняться и, следовательно, предположения, на которых основан этот простой анализ, станут несправедливыми. Это значит, что случай сильного поля необходимо проанализировать без этих предположений, что и будет осуществлено позже. Если же коэффициент γ мал, то

$$\Delta\rho_0 \approx b\gamma n_0 - \rho_0, \quad (7.62)$$

а область повышенной концентрации дырок теперь распространяется на расстояние порядка L_d от конца образца. Если для рассмотренного выше случая принять $\mathcal{E} = 10\mathcal{E}_c = 500 \text{ В} \cdot \text{м}^{-1}$, то получим ($L_p = 0,05 \text{ см}$) $L_d = 0,5 \text{ см}$. Таким образом, дырки втягиваются в полупроводник на большую глубину. Для отрицательного \mathcal{E} не существует решения, аналогичного уравнению (7.58), и в этом случае необходимо учитывать как проводимость, так и диффузию.

Прежде чем приступить к отысканию общего решения уравнения (7.40) при условии $n \gg p$, необходимо оценить относительную величину члена, зависящего от градиента электрического поля $d\mathcal{E}/dx$.

В случае слабого поля из уравнения (7.51) следует

$$\frac{d\mathcal{E}}{dx} = -\frac{(b-1)\gamma\mathcal{E}_\infty}{L_p} \exp\left(-\frac{x}{L_p}\right). \quad (7.63)$$

При $x=0$, где $d\mathcal{E}/dx$ имеет максимальное значение, получим

$$\rho\mu_n \frac{d\mathcal{E}}{dx} = -\frac{(b-1)(\rho_0 + \Delta\rho_0)\mu_n\gamma\mathcal{E}_\infty}{L_p}. \quad (7.64)$$

Сравним это выражение с диффузионным членом $(D_n/L_p^2)\Delta\rho_0$; последний при условии, что $\mathcal{E}_\infty \ll D_n/\gamma\mu_n L_p = \mathcal{E}_c/\gamma$, будет больше дрейфового. Таким образом, пока не выполняется условие $\gamma \approx 0$, картина оказывается такой же, как это предполагалось раньше. В случае сильного поля

$$\mathcal{E}_x = \frac{i}{\sigma + \Delta\sigma}, \quad (7.65)$$

где $\Delta\sigma = e(1+b)\Delta\rho\mu_h$. При $x=0$ имеем

$$\frac{1}{\mathcal{E}} \frac{d\mathcal{E}}{dx} \approx \frac{(b+1)\Delta\rho_0}{L_{d0}[bn_0 + (b+1)\Delta\rho_0]}. \quad (7.66)$$

Сравним теперь величины членов $p(d\mathcal{E}/dx)$ и $\mathcal{E}(dp/dx)$. При $x=0$ (где $d\mathcal{E}/dx$ максимально) первый из них оказывается меньше второго, поскольку

$$\frac{(b+1)(p_0 + \Delta\rho_0)}{bn_0 + (b+1)\Delta\rho_0} \ll 1. \quad (7.67)$$

Так как $p_0 \ll n_0$, последнее выражение будет справедливо при $\Delta\rho_0 \ll \ll n_0$, т. е. при том же условии, что и раньше. Таким образом, если $p_0 \ll n_0$ и $\Delta\rho_0$ столь мало, что изменение *общего* числа носителей заряда также оказывается малым, в уравнении (7.40) можно пренебречь членом, содержащим $d\mathcal{E}/dx$. В дальнейшем при рассмотрении случая сравнимых по величине токов — диффузионного и тока проводимости — мы будем предполагать, что это условие выполняется.

7.5.6. ОБЩЕЕ РЕШЕНИЕ ($p \ll n$)

Уравнение, которое необходимо решить для определения стационарного состояния, имеет теперь вид

$$D_h \frac{d^2\Delta p}{dx^2} - \mathcal{E}\mu_h \frac{d\Delta p}{dx} - \frac{\Delta p}{\tau_p} = 0. \quad (7.68)$$

При рассмотренных выше условиях электрическое поле \mathcal{E} можно считать постоянным. Тогда общим решением будет

$$\Delta p = A \exp\left(-\frac{x}{L_1}\right) + B \exp\left(-\frac{x}{L_2}\right); \quad (7.69)$$

$\lambda_1 = 1/L_1$ и $\lambda_2 = 1/L_2$ — корни квадратного уравнения

$$\lambda^2 + \frac{\lambda(\mathcal{E}/\mathcal{E}_c)}{L_p} - \frac{1}{L_p^2} = 0, \quad (7.70)$$

где, как и раньше, $\mathcal{E}_c = kT/eL_p$. Заметим, что один из корней уравнения (7.70) должен быть положительным, а другой — отрицательным и оба — действительными. В случае полубесконечного стержня только один из них имеет физический смысл. Корнями уравнения (7.70) будут

$$\lambda_{1,2} = -\frac{\mathcal{E}}{2\mathcal{E}_c} \frac{1}{L_p} \pm \frac{1}{2L_p} \left[\left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 + 4 \right]^{1/2}. \quad (7.71)$$

Если $\mathcal{E} \ll \mathcal{E}_c$, то $\lambda_{1,2} = \pm L_p^{-1}$, и для полубесконечного образца с $x \geq 0$ берется положительный корень. Это соответствует условию слабого поля. Если же $\mathcal{E} \gg \mathcal{E}_c$, то оба корня приближенно выра-

жаются равенствами

$$\lambda_1 = -\frac{\mathcal{E}}{\mathcal{E}_c} \frac{1}{L_p} = -\frac{e\mathcal{E}}{kT}, \quad (7.72)$$

$$\lambda_2 = \frac{\mathcal{E}_c}{\mathcal{E}} \frac{1}{L_p} = \frac{1}{L_{d_0}}. \quad (7.73)$$

Решение при $\lambda = \lambda_1$ соответствует случаю $\mathcal{E} < 0$. Тогда

$$\Delta p = \Delta p_0 \exp\left(-\frac{e|\mathcal{E}|x}{kT}\right). \quad (7.74)$$

Это соответствует обычному больцмановскому распределению дырок в задерживающем поле. Решение, отвечающее λ_2 , справедливо при $\mathcal{E} > 0$ и оказывается таким же, что и полученное раньше [см. формулу (7.58)]. Это означает, что в уравнении (7.71) знак минус берется при $\mathcal{E} < 0$, а плюс — при $\mathcal{E} > 0$.

Определим теперь величины L_u и L_d , заданные уравнениями:

$$\frac{1}{L_d} = \frac{1}{2L_p} \left[\left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 + 4 \right]^{1/2} - \frac{|\mathcal{E}|}{2L_p \mathcal{E}_c} \quad (7.75)$$

и

$$\frac{1}{L_u} = \frac{1}{2L_p} \left[\left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 + 4 \right]^{1/2} + \frac{|\mathcal{E}|}{2L_p \mathcal{E}_c}, \quad (7.76)$$

которые носят названия соответственно «диффузионной длины вдоль поля» и «диффузионной длины против поля». Далее имеем

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_d}\right) \quad (\mathcal{E} > 0) \quad (7.77)$$

и

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_u}\right) \quad (\mathcal{E} < 0). \quad (7.78)$$

Заметим, что при $\mathcal{E} = \mathcal{E}_c$ имеем $L_d = 2L_p / (\sqrt{5} - 1)$ и $L_u = 2L_p / (\sqrt{5} + 1)$. Рассмотрим теперь длинный, тонкий и однородный стержень полупроводникового материала. Поверхностными эффектами пока пренебрежем. Предположим, что в средней точке стержня ($x=0$) находится электрод, из которого с помощью низковольтного источника в образец инжектируются дырки, причем так, что в точке $x=0$ все время поддерживается избыточная концентрация дырок Δp_0 . Между концами образца приложена разность потенциалов, создающая вдоль него поле \mathcal{E} . Тогда, очевидно, получим

$$\begin{aligned} \Delta p &= \Delta p_0 \exp\left(-\frac{x}{L_d}\right) & (x \geq 0), \\ \Delta p &= \Delta p_0 \exp\left(\frac{x}{L_u}\right) & (x < 0). \end{aligned} \quad (7.79)$$

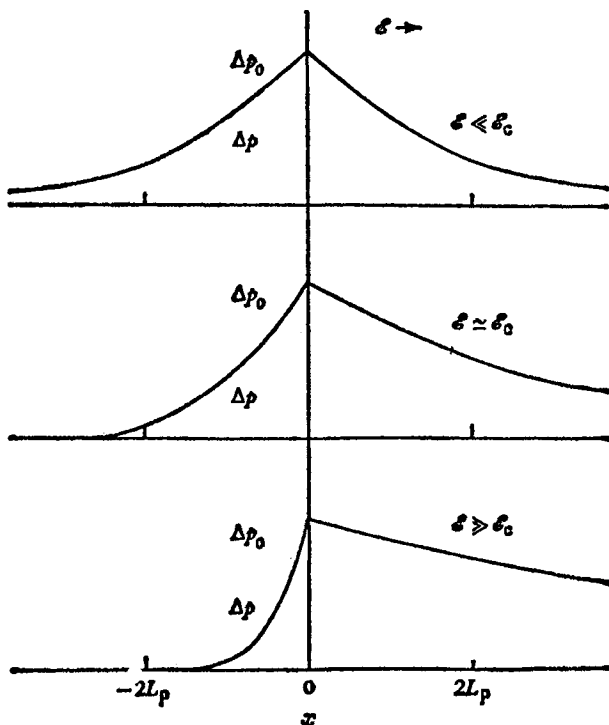


Рис. 7.4. Распределение концентраций инжектированных в точке $x=0$ дырок вдоль тонкого стержня.

На рис. 7.4 показано распределение инжектированных дырок вдоль однородного стержня для разных значений \mathcal{E} . Когда $\mathcal{E} \gg \mathcal{E}_c$ и L_p не очень мало, дырки проходят вдоль поля большое расстояние, в то время как влево от точки инжекции ($x < 0$) проникновение будет невелико.

Предположим, что в момент времени $t=0$ на электрод подан короткий импульс напряжения длительностью Δt , инжектирующий дырки в образец. Вблизи электрода образуется облако дырок. Это облако начнет перемещаться вдоль стержня и постепенно, по мере удаления от точки инжекции, будет спадать по плотности. Так как при $\mathcal{E} \gg \mathcal{E}_c$ дырки движутся со скоростью $\mathcal{E}\mu_n$, то можно ожидать, что с той же скоростью будет перемещаться и всё облако. Заметим, что облако дырок вызовет появление облака электронов. Это следует из условия сохранения электронейтральности. Позже мы увидим, что вследствие этого скорость движения облака изменится только при $n \approx p$ и что при $n \gg p$ облако движется со скоростью $\mathcal{E}\mu_n$. Этот факт лежит в основе сейчас уже ставшего классическим эк-

сперимента, впервые предложенного и осуществленного Хейнсом и Шокли [2]; эксперимент позволяет непосредственно определить подвижность и время жизни неосновных носителей заряда в германии. Ниже будет показано, что облако дырок можно регистрировать в любой точке образца, а время пролета и его плотность можно измерить как функцию расстояния от места инжекции. Следует отметить также, что все полученные выше результаты для дырок в полупроводниках n -типа в равной степени применимы для электронов в полупроводниках p -типа. В этом случае при помощи соотношения $L_n^2 = \tau_n D_e$ мы определим диффузионную длину L_n . Теперь следует приступить к исследованию нестационарных процессов, в которых происходит изменение Δp во времени.

7.6. Дрейф облака неосновных носителей заряда в электрическом поле

Если пренебречь рекомбинацией электронов и дырок, то в отсутствие электрического поля одномерное уравнение непрерывности для дырок (7.40) примет вид

$$\frac{\partial \Delta p}{\partial t} = D_h \frac{\partial^2 \Delta p}{\partial x^2}. \quad (7.80)$$

Это уравнение встречается в теории теплопроводности, и его решения хорошо известны. Например, если в момент времени $t=0$ избыточные дырки возникли в очень узкой области вблизи $x=0$, то решение для последующего момента времени t имеет вид

$$\Delta p = B t^{-1/2} \exp\left(-\frac{x^2}{4D_h t}\right). \quad (7.81)$$

Таким образом, к моменту времени t ширина облака окажется порядка $4(D_h t)^{1/2}$. Если полное число избыточных дырок в облаке равно N_p , то интегрирование соотношения (7.81) от $x=-\infty$ до $x=+\infty$ даст

$$B = \frac{N_p}{(4\pi D_h)^{1/2}}.$$

Для учета роли рекомбинации представим Δp в виде

$$\Delta p = f(x, t) \exp\left(-\frac{t}{\tau_p}\right) \quad (7.82)$$

и, подставив это выражение в уравнение (7.40) с $\mathcal{E}=0$, убедимся, что полученное выражение для функции $f(x, t)$ совпадает по форме

с уравнением (7.80). Таким образом, выражение

$$\Delta p = \frac{N_p}{(4\pi D_h t)^{1/2}} \exp\left(-\frac{t}{\tau_p} - \frac{x^2}{4D_h t}\right) \quad (7.83)$$

является решением интересующей нас задачи с учетом рекомбинации. Оно показывает, что, как и раньше, облако расширяется, а максимальная его плотность в момент времени t уменьшилась в $\exp(t/\tau_p)$ раз. Общее число неравновесных дырок в момент времени t теперь уже равно $N_p \exp(-t/\tau_p)$. Если вместо бесконечно узкого облака при $t=0$ мы имеем облако гауссовой формы вида

$$\Delta p = B \exp\left(-\frac{x^2}{x_0^2}\right), \quad (7.84)$$

то можем воспользоваться тем же уравнением, заменив t на $t+t_0$, где $t_0 = x_0^2/4D_h$.

Учтем теперь электрическое поле \mathcal{E} , но, как и раньше, предположим, что членом, содержащим $\partial\mathcal{E}/\partial x$, в уравнении (7.40) можно пренебречь. Сделаем подстановку

$$x' = x - \mathcal{E}\mu_h t. \quad (7.85)$$

Уравнение для $f(x, t)$ снова принимает такой же вид, как и (7.80), а решение определяется выражением

$$\Delta p = \frac{N_p \exp(-t/\tau_p)}{(4\pi D_h t)^{1/2}} \exp\left[-\frac{(x - \mathcal{E}\mu_h t)^2}{4D_h t}\right]. \quad (7.86)$$

Это подтверждает отмеченный выше результат, что облако смещается в направлении поля со скоростью $\mathcal{E}\mu_h$, расширяясь по мере движения, и что его плотность уменьшается в отношении $\exp(-t/\tau_p)$. Благодаря этому можно непосредственно измерить μ_h , τ_p и D_h , если в последующие моменты времени определять положение облака и наблюдать за изменением его формы.¹⁾ Найденную таким образом подвижность в общем случае называют амбиполярной дрейфовой подвижностью. Если $p \ll n$ или $n \ll p$, она оказывается равной соответственно дрейфовой подвижности дырок или электронов, но при $n \approx p$, очевидно, это уже не так.

7.7. Полупроводники с проводимостью, близкой к собственной

Мы обсудили некоторые детали процесса движения неосновных носителей заряда в примесных полупроводниках, у которых $p \ll n$

¹⁾ О способе регистрации положения и формы облака дырок см. разд. 7.13.—
Прим. ред.

или $n \ll p$, поскольку анализ при этих допущениях значительно проще, чем при условии $n \approx p$. При этом мы предполагали, что $\Delta n = \Delta p$ и что величиной $d\mathcal{E}/dx$ из-за малости можно пренебречь. Оба эти допущения взаимозависимы, поскольку $d\mathcal{E}/dx$ и $\Delta n - \Delta p$ связаны уравнением Пуассона. В этом разделе мы исследуем движение носителей заряда, когда эти ограничения сняты. Мы уже указывали, что предыдущий анализ оказывается несостоятельным, если инжекция неосновных носителей в полупроводниках n -типа такова, что Δp становится сравнимым с n (при этом Δp может быть гораздо больше p).

Вернемся на время к случаю слабого поля, когда \mathcal{E} нам известно и, следовательно, можно вычислить $\Delta n - \Delta p$. Используя уравнение Пуассона [выражение (7.25)] и воспользовавшись величиной \mathcal{E} , определяемой из уравнения (7.51), получим при $x=0$

$$\frac{\Delta n - \Delta p}{p} = \frac{\varepsilon \varepsilon_0 (b-1) \gamma \mathcal{E}_\infty}{c L_p p}. \quad (7.87)$$

Положив $\varepsilon=10$ и, как прежде, $L_p = 5 \cdot 10^{-4}$ м, $p = 3 \cdot 10^{18}$ м $^{-3}$, получим $(\Delta n - \Delta p)/p \approx 10^{-5}$ ($\gamma \mathcal{E}_\infty / \mathcal{E}_c$). Отсюда видно, что в этом случае приближение $\Delta n = \Delta p$ вполне удовлетворительно. В случае сильного поля выполнялось соотношение

$$\frac{d\mathcal{E}}{dx} \approx \frac{\Delta p_0}{n_0} \frac{\mathcal{E}}{b L_{d_0}} = \frac{\Delta p_0}{n_0} \frac{\mathcal{E}_c}{b L_p} \quad (7.88)$$

[ср. с уравнением (7.66)]. Поэтому можно записать

$$\frac{\Delta n - \Delta p}{p} = \frac{\Delta p_0}{n_0} \frac{\varepsilon_0 \varepsilon \mathcal{E}_c}{e b p L_p} = \quad (7.89)$$

$$= 10^{-5} \frac{\Delta p_0}{n_0}. \quad (7.89a)$$

Мы уже видели, что если $\Delta p_0 \approx n_0$, то нельзя пренебрегать в уравнении (7.40) членом, содержащим $d\mathcal{E}/dx$. Тем не менее, хотя $(\Delta n - \Delta p)$ и кажется очень малым по сравнению с p , даже этого небольшого объема заряда оказывается достаточно для создания очень сильного поля. Таким образом, $\Delta n = \Delta p$ является хорошим приближением для всех членов в уравнениях (7.40) и (7.40а), за исключением тех, которые содержат $d\mathcal{E}/dx$, и, следовательно, при рассмотрении общего случая желательно исключить эти члены.

Прежде всего рассмотрим более подробно рекомбинационные члены. Если нет захвата ловушками электронов или дырок, т. е. если электроны и дырки генерируются и рекомбинируют парами, то из уравнений (7.40) и (7.40а) в отсутствие электрического поля

или градиентов концентрации следует

$$\frac{d\Delta p}{dt} = \frac{dp}{dt} = \mathcal{R}_0 - \frac{p}{\tau_p'} = -\frac{\Delta p}{\tau_p}, \quad (7.90)$$

$$\frac{d\Delta n}{dt} = \frac{dn}{dt} = \mathcal{R}_0 - \frac{n}{\tau_n'} = -\frac{\Delta n}{\tau_n}, \quad (7.91)$$

где \mathcal{R}_0 — темп тепловой генерации, p/τ_p' — темп рекомбинации дырок, n/τ_n' — темп рекомбинации электронов. Из этих уравнений следуют *определения* параметров τ_n и τ_p . Очевидно, если $n \gg p$, τ_p является временем жизни неосновных носителей заряда и не зависит от p . Обозначив через τ_{n_0}' и τ_{p_0}' значения τ_n' и τ_p' , соответствующие равновесным концентрациям n_0 и p_0 , получим $\tau_p' = \tau_p = \tau_{p_0} = p_0/\mathcal{R}_0$ при $n \gg p$. Если же не делать этого допущения, то темп генерации \mathcal{R}_0 определяется соотношением

$$\mathcal{R}_0 = \frac{p_0}{\tau_{p_0}'} = \frac{n_0}{\tau_{n_0}'}. \quad (7.92)$$

Таким образом, из уравнений (7.91) и (7.92) следует

$$\frac{1}{\tau_p} = \frac{\frac{p_0}{\tau_{p_0}'} - \frac{p}{\tau_p'}}{p_0 - p}, \quad (7.93)$$

$$\frac{1}{\tau_n} = \frac{\frac{n_0}{\tau_{n_0}'} - \frac{n}{\tau_n'}}{n_0 - n}. \quad (7.94)$$

В общем случае τ_n и τ_p являются функциями n и p , причем точная форма этой зависимости определяется тем, какой из механизмов рекомбинации преобладает. Если основную роль играет излучательная рекомбинация, то ее темп будет пропорционален как n , так и p , и поэтому

$$\frac{n}{\tau_n'} = \frac{p}{\tau_p'} = Anp = \mathcal{R}_0 \frac{np}{n_0 p_0}. \quad (7.95)$$

Отсюда получаем

$$\frac{1}{\tau_n} = \mathcal{R}_0 \frac{np - n_0 p_0}{n_0 p_0 (n - n_0)} \quad (7.96)$$

$$\frac{1}{\tau_p} = \mathcal{R}_0 \frac{np - n_0 p_0}{n_0 p_0 (p - p_0)}. \quad (7.97)$$

Чтобы $dn/dt = dp/dt$, должно выполняться условие $(p - p_0)/\tau_p = (n - n_0)/\tau_n$. Величины, входящие в уравнения (7.96) и (7.97), удовлетворяют этому условию. Уравнение (7.97) можно записать в виде

$$\frac{1}{\tau_p} = \frac{\mathcal{R}_0}{n_0 p_0} \left(n_0 + p_0 \frac{\Delta n}{\Delta p} + \Delta n \right), \quad (7.98)$$

и если $p_0 \ll n_0$, а также $\Delta n \ll n_0$, то получим

$$\frac{1}{\tau_p} = \frac{\mathcal{R}_0}{p_0}, \quad (7.99)$$

так что τ_p — постоянная величина. Аналогично имеем

$$\frac{1}{\tau_n} = \frac{\mathcal{R}_0}{n_0 p_0} \left(p_0 + n_0 \frac{\Delta p}{\Delta n} + \Delta p \right). \quad (7.100)$$

Ниже будет показано, что в уравнениях (7.98) и (7.100) можно положить $\Delta n = \Delta p$ и тогда

$$\frac{1}{\tau_n} = \frac{1}{\tau_p} = \frac{\mathcal{R}_0}{n_0 p_0} (n_0 + p_0 + \Delta p). \quad (7.101)$$

Если Δp не мало по сравнению с n_0 , то τ_p уже не постоянно, и переход к равновесию не будет подчиняться экспоненциальному закону. Вывод, что $\tau_n = \tau_p$, если $\Delta n = \Delta p$, является, однако, весьма общим и вытекает из условия $dn/dt = dp/dt$.

Вернемся теперь к уравнениям непрерывности и запишем их в виде

$$\frac{\partial \Delta p}{\partial t} = -\frac{\Delta p}{\tau_p} + D_h \frac{\partial^2 \Delta p}{\partial x^2} - \mu_h p \frac{\partial \mathcal{E}}{\partial x} - \mu_h \mathcal{E} \frac{\partial \Delta p}{\partial x}, \quad (7.102)$$

$$\frac{\partial \Delta n}{\partial t} = -\frac{\Delta n}{\tau_n} + D_e \frac{\partial^2 \Delta n}{\partial x^2} + \mu_e n \frac{\partial \mathcal{E}}{\partial x} + \mu_e \mathcal{E} \frac{\partial \Delta n}{\partial x}. \quad (7.103)$$

Для нахождения Δn , Δp и \mathcal{E} мы имеем и третье уравнение, а именно уравнение Пуассона

$$\frac{\partial \mathcal{E}}{\partial x} = e \frac{\Delta p - \Delta n}{\epsilon \epsilon_0}. \quad (7.104)$$

Подставив $\partial \mathcal{E} / \partial x$ в уравнения (7.102) и (7.103), получим

$$\frac{\partial \Delta p}{\partial t} = -\frac{\Delta p}{\tau_p} + D_h \frac{\partial^2 \Delta p}{\partial x^2} + \frac{\sigma_h}{\epsilon \epsilon_0} (\Delta n - \Delta p) - \mu_h \mathcal{E} \frac{\partial \Delta p}{\partial x} \quad (7.105)$$

и

$$\frac{\partial \Delta p}{\partial t} = -\frac{\Delta n}{\tau} + D_e \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\sigma_e}{\epsilon \epsilon_0} (\Delta n - \Delta p) + \mu_e \mathcal{E} \frac{\partial \Delta n}{\partial x}, \quad (7.106)$$

где σ_e и σ_h — соответственно электронная и дырочная проводимости. Мы уже видели, что величину $|\Delta n - \Delta p|$ можно считать малой. Было бы, однако, неправильно опустить в уравнениях (7.105) и (7.106) члены, содержащие эту разность, так как коэффициенты $\sigma_e / \epsilon \epsilon_0$ и $\sigma_h / \epsilon \epsilon_0$ велики. Именно из-за этих коэффициентов допущение $\Delta n = \Delta p$, т. е. игнорирование роли объемного заряда, может оказаться несостоятельным. К счастью, от этих членов можно освободиться. Умножив уравнение (7.105) на σ_e , а уравнение (7.106) на

σ_h и сложив их, получим

$$\begin{aligned} \sigma_e \frac{\partial \Delta p}{\partial t} + \sigma_h \frac{\partial \Delta n}{\partial t} = & - \left(\frac{\sigma_e \Delta p}{\tau_p} + \frac{\sigma_h \Delta n}{\tau_n} \right) + \\ & + \left(\sigma_e D_h \frac{\partial^2 \Delta p}{\partial x^2} + \sigma_h D_e \frac{\partial^2 \Delta n}{\partial x^2} \right) + \sigma_h \mu_e \mathcal{E} \frac{\partial \Delta n}{\partial x} - \\ & - \sigma_e \mu_h \mathcal{E} \frac{\partial \Delta p}{\partial x}. \end{aligned} \quad (7.107)$$

Теперь, когда мы исключили члены с большим множителем при $(\Delta n - \Delta p)$, можем с большим основанием использовать приближение $\Delta n = \Delta p$. Несколько позднее мы вычислим $(\Delta n - \Delta p)$ и покажем, что эта разность мала, но что $(\sigma/\epsilon\epsilon_0)(\Delta n - \Delta p)$ оказывается малым лишь при определенных условиях. Теперь, полагая в уравнении (7.107) $\Delta n = \Delta p$, получаем после деления на $(\sigma_e + \sigma_h)$

$$\frac{\partial \Delta p}{\partial t} = - \frac{\Delta p}{\tau_p} + D \frac{\partial \Delta p}{\partial x^2} - \mu \mathcal{E} \frac{\partial \Delta p}{\partial x}, \quad (7.108)$$

где

$$D = \frac{\sigma_e D_h + \sigma_h D_e}{\sigma_e + \sigma_h} = \quad (7.109)$$

$$= \frac{n \mu_e D_h + p \mu_h D_e}{n \mu_e + p \mu_h}. \quad (7.109a)$$

Воспользовавшись соотношением Эйнштейна, запишем уравнение (7.109a) в виде

$$D = \frac{n + p}{\frac{n}{D_h} + \frac{p}{D_e}}. \quad (7.109b)$$

Имеем также

$$\mu = \frac{\sigma_e \mu_h - \sigma_h \mu_e}{\sigma_e + \sigma_h} = \quad (7.110)$$

$$= \frac{n - p}{\frac{n}{\mu_h} + \frac{p}{\mu_e}}. \quad (7.110a)$$

Принято называть D коэффициентом амбиполярной диффузии, а μ амбиполярной дрейфовой подвижностью. Величины D и μ описывают движение облака неравновесных дырок. Это следует из того факта, что уравнение (7.108) оказалось того же вида, что и детально рассмотренное в разд. 7.5.6, но с другими значениями коэффициентов диффузии и подвижности. Только при $p \ll n$ значение μ совпадает с μ_h и D с D_h . Аналогично при $n \ll p$ значение μ совпадает с μ_e и D с D_e . Таким образом, получено обоснование рассмотренного выше уравнения (7.68) с μ_e , D_e или μ_h и D_h ; тем самым мы подтвердили правильность приближения, согласно которому членами в уравнении (7.40), содержащими $\partial \mathcal{E} / \partial x$, можно пренебречь. Изложенный выше

анализ совпадает в основных чертах с ходом рассуждений Ван Рубрека [3]; аналогичные уравнения рассмотрел также Ганн [4], использовавший в качестве переменных параметров $N = (bn+p)n_1$ и $Q = (p-n)/n_1$; он преобразовал уравнения к такому виду, где значение величины $(\Delta p - \Delta n)$ стало еще более очевидным.

Использование полупроводников с почти собственной проводимостью приобрело важное значение после изобретения фоторезисторов — детекторов инфракрасного излучения. Свойства таких полупроводников были рассмотрены Риттнером [5]. В приведенном выше анализе мы пренебрегали захватом электронов ловушками с глубокими уровнями. Исследование проблемы было продолжено Риттнером, который учел условия, когда и возбуждение электрона из ловушки, и захват его ловушкой оказываются существенными. Позднее при рассмотрении фотопроводимости мы учтем роль ловушек (см. разд. 10.9.2). Материалы с почти собственной проводимостью имеют большое значение в случае некоторых типов транзисторов и в этом аспекте исследовались Ганном [4]. Когда $n = p = n_1$, коэффициент амбиполярной диффузии D имеет постоянную величину, определяемую уравнением

$$D = \frac{2D_e D_h}{D_e + D_h} = \frac{2bD_h}{b+1}. \quad (7.111)$$

Если отношение подвижностей носителей заряда b , как, например, в InSb, велико, то коэффициент амбиполярной диффузии в собственном полупроводнике приблизительно равен коэффициенту диффузии дырок, несмотря на то что коэффициент диффузии электронов имеет значительно большую величину.

Для собственных материалов таким же образом получаем $\mu = 0$, что несколько неожиданно. Подвижность μ не является скоростью дрейфа частиц в электрическом поле, а представляет собой скорость дрейфа концентрационного возмущения¹⁾. В случае материала n -типа величина μ положительна и облако движется в направлении, в котором двигались бы *положительные* заряды, т. е. в направлении движения избыточных *дырок*. Если же материал p -типа, то облако движется в противоположную сторону, т. е. в ту, в которую двигались бы *электроны*. Отсюда видно, что в электрическом поле квазинейтральное облако носителей движется в направлении, в котором двигались бы неосновные носители заряда²⁾. Если $n \gg p$ или $p \gg n$, то скорость дрейфа квазинейтрального облака в точности равна скорости дрейфа неосновных носителей заряда, но при $n \rightarrow p$ скорость в том же поле становится меньше и в конце концов (при $n = p$)

¹⁾ Возмущение представляет собой квазинейтральное облако неравновесных носителей заряда, в котором $\Delta n \sim \Delta p$. — Прим. ред.

²⁾ Этот результат получен Херрингом [6] без учета рекомбинации. См. также [7].

скорость дрейфа становится равной нулю. Именно эта зависимость μ от n и p вызывает необходимость делать различие между амбиполярной дрейфовой подвижностью μ и дрейфовыми подвижностями μ_e и μ_h .

Когда $n_0 \approx p_0$ и $\Delta n \approx \Delta p \ll n_0$, то n и p в уравнениях (7.109) и (7.110) можно заменить их равновесными значениями. Тогда при малых уровнях инжекции μ и D будут постоянны и проведенный выше анализ останется справедливым, если заменить μ_h на μ и D_h на D и т. д. Диффузионную длину L можно опять определить при помощи соотношения

$$L^2 = D\tau, \quad (7.112)$$

где в данном приближении $\tau_n = \tau_p = \tau$. В этом случае D и L носят название соответственно коэффициента амбиполярной диффузии и амбиполярной диффузионной длины. Стационарное решение уравнения (7.108) запишется в виде

$$\Delta n = \Delta p = \Delta p_0 \exp\left(-\frac{x}{L'}\right), \quad (7.113)$$

где L' — либо «диффузионная длина вдоль поля», либо «диффузионная длина против поля».

В качестве второго приближения примем

$$\Delta n = \Delta p + \eta \quad (7.114)$$

и вернемся к уравнению (7.105) для определения величины члена, учитывающего объемный заряд. Для облегчения анализа предположим, что поле слабое, и рассмотрим только диффузию. Тогда можно положить $L' = L = (D\tau)^{1/2}$; в результате получим

$$\frac{\sigma_h \eta}{\epsilon \epsilon_0} = \frac{\Delta p}{\tau_p} \left(1 - \frac{D_h}{D}\right) = \frac{\Delta p}{\tau_p} \frac{1+b}{b} \frac{p}{n+p}. \quad (7.115)$$

Отсюда видно, что член в уравнении (7.105), характеризующий объемный заряд, пренебрежимо мал только при $p \ll n$. Заметим, что подобным же членом в уравнении (7.106) при этом же условии уже нельзя пренебречь, так как

$$\frac{\sigma_e}{\epsilon \epsilon_0} \eta = \frac{\Delta p}{\tau_p} (1+b) \frac{n}{n+p}, \quad (7.116)$$

что сравнимо с величиной $\Delta p/\tau_p$.

Однако при $p \ll n$ достаточно одного лишь уравнения (7.105) без члена с $|\Delta n - \Delta p|$, а вместо (7.106) можно воспользоваться просто условием $\Delta n = \Delta p$. Аналогично если $n \ll p$, то достаточно одного уравнения (7.106). Как и раньше, на основе типичного численного примера легко показать, что $\eta/\Delta p \sim 10^{-5}$, и сделать вывод, что пока поле не слишком велико, т. е. меньше 10^7 В·м⁻¹, всегда $\Delta n \approx \Delta p$.

Небольшой объемный заряд, возникающий при сравнимых по величине n и p , оказывается, однако, достаточным для возникновения сильного поля, нарушающего это условие.

7.7.1. ПРИБЛИЖЕНИЕ СЛАБОГО ПОЛЯ

Предположим теперь, что n_0 , p_0 , $n_0 - p_0$ сравнимы по величине, а $n_0 > p_0$. Если приложено слабое поле, так что $\Delta p \ll n$ или p , то величины D и μ остаются постоянными, но отличаются от D_e , D_h и μ_e , μ_h . Решение при $\mathcal{E} > 0$ для полубесконечного стержня дается уравнением (7.113) с заменой L' на L'_d , где L'_d — диффузионная длина носителей заряда при движении их по полю, получаемая подстановкой в уравнение (7.75) величины D вместо D_h , так что дырочный ток

$$j_h = \gamma j = e\mu_h (p_0 + \Delta p) \mathcal{E} + \frac{eD_h}{L'_d} \Delta p. \quad (7.117)$$

Вблизи $x=0$ для полного тока можно приближенно записать $j = e[\mu_e(n_0 + \Delta n_0) + \mu_h(p_0 + \Delta p_0)]\mathcal{E}$. Отсюда, учитывая, что $\Delta n_0 = \Delta p_0$, найдем поле

$$\mathcal{E} = \frac{j}{e\mu_h [bn_0 + p_0 + (b+1)\Delta p_0]}; \quad (7.118)$$

тогда при $x=0$ получим

$$\gamma j = j \frac{p_0 + \Delta p_0}{bn_0 + p_0 + (b+1)\Delta p_0} + \frac{eD_h}{L'_d} \Delta p_0. \quad (7.119)$$

Если Δp_0 мало, то это уравнение можно переписать в виде

$$(\gamma - \gamma_0)j = \frac{eD_h}{L'_d} \Delta p_0. \quad (7.120)$$

где

$$\gamma_0 = \frac{p_0}{bn_0 + p_0}. \quad (7.121)$$

При $\gamma = \gamma_0$ получим $\Delta p = 0$; такой контакт называется «омическим». Если $\gamma > \gamma_0$, то $\Delta p > 0$, и будет происходить инжекция дырок; это условие обычно выполняется в случае сильно примесных полупроводников n -типа, если только величина γ не чрезвычайно мала. У почти собственных материалов очень легко добиться осуществления условий, когда $\gamma < \gamma_0$, и мы видим, что в этом случае происходит истощение неосновных носителей заряда. Особенно интересная ситуация наблюдается при очень малых γ и при наложении сильного положительного смещения. Если в уравнении (7.117) пренебречь диффузионным членом, то при $\gamma \rightarrow 0$ окажется, что $\Delta p \rightarrow -p_0$. Это означает, что из объема образца вблизи $x=0$ все неосновные носители удалены. В сильном поле L'_d велико и область обеднения носителями

заряда может распространиться на весь образец, длина которого мала по сравнению с L'_d . Это явление известно под названием *экссклюзии* носителей заряда.

7.7.2. ЭКСКЛЮЗИЯ НОСИТЕЛЕЙ ЗАРЯДА

Об экссклюзии носителей заряда впервые сообщил Гибсон [8]. Артур, Бэрдсли, Браун и Гибсон [9], а также Брэй [10] изучали ее на германии. (В оригинальных работах это явление называлось эксстракцией носителей, но это было обусловлено некоторой путаницей с эффектом «вытягивания» носителей заряда в слабом поле.) Экссклюзия — весьма поразительный эффект. Например, если $n_0 = 1,1 n_i$ и $p_0 = 0,9 n_i$, то число носителей, оставшееся после экссклюзии, составит $0,2 n_i$, а проводимость при $b \approx 2$ должна уменьшиться по крайней мере в 10 раз. На опыте наблюдалось еще большее уменьшение, однако для достижения такого результата необходимо соблюсти еще ряд условий.

Пусть мы имеем образец длиной l . Тогда если диффузионная длина по полю L'_d значительно меньше l , то обедненная носителями заряда область не будет охватывать весь образец. В приближении сильного поля имеем $L'_d = \mu_n \tau_p$, и для полной экссклюзии получаем условие

$$\mu_n \tau_p \gg l. \quad (7.122)$$

Итак, после завершения процесса экссклюзии, когда основная часть дырок уже вытянута, получим $n \gg p$, так что $\mu \approx \mu_n$. Пусть V_e — такая внешняя разность потенциалов, что $\mu_n \tau_p = l$. Тогда, используя соотношение Эйнштейна, легко получим

$$V_e = \frac{kT}{be} \left(\frac{l}{L_p} \right)^2. \quad (7.123)$$

Таким образом, для осуществления почти полной экссклюзии необходимо приложить разность потенциалов, несколько превышающую V_e .

Если L_d сравнимо по величине с l , то нам следовало бы воспользоваться решением вида

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_d}\right) + \Delta p_1 \exp\left(\frac{x-l}{L_u}\right). \quad (7.124)$$

В приближении сильного поля $L_u \ll L_d$, так что концентрация меняется только вблизи $x=l$. Как будет показано ниже, на величине Δp это скажется заметным образом только тогда, когда коэффициент γ для контакта, расположенного в точке $x=l$, мал. Следовательно, законно предположить, что эффект, связанный с L_u , будет мал, и тогда, записав Δp в виде

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_d}\right), \quad (7.125)$$

мы должны предположить также, что $L_d \gg l$. Определив время пробега t как $l/\mathcal{E}\mu$, получим $L_d = l\tau_p/t$. Поэтому наше допущение эквивалентно условию $t \ll \tau_p$. Из уравнения (7.119), пренебрегая диффузионным членом, при $\gamma \ll 1$ получаем

$$\Delta p_0 = -p_0 + \gamma (bn_0 - p_0), \quad (7.126)$$

так что уравнение (7.125) можно переписать в виде

$$p(x) = p_0 - [p_0 - \gamma (bn_0 - p_0)] \exp\left(-\frac{x}{L_d}\right). \quad (7.127)$$

При $x=l$ получим приближенно

$$p(l) = \gamma (bn_0 - p_0) + \frac{p_0 l}{\tau_p}; \quad (7.128)$$

для того чтобы $p(l)$ было намного меньше p_0 , должны выполняться условия

$$\gamma \ll \frac{p_0}{bn_0 - p_0} \quad \text{и} \quad t \ll \tau_p.$$

Второе условие эквивалентно $V \gg V_e$, где V — внешняя разность потенциалов. Ниже мы рассмотрим методы получения γ малой величины.

Теперь видно, почему при малых коэффициентах инжекции γ при сильном положительном смещении происходит резкое падение концентрации носителей заряда: поле настолько быстро уносит дырки из образца, что их убыль не может быть возмещена тепловой генерацией или инжекцией с положительного электрода. При больших величинах γ инжекция дырок оказывается достаточной для сохранения в образце их концентрации. Более полный теоретический анализ явления эксклюзии носителей заряда дал Рид [11].

В работе [9] рассмотрено интересное применение явления эксклюзии носителей для изучения изменения амбиполярной дрейфовой подвижности μ при приближении к области собственной проводимости. Это дало возможность проверить уравнение (7.110а).

7.7.3. АККУМУЛЯЦИЯ НОСИТЕЛЕЙ ЗАРЯДА

Когда коэффициент γ мал, а направление поля таково, что неосновные носители заряда (дырки) движутся к электроду, расположенному при $x=0$, можно наблюдать явление, обратное эксклюзии носителей. Поскольку коэффициент γ мал, то при нормальной концентрации дырок они не успевают достаточно быстро проникать в электрод, вблизи которого, таким образом, происходит накопление дырок и создается высокая их концентрация. Соответствующим этому случаю решением будет

$$\Delta p = \Delta p_0 \exp\left(-\frac{x}{L_u}\right), \quad (7.129)$$

где при сильном поле, как было показано, $L_u = kT/e|\mathcal{E}|$. Теперь вблизи $x=0$ получим

$$\gamma j = j_h = -(\rho_0 + \Delta\rho_0) \mu_h |\mathcal{E}| + \frac{\Delta\rho_0 D_h}{L_u} = -\rho_0 \mu_h |\mathcal{E}| \quad (7.130)$$

и

$$(1 - \gamma) j = j_e = -(n_0 + \Delta\rho_0) \mu_e |\mathcal{E}| - \frac{\Delta\rho_0 D_e}{L_u} = -(n_0 + 2\Delta\rho_0) b \mu_h |\mathcal{E}|, \quad (7.131)$$

так что

$$\gamma = \frac{\rho_0}{bn_0 + \rho_0 + 2b\Delta\rho_0} \quad (7.132)$$

и

$$\Delta\rho_0 = \frac{\rho_0(1 - \gamma) - \gamma bn_0}{2b\gamma}. \quad (7.133)$$

Из уравнения (7.133) следует, что при малых γ величина $\Delta\rho_0$ становится очень большой. Явление аккумуляции носителей заряда в германии исследовали Артур, Гибсон и Ганн [12]. Вместо коэффициента инжекции γ эти авторы для описания свойств контакта ввели величину v_1 , называемую скоростью утечки в контакте. Она определяется соотношением

$$ev_1 \Delta\rho_0 = j_h = \gamma j, \quad (7.134)$$

так что $\gamma = ev_1 \Delta\rho_0 / j$. Ни γ , ни v_1 не зависят от j , но когда v_1 мало; то и γ также мало. До сих пор мы рассматривали только одномерный случай. Случай радиального потока, который более соответствует точечному контакту, рассмотрен Ганном [4].

7.8. Сопоставление различных контактных явлений

На данном этапе интересно сопоставить различные явления, относящиеся к контактам с полупроводником. Лоу [13] выделил и проанализировал четыре различных эффекта, связанных с рассмотренными явлениями. Инжекция и экстракция неосновных носителей заряда представляют собой эффекты слабого поля и определяются главным образом диффузией; инжекция имеет место у положительного, а экстракция — у отрицательного контакта в случае полупроводника n -типа и наоборот у полупроводников p -типа. Они связаны с величиной коэффициента γ ; у материала n -типа он близок к единице, а у материала p -типа — близок к нулю. Эти явления характерны в основном для примесных полупроводников. Эксклюзия и аккумуляция носителей заряда являются эффектами сильного поля. В слабо легированном полупроводнике n -типа, если величина

коэффициента γ близка к нулю и контакт положителен, происходит эксклюзия, а если контакт отрицателен — аккумуляция носителей заряда. В слабо легированном полупроводнике p -типа, если коэффициент γ близок по величине к единице и контакт отрицателен, происходит эксклюзия, а если контакт положителен — аккумуляция носителей заряда. Из этих четырех эффектов наиболее важным с технической точки зрения, несомненно, является инжекция, поскольку на ней основана работа большинства типов транзисторов и многих других приборов. Остальные эффекты, однако, также имеют большое прикладное значение, поскольку определяют детали работы этих устройств, а также находят и специальные применения.

Все эти явления были исследованы экспериментально Гарриком [14], применившим превосходную технику эксперимента. На образец падал очень узкий пучок инфракрасного излучения из той спектральной области, где основную роль играет поглощение свободными носителями (см. разд. 10.2). По измеренному поглощению определялась концентрация носителей и изменение ее по образцу. Благодаря этому Гаррику удалось наблюдать большинство рассмотренных в предыдущих разделах явлений, обусловленных довольно слабым электрическим полем. Он сумел также проверить полученные соотношения для почти собственных полупроводников.

7.9. $p-n$ -переход

Приступим теперь к рассмотрению ряда явлений, возникающих в неоднородном образце. В этом случае, как отмечалось в разд. 7.1, в образце могут существовать значительные объемные заряды, обуславливающие появление сильных электрических полей. Именно в этом и заключается резкое отличие неоднородного полупроводника от однородного, в котором объемный заряд в толще образца должен быть малым. Примером наиболее сильной из всех встречающихся неоднородностей может служить полупроводник, в котором концентрации доноров N_d и акцепторов N_a меняются так, что в каком-то месте образца происходит смена типа проводимости с электронной на дырочную. Необходимо подчеркнуть, что здесь речь идет об одном кристалле, а не о контактах между двумя отдельными кристаллами разного типа.

Пусть, например, в определенной области кристалла (рис. 7.5) существует такой линейный градиент концентраций доноров и акцепторов, что слева от переходной области $N_a > N_d$, а справа от нее $N_d > N_a$. Слева от переходной области будет полупроводник p -типа с эффективной концентрацией акцепторов $N_a - N_d$, а справа — полупроводник n -типа с эффективной концентрацией доноров $N_d - N_a$ (см. разд. 4.3). Подобные переходы между материалами с проводимостями n - и p -типа носят название $p-n$ -переходов. Они имеют чрезвычайно важное прикладное значение для многих современных

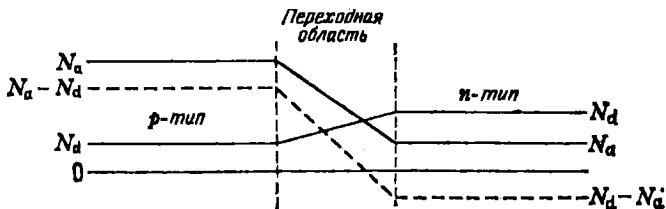


Рис. 7.5. Переход между полупроводниками n - и p -типа.

полупроводниковых приборов. Интересоваться ими начали с тех пор, как в 1938 г. Давыдов [15] рассмотрел обусловленные ими явления выпрямления переменного тока и возникновения фото-э. д. с.; прикладное значение p — n -переходов в течение многих лет недооценивалось. Еще до появления транзисторов ряд экспериментов с p — n -переходами в германии осуществили, например, Бенцер [16], а также Беккер и Фэн [17]. Роль p — n -переходов в фотопроводимости и возникновении фото-э. д. с. в пленках PbS обсуждалась Сосновским, Старкевичем и Симпсоном [18]. Математическую теорию p — n -переходов типа, представленного на рис. 7.5, в 1949 г. разработал Шокли [19]. На основе этой теории он предложил довольно детальную схему транзистора с p — n -переходами, принцип действия которого оказался в замечательном согласии с предсказаниями теории. Это позволило транзисторам на p — n -переходах занять главное место в технологии транзисторов, которое они продолжают сохранять и поныне.

Математический анализ перехода с линейным градиентом концентрации доноров и акцепторов оказывается весьма сложным. Кроме того, выяснилось, что наибольшее практическое значение имеют переходы с очень тонкой переходной областью. Позже будут рассмотрены критерии понятия тонкого перехода. Для многих практических целей оказывается вполне достаточным рассматривать подобный изображенному на рис. 7.6 бесконечно узкий переход от полупроводника n -типа к полупроводнику p -типа. На рис. 7.6 в n -области показаны только доноры, а в p -области — только акцепторы. Если имеются оба типа центров, то N_a и N_d можно просто заменить *эффективным* числом доноров или акцепторов. Предположим, что энергетические уровни доноров и акцепторов соответственно расположены так близко к зоне проводимости и к валентной зоне, что все они полностью ионизованы. Будем считать также $N_a \gg n_1$ и $N_d \gg n_1$. Пусть справа от перехода $n = n_n$ и $p = p_n$, а слева от него $n = n_p$ и $p = p_p$. Предположим сначала, что переход еще не сформировался¹⁾. Тогда в правой части будем иметь приближенно

¹⁾ Другими словами, предположим, что p - и n -части приведены в идеальное соприкосновение, а необходимое равновесное распределение зарядов и поля вблизи границы раздела еще не сформировалось. — *Прим. ред.*

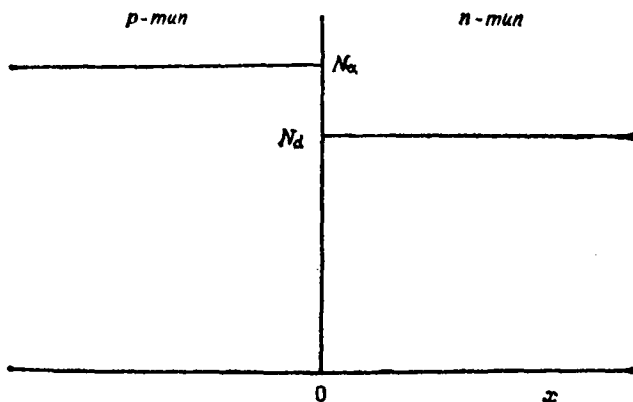
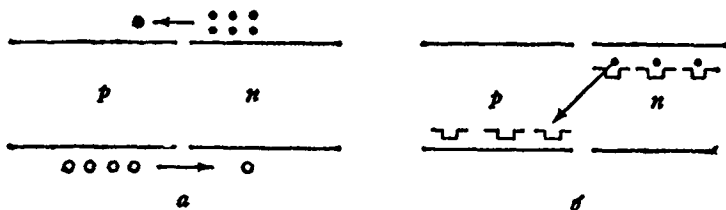
Рис. 7.6. Тонкий p — n -переход.

Рис. 7.7. Диффузия электронов и дырок через переход.

$n_n = N_d$, $n_n \gg p_n = n_i^2/N_d$, в то время как в левой части $p_p = N_a$, $p_p \gg n_p = n_i^2/N_a$. Вместе с тем $n_n \gg n_p$ и $p_n \ll p_p$. Сразу после соприкосновения n - и p -частей вблизи границы перехода должны возникнуть большие градиенты концентрации как электронов, так и дырок. В результате начнется диффузия электронов из n -области в p -область и дырок из p -области в n -область. Это иллюстрируется рис. 7.7,а. Другой способ описания этого процесса дает схема энергетических уровней, подобная изображенной на рис. 7.7,б. При образовании p — n -перехода электроны зоны проводимости n -области начнут переходить на более низкие уровни, заполненные дырками в p -области. Этот процесс эквивалентен переходу дырок в n -область. Он будет происходить даже при очень низких температурах, когда доноры и акцепторы не ионизованы, так как электроны, занимающие донорные уровни, перейдут на пустые акцепторные уровни. Этот процесс не может продолжаться бесконечно, так как уход электронов из n -полупроводника и приток в него дырок приведет к образованию объемного заряда; аналогичный объемный заряд образуется также и в p -области. Эти объемные заряды создают

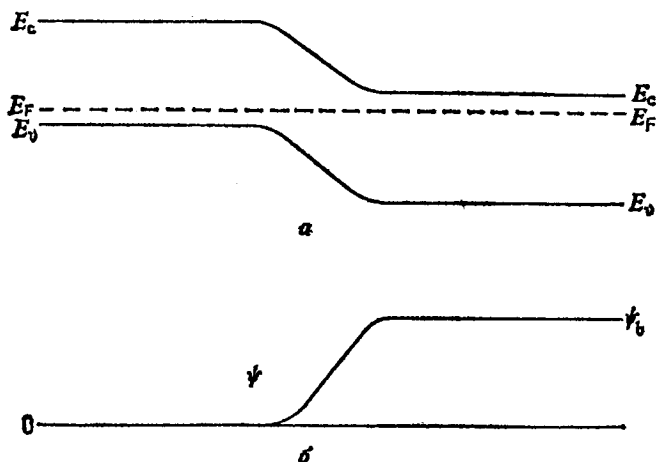


Рис. 7.8. Энергетические зоны и электростатический потенциал ψ в p - n -переходе.

сильное электрическое поле такого направления, которое будет препятствовать диффузионному движению электронов и дырок, так что в конце концов устанавливается равновесие.

В условиях равновесия наличие в p -области отрицательного заряда приводит к повышению энергии ее электронных состояний. Величину смещения энергетических уровней в равновесном образце легко вычислить из условий постоянства положения уровня Ферми (рис. 7.8). Если обозначить величины, относящиеся к материалу p -типа, индексом a , а величины, относящиеся к материалу n -типа, индексом b , то относительное повышение системы уровней в n -материале относительно уровней в p -материале будет равно $E_{F_a} - E_{F_b}$.

Из разд. 7.2 следует, что электростатический потенциал ψ отличается от энергии $-E_c/e$ только на постоянную величину. Тогда, если обозначить с помощью индексов a и b значения различных величин вдали от области перехода, получим

$$E_{c_a} - E_{c_b} = E_{v_a} - E_{v_b} = e(\psi_b - \psi_a). \quad (7.135)$$

Для невырожденного случая (см. разд. 4.3) имеем

$$\left. \begin{aligned} n_n &= N_c \exp\left(\frac{E_F - E_{cb}}{kT}\right), \\ p_n &= N_v \exp\left(\frac{E_{vb} - E_F}{kT}\right), \\ p_p &= N_v \exp\left(\frac{E_{va} - E_F}{kT}\right), \\ n_p &= N_c \exp\left(\frac{E_F - E_{ca}}{kT}\right), \end{aligned} \right\} \quad (7.136)$$

$$n_i^* = N_c N_v \exp\left(-\frac{E_{ca} - E_{vb}}{kT}\right) = N_c N_v \exp\left(-\frac{\Delta E}{kT}\right), \quad (7.137)$$

где N_c , N_v — величины, определенные в разд. 4.2. Из уравнений (7.135) и (7.136), используя электростатический потенциал, получаем

$$\frac{n_p}{n_n} = \frac{p_n}{p_p} = \exp\left[\frac{e(\psi_a - \psi_b)}{kT}\right], \quad (7.138)$$

а также

$$E_{cb} - E_{va} = -kT \ln\left(\frac{n_n}{N_c}\right) - kT \ln\left(\frac{p_p}{N_v}\right), \quad (7.139)$$

или, используя соотношение (7.135),

$$e(\psi_b - \psi_a) = E_{ca} - E_{cb} = \Delta E + kT \ln\left(\frac{n_n p_p}{N_c N_v}\right). \quad (7.140)$$

Когда n_n и p_p значительно превышают n_i , имеем $e(\psi_b - \psi_a) \approx \Delta E$. Для вычисления электростатического поля в переходе необходимо знать эффективную область изменения ψ по обе стороны от контакта в областях с проводимостями n - и p -типа.

Объемный заряд ρ , возникший вследствие потока электронов через переход, определяется выражением

$$\rho = e(p - n + N_d - N_a). \quad (7.141)$$

Если нуль электростатического потенциала выбрать так, что $\psi = 0$, когда электрон обладает энергией E_{ca} , т. е. $\psi_a = 0$, то концентрации электронов и дырок можно выразить следующим образом:

$$p = N_a \exp\left(-\frac{e\psi}{kT}\right), \quad (7.142)$$

$$n = N_d \exp\left[\frac{e(\psi - \psi_b)}{kT}\right]. \quad (7.143)$$

Объединяя уравнения (7.141), (7.142), (7.143), получаем зависимость ρ от ψ . Уравнение Пуассона, в случае когда ψ есть функция одной координаты, можно записать в виде

$$\epsilon \epsilon_0 \frac{d^2 \psi}{dx^2} = -\rho(\psi). \quad (7.144)$$

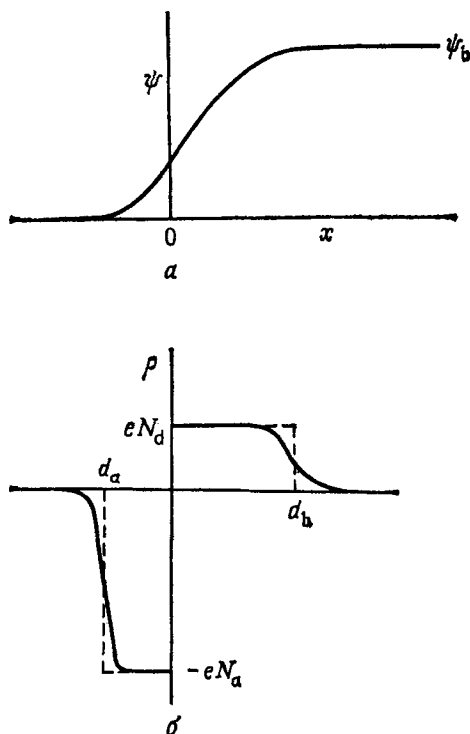


Рис. 7.9. Объемный заряд p - n -перехода.

Для полученной выше зависимости $\rho(\psi)$ точно решить это уравнение трудно, однако приближенное значение можно найти следующим образом ¹⁾. Заметим, что концентрация носителей заряда n экспоненциально зависит от ψ , так что при отрицательном $(\psi - \psi_b)$ величина n быстро уменьшается, а значение ρ при этом резко возрастает. В этой области концентрация зарядов быстро достигает величины eN_d справа от перехода и $-eN_a$ слева от него. Эта ситуация иллюстрируется рис. 7.9, на котором показана зависимость ρ и ψ от x . На рисунке видно, что с достаточной степенью точности можно считать ρ постоянным и равным $-eN_a$ в интервале от $x = -d_a$ до $x = 0$ слева от контакта, а в интервале от $x = 0$ до $x = d_b$ справа от контакта $\rho = eN_d$; в других областях $\rho = 0$. Расстояния d_a и d_b определяются из уравнения Пуассона, которое теперь можно записать

¹⁾ См., например, [20].

В виде

$$\begin{aligned} \frac{d^2\psi}{dx^2} &= 0 & (-\infty < x < -d_a), \\ &= \frac{e}{\epsilon\epsilon_0} N_a & (-d_a \leq x < 0), \\ &= -\frac{e}{\epsilon\epsilon_0} N_d & (0 \leq x < d_b), \\ &= 0 & (d_b \leq x < \infty). \end{aligned} \quad (7.145)$$

Граничные условия в этом случае следующие:

$$\begin{aligned} \psi &= 0 & (x = -d_a), \\ \frac{d\psi}{dx} &= 0 & (x = -d_a), \\ \psi &= \psi_b & (x = d_b), \\ \frac{d\psi}{dx} &= 0 & (x = d_b). \end{aligned}$$

Кроме того, производная $d\psi/dx$ непрерывна при $x=0$ (рис. 7.10). Очевидно, из условия сохранения электронной нейтральности $p-n$ -перехода в целом получим

$$d_a N_a = d_b N_d. \quad (7.146)$$

Для электрического поля \mathcal{E} имеем

$$\left. \begin{aligned} \mathcal{E} &= -\frac{d\psi}{dx} = -\frac{e}{\epsilon\epsilon_0} N_a (d_a + x) & (-d_a < x \leq 0), \\ \mathcal{E} &= \frac{e}{\epsilon\epsilon_0} N_d (x - d_b) & (d_b \geq x > 0), \\ \mathcal{E} &= 0 & \text{для всех остальных областей.} \end{aligned} \right\} \quad (7.147)$$

Условие непрерывности \mathcal{E} при $x=0$ эквивалентно соотношению (7.146). Интегрируя уравнения (7.147) с использованием граничных условий при $x=-d_a$ и $x=d_b$, получаем

$$\begin{aligned} \psi &= \frac{e}{2\epsilon\epsilon_0} N_a (d_a + x)^2 & (-d_a \leq x < 0), \\ \psi &= \psi_b - \frac{e}{2\epsilon\epsilon_0} N_d (d_b - x)^2 & (0 \leq x \leq d_b). \end{aligned} \quad (7.148)$$

Из условия непрерывности ψ при $x=0$ следует

$$\psi_b = \frac{e}{2\epsilon\epsilon_0} (N_d d_b^2 + N_a d_a^2) \quad (7.149)$$

и теперь, поскольку ψ_b считается известным, из уравнений (7.146) и (7.149) находим d_a и d_b .

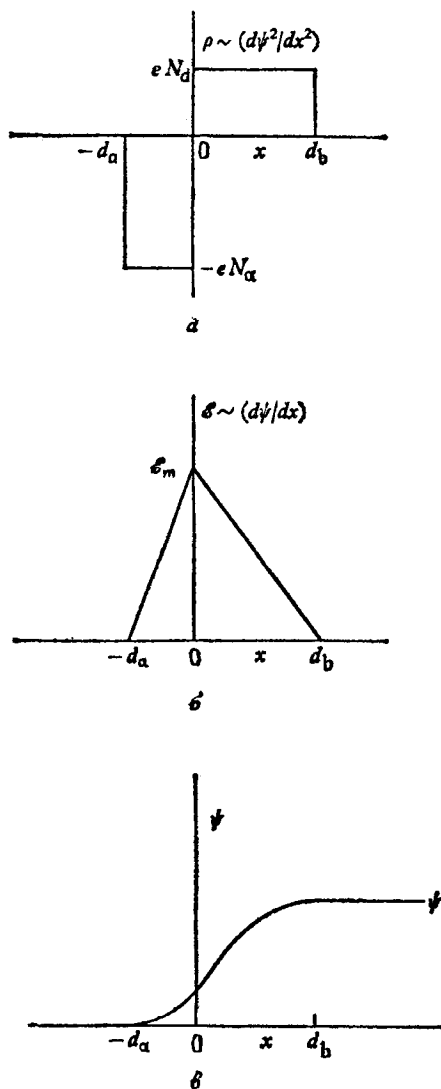


Рис. 7.10. Объемный заряд, электрическое поле и электростатический потенциал тонкого $p-n$ -перехода.

Если одна часть перехода легирована значительно сильнее, чем другая, то соотношения эти можно несколько упростить. Пусть, например, $N_a \gg N_d$. Подобный переход обычно обозначается как $p^+ - n$ -переход. Тогда из уравнения (7.146) следует, что $d_a \ll d_b$ и

может быть принято равным нулю. В этом случае

$$d_b^2 = \frac{2\epsilon\epsilon_0}{eN_d} \psi_b \approx \frac{2\epsilon\epsilon_0 \Delta E}{e^2 N_d} \quad (7.150)$$

(заметим, что $\psi_b > 0$). Например, если принять $\Delta E = 1$ эВ, $N_d = 10^{21}$ м⁻³ и $\epsilon = 10$, то получим $d_b \approx 1$ мкм. Тогда максимальная величина \mathcal{E} составит $2 \cdot 10^6$ В·м⁻¹. Отсюда видно, что вблизи перехода существует довольно сильное поле.

Теперь можно более точно сформулировать понятие тонкого перехода: переход, приблизительно отвечающий рассмотренным выше реальным условиям, можно назвать тонким, если d (толщина переходного слоя между n - и p -областями) мала по сравнению с расстоянием $d_a + d_b$, вычисленным для идеального перехода. Переход, для которого это условие не выполняется, обычно называется толстым или широким переходом и требует более детального рассмотрения [19].

7.9.1. БАРЬЕРНАЯ ЕМКОСТЬ

Из предыдущего анализа ясно, что сопротивление p — n -перехода будет велико по сравнению с сопротивлением слоя полупроводника n - или p -типа той же площади и толщины, что и область объемного заряда. Более того, переход будет обладать свойствами конденсатора, поскольку он состоит из двух «обкладок» из материала с высокой проводимостью, разделенных областью полупроводника с низкой проводимостью. Емкость C_b на единицу площади, известная под названием барьерной емкости, определяется уравнением

$$C_b = \frac{\epsilon\epsilon_0}{d_a + d_b}. \quad (7.151)$$

Когда $N_a \gg N_d$, имеем

$$C_b = \left(\frac{\epsilon\epsilon_0 e N_d}{2\psi_b} \right)^{1/2} \approx \left(\frac{\epsilon\epsilon_0 e^2 N_d}{2\Delta E} \right)^{1/2}. \quad (7.152)$$

Тогда для рассмотренного выше примера получим $C_b = 0,01$ мкФ·см².

7.9.2. ВОЛЬТАМПЕРНАЯ ХАРАКТЕРИСТИКА

Допустим, что к переходу приложена разность потенциалов V_n , положительными значениями V_e будут те, при которых p -область заряжена положительно. Падение напряжения вследствие высокого сопротивления слоя объемного заряда, простирающегося от $x = -d_a$ до $x = d_b$, будет иметь место почти исключительно в этой области. Тогда приближение, согласно которому электростатический

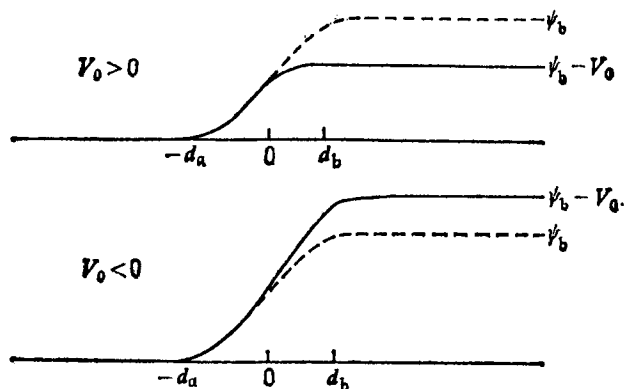


Рис. 7.11. Влияние внешнего напряжения на потенциальный барьер p - n -перехода.

потенциал ψ принимает значения (рис. 7.11)

$$\begin{aligned} \psi &= 0 & (x \leq -d_a), \\ \psi &= \psi_b - V_0 & (x \geq d_b), \end{aligned} \quad (7.153)$$

оказывается достаточно хорошим. Так как теперь равновесия нет, уровень Ферми не имеет постоянного значения. В стационарном случае и в отсутствие вырождения можно заменить ψ на $\psi - V_0$, так что [см. (7.138)] при $\psi_a = 0$ получим

$$\frac{p_n}{p_p} = \frac{n_p}{n_n} = \exp \left[\frac{-e(\psi_b - V_0)}{kT} \right]. \quad (7.154)$$

Концентрация избыточных дырок Δp_0 при $x = d_b$ равна

$$\Delta p_0 = p_n \left[\exp \left(\frac{eV_0}{kT} \right) - 1 \right], \quad (7.155)$$

аналогично концентрация избыточных электронов Δn_0 при $x = -d_a$ равна

$$\Delta n_0 = n_p \left[\exp \left(\frac{eV_0}{kT} \right) - 1 \right]. \quad (7.156)$$

Если предположить, что приложенное поле не слишком велико, то эти избыточные концентрации будут сохраняться справа и слева от перехода лишь до расстояний порядка L_p и L_n . Было показано, что в любом случае падение напряжения имеет место в основном в области объемного заряда, так что ток вблизи перехода определяется главным образом диффузией. Можно предположить, что L_p и L_n значительно больше d_a и d_b , и считать, что Δn_0 и Δp_0 являются избыточными концентрациями *неосновных* носителей заряда для

обеих частей перехода. Заметим, что при $V_0 > 0$ как Δn_0 , так и Δp_0 соответствуют увеличению концентрации неосновных носителей.

Используя уравнения (7.155) и (7.156), легко вычислить дырочный ток в правой части перехода и электронный ток в его левой части. Тогда из уравнений (7.155) и (7.156) следует

$$j_e = \frac{en_p D_e}{L_n} \left[\exp\left(\frac{eV_0}{kT}\right) - 1 \right] \quad (x < 0), \quad (7.157)$$

$$j_h = \frac{ep_n D_h}{L_p} \left[\exp\left(\frac{eV_0}{kT}\right) - 1 \right] \quad (x > 0). \quad (7.158)$$

Теперь, поскольку слой объемного заряда достаточно узок, можно предположить, что внутри слоя нет заметной рекомбинации. Если это не так, то следует видоизменить теорию. Таким образом, электронный и дырочный токи в переходе можно считать непрерывными и записать тогда плотность полного тока j в виде

$$j = j_s \left[\exp\left(\frac{eV_0}{kT}\right) - 1 \right], \quad (7.159)$$

где

$$j_s = e \left(\frac{p_n D_h}{L_p} + \frac{n_p D_e}{L_n} \right), \quad (7.160)$$

L_p и D_h — диффузионная длина и коэффициент диффузии дырок в n -области; L_n и D_e — соответствующие величины для электронов в p -области. Когда $V_0 > 0$, этот ток быстро растет при увеличении V_0 . С другой стороны, если $V_0 < 0$, то плотность тока насыщается и стремится к величине $-j_s$. Уравнение (7.159) вывел Шокли [19], поэтому его часто называют уравнением Шокли. Оно показывает, что p — n -переход обладает сильным выпрямляющим действием и фактически ведет себя как идеальный выпрямитель переменного тока. Вольтамперные характеристики большинства выпрямителей описываются уравнением вида

$$i = i_0 \left[\exp\left(\frac{\beta e V_0}{kT}\right) - 1 \right], \quad (7.161)$$

где i_0 и β — константы. Выпрямитель называют *идеальным*, если $\beta = 1$. Уравнение (7.159) экспериментально проверялось многими исследователями, и, если только переходная область оказывалась узкой в указанном выше смысле, совпадение теории с экспериментом было превосходным. Уравнения (7.159) и (7.160) можно также выразить через проводимости σ_n и σ_p соответственно n - и p -областей перехода и σ_1 — проводимость собственного полупроводника.

В этом случае получим

$$\sigma_n = e(bn_n + p_n)\mu_h \approx ebn_n\mu_h = \frac{ebn_n^2\mu_h}{p_n}, \quad (7.162)$$

$$\begin{aligned} \sigma_p &= e(bn_p + p_p)\mu_h \approx ep_p\mu_h = \frac{en_p^2\mu_h}{n_p}, \\ \sigma_i &= en_i(b+1)\mu_h. \end{aligned} \quad (7.163)$$

Используя соотношение Эйнштейна, получим для плотности тока насыщения

$$j_s = \frac{b\sigma_i^2}{(1+b)^2} \left(\frac{1}{\sigma_n L_p} + \frac{1}{\sigma_p L_n} \right) \frac{kT}{e}. \quad (7.164)$$

Практический интерес представляет случай, когда $N_a \gg N_d$ и $N_d \gg n_i$. При этих условиях $\sigma_p \gg \sigma_n$ и $j_e \ll j_h$. Такой переход называется $p^+ - n$ -переходом. Когда $\gamma \approx 1$, он обладает свойством контакта, очень эффективно инжектирующего дырки. Необходимо, однако, отметить, что условием, для того чтобы $\gamma \approx 1$, является неравенство $\sigma_p L_n \gg \sigma_n L_p$, т. е. диффузионная длина L_n в сильно легированной p -области не должна быть малой. На практике это обычно означает, что сильное легирование не должно приводить к чрезмерному уменьшению времени жизни электронов τ_n . Легко видеть, что ток насыщения теперь можно представить в виде

$$j_s = \frac{bkT\sigma_i^2}{e(1+b)^2\sigma_n L_p} \quad (7.165)$$

и что свойства перехода полностью определяются n -областью.

Коэффициент инжекции γ дается уравнением

$$\gamma = \frac{j_h}{j} = \frac{\sigma_p L_n}{\sigma_p L_n + \sigma_n L_p}. \quad (7.166)$$

Если $\sigma_p L_n \ll \sigma_n L_p$, то $\gamma \approx 0$ и переход становится эффективным контактом для инжекции электронов в область полупроводника p -типа, причем инжектирующим электродом служит n -область, заряженная отрицательно по отношению к p -области. Переход такого типа известен как $n^+ - p$ -переход. Он обладает свойствами, аналогичными свойствам $p^+ - n$ -перехода.

Интересно рассмотреть уравнения (7.157) — (7.164) с более физической точки зрения. Ток в «прямом» направлении образуется вследствие инжекции дырок в n -область и электронов в p -область. На рис. 7.11 хорошо видно, как это происходит. Потенциальный барьер при наличии равновесия показан штриховой линией. Он удерживает основную массу электронов в n -области и основную массу дырок в p -области. При наложении положительной разности потенциалов V_0 высота барьера понижается на величину V_0 , что на рис. 7.11 изображено сплошной линией. Теперь больше электро-

нов может перейти из n -области в p -область, создавая электронный ток в прямом направлении. Аналогично возросшее число дырок, переходящих из p -области в n -область, создает прямой дырочный ток. С другой стороны, если $V_0 < 0$, то высота барьера возрастает, а токи в этом направлении уменьшаются до значений обратных токов насыщения. Когда V_0 велико и отрицательно, никакого потока электронов из n -области в p -область нет, остается только поток зарядов, обусловленный диффузией электронов из p -области по направлению к потенциальному барьеру, пройдя который, они скатываются в n -область, и аналогичный диффузионный поток дырок из n -области. Порядок величины плотности тока насыщения можно легко оценить, исходя из величины этих диффузионных токов. Дырки в n -области генерируются теплом в темпе p_n/τ_p на единицу объема, а затем они диффундируют по направлению к барьеру на расстояние порядка диффузионной длины L_p . Поэтому дырочный ток будет порядка $eL_p p_n/\tau_p$, что, согласно уравнению (7.160), эквивалентно $e p_n D_n/L_p$. Аналогично электронный ток насыщения будет порядка $e n_p D_e/L_n$.

Заметим, что в случае p^+ — n -перехода обратный ток насыщения пропорционален p_n , т. е. концентрации неосновных носителей заряда в n -области. Таким образом, ток насыщения служит мерой концентрации неосновных носителей. Если она искусственно повышена, например освещением или инжекцией из другого контакта, то соответствующим образом возрастет ток насыщения. Небольшой контакт в виде p^+ — n -перехода к n -области, к которому приложено отрицательное смещение, называют *коллектором*, поскольку он собирает неосновные носители заряда из окружающей его области, а его ток насыщения служит мерой концентрации неосновных носителей. Подобный контакт можно использовать, например, для детектирования подходящего к нему облака дырок (см. разд. 7.6). При освещении контакта этого типа обнаруживается заметное изменение проводимости. Это явление, как известно, используется в фотодиодах.

Как показывает опыт, обратный ток насыщения при неограниченном увеличении отрицательного напряжения не остается неизменным, а очень медленно растет вплоть до критического напряжения $-V_0$. При этом значении потенциала ток начинает быстро увеличиваться. Иногда этот потенциал назывался напряжением Зинера, так как считалось, что при этом напряжении электрическое поле способно создавать электронно-дырочные пары путем возбуждения электронов из заполненной зоны в зону проводимости, т. е. имеет место процесс, который носит название эффекта Зинера. Теперь полагают, что в большинстве случаев резкий рост тока обусловлен ударной ионизацией и образованием лавины (см. разд. 12.3). Характеристика пробоя вполне обратима, а сам эффект используется различными способами для создания стабилизаторов напряже-

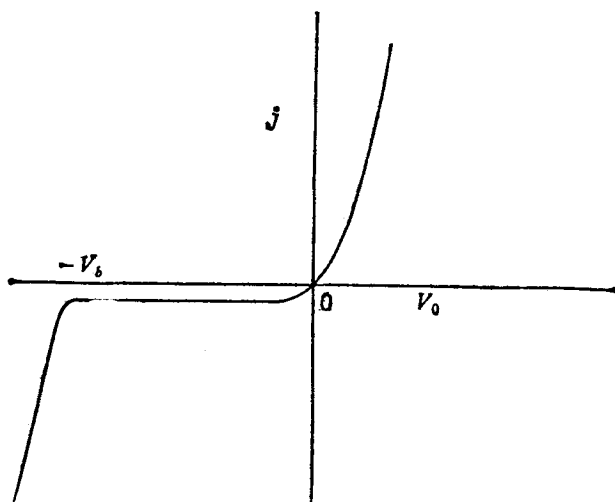


Рис. 7.12. Вольт-амперная характеристика диода с p — n -переходом.

ния и т. п. Типичная вольт-амперная характеристика p — n -перехода изображена на рис. 7.12. Величину V_b можно варьировать путем изменения состава области с меньшим содержанием примесей.

Напряженность электрического поля достигает своего максимального значения \mathcal{E}_m в области p — n -перехода. Это значение определяется из уравнения (7.147) после подстановки в него величины d_b

$$\mathcal{E}_m^2 = \frac{2e |V_t| N_d}{\epsilon \epsilon_0} = \frac{2 |V_t| \sigma_n}{\epsilon \epsilon_0 \mu_n}, \quad (7.167)$$

где V_t — полная разность потенциалов на переходе, которая приблизительно равна V_0 , когда $|eV_0|$ велико по сравнению с ΔE . Если через \mathcal{E}_a обозначить потенциал, при котором начинается возникновение лавины, то максимальное обратное напряжение определится уравнением

$$V_m = \frac{\epsilon \epsilon_0 \mathcal{E}_a^2 \mu_n}{2 \sigma_n}, \quad (7.168)$$

из которого видно, что $V_m \sim \sigma_n^{-1}$. Если положить $\epsilon = 10$, $\mathcal{E}_a = 10^7$ В·м⁻¹, $\sigma_n = 10^2$ Ом⁻¹·м⁻¹ и $\mu_n = 0,2$ м²·В⁻¹·с⁻¹, то $V_m = 10$ В.

Ток насыщения в обратном направлении пропорционален σ_i^2 и оказывается малым для веществ с широкой запрещенной зоной, у которых соответственно мало σ_i . Например, он значительно меньше у Si, чем у Ge, что делает кремний более перспективным материалом для производства выпрямителей. В настоящее время выпрямители с p — n -переходом быстро вытесняют все другие типы выпрямителей.

Воспользовавшись уравнениями (7.159) и (7.164), можно вычислить проводимость перехода. Если i — ток, а A — площадь перехода, то

$$i = A j_s \left[\exp \left(\frac{eV_0}{kT} \right) - 1 \right] = \quad (7.169)$$

$$= i_s \left[\exp \left(\frac{eV_0}{kT} \right) - 1 \right], \quad (7.169a)$$

где $i_s = A j_s$ — ток насыщения. Дифференциальная проводимость G определяется соотношением

$$G = \frac{di}{dV_0} = \frac{ei_s}{kT} \exp \left(\frac{eV_0}{kT} \right). \quad (7.170)$$

При слабом токе, полагая $V_0 \rightarrow 0$, получаем

$$G_0 = \frac{ei_s}{kT} = \quad (7.171)$$

$$= \frac{Ab\sigma_i^2}{(1+b)^2} \left(\frac{1}{\sigma_n L_p} + \frac{1}{\sigma_p L_n} \right). \quad (7.172)$$

Для $p^+ - n$ -перехода

$$G_0 = \frac{Ab}{(1+b)^2} \left(\frac{\sigma_i}{L_p} \right) \left(\frac{\sigma_i}{\sigma_n} \right). \quad (7.173)$$

Это примерно в σ_n/σ_i раз меньше проводимости слоя собственного полупроводника толщиной L_p .

7.9.3. ВЫСОКОЧАСТОТНЫЕ СВОЙСТВА $p - n$ -ПЕРЕХОДА

В разд. 7.9.1 было показано, что $p - n$ -переход обладает барьерной емкостью, определяемой уравнением (7.152), которая шунтируется проводимостью G . Однако уравнение (7.152) следует видоизменить, учитывая, что изменение падения электростатического потенциала на переходе при приложении к нему напряжения V_0 равно $\psi - V_0$. Тогда для барьерной емкости $p^+ - n$ -перехода получим соотношение

$$C_b = \left[\frac{A \epsilon \epsilon_0 e N_d}{2(\psi_b - V_0)} \right]^{1/2}. \quad (7.174)$$

Таким образом, при наложении обратного смещения барьерная емкость уменьшается. В этом случае G мало, следовательно, емкость становится более существенной, чем при прямом смещении.

Существует и другой эффект, обуславливающий появление реактивной составляющей в импедансе $p - n$ -перехода; так как диффузия электронов и дырок — процесс инерционный, то должен иметь место эффект накопления, который, как будет показано, также эк-

вивалентен появлению у перехода емкости, включенной параллельно C_b . Если к переходу приложено напряжение вида $V = V_0 e^{i\omega t}$, то можно предположить, что и решение диффузионного уравнения будет периодическим во времени. Следовательно, мы получим

$$D_n \frac{d^2 \Delta p}{dx^2} = \left(\frac{1}{\tau_p} + i\omega \right) \Delta p. \quad (7.175)$$

Решение этого уравнения совпадает по форме с найденным раньше, если заменить L_p на $L_p / (1 + i\omega\tau_p)^{1/2}$. Отсюда получим выражение для тока i :

$$\hat{i} = i_s \left[\exp \left(\frac{eV_0 e^{i\omega t}}{kT} \right) - 1 \right] (1 + i\omega\tau_p)^{1/2}, \quad (7.176)$$

и если $eV_0 \ll kT$, а $\omega\tau_p \ll 1$, то приближенно

$$\hat{i} = G_0 V_0 e^{i\omega t} \left(1 + \frac{1}{2} i\omega\tau_p \right). \quad (7.177)$$

Таким образом, обозначив через C_d диффузионную емкость, получим

$$C_d = \frac{1}{2} G_0 \tau_p, \quad (7.178)$$

и тогда полная проводимость перехода при слабых переменных токах равна

$$G_0 + i\omega (C_b + C_d). \quad (7.178a)$$

До сих пор мы рассматривали плоскостные $p-n$ -переходы. В случае переходов, образуемых точечными контактами, задача сильно усложняется, однако, если предположить, что подобный переход сферически-симметричен, можно вычислить диффузионный ток, используя для этого диффузионное уравнение соответствующего вида. В этом случае очень важно сопротивление полупроводника вблизи перехода, поскольку им определяется так называемое сопротивление растекания. Сопротивление полусферического контакта с радиусом r_0 равно $1/2 \rho \sigma r_0$. Оно резко меняется при пропускании тока в прямом направлении, поскольку σ меняется из-за инжекции, а это в свою очередь вызывает изменение падения напряжения в области объемного заряда. Все это приводит к тому, что вольтамперная характеристика оказывается несколько другого вида, чем это следует из уравнения (7.159). Хениш [21] рассмотрел различные варианты и дал исчерпывающий анализ свойств выпрямляющих контактов. Имеется также весьма обширная литература по $p-n$ -переходам, связанная с разработкой транзисторов.

7.10. $n^+ - n$ - и $p^+ - p$ -переходы

Другим типом переходов, имеющим большое практическое значение, является переход между двумя частями полупроводника с одинаковым типом проводимости, но с сильно отличающимися эффективными концентрациями доноров или акцепторов. Если одна часть легирована сильнее, чем другая, то такой переход называется либо $n^+ - n$ -переходом, либо $p^+ - p$ -переходом. Схема энергетических уровней такого перехода n -типа (в равновесии) показана на рис. 7.13. Обозначим, как и прежде, левую и правую части перехода буквами a и b и предположим, что $N_{da} \gg N_{db}$. Изменение электростатического потенциала ψ вдоль перехода получим точно так же, как и для $p - n$ -перехода. В этом случае

$$\begin{aligned} N_{da} &= N_c \exp\left(-\frac{E_{ca} - E_F}{kT}\right), \\ N_{db} &= N_c \exp\left(-\frac{E_{cb} - E_F}{kT}\right), \end{aligned} \quad (7.179)$$

так что

$$\psi_b - \psi_a = -\frac{1}{e} (E_{cb} - E_{ca}) = -\frac{kT}{e} \ln\left(\frac{N_{da}}{N_{db}}\right), \quad (7.180)$$

и если, как и прежде, положить $\psi_a = 0$, то для ψ_b получим

$$\psi_b = -\frac{kT}{e} \ln\left(\frac{N_{da}}{N_{db}}\right). \quad (7.181)$$

В некоторых отношениях $n^+ - n$ -переход отличается от $p - n$ -перехода. Например, у $n^+ - n$ -перехода отсутствует область с высоким сопротивлением, а концентрация основных носителей заряда

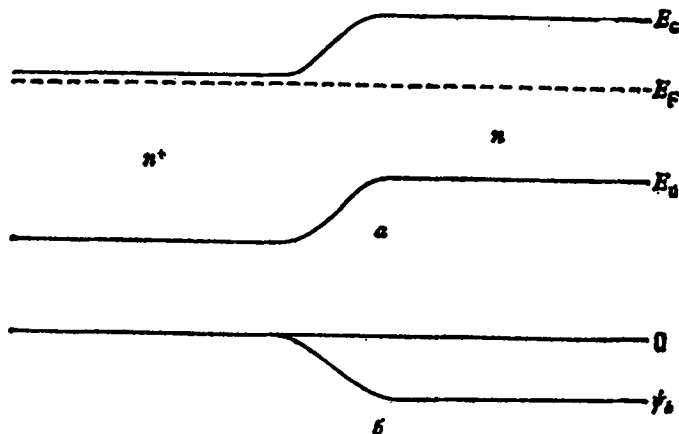


Рис. 7.13. Энергетические зоны и электростатический потенциал в $n^+ - n$ -переходе.

плавно меняется от N_{da} до N_{db} . Следовательно, при наложении внешнего напряжения падение потенциала происходит не только в узкой области, как это в основном имеет место в $p-n$ -переходе. Когда на переход подано положительное напряжение (n^+ -область положительна), то потенциальный барьер в этом случае повышается, а дырки смещаются вправо от перехода. Следовательно, дырочный ток возникает лишь благодаря диффузии дырок в левой части по направлению к переходу. (Напряженность электрического поля в низкоомной n^+ -области оказывается очень малой.) Плотность дырочного тока можно определить точно так же, как и для $p-n$ -перехода, что дает

$$j_h = \frac{e p_a L_{pa}}{\tau_{pa}} = \frac{e n_1^2 L_{pa}}{N_{da} \tau_{pa}}. \quad (7.182)$$

Если N_{da} велико, ток оказывается малым, т. е. у такого перехода $\gamma \approx 0$. Следовательно, переходы этого типа оказываются очень полезными в качестве контактов, когда необходимо избежать инжекции дырок в полупроводник n -типа; такого рода переходы часто используются, например, в качестве базовых контактов диодов.

Когда на n^+-n -переход смещение подано так, что n^+ -область заряжается отрицательно, дырки начнут смещаться в сторону этой области, но дырочный ток не сможет протекать в n^+ -области, если только в ней не существует достаточно высокой концентрации избыточных дырок, которая способствовала бы появлению диффузионного тока. Из рассмотрения эффекта аккумуляции носителей мы уже знаем, каким образом может возникнуть такая большая концентрация у контакта с $\gamma \approx 0$. Если Δp_a — избыточная концентрация дырок слева от перехода, а Δp_b — избыточная концентрация справа от него, то

$$\frac{p_a + \Delta p_a}{p_b + \Delta p_b} = \frac{p_a}{p_b} = \exp\left(\frac{e\psi_b}{kT}\right). \quad (7.183)$$

Поскольку в данном случае на переходе заметного падения напряжения нет, то, следовательно,

$$\frac{\Delta p_a}{\Delta p_b} = \frac{p_a}{p_b} = \frac{N_{db}}{N_{da}}. \quad (7.184)$$

Отсюда плотность дырочного тока j_h равна

$$j_h = - \frac{e \Delta p_a D_{ha}}{L_{pa}} = - \frac{e \Delta p_a L_{pa}}{\tau_{pa}}. \quad (7.185)$$

Таким образом, получим

$$\Delta p_a = |j_h| \frac{\tau_{pa}}{e L_{pa}} \quad (7.186)$$

и

$$\Delta p_b = \frac{|j_h| \tau_{pa}}{e L_{pa}} \frac{N_{da}}{N_{db}}, \quad (7.187)$$

откуда видно, что Δp_b может стать очень большим, если $N_{da} \gg N_{db}$. Если ввести «скорость утечки» v_1 , определенную уравнением (7.134), то получим

$$v_{1a} \Delta p_a = v_{1b} \Delta p_b, \quad (7.188)$$

и тогда из уравнения (7.185) следует, что $v_{1a} = D_{ha}/L_{pa} = (D_{ha}/\tau_{pa})^{1/2}$. Например, если $D_{ha} = 25 \text{ см}^2 \cdot \text{с}^{-1}$ и $\tau_{pa} = 10^{-4} \text{ с}$, то $v_{1a} = 500 \text{ см} \cdot \text{с}^{-1}$. Пусть v_d — скорость дрейфа дырок в электрическом поле по направлению к переходу. Тогда

$$j_h = ev_d p_b = ev_{1a} \Delta p_a = ev_{1a} \Delta p_b \frac{N_{db}}{N_{da}}, \quad (7.189)$$

так что

$$\frac{\Delta p_b}{p_b} = \left(\frac{v_d}{v_{1a}} \right) \left(\frac{N_{da}}{N_{db}} \right). \quad (7.190)$$

Приведенная выше величина D_{ha} соответствует $\mu_h = 10^3 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, так что для $\mathcal{E}_b = 0,5 \text{ В} \cdot \text{см}^{-1}$ имеем $v_d = v_{1a}$. При этом условии $\Delta p_b/p_b = N_{da}/N_{db}$, очевидно, может быть очень большим.

Полная математическая теория $n^+ - n^-$ и $p^+ - p^-$ переходов разработана Ганном [4].

7.11. Поверхностные свойства полупроводников

Рассмотрим теперь свойства контактов между металлами и полупроводниками. Как выяснилось, свойства таких контактов сильно зависят от условий на поверхности полупроводника. Это не является неожиданным для прижимных контактов, поскольку в этом случае очевидно влияние на свойства контакта поверхностных оксидных слоев и других загрязнений поверхности. Роль поверхностных свойств, однако, оказалась более существенной для всех контактов между металлами и полупроводниками, чем для контактов между двумя металлами. Они оказывают заметное влияние на выпрямительные и инжекционные свойства контакта. Когда мы приступим к рассмотрению сварных и плавных контактов, то увидим, что положение еще более осложняется, так как процесс нанесения контакта почти всегда сопровождается изменением свойств полупроводника в окрестности контакта. Наиболее контролируемые контактами являются $p^+ - n^-$ или $n^+ - n^-$ переходы, поскольку они являются частью монокристалла.

Таким образом, прежде чем приступить к рассмотрению контакта между металлом и полупроводником, необходимо уделить некоторое внимание поверхностным свойствам полупроводников. Наиболее важным свойством поверхности является то, что она обычно вызывает появление поверхностных энергетических уровней, рас-

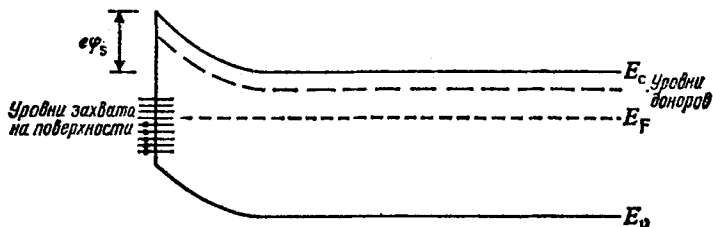


Рис. 7.14. Влияние ловушек для электронов на поверхности полупроводника *n*-типа.

положенных в зоне запрещенных энергий между валентной зоной и зоной проводимости. Такие поверхностные состояния, предсказанные Таммом [22], существуют даже в идеальном кристалле и обусловлены наличием границы кристалла. Однако на практике имеют дело с поверхностными состояниями, возникшими главным образом вследствие адсорбции примесей на поверхности. Такие электроотрицательные примеси, как, например, кислород, ведут себя как акцепторы электронов. Имеются также многочисленные доказательства того, что на поверхности могут находиться центры рекомбинации и прилипания для дырок. Энергетические уровни всех этих поверхностных состояний обычно располагаются значительно ниже зоны проводимости или выше валентной зоны и могут даже оказаться распределенными по запрещенной зоне.

Теоретическое исследование таких поверхностных состояний, а также влияния адсорбированных на поверхности полупроводника примесей было проведено рядом авторов. Обширный обзор работ, посвященных этому вопросу, дали Дэвисон и Левин [23]. Они обсудили различные методы исследований изменений кристаллической структуры, обусловленных наличием поверхности.

Рассмотрим, что произойдет в полупроводнике *n*-типа, донорные уровни которого расположены в непосредственной близости от края зоны проводимости, если на поверхности его имеется ряд акцепторных состояний, лежащих ниже уровня Ферми в объеме материала. Эта ситуация изображена на рис. 7.14. Если температура достаточно низка, так что электроны из зоны проводимости или с донорных уровней оказываются не полностью ионизованными, то они стремятся заполнить те поверхностные состояния, которые находятся ниже уровня Ферми. Пусть N_s — число соответствующих поверхностных состояний на единицу поверхности, расположенных ниже уровня Ферми. Тогда на каждой единице площади поверхности возникнет отрицательный заряд величиной $-eN_s$. Как было уже показано в случае *p-n*-перехода, это вызовет появление обедненного носителями слоя, из которого удалены почти все электроны и который будет обладать положительным объемным зарядом, приблизительно рав-

ным eN_d , где N_d — концентрация доноров. В результате появится сильное электрическое поле, направленное к поверхности, препятствующее дальнейшему переходу электронов на поверхности уровня. В конце концов энергетические уровни на поверхности полупроводника поднимутся, как показано на рис. 7.14. Если величину подъема зоны проводимости на поверхности обозначить через $e\varphi_s$, то φ_s — поверхностный потенциал. Для вычисления φ_s необходимо решить уравнение Пуассона с соответствующими граничными условиями. Здесь, однако, будет дана приближенная теория, основанная, как и в случае теории $p-n$ -перехода, на предположении, что в области от $x=0$ (поверхность) до $x=d$, где $d=N_s/N_d$, объемный заряд ρ равен eN_d . Далее решается уравнение Пуассона с граничными условиями $\mathcal{E}=0$ при $x=d$ и $\mathcal{E}=-eN_s/\epsilon\epsilon_0$ при $x=0$. В результате получаем

$$\mathcal{E} = -\frac{eN_d}{\epsilon\epsilon_0}(d-x) = -\frac{eN_s}{\epsilon\epsilon_0 d}(d-x). \quad (7.191)$$

Если при $x=d$ положить $\psi=0$, получим уравнение для электростатического потенциала ψ :

$$\psi = -\frac{eN_d}{2\epsilon\epsilon_0}(d-x)^2. \quad (7.192)$$

Поверхностный потенциал φ_s , очевидно, равен величине $-\psi$ при $x=0$. Поэтому

$$\varphi_s = \frac{eN_s d}{2\epsilon\epsilon_0} = \frac{eN_s^2}{2\epsilon\epsilon_0 N_d}. \quad (7.193)$$

Энергия E_s , на которую сместятся энергетические зоны на поверхности, определяется уравнением

$$E_s = e\varphi_s = \frac{e^2 N_s^2}{2\epsilon\epsilon_0 N_d}. \quad (7.194)$$

Например, если $\epsilon=10$, $N_d=2 \cdot 10^{21} \text{ м}^{-3}$ и $N_s=10^{16} \text{ м}^{-2}=10^{11} \text{ см}^{-2}$, то получим $\varphi_s \approx 0,5 \text{ В}$.

Более строгая теория поверхностного барьера, в которой распределение объемного заряда вблизи поверхности определяется с помощью уравнения Пуассона, разработана Кингстоном и Нейштадтером [24]. Они предполагали, что в области с проводимостью n -типа все доноры полностью ионизованы и что при определении концентрации электронов можно пользоваться уравнением Больцмана. Результаты исследования представлены ими в численном виде.

Если поверхностные состояния способны захватывать дырки, то в результате энергия электрона, соответствующая краю зоны, *понижится*. Для объяснения поверхностных эффектов, обнаруженных после выдерживания полупроводника в различного рода газовых средах, необходимо учитывать захват и электронов, и дырок.

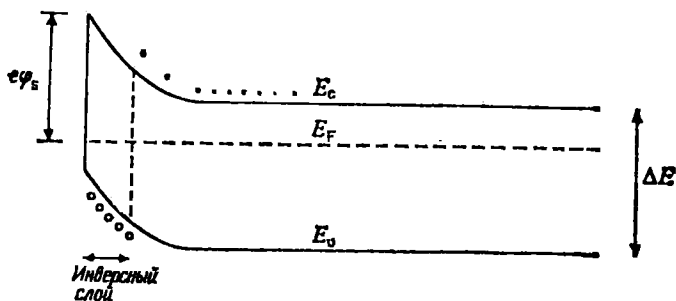


Рис. 7.15. Инверсный слой p -типа на поверхности полупроводника n -типа.

В случае полупроводника p -типа зоны также будут изгибаться вниз, если на поверхности захватываются дырки, и изгибаться вверх, если захватываются электроны. Так как свойства поверхности, помещаемой в газové среды, очень трудно контролировать, то часто изучают ее, помещая в раствор электролита. Браттейн и Гарретт [25] провели таким способом соответствующие исследования германия и показали, что при помощи соответствующих электролитов можно получить поверхности как p -, так и n -типа. Браттейн и Бардин [26], используя классический метод Кельвина, исследовали контактный потенциал поверхности германия. Этим способом удалось легко наблюдать изменения величины ϕ_s . Изучались изменения ϕ_s , вызванные атмосферой, и было обнаружено, что они очень велики, причем при циклическом изменении состава атмосферы контактный потенциал последовательно менялся от низких величин до относительно высоких. Изучалось также влияние освещения поверхности.

Если в случае полупроводника n -типа поверхностный потенциал ϕ_s очень велик, то вблизи поверхности изгиб валентной зоны и зоны проводимости может быть настолько значительным, что валентная зона окажется ближе к уровню Ферми, чем зона проводимости. Тогда вблизи поверхности образуется слой с преобладающей проводимостью p -типа. Если валентная зона окажется при этом очень близко от уровня Ферми, то концентрация дырок может стать весьма высокой. Слои такого типа называют инверсными; энергетическая схема такого слоя изображена на рис. 7.15. Подобные слои играют важную роль в технологии полупроводников. Например, слой p -типа на поверхности образца с p — n -переходом существенно образом меняет характеристики перехода.

Интересно отметить, что именно изучение Бардином [27] поверхностных состояний привело к открытию транзисторов. Бардин пришел к выводу, что для объяснения того факта, что выпрямляющие свойства контакта металл — полупроводник почти не зависят от

свойств металла, но очень сильно зависят от обработки поверхности полупроводника, необходимо допустить существование поверхностных состояний. Он показал также, что подобные состояния объясняют результаты, полученные при исследовании явления, известного под названием эффекта поля, которое будет рассмотрено ниже. Эти соображения привели к представлению об инверсном слое, а затем и к понятию об инжекции дырок.

7.11.1. ЭФФЕКТ ПОЛЯ

Если одной из обкладок плоскопараллельного конденсатора является пластинка из полупроводника n -типа, а другой — металлическая пластинка и если металлическая пластинка заряжена положительно, то полупроводник должен зарядиться отрицательно, т. е. в нем должны появиться избыточные электроны. Можно ожидать, что эти электроны, как и электроны, нормально заполняющие зону проводимости, будут принимать участие в проводимости. Таким образом, при зарядке конденсатора должно наблюдаться увеличение проводимости полупроводника. Эксперимент такого рода был осуществлен в 1948 г. Шокли и Пирсоном [28] в лаборатории фирмы «Белл телефон» (США) и оказался решающим шагом на пути к изобретению транзистора¹⁾. Искомый эффект был обнаружен, но его величина оказалась на порядок меньше ожидаемой. Чтобы объяснить это, Бардин предположил, что 90% избыточных электронов захватываются поверхностными состояниями и поэтому лишены возможности участвовать в проводимости²⁾.

Эффект поля стал мощным средством исследования поверхностных свойств полупроводников. Установлено, что изменение проводимости при наложении перпендикулярного поверхности электрического поля сильно зависит от поверхностного потенциала ϕ_s и может быть как положительным, так и отрицательным. Более того, установлено, что это изменение не остается постоянным во времени; хотя поле и не выключается, но проводимость постепенно, с постоянной времени τ_f возвращается к первоначальной величине. Большая часть опубликованных работ, посвященных этому эффекту, относится преимущественно к исследованию поверхности кристаллов германия и кремния. Исчерпывающий анализ использования эффекта поля для определения различных параметров поверхности дан Лоу [29]. Ряд обзорных статей, относящихся к данной проблеме, вышел под редакцией Кингстона [30]. Обзор результатов исследований эффекта поля составил также Годфрой [31]. Основной результат этих исследований сводится к тому, что, по-видимому, бывает два

¹⁾ Эти эксперименты имели решающее значение для открытия полевого транзистора.— *Прим. ред.*

²⁾ Интересные сведения о роли, которую сыграл эффект поля на ранней стадии исследования транзисторов, приведены в § 2.1 книги Шокли [7].

разных типа поверхностных состояний полупроводников; эти состояния характеризуются совершенно различными значениями τ_f . Состояния, известные под названием «медленных», характеризуются величинами τ_f , лежащими в пределах от нескольких секунд до многих минут; они локализованы, по-видимому, на внешней поверхности слоя окисла полупроводника, и большая величина τ_f обусловлена временем, затрачиваемым электроном на прохождение через этот слой¹⁾. Эти состояния подробно исследовал Моррисон (см. [30], стр. 169). Другому типу состояний, известных под названием «быстрых», соответствуют времена τ_f в пределах от нескольких микросекунд до нескольких миллисекунд. Состояния локализованы, по-видимому, на границе раздела между слоем окисла и полупроводником. Именно эти состояния главным образом и ответственны за рекомбинацию электронов и дырок на поверхности²⁾ (см. разд. 9.8). Электрохимия поверхности и влияние травления на поверхностные состояния рассматривались в ряде работ различных авторов [34].

Инверсные слои на поверхности полупроводников играют важную роль, поскольку определяют поведение поверхности в сильных электрических полях вплоть до пробоя. Этот вопрос изучали Гарретт и Браттейн [35], которые показали, что пробой обусловлен процессами лавинного размножения носителей заряда и что он сильно зависит от величины поверхностного потенциала.

Электрические свойства поверхностей полупроводников, в том числе и эффект поля, являющийся физической основой работы одного из наиболее широко используемых типов транзисторов — полевого транзистора, рассмотрены в обстоятельном обзоре Уоткинса [36].

7.12. Контакты металл — полупроводник

После всего сказанного выше не будет неожиданным то обстоятельство, что поверхностные состояния заметно влияют на свойства контактов металл — полупроводник, особенно на свойства контактов с малой площадью. Интересно, однако, выяснить, что следовало бы ожидать в случае контакта между металлом и полупроводником в отсутствие таких состояний. Если работа выхода W_s примесного полупроводника n -типа, донорные уровни у которого лежат вблизи дна зоны проводимости, меньше работы выхода W_m металла, образующего контакт, то электроны начнут переходить в металл. Это иллюстрируется рис. 7.16. По-видимому, эти условия обычно и реализуются, поэтому следовало бы ожидать, что вблизи контакта энергетические уровни полупроводника, как показано на рис. 7.16, изог-

¹⁾ См., например, [32].

²⁾ См., например, [33].

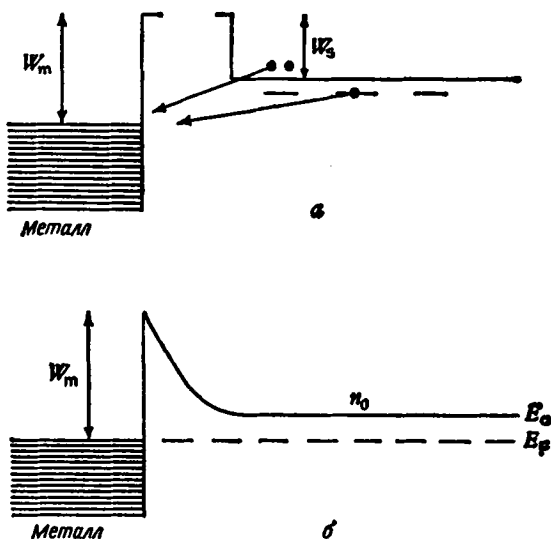


Рис. 7.16. Контакт металла с полупроводником.

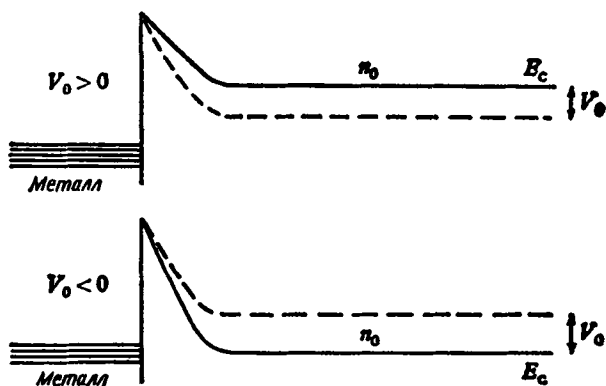


Рис. 7.17. Влияние внешнего напряжения на контакт металл — полупроводник.

нутя вверх. Величина подъема уровней должна была бы определяться разностью работ выхода металла и полупроводника. Можно, как и раньше, ввести поверхностный потенциал ϕ_s , и тогда величина ϕ_s равнялась бы примерно $W_m - W_s$. Однако экспериментально показано, что это не так и что величина ϕ_s почти не зависит от W_m . Таким образом, значение ϕ_s определяется в основном поверхностными состояниями и почти не зависит от природы металла. Следовательно, в изолирующем слое на поверхности должно суще-

ствовать электрическое поле, обуславливающее смещение уровней Ферми в металле и полупроводнике до тех пор, пока они не совпадут.

Наложение потенциала, сообщающего металлу положительный относительно полупроводника (n -типа) заряд, приводит к снижению потенциального барьера, препятствующего переходу электронов из полупроводника в металл, и электроны легко проникают в металл [37, 38]. (В металле имеет место очень небольшое падение потенциала.) Если же металл заряжен отрицательно относительно полупроводника, то высота барьера, который должны преодолеть электроны для перехода в металл, повышается, в соответствии с чем этот поток по существу прекращается; остается лишь небольшой поток электронов из металла в полупроводник. Количественно это можно представить следующим образом. Пусть $-j_s$ — ток электронов из металла в полупроводник. Он не зависит от приложенного напряжения и определяется лишь высотой потенциального барьера относительно уровня Ферми. Если предположить, что падение потенциала V_0 происходит в основном лишь в области объемного заряда (рис. 7.17), то эффективная высота барьера окажется равной $e(\varphi_s - V_0)$, и тогда плотность электронного тока из полупроводника в металл запишется следующим образом:

$$-j_e = j_0 \exp \left[-\frac{e(\varphi_s - V_0)}{kT} \right].$$

При $V_0=0$ имеем $j_e=j_s$, так что

$$-j_s = j_0 \exp \left(-\frac{e\varphi_s}{kT} \right). \quad (7.195)$$

Таким образом, плотность полного тока j определяется выражением

$$j = -j_s + j_e = j_0 \left[\exp \left(\frac{V_0}{kT} \right) - 1 \right], \quad (7.196)$$

которое имеет тот же вид, что и для $p-n$ -перехода, однако величину тока насыщения j_0 теперь найти не столь просто.

Существует ряд теорий контакта, позволяющих вычислить j_0 . Условия их применимости и суть сделанных в них приближений проанализировал Хениш [21]. Основное отличие этих теорий друг от друга связано с шириной слоя пространственного заряда и вероятностью столкновений электронов и дырок при прохождении через этот слой. Если слой тонкий, то применима так называемая диодная теория, согласно которой

$$j_0 = ne \left(\frac{kT}{2\pi m_e} \right)^{1/2}, \quad (7.197)$$

где m_e — эффективная масса электрона. Если процесс изготовления контакта металл — полупроводник не выполнен с большой тща-

тельностью в условиях высокого вакуума, то неизбежно образуется тонкий слой окисла и получается контакт типа металл — окисел — полупроводник (МОП). Это оказывает заметное влияние на вольтамперную характеристику. Эта теория была в дальнейшем развита в направлении учета роли слоя окисла и сил изображения для электронов в полупроводнике, а также учета роли других факторов. Рассматривались также туннельные переходы электронов сквозь потенциальный барьер.

Современная теория выпрямления, учитывающая свойства контактов металл — полупроводник, показывает, что задача расчета вольтамперной характеристики подобного выпрямляющего контакта чрезвычайно сложна. Например, в рассмотренной выше упрощенной теории предполагалось, что неосновные носители заряда не играют существенной роли. С другой стороны, нам известно, что могут возникать инверсные слои и что некоторые контакты металл — полупроводник (p -типа) обладают большой величиной коэффициента инжекции γ . Лишь в очень редких случаях вольтамперные характеристики контактов металл — полупроводник описываются уравнением вида (7.196). Довольно часто эти характеристики можно описать уравнением такого типа, но с заменой e/kT на $\beta e/kT$, где множитель β обычно весьма заметно отличается от единицы. Это осложнение находится в резком контрасте с изящной простотой математического описания выпрямителей с p — n -переходами. По этой причине при создании вплавных и сварных контактов металл — полупроводник стремятся получить их в виде p^+ — n - или n^+ — n -переходов. Обычно мы имеем дело с контактами типа металл — p^+ — n или металл — n^+ — n , причем контактирующие с металлом p^+ - или n^+ -области обладают очень низким сопротивлением. Такие контакты получают путем сплавления металла с небольшой присадкой элементов III или V групп. Например, если необходимо, чтобы контакт золотой проволоочки и германия n -типа имел большую величину γ , то используется сплав золота с In или Ga; в процессе сплавления или вжигания в Ge золотой проволоочки происходит диффузия In или Ga, которые, проникая на небольшое расстояние в германий, образуют p^+ — n -переход. В этом случае, как известно, $\gamma \approx 1$. Если же требуется получить неинжектирующий контакт, тогда для этой цели можно использовать сплав золота с сурьмой, который дает n^+ — n -переход с коэффициентом инжекции $\gamma \approx 0$. Многие из обследованных в прошлом контактов металл — полупроводник были непреднамеренно получены именно этим способом, и поэтому теперь благодаря теории p^+ — p - и n^+ — n -переходов их свойства стали более понятными. Контакты этого типа могут быть получены с вольтамперной характеристикой, близкой к линейной; выпрямительные свойства очень слабы, либо вовсе отсутствуют. Такие контакты называют омическими. Для выпрямителя, в частности, один контакт должен быть выпрямляющим, а другой — омическим [39].

Большинство ранних работ по выпрямляющим контактам было выполнено на Si и Ge. В дальнейшем исследовались контакты различных металлов с целым рядом полупроводников, в том числе с некоторыми тройными соединениями [40], такими, как AgGaSe_2 . Были обнаружены некоторые интересные отличия их свойств от контактов на Si и Ge, но и для них вольтамперные характеристики очень хорошо описывались уравнением (7.196). Хотя во многих случаях характеристики выпрямляющих контактов слабо зависят от вида металла, который используется для их изготовления, так как большее влияние оказывают поверхностные состояния, существуют и некоторые исключения. Хейне [41] показал, что зонная структура металла в некоторых случаях может играть заметную роль. Это существенно, в частности, для переходных металлов, таких, как Ni. Сведения о некоторых более поздних работах, посвященных исследованиям контактов типа полупроводник — металл, приведены в сообщении конференции, посвященной этой проблеме [42].

В первые годы после открытия транзисторов, когда широко применялись подобные контакты, велись интенсивные исследования явления инжекции неосновных носителей заряда контактами металл — полупроводник. Методы измерения коэффициента инжекции γ предложили Шокли, Пирсон и Хейнс [43], а также Вальдес [44]. Теории инжекции и выпрямления контактами металл — полупроводник разработаны Ганном [45] и Бэнбери [46]. Экспериментальное исследование зависимости γ от поверхностных свойств и от освещенности провели Хогарт [47] (контакты с малой площадью) и Хениш [48] (контакты с большой площадью).

Если предположить, что γ равно отношению $p/(nb+p)$, где n и p — концентрации носителей заряда вблизи контакта, то, используя величины n и p , вычисленные с помощью уравнения Больцмана, легко показать, что

$$\gamma = \frac{1}{1 + b \exp\left(\frac{\Delta E - 2e\varphi_s}{kT}\right)}. \quad (7.198)$$

Выражение этого типа было получено Хогартом [47]; из него должно следовать, что γ всегда близко к 1 или 0, за исключением лишь ограниченной области значений φ_s . Аналогичное выражение, полученное Хенишем (см. [48], стр. 226), содержит коэффициент диффузии и время жизни неосновных носителей заряда. Оно приводит к точно таким же следствиям и, в частности, показывает, что коэффициент γ действительно очень мал, пока $e\varphi_s < \Delta E/2$. Из обеих формул следует слишком сильная зависимость γ от φ_s , однако никакой сильной зависимости γ от поверхностных условий не обнаружено. На практике в случае хороших контактов металл — полупроводник очень редко встречаются значения γ меньше 0,1. Для веществ n -типа обычные значения γ лежат между 0,1 и 0,5, в то время как для полу-

проводников p -типа γ принимает значения от 0,5 до 0,9. Для контакта, инжектирующего дырки, увеличение γ с 0,5 до 1,0 означает четырехкратный рост суммарной концентрации неравновесных носителей при заданном значении тока, поскольку каждая избыточная дырка увлекает за собой по одному избыточному электрону. Таким образом, вплавные контакты в случае использования их для инъекции обладают существенным преимуществом перед простыми контактами металл—полупроводник, поскольку у них, как отмечалось выше, значения коэффициента γ близки к единице. Вопрос о токовой инъекции был рассмотрен Лампертом и Марком [49].

7.13. Дрейфовая подвижность электронов и дырок

Наиболее прямой метод измерения дрейфовой подвижности электронов и дырок основан на рассмотренных в разд. 7.6 опытах Хейнса и Шокли [2], посвященных исследованию явления дрейфа носителей заряда, а также на последующих более совершенных опытах Шокли, Пирсона и Хейнса [43]. Сущность метода состоит в инъекции коротким импульсом напряжения неосновных носителей заряда в тонкий стержень из полупроводника и в смещении образовавшегося облака неосновных носителей электрическим полем вдоль стержня с последующей регистрацией прохождения облака возле коллектора — контакта, чувствительного к концентрации неосновных носителей вблизи него. Одна из наиболее распространенных схем этого опыта показана на рис. 7.18. Прохождение облака возле коллекторного контакта (см. разд. 7.9.2) наблюдается при помощи катодного осциллографа. Если импульс положительного напряжения, создающего электрическое поле \mathcal{E} , приложен к эмиттерному контакту, то почти сразу, точнее спустя время, необходимое для распространения электромагнитного сигнала по стержню, на коллекторном электроде появится аналогичный импульс. Позже, через отрезок времени порядка $d/\mathcal{E}\mu_d$, где d — расстояние между эмиттером и коллектором, у коллектора появится облако неосновных носителей заряда, который приведет к возникновению импульса напряжения в выходной цепи, поскольку дырки изменяют величину коллекторного тока. Первый импульс на коллекторе является точной копией импульса напряжения, поданного на инжектирующий электрод, если, конечно, полоса пропускания частот усилителя измерительной схемы достаточно широка для того, чтобы импульс воспроизводился без заметных искажений. Однако импульс от неосновных носителей заряда, как показано на рис. 7.18, несколько сгладится из-за их диффузии. Время t между серединами обоих импульсов, равное $d/\mathcal{E}\mu_d$, дает возможность непосредственно измерить $\mu_d \equiv \mu$, при этом предполагается, что $\mathcal{E} \gg \mathcal{E}_c$ (см. разд. 7.5.5).

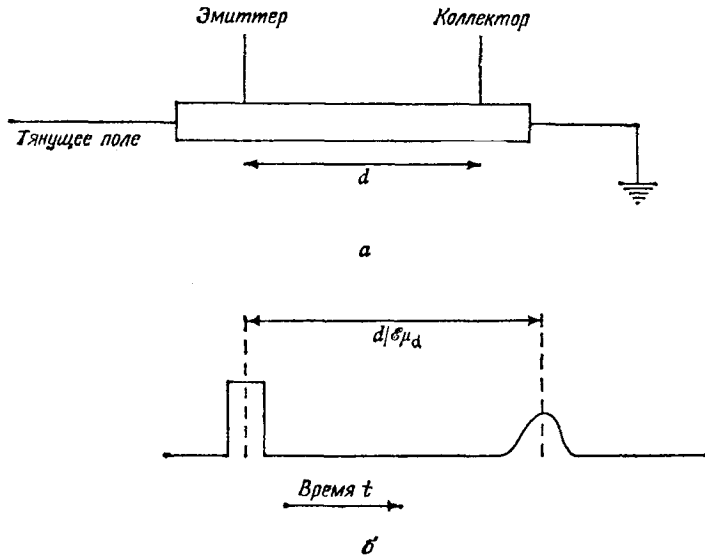


Рис. 7.18. Схема опыта для измерения дрейфовой подвижности (а) и форма сигнала на коллекторе (б).

В германии n -типа $\mu_{dh} = 1900 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, так что при $\mathcal{E} = 10 \text{ В} \cdot \text{см}^{-1}$ (что является типичной величиной в подобного рода экспериментах) получим $t = 30 \text{ мкс}$. Если время жизни неосновных носителей намного превышает это значение, то уменьшение концентрации избыточных носителей из-за рекомбинации будет незначительным. Если, как это обычно и делают, использовать короткие импульсы длительностью порядка 1 мкс, то оба импульса на коллекторе будут достаточно четко разделены во времени. В случае же более длительных импульсов, которые использовались в первых опытах, на коллекторе будет наблюдаться импульс более сложной формы. Необходимо проанализировать форму этого импульса, чтобы иметь возможность определить t — время движения облака неосновных носителей заряда.

Если число инжектированных с эмиттера дырок велико, то поле, в котором движется импульс, не будет однородным и импульс приобретет несимметричную форму. Это может привести к ошибке в определении момента прихода облака к коллектору. Однако, используя короткие импульсы и низкие концентрации неосновных носителей, можно получить достаточно точные значения величины амбиполярной дрейфовой подвижности. Как было указано в разд. 7:7, в веществах с высокой концентрацией примесей амбиполярная дрейфовая подвижность μ совпадает по величине с дрейфовой подвиж-

ностью, но может заметно отличаться от нее у образцов с почти собственной проводимостью. Варьируя содержание примесей в веществе образца, можно этим способом определить μ_e и μ_h , а также их зависимость от температуры и концентрации примесей.

7.14. Гетеропереходы

Все возрастающее значение в технике, особенно с расширением области применения и развитием оптоэлектронных приборов — фотодиодов и лазеров (см. разд. 14.6), приобретает тип полупроводникового перехода, называемый гетеропереходом. Это переход между двумя различными полупроводниками. Ими могут быть совершенно различные материалы, такие, например, как Ge и GaAs, хотя существуют трудности в изготовлении перехода, если параметры кристаллической решетки для этих материалов сильно отличаются. Чаще один из материалов представляет собой полупроводниковое соединение, например GaAs, а другой — его сплав с родственным соединением, таким, как $\text{GaAs}_x\text{P}_{1-x}$. Последнее имеет более широкую запрещенную зону, чем первое, и это является характерной особенностью гетероперехода. Свойства гетеропереходов впервые были изучены Андерсоном [50], который оценил потенциальные возможности их технического применения и описал основные принципы работы гетеропереходов.

Существует два различных типа гетеропереходов. В первом один полупроводник — p -типа, а другой — n -типа, таким образом получается увеличение барьера p — n -перехода. Во втором оба материала n -типа или оба p -типа. Такие переходы в какой-то мере сходны с n^+ — n - и p^+ — p -переходами.

Рассмотрим сначала p — n -гетеропереход. Предположим, что материал p -типа имеет более широкую запрещенную зону, чем материал n -типа, т. е. $\Delta E_p > \Delta E_n$. Поэтому вид энергетических зон должен быть изменен по сравнению с тем, что показано на рис. 7.8. Вид энергетических зон гетероперехода в случае равновесия, когда энергия уровня Ферми одинакова по обе стороны перехода, показан на рис. 7.19, а. На рис. 7.19, б показано также влияние напряжения V_0 , приложенного в «прямом» направлении, т. е. так, чтобы дырки двигались слева направо. Можно видеть, что потенциальный барьер, который дырки должны преодолеть при продвижении из левой от p — n -перехода части кристалла в правую, уменьшается, и если V_0 достаточно велико, барьер может быть понижен до нуля, как и для обычного p — n -перехода, так что дырочный ток в прямом направлении сильно возрастает. С другой стороны, потенциальный барьер, который должны преодолеть электроны при движении из правой от перехода части кристалла в левую, останется еще существенным, если ΔE_p заметно выше, чем ΔE_n . Следовательно, элект-

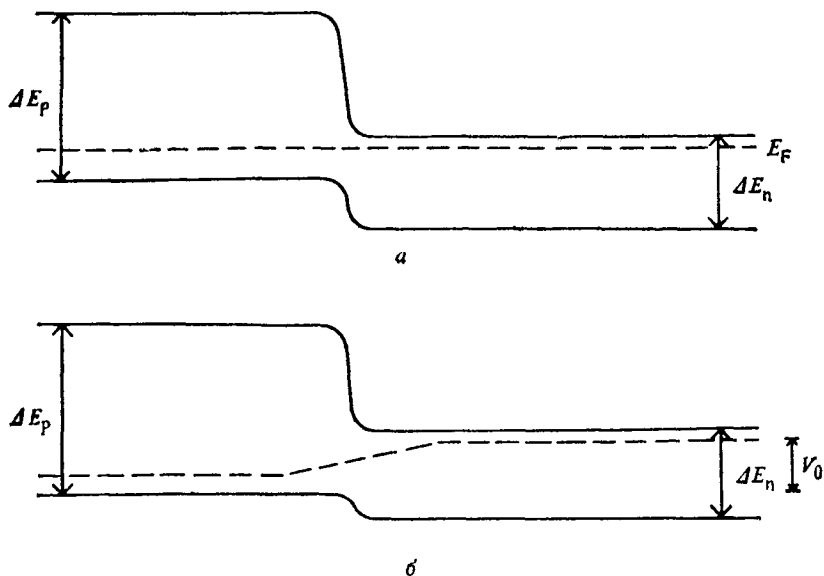


Рис. 7.19. Энергетические зоны идеального p — n -гетероперехода, находящегося в термодинамическом равновесии (а) и при наложении внешнего напряжения (б).

ронный ток будет довольно мал, даже если дырочный ток велик. Такой переход ведет себя подобно p^+ — n -переходу с $\gamma \approx 1$.

Одним из важных применений p — n -гетеропереходов является их использование в качестве эффективных инжекторов дырок в материал n -типа, что устраняет необходимость в сильно легированной области p -типа. При использовании переходов в приборах, где желательно вывести рекомбинационное излучение, возникающее в результате инжекции дырок, гетеропереходы имеют ряд преимуществ. Так, область p -типа, имеющая более широкую запрещенную зону, прозрачна для рекомбинационного излучения из области n -типа. Отсутствует здесь и поглощение свободными носителями, которое имел бы место в сильно легированной p^+ -области. p - и n -области перехода могут, конечно, поменяться ролями, образовав переход, инжектирующий электроны, с $\gamma \approx 0$. Как и следовало ожидать, вольтамперная характеристика p — n -гетероперехода в прямом направлении очень похожа на вольтамперную характеристику обычного p — n -перехода. Она описывается как функция V_0 уравнением вида (7.159), но с модифицированным некоторым образом значением j_s . Однако ток в обратном направлении, хотя он и много меньше «прямого» тока, не насыщается, как в обычном p — n -переходе, но продолжает расти при увеличении $-V_0$. Причина этого отличия заключается главным образом в свойствах переходного слоя, соеди-

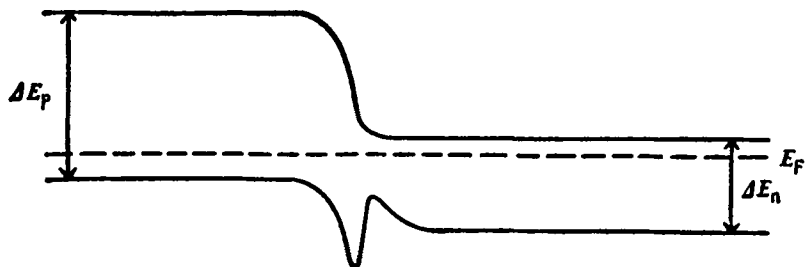


Рис. 7.20. Вид энергетических зон p — n -гетероперехода, показывающий влияние ловушек в переходной области.

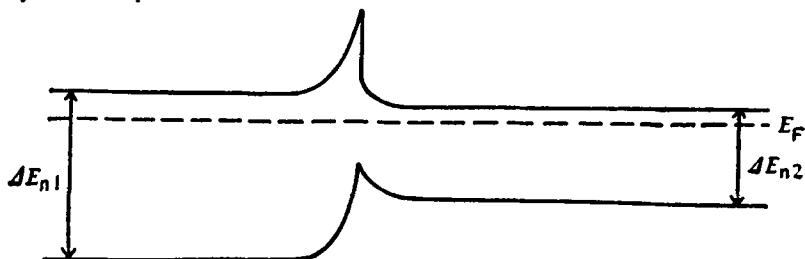


Рис. 7.21. Вид энергетических зон n — n -гетероперехода при наличии ловушек в переходной области.

нящего два полупроводника в гетеропереход. Из-за различия в величине постоянной решетки в переходном слое будет существовать некоторое механическое напряжение, которое вызывает появление в этом слое ловушек как для дырок, так и для электронов, подобно тому, как это имеет место на поверхности полупроводника (см. разд. 7.11). Эти ловушки удерживают уровень Ферми в переходном слое вблизи середины между валентной зоной и зоной проводимости обоих полупроводников и вызывают подъем зон со стороны n -типа и их опускание со стороны p -типа, так же как в случае контакта металл — полупроводник (см. разд. 7.12). В результате вид энергетических зон изменится по сравнению с тем, что показано на рис. 7.19, а, и станет таким, как на рис. 7.20.

Влияние состояний в переходном слое обсуждалось Олдохэмом и Милнсом [51]. Свойства переходов из GeGaAs, обусловленные наличием этих состояний, описаны Перлманом и Фойхтом [52]. На рис. 7.20 можно видеть, что при движении как в «прямом», так и в «обратном» направлении носители заряда должны преодолеть потенциальный барьер. Считается, что ток через барьер обусловлен главным образом туннелированием дырок. Это подтверждается тем, что температурная зависимость величины j_s , найденной из «прямой» ветви вольтамперной характеристики, оказывается слабой (см. разд. 14.4.3).

Вид энергетических зон для n - n -гетероперехода показан на рис. 7.21. Здесь мы учли влияние состояний в переходном слое. Можно видеть, что для электронов существует барьер типа Шоттки, подобный барьеру контакта металл — полупроводник, но обычно более сложный по форме. Действительно, край валентной зоны может иметь подъем с последующим спадом. Характеристики таких переходов обсуждались Хинкли, Редикером и Ядусом [53], которые показали, что прямая ветвь вольтамперной характеристики дается уравнением вида

$$i = AT^2 \exp\left(-\frac{e\phi}{kT}\right) \left[\exp\left(\frac{\beta e V_0}{kT}\right) - 1 \right], \quad (7.199)$$

где A' — константа, зависящая от материала перехода, ϕ — потенциал, также зависящий от материала, и β — константа порядка единицы. Это уравнение очень похоже на уравнение (7.196) для контакта металл — полупроводник. Переходы такого типа между GaAs и InSb, характеристики которых измерял Хинкли с сотрудниками [53], обнаруживают очень хорошее выпрямление порядка 10^8 . Найдено, что $\phi \approx 0,8$ В; это сходно со значениями, которые получаются для контактов металл — полупроводник. Насыщения обратного тока снова не наблюдалось.

Вопрос о гетеропереходах подробно рассмотрен Тансли [54]. Он описал различные сложные формы энергетических зон, которые могут получаться из-за влияния различного вида состояний в переходном слое. Он исследовал также большое число комбинаций полупроводников, чтобы выявить наилучшие с точки зрения соответствия параметров кристаллической решетки. Несмотря на то что для получения более совершенной структуры с возможно меньшим влиянием «переходных» состояний такой подбор, несомненно, важен, хорошо работают и некоторые переходы с большим рассогласованием параметров решетки, что явилось неожиданным. Хинкли с сотрудниками использовали, например, GaAs и InSb, которые отличаются по величине постоянной решетки на 18%, а по ширине запрещенной зоны в 8 раз (1,4/0,18).

ЛИТЕРАТУРА

1. *Smith R. A.* The Wave Mechanics of Crystalline Solids, Chapman and Hall, 1969.
2. *Haynes J. R., Shockley W.*, Phys. Rev., 75, 691 (1949).
3. *Van Roosbroeck W.*, Bell Syst. Techn. Journ., 29, 560 (1950); Phys. Rev., 91, 282 (1953).
4. *Gunn J. B.*, Journ. Electron. Control, 4, 17 (1958).
5. *Rittner E. S.*, Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 215.
6. *Herring C.*, Bell Syst. Techn. Journ., 28, 401 (1949).
7. *Shockley W.*, Electrons and Holes in Semiconductors, Van Nostrand, 1950. (Имеется перевод: Шокли В. Теория электронных полупроводников.— М.: ИЛ, 1953.)
8. *Gibson A. F.*, Physica, 20, 1058 (1954).

9. Arthur J. B., Bardsley W., Brown M. A. C. S., Gibson A. F., Proc. Phys. Soc., **B68**, 43 (1955).
10. Bray R., Phys. Rev., **100**, 1047 (1955).
11. Read W. T., Bell Syst. Techn. Journ., **35**, 1239 (1956).
12. Arthur J. B., Gibson A. F., Gunn J. B., Proc. Phys. Soc., **B69**, 697 (1956).
13. Low G. G. E., Proc. Phys. Soc., **B68**, 310 (1955).
14. Harrick N. J., Phys. Rev., **101**, 491 (1956); **103**, 1173 (1956).
15. Давыдов Б.—ЖТФ, **5**, 7987 (1938).
16. Benzer S., Phys. Rev., **72**, 1267 (1947).
17. Becker M., Fan H. Y., Phys. Rev., **75**, 1631 (1949).
18. Sosnowski L., Starkiewicz J., Simpson O., Nature, **159**, 818 (1947).
19. Shockley W., Bell Syst. Techn. Journ., **28**, 435 (1949).
20. Elliott R. J., Gibson A. F., Solid State Physics, Macmillan, 1974.
21. Henisch H. K., Rectifying Semiconductor Contacts, Oxford University Press, 1957.
22. Тамм И. Е., Zs. Phys., **76**, 849 (1932).
23. Davison S. G., Levine J. D., Solid State Physics, Academic Press, 1970, **25**, p. 1.
24. Kingston R. H., Neustadter S. F., Journ. Appl. Phys., **26**, 718 (1955).
25. Brattain W. H., Garrett C. G. B., Physica, **20**, 885 (1954); Bell Syst. Techn. Journ., **34**, 129 (1955).
26. Brattain W. H., Bardeen J., Bell Syst. Techn. Journ., **32**, 1 (1953).
27. Bardeen J., Phys. Rev., **71**, 717 (1947).
28. Shockley W., Pearson G. L., Phys. Rev., **74**, 223 (1948).
29. Low G. G. E., Proc. Phys. Soc., **B69**, 1331 (1956).
30. Semiconductor Surface Physics, ed. Kingston R. H., Pennsylvania University Press, Oxford University Press, 1957.
31. Godefroy L. R., Progress in Semiconductors, Heywood, **1**, 1956, p. 195.
32. Lasser M., Wysocki C., Bernstein B., Phys. Rev., **105**, 491 (1957).
33. Many A., Gerlich D., Phys. Rev., **107**, 404 (1957).
34. The Electrochemistry of Semiconductors, ed. Holmes P. J., Academic Press, 1973.
35. Garrett C. G. B., Brattain W. H., Journ. Appl. Phys., **27**, 299 (1956).
36. Watkins T. B., Progress in Semiconductors, Heywood, **5**, 1960, p. 1.
37. Schottky W., Hartmann H., Zs. Techn. Phys., **16**, 512 (1935).
38. Schottky W., Naturwiss., **26**, 843 (1938).
39. Ohmic Contacts to Semiconductors, ed. Schwarz B., Electrochem. Soc., N. Y., 1969.
40. Robinson P., Wilson J. I. B., Ternary Compounds, ed. Holah G., Inst. of Phys., London, 1977, p. 229.
41. Heine V., Phys. Rev., **138**, A1, 689 (1965).
42. Metal-Semiconductors Contacts, ed. Pepper M., Inst. Phys. Conf. Series, London, 1974.
43. Shockley W., Pearson G. L., Haynes J. R., Bell Syst. Techn. Journ., **28**, 344 (1949).
44. Valdes B., Proc. IRE, **40**, 1420 (1952).
45. Gunn J. B., Proc. Phys. Soc., **B67**, 575 (1954).
46. Banbury P. C., Proc. Phys. Soc., **B66**, 833 (1953).
47. Hogarth C. A., Proc. Phys. Soc., **B66**, 845 (1953).
48. Henisch H. K., Proc. Phys. Soc., **B66**, 841 (1953).
49. Lampert M. A., Mark P., Current Injection in Solids, Academic Press, 1970. [Имеется перевод: Ламперт М., Марк П. Инжекционные токи в твердых телах.— М.: Мир, 1973.]
50. Anderson R. L., Solid State Electron., **5**, 341 (1962).
51. Oldham W. G., Milnes A. G., Solid State Electron., **7**, 153 (1964).
52. Perlman S., Feucht D. L., Solid State Electron., **7**, 911 (1964).
53. Hinkley E. D., Rediker R. H., Jadus D. K., Appl. Phys. Lett., **6**, 144 (1965).
54. Tansley T. L., Semiconductors and Semimetals, Academic Press, **7A**, 1971, p. 293.

РАССЕЯНИЕ ЭЛЕКТРОНОВ И ДЫРОК

8.1. Изменение состояния

В разд. 5.1 были введены времена релаксации τ_e и τ_h , которые определяют скорость изменения волновых векторов электронов и дырок и, следовательно, их подвижность в электрическом поле. В случае скалярной эффективной массы m_e , m_h величина подвижности дается простыми выражениями $e\langle\tau_e\rangle/m_e$ или $e\langle\tau_h\rangle/m_h$, где знак $\langle \rangle$ означает соответствующим образом выполненное усреднение по распределению носителей по энергии. В разд. 7.4 были введены значения τ_n и τ_p , названные временами жизни неосновных носителей заряда (соответственно электронов и дырок). Эти величины определяют скорость, с которой происходит изменение состояния электронов и дырок вследствие рекомбинации или захвата. Они связаны с темпом генерации электронно-дырочных пар, а также темпом возбуждения электронов и дырок с примесных центров. В этой главе мы рассмотрим различные механизмы, определяющие значения τ_e и τ_h и связанных с ними величин, а в следующей главе обсудим различные процессы, которые определяют τ_n и τ_p .

8.2. Механизмы рассеяния

Проблема определения τ_e , τ_h как функции волнового вектора k является основной проблемой кинетической теории, и для ее разрешения необходимо привлечь квантовую теорию процессов столкновений; последняя использует довольно сложный аппарат волновой механики, на деталях которого мы не имеем здесь возможности останавливаться. Поэтому ограничимся лишь ссылками на некоторые статьи, в которых разрабатывалась квантовомеханическая теория различных механизмов рассеяния. Несмотря на то что для детального определения τ в каждом отдельном случае требуется применение квантовомеханической теории в полном объеме, иногда оказывается возможным выяснить *зависимость* τ от энергии E

значительно более простыми средствами по крайней мере в тех сравнительно простых случаях, когда τ функционально зависит лишь от энергии E и не зависит от направления движения рассеиваемого электрона в кристалле. Как было показано в подразд. 5.3.2, зависимость τ от направления в определенных случаях можно учесть, приняв соответствующие значения эффективной массы. За исключением выражения для электропроводности, которое содержит $\langle \tau \rangle$ (или $\langle \tau^{-1} \rangle$ для случая сильного магнитного поля), во все формулы разд. 5.3 входят лишь отношения вида

$$\frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}, \quad \frac{\langle \tau^3 \rangle \langle \tau \rangle}{\langle \tau^2 \rangle^2} \text{ и т. д.},$$

которые не зависят от абсолютного значения τ , и для их вычисления достаточно знать лишь зависимость τ от энергии E . Поэтому главная задача заключается именно в том, чтобы определить эту зависимость.

8.3. Рассеяние на колебаниях атомов решетки

Как мы уже говорили, наиболее фундаментальным процессом рассеяния электронов в кристалле является их взаимодействие с тепловыми колебаниями атомов кристалла. Эти колебания имеются даже в совершенном кристалле. Другие виды рассеяния связаны с наличием различного рода несовершенств кристалла, которые обсуждались в разд. 3.1.

Часто утверждают, что атомы совершенного кристалла занимают в нем положения, образующие идеальную периодическую решетку. Необходимо, однако, уточнить, как следует понимать выражение «занимать положение» в кристаллической решетке. Это, конечно, не означает, что в каждый данный момент времени атомы в кристалле образуют идеальную периодическую решетку. При обычных температурах атомы совершают небольшие колебания относительно фиксированных положений равновесия, причем средняя энергия колебаний определяется законом равномерного распределения энергии по степеням свободы. Согласно классической механике, при абсолютном нуле температуры атомы должны покоиться в положениях равновесия, и в этих случаях их можно рассматривать как узлы идеальной решетки. Однако, согласно квантовой теории, даже при абсолютном нуле температуры атомы в кристаллах по аналогии с атомами в молекулах обязаны совершать небольшие так называемые нулевые колебания около своих положений равновесия [см. ниже формулу (8.4)]. Такие нулевые колебания являются *упорядоченными* и поэтому не вызывают рассеяния носителей заряда. Поэтому относящееся к атомам кристалла выражение «занимают

определенное место» следует понимать в том смысле, что среднее положение каждого атома в идеальном кристалле, если провести усреднение за достаточно большой промежуток времени (по сравнению с периодом колебаний), совпадает с соответствующим узлом идеальной периодической решетки. С достаточной степенью точности можно воспользоваться представлением о том, что при абсолютном нуле температуры атомы покоятся в узлах решетки и что их беспорядочные смещения при более высоких температурах можно рассматривать как случайные колебания около соответствующих положений равновесия. Такие колебания обычно носят название «колебаний решетки», хотя в некотором смысле это не совсем удачный термин, поскольку сама решетка на самом деле фиксирована и не колеблется, будучи геометрической совокупностью точек в пространстве. Однако такая терминология давно стала общепринятой в специальной литературе. Средняя амплитуда таких тепловых колебаний решетки, за исключением области весьма низких температур, пропорциональна $T^{1/2}$, где T — абсолютная температура. Это следует из того, что энергия колебаний каждого атома в первом приближении равна $3kT$. Следствием равномерного распределения энергии является также хорошо известный закон Дюлонга и Пти, согласно которому теплоемкость одноатомных кристаллических твердых тел при обычных температурах не зависит от температуры и равна $3R$, где R — газовая постоянная. При более низких температурах, как следует из квантовой теории, теплоемкость становится пропорциональной T^3 .

Существует два эквивалентных метода построения теории колебаний атомов кристалла. Первым из них является классическая теория малых колебаний. Кристалл, содержащий N атомов, может рассматриваться как динамическая система с $3N$ степенями свободы. Поэтому ее движение может быть описано с помощью $3N$ нормальных координат Q_r ($r=1, 2, \dots, 3N$). Нормальные координаты Q суть функции конфигурационных координат (x_r, y_r, z_r) ($r=1, 2, \dots, N$), характеризующих смещения N атомов относительно своих положений равновесия. Каждая нормальная координата описывает определенную конфигурацию всех атомов кристалла, совершающую колебания по простому гармоническому закону с периодом $T_s = 2\pi/\omega_s$. Такое колебание носит название нормального колебания (или моды), а период T_s называют периодом нормального колебания. Координата Q_s удовлетворяет простому уравнению

$$\ddot{Q}_s + \omega_s^2 Q_s = 0 \quad (s = 1, \dots, 3N). \quad (8.1)$$

Введение совокупности независимых друг от друга нормальных координат Q_s позволяет простым образом применить квантовую теорию к рассмотрению задачи о колебаниях решетки. В случае малых колебаний движение совокупности атомов кристалла можно считать линейной суперпозицией движений, соответствующих отдель-

ным нормальным колебаниям. Уравнение (8.1) есть не что иное, как классическое уравнение движения гармонического осциллятора, квантовый аналог которого имеет хорошо известные решения (см. III, гл. 8). Спектр разрешенных энергий E_{n_s} осциллятора определяется в квантовой механике формулой

$$E_{n_s} = \hbar \left(n + \frac{1}{2} \right) \omega_s = h \left(n + \frac{1}{2} \right) \nu_s, \quad (8.2)$$

где n — любое положительное целое число (включая 0), ν_s — частота s -го нормального колебания. Поэтому полную колебательную энергию кристалла можно записать в виде

$$E_v = h \sum \left(n_s + \frac{1}{2} \right) \nu_s, \quad (8.3)$$

где $n_1, n_2, \dots, n_s, \dots$ — некоторая последовательность положительных целых чисел. Заметим, что минимальная энергия колебаний, связанная с s -м нормальным колебанием, равна $(1/2)\hbar\nu_s$ и представляет собой энергию так называемых нулевых колебаний данной частоты. Полная энергия нулевых колебаний всего кристалла W_{0v} определяется как

$$W_{0v} = \frac{1}{2} \hbar \sum \nu_s. \quad (8.4)$$

Волновая функция, соответствующая минимальной энергии кристалла при абсолютном нуле температуры, распределена в небольшой области, прилегающей к равновесной конфигурации атомов, и описывает колебания с небольшой, но конечной амплитудой. Таким образом, согласно квантовой теории, состояние кристалла с атомами, покоящимися в соответствующих положениях равновесия, невозможно даже при абсолютном нуле температуры.

Другой метод рассмотрения вопроса о малых колебаниях сводится к использованию представления о бегущих волнах, когда движение каждого атома кристалла описывается с помощью суперпозиции бегущих волн смещения. Если под \mathbf{R} понимать вектор смещения атома, положение равновесия которого характеризуется вектором \mathbf{r} , то каждую бегущую волну смещений можно записать в виде

$$\mathbf{R} = \mathbf{A} \exp [i (\mathbf{K} \cdot \mathbf{r} - \omega t)], \quad (8.5)$$

где \mathbf{K} — волновой вектор. Зависимость ω от \mathbf{K} определяется конкретной структурой кристалла и носит название закона дисперсии для кристалла. Учет надлежащих граничных условий показывает, что в кристалле имеется вполне определенный набор разрешенных частот колебаний ω , причем разрешенные значения ω , как и следовало ожидать, совпадают с частотами ω_s нормальных колебаний. Более того, амплитуды A_s бегущих волн в выражении (8.5), соответствующие частотам ω_s , тесно связаны с нормальными коор-

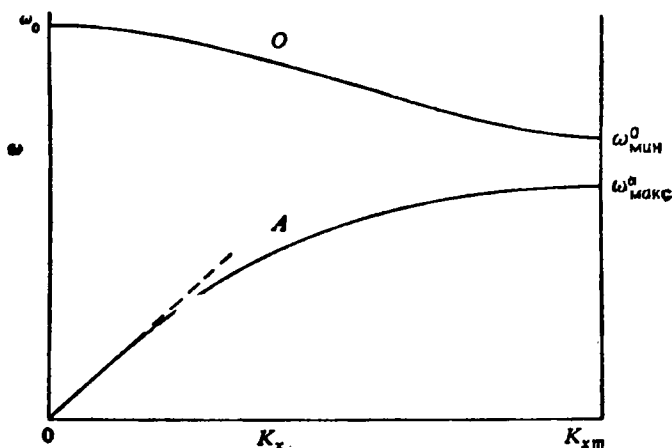


Рис. 8.1. Кривые дисперсии для спектра колебаний кристалла с двумя атомами на одну элементарную ячейку.

динатами Q_s (см. [1], гл. 8). Длина волны λ_s колебаний с частотой ω_s равна $2\pi/K_s$. При малых значениях K , соответствующих длинным волнам, формула (8.5) описывает обычные звуковые волны в кристалле. Вид зависимости ω от K напоминает зависимость E от k для электронов в кристалле. Вследствие периодичности кристаллической решетки значения волнового вектора K , так же как и волнового вектора электрона k , ограничены первой зоной Бриллюэна. Разрешенные значения K , определяющие спектр частот ω , для бесконечного кристалла можно найти тем же путем, что и разрешенные значения k (см. разд. 2.2). В кристалле, содержащем N_a атомов на одну элементарную ячейку, кривые $\omega = \omega(K)$ разбиваются на $3N$ ветви, а не на бесконечное число ветвей, как в случае электронов. На рис. 8.1 представлены дисперсионные кривые колебаний решетки для определенного направления в кристалле, содержащем (подобно Ge или PbS) по два атома на элементарную ячейку. На рисунке показаны две ветви разрешенных частот A и O , разделенные областью запрещенных частот в интервале $\omega_{\max}^0, \omega_{\min}^0$. При $K = K_{xm}$, где K_{xm} — значение K на границе первой зоны в рассматриваемом направлении, касательная к кривой $\omega(K_x)$ параллельна оси K_x ; то же имеет место и для O -ветви при $K_x = 0$ и при $K_x = K_{xm}$. Кривая, изображающая ветвь A вблизи $K = 0$, переходит в прямую с наклоном u_a . Фазовая скорость волн смещения равна ω/K и стремится к u_a при $K \rightarrow 0$ в случае A -ветви. Групповая скорость этих волн равна $d\omega/dK$ и стремится к u_a для A -ветви при малых значениях K . Поэтому A -ветвь при малых значениях K соответствует обычным звуковым волнам, причем скорость распространения этих волн в точности равна скорости звука u_a в кристалле. Из макроscopic-

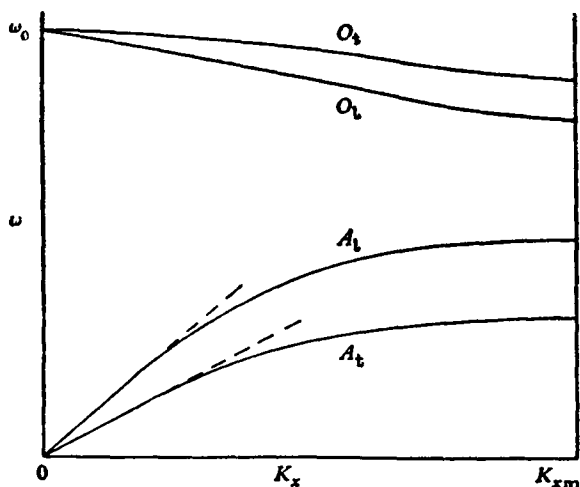


Рис. 8.2. Кривые дисперсии для продольных и двух вырожденных поперечных ветвей колебаний кристалла.

ческой теории звука в твердых телах известно, что в изотропной среде имеются два типа звуковых волн; первому типу соответствует волна сжатий и растяжений в твердом теле, второму — распространение деформаций сдвига. В первом случае мы имеем дело с продольными волнами, направление распространения которых совпадает с направлением колебаний частиц твердого тела, второму типу колебаний соответствуют поперечные волны, так как в этом случае движение частиц перпендикулярно направлению распространения волны. Любая поперечная волна может быть представлена в виде суммы двух независимых поперечных волн сдвига, колебания в которых происходят соответственно по двум взаимноперпендикулярным направлениям. Поэтому общий вид уравнения для смещения R можно записать в виде

$$R = \sum R_i = \sum_i \sum_s A_{is} \exp [i(\mathbf{K}_{is} \cdot \mathbf{r} - \omega_{is} t)], \quad (8.6)$$

где индексом $i=1, 2, 3$ нумеруются соответственно продольная и две поперечные волны. В общем случае кривые $\omega = \omega(\mathbf{K})$ для волн с различной поляризацией отличаются друг от друга, однако в случае изотропной среды две поперечные волны сдвига вырождены, поскольку им соответствуют одинаковые значения частоты ω . Такое вырождение различных ветвей в спектре колебаний может иметь место и для некоторых направлений в анизотропном кристалле, характеризующихся высокой степенью симметрии. Вырождение поперечных ветвей колебаний в кристалле показано на рис. 8.2, где кривые A_1, O_1 соответствуют продольным колебаниям, а $A_1,$

O_1 — поперечным. Кривые, отличающиеся только поляризацией, совпадают, вообще говоря, при $K=0$ для неполярных кристаллов. Как уже было сказано, ветви A при малых K соответствуют распространению акустических волн. Поэтому A -ветви и называются акустическими. В случае длинных акустических волн оба атома в элементарной ячейке движутся в одном направлении, как и атомы в ряде соседних ячеек. Ветви O соответствуют колебаниям, при которых два атома в элементарной ячейке движутся в противоположных направлениях. Случай $K=0$ соответствует колебанию, при котором все атомы типа 1¹⁾ во всем кристалле колеблются синфазно, так же как и все атомы типа 2, но движение этих двух систем атомов происходит в противоположных направлениях. Понятно, что такому виду колебаний соответствует стоячая волна; на рис. 8.1 и 8.2 легко видеть, что групповая скорость волн для ветвей типа O обращается в нуль при $K=0$. Такие колебания носят название оптических вследствие того, что они обычно сильно взаимодействуют с электромагнитными волнами (инфракрасным излучением). Такое взаимодействие имеет место в тех случаях, когда два атома в элементарной ячейке обладают зарядами противоположного знака, как, например, в полярных кристаллах. В процессе колебаний, смещаясь в противоположных направлениях, ионы порождают сильные электрические дипольные моменты, взаимодействующие с электромагнитным полем (см. [1], § 8.1). Частота ω_0 этих колебаний при $K=0$ обычно лежит в далекой инфракрасной области спектра, так что длина волны для таких колебаний значительно больше расстояния между атомами в кристалле. Поэтому взаимодействие оптических колебаний с излучением велико лишь при малых значениях K . При малых же длинах волн $2\pi/K$ по сравнению с длиной волны инфракрасного излучения это взаимодействие в среднем исчезает. Вследствие этого в полярных кристаллах наблюдается резко выраженное поглощение излучения с частотой $\nu_0 = \omega_0/2\pi$, которую называют частотой *остаточных лучей*. Различие между акустическими и оптическими колебаниями становится менее отчетливым при $K_x \rightarrow K_{xм}$, однако поведение этих двух ветвей различается при малых K . Заметим, что на границе зоны Бриллюэна, так же как и в случае электронов, групповая скорость равна нулю. Это означает, что вблизи границы зоны колебания имеют характер стоячих волн. В некоторых полупроводниках соответствующие краю зоны значения ω , которые мы обозначим через ω_{max}^0 и т. д., могут играть существенную роль в теории поглощения инфракрасного излучения электронами.

На рис. 8.2 показаны две ветви оптических колебаний, совпадающие по частоте при $K=0$. Такая ситуация имеет место в кристаллах типа Si и Ge, которые являются неполярными. Для полярных сое-

¹⁾ Для удобства обсуждения атомы, входящие в каждую элементарную двухатомную ячейку, названы соответственно атомами типа 1 и типа 2. — *Прим. ред.*

динений, таких, как полупроводники типа $A^{III}B^V$, две ветви оптических колебаний не совпадают и при $K=0$. У них имеются две частоты оптических колебаний, ω_{LO} и ω_{TO} , соответствующие нулевому значению K . Если рассмотреть взаимодействие электромагнитных волн частоты $\omega/2\pi$ с колебаниями решетки, можно найти, что существует определенное соотношение, связывающее частоты ω_{LO} и ω_{TO} со статической диэлектрической проницаемостью ϵ_s и диэлектрической проницаемостью на высоких частотах ϵ_h . Это соотношение $\omega_{LO}^2/\omega_{TO}^2 = \epsilon_s/\epsilon_h$ известно как соотношение Лиддана — Закса — Теллера. Так как ϵ_s в отличие от ϵ_h включает в себя поляризацию, обусловленную движением ионов, то в общем случае для полярных полупроводников $\epsilon_s > \epsilon_h$, так что для них $\omega_{LO} > \omega_{TO}$. Если пренебречь затуханием колебаний решетки, то получим, что $\epsilon \rightarrow \infty$ при $\omega = \omega_{TO}$ и $\epsilon = 0$ при $\omega = \omega_{LO}$. При $\omega_{TO} < \omega < \omega_{LO}$ величина ϵ отрицательна, так что в этой полосе частот будет иметь место сильное отражение. С другой стороны, анализ, проведенный с учетом затухания, показывает, что на частоте $\omega = \omega_{TO}$ существует также очень сильное поглощение. Это поглощение в области остаточных лучей. Чтобы найти величины ω_{LO} и ω_{TO} , необходимы измерения как поглощения, так и отражения или излучательной способности (см. разд. 10.8 и 14.3).

Поскольку различие между частотами ω_{LO} и ω_{TO} зависит от эффективного заряда иона e^* , такие измерения дают способ измерения e^* . На низких частотах поляризация, пересчитанная на ионную пару, дается выражением $e^{*2}/4\pi\epsilon_0 M_r \omega_{TO}^2$, где M_r — приведенная масса ионной пары (см. [1] § 8.18). (См. также разд. 13.5 и 14.3.)

Квантование колебаний, соответствующих бегущим волнам, выполняется так же, как и ранее. Каждому индексу i соответствует N разрешенных значений частоты, что в сумме составляет $3N$ колебаний. Энергия по-прежнему определяется равенствами, подобными (8.2). Обмен энергиями между колебательными состояниями кристалла и внешним полем излучения может произойти лишь порциями, кратными $\hbar\nu_s$, где ν_s — одна из разрешенных частот. Аналогичная ситуация имеет место, естественно, и при обмене энергией между электронами и колебаниями решетки. Этот энергетический обмен вполне подобен обмену энергией между колебаниями решетки и полем излучения, в котором происходит рождение или уничтожение фотонов. В следующем разделе мы убедимся в существовании тесной аналогии между фотонами и колебаниями решетки.

8.4. Фононы

Исследуем теперь воздействие, которое оказывает бегущая волна, описываемая формулой (8.5), на движение электрона в зоне проводимости. Рассмотрим сперва волну сжатий и растяжений.

Распространение волны такого вида приводит к тому, что межатомные расстояния в кристалле периодически изменяются и как следствие этого на периодический потенциал в кристалле налагается слабое периодическое возмущение. При малых значениях K скорость распространения этого возмущения по кристаллу совпадает со скоростью звука ($\sim 10^5$ см·с $^{-1}$). При комнатных температурах электрон движется в среднем со скоростью порядка 10^7 см·с $^{-1}$, так что периодический возмущающий потенциал, вызывающий рассеяние электрона, перемещается медленно по отношению к электронной волне. При рассеянии электронной волны возникает небольшое доплеровское смещение частоты, и, следовательно, энергия электрона испытывает небольшое изменение, которое можно считать.

Рассмотрим электрон, который имел до рассеяния волновой вектор \mathbf{k} и описывался волновой функцией следующего вида:

$$\Psi_i = u(r) \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (8.7)$$

при этом $\hbar\omega = E$, где E — энергия электрона. Как будет видно из дальнейшего, основную роль играет рассеяние на длинноволновых акустических колебаниях, поэтому функцию $u(r)$ можно заменить ее средним значением и рассматривать как постоянное число. После рассеяния волновой вектор электрона становится равным \mathbf{k}' и соответствующая волновая функция приобретает вид

$$\Psi_s = u'(r) \exp [i(\mathbf{k}' \cdot \mathbf{r} - \omega' t)]. \quad (8.8)$$

Сжатия и растяжения в кристалле, обусловленные волнами смещения (8.5), достигают своего максимума в плоскостях, отстоящих друг от друга на расстояние $\lambda_s = 2\pi/K_s$. Поэтому критерием сильного отражения служит хорошо известное условие Брэгга

$$\mathbf{k} - \mathbf{k}' = \mathbf{K}_s. \quad (8.9)$$

В этом можно убедиться следующим образом: пусть направление движения электрона до рассеяния образует угол φ с вектором \mathbf{K}_s , т. е. с нормалью к рассеивающим плоскостям. Тогда $\mathbf{k} \cdot \mathbf{K}_s = kK_s \cos\varphi$. Аналогично $\pi - \psi$ представляет собой угол между направлением

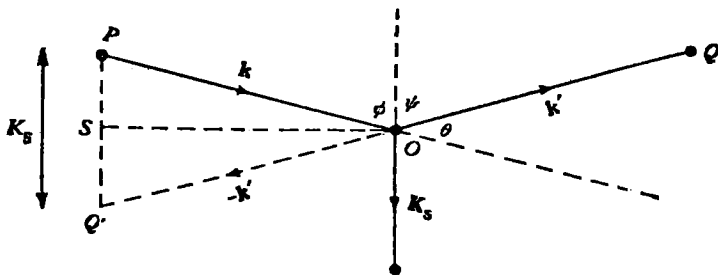


Рис. 8.3. Рассеяние волн системой параллельных плоскостей.

движения электрона после рассеяния и вектором K_s , так что $k'K_s = k'K_s \cos\psi$, где k , k' и K_s — модули векторов k , k' и K_s (рис. 8.3). Чтобы фазы всех локальных волн, отраженных от различных точек одной и той же плоскости, совпадали, необходимо, чтобы $k \sin\psi = k' \sin\psi$. На рис. 8.3 отрезок PO представляет собой вектор k , OQ — вектор k' , OQ' — вектор $-k'$, а отрезок OS — перпендикуляр к K_s в плоскости OPQ' . Согласно приведенному выше соотношению, PQ' составляет прямой угол с OS . Отсюда можно заключить, что

$$k - k' = AK_s, \quad (8.10)$$

где A — некоторая постоянная. Теперь, для того чтобы совпадали фазы волн, отраженных от различных плоскостей, отстоящих друг от друга на λ_s , необходимо, чтобы $A=1$. В этом легко убедиться, положив $\psi=0$, так как в этом случае

$$k + k' = \frac{2\pi}{\lambda_s} = K_s,$$

так что $A=1$. Поэтому условие рассеяния определяется формулой (8.9). Угол рассеяния электрона θ равен $(\pi - \varphi - \psi)$. Поскольку доплеровское смещение мало, то $k' \approx k$, $\varphi \approx \psi$, $\theta \approx \pi - 2\varphi$, так что

$$K_s = |k' - k| = 2k \sin \frac{1}{2} \theta. \quad (8.11)$$

Для того чтобы фаза падающей и отраженной волн в точке, движущейся вместе с волной, оставалась постоянной (условие отражения), должны совпадать фазы в (8.7) и (8.8). Таким образом, после замены τ на $(K_s/K_s)ut$, где $u = \omega_s/K_s$, получим

$$\frac{(k \cdot K_s)u}{K_s} - \omega = \frac{(k' \cdot K_s)u}{K_s} - \omega'. \quad (8.12)$$

Этой формулой и определяется доплеровское смещение в отраженной волне. Из (8.12) следует

$$\omega - \omega' = \frac{[(k - k') \cdot K_s]u}{K_s}, \quad (8.13)$$

или, используя формулу (8.11),

$$\omega - \omega' = K_s u = \omega_s. \quad (8.14)$$

Аналогичным образом для волн, бегущих в обратном направлении, можно записать

$$\omega - \omega' = -K_s u = -\omega_s. \quad (8.15)$$

Таким образом, разность $\omega - \omega'$ может принимать два возможных значения:

$$\omega - \omega' = \pm \omega_s. \quad (8.16)$$

Умножая обе части равенства (8.16) на \hbar , получаем для изменения энергии электрона при рассеянии формулу

$$E - E' = \pm \hbar \nu_s. \quad (8.17)$$

Физический смысл этой формулы заключается в том, что при рассеянии электрон может терять или приобретать лишь дискретную порцию энергии $\hbar \nu_s$, на которую может измениться полная энергия кристалла.

Остается еще доказать, что в каждом акте рассеяния изменение энергии электрона мало по сравнению со средней энергией электрона, так что действительно можно положить $k \approx k'$. Комбинируя равенства (8.11) и (8.17), можно показать, что изменение энергии $\delta E = E' - E$ при рассеянии электрона на угол θ составляет

$$\delta E \approx \pm 2m\upsilon v \sin\left(\frac{1}{2}\theta\right), \quad (8.18)$$

где v — скорость электрона, причем положительный знак соответствует поглощению, а отрицательный — излучению энергии. Наибольшему изменению энергии при рассеянии соответствует значение $\theta = \pi$. В этом случае $|\delta E|/E \approx 4u/v$. Так как отношение u/v мало, за исключением случая низких температур, то действительно $|\delta E|/E \ll 1$.

Рассмотрим теперь изменение квазиимпульса \mathbf{P} электронов при рассеянии. Из (8.10) следует, что

$$\mathbf{P} - \mathbf{P}' = \hbar(\mathbf{k} - \mathbf{k}') = \hbar \mathbf{K}_s. \quad (8.19)$$

Таким образом, $\hbar \mathbf{K}_s$ можно трактовать как импульс, приобретаемый решеткой вследствие соударения с электроном, так что соответствующий прирост энергии составляет $\hbar \nu_s$. Приобретенный решеткой импульс можно записать также в виде \hbar/λ_s . Этот процесс аналогичен возникновению фотона с энергией $\hbar \nu$ и импульсом \hbar/λ . Кванты колебаний решетки, которые можно рассматривать как возникшие в результате взаимодействия электронов с атомами кристалла, называются *фононами*, а процессы, аналогичные рассмотренным выше, — однофононными процессами. Кроме однофононных процессов рассеяния, возможно также рассеяние с образованием или поглощением более чем одного фонона, поскольку колебания решетки не описываются в точности простыми гармоническими волнами, т. е. эффекты ангармоничности дают члены с частотами $2\nu_s$ и т. д. Однако вероятность таких процессов весьма мала по сравнению с вероятностью однофононных процессов, и они здесь рассматриваться не будут.

Теперь можно понять, почему рассеяние электронов на нулевых колебаниях решетки невозможно. Вблизи абсолютного нуля температуры колебания решетки находятся в своем наименьшем энергетическом состоянии и поэтому не могут передать электрону даже

одного кванта энергии $h\nu_s$. С другой стороны, энергия электрона также слишком мала для того, чтобы фононы возникли за счет собственной энергии электрона, поэтому электрон не может рассеяться, т. е. изменить значение своего волнового вектора k . В дальнейшем (см. разд. 10.5.3 и 10.8) мы рассмотрим важную роль фононов в процессах поглощения инфракрасного излучения с помощью механизма межзонных переходов в полупроводниках. Теперь же мы займемся более подробным рассмотрением процессов рассеяния электронов на колебаниях решетки, чтобы выяснить зависимость времени релаксации электрона от энергии, а также от температуры кристалла.

8.5. Рассеяние на колебаниях решетки

Когда движущийся электрон взаимодействует с волной смещений, он может либо рассеяться волной, либо остаться невозмущенным. Если рассеяние произошло, то существует большая вероятность того, что угол рассеяния окажется равным $\pi - 2\varphi$, где φ — угол между исходным направлением квазимпульса электрона и нормалью к фронту волны. Не всякие волны смещения, движущиеся в одном и том же направлении, могут рассеивать электрон, рассеяние произойдет лишь на тех волнах, частота ν_s которых удовлетворяет условию (8.17). Максимально допустимое значение ν_s , согласно этому условию, определяется из соотношения $h\nu_s/E_e \approx 4u/v$, где E_e — энергия электрона. При $E_e = kT$ и $T = 300\text{K}$ имеем $h\nu_s \approx kT/10 = k\theta_s$, где $\theta_s = 30\text{K}$. Верхний предел акустической ветви в спектре колебаний обычно соответствует энергии $k\theta$, где θ — температура порядка $100 \div 300\text{K}$, так что эффективно рассеивают лишь колебания, относящиеся к начальному участку акустической ветви, когда длина волны значительно больше межатомных расстояний в кристалле.

Поскольку волны смещения могут свободно распространяться в кристалле по всем направлениям, то рассеяние должно быть изотропным, если существует равная вероятность столкновения электрона с соответствующими рассеивающими фононами. Согласно принципу детального равновесия, вероятности поглощения и испускания фонона в состоянии термодинамического равновесия одинаковы, кроме того, вероятность поглощения фонона пропорциональна числу имеющихся фононов. Нас интересуют фононы с энергиями в интервале от нуля до максимального значения E_m , причем отношение среднего числа фононов с максимальной энергией к числу фононов с энергиями вблизи нуля равно $\exp(-E_m/kT)$. За исключением области низких температур, обычно $E_m/kT \ll 1$. Поэтому в интервале энергии $(0, k\theta_s)$, в котором в основном и происходят процессы рассеяния под разными углами, плотность фононов прак-

тически постоянна. Таким образом, рассеяние в этом случае должно быть изотропным. Однако при очень низких температурах, когда число достаточно энергичных фононов сильно уменьшается, вероятность рассеяния под большим углом может стать очень малой. Однако, как мы увидим, при этих температурах основную роль, вероятно, играет рассеяние на ионизованных примесях.

Амплитуда электронной волны, рассеянной слабым периодическим потенциалом волны смещений, пропорциональна этому потенциалу; последний в свою очередь пропорционален амплитуде волны смещения. Вероятность рассеяния пропорциональна квадрату амплитуды электронной волны, а следовательно, и квадрату амплитуды волны смещений. Вместе с тем квадрат амплитуды волны есть мера ее энергии; в случае длинных акустических волн эта энергия по порядку величины равна kT . Таким образом, сечение рассеяния σ_1 должно быть пропорционально T ; поэтому средняя длина свободного пробега l , обусловленная рассеянием электронов на продольных акустических колебаниях, определяется зависимостью вида

$$\frac{1}{l} = B_1 T, \quad (8.20)$$

где B_1 — некоторая постоянная.

Для вычисления постоянной B_1 требуется подробный квантовомеханический расчет. Такой расчет был произведен Шокли и Бардином [2], он полностью подтверждает те выводы, к которым мы пришли на основе элементарных соображений. Средняя длина свободного пробега l для полупроводника со сферическими поверхностями постоянной энергии равна (см. [1], § 13.4.4)

$$l = \frac{\rho u_1^2 h^4}{16\pi^3 m_e^2 E_1^2 k T}, \quad (8.21)$$

где ρ — плотность, u_1 — скорость длинных волн сжатия в кристалле, E_1 — энергия, определяемая равенством

$$\Delta E_c = \frac{E_1 \Delta V}{V_0}. \quad (8.22)$$

В этой формуле ΔE_c означает величину сдвига энергии дна зоны проводимости, обусловленную малым изменением ΔV первоначального объема кристалла V_0 . Величину E_1 называют деформационным потенциалом. Время релаксации $\tau = l/v$ определяется тогда выражением

$$\tau = \frac{\rho u_1^2 h^4}{8\pi^3 (2m_e)^{3/2} k T E_1^2 E^{1/2}}. \quad (8.23)$$

Таким образом, время τ можно представить в виде

$$\tau = a E^{-1/2} T^{-1}, \quad (8.24)$$

где a — постоянное число. Выше (разд. 5.3.2) мы уже воспользовались таким видом зависимости τ от энергии при рассеянии на колебаниях решетки.

Когда поверхности равной энергии не имеют сферической формы, эффективная масса m_e должна быть заменена усредненной соответствующим образом величиной. Множитель $E^{1/2}m_e^{3/2}$ в знаменателе формулы (8.23) появляется из статического веса тех состояний, в которые попадает электрон после рассеяния. Мы уже знаем (см. разд. 4.2), что в полупроводниках с одним минимумом энергии в зоне проводимости этот множитель равен $E^{1/2}(m_1m_2m_3)^{1/2}$, если изоэнергетические поверхности имеют вид трехосных эллипсоидов. В полупроводниках с кубической симметрией и сферическими поверхностями равной энергии он равен $E^{1/2}m_T m_L^{1/2}$. Наконец, из формулы (8.24) следует, что подвижность μ , обусловленная рассеянием на акустических колебаниях решетки, может быть представлена в виде $\mu_1 = \mu_0 (T_0/T)^{3/2}$, где μ_0 и T_0 — некоторые постоянные. Используя формулу (5.97) для среднего значения $\langle \tau \rangle$ времени релаксации τ , получаем

$$\mu_l = \frac{2^{3/2} \pi^{1/2} e \hbar^4 \rho \mu_l^2}{3 m_e^{3/2} E_1^2 (kT)^{3/2}}. \quad (8.25)$$

Аналогичным образом можно рассмотреть и рассеяние положительных дырок. Формула для подвижности дырок совпадает с (8.25) с соответствующей заменой m_e эффективной массой m_h дырок и E_1 — величиной деформационного потенциала для валентной зоны. В полупроводниках со сферическими изоэнергетическими поверхностями как в валентной зоне, так и в зоне проводимости имеем

$$\frac{\mu_{el}}{\mu_{hl}} = \left(\frac{m_h}{m_e} \right)^{3/2} \left(\frac{E_{1h}}{E_{1e}} \right)^2. \quad (8.26)$$

Если $E_{1h} \approx E_{1e}$, то отсюда следует, что

$$\frac{\mu_{el}}{\mu_{hl}} \approx \left(\frac{m_h}{m_e} \right)^{3/2}. \quad (8.27)$$

Это соотношение очень часто используется в литературе без каких-либо указаний на пределы его применимости.

Из сказанного видно, что для эллипсоидальных поверхностей равной энергии при вычислении μ (см. разд. 5.2.1) необходимо заменить множитель $1/m_e$ на

$$\frac{1}{3} \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right).$$

Отсюда следует, что в формулу (8.25) необходимо подставить

$$\frac{1}{m_e^{3/2}} = \frac{1}{3 (m_1 m_2 m_3)^{1/2}} \left[\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right]. \quad (8.28)$$

В кристаллах с кубической симметрией и сферическими поверхностями равной энергии

$$\frac{1}{m_e^{*1/2}} = \frac{1}{3m_T m_L^{1/2}} \left(\frac{2}{m_T} + \frac{1}{m_L} \right). \quad (8.28a)$$

Необходимо, вообще говоря, учесть также и рассеяние, обусловленное поперечными типами колебаний, т. е. колебаниями сдвига. Зависимость времени релаксации τ от энергии и температуры для такого механизма рассеяния та же, что и при рассеянии на продольных колебаниях, однако полное выражение для τ_s несколько сложнее. Рассеяние на поперечных акустических колебаниях решетки теоретически рассмотрено Херрингом [3], который показал, что рассеяние на поперечных колебаниях в Si и Ge может дать довольно большой вклад по сравнению с рассеянием на продольных колебаниях. Однако в полупроводниках со сферическими поверхностями равной энергии относительный вклад поперечных колебаний в рассеяние должен быть значительно меньше.

Рассеяние дырок акустическими колебаниями представляет собой намного более сложный процесс, если валентная зона вырождена при $k=0$, как в Si и Ge. Даже если аппроксимировать зонную структуру наиболее простым способом, введя две скалярные эффективные массы m_{h1} , m_{h2} (т. е. легкие и тяжелые дырки, см. разд. 2.3), то существует возможность еще и такого процесса, когда легкая дырка, претерпев рассеяние, попадает в зону тяжелых дырок и наоборот. Если τ_{h1} и τ_{h2} — средние времена релаксации для двух типов дырок, то можно показать, что (см. [1], § 13.4.4)

$$\frac{1}{\tau_1} = \frac{1}{\tau_2} = \frac{1}{\tau_0} \left(\frac{m_{h1}}{m_s} \right)^{3/2} + \frac{1}{\tau_0} \left(\frac{m_{h2}}{m_s} \right)^{3/2}, \quad (8.29)$$

где τ_0 — время релаксации для дырки со скалярной эффективной массой m_s в «сферической» зоне с тем же деформационным потенциалом. Подвижности μ_{h1} и μ_{h2} двух типов дырок даются тогда соотношениями

$$\frac{e}{\mu_{h1}} = \frac{m_s}{\tau_0} \left[\left(\frac{m_{h1}}{m_s} \right)^{3/2} + \left(\frac{m_{h2}}{m_s} \right)^{3/2} \left(\frac{m_{h1}}{m_s} \right) \right], \quad (8.30)$$

$$\frac{e}{\mu_{h2}} = \frac{m_s}{\tau_0} \left[\left(\frac{m_{h2}}{m_s} \right)^{3/2} + \left(\frac{m_{h1}}{m_s} \right)^{3/2} \left(\frac{m_{h2}}{m_s} \right) \right]. \quad (8.30a)$$

Если $m_{h2} \ll m_{h1}$ (легкие дырки), то можно видеть, что время релаксации легких дырок при рассеянии в зону тяжелых дырок практически равно времени релаксации для рассеяния тяжелых дырок. Поэтому избыточные легкие дырки оказываются быстрорелаксирующими.

Для описания поперечных колебаний сдвига необходимо использовать отдельные деформационные потенциалы, которые зависят от различных компонент тензора упругой деформации и очень сложны.

Имея в распоряжении такое большое количество параметров, довольно трудно разделить различные компоненты решеточного рассеяния, однако исследованию этого вопроса было посвящено большое количество экспериментальных и теоретических работ, так что в хорошо изученных полупроводниках эти процессы теперь достаточно хорошо поняты.

Существует также рассеяние, обусловленное длинноволновыми оптическими колебаниями с частотами, близкими к частоте *остаточных лучей*. Долго считалось, что эти колебания должны играть важную роль в полярных полупроводниках, в то время как в элементарных полупроводниках их вклад в рассеяние должен быть мал. Этого следует ожидать для низких температур, поскольку в этом случае количество равновесных оптических фононов слишком мало, чтобы вызвать существенное рассеяние, обусловленное их поглощением. В то же время будет несущественным и рассеяние с испусканием оптических фононов, так как в этом случае будет слишком мало электронов, обладающих энергией, достаточной для возбуждения оптического фонона. Херринг [3] показал, однако, что при более высоких температурах в Si и Ge оптические фононы могут давать существенный вклад в рассеяние.

Несферичность изоэнергетических поверхностей приводит к большим трудностям в описании рассеяния из-за усложнения деформационных потенциалов и необходимости учета «междолинного» рассеяния (см. разд. 8.7). Проблема определения относительной величины вкладов в рассеяние различных типов колебаний решетки весьма сложна, и этому вопросу посвящено очень много работ. Обширные обзоры этих работ сделали Блатт [4], Содха [5], Молл [6] и Эстер М. Конуэлл [7], которая рассмотрела также условия, когда даже приближенно нельзя считать, что электроны находятся в равновесии с кристаллической решеткой.

Позже с помощью теории псевдопотенциала были вычислены значения различных параметров для Si, Ge и GaAs. Фосетт с соавторами дали критический обзор [8] предшествующих работ.

Виды деформационного потенциала для энергетических зон с еще более сложной структурой обсуждали Пикус и Бир [9]. Влияние отклонений зависимости $E(\mathbf{k})$ от квадратичного закона, имеющих более общий характер, рассмотрено рядом авторов и подробно обсуждалось Сосновским [10]. Отмечалось, в частности, что для полупроводниковых соединений типа $A^{III}B^V$, таких, как InSb, GaAs и т. д., для которых форма изоэнергетических поверхностей близка к «сферической», E не является *квадратичной* функцией k . В этом случае можно использовать эффективную массу, которая должна зависеть от средней энергии носителей заряда и, следовательно, от концентрации введенных примесей. Экспериментальное подтверждение этого эффекта получено Галавановым и др. [11], а также Бичевским и др. [12].

8.6. Рассеяние на оптических колебаниях в полярных кристаллах

Вследствие большого дипольного момента, порождаемого оптическими колебаниями в полярных кристаллах, взаимодействие электронов с оптической ветвью колебаний в таких кристаллах должно быть значительно более сильным, чем в неполярных кристаллах. Поэтому в полярных кристаллах оптические колебания должны играть в рассеянии намного более важную роль. Приближенные формулы, определяющие вклад рассеяния на оптических колебаниях в подвижность, выведены в работах Фрелиха и Мотта [13], Говарта и Зондгеймера [14], а также Лоу и Пайнса [15]. Применение результатов этих работ к полупроводникам обсуждено в работе Петрица и Скэнлона [16], которые указали на ряд необходимых уточнений. Сила взаимодействия между электроном и оптическим колебанием характеризуется константой взаимодействия α , которая определяется следующим образом:

$$\alpha = \frac{e^2}{2\hbar} \left(\frac{m_0}{2\hbar\nu_e} \right)^{1/2} \left(\frac{m_e}{m_0} \right)^{1/2} \frac{\epsilon - \epsilon'}{\epsilon\epsilon'}. \quad (8.31)$$

Частота ν_e равна $(\epsilon/\epsilon')^{1/2}\nu_0$, где ν_0 — частота остаточных лучей, соответствующая поперечным оптическим колебаниям при $k=0$; ϵ' — высокочастотная диэлектрическая проницаемость. Отличие между ϵ и ϵ' связано с тем, что ϵ' не содержит поляризации, обусловленной движением ионов. Выражение для α можно записать также и следующим образом:

$$\alpha = \frac{l_e}{a_0} \left(\frac{m_e}{m_0} \right) \frac{\epsilon_0(\epsilon - \epsilon')}{\epsilon\epsilon'}, \quad (8.31a)$$

где $a_0 = 4\pi\epsilon_0\hbar^2/m_0e^2$ — боровский радиус, l_e — длина, равная $\hbar/(2m_e \times \hbar\nu_e)^{1/2}$. В неполярных кристаллах величина α незначительна в силу малости $\epsilon - \epsilon'$, однако в полярных кристаллах она может принять значение лишь немногим меньше единицы. Например, для InSb $\alpha = 0,02$, для GaAs $\alpha = 0,06$, в то время как для PbS $\alpha = 0,28$, а для CdTe $\alpha = 0,35$. Теория Фрелиха и Мотта, а также ее обобщение Говартом и Зондгеймером применимы лишь при $\alpha < 1$. Значительно более сложный расчет Лоу и Пайнса проведен в приближении, когда $\alpha > 1$, однако при этом не получено явного выражения для времени релаксации, за исключением области низких температур. Выражение для подвижности $\mu_{\text{опт}}$, обусловленной рассеянием на оптических колебаниях, полученное Петрицем и Скэнлоном, можно записать в виде

$$\frac{1}{\mu_{\text{опт}}} = \frac{3\pi^{3/2}\alpha m_e \nu_e}{2e} [f(z)]^{-1}, \quad (8.32)$$

где

$$f(z) = \chi(z) \frac{e^z - 1}{z^{1/2}},$$

$$z = \frac{h\nu_e}{kT}.$$

Здесь $\chi(z)$ — медленно меняющаяся функция, принимающая значение $\chi=1,0$ для $z=0$ и $z \approx 3$, и $\chi=0,6$ для $z=1$. Экспоненциальный множитель, входящий в $f(z)$, приводит к тому, что при низких температурах величина $1/\mu_{\text{опт}}$ становится очень малой. Это обусловлено тем, что возбуждение оптических колебаний требует энергии по крайней мере порядка $h\nu_e$ и, следовательно, маловероятно при низких температурах. Петриц и Скэнлон [16] использовали это выражение (или эквивалентное ему), чтобы обсудить вопрос о подвижности электронов в халькогенидах свинца, в первую очередь в PbS, и показали, что при комнатной температуре величину подвижности определяет главным образом рассеяние на оптических колебаниях.

8.7. Междолинное рассеяние

Другой механизм рассеяния, имеющий значение для полупроводников, зона проводимости которых содержит более одного минимума энергии, — это так называемое «междолинное» рассеяние, связанное с перебросами электрона из одного минимума в другой. Электрон с волновым вектором \mathbf{k} , приблизительно равным волновому вектору \mathbf{k}_1 , соответствующему одному из минимумов, может в результате рассеяния оказаться в состоянии с волновым вектором \mathbf{k}' , близким к \mathbf{k}_2 , т. е. волновому вектору в другом минимуме. Однако, в связи с тем что разность $(\mathbf{k}_1 - \mathbf{k}_2)$, вообще говоря, не мала, такие переходы сопровождаются значительным изменением квазиимпульса. Это изменение квазиимпульса при рассеянии приблизительно равно $\hbar(\mathbf{k}_1 - \mathbf{k}_2)$, и для его осуществления необходимо испускание или поглощение фонона с довольно большой энергией. Такое рассеяние должно быть весьма неупругим, а потому и маловероятным при низких температурах из-за малости как числа фононов с большой энергией, которые могли бы участвовать в рассеянии, так и числа электронов, способных испускать такие быстрые фононы. Херринг [3] подробным образом рассмотрел влияние такого рода рассеяния на подвижность электронов в германии и кремнии n -типа. Херринг записывает выражение для τ в виде

$$\tau^{-1} = W_a + W_e, \quad (8.33)$$

где

$$W_a = W \frac{\left(1 + \frac{E}{\hbar\omega}\right)^{1/2}}{\exp\left(\frac{\hbar\omega}{kT}\right) - 1}$$

и

$$W_e = W \frac{\left(\frac{E}{\hbar\omega} - 1\right)^{1/2} \exp\left(\frac{\hbar\omega}{kT}\right)}{\exp\left(\frac{\hbar\omega}{kT}\right) - 1} \quad (E \geq \hbar\omega),$$

$$W_e = 0, \quad (E < \hbar\omega);$$

W_a — вероятность процесса поглощения фонона; множитель

$$\left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1\right]^{-1}$$

пропорционален N_p — числу фононов с энергией $\hbar\omega$ при температуре T . Соответствующий множитель $[\exp(-\hbar\omega/kT) - 1]^{-1}$, фигурирующий в выражении для вероятности испускания фонона W_e , пропорционален $(1 + N_p)$ (см. разд. 14.6). Квант $\hbar\omega$ равен изменению энергии колебаний решетки, которое сопутствует изменению волнового вектора при переходе электрона из окрестности одного минимума энергии в окрестность другого минимума. Можно убедиться в том, что при $\hbar\omega \gg kT$ вероятность W_a очень мала. Одновременно мал также и вклад W_e , так как $W_e = 0$ для всех энергий $E < \hbar\omega$. Необходимо также отметить, что в этом случае нет простой степенной зависимости τ от E .

8.8. Межэлектронное рассеяние

Обычно считается, что взаимодействие между электронами мало сказывается на подвижности носителей заряда в полупроводниках, по крайней мере вне области сильного вырождения. Но даже и в этом случае вклад межэлектронного рассеяния, по-видимому, невелик, если судить по тому, что электрон-электронное взаимодействие оказывает незначительное влияние на величину сопротивления металлов. Вместе с тем электрон-электронные и электрон-дырочные взаимодействия играют важную роль в процессах установления стационарного состояния в сильных электрических полях. Поскольку соударяющиеся частицы в этом случае имеют массы одного порядка, процесс обмена энергиями между носителями заряда происходит значительно быстрее, чем процесс обмена энергией между электронами и примесными центрами или колебаниями решетки. Эти вопросы рассмотрели Фрелих и Паранджайл [17] в

связи с теорией пробоя диэлектриков. Влияние электрон-электронного взаимодействия на подвижность исследовали Фрелих, Паранджайп, Купер и Накаджима [18], а также Содха и Истмен [19]. В этих работах предложены и обсуждены опыты, в которых межэлектронное рассеяние могло бы быть в принципе выделено среди других видов рассеяния. Однако имеющийся пока весьма скудный экспериментальный материал позволяет сказать о величине этого эффекта только то, что обычно она мала.

8.9. Рассеяние на ионизованных примесях

Рассеяние электрона в кулоновском поле ионизованного атома примеси можно рассматривать как с классической точки зрения, так и на основе квантовой механики; оба метода дают один и тот же результат ¹⁾. Дифференциальное сечение рассеяния под углом θ в телесном угле $d\omega$ имеет в этом случае вид

$$\sigma(\theta) d\omega = \frac{1}{4} R^2 \operatorname{cosec}^4\left(\frac{\theta}{2}\right) d\omega, \quad (8.34)$$

где

$$R = Ze^2/4\pi\epsilon_0 m_e v^2,$$

ϵ — диэлектрическая проницаемость. Отсюда следует, что это рассеяние весьма анизотропно; преобладающее значение имеет рассеяние под малыми углами. Для вычисления времени релаксации τ необходимо произвести усреднение (8.34) по всем возможным углам рассеяния, используя формулу (5.8). Однако если подставить приведенное выше выражение для $\sigma(\theta)$ в уравнение (5.8) и попытаться проинтегрировать по θ , чтобы получить σ_c , то мы получим расходящийся интеграл, т. е. бесконечное сечение рассеяния, что, естественно, никак не соответствует реальному положению вещей в кристалле. На самом деле рассеяние на ионизованной примеси обладает конечным сечением, так как кулоновский характер поля ионизованного центра не сохраняется на больших расстояниях от центра. Вместе с тем расходимость интеграла, определяющего σ_c , обусловлена именно процессами рассеяния под большими прицельными расстояниями ²⁾. Кулоновское поле каждого примесного центра в кристалле на определенном расстоянии от него оказывается искаженным полем ближайших примесных ионов, а также вследствие экранирующего действия свободных электронов и дырок.

¹⁾ См., например, книгу Мотта и Мессис [20].

²⁾ Под прицельным расстоянием понимают минимальное расстояние между рассеивающим центром и траекторией частицы, если бы частица двигалась все время свободно по прямой и не испытывала отклонения в поле рассеивающего центра.

В соответствии с этим имеются два подхода к решению этого вопроса. Конуэлл и Вайскопф [21] предположили, что кулоновское поле каждого рассеивающего центра перестает эффективно рассеивать за пределами радиуса r_m , определяемого из условия $(2r_m)^{-3} = N_1$, где N_1 — концентрация ионизованных примесных центров, а величина r_m равна половине среднего расстояния между примесями. Угол отклонения θ_m , соответствующий прицельному расстоянию r_m , можно определить из хорошо известного соотношения

$$\operatorname{tg} \left(\frac{\theta_m}{2} \right) = \frac{R}{r_m}. \quad (8.35)$$

Поэтому для вычисления сечения рассеяния для проводимости σ_c необходимо подставить значение $\sigma(\theta)$ из формулы (8.34) в уравнение (5.8) и проинтегрировать по углам в пределах от $\theta = \theta_m$ до $\theta = \pi$. В результате получим

$$\begin{aligned} \sigma_c &= 2\pi \int_{\theta_m}^{\pi} (1 - \cos \theta) \sigma(\theta) \sin \theta d\theta = \\ &= -4\pi R^2 \ln \sin \left(\frac{\theta_m}{2} \right) = 2\pi R^2 \ln \left(1 + \frac{r_m^2}{R^2} \right). \end{aligned} \quad (8.36)$$

Отношение r_m/R можно записать в виде $r_m/R = 2E/E_m$, где $E_m = Ze^2/4\pi\epsilon\epsilon_0 r_m$ представляет собой потенциальную энергию электрона в поле примесного центра на расстоянии r_m от последнего, взятую с обратным знаком. Оценим величину E_m . При $N = 10^{15} \text{ см}^{-3}$, $2r_m = 10^{-5} \text{ см}$, $Z = 1$ и $\epsilon = 10$ получим $E_m = 2,88 \cdot 10^{-3} \text{ эВ}$. При $T = 300 \text{ К}$ имеем $E = 3/2 kT = 0,039 \text{ эВ}$, так что $r_m/R \approx 27$. Таким образом, при комнатных температурах отношение r_m^2/R^2 велико. Чтобы определить время релаксации τ , воспользуемся соотношением $\tau_1^{-1} = N_1 \sigma_c v$ (см. разд. 5.1). Тогда получаем

$$\frac{1}{\tau_1} = \frac{Z^2 e^4 N_1}{16\pi (2m_e)^{1/2} \epsilon^2 \epsilon_0^2 E^{3/2}} \ln \left[1 + \left(\frac{2E}{E_m} \right)^2 \right]. \quad (8.37)$$

Для вычисления вклада в подвижность, вносимого рассеянием на ионизованных примесях, необходимо рассчитать $\langle \tau_1 \rangle$ по формуле

$$\langle \tau_1 \rangle = \frac{16\pi (2m_e)^{1/2} \epsilon^2 \epsilon_0^2}{Z^2 e^4 N_1} \cdot \frac{\int_0^{\infty} E^3 \exp \left(-\frac{E}{kT} \right) \left\{ \ln \left[1 + \left(\frac{2E}{E_m} \right)^2 \right] \right\}^{-1} dE}{\int_0^{\infty} E^{3/2} \exp \left(-\frac{E}{kT} \right) dE}. \quad (8.38)$$

Логарифмический член представляет собой сравнительно медленно меняющуюся функцию, поэтому его можно вынести за знак интеграла с заменой E на $E_{\max} = 3 kT$. E_{\max} представляет собой значение

E в максимуме остающейся под знаком интеграла части подынтегральной функции. Таким путем можно получить формулу

$$\mu_1 = \frac{e}{m_e} \langle \tau_1 \rangle = \frac{64\pi^{1/2} \varepsilon^2 \varepsilon_0^2 (2kT)^{3/2}}{N_1 Z^2 e^3 m_e^{1/2}} \left\{ \ln \left[1 + \left(\frac{12\pi e \varepsilon_0 kT}{Ze^2 N_1^{1/2}} \right)^2 \right] \right\}^{-1}, \quad (8.39)$$

которую обычно называют формулой Конуэлл — Вайскопфа¹⁾.

Другой подход к проблеме «отсечения» далекой части кулоновского поля использован в работах Дингля [22] и Брукса²⁾. Ими учтено экранирование кулоновского поля ионизованного примесного центра свободными электронами и дырками. Поле заряженного центра обуславливает изменение средней плотности свободных зарядов вокруг центра. Пусть n — предельная концентрация свободных зарядов при больших r , а n' — переменная концентрация носителей заряда на близких расстояниях r от центра. Очевидно, что если на малых расстояниях от центра электростатический потенциал φ обладает сферической симметрией, то он должен подчиняться уравнению Пуассона

$$\frac{1}{r} \frac{d^2 (r\varphi)}{dr^2} = \frac{e (n' - n)}{\varepsilon \varepsilon_0}. \quad (8.40)$$

Считая систему свободных носителей заряда невырожденной, можем использовать функцию Больцмана

$$n' = n \exp \left(\frac{e\varphi}{kT} \right). \quad (8.41)$$

Уравнение (8.40) с учетом (8.41) приобретает следующий вид:

$$\frac{1}{r} \frac{d^2 (r\varphi)}{dr^2} = \frac{ne}{\varepsilon \varepsilon_0} \left[\exp \left(\frac{e\varphi}{kT} \right) - 1 \right]. \quad (8.42)$$

Ранее мы уже выяснили, что на расстояниях порядка r_m , где учет экранирования становится существенным, $e\varphi \ll kT$. В этом приближении выражение (8.42) упрощается:

$$\frac{d^2 (r\varphi)}{dr^2} = \frac{e^2 n}{\varepsilon \varepsilon_0 kT} (r\varphi). \quad (8.43)$$

Решение этого уравнения имеет вид

$$\varphi = \varphi_0 \frac{e^{-r/d}}{r},$$

где $d^2 = \varepsilon_0 e kT / e^2 n$. При $r \rightarrow 0$ $\varphi \rightarrow Ze / 4\pi \varepsilon \varepsilon_0 r$, так что

$$\varphi = \frac{Ze}{4\pi \varepsilon \varepsilon_0 r} e^{-r/d}. \quad (8.44)$$

¹⁾ См. [23]. Аналогичная теория применена к примесям в металлах Моттом [24].

Длина d носит название дебаевской длины. Потенциал, аналогичный (8.44), определяет также поле вблизи иона в электролите.

Дифференциальное сечение рассеяния $\sigma(\theta)$ в поле, описываемом потенциалом φ , определяется выражением [25]

$$\sigma(\theta) = \frac{1}{4} R^2 \left[\sin^2 \frac{1}{2} \theta + \frac{\hbar^2}{(2m_e v d)^2} \right]^{-2}, \quad (8.45)$$

где R определено формулой (8.34). При такой зависимости $\sigma(\theta)$ можно произвести усреднение по углам θ от 0 до π . Интегральное сечение σ_c получается в виде

$$\begin{aligned} \sigma_c &= 4\pi R^2 \int_0^\pi \frac{\sin^4 \frac{\theta}{2} d\left(\sin \frac{\theta}{2}\right)}{\left[\sin^2 \frac{\theta}{2} + \left(\hbar^2/2m_e v d\right)^2\right]^2} = \\ &= 2\pi R^2 \left[\ln(1 + \beta^2) - \frac{\beta^2}{1 + \beta^2} \right], \quad (8.46) \end{aligned}$$

где

$$\frac{1}{\beta^2} = \frac{\hbar^2}{4m_e^2 v^2 d^2} = \frac{\hbar^2 e^2 n}{8m_e \epsilon \epsilon_0 k T E}.$$

Величина β^2 , как правило, велика, и поэтому формулу (8.46) можно переписать в виде

$$\sigma_c = 2\pi R^2 \ln(\beta^2). \quad (8.46a)$$

Предположим, что $n = N_1$, тогда время релаксации τ_1 определяется зависимостью

$$\frac{1}{\tau_1} = \frac{Z^2 e^4 N_1 \ln(\beta^2)}{16\pi (2m_e)^{1/2} E^{3/2} e^2 \epsilon_0^2}. \quad (8.47)$$

После усреднения по E получим для подвижности выражение

$$\mu_1 = \frac{e}{m_e} \langle \tau_1 \rangle = \frac{64\pi^{1/2} e^2 \epsilon_0^2 (2kT)^{3/2}}{N_1 Z^2 e^3 m_e^{1/2}} \left[\ln \left(\frac{24m_e k^2 T^2 \epsilon \epsilon_0}{e^2 \hbar^2 N_1} \right) \right]^{-1}. \quad (8.47a)$$

Если полупроводник содержит как доноры, так и акцепторы в ионизованном состоянии и в сопоставимых концентрациях, то формула (8.47a) несколько изменится. Вместо величины $N_1 Z^2$ вне квадратных скобок необходимо подставить величину $Z_D^3 N_{ID} + Z_A^3 N_{IA}$, где Z_D, A — зарядовое число соответствующих ионов примеси. Что касается квадратных скобок, то здесь под N_1 следует понимать полную концентрацию свободных носителей заряда. Рассеяние на ионизованных примесях в условиях вырождения исследовал Мэнсфилд [26], который рассмотрел также вопрос о совместном решеточно-примесном рассеянии в условиях вырождения.

Выражения (8.39) и (8.47a) для подвижности μ_1 отличаются в основном лишь соответствующими логарифмическими множителями. Величину $r'_m = R\beta$ можно считать эквивалентным радиусом

«отсечения» кулоновского поля во втором методе. Полагая $E = 3kT$, имеем

$$r_m'^2 = \frac{Z^2 e^2 m_e}{24 \pi^2 \epsilon \epsilon_0 \hbar^2 N_1} = \frac{Z^2 (m_e/m_0)}{6\pi(\epsilon) a_0 N_1} \quad (a_0 = 0,53 \cdot 10^{-8} \text{ см}).$$

При $Z=1$ и $m_e=m_0$, $\epsilon=10$ получаем (концентрация N_1 взята в см^3)

$$r_m'^2 \approx \frac{10^6}{N_1} \text{ см}^2.$$

Для концентрации $N_1 \approx 10^{15} \text{ см}^{-3}$ нет заметной разницы между множителями $\ln(r_m'^2/R^2)$ и $\ln(r_m^2/R^2)$. Однако выражение (8.47а) для μ_1 предпочтительнее, так как оно основано на более точных теоретических предпосылках. Кроме того, внося соответствующие поправки в (8.47а), можно учесть то обстоятельство, что при низких температурах, где примесное рассеяние играет наиболее важную роль, многие примесные центры перестают быть ионизованными.

Когда поверхности постоянной энергии не обладают сферической симметрией, точный расчет времени релаксации τ_1 для рассеяния на ионизованных примесях затруднен. Поскольку основной вклад в рассеяние в этом случае вносит рассеяние под малыми углами, то можно ожидать, что для эллипсоидальных изознергетических поверхностей множитель $(m_e)^{-1/2}$ в формуле (8.47а) должен быть заменен на $m_1^{-1/2}$, $m_2^{-1/2}$ и т. д., если соответственно рассматривается движение по одному из главных направлений. Однако не совсем ясно, как при этом следует проводить усреднение при вычислении подвижности, например, в случае полупроводников с кубической симметрией. Можно только надеяться, что эффективная масса m^* , определяемая соотношением

$$\frac{3}{(m^*)^{1/2}} = \frac{1}{m_L^{1/2}} + \frac{2}{m_T^{1/2}}, \quad (8.48)$$

не приведет к значительным ошибкам, если ее использовать в формулах (8.39) и (8.47а). Подлинный учет анизотропии рассеяния приводит к необходимости введения времени релаксации в тензорном виде; Хэм [27] указал на то, что τ_1 должно быть тензором, диагонализированным так же, как тензор обратной эффективной массы. Гликсман [28] показал, что такое уточнение может быть необходимым для интерпретации результатов исследования рассеяния на ионизованной примеси в германии n -типа (см. разд. 5.3.2).

8.10. Рассеяние на нейтральных примесях

До сих пор речь шла только о рассеянии на ионизованной примеси. Следует ожидать, что нейтральная примесь рассеивает слабее, чем ионизованная примесь. Вместе с тем рассеянием на нейтральных

атомах примеси пренебречь нельзя, особенно при очень низких температурах, когда число нейтральных атомов примеси намного превышает число ионизованных примесных атомов. В разд. 3.4.1 было показано, что эффективный радиус орбиты электрона, связанного с примесным центром, намного превышает постоянную решетки, так что полное экранирование поля примесного иона связанным электроном имеет место лишь на значительном расстоянии от иона. Влияние рассеяния на нейтральных атомах примеси рассмотрели Бардин и Пирсон [29]. Они показали, что эффект рассеяния можно рассчитать таким же образом, как и рассеяние электронов на нейтральных атомах водорода. Существенную роль играют при этом два рода процессов: прямое упругое рассеяние и обменное рассеяние, в котором падающий электрон обменивается местом с электроном на примесном центре. В этих условиях обычно используемые приближенные методы вычисления атомных сечений рассеяния на самом деле неприменимы. Эргинсой [30] применит предложенную Месси и Моисеевичем [31] теорию рассеяния медленных электронов на атомах и показал, что при условии, когда $E < \frac{1}{4} E_1$, где E_1 — энергия ионизации примеси, выражение для сечения рассеяния σ_c имеет вид

$$\sigma_c = \frac{20a_1}{k}, \quad (8.49)$$

где k — модуль волнового вектора рассеиваемого электрона, $a_1 = (m_0/m_e)e a_0$ — радиус первой боровской орбиты для внешнего электрона атома примеси. Время релаксации τ_N при рассеянии на нейтральных атомах примеси выражается соотношением

$$\frac{1}{\tau_N} = \frac{20a_1 v N_n}{k} = \frac{20a_1 \hbar N_n}{m_e}, \quad (8.50)$$

где N_n — концентрация нейтральной примеси. Для примесей, к которым применима «водородоподобная» модель, величина a_1 дается выражением (3.4). Величину τ_N можно выразить через энергию связи электрона W_n на нейтральной примеси, которая дается формулой (3.3), в виде

$$\frac{1}{\tau_N} = \frac{5N_n \hbar e^2}{2\pi \epsilon_0 W_n m_e}. \quad (8.50a)$$

Поскольку τ_N не зависит от энергии, средняя величина $\langle \tau_N \rangle$ равна τ_N и подвижность μ_N дается выражением $e\tau_N/m_e$. Можно видеть, таким образом, что время релаксации при рассеянии на нейтральных примесях не зависит от температуры и не зависит также от энергии падающего электрона или дырки, по крайней мере в этом приближении. Поэтому такой вид рассеяния может быть существен при очень низких температурах. При вычислении величины a_1 для полупроводников с эллипсоидальными поверхностями равной

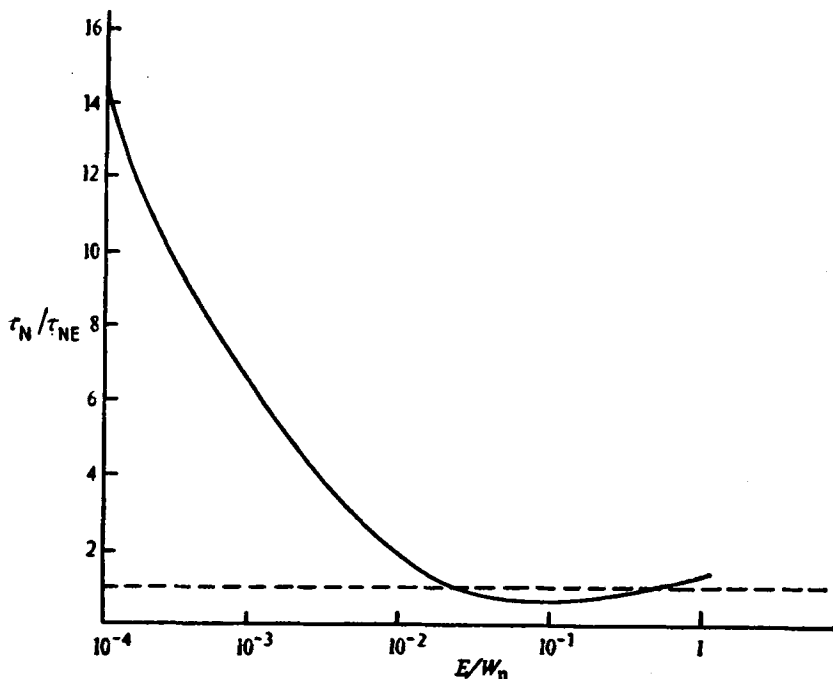


Рис. 8.4. Зависимость отношения значений времени релаксации при рассеянии на нейтральной примеси, полученных Мак-Гиллом и Бароном (τ_N) и вычисленных по формуле Эргинсоа (τ_{NE}), от энергии электрона E , отнесенной к величине энергии связи на примесном центре W_n [32].

энергии эффективная масса m_e заменяется величиной $(m_1 m_2^2)^{1/3}$, в то время как в знаменателе формулы (8.50) используется эффективная масса, определяющая величину подвижности.

При вычислении времени релаксации, обусловленного рассеянием нейтральными примесями, Эргинсой рассматривал фазовый сдвиг в волнах, рассеянных атомом в s -состоянии. Чтобы получить лучшее приближение, Мак-Гилл и Барон [32] использовали члены более высокого порядка. Ими найдено, что в то время как в рассматриваемой области энергий электрона, претерпевающего рассеяние, время релаксации почти постоянно, как это следует из формулы Эргинсоа, для значений энергии E , соответствующих условию $(E/W_n) < 0,01$, это уже не так, и τ_N обнаруживает существенное увеличение. Это иллюстрируется рис. 8.4, на котором отношение τ_N/τ_{NE} показано в зависимости от E/W_n ; τ_{NE} — постоянное значение, даваемое формулой Эргинсоа.

Поскольку теперь τ_N зависит от энергии E , то, чтобы получить μ_N , нужно вычислить среднее значение $\langle \tau_N \rangle$, как указано в разд. 3.1

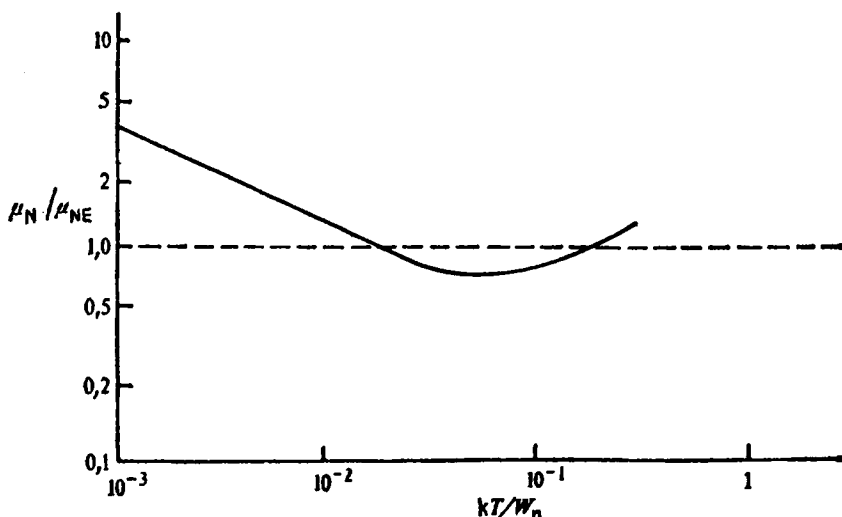


Рис. 8.5. Температурная зависимость отношения подвижности, обусловленной рассеянием на нейтральной примеси (μ_N), к значению, полученному с помощью формулы Эргинсоя для τ_N [32].

гл. 5. Полученная таким образом величина μ_N , отнесенная к значению μ_{NE} , даваемому формулой Эргинсоя, показана на рис. 8.5 в зависимости от величины kT/W_n . Можно видеть, что при $kT/W_n > 0,01$ отличие от значения Эргинсоя невелико, но, если kT/W_n порядка 0,001, значение подвижности приблизительно в четыре раза выше, чем значение, вычисленное по Эргинсою. При $W_n = 0,1$ эВ значения, вычисленные по Эргинсою, будут справедливы для области температур $T > 12$ К, а при $W_n = 0,01$ эВ для $T > 1,2$ К. Для более глубоких примесей необходимы более тщательные вычисления. Это было показано экспериментами Барона, Янга и Мак-Гилла [33] для примесей Al, As, Ga и In в Si. Даже если водородоподобная модель не дает правильного значения для энергии связи, вычисленная величина τ_N все же оказывается в хорошем согласии с экспериментально найденным значением.

Бароном, Янгом и Мак-Гиллом была вычислена также величина $r = \langle \tau_N^2 \rangle / \langle \tau_N \rangle^2$, которая появляется в выражении для коэффициента Холла (см. разд. 5.3.3). Для значений kT/W_n , больших 0,01, r лишь немного превышает единицу (как и следует ожидать для случая τ_N , не зависящего от E), но возрастает до значения 1,17, если $kT/W_n = 0,001$.

Рассеяние на нейтральных примесях обсуждалось также Благосклонской, Гершензоном, Ладыженским и Поповой [34], которые показали, что оно может быть доминирующим видом рассеяния при

низких температурах. Хорошо известно, что рассеяние электронов на атомах водорода зависит от направления спина [31]. При очень низких температурах спины электронов проводимости в сильном магнитном поле могут быть почти полностью поляризованы (см. разд. 12.6); в этих условиях следует поэтому ожидать заметной зависимости подвижности от направления спина. Этот вопрос подробно обсуждался Соломоном [35], а Максвелл и Хониг [36] продемонстрировали существование этой зависимости экспериментально.

8.11. Рассеяние на дислокациях

Искажения кристаллической решетки, вызванные дислокациями¹⁾, также приводят к рассеянию электронов и дырок. Этот вид рассеяния теоретически рассмотрен Декстером и Зейтцем [37], которым удалось показать, что вклад рассеяния на дислокациях пренебрежимо мал вплоть до концентраций дислокаций в кристаллах порядка 10^8 см^{-2} . Другой механизм рассеяния, обусловленный дислокациями, указал Рид [38], разработавший также и теорию этого вопроса. Речь идет о дислокациях, действующих в качестве акцепторных центров (см. разд. 3.1.3). Дислокация в кристалле *n*-типа ведет себя подобно линейному отрицательному заряду и создает вокруг себя положительный пространственный заряд (см. разд. 7.9). Падающие на дислокацию электроны испытывают с ее стороны отталкивание, приводящее к их рассеянию. Для расчета этого рассеяния каждую отдельную дислокацию можно рассматривать как заряженный цилиндр радиуса *R*, причем Рид показал, что при концентрации электронов в германии порядка 10^{15} см^{-3} *R* по порядку величины равно $3 \cdot 10^{-5}$ см. Рассеяние такими заряженными цилиндрами должно быть сильно анизотропным, достигая максимума, когда электрон движется под прямым углом к оси цилиндра, и исчезая при движении параллельно цилиндру. С помощью пластической деформации изгиба можно создать в кристалле упорядоченное расположение дислокаций в виде сетки параллельных дислокаций (см. разд. 3.1.3). Следует отметить, что в этом случае должно наблюдаться анизотропное рассеяние.

Проведя усреднение по всевозможным направлениям, Рид показал, что рассеяние на дислокациях можно описать с помощью времени релаксации

$$\tau_d = \frac{3}{8RNv}, \quad (8.51)$$

где *v* — скорость рассеиваемого электрона, а *N* — число дислокаций на единицу поверхности. Приняв $v = 10^7 \text{ см} \cdot \text{с}^{-1}$, $R = 3 \cdot 10^{-5} \text{ см}$,

¹⁾ См. разд. 3.1.3.

$N=10^6 \text{ см}^{-2}$, получаем $\tau_d=1,25 \cdot 10^{-9}$ с. Как видно из этой формулы, τ_d таким же образом зависит от скорости электрона, как и время релаксации τ_1 для решеточного рассеяния. Риду удалось показать, что, несмотря на анизотропный характер рассеяния на дислокациях, его можно комбинировать с решеточным рассеянием обычным способом, т. е. путем сложения соответствующих обратных времен релаксации. Проведенная выше оценка τ_d дает численное значение на три порядка большее, чем время релаксации для решеточного рассеяния при комнатных температурах. Таким образом, весьма маловероятно, чтобы рассеяние на дислокациях играло существенную роль при комнатных температурах, по крайней мере если плотность дислокаций не превышает 10^9 см^{-2} . При этом необходимо отметить, что кристалл, содержащий 10^9 см^{-2} дислокаций, должен быть очень сильно деформированным. Однако при низких температурах рассеяние на дислокациях может стать весьма существенным. Имеется пока мало экспериментальных данных о рассеянии на дислокациях, хотя некоторые опыты, проведенные Пирсоном, Ридом и Моринном [39], как будто указывают на существование такого низкотемпературного механизма рассеяния, однако есть сомнения относительно интерпретации этих экспериментальных данных.

При высоких концентрациях примесей (для Si и Ge $\geq 10^{19} \text{ см}^{-3}$) в кристалле могут возникнуть заметные механические напряжения, приводящие к появлению дислокаций. Дэга и Кокль [40] объясняют наличием дислокаций наблюдаемые при таких концентрациях отклонения измеренных величин подвижности от значений, вычисленных по формуле (8.47а).

8.12. Вклад различных видов рассеяния в величину подвижности

Для любого конкретного полупроводника определение вкладов различных типов рассеяния в разных областях температур в величину подвижности является очень сложным процессом. Относительные вклады различных типов рассеяния в сильной степени изменяются от одного полупроводника к другому, так что невозможно дать сколько-нибудь общий результат. В широком смысле, однако, можно сказать, что колебания решетки различных типов дают основной вклад в рассеяние в области высоких температур, а в более чистых материалах этот вид рассеяния может преобладать и при более низких температурах. Для полярных полупроводников очень важным будет рассеяние на оптических колебаниях. Примесное рассеяние, конечно, в сильной степени связано с содержанием примеси, а рассеяние на дислокациях — с деформацией

кристалла. Очевидно, при очень низких температурах доминирует рассеяние, обусловленное несовершенствами кристалла, и очень важным будет рассеяние на нейтральных примесях. Более подробное обсуждение конкретных полупроводниковых материалов мы отложим до гл. 13.

ЛИТЕРАТУРА

1. *Smith R. A.*, The Wave-Mechanics of Crystalline Solids, Chapman and Hall, 1969.
2. *Shockley W., Bardeen J.*, Phys. Rev., 77, 407 (1950); 80, 72 (1950).
3. *Herring C.*, Bell Syst. Techn. Journ., 34, 237 (1955).
4. *Blatt F. J.*, Solid State Physics, Academic Press, 4, 1957, p. 199. [Имеется перевод: Блатт Ф. Дж. Теория подвижности электронов в твердых телах.— М.— Л.: Физматгиз, 1963.]
5. *Sodha M. S.*, Progress in Semiconductors, Heywood, 3, 1958, p. 153.
6. *Moll J. L.*, Physics of Semiconductors, McGraw-Hill, 1964.
7. *Conwell Esther M.*, Solid State Physics, Academic Press, suppl. 9, 1967. [Имеется перевод: Конуэлл Э. Кинетические свойства полупроводников в сильных электрических полях.— М.: Мир, 1970.]
8. *Fawcett W., Paige E. G. S.*, Journ. Phys. C: Solid State Phys., 4, 1801 (1971); *Herbert D. C., Fawcett W., Lettington A. H., Jones D.*, Proc. XIth Intern. Conf. Phys. Semicond., Polish. Sci. Publ., 1972, p. 1221.
9. *Picus G. E., Bir G. L.*, Proc. Vth Intern. Conf. Phys. Semicond., Czechoslovak Akad. Sci., 1960, p. 89.
10. *Sosnowski L.*, Proc. VIIth Intern. Conf. Phys. Semicond., Dunod, 1964, p. 341.
11. *Виноградова К. И., Галаванов В. В., Наследов Д. И.*, ФТТ, 4, 1673 (1962).
12. *Byszeowski B., Kotodziejczak J., Żukotynski S.*, Phys. Stat. Sol., 3, 1880 (1963).
13. *Fröhlich H., Mott N. F.*, Proc. Roy. Soc. A, 171, 496 (1939).
14. *Howarth D., Sondheimer E. H.*, Proc. Roy. Soc. A, 219, 53 (1953).
15. *Low F. E., Pines D.*, Phys. Rev., 91, 193 (1953).
16. *Petritz R. L., Scanlon W. W.*, Phys. Rev., 97, 1620 (1955).
17. *Fröhlich H., Paranjape B. V.*, Proc. Phys. Soc., B69, 21 (1956).
18. *Fröhlich H., Paranjape B. V., Kuper C. G., Nakajima S.*, Proc. Phys. Soc., B69, 842 (1956).
19. *Sodha M. S., Eastman P. C.*, Zs. Phys., 150, 242 (1958).
20. *Mott N. F., Messey H. S. W.*, Theory of Atomic Collisions, Oxford University Press, 1933, ch. 3. [Имеется перевод: Мотт Н., Мессси Г. Теория атомных столкновений.— М.: Мир, 1969.]
21. *Conwell Ester M., Weisskopf V. F.*, Phys. Rev., 77, 388 (1950).
22. *Dingle R. B.*, Phil. Mag., 46, 831 (1955).
23. *Brooks H.*, Advances in Electronics and Electron Physics, 7, 85 (1955).
24. *Mott N. F.*, Proc. Camb. Phil. Soc., 32, 281 (1936).
25. *Schiff L. I.*, Quantum Mechanics, McGraw-Hill, 1955, p. 170. [Имеется перевод: Шифф Л. Квантовая механика.— М.: ИЛ, 1959.]
26. *Mansfield R.*, Proc. Phys. Soc., B69, 76 (1956).
27. *Ham F.*, Phys. Rev., 100, 1251 (1955).
28. *Glicksman M.*, Phys. Rev., 108, 264 (1957).
29. *Bardeen J., Pearson G. L.*, Phys. Rev., 75, 865 (1949).
30. *Erginsoy C.*, Phys. Rev., 79, 1013 (1950).
31. *Massey H. S. W., Moisewitsch B. L.*, Phys. Rev., 78, 180 (1950),

32. *McGill T. C., Baron R.*, Phys. Rev., **11**, 5208 (1975).
33. *Baron R., Young M. H., McGill T. C.*, Proc. XIII th Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 1158.
34. *Благосклонская Л. Е., Гершензон Е. М., Ладъжинский Ю. П., Попова А. П.*, ФТТ, **11**, 2967 (1969).
35. *Solomon I.*, Proc. XI th Intern. Conf. Phys. Semicond., Polish Sci. Publ., 1972, p. 1.
36. *Maxwell R., Honig A.*, Phys. Rev. Lett., **17**, 188 (1966).
37. *Dexter D. L., Seitz F.*, Phys. Rev., **86**, 964 (1952).
38. *Read W. T.*, Phil. Mag., **45**, 775, 1119 (1954); **46**, 111 (1954).
39. *Pearson G. L., Read W. T., Morin F. J.*, Phys. Rev., **93**, 666 (1954).
40. *Daga P. O., Khokle W. S.*, Journ. Phys. C: Solid State Phys., **4**, 190 (1971); **5**, 3473 (1972).

РЕКОМБИНАЦИЯ ЭЛЕКТРОНОВ И ДЫРОК

9.1. Механизмы рекомбинации

В предыдущей главе были рассмотрены различные механизмы, определяющие времена релаксации τ_e и τ_h . Теперь следует провести подробное обсуждение процессов, посредством которых происходит рекомбинация электронов и дырок при восстановлении нарушенного равновесного состояния. Процессы рекомбинации можно разделить на три основных типа: а) прямая рекомбинация, б) рекомбинация через центры захвата (ловушки), в) поверхностная рекомбинация¹⁾. В первом случае воссоединение электрона с дыркой происходит благодаря однократному переходу электрона из состояния в зоне проводимости в пустое состояние валентной зоны. При этом каким-либо способом должен рассеяться избыток энергии ΔE . Простейшим процессом рассеяния этой энергии является излучение, которое может сопровождаться или не сопровождаться испусканием или поглощением фононов.

Вместо того чтобы непосредственно рекомбинировать с дыркой, электрон может сначала образовать с дыркой экситон (см. разд. 3.5). Затем экситон может исчезнуть, испустив при этом квант с энергией, несколько *меньшей*, чем ширина запрещенной зоны ΔE . Вклад свободных экситонов в излучательную рекомбинацию учитывается, как будет видно из дальнейшего, тем, что при вычислении вероятности излучательных переходов экситонное излучение рассматривается совместно с излучением, возникающим при межзонных переходах. Однако если происходит захват экситона, то экситон может длительное время оставаться неподвижным и не давать вклада в излучательное время жизни (за исключением того, что процесс

¹⁾ Обычно процессы рекомбинации неравновесных носителей заряда делят на два основных типа: 1) прямую рекомбинацию, или рекомбинацию зона — зона и 2) рекомбинацию с участием примесных центров или (и) дефектов. В то же время рекомбинация может происходить как на поверхности полупроводникового образца, так и в его объеме. При этом на поверхности, так же как и в объеме, могут иметь место оба типа рекомбинации.— *Прим. ред.*

захвата связывает электроны и дырки в пары). В этом случае излучение часто наблюдается в виде серии очень узких линий, так что его легко отделить от непрерывного спектра излучения, который дают межзонные переходы (см. разд. 10.14).

Электрон может рекомбинировать с дыркой, передавая избыточную энергию E другому электрону зоны проводимости или дырке в валентной зоне. Возникшие таким образом носители заряда с высокой энергией будут передавать ее другим свободным носителям. Этот процесс называется рекомбинацией Оже, поскольку по своему принципу он подобен известному эффекту Оже, при котором атом переходит в состояние с меньшей энергией, отдавая избыток энергии в виде кинетической энергии электрона, а не в виде излучения¹⁾. Одно время считалось, что вероятность такого процесса рекомбинации слишком мала, чтобы он мог играть в полупроводниках существенную роль, однако теперь известно, что это не так. Мы рассмотрим рекомбинацию Оже после излучательной рекомбинации.

Третий процесс, с помощью которого электрон может отдать энергию ΔE , — это передача ее некоторому количеству фононов. Даже фононы с максимальной энергией — оптические фононы имеют энергию порядка 0,1 эВ, так что при $\Delta E = 1$ эВ в таком процессе должны участвовать около десяти фононов. Считается, что вероятность излучения такого большого количества фононов при межзонных переходах очень мала, однако если в процессе рекомбинации участвуют ловушки, многофононные процессы могут оказаться важными.

Ко второму типу процессов рекомбинации относятся такие, при которых переход электрона из зоны проводимости в валентную зону осуществляется двумя последовательными актами, которые могут быть разделены относительно большим промежутком времени. Сначала электрон может быть захвачен глубокой ловушкой, энергетический уровень которой расположен вблизи середины запрещенной зоны. Вслед за тем заполненное состояние может захватить дырку, и этим процесс рекомбинации завершается. Как будет показано, вероятность излучательных переходов быстро уменьшается по мере увеличения излучаемой энергии, так что двухступенчатые переходы, в процессе которых рассеивается существенно меньшая энергия, становятся значительно более вероятными, чем одноступенчатые. Двухступенчатые переходы необходимо ясно отличать от двойных переходов, рассматриваемых в связи с непрямыми переходами, так как последние столь быстро следуют друг за другом, что энергия промежуточного состояния становится неопределенной. Третий из названных выше процессов сходен со вторым, но центры рекомбинации в этом случае находятся на поверхности полупроводника. Очевидно, в этом случае для безызлучательного процесса, в

¹⁾ См., например, [1].

котором энергия передается колебаниям решетки, требуется меньшее количество фононов, и он будет более вероятен, чем прямой многофононный процесс.

9.2. Излучательная рекомбинация

Рассмотрим сначала излучательную рекомбинацию. При определении темпа рекомбинации не будем делать различия между прямыми и непрямыми переходами (см. разд. 10.5.1 и 10.5.3), поскольку он зависит лишь от полного коэффициента поглощения, связанного с межзонными переходами. Пусть $\alpha(\nu)$ — коэффициент поглощения на частоте ν . Тогда средняя длина свободного пробега фотона с частотой ν до возбуждения межзонного перехода просто равна α^{-1} . Следовательно, среднее время жизни $\tau(\nu)$ такого фотона равно $(\alpha\nu_g)^{-1}$, где ν_g — скорость перемещения фотона в среде. Ясно, что она равна групповой скорости, т. е. скорости перемещения группы фотонов в среде. Скорость ν_g равна $d\nu/d(1/\lambda) = cd\nu/d(n\nu)$, где n — показатель преломления. Отсюда

$$\frac{1}{\tau(\nu)} = \frac{\alpha c d\nu}{d(n\nu)}. \quad (9.1)$$

Выражая α через показатель поглощения κ , получаем $\alpha = 4\pi n \kappa \nu / c$ (см. разд. 10.1), так что

$$\frac{1}{\tau(\nu)} = \frac{4\pi n \kappa d\nu}{d(n\nu)}. \quad (9.2)$$

Для того чтобы найти темп генерации в условиях теплового равновесия для области частот от ν до $\nu + d\nu$, следует $1/\tau(\nu)$ умножить на число фотонов, частоты которых соответствуют данному интервалу частот. Число их дается хорошо известной формулой Планка, которая обычно выражается как функция длины волны, так как в этом случае нет зависимости от показателя преломления. Число фотонов $N'(\lambda)d\lambda$ с длинами волн в области от λ до $\lambda + d\lambda$ дается выражением

$$N'(\lambda) d\lambda = \frac{8\pi}{\lambda^4} \frac{d\lambda}{\exp\left(\frac{h\nu}{kT}\right) - 1}, \quad (9.3)$$

из которого легко получить $N(\nu)d\nu$ — число фотонов, имеющих частоты в интервале от ν до $\nu + d\nu$. Поскольку $\lambda = c/n\nu$, то

$$-d\lambda = \frac{c}{n^2\nu^2} \frac{d(n\nu)}{d\nu} d\nu, \quad (9.4)$$

а после подстановки этой величины в выражение (9.3) получим

$$N(\nu) d\nu = \frac{8\pi\nu^2 \left[n^2 \frac{d(n\nu)}{d\nu} \right] d\nu}{c^3 \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]}. \quad (9.5)$$

Темп рекомбинации в условиях теплового равновесия равен темпу генерации для каждого интервала dv . Если обозначить его через $\mathcal{R}(v)dv$, то получим

$$\mathcal{R}(v)dv = \frac{N(v)}{\tau(v)} dv = \frac{32\pi^2 n^3 v^3 \kappa dv}{c^3 \left[\exp\left(\frac{hv}{kT}\right) - 1 \right]}. \quad (9.6)$$

Это выражение дает спектральную плотность излучения, которая должна наблюдаться при излучательном процессе рекомбинации.

Чтобы найти полный темп рекомбинации \mathcal{R} , необходимо проинтегрировать (9.6) по всем значениям v . Таким образом получим

$$\mathcal{R} = \frac{32\pi^2 c^3 k^4 T^4}{h^4} \int_0^\infty \frac{\kappa n^3 x^3 dx}{e^x - 1}, \quad (9.7)$$

где $x = hv/kT$. Множитель перед интегралом можно выразить численно в виде $1,785 \cdot 10^{28} (T/300)^4 \text{ м}^{-3} \cdot \text{с}^{-1}$. Эту формулу вывели Русбек и Шокли [2], которые использовали ее для вычисления темпа излучательной рекомбинации в германии. Подынтегральная функция в правой части выражения (9.7) имеет заметную величину только в довольно узкой области частот вблизи края собственной полосы поглощения, поскольку e^x быстро увеличивается с ростом частоты v , а κ быстро уменьшается при переходе к частотам, лежащим ниже частоты, соответствующей краю поглощения (см. разд. 10.5). Таким образом, ясно, что испускание света, соответствующее прямой рекомбинации, ограничено узким интервалом длин волн в области края собственной полосы поглощения.

Теперь можно вычислить время жизни неосновных носителей заряда. Так как темп неравновесной рекомбинации \mathcal{R}' пропорционален p и n , то можно записать $\mathcal{R}' = Anp$. В невырожденном случае при тепловом равновесии имеем также $pn = n_i^2$, так что $A = \mathcal{R}/n_i^2$. Если обозначить $p = p_0 + \Delta p$ и $n = n_0 + \Delta n$, то при $p_0 \ll n_0$ и $\Delta p \ll n_0$, согласно определению τ_p ,

$$\mathcal{R}' - \mathcal{R} = \frac{\Delta p}{\tau_p} = \frac{\mathcal{R}}{n_i^2} (pn - n_i^2) = \frac{\mathcal{R}n_0 \Delta p}{n_i^2}, \quad (9.8)$$

и мы получим

$$\frac{1}{\tau_p} = \frac{\mathcal{R}}{p_0}. \quad (9.9)$$

Для собственного полупроводника, когда $\Delta p = \Delta n \ll n_i$,

$$\frac{1}{\tau_i} = \frac{2\mathcal{R}}{n_i}, \quad (9.10)$$

так что

$$\frac{\tau_p}{\tau_i} = \frac{2p_0}{n_i}. \quad (9.11)$$

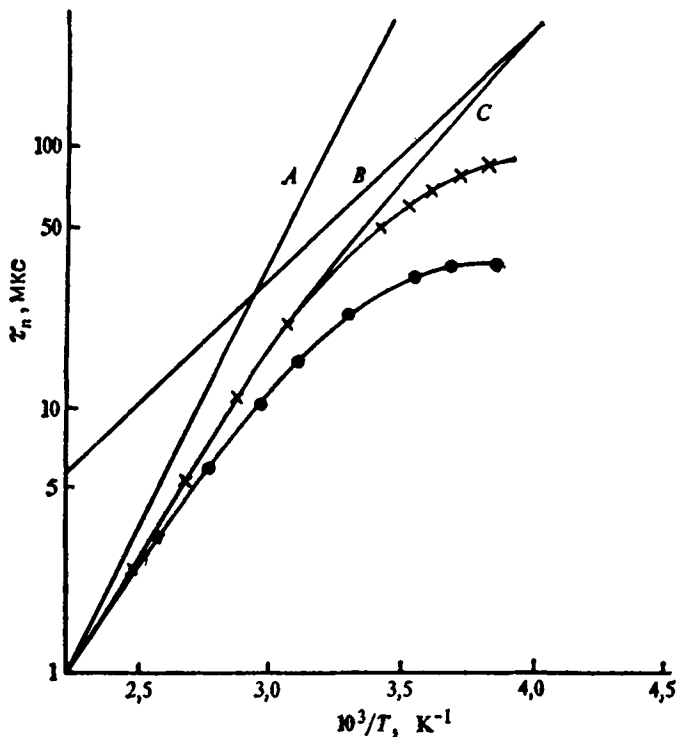


Рис. 9.1. Время жизни неосновных носителей в почти собственном теллуре [3]. A — τ_R , B — τ_a , C — τ_n ($\tau_n^{-1} = \tau_R^{-1} + \tau_a^{-1}$). Крестики — экспериментальные результаты, полученные на образце почти собственного Te, точки — экспериментальные значения для менее чистого образца.

Полученная Русбреком и Шокли величина темпа рекомбинации для германия при $T=300\text{К}$ составляет $1,57 \cdot 10^{19} \text{ м}^{-3} \cdot \text{с}^{-1}$.

На рис. 9.1 показана вычисленная Блейкмором [3] температурная зависимость величины τ_n для почти собственного теллура, когда $\tau_n \approx \tau_i$. Там же показаны экспериментально определенные значения. Можно видеть, что τ_i повышается по мере уменьшения температуры главным образом из-за множителя n_i . При температурах порядка 400К излучательный процесс обуславливает основную долю рекомбинации; при 400 К τ_n имеет величину около 2 мкс.

Темп рекомбинации \mathcal{R} быстро растет по мере уменьшения ΔE , а также при увеличении температуры. Этот вид рекомбинации должен быть, по-видимому, наиболее важен для полупроводников с узкой запрещенной зоной при комнатной температуре и выше. При более низких температурах значительно больший темп реком-

бинации будут давать другие процессы, такие, как рекомбинация Оже (см. разд. 9.3) и рекомбинация через ловушки.

Патриция Х. Брилл и Рут Ф. Шварц [4] измерили величину \mathcal{R} в германии при комнатной температуре. Она оказалась равной $2,5 \cdot 10^{19} \text{ м}^{-3} \cdot \text{с}^{-1}$. Для собственного германия это значение дает величину τ_1 порядка 1 с, что много больше наблюдаемых величин τ_n или τ_p . Тем не менее величину \mathcal{R} важно знать, даже если время жизни неосновных носителей определяется другим механизмом рекомбинации, чтобы иметь возможность оценить вероятность испускания излучения из полупроводника.

Даже если излучательный процесс относительно неэффективен как механизм рекомбинации, он может иметь важное прикладное значение для изготовления источников излучения. По этой причине широко исследовалась излучательная рекомбинация таких материалов, как GaAs (см. разд. 10.14). Обширный обзор работ по излучательной рекомбинации в ряде элементарных полупроводников и полупроводниковых соединений, в том числе в некоторых соединениях типа $A^{III}B^V$, был сделан Варшни [5].

При очень низких температурах важную роль будет играть рекомбинация через мелкие ловушки, энергетические уровни которых расположены вблизи зоны проводимости (в случае доноров) или вблизи валентной зоны (в случае акцепторов).

Из предыдущего анализа следует, что темп излучательной рекомбинации зависит от *полного* числа межзонных переходов на любой заданной частоте ν . Относительные величины вкладов, связанных с прямыми и непрямыми переходами, можно определить лишь при наличии данных о соответствующих вероятностях переходов.

9.3. Рекомбинация Оже

Вычисление темпа рекомбинации Оже требует подробного квантовомеханического анализа процесса столкновений, описанного в разд. 9.1. Такой анализ выполнили Ландсберг и Битти [6]. Для случая собственного полупроводника они получили выражение

$$\tau_{ai} = \frac{A \varepsilon^2 m_0 (m_v + 2m_c) (m_c + m_v)^{1/2}}{m_c m_v^{3/2}} \left(\frac{\Delta E}{kT} \right)^{3/2} \exp \left[\frac{\Delta E}{kT} \left(\frac{2m_c + m_v}{m_c + m_v} \right) \right], \quad (9.12)$$

где, как обычно, ε — диэлектрическая проницаемость, а A — константа, приблизительно равная $4 \cdot 10^{-16} \text{ с}^{-1}$. Можно видеть, что τ_{ai} быстро растет, когда температура становится достаточно малой, и таким образом, рекомбинация Оже, как и излучательная рекомбинация, будет наиболее важна при высоких температурах. При этом

время жизни τ_a увеличивается не так быстро, как в случае излучательной рекомбинации. Это было хорошо проиллюстрировано вычислениями и экспериментами Блекмора с почти собственным теллуrom — рис. 9.1. На этом рисунке можно видеть, что при понижении температуры экспериментальные значения τ_a становятся заметно меньшими, чем время жизни, обусловленное только излучательной рекомбинацией τ_R , но хорошо согласуются с величиной времени жизни, которое получается при учете как рекомбинации Оже, так и излучательной рекомбинации. Результирующее время жизни определяется выражением

$$\frac{1}{\tau_n} = \frac{1}{\tau_R} + \frac{1}{\tau_a}. \quad (9.13)$$

Вычисленные таким образом значения времени жизни для почти собственного теллура хорошо согласуются с измеренными значениями всюду, за исключением области низких температур. Для менее чистого образца величина τ_n при понижении температуры возрастает менее быстро и достигает приблизительно постоянного значения, которое зависит от концентрации примеси.

В условиях термодинамического равновесия рекомбинация Оже в точности сбалансирована соответствующим обратным процессом — генерацией электронно-дырочных пар быстрыми электронами и быстрыми дырками. Пусть G_e и G_h — темпы генерации для этих двух процессов, тогда в равновесии должны выполняться равенства

$$G_e = R_{ae}, \quad G_h = R_{ah}, \quad (9.14)$$

где R_{ae} и R_{ah} — темпы рекомбинации для двух процессов, в которых энергия передается соответственно электронам и дыркам. Тогда имеем

$$R_a = R_{ae} + R_{ah} = G_{ae} + G_{ah}. \quad (9.14a)$$

Пусть R'_{ae} и R'_{ah} — темпы рекомбинации для неравновесных условий. Для невырожденного случая величина R'_{ae} пропорциональна n^2 , поскольку в процессе рекомбинации участвуют два электрона, и, кроме того, пропорциональна p , поэтому для R'_{ae} имеем

$$R'_{ae} = \frac{n^2 p R_{ae}}{n_0 p_0}, \quad (9.15)$$

где n_0 и p_0 — равновесные концентрации электронов и дырок.

С другой стороны, темп генерации G'_{ae} пропорционален n , поскольку он зависит только от числа имеющихся быстрых электронов, так что

$$G'_{ae} = \frac{G_{ae} n}{n_0} = \frac{R_{ae} n}{n_0}. \quad (9.16)$$

В неравновесных условиях результирующий темп рекомбинации дается выражением

$$R'_{ae} - G'_{ae} = G_{ae} \left(\frac{np}{n_0 p_0} - 1 \right) \frac{n}{n_0} \quad (9.17)$$

для электронов и аналогичным выражением для дырок. Таким образом, для времени жизни τ_a имеем

$$\frac{1}{\tau_a} = \frac{n(n+p_0)G_{ae}}{n_0^2 p_0} + \frac{p(p+n_0)G_{ah}}{p_0^2 n_0}. \quad (9.18)$$

Оказывается, что в этом выражении обычно преобладает какой-либо один член.

Действительно, если $m_e \ll m_h$, то первый член будет намного больше второго. В этом случае можно записать

$$\frac{1}{\tau_a} = \frac{n(n+p_0)G_{ae}}{n_0 n_i^2}. \quad (9.19)$$

Если $\Delta n = \Delta p$ мало, то

$$\frac{1}{\tau_a} = \frac{(n_0 + p_0)G_{ae}}{n_i^2}. \quad (9.20)$$

Поскольку очевидно, что G_{ae} пропорционально n_0 , запишем

$$G_{ae} = n_0 g_{ae} \quad (9.21)$$

и

$$\frac{1}{\tau_a} = \frac{n_0(n_0 + p_0)}{n_i^2} g_{ae}. \quad (9.22)$$

Для собственных материалов при этих же условиях имеем

$$\frac{1}{\tau_i} = 2g_{ae}, \quad (9.23)$$

и можем записать

$$\frac{\tau_i}{\tau_a} = \frac{n_0(n_0 + p_0)}{2n_i^2}. \quad (9.24)$$

Это выражение показывает связь величины τ_a со степенью легирования полупроводника.

Можно записать более общее выражение

$$\frac{\tau_i}{\tau_a} = \frac{(n+p_0)(n+c_a p)}{2n_i^2}, \quad (9.25)$$

где

$$c_a = \frac{g_{ah}}{g_{ae}} \quad (9.26)$$

и

$$G_{ah} = p_0 g_{ah}. \quad (9.27)$$

Ясно, что выражение (9.25) сводится к (9.24), когда $c_a p \ll n$. Если, как обычно, $c_a \ll 1$, то (за исключением случая материалов сильно выраженного p -типа, где $p \gg n$) применимо более простое выражение (9.24).

Если $\Delta n = \Delta p$ много меньше, чем n или p , уравнение (9.25) может быть записано в более простом виде:

$$\frac{\tau_i}{\tau_a} = \frac{n_0^2}{2n_i^2} + \frac{1}{2} (1 + c_a) + \frac{c_a n_i^2}{2n_0^2} = \quad (9.28)$$

$$= \frac{n_i^2}{2p_0^2} + \frac{1}{2} (1 + c_a) + \frac{c_a p_0^2}{2n_i^2}. \quad (9.29)$$

Из приведенных выражений можно видеть, что τ_a имеет максимальное значение $(1 + c_a)\tau_i$ при $n_0 = n_i$ и уменьшается, если n_0 или p_0 возрастают свыше величины n_i . В материале p -типа, однако, τ_a снижается значительно медленнее, т. к. обычно $c_a \ll 1$.

Подробное обсуждение поведения τ_a в зависимости от концентрации носителей заряда и температуры дал Блекмор [7].

Если существует отщепленная валентная зона, обусловленная связью спина с орбитальным движением дырки (см. разд. 2.3), а величина спин-орбитального расщепления ΔE_s приблизительно равна ширине запрещенной зоны ΔE , то рожденные в процессе рекомбинации Оже дырки с высокой энергией могут перейти в отщепленную зону практически без изменения квазиимпульса. Вероятность рекомбинации Оже в этих условиях очень высока. Такая ситуация реализуется, например, в соединении GaSb, в котором этот процесс исследовали Бенц и Конрад [8]. Заметное число дырок, оказавшихся в отщепленной зоне в результате рекомбинации Оже, при низких температурах, по-видимому, осуществляет излучательный переход на мелкие нейтральные доноры. Такую рекомбинацию в некотором смысле можно считать комбинированным процессом рекомбинации Оже и рекомбинации с участием ловушек.

В сильно легированном полупроводнике при наличии вырождения темп рекомбинации уже не пропорционален $n^2 p$, и уравнение (9.15) неприменимо. Этот случай рассмотрел Хог [9], который показал, что в определенных условиях следует ожидать зависимости вида $R'_{ae} \sim n^{1/2} p$.

Вопрос о межзонной рекомбинации Оже достаточно полно обсужден Ландсбергом [10] в обзоре, посвященном безызлучательным рекомбинационным процессам.

9.4. Рекомбинация через ловушки

Теорию рекомбинации через локальные центры захвата (ловушки) разработали Шокли и Рид [11]. Ниже следует упрощенный вариант их теории. Прежде всего вычислим темп захвата электро-

нов из зоны проводимости. Он будет пропорционален n и $(1-f_t)$, где f_t — доля ловушек, которые заполнены электронами. Следовательно, темп захвата \mathcal{R}_c можно записать в виде

$$\mathcal{R}_c = C_n n (1-f_t), \quad (9.30)$$

где C_n — вероятность захвата одного электрона, когда все ловушки пусты. Значение C_n должно быть пропорционально концентрации ловушек N_t . Темп выброса электронов из ловушек \mathcal{R}_e должен быть пропорционален f_t и, следовательно, его можно представить в виде

$$\mathcal{R}_e = C'_n f_t. \quad (9.31)$$

В состоянии термодинамического равновесия $\mathcal{R}_c = \mathcal{R}_e$, так что

$$C'_n = \frac{C_n n_0 (1-f_{t_0})}{f_{t_0}},$$

где n_0 и f_{t_0} — равновесные значения n и f_t . Определить f_{t_0} можно из функции распределения Ферми

$$\frac{1-f_{t_0}}{f_{t_0}} = \exp \frac{E_t - E_F}{kT}, \quad (9.32)$$

где E_t — энергетический уровень ловушки. Рассмотрим пока одинаковые ловушки с одним энергетическим уровнем. При условии отсутствия вырождения¹⁾ получим

$$n_0 = N_c \exp \frac{E_F}{kT} \quad (9.33)$$

[см. уравнение (4.29)], следовательно,

$$C'_n = C_n N_c \exp \frac{E_t}{kT} = n_1 C_n, \quad (9.34)$$

где n_1 — концентрация электронов в зоне проводимости, когда уровень Ферми совпадает с уровнем ловушек. Теперь результирующий темп захвата электронов можно записать в виде

$$\mathcal{R}_n = C_n [(1-f_t) n - n_1 f_t]. \quad (9.35)$$

Точно таким же образом вычисляется результирующий темп захвата дырок \mathcal{R}_p , который можно представить как

$$\mathcal{R}_p = C_p [f_t p - p_1 (1-f_t)]. \quad (9.36)$$

Если имеется какой-либо внешний источник, генерирующий электронно-дырочные пары с постоянным темпом \mathcal{R} , то после установления стационарного состояния электроны и дырки будут захватываться в одном и том же темпе, так как они рекомбинируют парами. Следовательно, $\mathcal{R}_n = \mathcal{R}_p = \mathcal{R}$. Из этого соотношения можно опре-

¹⁾ Влияние вырождения проанализировал Ландсберг [12].

делить долю занятых электронами ловушек f_t . В общем случае f_t отличается от равновесной величины f_{t_0} . Приравнявая \mathcal{R}_n и \mathcal{R}_p , получаем

$$f_t = \frac{nC_n + p_1C_p}{C_n(n + n_1) + C_p(p + p_1)}. \quad (9.37)$$

Таким образом, темп рекомбинации \mathcal{R} определяется выражением

$$\mathcal{R} = \frac{C_n C_p (pn - n_1^2)}{C_n(n + n_1) + C_p(p + p_1)}, \quad (9.38)$$

поскольку $n_1 p_1 = n_1^2$. Если обозначить $n = n_0 + \Delta p$ и $p = p_0 + \Delta p$ и предположить, что $\Delta n = \Delta p$, то получим

$$\mathcal{R} = \frac{n_0 \Delta p + p_0 \Delta p + \Delta p^2}{C_p^{-1}(n_0 + n_1 + \Delta p) + C_n^{-1}(p_0 + p_1 + \Delta p)}. \quad (9.39)$$

Итак, если τ определяется соотношением $\mathcal{R} = \Delta p / \tau$, то для τ имеем

$$\frac{1}{\tau} = \frac{n_0 + p_0 + \Delta p}{C_p^{-1}(n_0 + n_1 + \Delta p) + C_n^{-1}(p_0 + p_1 + \Delta p)}. \quad (9.40)$$

Когда Δp мало, а $n_0 \gg p_0$ и $n_0 \gg n_1$, то

$$\frac{1}{\tau} = \frac{1}{\tau_{p_0}} \equiv C_p. \quad (9.41)$$

Аналогично если $p_0 \gg n_0$ и $p_0 \gg p_1$, то

$$\frac{1}{\tau} = \frac{1}{\tau_{n_0}} \equiv C_n. \quad (9.42)$$

Таким образом, выражение (9.40) можно переписать в виде

$$\frac{1}{\tau} = \frac{n_0 + p_0 + \Delta p}{\tau_{p_0}(n_0 + n_1 + \Delta p) + \tau_{n_0}(p_0 + p_1 + \Delta p)}. \quad (9.43)$$

Если теперь Δp мало, то

$$\tau = \tau_{p_0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n_0} \frac{p_0 + p_1}{n_0 + p_0}. \quad (9.44)$$

В случае сильно легированных полупроводников n - или p -типа величины τ_p и τ_n являются постоянными, не зависящими от концентрации носителей заряда. В этом и состоит отличие рекомбинации через ловушки от излучательной рекомбинации, где τ_p пропорционально p_0^{-1} .

На рис. 9.2, а показана зависимость времени жизни носителей заряда от их концентрации, вытекающая из выражения (9.44) для малых величин Δp . Значения τ проходят через максимум при определенном значении n_0 , зависящем от n_1 . Если $\tau_{n_0} = \tau_{p_0}$, то максимум

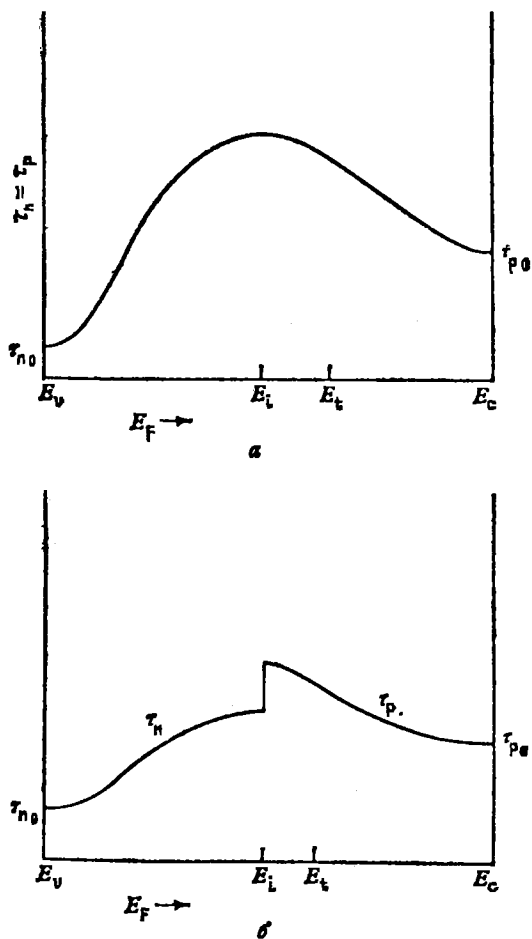


Рис. 9.2. Зависимость времени жизни, обусловленного рекомбинацией через ловушки, от концентрации носителей, выраженной через уровень Ферми.

достигается при $n_0 = p_0 = n_1$. Тогда имеем

$$\tau = \tau_{p_0} \left(1 + \frac{n_1 + p_1}{n_0 + p_0} \right). \quad (9.45)$$

При $n = n_1$

$$\tau_1 = \tau_{p_0} \left(1 + \frac{n_1 + p_1}{2n_1} \right), \quad (9.46)$$

или

$$\tau_1 = \tau_{p_0} \left(1 + \operatorname{ch} \frac{E_t - E_1}{kT} \right), \quad (9.47)$$

где E_i — уровень Ферми в собственном полупроводнике. Отсюда видно, что если $E_i - E_i \gg kT$, то $\tau_i \gg \tau_{p_0}$. Если же $E_i = E_i$, то $\tau_i = 2\tau_{p_0}$.

Зависимость τ от концентрации носителей заряда в германии как n -, так и p -типа исследовали Бартон, Халл, Морин и Северинс [13], которые получили прекрасное согласие с уравнением (9.44). Изменения n и p осуществляются введением в кристалл германия примесей Sb и In, а изменения концентрации ловушек — введением Cu и Ni, которые создают центры рекомбинации, энергетические уровни которых располагаются вблизи середины запрещенной зоны. В этом случае оказалось, что $\tau_{p_0}/\tau_{n_0} = 0,1$.

Когда N_t становится сравнимым с n_0 или p_0 , предположение $\Delta n = \Delta p$ не выполняется и вместо него имеет место соотношение

$$\Delta n = \Delta p - N_t \Delta f_t. \quad (9.48)$$

Более того, в таком случае нельзя считать $\tau_n = \tau_p$, так что

$$\mathcal{R} = \frac{\Delta p}{\tau_p} = \frac{\Delta n}{\tau_n}. \quad (9.49)$$

Когда $n > p$, то мы обычно имеем дело с τ_p , а когда $n < p$, то с τ_n . Рассмотрим теперь случай только малых значений Δn и Δp и не будем при этом пренебрегать Δf_t . Из равенства \mathcal{R}_n и \mathcal{R}_p получим [см. выражения (9.35) и (9.36)]

$$C_n [(1-f_t) \Delta n - (n_0 + n_1) \Delta f_t] = C_p [f_t \Delta p + (p_0 + p_1) \Delta f_t],$$

так что

$$\Delta f_t = \frac{\tau_{p_0} (1-f_t) \Delta n + \tau_{n_0} f_t \Delta p}{\tau_{n_0} (p_0 + p_1) + \tau_{p_0} (n_0 + n_1)}. \quad (9.50)$$

В этом приближении можно воспользоваться равновесной величиной f_t , а именно $f_t = n_0 / (n_0 + n_1) = 1 - [p_0 / (p_0 + p_1)]$, что дает

$$\Delta f_t = \frac{\tau_{p_0} (n_0 + n_1) p_0 \Delta n + \tau_{n_0} (n_0 + n_1) n_0 \Delta p}{(n_0 + n_1) (p_0 + p_1) [\tau_{n_0} (p_0 + p_1) + \tau_{p_0} (n_0 + n_1)]}. \quad (9.51)$$

Отсюда видно, что $N_t \Delta f_t \ll \Delta n$ или Δp , если $N_t \ll (n_0 + n_1)$ или $(p_0 + p_1)$. Подставляя в уравнение (9.38) Δn и Δf_t , выраженные через Δp , получаем

$$\tau_p = \frac{\tau_{n_0} (p_0 + p_1) + \tau_{p_0} [(n_0 + n_1) + N_t n_1 / (n_1 + n_0)]}{n_0 + p_0 + N_t n_1 n_0 / (n_1 + n_0)^2} \quad (9.52)$$

и

$$\tau_n = \frac{\tau_{p_0} (n_0 + n_1) + \tau_{n_0} [(p_0 + p_1) + N_t p_1 / (p_1 + p_0)]}{n_0 + p_0 + N_t p_1 p_0 / (p_1 + p_0)^2}. \quad (9.53)$$

При $n_0 \rightarrow n_1$ или $p_0 \rightarrow p_1$ уже не будет иметь место $\tau_p \rightarrow \tau_n$. Если построить графики τ_p для $n > n_1$ и τ_n для $n < n_1$, то получим зависимость, представленную на рис. 9.2, б. Максимальное значение ве-

личин τ_n и τ_p и в этом случае будет наблюдаться у почти собственного, хотя и не обязательно точно собственного полупроводника.

Времена жизни, определяемые выражениями (9.52) и (9.53), выведены для стационарного случая. Если члены, содержащие N_1 , существенны, то вовсе не очевидно, что и для переходных условий получится то же самое время жизни. Если число электронов на ловушках меняется, то скорость изменения Δn не обязательно совпадает со скоростью изменения Δp . Этот эффект может оказаться существенным при определении темпа затухания фотопроводимости после выключения возбуждающего излучения. Он будет рассматриваться в разд. 10.9.2. В приведенном выше расчете значения C_n , C_p и связанные с ними времена жизни τ_{n0} , τ_{p0} были введены как неизвестные параметры. Различные попытки вычислить эти значения имели лишь ограниченный успех. Мы обсудим методы, используемые для оценки истинных значений этих величин. Однако приведенные ранее формулы дают возможность оценить их на основании экспериментальных исследований изменений τ_n и τ_p в зависимости от равновесной концентрации носителей и интенсивности внешнего возбуждения. Мэни и Брей [14] описали несколько способов таких оценок.

Для объяснения того, каким образом свободный электрон или дырка могут потерять значительное количество энергии при захвате ловушкой, были предложены различные механизмы. Если считать, что в процессе захвата электрона вся его энергия передается решётке сразу, то при этом, как мы видели, должно одновременно испускаться значительное количество фононов. Чтобы преодолеть эту трудность, Лэкс [15] предположил, что электрон (или дырка) захватывается на высокорасположенное возбужденное состояние примесного центра. В этом случае количество энергии, отдаваемой электроном при захвате, невелико, и такой процесс может идти с испусканием лишь одного акустического фонона. Затем в процессе захвата электрон совершает серию переходов между возбужденными состояниями центра, при этом для осуществления каждого перехода требуется испускание только одного фонона. Расчет Лэкса был преимущественно «классическим». Позже его детализировали Браун и Родригес [16], которые применили квантово-механический подход и учли влияние конкретного вида зонной структуры. Они сравнили вычисленные значения с экспериментальными результатами для «водородоподобных» ловушек в Si и Ge при низких температурах. Хотя при этом и было получено некоторое согласие, его все же нельзя считать очень хорошим.

По-видимому, должен иметь место еще какой-то механизм рекомбинации. Карпова и Калашников [17] предположили, что таким механизмом может быть рекомбинация Оже, при которой электрон захватывается ловушкой, а не рекомбинирует с дыркой валентной зоны. Теоретически такой процесс рассмотрели Бонч-Бруевич и

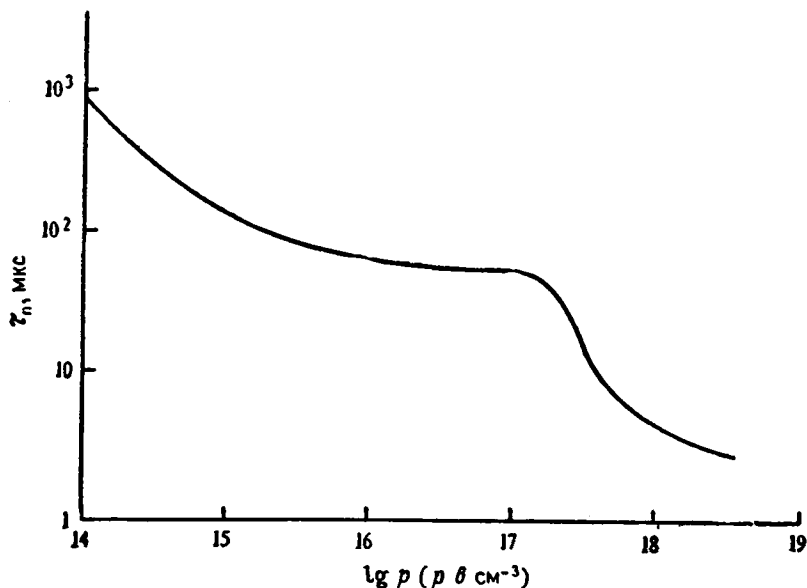


Рис. 9.3. Зависимость τ_n от концентрации дырок p [24].

Гуляев [18], а также Ландсберг, Эванс и Рис-Робертс [19]. Этот механизм объясняет некоторые особенности изменения τ_n и τ_p для содержащего рекомбинационные центры германия, обнаруженные Карповой и Калашниковым [17], в частности быстрый спад значений τ_n и τ_p при концентрациях носителей, превышающих 10^{17} см^{-3} .

Различные более сложные механизмы рекомбинации Оже с участием ловушек были предложены и теоретически рассмотрены рядом авторов, в том числе Шейнкманом [20], и обсуждены Ландсбергом [10] в обзорной статье, посвященной рекомбинационным процессам в полупроводниках.

Другое предположение о возможном механизме передачи в процессе рекомбинации избыточной энергии решетке кристалла основано на учете влияния смещения центра рекомбинации относительно его равновесного положения на потенциальную энергию центра. Это влияние в общем случае будет различным в зависимости от того, занят ли центр рекомбинации электроном или дыркой или нет. Тогда, подобно тому, как это имеет место в молекуле, зависимость потенциальной энергии центра от величины смещения будет изображаться графически соответственно двумя различными кривыми, которые стремятся пересечься. В точке, где они пересекаются (или сближаются), может иметь место безызлучательный переход носителя из одного состояния в другое. Если точка пересечения

кривых не соответствует равновесному положению занятого центра рекомбинации, то переход в незанятое состояние вызовет сильные колебания, так как решетка будет возвращаться в равновесное положение, соответствующее незанятому центру. Это эквивалентно испусканию некоторого количества фононов.

Эту идею рассмотрели с позиций квантовой механики Энглман и Джортнер [21] и развили далее Мотт, Девис и Стрит [22] для описания излучательной рекомбинации на ловушках, в частности в аморфных полупроводниках. Подобный метод вычисления применили также Ярос и Брэнд [23], которые использовали его для вычисления темпа рекомбинации через центры с глубокими уровнями в GaP и GaAs.

Зависимости τ_n и τ_p от концентрации носителей заряда очень разнообразны, но обычно существует область «плато», в которой значение τ почти постоянно. Значение τ увеличивается при понижении концентрации носителей относительно этой области и быстро падает, когда концентрация становится выше некоторого критического значения (10^{17} см^{-3} для Ge). На рис. 9.3 показана для примера зависимость τ_n от концентрации носителей для легированных бором образцов германия, содержащих рекомбинационные центры Cu. Эту зависимость получил Калашников [24], который подробно обсудил значения коэффициентов рекомбинации для многих примесей в Si и Ge

9.5. Рекомбинация через экситоны

Исследования излучения полупроводников в ближней ИК-области спектра показали, что наряду с межзонными переходами имеет место рекомбинация через экситоны, которая дает большую часть рекомбинационного излучения (см. разд. 10.14). Это указывает на то, что до рекомбинации электроны и дырки с большой вероятностью связываются в экситоны и излучение возникает в процессе высвечивания экситонов. Оно возникает при высвечивании не только свободных экситонов, но также экситонов, связанных с донорами и акцепторами, что проявляется в первую очередь в полупроводниках с довольно широкой запрещенной зоной (таких, как GaP) (см. разд. 10.14).

Спектр излучения свободных экситонов сливается со спектром, возникающим при межзонной рекомбинации. В использованном ранее методе вычисления излучательного времени жизни проводилось интегрирование по всему спектру поглощения. Таким образом, наряду с межзонной рекомбинацией учитывалась и рекомбинация через экситоны. Если, однако, мы можем отделить вклад экситонов в поглощение, что для хорошо очищенных полупроводников, по-видимому, возможно, то из этих данных можно определить вре-

мя жизни свободных экситонов. Экситоны, связанные с некоторыми донорами и акцепторами, из-за малой величины энергии связи будут существовать при очень низких температурах. Рекомбинационное излучение связанных экситонов имеет вид очень узкой линии, которую легко идентифицировать (см. разд. 10.14). Эта линия может наблюдаться и в поглощении, что дает возможность вычислить излучательное время жизни связанного экситона тем же способом, который описан в разд. 9.2, интегрируя коэффициент поглощения по ширине линии. Величина времени жизни, измеренная по скорости затухания рекомбинационного излучения при использовании коротких возбуждающих импульсов оказывается на несколько порядков меньше, чем вычисленное время жизни. Нельсон [25] нашел, в частности, что для экситонов, связанных на атомах S в GaP, при 4,2 К один квант излучается приблизительно за 700 актов рекомбинации, в то время как для As в Si при 1,6 К это соотношение составляет приблизительно один к 8000. Должен существовать, следовательно, намного более эффективный процесс рекомбинации, который Нельсон отождествил с рекомбинацией Оже, сопровождающейся выбросом быстрых электронов из экситонного комплекса.

Этот процесс больше похож на первоначальный эффект Оже, чем на оже-рекомбинацию свободных электронов и дырок. Хотя детальный квантовомеханический анализ этого процесса еще не проведен, численные оценки могут быть получены с использованием хорошо известного коэффициента внутренней конверсии для атомов, но с учетом эффективных масс электронов и дырок, а также диэлектрической проницаемости. Такие оценки по порядку величины хорошо согласуются с экспериментальными результатами, почти не оставляя сомнений в том, что рекомбинация Оже является в этом случае доминирующим процессом. Это подкрепляет точку зрения, согласно которой процесс Оже играет важную роль и в рекомбинации носителей заряда через ловушки.

Вопрос о рекомбинации экситонов подробно обсуждал Ландсберг [10] в обзоре процессов рекомбинации, на который мы уже ссылались.

9.6. Рекомбинация на дислокациях

Одним из видов ловушек, представляющих особый практический интерес, являются ловушки, связанные с дислокациями. Как уже отмечалось в разд. 3.1.3, дислокация создает цепь ловушек, проходящую по ненасыщенным связям вдоль дислокаций. Влияние связанного с дислокациями объемного заряда исследовал Моррисон [26], который показал, что этот заряд может заметно влиять на постоянные рекомбинации C_n и C_p . Опыты Мак-Килви [27] с пластически деформированными кристаллами германия показали, что время жизни неосновных носителей заряда в полупроводнике

n -типа начинает зависеть от присутствия краевых дислокаций, если их концентрация превышает 10^4 см^{-2} . Такой же результат получили Вертгейм и Пирсон [28]. Согласно этим измерениям, влияние дислокаций на время жизни неосновных носителей в германии можно выразить следующим образом:

$$\begin{aligned}\tau_{pd} &= \frac{2,5}{N} \text{ с} & (n\text{-тип}), \\ \tau_{nd} &= \frac{0,7}{N} \text{ с} & (p\text{-тип}),\end{aligned}\tag{9.54}$$

где N — плотность дислокаций на 1 см^2 . Эксперименты Блекмора, Шульца и Номуры [29] на Te показали, что темп рекомбинации практически линейно связан с плотностью дислокаций при $N < 5 \times 10^5 \text{ см}^{-2}$, однако при дальнейшем увеличении плотности дислокаций он возрастает менее быстро. При плотности дислокаций $\sim 10^6 \text{ см}^{-2}$ время жизни носителей при комнатной температуре понижалось до величины $\sim 10^{-8} \text{ с}$, в то время как в недеформированных кристаллах оно составляло $\sim 10^{-5} \text{ с}$.

9.7. Рекомбинация через доноры или акцепторы при низких температурах

Концентрация дырок в полупроводнике n -типа при низких температурах столь мала, что ею можно пренебречь, а поэтому интерес представляет только темп захвата возбужденных электронов, т. е. *основных* носителей заряда, донорами. Можно предположить, что захватывающими центрами являются сами доноры, а концентрации p_0 и p_1 ничтожно малы, тогда имеем $n = (1 - f_t) N_d$. Темп захвата электронов донорами определяется выражением (9.35). Если предположить далее, что $n \ll N_d$, то, поскольку C_n пропорционально $N_t (= N_d)$, получим

$$\mathcal{R}_n = B_d (n^2 - n_0^2),\tag{9.55}$$

где n_0 — равновесная концентрация электронов, а B_d — некоторая постоянная. Следовательно, получим

$$\frac{1}{\tau_n} = B_d (2n_0 + \Delta n).\tag{9.56}$$

При низких температурах n_0 очень мало, так что при значительном увеличении концентрации возбужденных носителей заряда величина $\tau_n^{-1} \sim \Delta n$, а темп захвата пропорционален Δn^2 , а не Δn . Тогда

$$\frac{d\Delta n}{dt} = - B_d \Delta n^2.\tag{9.57}$$

Решением этого уравнения при условии, что $\Delta n = \Delta n_0$ при $t=0$ будет

$$\Delta n = \frac{\Delta n_0}{1 + \Delta n_0 / B_d t} \quad (9.58)$$

Следовательно, концентрация избыточных электронов спадает по гиперболическому закону, а не по экспоненте, как обычно.

Горник [30] наблюдал излучение мелких донорных примесей в InSb в области миллиметровых волн. Это означает, что некоторые центры должны захватывать носители, испуская при этом избыточную энергию в виде излучения. Этот процесс, по-видимому, менее эффективен, чем рекомбинация через ловушки с более глубокими энергетическими уровнями. Вероятно, более существенным механизмом рекомбинации в данном случае является рекомбинация типа Оже, которую рассмотрели Браун и Родригес [16] применительно к донорам в Ge и Si.

9.8. Поверхностная рекомбинация

Помимо рекомбинации в объеме полупроводника, электроны и дырки могут диффундировать к поверхности и там рекомбинировать. При определенных условиях темп рекомбинации на поверхности может даже оказаться намного больше темпа рекомбинации в объеме. Действительно, во многих практических случаях необходимо принимать специальные меры предосторожности, чтобы темп поверхностной рекомбинации не стал чрезмерно высоким. Например, если темп поверхностной рекомбинации вблизи точечного контакта слишком велик, то она может отрицательно повлиять на свойства контакта. Если поверхностная рекомбинация не уменьшена специальной обработкой поверхности, она может также играть главную роль в тонких пленках, которые используются, в частности, в фоторезисторах ИК-диапазона. Это может свести на нет преимущества большого объемного времени жизни (см. подразд. 10.9.3).

В равновесном состоянии поток дырок, подходящих к поверхности кристалла n -типа, должен быть равен потоку дырок, движущихся в обратном направлении, причем этот поток включает как отраженные от поверхности, так и вновь возникшие дырки. Но при отклонении от равновесного состояния эти потоки уже не компенсируют друг друга. Можно ожидать, что при малых отклонениях от равновесия число вновь появившихся дырок не будет зависеть от величины отклонения. Пусть средняя вероятность отражения дырки при столкновении с поверхностью равна r . Тогда число дырок, падающих на единичную площадку поверхности за единицу времени, равно $\frac{1}{4} v_t p$, где v_t — средняя тепловая скорость дырок. Число дырок, ушедших с поверхности за единицу времени, будет равно $\frac{1}{4} r v_t p + S$, где S — темп возникновения дырок на единичной

площадке поверхности. В состоянии равновесия будем иметь соотношение

$$\frac{1}{4} v_t p_0 = \frac{1}{4} r v_t p_0 + S_0, \quad (9.59)$$

так что

$$S_0 = \frac{1}{4} (1-r) v_t p_0. \quad (9.60)$$

Таким образом, результирующий темп исчезновения дырок у поверхности, S_a , определяется выражением

$$S_a = \frac{1}{4} (1-r) v_t (p - p_0), \quad (9.61)$$

которое можно записать в виде

$$S_a = s \Delta p, \quad (9.62)$$

где s имеет размерность скорости и называется скоростью поверхностной рекомбинации. Ясно, что в стационарных условиях темп поглощения дырок поверхностью равен темпу их рекомбинации с электронами, в противном случае на поверхности происходило бы непрерывное накопление заряда.

Мы уже упоминали о двух типах ловушек, которые, как известно, имеются на поверхности полупроводника, а именно о так называемых «медленных» и «быстрых» состояниях (см. подразд. 7.11.1). Последние, будучи расположены дальше от поверхности полупроводника, с большей вероятностью действуют как центры рекомбинации. Их энергетические уровни лежат ниже зоны проводимости, но выше валентной зоны, и поэтому они могут действовать подобно ловушкам в процессе рекомбинации Шокли — Рида. Свойства этих ловушек изучал ряд авторов (см. подразд. 7.11.1), в частности Браттейн и Бардин [31], которые показали, что для объяснения результатов, полученных на германии, необходимо предположить наличие ловушек как для электронов, так и для дырок, и что электронные ловушки не могут находиться очень далеко от дна зоны проводимости, а дырочные ловушки — от верхнего края валентной зоны. Если в уравнении (9.43) положить либо $n_1 \gg p_1$, либо $n_1 \ll p_1$, то

$$\mathcal{R} \sim (n_0 + p_0 + \Delta p),$$

что будет приблизительно верно даже в том случае, когда в процессе рекомбинации участвуют два типа ловушек, каждый из которых удовлетворяет этому условию. Применив это соображение к поверхностным ловушкам, получим, что

$$s = B_s (n_0 + p_0 + \Delta p), \quad (9.63)$$

где B_s — постоянная. На основании более строгого анализа Браттейн и Бардин получили выражения такого же вида. Зависимость

с от n_0 и p_0 в случае германия исследовал Шульц [32], который обнаружил довольно хорошее согласие с формулой (9.63).

Определение величины B_s для поверхности является трудной задачей, которая обсуждалась рядом авторов. Здесь необходимо устранить различного рода неоднозначности, которые оставляют сомнения в правильности определения искомых параметров. Маргонинский и Вальцер [33] описали метод, который позволяет однозначно определить рекомбинационные характеристики и величину s . Этот метод объединяет рекомбинационные измерения с импульсными измерениями эффекта поля. Он применялся главным образом на Ge и Si.

Химическая природа центров захвата и рекомбинации и роль конкретных химических элементов в образовании ловушек, с участием которых происходит безызлучательная рекомбинация электронов и дырок, обсуждал Моррисон [34]. Величины скорости поверхностной рекомбинации, которые получили Браттейн и Бардин, для Ge находились в пределах от 10^6 до 10^3 см·с⁻¹, в то же время при условии, что все дырки, достигая поверхности, рекомбинируют, эта величина при $T=300$ К, должна составлять $s \approx \frac{1}{4} v_t \approx 2 \cdot 10^6$ см·с⁻¹. Было найдено, что шлифованные, а также обработанные пескоструйным способом поверхности обладают большими величинами скорости поверхностной рекомбинации, в то время как полировка в соответствующем травящем растворе обеспечивает низкие значения s .

Было исследовано также влияние поверхностной рекомбинации на эффект поля (см. подразд. 7.11.1). Ряд работ на эту тему включен в сборники [35], тематика которых связана с проблемами поверхностной рекомбинации. Позже эти вопросы рассматривали Мэни, Голдстейн и Гровер [36], а также Франкль [37].

Рассмотрим теперь, каким образом поверхностная рекомбинация влияет на процесс восстановления равновесия. Ясно, что на расстояниях от поверхности, превышающих диффузионную длину, влияние поверхности будет незначительным. Рассмотрим плоскую поверхность и пусть x — расстояние от поверхности до некоторой точки в полупроводнике. Предположим также, что избыточные носители заряда генерируются равномерно по всему объему образца с темпом \mathcal{R}_e , тогда распределение концентрации избыточных дырок определяется уравнением

$$D_h \frac{d^2 \Delta p}{dx^2} = \frac{\Delta p}{\tau_p} - \mathcal{R}_e. \quad (9.64)$$

Влияние поверхностной рекомбинации учтем через граничное условие при $x=0$, где справедливо соотношение

$$D_h \frac{d \Delta p}{dx} = s \Delta p. \quad (9.65)$$

В качестве второго граничного условия примем, что $\Delta p \rightarrow \mathcal{R}_e \tau_p$, если $x \rightarrow \infty$. Тогда решение, удовлетворяющее уравнению (9.64), будет иметь вид

$$\Delta p = \Delta p_1 e^{-x/L_p} + \mathcal{R}_e \tau_p. \quad (9.66)$$

Используя граничное условие при $x=0$, получаем

$$-\frac{D_h \Delta p_1}{L_p} = s (\Delta p_1 + \mathcal{R}_e \tau_p), \quad (9.67)$$

откуда

$$\Delta p_1 = -\frac{s \mathcal{R}_e \tau_p^2}{L_p + s \tau_p}, \quad (9.68)$$

так как $D_h/L_p = L_p/\tau_p$. Таким образом, имеем

$$\Delta p = \mathcal{R}_e \tau_p \left[1 - \frac{\tau_p \exp(-x/L_p)}{L_p + s \tau_p} \right], \quad (9.69)$$

а величина Δp при $x=0$, Δp_0 , определяется выражением

$$\Delta p_0 = \mathcal{R}_e \tau_p \left(1 - \frac{s \tau_p}{L_p + s \tau_p} \right). \quad (9.70)$$

Уменьшение величины Δp вблизи поверхности, обусловленное поверхностной рекомбинацией, сказывается, как и ожидалось, только на расстояниях не более нескольких L_p от поверхности. Степень уменьшения Δp зависит от отношения s/v_p , где v_p — скорость, равная L_p/τ_p . При $L_p = 0,05$ см и $\tau_p = 100$ мкс, что типично для германия, получим $v_p = 500$ см·с⁻¹. Если $s \gg v_p$, то Δp_0 очень мало, и уменьшенные концентрации избыточных дырок вблизи поверхности будут значительным.

9.9. Среднее время жизни в нитевидных образцах и тонких полосках

Общее решение задачи о движении электронов и дырок под действием электрического поля в однородном кристалле произвольной формы и с учетом влияния скорости поверхностной рекомбинации элементарными методами получить нельзя. Некоторые математические трудности, связанные с такого рода задачей, проанализировал Ван Русбрек [38], который получил довольно общие решения. Однако для ряда образцов простых геометрических форм можно дать простые решения, из которых нетрудно усмотреть общие черты эффекта поверхностной рекомбинации. Подобным примером является тонкий длинный стержень или нить с прямоугольным поперечным сечением. Эту задачу рассмотрел также Шокли [39], показавший что для ряда случаев, относящихся к движению электронов в ни-

тевидных образцах, можно получить простые аналитические решения.

Рассмотрим более простую задачу. Пусть имеется длинная тонкая полоска, длина и ширина которой значительно больше ее толщины. Направим ось z по длине полоски параллельно ее краям, ось y — поперек полоски, а ось x — перпендикулярно плоскости полоски. Полоска ограничена плоскостями $x = \pm a$, так что толщина ее равна $2a$. Для простоты предположим, что скорости поверхностной рекомбинации одинаковы для обеих сторон полоски.

Прежде всего определим среднюю концентрацию носителей заряда, создаваемых источником, который равномерно по всему объему образца генерирует электронно-дырочные пары с темпом \mathcal{R}_e . Этого можно добиться, например, облучая полоску светом, коэффициент поглощения α которого значительно меньше a^{-1} . Теперь нужно решить уравнение (9.64) с граничными условиями (9.65) при $x = a$ и $x = -a$. Соответствующее решение (из соображений симметрии) имеет вид

$$\Delta p = A \operatorname{ch} \left(\frac{x}{L_p} \right) + \mathcal{R}_e \tau_p. \quad (9.71)$$

Используя граничные условия, получаем

$$\Delta p = \mathcal{R}_e \tau_p \left[1 - \frac{s \tau_p \operatorname{ch} (x/L_p)}{s \tau_p \operatorname{ch} (a/L_p) + L_p \operatorname{sh} (a/L_p)} \right]. \quad (9.72)$$

Средняя величина Δp , т. е. $\overline{\Delta p}$, дается выражением

$$\overline{\Delta p} = \mathcal{R}_e \tau_p \left[1 - \frac{(L_p/a) s \tau_p \operatorname{sh} (a/L_p)}{s \tau_p \operatorname{ch} (a/L_p) + L_p \operatorname{sh} (a/L_p)} \right]. \quad (9.73)$$

При помощи уравнения $\overline{\Delta p} = \mathcal{R}_e \overline{\tau_p}$ можем определить среднюю величину $\overline{\tau_p}$ для полоски. Если $a \ll L_p$, то величина Δp оказывается примерно постоянной и равной

$$\Delta p = \mathcal{R}_e \tau_p \frac{a}{s \tau_p + a}, \quad (9.74)$$

так что

$$\frac{1}{\tau_p} = \frac{1}{\tau_p} + \frac{s}{a}. \quad (9.75)$$

Решение соответствующей задачи для стержня прямоугольного сечения оказывается более сложным, так как приходится использовать неоднородные граничные условия. Если размеры прямоугольного сечения составляют $2a$ и $2b$ и $a \ll L_p$, $b \ll L_p$, то можно показать, что приблизительно выполняется соотношение

$$\frac{1}{\tau_p} = \frac{1}{\tau_p} + \frac{s}{a} + \frac{s}{b}. \quad (9.76)$$

Рассмотрим теперь, как первоначально равномерная по объему концентрация избыточных носителей заряда уменьшается со временем после выключения источника генерации. Для этого решим уравнение

$$D_n \frac{\partial^2 \Delta p}{\partial x^2} = \frac{\Delta p}{\tau_p} + \frac{\partial \Delta p}{\partial t}. \quad (9.77)$$

Решение, очевидно, должно содержать как t , так и x . Решения этого уравнения, рассмотренные в разд. 7.6, здесь использовать нельзя, поскольку они не описывают первоначальное равномерное распределение избыточных носителей заряда. Будем искать решение в виде $\Delta p = \exp(\alpha x - \beta t)$ и убедимся, что такое решение действительно существует. Соотношение между α и β находится подстановкой Δp в уравнение (9.77), что дает

$$D_n \alpha^2 = \frac{1}{\tau_p} - \beta. \quad (9.78)$$

Из физических соображений следует, что спад будет более быстрый, чем $1/\tau_p$, так что величина $1/\tau_p - \beta$ будет отрицательной. Таким образом, мы приходим к решению в тригонометрических функциях, причем из условий симметрии его можно записать в следующем виде:

$$\Delta p = A \cos \alpha x \exp(-\beta t). \quad (9.79)$$

Теперь соотношение между α и β будет иметь вид

$$D_n \alpha^2 = \beta - \frac{1}{\tau_p}. \quad (9.80)$$

Используя граничное условие (9.65), получаем уравнение

$$D_n \alpha a \sin \alpha a = a s \cos \alpha a. \quad (9.81)$$

Уравнение

$$\eta \operatorname{tg} \eta = \frac{as}{D_n} \quad (9.82)$$

является трансцендентным с бесконечным числом корней $\eta_1, \eta_2, \eta_3, \dots$ и т. д. Для корня η_r получим $\alpha_r = \eta_r/a$ и тогда β дается выражением

$$\beta_r = \frac{1}{\tau_p} + D_n \left(\frac{\eta_r}{a} \right)^2. \quad (9.83)$$

Начальное значение Δp при $x=0$, которое обозначим Δp_0 , можно представить в виде ряда Фурье

$$\Delta p_0 = \sum A_r \cos \eta_r x \quad (-a \leq x \leq a), \quad (9.84)$$

причем нет необходимости определять значения коэффициентов A_r , поскольку справедливы неравенства $\eta_1 < \eta_2 < \eta_3 < \dots$, а из уравнения (9.79) следует, что решения, соответствующие корням высших

порядков, затухают во времени гораздо быстрее, чем решение, соответствующее первому корню, поэтому после первоначального «переходного» процесса ими можно пренебречь. Общее решение изучали Висванатан и Бэтти [40]. Если обозначить $\beta_1 = 1/\tau_1$, где τ_1 — эффективное время жизни в полоске, то получим

$$\frac{1}{\tau_1} = \frac{1}{\tau_p} + \frac{\eta_1^2 D_n}{a^2}. \quad (9.85)$$

Для больших значений s наименьший корень уравнения (9.82) стремится к величине $\eta_1 = \pi/2$. Тогда имеем

$$\frac{1}{\tau_1} \rightarrow \frac{1}{\tau_p} + \frac{\pi^2 D_n}{4a^2}. \quad (9.86)$$

Для малых величин s имеем $\eta_1^2 = as/D_n$, так что.

$$\frac{1}{\tau_1} = \frac{1}{\tau_p} + \frac{s}{a}, \quad (9.87)$$

т. е. то же значение, что было получено для $\overline{\tau_p^{-1}}$ при малых величинах s . Однако в общем случае времена жизни τ_1 и $\overline{\tau_p}$ не одинаковы. Величину τ_1 обычно называют временем жизни в полоске.

Задачу для нитевидного образца с прямоугольной формой поперечного сечения можно решить точно таким же способом [39]. В этом случае получим

$$\frac{1}{\tau_1} = \frac{1}{\tau_p} + D_n \left(\frac{\eta_1^2}{a^2} + \frac{\zeta_1^2}{b^2} \right), \quad (9.88)$$

где ζ_1 — наименьший корень уравнения

$$\zeta \operatorname{tg} \zeta = \frac{bs}{D_n}. \quad (9.89)$$

При малых величинах s уравнение (9.88) принимает ту же форму, что и уравнение (9.76), а $\tau_p = \overline{\tau_p}$. Уравнения оказываются точно такими же, когда неосновными носителями заряда являются электроны.

Рассмотрим далее движение неосновных носителей заряда вдоль полоски под действием электрического поля \mathcal{E}_z . Для стационарного случая найдем решение уравнения

$$D_p \left(\frac{\partial^2 \Delta p}{\partial x^2} + \frac{\partial^2 \Delta p}{\partial z^2} \right) - \mu_n \mathcal{E}_z \frac{\partial \Delta p}{\partial z} - \frac{\Delta p}{\tau_p} = 0. \quad (9.90)$$

Из предыдущего анализа следует, что соответствующее решение, стремящееся к нулю при $z \rightarrow \infty$, имеет вид

$$\Delta p = e^{-\nu z} \cos \alpha x. \quad (9.91)$$

Для решения, которое затухает с расстоянием наименее быстро, α равно η_1/a , так что после подстановки в уравнение (9.90) получим

$$D_p (\gamma^2 - \alpha^2) + \mu_h \gamma \mathcal{E}_z - \frac{1}{\tau_p} = 0.$$

Если обозначить $\gamma = 1/L_d$, или $1/L_{u1}$, соответственно для положительного или отрицательного \mathcal{E}_z , то после замены τ_p на τ_1 получим такие же уравнения, как (7.75) и (7.76). В частности, дрейфовая длина дырок в сильном поле, L_1 , запишется в виде

$$L_1 = \mathcal{E}_z \mu_h \tau_1. \quad (9.92)$$

Подобное же уравнение можно написать и для электронов. Это уравнение применимо также в случае нитевидного образца с поперечным сечением прямоугольной формы. Величину L_1 можно легко определить, измерив с помощью коллекторного контакта (см. разд. 7.9.2) распределение концентрации неосновных носителей заряда вдоль полоски. Сопоставляя L_1 с дрейфовой длиной L_d неосновных носителей заряда в толстом образце, характеризующей объемные свойства полупроводника, можно найти скорость поверхностной рекомбинации s . Мэни [41], измеряя τ_1 в нитевидных образцах прямоугольного сечения с высокой скоростью поверхностной рекомбинации, использовал уравнение (9.88) при $\eta_1 = \zeta_1 = \pi/2$ и смог осуществить *прямое* измерение коэффициентов диффузии D_h и D_e электронов и дырок.

ЛИТЕРАТУРА

1. *Chattarji D.*, The Theory of Auger Transitions, Academic Press, 1976.
2. *Roosbroeck W. Van, Shockley W.*, Phys. Rev., **94**, 1558 (1954).
3. *Blakemore J. S.*, Proc. Vth Intern. Conf. Phys. Semicond., Czechoslovak Sci. Publ., 1961, p. 981.
4. *Brill Patricia H., Schwarz Ruth F.*, Proc. IVth Intern. Conf. Phys. Semicond., Journ. Phys. Chem. Solids, **8**, 75 (1959).
5. *Varshni V. P.*, Phys. Stat. Sol., **19**, 459 (1967); **20**, 9 (1967).
6. *Landsberg P. T., Beattie A. R.*, Journ. Phys. Chem. Solids, **8**, 73 (1959); Proc. Roy. Soc., **A249**, 16 (1959); **A258**, 486 (1960).
7. *Blakemore J. S.*, Semiconductor Statistics, Pergamon Press, 1962. [Имеется перевод: Блекмор Дж. Статистика электронов в полупроводниках.— М.: Мир, 1964.]
8. *Benz G., Conradt R.*, Proc. XIIth Conf. Phys. Semicond., Teubner, 1974, p. 1262.
9. *Haug A.*, Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 1106.
10. *Landsberg P. T.*, Phys. Stat. Sol., **41**, 457 (1970).
11. *Shockley W., Read W. T.*, Phys. Rev., **87**, 835 (1952).
12. *Landsberg P. T.*, Proc. Phys. Soc., **B70**, 282 (1957).
13. *Burton J. A., Hull G. W., Morin F. J., Severiens J. C.*, Journ. Phys. Chem., **57**, 853 (1953).
14. *Many A., Bray R.*, Progress in Semiconductors, Heywood, **3**, 1958, p. 117.
15. *Lax M.*, IVth Intern. Conf. Phys. Semicond., Journ. Phys. Chem. Solids, **8**, 66 (1959).
16. *Brown R. A., Rodriguez S.*, Phys. Rev., **153**, 890 (1967).

17. Карпова И. В., Калашников С. Г., Proc. VIth Intern. Conf. Phys. Semicond., Inst. Phys. and Phys. Soc., London, 1962, p. 880.
18. Бонч-Бруевич В. Л., Гуляев Ю. В., ФТТ, 2, 465 (1960).
19. Landsberg P. T., Evans D. A., Rhys-Roberts C., Proc. Phys. Soc., 83, 325 (1964).
20. Шейнкман М. К., ФТТ, 5, 2780 (1963); 7, 28 (1965).
21. Englman R., Jortner J., Molec. Phys., 18, 145 (1970).
22. Mott N. F., Davis A. E., Street R., Phil. Mag., 32, 961 (1975).
23. Jaros M., Brand S., Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 1090.
24. Калашников С. Г. Proc. IVth Intern. Conf. Phys. Semicond., Journ. Phys. Chem. Solids, 8, 52 (1959); Proc. Vth Intern. Conf. Phys. Semicond., Czechoslovak Acad. Sci., 1961, p. 241.
25. Nelson D. F., Electronic Structure in Solids, ed. Hademenakis E. D., Plenum Press, 1969, p. 122.
26. Morrison S. R., Phys. Rev., 104, 619 (1956).
27. McKelvey J. P., Phys. Rev. 106, 910 (1957).
28. Wertheim G. K., Pearson G. L., Phys. Rev., 107, 694 (1957).
29. Blakemore J. S., Schultz J. Y., Nomura K. C., Journ. Appl. Phys., 31, 2226 (1960).
30. Gornik E., Phys. Rev. Lett., 29, 595 (1972).
31. Brattain W. H., Bardeen J., Bell Syst. Techn. Journ., 32, 1 (1953).
32. Schultz B. H., Physica, 20, 1031 (1954).
33. Margoninski Y., Walzer Y., Phys. Rev., 156, 903 (1967).
34. Morrison S. R., Treatise on Solid State Chemistry, ed. Hannay N. B., Plenum Press, 6B, 1976, p. 203.
35. Semiconductor Surface Physics, ed. Kingston R. H., Penn. Univ. Press, 1957; Surface Properties of Semiconductors, ed. Frumkin A. N., Consultants Bureau, N. Y., 1964.
36. Many A., Goldstein Y., Grover N. B., Semiconductor Surfaces, North Holland, 1965.
37. Frankl D. R., Electronic Properties of Semiconductors Surfaces, Pergamon Press, 1967.
38. Roosbroeck W. Van, Journ. Appl. Phys., 26, 380 (1955).
39. Shockley W., Electrons and Holes in Semiconductors, Van Nostrand, 1950, p. 322. [Имеется перевод: Шокли В. Теория электронных полупроводников.— М.: ИЛ, 1953.]
40. Viswanathan S., Batley J. F., Journ. Appl. Phys., 25, 99 (1954).
41. Many A., Physica, 20, 989 (1954).

ОПТИЧЕСКИЕ И ВЫСОКОЧАСТОТНЫЕ ЯВЛЕНИЯ

10.1. Оптические характеристики полупроводников

Многие полупроводники обладают характерным блеском полированных металлов и почти не отличаются от них по своему внешнему виду. Это сходство обусловлено их высокой отражательной способностью. В видимой области спектра полупроводники, как и металлы, обычно сильно поглощают, имея величины коэффициента поглощения порядка 10^6 см^{-1} . Оптические константы ряда полупроводников были измерены в широкой области длин волн. Характерной особенностью всех достаточно чистых полупроводников является то, что на определенной длине волны, обычно в ближней или средней ИК-области, коэффициент поглощения начинает быстро уменьшаться, и на более длинных волнах материал становится вполне прозрачным. Этот участок существенного понижения поглощения называется краем собственного поглощения. В литературе иногда его называют также краем фундаментального поглощения. Прозрачность полупроводников за длинноволновым краем поглощения обычно проявляется лишь тогда, когда полупроводники очищены настолько, что поглощение излучения свободными носителями заряда становится невелико по сравнению с собственным поглощением. При наличии большого количества примесей полупроводники обычно непрозрачны в широкой спектральной области — от ультрафиолета до радиоволн.

Для описания оптических свойств материалов обычно используется комплексный показатель преломления n^* , который может быть записан через действительную и мнимую части в виде

$$n^* = n(1 - ik). \quad (10.1)$$

Физический смысл n^* состоит в следующем. Уравнение, описывающее плоскую электромагнитную волну с частотой ν и длиной волны λ , распространяющуюся в материале в направлении x , имеет вид

$$\psi = A \exp \left[2\pi i \nu \left(t - \frac{n^* x}{c} \right) \right], \quad (10.2)$$

где c — скорость света в вакууме. Разделив n^* в (10.2) на действительную и мнимую части, имеем

$$\psi = A \exp\left(-\frac{2\pi\nu n_k x}{c}\right) \exp\left[2\pi i\nu\left(t - \frac{nx}{c}\right)\right]. \quad (10.3)$$

Фазовая скорость V равна c/n , а длина волны λ , если поглощение не слишком велико, равна $c/\nu n$. Выразив $c/\nu n$ через λ , получаем

$$\psi = A \exp\left(-\frac{\alpha x}{2}\right) \exp\left[2\pi i\nu\left(t - \frac{x}{V}\right)\right], \quad (10.4)$$

$$\alpha = \frac{4\pi k}{\lambda} = \frac{4\pi n_k}{\lambda_0}, \quad (10.5)$$

где λ_0 — длина волны в вакууме при частоте ν ; k обычно называют показателем поглощения, а α — коэффициентом поглощения¹⁾. Используя классическую теорию электромагнетизма, n^* можно связать с комплексной диэлектрической проницаемостью ϵ^* , которая определяется равенством

$$\epsilon^* = \epsilon - \frac{i\sigma}{2\pi\nu\epsilon_0}, \quad (10.6)$$

где ϵ — обычная диэлектрическая проницаемость, а σ — электропроводность. Комплексный показатель преломления n^* и ϵ^* связаны соотношением

$$(n^*)^2 = \epsilon^*. \quad (10.7)$$

Равенства (10.6) и (10.7) непосредственно следуют из уравнения распространения электромагнитных волн в проводящей среде. Сравнивая действительные и мнимые части равенства (10.7), получаем

$$\epsilon = n^2(1 - k^2), \quad (10.8)$$

$$\sigma = 4\pi n^2 k \nu \epsilon_0, \quad (10.9)$$

а из выражений (10.5) и (10.9) при $\epsilon > 0$ имеем

$$\alpha = \frac{\sigma}{n^2 \lambda \nu \epsilon_0} = \frac{\sigma}{n \epsilon_0 c}. \quad (10.10)$$

Для характеристики оптических свойств материала чаще всего используют величины $n^2(1 - k^2)$ и $2n^2k$, которые удобно представлять графически. Они могут быть получены из измерений коэффициента отражения. При нормальном падении излучения коэффициент отражения R определяется выражением

$$R = \left| \frac{1 - n^*}{1 + n^*} \right|^2 = \frac{(1 - n)^2 + n^2 k^2}{(1 + n)^2 + n^2 k^2}. \quad (10.11)$$

¹⁾ Некоторые авторы называют показателем поглощения величину n_k , а не k .

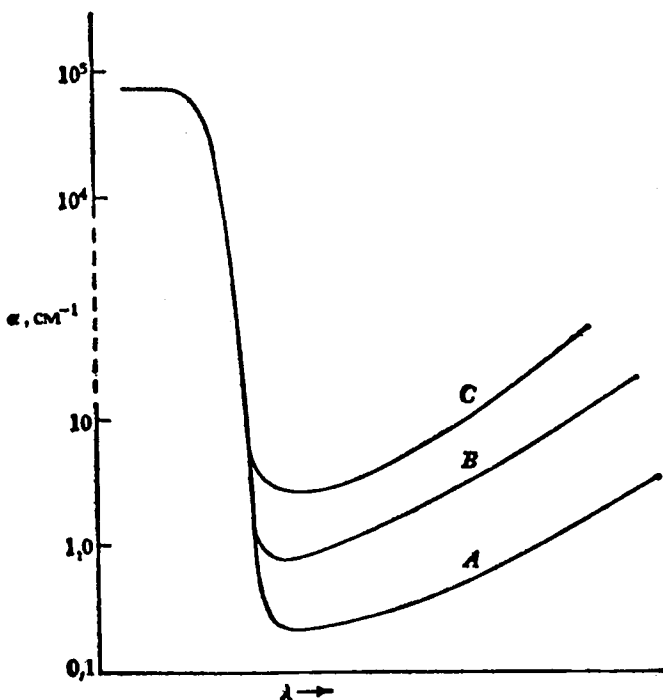


Рис. 10.1. Зависимость коэффициента поглощения α от длины волны вблизи края собственного поглощения для образцов с различной степенью чистоты. Кривая А соответствует наиболее чистому образцу.

Из этой формулы видно, что, если n и k велики, R мало отличается от единицы. Для углов падения, отличных от нормального, коэффициент отражения зависит от поляризации. Проведя измерения при различных углах, можно определить как n , так и k , если только значения k не слишком малы.

Если падающее излучение не может вызвать переход электронов из валентной зоны в зону проводимости, то можно предположить, что оптические эффекты будут обусловлены переходами внутри зоны. В этом случае, как мы увидим, взаимодействие излучения со свободными носителями заряда можно приближенно рассматривать с помощью классической теории электромагнетизма, учитывая, что электроны и дырки обладают эффективными массами. Если, однако, излучение может возбуждать электроны из валентной зоны в зону проводимости, должны иметь место заметные квантовые эффекты. В этом случае можно ожидать сильного увеличения поглощения, и оно действительно наблюдается. Типичная наблюдаемая зависимость коэффициента поглощения от длины волны излу-

чения показана на рис. 10.1. Можно видеть, что величина коэффициента поглощения проходит через минимум, в котором его значение зависит от содержания примесей. Оно будет минимальным для наиболее чистых образцов. Длина волны λ_e и частота ν_e , соответствующие краю собственного поглощения, приближенно определяются равенством

$$\frac{hc}{\lambda_e} = h\nu_e = \Delta E, \quad (10.12)$$

где ΔE — ширина запрещенной зоны, поскольку ΔE — это энергия, необходимая для возбуждения электрона из состояния у потолка валентной зоны в состояние у дна зоны проводимости. Как мы увидим, эти переходы не всегда возможны без участия других процессов изменения энергии, и в этом случае уравнение (10.12) следует видоизменить. Если не говорить, однако, об очень точных измерениях, это уравнение можно считать вполне хорошим приближением. На рис. 10.1 видно, что положение края поглощения точно не определяется. Обычно его можно определить без особых трудностей с точностью около 0,1 эВ, но для получения большей точности необходимо провести более детальный анализ типов электронных переходов, участвующих в процессе поглощения. Из вида кривых, показанных на рис. 10.1, ясно также, что, для того чтобы получить точное значение положения края собственного поглощения, следует использовать образцы самой высокой степени чистоты. Кроме того, они должны представлять собой по возможности совершенные монокристаллы, поскольку внутренние напряжения приводят к размытию края поглощения.

При изменении длины волны света в коротковолновую сторону от края поглощения коэффициент поглощения быстро возрастает и величина его может достигь нескольких единиц $\times 10^4 \text{ см}^{-1}$. При дальнейшем уменьшении длины волны коэффициент поглощения может оставаться почти постоянным, но иногда наблюдается дальнейшее возрастание до значений, превышающих 10^5 см^{-1} . Сам край поглощения может иметь некоторую тонкую структуру, однако прежде чем обсуждать экспериментальные результаты (см. разд. 10.5.2), мы должны рассмотреть теорию собственного поглощения. Сначала, однако, рассмотрим теорию поглощения излучения свободными носителями заряда.

10.2. Поглощение свободными носителями заряда

На частотах, недостаточно высоких, чтобы вызвать межзонные переходы или образование экситонов, поглощение обусловлено возбуждением колебаний решетки или переходами между состоя-

ниями в одной зоне. Рассмотрим сначала второй механизм поглощения. Понятно, что в полностью заполненной зоне такие переходы невозможны. Для зоны проводимости при отсутствии вырождения интенсивность поглощения пропорциональна числу свободных электронов, а в случае валентной зоны, как мы увидим ниже, существует, кроме этого, некоторое дополнительное поглощение, пропорциональное числу свободных дырок. Поглощение свободными носителями можно рассматривать, используя квантовую теорию вероятностей переходов между состояниями с различными энергиями, однако при этом необходимо учитывать не только электронные переходы, но и переходы между различными энергетическими состояниями решетки. Такое рассмотрение выполнил ряд авторов, в том числе Фэн [1, 3] и Бардин [2]. За исключением случая, когда имеет место вырождение зон, как, например, в Ge *p*-типа, полуклассическое рассмотрение дает по существу те же результаты, что и квантовая теория¹⁾. Так как первый метод более прост, то здесь мы будем следовать полуклассическому методу, в котором коэффициент поглощения выводится из выражения для электропроводности полупроводника на высоких частотах.

Взяв теперь, какие изменения необходимо внести в теорию электропроводности, рассмотренную в разд. 5.2.1 для случая переменного электрического поля. Уравнение движения электрона в переменном поле с частотой $\omega/2\pi$, направленном по оси *x*, имеет вид

$$m_e \dot{v}_x = -e\mathcal{E}_x = -e\mathcal{E}e^{i\omega t}. \quad (10.13)$$

Соответствующее решение этого уравнения есть

$$v_x = v_{x0} + \frac{e\mathcal{E}}{i\omega m_e} (e^{i\omega t_0} - e^{i\omega t}), \quad (10.14)$$

где v_{x0} — скорость в момент $t=t_0$ ($t_0 < t$). Умножая правую часть уравнения (10.14) на $\tau^{-1}e^{-\frac{(t-t_0)}{\tau}}$ и интегрируя по t_0 в пределах от $t=-\infty$ до t , а также полагая, что $\bar{v}_{x0} = 0$, получаем

$$\bar{v}_x = - \left[\frac{e\tau}{m_e(1+\omega^2\tau^2)} - \frac{i\omega\tau^2}{m_e(1+\omega^2\tau^2)} \right] \mathcal{E}_x, \quad (10.15)$$

где τ — время релаксации для электрона с энергией E . Плотность тока j_x получается усреднением по различным значениям E , в результате чего можем записать

$$j_x = \frac{n_e e^2}{m_e} \left(\left\langle \frac{\tau}{1+\omega^2\tau^2} \right\rangle - i\omega \left\langle \frac{\tau^2}{1+\omega^2\tau^2} \right\rangle \right) \mathcal{E}_x, \quad (10.16)$$

где n_e — концентрация электронов. Вместо комплексной диэлектрической проницаемости ϵ^* , определяемой уравнением (10.6), мы мо-

¹⁾ См., например, [4].

жем ввести комплексную проводимость σ^* с помощью уравнения $j_x = \sigma^* \mathcal{E}_x$. Вместо формулы (10.6) тогда имеем

$$\sigma^* = i\omega \epsilon_e^* \epsilon_0 = \sigma + i\omega \epsilon_e \epsilon_0, \quad (10.17)$$

где ϵ_e обозначает электронную часть диэлектрической проницаемости. Поляризация атомов решетки, а в полярных кристаллах их смещение относительно равновесного положения также дают вклад в диэлектрическую проницаемость, который мы обозначим ϵ_L . Тогда полная диэлектрическая проницаемость будет равна

$$\epsilon = \epsilon_L + \epsilon_e. \quad (10.18)$$

Из уравнения (10.16) получим

$$\sigma = \frac{n_e e^2}{m_e} \left\langle \frac{\tau}{1 + \omega^2 \tau^2} \right\rangle, \quad (10.19)$$

$$\sigma = \sigma_0 \frac{\left\langle \frac{\tau}{1 + \omega^2 \tau^2} \right\rangle}{\langle \tau \rangle}, \quad (10.20)$$

где σ_0 — проводимость на низких частотах. Кроме того,

$$\epsilon = \epsilon_L - \frac{n_e e^2}{\epsilon_0 m_e} \left\langle \frac{\tau^2}{1 + \omega^2 \tau^2} \right\rangle, \quad (10.21)$$

или

$$\epsilon = \epsilon_L - \frac{\sigma_0 \langle \tau \rangle}{\epsilon_0} \frac{\left\langle \frac{\tau^2}{1 + \omega^2 \tau^2} \right\rangle}{\langle \tau \rangle^2}. \quad (10.22)$$

При $\omega \langle \tau \rangle \ll 1$, пренебрегая малыми членами, зависящими от ω , имеем $\sigma = \sigma_0$ и

$$\epsilon = \epsilon_1 = \epsilon_L - \frac{\sigma_0 \langle \tau^2 \rangle}{\epsilon_0 \langle \tau \rangle}, \quad (10.23)$$

где σ_0 и ϵ_1 — низкочастотные значения проводимости и диэлектрической проницаемости соответственно. Величину $\langle \tau^2 \rangle / \langle \tau \rangle^2$ мы уже использовали при рассмотрении эффекта Холла (см. разд. 5.3.3) и обозначили ее через r , поэтому формулу (10.23) можем записать в виде

$$\epsilon_1 = \epsilon_L - \frac{r \sigma_0}{\epsilon_0} \langle \tau \rangle. \quad (10.24)$$

Отметим, что в первом приближении вклад электронов в диэлектрическую проницаемость отрицателен и величина его не зависит от частоты. В следующем приближении

$$\sigma = \sigma_0 (1 - a \omega^2 \langle \tau \rangle^2), \quad (10.25)$$

где

$$a = \frac{\langle \tau^3 \rangle}{\langle \tau \rangle^3}, \quad (10.26)$$

а также

$$\epsilon = \epsilon_1 + b \frac{\sigma_0}{\epsilon_0} \omega^2 \langle \tau \rangle^3, \quad (10.27)$$

где

$$b = \frac{\langle \tau^4 \rangle}{\langle \tau \rangle^4}. \quad (10.28)$$

Один из первых методов прямого измерения $\langle \tau \rangle$ и, следовательно, определения величины эффективной массы, предложили Бенедикт и Шокли [5]. Этот метод основан на сравнении диэлектрической проницаемости ϵ очень чистого материала с диэлектрической проницаемостью образца с известной концентрацией носителей. Первые эксперименты были выполнены с германием на частоте сантиметрового диапазона волн.

Другой предельный случай имеет место при $\omega \langle \tau \rangle \gg 1$. При длине волны $\lambda = 10$ мкм $\omega = 2 \cdot 10^{14}$ рад·с⁻¹ и при $T = 300$ К $\omega \tau_c \approx 260$, так что в ближней ИК-области это условие хорошо выполняется, за исключением, быть может, очень низких температур. В этом пределе, заменяя m_0 на дрейфовую эффективную массу m_c , получим

$$\sigma = \frac{n_c e^2}{m_c \omega^2} \left\langle \frac{1}{\tau} \right\rangle = \frac{g}{\omega^2 \langle \tau \rangle^2} \frac{n_c e^2 \langle \tau \rangle}{m_c}, \quad (10.29)$$

или

$$\sigma = \frac{g \sigma_0}{\omega^2 \langle \tau \rangle^2}, \quad (10.30)$$

где

$$g = \langle \tau \rangle \left\langle \frac{1}{\tau} \right\rangle. \quad (10.31)$$

Если τ можно записать в виде αE^{-s} , то имеем

$$g = \frac{\Gamma\left(\frac{5}{2} - s\right) \Gamma\left(\frac{5}{2} + s\right)}{\left[\Gamma\left(\frac{5}{2}\right)\right]^2} \quad (10.32)$$

[ср. с уравнением (5.162)]. Формулу (10.29) можно записать в виде

$$\frac{\sigma}{\sigma_0} = g \frac{e^2 / \omega^2 \mu_c^2}{m_c^2}, \quad (10.33)$$

где μ_c — подвижность, определенная из проводимости на постоянном токе. Используя формулу (10.10), можно получить выражение для коэффициента поглощения

$$\alpha = \sigma_0 g \frac{e^2 / \omega^2 \mu_c^2}{m_c^2 \epsilon_0 n}, \quad (10.34)$$

где n — показатель преломления. Это выражение можно записать также в виде

$$\alpha = \frac{n_e \lambda_0^2 g e^3}{4\pi^2 \epsilon_0 m_c^2 c^3 \mu_c}, \quad (10.35)$$

где n_e — концентрация электронов в зоне проводимости. Отсюда следует, что коэффициент поглощения α пропорционален концентрации свободных носителей заряда n_e , квадрату длины волны излучения λ_0 в вакууме и обратно пропорционален дрейфовой подвижности. Измерения поглощения на свободных носителях заряда, проведенные на длинах волн за краем собственного поглощения, позволили проверить справедливость формулы (10.35) для многих полупроводников. Минимум на кривых поглощения, показанных на рис. 10.1, обусловлен комбинацией собственного поглощения, которое уменьшается с увеличением длины волны, и поглощения на свободных носителях, которое растет пропорционально λ^2 . Явным исключением является поглощение свободными носителями в германии p -типа, которое мы рассмотрим позже. Величина поглощения в минимуме увеличивается при повышении концентрации свободных носителей. Одно время считалось, что для германия n -типа формула (10.35) дает слишком малую величину коэффициента поглощения, заниженную приблизительно в 100 раз¹⁾. Это было еще до того, как стала известна зонная структура германия или выполнены эксперименты по циклотронному резонансу, позволившие определить эффективные массы. Как выяснилось позже, множитель $g(m_0/m_c)^2$ для германия n -типа приблизительно равен 70, что приводит формулу (10.35) в согласие с экспериментальными данными. Искусственно повышая концентрацию носителей заряда с помощью оптической инжекции (генерируя электронно-дырочные пары светом) или инжекцией из электрода (см. разд. 7.5.3), удалось показать, что процессы поглощения электронами и дырками аддитивны. Коэффициент внутризонного поглощения на дырках также определяется формулой (10.35), однако с соответствующими для дырок значениями g , m_c и μ_c .

В то время как поглощение свободными носителями заряда в средней ИК-области, как мы видели, пропорционально λ^2 , на более коротких волнах в ближней области ИК-излучения это уже не так. Беккер и Фэн [6] показали, что в случае рассеяния на акустических колебаниях $\alpha \sim \lambda^{1.6}$, а при рассеянии на оптических колебаниях $\alpha \sim \lambda^{2.6}$, в то же время при рассеянии на примесях $\alpha \sim \lambda^{3+3.5}$. Поскольку в общем случае в какой-то мере будут присутствовать все эти виды рассеяния, зависимость $\alpha(\lambda)$ не будет выражаться простым степенным законом. В любом случае, однако, поглощение довольно быстро возрастает с увеличением длины волны.

¹⁾ См., например, [6, 7].

Совершая переход в зоне проводимости в состояние с другой энергией, электрон должен изменить свой волновой вектор \mathbf{k} и, следовательно, квазиимпульс \mathbf{P} . Как мы увидим позже, изменение квазиимпульса при таком переходе слишком велико по сравнению с импульсом поглощенного фотона, поэтому оно должно быть обеспечено кристаллической решеткой или примесью. Другими словами, электрон в процессе поглощения должен претерпеть рассеяние. Поэтому не является неожиданным, что коэффициент поглощения должен зависеть от вида рассеяния. На более длинных волнах эта зависимость ослабевает и коэффициент поглощения пропорционален λ^2 .

В случае $\omega \langle \tau \rangle \ll 1$ и $\epsilon_e \ll \epsilon_L$ коэффициент поглощения равен

$$\alpha = \frac{\sigma_0}{\epsilon_0 c \pi} = \frac{\sigma_0}{\epsilon_0 c \sqrt{\epsilon_L}}, \quad (10.36)$$

или

$$\alpha = 375 \frac{\sigma_0}{\sqrt{\epsilon_L}} \text{ м}^{-1}, \quad (10.36a)$$

где σ_0 выражается в $\text{Ом}^{-1} \cdot \text{м}^{-1}$ (или α в см^{-1} , если σ в $\text{Ом}^{-1} \cdot \text{см}^{-1}$). Это соотношение было подтверждено Гибсоном [8] для частот до $3,9 \cdot 10^{10}$ Гц ($\lambda_0 = 8,3$ мм).

Для материала, имеющего электропроводность порядка $0,1 \text{ Ом}^{-1} \cdot \text{см}^{-1}$ и собственную диэлектрическую проницаемость порядка 10, электронная составляющая диэлектрической проницаемости довольно мала. Например, при $\langle \tau \rangle = 10^{-12} \text{ с}$ $\epsilon/\epsilon_L \approx 1$. Однако для материала с более высокой проводимостью электронная составляющая может преобладать и ϵ станет отрицательным. Если проводимость настолько высока, что $|\epsilon_e| \gg \epsilon_L$, то приближенно имеем

$$\epsilon \approx -r \frac{\sigma_0}{\epsilon_0} \langle \tau \rangle. \quad (10.37)$$

Комплексная диэлектрическая проницаемость в этом случае дается выражением

$$\epsilon^* = -\frac{\sigma_0 r}{\epsilon_0} \langle \tau \rangle - \frac{i \sigma_0}{\epsilon_0 \omega}. \quad (10.38)$$

Воспользовавшись связью между n и κ , запишем

$$n^2 (1 - \kappa^2) = -\frac{\sigma_0 r \langle \tau \rangle}{\epsilon_0}, \quad (10.39)$$

$$2n^2 \kappa = \frac{\sigma_0}{\omega \epsilon_0}, \quad (10.40)$$

откуда получаем уравнение для κ

$$\frac{2\kappa}{\kappa^2 - 1} = \quad (10.41)$$

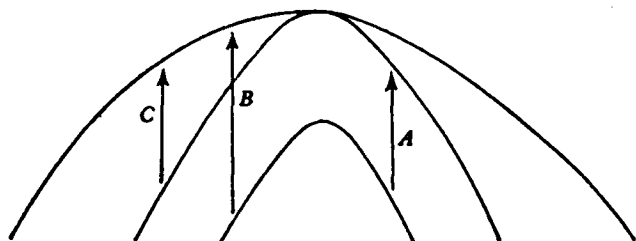


Рис. 10.2. Переходы между различными ветвями валентной зоны.

Предположим, что $\omega(\tau) \ll 1$, так что $\kappa \approx 1$. Тогда n имеет значение

$$n = \left(\frac{\sigma_0}{2\omega\epsilon_0} \right)^{1/2}, \quad (10.42)$$

а коэффициент поглощения равен

$$\alpha = \left(\frac{2\sigma_0\omega}{\epsilon_0 c^2} \right)^{1/2}. \quad (10.43)$$

В этом случае материал ведет себя подобно металлу, для которого имеет место хорошо известный скин-эффект. При этом излучение проникает в вещество на глубину скин-слоя, толщина которого обратно пропорциональна корню квадратному из частоты и из электропроводности.

В германии p -типа поглощение свободными носителями заряда не подчиняется закону $\alpha \sim \lambda^2$ и в области длин волн между 2 и 10 мкм имеет хорошо выраженную структуру, которую наблюдали Бриггс и Флетчер [9], а также Кайзер, Коллинз и Фэн [10]. Величина коэффициента поглощения в этом случае приблизительно на порядок выше, чем в германии n -типа при той же степени легирования. Рассматривая переходы между отдельными перекрывающимися зонами, образующими валентную зону германия, Кан [11] смог детально объяснить наблюдаемую структуру. Структура валентной зоны германия и типы переходов, ответственных в этом случае за поглощение, показаны на рис. 10.2. Считается, что валентная зона состоит из двух близких к «сферическим» зон, вырожденных при $k=0$, и третьей, отщепленной зоны. Результаты теоретического анализа согласуются с результатами измерений поглощения, если величину отщепления принять равной 0,28 эВ. В кремнии p -типа подобной тонкой структуры не наблюдается и поглощение довольно хорошо подчиняется закону $\alpha \sim \lambda^2$ [12]. Считается, что величина отщепления в этом случае значительно меньше ($\sim 0,05$ эВ).

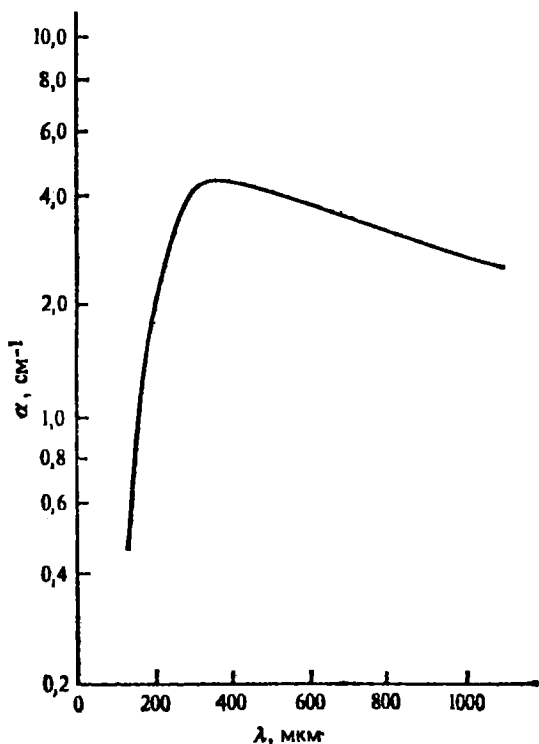


Рис. 10.3. Зависимость коэффициента поглощения α от длины волны для Si p -типа при $T=2,5$ К [13].

$N_d=1,4 \cdot 10^{17}$ см⁻³, $K=N_a/N_d=0,13$.

10.2.1. ПОГЛОЩЕНИЕ ИЗЛУЧЕНИЯ В ДАЛЕКОЙ ИК-ОБЛАСТИ ПРИ НИЗКИХ ТЕМПЕРАТУРАХ

В разд. 5.4 было показано, что в некоторых полупроводниках даже при очень низких температурах, когда все носители заряда выморожены на донорных или акцепторных центрах, остается заметная электропроводность. Считается, что эта электропроводность обусловлена прыжковым процессом, который может иметь место, если часть доноров или акцепторов не будет заполнена носителями из-за наличия компенсации. В таких условиях, согласно выражению (10.30), следует ожидать некоторого остаточного поглощения в далекой ИК-области, хотя из-за высокой частоты излучения нельзя ожидать, что поглощение будет простым образом связано с проводимостью на постоянном токе.

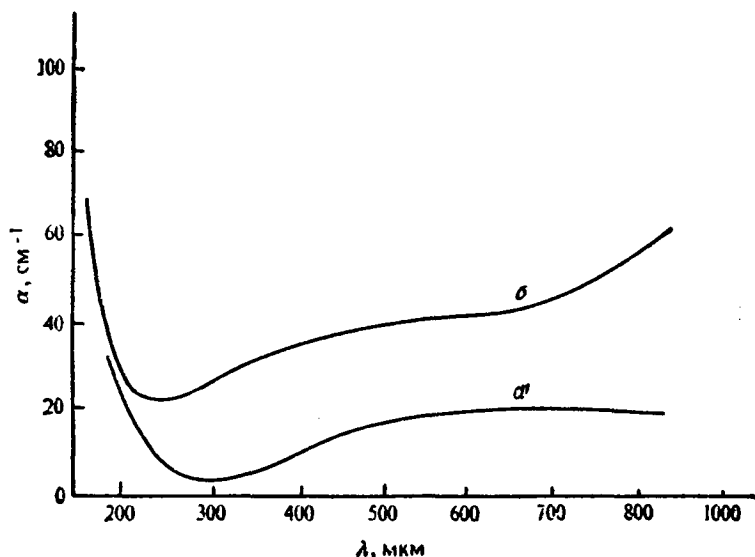


Рис. 10.4. Зависимость коэффициента поглощения α от длины волны для Ge p -типа при 4,3 К [15].

а) $N_a = 9 \cdot 10^{15} \text{ см}^{-3}$, $K = N_d/N_a = 0,11$; б) $N_a = 1,6 \cdot 10^{16} \text{ см}^{-3}$, $K = 0,50$.

Такое поглощение обнаружили в кремнии Нойрингер, Милвард и Аггарвал [13] в области длин волн (200—1000) мкм, рис. 10.3. Основанный на прыжковом механизме теоретический анализ Блиновского и Мисельского [14] дает очень хорошее согласие с экспериментальными результатами. В компенсированном германии p -типа подобное поглощение при низкой температуре в далекой ИК-области наблюдали Смит и др. [15], рис. 10.4. Было найдено, что поглощение зависит от степени компенсации $K = N_d/N_a$, но величина его значительно больше, чем дает теория Блиновского и Мисельского. Качмарек и Гортел [16] усовершенствовали эту теорию, учтя вырождение валентной зоны, и получили для p -Ge намного лучшее согласие с результатами измерений, хотя и не настолько точное, как для кремния. Рост поглощения на длинах волн, меньших 200 мкм, вызван ионизацией примесных центров с мелкими уровнями.

10.3. Плазменный резонанс

Если выполняется условие $\omega(\tau) \gg 1$, диэлектрическая проницаемость ϵ дается выражением

$$\epsilon = \epsilon_L - \frac{ne^2}{m_c \omega^2 \epsilon_0}, \quad (10.44)$$

а проводимость σ — формулой (10.33). Особый интерес представляет случай $\epsilon = 0$. Частота ω_p , на которой $\epsilon = 0$, определяется формулой

$$\omega_p = e \left(\frac{n}{m_e \epsilon_L \epsilon_0} \right)^{1/2} \quad (10.45)$$

и известна как плазменная частота. Это частота, с которой бесстолкновительная плазма электронов и положительных ионов может колебаться как одно целое. Если произведение $\omega \langle \tau \rangle$ не очень велико, то резонанс не будет резким, а коэффициент поглощения не будет неограниченно большим. На основании уравнений (10.21), (10.23) довольно легко показать, что резонанс не может быть резким, если $\omega \langle \tau \rangle$ не будет значительно больше единицы, скажем $\omega \langle \tau \rangle \sim 10$ на резонансной частоте.

10.4. Высокочастотные эффекты в магнитном поле

При наличии магнитного поля уравнения электропроводности для высоких частот нужно видоизменить. Для относительно низких частот, удовлетворяющих условию $\omega \langle \tau \rangle \ll 1$, формула для постоянной Холла в слабом поле не изменяется. Барлоу [17] провел измерения постоянной Холла в Ge в интервале частот вплоть до 10^{10} Гц и получил по существу те же значения, что и при измерениях на постоянном токе.

Вывод уравнений движения электронов в переменном электрическом и постоянном магнитном полях проще всего провести в два этапа. Уравнение движения электрона в отсутствие магнитного поля после усреднения по различным временам свободного пробега можно записать в виде

$$m_e \frac{d\bar{v}_x}{dt} + m_e \frac{\bar{v}_x}{\tau} = -e \mathcal{E}_x. \quad (10.46)$$

Это уравнение, очевидно, справедливо для случая постоянного поля, поскольку оно приводит к стационарному решению

$$\bar{v}_x = -\frac{e\tau \mathcal{E}_x}{m_e}. \quad (10.47)$$

Для случая переменного поля частоты $\omega/2\pi$ имеем

$$\bar{v}_x = -\frac{e\tau \mathcal{E}_x}{m_e (1 + i\omega\tau)}. \quad (10.48)$$

Это в точности эквивалентно выражению (10.15), и в таком случае мы можем провести усреднение по значениям τ , считая его функцией энергии E . Если магнитное поле направлено по оси z , уравнения

движения в плоскости (x, y) , соответствующие (10.46), имеют вид

$$\begin{aligned}\frac{d\bar{v}_x}{dt} + \frac{\bar{v}_x}{\tau} &= -\frac{e}{m_e} \mathcal{E}_x - \omega_c \bar{v}_y, \\ \frac{d\bar{v}_y}{dt} + \frac{\bar{v}_y}{\tau} &= -\frac{e}{m_e} \mathcal{E}_y + \omega_c \bar{v}_x,\end{aligned}\quad (10.49)$$

где $\omega_c = eB/m_e$. Теперь мы можем использовать эти уравнения для определения \bar{v}_x и \bar{v}_y .

10.4.1. ЦИКЛОТРОННЫЙ РЕЗОНАНС

Воспользуемся теперь системой уравнений (10.49) для исследования циклотронного резонанса, который ранее (разд. 2.5) мы обсуждали без учета явлений релаксации. Проще всего провести анализ для случая электрического поля с круговой поляризацией. Для правосторонней круговой поляризации имеем

$$\begin{aligned}\mathcal{E}_x &= \mathcal{E} \cos \omega t, \\ \mathcal{E}_y &= \mathcal{E} \sin \omega t.\end{aligned}\quad (10.50)$$

Умножив второе уравнение системы (10.49) на i и сложив его с первым, получим

$$\frac{d}{dt} (\bar{v}_x + i\bar{v}_y) + \left(\frac{1}{\tau} - i\omega_c\right) (\bar{v}_x + i\bar{v}_y) = -\frac{e\mathcal{E}}{m_e} e^{i\omega t}.\quad (10.51)$$

Решение уравнения (10.51), соответствующее вынужденному колебанию, дается выражением

$$\bar{v}_x + i\bar{v}_y = -\frac{\tau e (\mathcal{E}_x + i\mathcal{E}_y)/m_e}{1 + i(\omega - \omega_c)\tau}.\quad (10.52)$$

Отсюда следует, что средняя скорость не совпадает по фазе с полем, за исключением случая $\omega = \omega_c$. Сравнивая с равенствами (10.15), (10.17), (10.18), получаем

$$\sigma_r = \frac{ne^2}{m_e} \left\langle \frac{\tau}{1 + (\omega - \omega_c)^2 \tau^2} \right\rangle,\quad (10.53)$$

$$\varepsilon_r = \varepsilon_L - \frac{ne^2}{m_e \varepsilon_0} \left\langle \frac{\tau^2}{1 + (\omega - \omega_c)^2 \tau^2} \right\rangle \left(\frac{\omega - \omega_c}{\omega} \right).\quad (10.54)$$

Для левосторонней круговой поляризации аналогичным образом получаем

$$\sigma_l = \frac{ne^2}{m_e} \left\langle \frac{\tau}{1 + (\omega + \omega_c)^2 \tau^2} \right\rangle,\quad (10.55)$$

$$\varepsilon_l = \varepsilon_L - \frac{ne^2}{m_e \varepsilon_0} \left\langle \frac{\tau^2}{1 + (\omega + \omega_c)^2 \tau^2} \right\rangle \left(\frac{\omega + \omega_c}{\omega} \right).\quad (10.56)$$

Очевидно, для правосторонней круговой поляризации при $\omega_c \langle \tau \rangle = = V m_e \gg 1$ мы имеем острый резонанс на частоте $\omega = \omega_c = eB/m_e$. Это и есть циклотронный резонанс, который мы уже обсуждали в разд. 2.5. Величина σ на частоте резонанса равна величине электропроводности на постоянном токе, а при увеличении или уменьшении частоты σ спадает до намного меньшего значения, если $\omega_c \langle \tau \rangle \gg 1$. Таким образом, вблизи резонанса будет иметь место заметное увеличение коэффициента поглощения. Ширина резонанса составляет величину порядка $1/\langle \tau \rangle$ рад·с⁻¹, и, если не выполняется условие $\omega_c \langle \tau \rangle \gg 1$, резонанс будет выражен очень слабо. Поскольку для полей порядка 0,1 Т (10³ Гс) значение ω_c лежит в сантиметровом диапазоне ($\omega/2\pi = 2,8 \cdot 10^9$ Гц для $m_e = m_0$) τ должно быть несколько выше, чем 10⁻¹⁰ с; этого можно достичь только при низких температурах для чистых монокристаллов. Смысл условия $\omega_c \langle \tau \rangle \gg 1$, выполнение которого необходимо для получения острого резонанса, легко понять. Оно просто означает, что частица может несколько раз описать свою периодическую орбиту, прежде чем столкновение нарушит синфазность ее движения с возбуждающим электрическим полем. Пока частица движется в фазе с переменным полем, она получает от него энергию. Частота резонанса зависит только от эффективной массы m_e , заряда электрона и магнитной индукции B . Поэтому циклотронный резонанс представляет собой метод непосредственного измерения m_e . Пока мы имеем дело с простым случаем скалярной эффективной массы. Позже мы рассмотрим более общий случай тензорной эффективной массы.

Отметим, что в случае электронов резонанс имеет место только при правосторонней круговой поляризации и отсутствует при левосторонней. Различие в диэлектрической проницаемости для двух поляризаций приводит к двум значениям показателя преломления n^+ , n^- , которые немного отличаются друг от друга. А это приводит к эффекту Фарадея — вращению плоскости поляризации плоскополяризованного излучения при прохождении через кристалл с электронной проводимостью. Сделанный выше расчет относится к электронам. В случае дырок необходимо заменить ω_c на $-\omega_c$, что дает резонанс для левосторонней круговой поляризации. Это обстоятельство используется для определения знака носителей заряда, с которыми связан наблюдаемый резонанс. Еще одно условие наблюдения резонанса заключается в том, что переменное поле должно проникать в объем образца, иными словами, коэффициент поглощения α на рабочей частоте не должен быть слишком велик. Необходимые для этого предпосылки уже были выяснены нами раньше. Эффект Фарадея наблюдался в целом ряде полупроводников [18], но лучше всего он выражен в полупроводниках с малыми эффективными массами электронов, например, в таких, как InSb. Явление циклотронного резонанса подробно обсуждали Хогтон и С. Смит [19], а также Пиллер [20].

При наличии нескольких замкнутых поверхностей постоянной энергии, как, например, в германии *n*-типа, уравнения движения для электронов становятся крайне сложными. К счастью, каждая из поверхностей может рассматриваться отдельно, так как, за исключением некоторых избранных направлений магнитного поля, резонансные частоты для различных поверхностей неодинаковы. Чтобы найти резонансную частоту, можно предположить, что $\langle \tau \rangle$ бесконечно велико, и просто найти условие периодического движения электронов или дырок в постоянном магнитном поле. Частота этого периодического движения, являясь циклотронной частотой, будет в то же время и резонансной частотой. Направим оси координат x, y, z вдоль главных осей одного из эллипсоидов изоэнергетической поверхности, которым соответствуют главные значения эффективной массы m_1, m_2 и m_3 . Уравнения движения электронов в постоянном магнитном поле с компонентами вектора магнитной индукции по осям B_x, B_y, B_z имеют вид

$$\left. \begin{aligned} m_1 \frac{dv_x}{dt} &= -eB_z v_y + eB_y v_z, \\ m_2 \frac{dv_y}{dt} &= -eB_x v_z + eB_z v_x, \\ m_3 \frac{dv_z}{dt} &= -eB_y v_x + eB_x v_y. \end{aligned} \right\} \quad (10.57)$$

Для периодического движения по замкнутой орбите с частотой $\omega/2\pi$ можем считать, что каждая из составляющих скорости v_x, v_y, v_z пропорциональна $e^{i\omega t}$. Тогда из (10.57) получаем три однородных уравнения для v_x, v_y и v_z , которые имеют ненулевое решение, только если определитель системы равен нулю. Существуют два решения

$$\omega = 0 \quad (10.58)$$

и

$$\omega^2 = \frac{e^2}{m_1 m_2 m_3} (m_1 B_x^2 + m_2 B_y^2 + m_3 B_z^2). \quad (10.59)$$

Для первого решения имеем $v_x : v_y : v_z = B_x : B_y : B_z$, что означает прямолинейное равномерное движение по направлению магнитного поля. Если магнитное поле направлено вдоль одной из главных осей эллипсоида, получим

$$\left. \begin{aligned} \text{либо} \quad \omega = \omega_1 &= \frac{eB}{(m_2 m_3)^{1/2}}, \\ \omega = \omega_2 &= \frac{eB}{(m_1 m_3)^{1/2}}, \\ \text{либо} \quad \omega = \omega_3 &= \frac{eB}{(m_2 m_1)^{1/2}}. \end{aligned} \right\} \quad (10.60)$$

При произвольном направлении поля

$$\omega = (l^2\omega_1^2 + m^2\omega_2^2 + n^2\omega_3^2)^{1/2}, \quad (10.61)$$

где l, m, n — направляющие косинусы поля относительно главных осей эллипсоида.

Приведенный выше полуклассический анализ становится неприемлем, если $\omega_c \langle \tau \rangle \gg 1$. В этом случае следует использовать квантовую теорию. Мы отложим квантовый анализ до разд. 12.5.2, где рассмотрим метод эффективной массы для случая сильного магнитного поля. Можно сказать, однако, что условие резонанса, полученное с помощью квантовой теории, совпадает с условием (10.61).

Для того чтобы иметь возможность вычислить ширину резонанса, требуется более детальный анализ¹⁾. Такой анализ с учетом зависимости времени релаксации от энергии для «многодолинного» полупроводника выполнили Голд, Буллис и Кэмпбелл [23], а Цайгер, Лэкс и Декстер [24] рассмотрели случай гофрированных изоэнергетических поверхностей, которые, как было выяснено, имеют место в германии и кремнии p -типа.

К настоящему времени циклотронный резонанс наблюдался для большого числа полупроводников, причем ряд усовершенствований в методике очистки полупроводниковых материалов позволил для большинства из них обеспечить выполнение условия $\omega_c \langle \tau \rangle \gg 1$. Исследование циклотронного резонанса является не только наиболее прямым методом определения эффективной массы в экстремумах энергетических зон, но позволяет также в общих чертах определить форму изоэнергетических поверхностей. Без этого ценного метода исследований наши знания о полупроводниках были бы намного беднее. Обширный обзор работ на эту тему был дан Лэксом и Мавродесом [25].

10.4.2. КВАНТОВАНИЕ В МАГНИТНОМ ПОЛЕ

Выше уже отмечалось, что использованный нами способ рассмотрения движения электронов в сильном магнитном поле неприменим при условии $\omega_c \langle \tau \rangle \gg 1$, которое можно также выразить в виде $V_H \gg 1$. При выполнении этого условия, которое одновременно является условием наблюдения циклотронного резонанса, следует ожидать, что движение электронов должно быть периодическим и, следовательно, квантованным. К счастью, учет квантования движения электронов приводит к тем же значениям резонансных частот, что и полуклассический вывод, который был дан в разд. 10.4.1. Это можно показать следующим простым способом. Электрон, движущийся по круговой орбите радиуса r с периодом $2\pi/\omega_c$ будет обладать моментом количества движения $m_e r^2 \omega_c$, где $\omega_c = eB/m_e$, а согласно эле-

¹⁾ См., например, [21, 22].

ментарной квантовой теории момент количества движения должен быть равен $s\hbar$, где s — целое число, большее или равное единице. Таким образом, имеем

$$r^2 = \frac{s\hbar}{m_e\omega_c}. \quad (10.62)$$

Кинетическая энергия E электрона определяется равенством

$$E = \frac{1}{2} m_e r^2 \omega_c^2 = \frac{1}{2} s\hbar\omega_c, \quad (10.63)$$

так что наименьшая энергия электрона в зоне проводимости теперь отлична от нуля и равна $\frac{1}{2}\hbar\omega_c$. Дно зоны проводимости, следовательно, поднимается на величину, пропорциональную индукции магнитного поля. Чисто квантовомеханический анализ, выполненный Ландау [26], показывает, что для свободных электронов разрешенными являются только *нечетные* значения s . Разрешенные уровни энергии тогда даются выражением

$$E_s = \left(s + \frac{1}{2}\right) \hbar\omega_c, \quad (10.64)$$

где s теперь *любое* целое положительное число. Если магнитное поле направлено вдоль оси z , то квантуется лишь движение по осям x и y , а движение частицы по направлению оси z останется свободным. Выражение для энергии электрона в зоне проводимости тогда имеет вид (см. разд. 12.5)

$$E = \left(s + \frac{1}{2}\right) \hbar\omega_c + \left(\frac{\hbar^2}{2m_e}\right) k_z^2. \quad (10.64a)$$

Переходы между двумя уровнями, имеющими квантовые числа s и $s+1$, могут возбуждаться излучением с энергией кванта $\hbar\omega_c$ и, следовательно, с частотой $\omega_c/2\pi$. Это и есть циклотронная частота.

Позже мы более детально рассмотрим влияние сильного магнитного поля на энергетические уровни электронов в полупроводнике. Квантовомеханический анализ также приводит к выражению (10.61) в качестве условия резонанса, но, кроме того, обнаруживает некоторые новые особенности (см. разд. 12.5.2).

10.4.3. СДВИГ ЭНЕРГЕТИЧЕСКИХ ЗОН В МАГНИТНОМ ПОЛЕ

Чем меньше величина эффективной массы электрона m_e , тем больше частота ω_c и тем больше сдвиг дна зоны проводимости в магнитном поле. В InSb, где $m_e = 0,013 m_0$, при $B = 1$ Т величина сдвига составляет $3,85 \cdot 10^{-3}$ эВ. Подобный сдвиг для потолка валентной зоны будет значительно меньшим из-за того, что эффективная масса дырок m_h намного больше, чем m_e . Этот эффект наблюда-

ли в InSb Бурштейн и др. [27] в виде сдвига края собственного поглощения в магнитных полях до 6Т. Цвердлинг и др. [28] наблюдали этот эффект в InSb и InAs, используя импульсное магнитное поле индукцией до 25Т. Найденные величины сдвига в сильных полях несколько меньше значений, которые дает формула (10.64), что указывает на возрастание эффективной массы электронов при увеличении их энергии.

10.5. Собственное поглощение

В предыдущих разделах были рассмотрены оптически индуцированные переходы между различными энергетическими состояниями, принадлежащими *одной* зоне, либо зоне проводимости, либо валентной зоне. Теперь рассмотрим переходы, в которых электрон, поглощая фотон с энергией, приблизительно равной ширине запрещенной зоны ΔE , возбуждается из валентной зоны в зону проводимости. Такое поглощение было названо (см. разд. 10.1) собственным поглощением. Необходимо различать два типа переходов: переходы, в которых участвует только фотон (или фотоны), и переходы, при которых изменение энергии электрона при поглощении кванта излучения сопровождается также изменением энергии кристаллической решетки, т. е. переходы с участием фононов. Будем считать, что в переходе участвует только *один* фотон, отложив пока рассмотрение многофотонных переходов, которые имеют намного меньшую вероятность (см. разд. 10.5.5). Переходы, в которых не

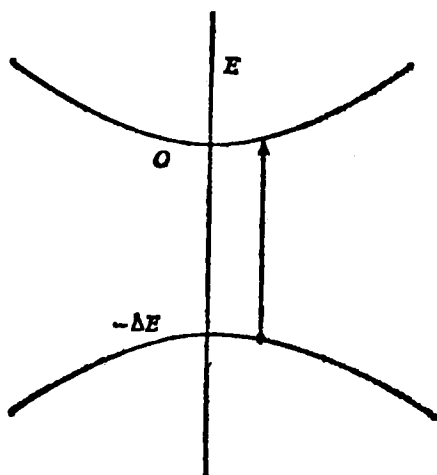


Рис. 10.5. Прямые переходы, $k_{\min} = k_{\max}$.

участвуют фононы, называют *прямыми*, а переходы, в которых участвуют фононы, — *непрямыми* переходами.

Необходимо также различать и два класса полупроводников, оптические свойства которых заметно отличаются друг от друга. К первому относятся полупроводники, у которых значения волновых векторов, соответствующие минимуму энергии в зоне проводимости, k_{\min} , и максимуму энергии в валентной зоне, k_{\max} , совпадают (при этом обычно $k_{\min} = k_{\max} = 0$). Ко второму классу относятся полупроводники, для которых $k_{\min} \neq k_{\max}$. Кремний и германий являются типичными полупроводниками второго класса. Примерами полупроводников первого класса являются интерметаллические соединения InSb и GaSb , халькогенид свинца PbS и т. д. Сначала мы рассмотрим полупроводники первого класса, поскольку энергетические переходы, отвечающие за собственное поглощение, в них несколько проще, чем в полупроводниках второго класса. Будем считать, что зоны не вырождены, как это показано на рис. 10.5, где принято $k_{\min} = k_{\max} = 0$.

10.5.1. ПРЯМЫЕ ПЕРЕХОДЫ, $k_{\min} = k_{\max}$

В процессе поглощения квазимпульс должен сохраняться (см. ниже). Импульс \mathbf{P}_p падающего фотона равен $(h/\lambda)\mathbf{i}_0$, где \mathbf{i}_0 — единичный вектор в направлении движения фотона до поглощения и λ — длина волны соответствующего излучения. Закон сохранения импульса имеет вид

$$\mathbf{P}_2 - \mathbf{P}_1 = (h/\lambda) \mathbf{i}_0, \quad (10.65)$$

где \mathbf{P}_1 и \mathbf{P}_2 — квазимпульсы электрона до и после поглощения фотона. Если обозначить через λ_e длину волны для электрона с квазимпульсом \mathbf{P}_1 ($|\mathbf{P}_1| = h/\lambda_e$), то при энергии электрона порядка kT , где $T = 300\text{K}$, значение λ_e окажется примерно равным $5 \cdot 10^{-7}$ см; если λ соответствует длине волны, лежащей в близкой ИК-области спектра, например $\lambda = 10^{-4}$ см, то $\lambda \gg \lambda_e$. В этом случае $|\mathbf{P}_1 - \mathbf{P}_2|$ — очень малая величина по сравнению с $|\mathbf{P}_1|$ и можно положить $\mathbf{P}_1 = \mathbf{P}_2$.

Сохранение квазимпульса является следствием законов квантовой механики. Вероятность перехода между двумя состояниями под действием возмущения F пропорциональна квадрату матричного элемента $|M_{if}|^2$, где

$$M_{if} = \int \psi_i F \psi_f^* dr, \quad (10.66)$$

а ψ_i и ψ_f — волновые функции начального и конечного состояний. В нашем случае F имеет вид

$$F = A \exp \frac{2\pi i (\mathbf{r} \cdot \mathbf{i}_0)}{\lambda},$$

а

$$\psi_i = U_{k_1}(\mathbf{r}) \exp [i(\mathbf{k}_1 \cdot \mathbf{r})], \quad (10.67)$$

$$\psi_f = U_{k_2}(\mathbf{r}) \exp [i(\mathbf{k}_2 \cdot \mathbf{r})] \quad (10.68)$$

(см. разд. 2.2). Таким образом, подынтегральная функция (10.66) содержит множитель

$$\exp \left[i \left(\mathbf{k}_1 - \mathbf{k}_2 + \frac{2\pi\mathbf{l}_0}{\lambda} \right) \cdot \mathbf{r} \right],$$

который очень быстро и периодически изменяется, за исключением случая, когда

$$\mathbf{k}_1 - \mathbf{k}_2 + \frac{2\pi\mathbf{l}_0}{\lambda} = 0. \quad (10.69)$$

При всех значениях \mathbf{k}_1 и \mathbf{k}_2 , не удовлетворяющих этому равенству, величина M_{if} ничтожно мала. Умножая обе части (10.69) на \hbar , получаем формулу (10.65), поскольку $\mathbf{k} = \mathbf{P}/\hbar$.

Таким образом, мы приходим к выводу, что единственными «разрешенными» переходами из зоны в зону в этом случае являются «вертикальные» переходы, которым соответствует «правило отбора» в виде $\Delta\mathbf{k} = 0$. Это показано на рис. 10.5. Интересно отметить, что электрон с квазиимпульсом \mathbf{P}_i из валентной зоны переходит в зону проводимости в состоянии с квазиимпульсом \mathbf{P}_f , оставляя в заполненной зоне дырку, движущуюся со скоростью $-\mathbf{P}_i/m_h$ (см. разд. 2.5). Таким образом, генерируемые светом свободные электрон и дырка должны двигаться в противоположных направлениях. В этом случае *минимальная* частота излучения, при которой еще возможен прямой переход, в точности определяется равенством (10.12). Таким образом, если в этой области частот нет никаких других типов переходов, которые могли бы конкурировать с только что рассмотренным, то частота, соответствующая краю собственного поглощения, определяется равенством (10.12). В связи с этим представляет значительный интерес выяснить вид спектра поглощения у края собственного поглощения со стороны коротких длин волн. Форма спектральной кривой поглощения зависит от того, имеются ли наряду с законом сохранения квазиимпульса еще и другие правила отбора или нет, т. е. от конкретного вида функций $U_{k_1}(\mathbf{r})$ и $U_{k_2}(\mathbf{r})$ в выражениях (10.67) и (10.68). Если эти функции таковы, что вероятность перехода P_h отлична от нуля, то P_h будет очень слабо зависеть от \mathbf{k} . Рассмотрим переход электрона из состояния в валентной зоне с волновым вектором \mathbf{k} , лежащим в интервале от \mathbf{k} до $\mathbf{k} + d\mathbf{k}$. Энергия фотона, индуцирующего переход, определяется соотношением

$$\hbar\nu = E_c(\mathbf{k}) - E_v(\mathbf{k}). \quad (10.70)$$

С другой стороны, вероятность поглощения кванта с частотой, лежащей в интервале $d\nu$, пропорциональна числу состояний в

валентной зоне с энергиями от $-\Delta E - E'$ до $-\Delta E - E' - dE'$. Допустим на время, что энергия у дна зоны проводимости вблизи $k=0$ очень слабо зависит от k , т. е. что $E_c(k)$ в этой области можно считать постоянным. Тогда из соотношения $E_v(k) = -\Delta E - E'$ следует, что

$$h\nu = -dE_v(k) = dE'. \quad (10.71)$$

Число переходов $N_i d\nu$ в интервале частот $d\nu$ определяется как

$$N_i d\nu = P_k N(E') dE',$$

где $N(E') dE'$ — число состояний с энергией, лежащей в интервале $E', E' + dE'$. Из предыдущего известно, что

$$N(E') = AE'^{1/2}, \quad (10.72)$$

где A — некоторая постоянная (см. разд. 2.3). Поскольку $E' = h\nu - \Delta E$ [из (10.70)], то коэффициент поглощения α_d можно записать в виде

$$\begin{aligned} \alpha_d &= A' P_k (h\nu - \Delta E)^{1/2}, & h\nu > \Delta E, \\ \alpha_d &= 0, & h\nu \leq \Delta E, \end{aligned} \quad (10.73)$$

где A' — некоторая постоянная. Если теперь учесть зависимость E_c от k , то получим

$$h d\nu = \frac{\partial E_c}{\partial k} dk - \frac{\partial E_v}{\partial k} dk.$$

Таким образом, учет зависимости E_c от k приводит лишь к умножению коэффициента поглощения на постоянную величину. Поскольку, как уже было сказано, вероятность P_k очень слабо зависит от k , коэффициент поглощения можно представить в виде

$$\begin{aligned} \alpha_d &= B (h\nu - \Delta E)^{1/2}, & h\nu > \Delta E, \\ \alpha_d &= 0, & h\nu \leq \Delta E, \end{aligned} \quad (10.74)$$

где B — некоторая постоянная. Это соотношение справедливо, естественно, лишь в ограниченной области значений $(h\nu - \Delta E)$. При более высоких значениях $(h\nu - \Delta E)$ коэффициент поглощения достигает области, где он почти не зависит от частоты ν падающего излучения. Детальный ход кривой $\alpha_d(\nu)$ на этом участке зависит от конкретной структуры зон.

В области частот, где величина $(h\nu - \Delta E)$ сравнима с энергией связи экситона (см. разд. 3.5), учет кулоновского взаимодействия между свободным электроном и дыркой, как показали Макфарлан и др. [29], приводит к изменению соотношений (10.74). В этом случае при $h\nu \rightarrow \Delta E$ поглощение не исчезает, а стремится к некоторому постоянному значению. При $h\nu < \Delta E$ кривая поглощения непрерывным образом сливается с поглощением, вызванным наиболее высо-

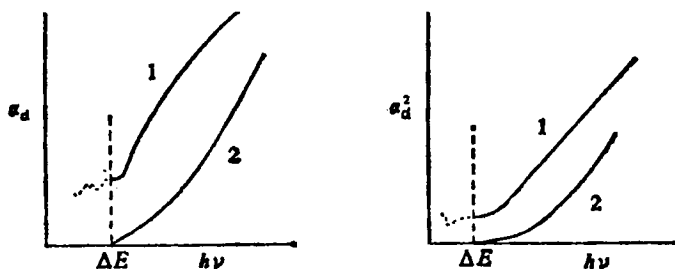


Рис. 10.6. Зависимость коэффициента поглощения от частоты для прямых переходов.

Кривые 1 относятся к «разрешенным» переходам, кривые 2 — к «запрещенным».

кими возбужденными состояниями экситона. Это показано на рис. 10.6 (см. также рис. 10.9).

Оптические переходы оказываются «разрешенными» в том случае, например, когда волновые функции электронов из валентной зоны построены на основе s -состояний индивидуальных атомов, а волновые функции электронов в зоне проводимости — на основе атомных p -состояний. Если бы последние обладали симметрией d -функций, то соответствующие переходы оказались бы «запрещенными». В этом случае легко показать, что, хотя вероятность перехода и обращается в нуль при $k=0$, но для всех $k \neq 0$ она пропорциональна k^2 , т. е. величине $(h\nu - \Delta E)$ (см. [30], § 13.3). Таким образом,

$$P_k = \text{const} \cdot (h\nu - \Delta E), \quad (10.75)$$

и коэффициент поглощения α_d' определяется выражением

$$\alpha_d' = C (h\nu - \Delta E)^{3/2}, \quad (10.76)$$

где C — постоянная. Значения постоянных B и C , которые можно вычислить с помощью известных методов квантовой механики, были найдены Бардином, Блаттом и Холлом [31]. Постоянную B можно записать в следующем виде (см. [30], § 13.3):

$$B = \frac{\pi e^2 (2m_r)^{3/2}}{nch^2 m_0 e_0} f_{if}, \quad (10.77)$$

где n — действительная часть показателя преломления; m_r — приведенная эффективная масса электрона и дырки, определяемая равенством

$$\frac{1}{m_r} = \frac{1}{m_e} + \frac{1}{m_h}; \quad (10.78)$$

f_{if} — множитель порядка единицы, который обычно называют «силой осциллятора». Формула (10.77) отличается от выражения, приведенного в работе Бардина, Блатта и Холла, тем, что она содержит

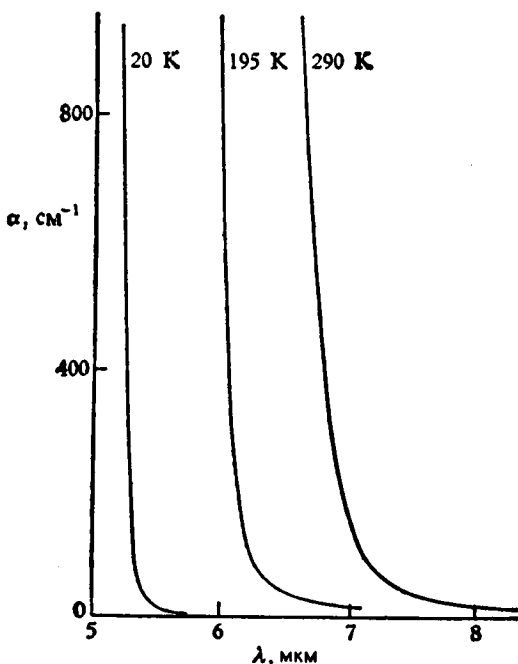


Рис. 10.7. Собственное поглощение в InSb [32].

приведенную массу m_r вместо m_h . Этим учтена зависимость E_c от k . При $m_e = m_h = m_0$, и полагая $p=4$, имеем

$$\alpha_d = 6,7 \cdot 10^4 (h\nu - \Delta E)^{1/2} f_{if} \text{ см}^{-1}, \quad (10.79)$$

где $h\nu$ и ΔE выражены в электрон-вольтах. В этих же обозначениях постоянная C равна (см. [30], § 13.3)

$$C = \frac{2}{3} B \left(\frac{2m_r}{m_0} \right) \frac{f'_{if}}{f_{if}} \frac{1}{h\nu}. \quad (10.80)$$

В этом случае имеем

$$\alpha'_d = 4,5 \cdot 10^4 f'_{if} \frac{(h\nu - \Delta E)^{1/2}}{h\nu} \text{ см}^{-1}. \quad (10.81)$$

При $h\nu = 1$ эВ и $(h\nu - \Delta E) = 0,01$ эВ, если $f_{if} = f'_{if} = 1$, имеем $\alpha_d = = 6,7 \cdot 10^3 \text{ см}^{-1}$ и $\alpha'_d = 45 \text{ см}^{-1}$.

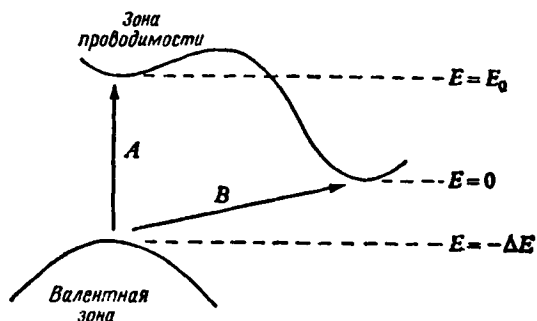
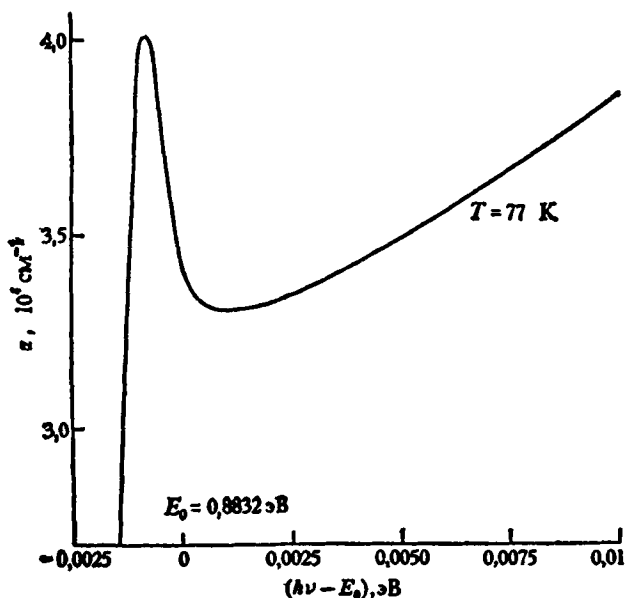
Для разрешенных прямых переходов график зависимости α_d^2 от $h\nu$ должен иметь вид прямой, из которой путем экстраполяции к $\alpha_d = 0$ можно получить значение ΔE . Для других переходов это правило неприменимо. Обычно графически изображают зависимость α_d от $h\nu$. Вид такого рода кривых для разрешенных и запрещенных переходов представлен на рис. 10.6. В большинстве случаев, когда наблюдалось поглощение, связанное с прямыми переходами, за-

висимость α_d от $(h\nu - \Delta E)$ была степенной, показатель степени которой оказывался ближе к $3/2$, чем к $1/2$.

В чистом соединении InSb обнаруживается весьма крутой рост поглощения. На основании ряда других соображений полагают (см. разд. 13.5), что InSb принадлежит к полупроводникам первого класса, о которых шла речь выше, хотя существуют некоторые осложнения, связанные с вырождением валентной зоны. Вместе с тем на кривых поглощения для InSb, полученных Робертсом и Кваррингтоном [32], обнаружены длинноволновые «хвосты» (рис. 10.7), происхождение которых можно толковать по-разному: например, они могли бы быть вызваны механическими напряжениями, приводящими к деформации кристаллической решетки. Эти эксперименты были проведены, однако, на монокристаллах высокого качества, так что такое объяснение кажется маловероятным. Некоторые авторы указывали также на то, что хвосты могли бы быть следствием переходов с участием фононов. Отложив пока обсуждение этой возможности, мы займемся рассмотрением полупроводников, для которых $k_{\min} \neq k_{\max}$.

10.5.2. ПРЯМЫЕ ПЕРЕХОДЫ, $k_{\min} \neq k_{\max}$

Прямые оптические переходы могут иметь место и в тех случаях, когда минимум энергии в зоне проводимости и максимум энергии в валентной зоне лежат при различных значениях волнового вектора k . Однако эти переходы уже не связаны с минимальным значением $h\nu$ для переходов из зоны в зону. Это иллюстрируется рис. 10.8, на котором принято, что максимум валентной зоны лежит при $k=0$, а минимум зоны проводимости расположен вблизи границы первой зоны Бриллюэна для определенного направления в кристалле. При этом k_{\min} действительно может быть равно k на границе зоны, как, например, для Ge (см. разд. 13.3), однако такое совпадение не является необходимым. Если энергия E_c , соответствующая дну зоны проводимости, гораздо меньше энергии E_v в зоне проводимости при $k=0$, то разность энергий, соответствующая прямому, или «вертикальному», переходу (обозначен буквой A на рис. 10.8), будет несколько больше разности энергий, соответствующей непрямо переходу (обозначен буквой B на рис. 10.8). Спектр поглощения в тонких монокристаллических образцах весьма чистого германия, полученный Дэшом и Ньюменом [33], указывает на крутой рост поглощения вблизи значения $h\nu$, соответствующего энергии 0,80 эВ, который можно связать с прямыми переходами между валентной зоной и зоной проводимости при $k=0$. Этот участок спектра поглощения исследовали в широком интервале температур Макфарлан и др. [29]. Эти авторы применили аппаратуру с более высоким спектральным разрешением и показали, что крутой рост поглощения в этой обла-

Рис. 10.8. Прямые и не прямые переходы, $k_{\min} \neq k_{\max}$ Рис. 10.9. Спектр поглощения, связанного с «вертикальными» переходами при $k=0$ в Ge [29].

сти связан не только с прямыми межзонами переходами, но и с экситонной линией поглощения, которую они сумели выделить из фона непрерывного спектра поглощения. Вид спектра поглощения при 77К показан на рис. 10.9. Непрерывное поглощение на этом рисунке при $h\nu \rightarrow \Delta E_0$, где ΔE_0 — ширина запрещенной зоны при $k=0$, стремится не к нулю, а к некоторому постоянному значению. Более того, вид кривой значительно отступает от закона $\alpha \sim (h\nu - \Delta E)^{1/2}$. Хорошее согласие с экспериментальной зависимостью коэффициента

поглощения от $h\nu$ удалось получить лишь при учете кулоновского взаимодействия между парой свободных частиц, образованных в процессе перехода. Поведение экситонной линии в сильном магнитном поле исследовали Цвердлинг, Лаура Рот и Лэкс [34].

Поглощение более длинноволнового излучения связано с непрямыми переходами типа *B* (рис. 10.8). Длинноволновая граница непрямых переходов представляет собой истинный край кривой собственного поглощения, однако энергетические соотношения осложняются здесь тем, что переходы типа *B* в отличие от переходов типа *A* происходят без сохранения квазиимпульса электрона. Закон сохранения импульса кристалла приводит к требованию обмена импульсами между электроном и решеткой, в результате чего должны испускаться или поглощаться один или большее число фононов. Легко показать, что многофононные процессы гораздо менее вероятны, чем однофононные. По этой причине ниже будет рассмотрен лишь случай однофононных переходов.

10.5.3. НЕПРЯМЫЕ ПЕРЕХОДЫ, $k_{\min} \neq k_{\max}$

Обозначим по-прежнему через \mathbf{P}_1 и \mathbf{P}_2 квазиимпульсы электрона соответственно до и после перехода. Поскольку импульсом фотона можно пренебречь, необходимо, чтобы решетка обеспечила импульс ($\mathbf{P}_1 - \mathbf{P}_2$). Этого можно достичь двумя путями: поглощением фонона с импульсом $-(\mathbf{P}_1 - \mathbf{P}_2)$ или испусканием фонона с импульсом $(\mathbf{P}_1 - \mathbf{P}_2)$. Для простоты положим $k_1 = 0$ и $k_2 = \mathbf{P}_2/\hbar = k_{\min}$. Тогда закон сохранения импульса требует, чтобы был испущен или поглощен фонон с волновым вектором k_{\min} . На основании дисперсионного соотношения между энергией фонона и его волновым числом, которое следует из вида спектра колебаний решетки (см. разд. 8.3 и 10.8), можно получить энергию E_p , соответствующую фонону с волновым числом k_{\min} . Минимальная частота, при которой может иметь место переход типа *B*, определяется условием

$$h\nu = \Delta E - E_p \quad (10.82)$$

в случае поглощения фонона, и условием

$$h\nu = \Delta E + E_p \quad (10.83)$$

в случае испускания фонона. Таким образом, минимальная частота, при которой имеет место оптический переход, определяется соотношением (10.12) в случае прямых переходов и соотношениями (10.82) и (10.83) в случае непрямых переходов. Минимальное значение $h\nu$ для непрямых переходов вытекает из равенства (10.82), которое и определяет частоту края собственного поглощения, если не говорить о многофононных переходах, обладающих весьма малой вероятностью. Поскольку величина энергии фонона E_p , как правило, мала по сравнению с ΔE , оба типа непрямых переходов

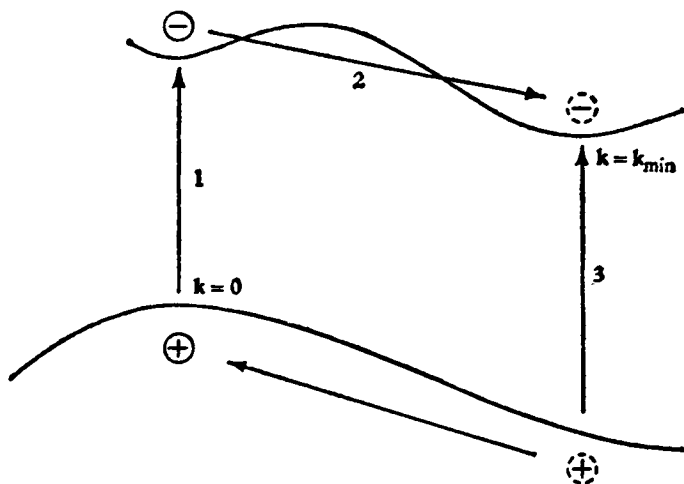


Рис. 10.10. Непрямые переходы электрона из валентной зоны в зону проводимости.

(с поглощением и испусканием фонона) сказываются на ходе кривой поглощения только в области частот, где $h\nu \approx \Delta E$.

Коэффициент поглощения для переходов, сопровождающихся испусканием или поглощением одного фонона, теоретически рассчитали Бардин, Блатт и Холл [31]. Они получили выражение для коэффициента поглощения, содержащее константу электрон-фононного взаимодействия, которую можно оценить с помощью времени релаксации для решеточного рассеяния. Эти выражения не содержат в явном виде члены, которые отдельным образом учитывают вклады различных фононов в поглощение. Точные формулы, определяющие эти вклады, даны в книге [30], § 13.5. При этом считается, что переходы происходят через «виртуальные» состояния, имеющие очень малое время жизни. Энергия не сохраняется при переходе в виртуальное состояние, в то же время импульс при этом сохраняется. Сохранение энергии имеет место для всего перехода в целом. Как при испускании, так и при поглощении фонона двойной переход может осуществляться двумя способами (см. рис. 10.10). В одном случае под действием излучения электрон без заметного изменения своего волнового вектора \mathbf{k} может покинуть состояние вблизи потолка валентной зоны ($\mathbf{k}=0$), где остается дырка с $\mathbf{k}=0$, и перейти в зону проводимости, занимая там состояние с *тем же* волновым вектором \mathbf{k} (переход 1 на рис. 10.10). Поскольку в этом состоянии электрон обладает несколько большей энергией, чем та, которая соответствует минимуму зоны проводимости, он тут же совершает переход в состояние вблизи минимума энергии, испуская или поглощая при этом фонон с импульсом, равным k_{\min} (переход 2 на рис. 10.10).

Время нахождения электрона в промежуточном состоянии считается настолько малым, что энергия его неопределенна и сохраняется лишь при двойном переходе. В другом случае электрон возбуждается излучением из глубоко лежащего состояния в валентной зоне и совершает «вертикальный» переход в состояние зоны проводимости, характеризуемое волновым вектором $\mathbf{k} \approx \mathbf{k}_{\min}$ (переход 3 на рис. 10.10). В валентной зоне остается глубоко лежащая дырка, которая совершает переход в состояние вблизи потолка валентной зоны, где $\mathbf{k}=0$, и одновременно испускает или поглощает фотон. Формула, которую получили Бардин, Блатт и Холл [31] для коэффициента поглощения, включает оба типа переходов, однако в ней отсутствует явное разделение на процессы, идущие с поглощением фотона, и на процессы, идущие с испусканием фотона. Вместе с тем, как показали Макфарлан и Робертс [35], для объяснения формы кривой поглощения у Ge и Si такое разделение необходимо. Действительно, процессы с поглощением и испусканием фотона имеют несколько различные длинноволновые границы, что видно из соотношений (10.82) и (10.83). Если коэффициент поглощения для случая не прямых переходов α_i записать в виде

$$\alpha_i = \alpha_e + \alpha_a, \quad (10.84)$$

где α_e и α_a — коэффициенты поглощения, связанные соответственно с испусканием и поглощением фотонов, то мы имеем

$$\alpha_e = 0, \quad h\nu < \Delta E + E_p, \quad (10.85)$$

$$\alpha_a = 0, \quad h\nu < \Delta E - E_p. \quad (10.86)$$

Вид кривой спектра поглощения (но не величину коэффициента поглощения) можно определить из весьма простых соображений. Если волновой вектор \mathbf{k}_{\min} существенно отличается от $\mathbf{k}=0$, то вероятность перехода из состояния с малым значением \mathbf{k} вблизи потолка валентной зоны в состояние с волновым вектором $\mathbf{k}_{\min} - \mathbf{k}'$ вблизи дна зоны проводимости ($k \ll k_{\min}$, $k' \ll k_{\min}$) будет медленно меняющейся функцией \mathbf{k} и \mathbf{k}' и может в первом приближении считаться постоянной. Коэффициент поглощения зависит в этом случае от плотностей состояний, из которых совершаются переходы, и от относительной вероятности испускания и поглощения фотонов. Ясно, что энергию E_p тоже можно считать постоянной величиной, равной энергии фотона с импульсом \mathbf{k}_{\min} . В случае прямого перехода задание величины волнового вектора \mathbf{k} в исходном состоянии полностью определяет значение волнового вектора \mathbf{k}' электрона в конечном состоянии. В случае не прямых переходов это не так, и \mathbf{k}' может принимать разные значения. Возникающий при этом избыточный импульс может быть поглощен фононом, обладающим импульсом, несколько отличающимся от импульса, соответствующего

условию $k' = 0$. Запишем соотношение для закона сохранения энергии

$$h\nu = \Delta E \pm E_p + E + E', \quad (10.87)$$

где E — энергия электрона в конечном состоянии в зоне проводимости, $-\Delta E - E'$ — начальная энергия электрона в заполненной зоне; начало отсчета энергии, как и ранее, берется у дна зоны проводимости. Теперь предположим, что значение E' фиксировано. Тогда для частоты, меняющейся в интервале от ν до $\nu + d\nu$, имеем $dE = = h d\nu$. Число состояний в этом интервале равно

$$N_c(E) dE = a E'^{1/2} dE = a (h\nu - \Delta E \pm E_p - E')^{1/2} dE,$$

где a — постоянная. Полное число пар состояний, между которыми могут осуществляться оптические переходы с частотой, лежащей в интервале $(\nu, \nu + d\nu)$, можно найти путем интегрирования по E' в валентной зоне в пределах того участка изменения E' , для которого выполняется равенство (10.87). Максимально возможное значение E'_m определяется равенством

$$E'_m = h\nu - \Delta E \pm E_p, \quad (10.88)$$

где знак минус соответствует испусканию, а знак плюс — поглощению фонона. Плотность состояний в валентной зоне $N_v(E') dE'$ в интервале энергий dE' можно записать в виде $a' E'^{1/2} dE'$. Тогда полное число пар состояний $N d\nu$, между которыми могут происходить оптические переходы для частот, лежащих в интервале $(\nu, \nu + d\nu)$, определяется выражением

$$N(\nu) d\nu = a a' h d\nu \int_0^{E'_m} (E'_m - E')^{1/2} E'^{1/2} dE' = D E_m'^2 d\nu, \quad (10.89)$$

где D — некоторая постоянная. Коэффициент поглощения для непрямых переходов, происходящих с поглощением фонона, оказывается, таким образом, пропорциональным $(h\nu - \Delta E + E_p)^2$. Он должен быть также пропорциональным количеству N_p имеющихся фононов с энергией E_p , которое определяется выражением (см. [30], § 8.16)

$$N_p = \frac{1}{\exp(E_p/kT) - 1}. \quad (10.90)$$

Итак, непрямые переходы, происходящие с поглощением фонона, дают следующий вклад в коэффициент поглощения:

$$\alpha_a = \frac{A (h\nu - \Delta E + E_p)^2}{\exp(E_p/kT) - 1} \quad (h\nu > \Delta E - E_p), \quad (10.91)$$

$$\alpha_a = 0 \quad (h\nu \leq \Delta E - E_p),$$

где A — функция, слабо меняющаяся с частотой ν (см. [30], § 13.5.2). Отношение вероятности испускания к вероятности поглощения фо-

нона равно $(N_p + 1)/N_p$. Поэтому вклад α_e в коэффициент поглощения определяется равенствами

$$\alpha_e = \frac{A(h\nu - \Delta E - E_p)^2}{1 - \exp(-E_p/kT)} \quad (h\nu > E + E_p),$$

$$\alpha_e = 0 \quad (h\nu \leq E + E_p). \quad (10.92)$$

Общее выражение для коэффициента поглощения α_i для непрямых переходов имеет вид

$$\alpha_i = \frac{A'}{\Delta E^2} \left[\frac{(h\nu - \Delta E - E_p)^2}{1 - \exp(-E_p/kT)} + \frac{(h\nu - \Delta E + E_p)^2}{\exp(E_p/kT) - 1} \right] \quad (h\nu > \Delta E + E_p),$$

$$\alpha_i = \frac{A'(h\nu - \Delta E + E_p)^2}{\Delta E^2 [\exp(E_p/kT) - 1]} \quad (\Delta E - E_p < h\nu \leq \Delta E + E_p),$$

$$\alpha_i = 0 \quad (h\nu \leq \Delta E - E_p),$$

$$(10.93)$$

где введено обозначение $A' = A\Delta E^2$.

До сих пор рассматривался лишь один тип фононов. На самом деле необходимо учесть все ветви колебаний решетки: две поперечные и одну продольную акустические ветви, а также оптические ветви. При учете двух акустических ветвей коэффициент поглощения можно записать в виде

$$\alpha_i = \alpha_{e1} + \alpha_{a1} + \alpha_{et} + \alpha_{at}; \quad (10.94)$$

величины α_{e1} и α_{et} определяются формулами типа (10.92), где A необходимо заменить соответственно на A_1 и A_t , а E_p — на E_{p1} и E_{pt} ; α_{a1} и α_{at} имеют вид, аналогичный (10.91), с заменой A соответственно на A_1 и A_t , а E_p — на E_{p1} и E_{pt} . В этом случае кривая поглощения распадается на четыре участка, разделенные следующими значениями $h\nu$: $(\Delta E - E_{p1})$, $(\Delta E - E_{pt})$, $(\Delta E + E_{pt})$, $(\Delta E + E_{p1})$.

Для простоты рассмотрим случай, когда основную роль играет лишь один тип фононов. Если изобразить зависимость $\alpha_a^{1/2}$ от $h\nu$ графически, то получим прямую, пересекающую ось $h\nu$ в точке $h\nu = \Delta E - E_p$, наклон которой равен $A^{1/2}[\exp(E_p/kT) - 1]^{-1/2}$. При очень низких температурах $E_p/kT \gg 1$ и наклон стремится к нулю. Это показано на рис. 10.11. Вклад α_a стремится к нулю при $T \rightarrow 0$ по той простой причине, что при низких температурах возбуждается очень мало фононов, и потому поглощение фонона становится весьма редким событием. Необходимо заметить, что ΔE зависит от T , так что точка пересечения прямой $\alpha_a^{1/2}$ с осью $h\nu$ также сдвигается с температурой. График зависимости $\alpha_e^{1/2}$ от $h\nu$ также представляет собой прямую, но с наклоном

$$A^{1/2} \left[1 - \exp\left(-\frac{E_p}{kT}\right) \right]^{-1/2}.$$

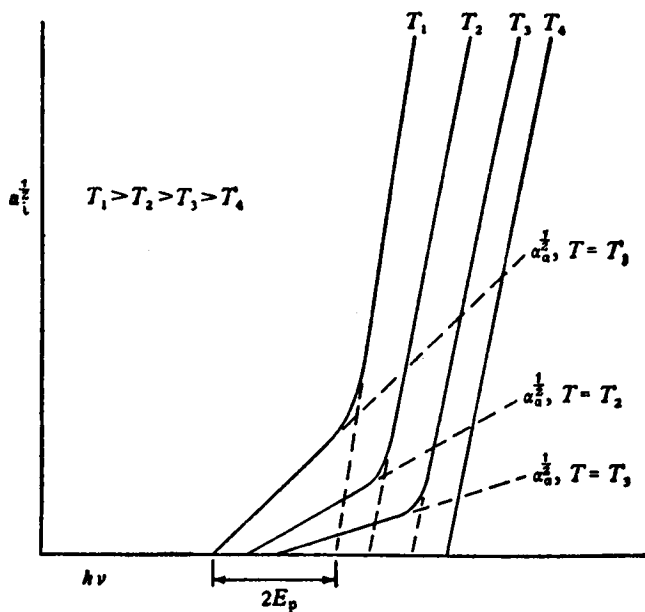


Рис. 10.11. Зависимость $\alpha_1^{1/2}$ от температуры.

Пунктиром показаны зависимости $\alpha_a^{1/2}$ и $\alpha_e^{1/2}$, сплошные кривые соответствуют $\alpha_1^{1/2} = (\alpha_a + \alpha_e)^{1/2}$.

В этом случае наклон прямых с понижением температуры слегка убывает, стремясь к постоянному значению $A^{1/2}$ при низких температурах. Прямые, соответствующие $\alpha_e^{1/2}$, на рис. 10.11 нанесены штриховыми линиями (линии с наиболее крутым наклоном). Заметим, что длина отрезка между точками пересечения прямых $\alpha_a^{1/2}$ и $\alpha_e^{1/2}$, соответствующих одной и той же температуре, с осью абсцисс равна $2E_p$, и что точка $h\nu = \Delta E$ находится как раз посередине этого отрезка. Сплошные кривые на рис. 10.11 представляют зависимость $\alpha_a^{1/2}$ от $h\nu$.

Если переходы из наивысшего состояния в валентной зоне «запрещены» в том смысле, который обсуждался выше, то в выражениях для α_e и α_a вводится дополнительный множитель $(h\nu - \Delta E \pm E_p)$ подобно тому, как это было сделано для прямых переходов (для которых $E_p = 0$) ([30], стр. 513).

Макфарлан и др. [36, 37] провели измерения коэффициента поглощения в чистых монокристаллах хорошего качества Ge и Si, пользуясь оптической аппаратурой высокой разрешающей способности. Для интерпретации опытных данных потребовалось применить выражение для α , в виде формулы (10.94), что указывает на

участие в поглощении фононов, принадлежащих к поперечным и продольным акустическим ветвям спектра колебаний решетки. В Ge состояния с минимальной энергией в зоне проводимости размещены на границе первой зоны Бриллюэна по направлениям $\langle 111 \rangle$. Значения E_{pt} и E_{pl} , соответствующие граничным значениям энергии поперечных и продольных акустических колебаний и определенные из сопоставления результатов эксперимента с теорией, оказались в случае германия равными 0,008 эВ ($\theta_t=90\text{K}$) и 0,027 эВ ($\theta_l=320\text{K}$). Эти значения θ_l и θ_t не очень хорошо согласуются со значениями констант, вычисленных на основе спектра колебаний решетки типа алмаза в работе Хси [38], который получил $\theta_l=150\text{K}$ и $\theta_t=350\text{K}$. Так, значение, найденное для θ_t , намного меньше, чем у Хси. Такое расхождение, однако, нельзя считать неожиданным, если учесть те приближения, которые были сделаны при вычислении спектра колебаний решетки типа алмаза. Недавние измерения спектра колебаний решетки в германии, проведенные Брукхаузом и Ингером [39] методом рассеяния нейтронов, подтверждают, однако, значения θ_t и θ_l , полученные в опытах по инфракрасному поглощению. Следует ожидать, что поскольку граничные энергии оптических колебаний ненамного превышают граничные энергии продольных акустических колебаний, то оптическая ветвь также должна вносить известный вклад в процесс поглощения. Однако поглощение в Ge с участием оптических фононов оказалось слабым. Это объясняется некоторыми правилами отбора, обусловленными условиями симметрии на границе зоны в направлениях $\langle 111 \rangle$ и приводящими к тому, что переходы с участием оптических фононов становятся маловероятными [40].

В случае Si значительный вклад в поглощение вносят как акустические, так и оптические фононы, причем для согласования теории с экспериментальными данными, полученными с помощью спектральной аппаратуры высокой разрешающей способности, необходимо было привлечь как продольные, так и поперечные типы двух видов колебаний. Энергии фононов и соответствующие им значения θ ($E_D=k\theta$) при $k=k_{\max}$ оказались следующими.

Акустические фононы	$E_{pt} = 0,0185$ эВ,	$E_{pl} = 0,0575$ эВ,
	$\theta_t = 212$ К,	$\theta_l = 670$ К,
Оптические фононы	$E_{pt} = 0,120$ эВ,	$E_{pl} = 0,091$ эВ,
	$\theta_t = 1420$ К,	$\theta_l = 1050$ К.

В настоящее время правильность такой интерпретации вызывает сомнения. Считается, что фонон с $\theta=212\text{K}$ действительно является поперечным акустическим фононом, однако фонон с $\theta=670\text{K}$ относится к оптическим фононам. Температуры $\theta=1050$ и 1420K , по-видимому, соответствуют комбинациям: поперечный акустический фонон + оптический фонон и оптический фонон + оптиче-

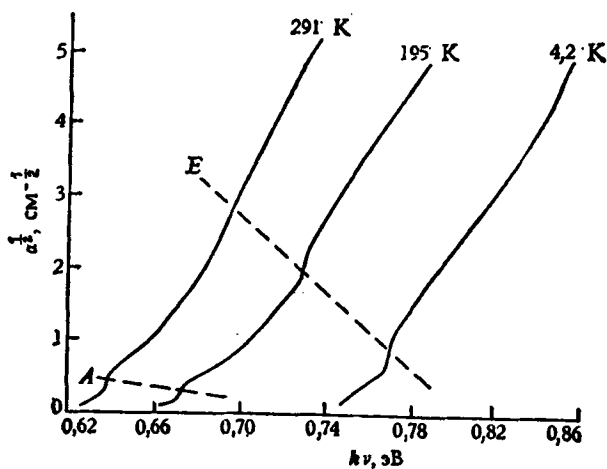


Рис. 10.12. Спектр поглощения для Ge [43].

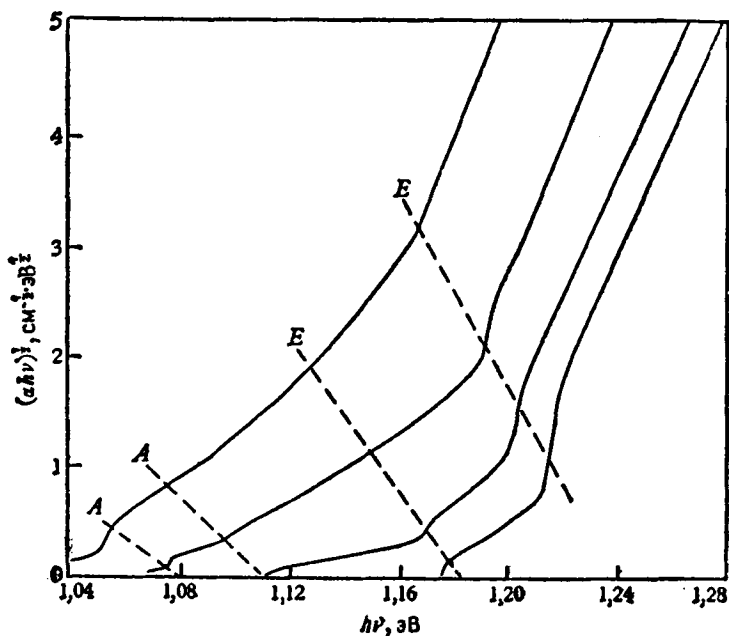


Рис. 10.13. Спектр поглощения для Si [37].

ский фонон. Если это так, то это означает, что продольные акустические фононы в Si не наблюдались [41].

Высокое спектральное разрешение, использованное в указанных выше работах, позволило обнаружить участки крутого роста поглощения в Ge и Si, которые не могут быть объяснены в рамках теории непрямых оптических переходов. Некоторые экспериментальные кривые для германия показаны на рис. 10.12, а для кремния — на рис. 10.13. Макфарлан и др. [36] приписали эти аномалии поглощению с образованием экситонов. Такой механизм поглощения будет рассмотрен в разд. 10.6.

Этому кругу вопросов было посвящено большое число экспериментальных работ. Это не является неожиданным, поскольку точное измерение поглощения вблизи собственного края дает не только наиболее точную информацию о ширине запрещенной зоны ΔE , но также информацию о величинах энергии фононов, соответствующих экстремумам в спектре колебаний решетки, дополняя, таким образом, результаты, полученные по рассеянию нейтронов. Эти вопросы широко обсуждались в ряде книг и обзоров [44—47], причем работы [46, 47] содержат большой объем справочных данных по спектрам поглощения ряда полупроводников.

10.5.4. НЕПРЯМЫЕ ПЕРЕХОДЫ, $k_{\min} = k_{\max}$

Если $k_{\min} = k_{\max}$, то наряду с прямыми «вертикальными» переходами могут иметь место также и непрямые переходы с испусканием или поглощением фонона. Однако вероятность таких переходов намного меньше вероятности прямых переходов. Если принять $k_{\min} = k_{\max} = 0$, то окажется, что существенную роль будут играть лишь фононы с малыми k . Акустические фононы с $K \approx 0$ обладают весьма малой энергией и могут не приниматься во внимание. Возможно, однако, поглощение или испускание оптического фонона с $K \approx 0$, соответствующего частоте колебаний ν_0 для *остаточных лучей*. При испускании такого фонона поглощение будет иметь место в области частот, примыкающей к предельной частоте прямых переходов с *коротковолновой* стороны, и поэтому такое поглощение будет почти незаметным на фоне значительно более сильного поглощения, обусловленного прямыми переходами. Однако непрямые переходом с поглощением оптического фонона должно соответствовать поглощение излучения, располагающееся с *длинноволновой* стороны относительно предельной частоты прямых переходов. Думке [48] высказал соображение, что небольшой длинноволновый «хвост» у границы собственного поглощения, обнаруженный на кривой поглощения в InSb (см. рис. 10.7), обусловлен именно этим механизмом. Такой вид поглощения должен был бы исчезнуть при низких температурах, когда оптические фононы практически не возбуждаются, однако имеются указания на то, что небольшой «хвост» на кривой поглощения сохраняется и при самых низких температурах.

10.5.5. ДВУХФОТОННОЕ ПОГЛОЩЕНИЕ

При обсуждении процессов оптических переходов до сих пор предполагалось, что возбуждение электрона из валентной зоны в зону проводимости вызывается поглощением одного фотона, поэтому фотон должен обладать энергией порядка ΔE или больше. Из общих принципов квантовой теории можно показать, однако, что такой переход может быть вызван двумя или несколькими фотонами, суммарная энергия которых достаточна для осуществления перехода. Вероятность процесса с участием нескольких фотонов быстро уменьшается по мере увеличения их числа, поэтому мы рассмотрим переходы с участием только двух фотонов. Следует рассмотреть также роль фононов в таких переходах.

Если два кванта, участвующие в двухфотонном процессе, имеют равные энергии, то каждый из них будет обладать вдвое меньшей энергией и, следовательно, вдвое большей длиной волны по сравнению с квантом, необходимым для возбуждения однофотонного прямого перехода. Так, например, для возбуждения электронно-дырочной пары в полупроводнике InSb, который при комнатной температуре имеет край собственного поглощения вблизи 7 мкм, необходимы два кванта излучения с длиной волны 14 мкм. Поскольку импульс двух фотонов, движущихся даже в одном направлении, все еще очень мал, такой переход будет сопровождаться очень малым изменением k , т. е. такой переход будет прямым, или, иначе говоря, «вертикальным». Так же как и для однофотонного перехода, вероятность перехода с участием двух фотонов будет зависеть от того, является ли он разрешенным или нет.

Вероятность возбуждения электрона при однофотонном процессе поглощения пропорциональна числу квантов в единице объема в возбуждающем пучке, т. е. его интенсивности I . Для двухфотонного процесса, однако, следует ожидать, что вероятность поглощения будет пропорциональна I^2 . Следовательно, здесь мы не можем ввести один постоянный коэффициент поглощения, так как его величина будет пропорциональна мощности. Поэтому коэффициент поглощения для двухфотонных переходов, α_2 , приводится к единице мощности и измеряется в $\text{см}^{-1}/(\text{Вт}/\text{см}^2)$, или $\text{см} \cdot \text{Вт}^{-1}$. Поскольку для двухфотонного процесса требуются высокие мощности, то в качестве единицы измерения поглощения обычно используется $\text{см} \cdot \text{МВт}^{-1}$. Так как вероятность однофотонного перехода по мере увеличения интенсивности растет линейно, а для двухфотонных переходов она возрастает квадратично, то ясно, что при некоторой величине мощности последние будут преобладать над первыми. Это произойдет лишь при очень высокой мощности порядка $\text{ГВт} \cdot \text{см}^{-2}$. Тем не менее двухфотонное поглощение можно уверенно наблюдать и при мощностях порядка $1 \text{ МВт} \cdot \text{см}^{-2}$, которые могут обеспечить и не самые мощные современные лазеры (см. разд. 14.6).

Двухфотонный переход можно возбудить и с помощью двух источников излучения с различными частотами ν_1 и ν_2 . При этом необходимо только, чтобы $h\nu_1 + h\nu_2 = E_t$, где E_t — энергия, требуемая для перехода. Теория таких переходов непосредственно следует из теории оптических переходов с участием фононов, если в последней фонон заменить фотоном. Квантом излучения с частотой ν_1 электрон возбуждается в виртуальное состояние с энергией E_1 над начальным состоянием, а затем возбуждается из него квантом с частотой ν_2 в конечное состояние с энергией $E - E_1$ над виртуальным состоянием. В виртуальном состоянии электрон находится очень короткое время, так что энергия сохраняется только в целом для всего перехода. Браунштейн [49] рассмотрел такие переходы для случая простых параболических зон, а Лоудон [50] включил в рассмотрение экситоны, которые должны играть определенную роль в случае прямого перехода (см. разд. 10.6). Теория этого вопроса была дана также Басовым и др. [51]. Влияние фононов, которые ответственны за не прямые переходы, применительно к этому случаю рассмотрел Ии [52], который показал, что, как можно было ожидать, вероятность таких переходов меньше, чем прямых, но ею все же нельзя пренебречь.

Если сумма энергий фотонов достаточно велика, чтобы при двухфотонном поглощении свободные электроны возбуждались с большей вероятностью, чем экситоны, то такое поглощение будет сопровождаться фотопроводимостью. Этот эффект был использован для определения величины поглощения, но потребовалась точная информация о процессах рекомбинации и захвата (см. разд. 10.9). Здесь мы имеем дело с прямым измерением поглощения. Баттон и др. [53], используя лазер на CO_2 , наблюдали такое поглощение для прямых переходов в InSb и PbTe . Двухфотонное поглощение наблюдали в ряде полупроводников, в том числе в ZnTe , GaAs , InP и InSb , Ли и Фэн [54], которые рассмотрели также влияние экситонов.

В результате большинства этих измерений значения коэффициентов поглощения α_2 были найдены с точностью до порядка величины. Гибсон и др. [55] показали, что для получения точных значений α_2 в интерпретации полученных результатов требуется большая тщательность. Главным образом это обусловлено наличием поглощения свободными носителями, рождающимися при двухфотонном процессе, и необходимостью учитывать их рекомбинацию. В этой работе [55] получены, по-видимому, наиболее точные значения α_2 для InSb и Ge и показано, каким образом можно интерпретировать результаты измерений Ли и Фэна [54]. Для InSb было найдено, что при комнатной температуре на длине волны 10,6 мкм α_2 составляет $0,2 \text{ см} \cdot \text{МВт}^{-1}$, что приблизительно на два порядка меньше значений, полученных ранее без учета поглощения на свободных носителях. Для Ge на длине волны 2,6 мкм найдено значение $\alpha_2 = 0,75 \text{ см} \cdot \text{МВт}^{-1}$, которое уменьшается практически до нуля при

увеличении λ до 3 мкм. Это поглощение соответствует прямому переходу при $k=0$, а не поглощению вблизи собственного края, которое намного слабее прямого, поскольку оптический переход на краю основного поглощения в Ge является непрямым.

10.6. Экситонное поглощение

Вопрос об экситонах уже обсуждался в разд. 3.5, а в подразд. 10.5.2 было указано, что образование экситонов с помощью механизма прямых переходов приводит к возникновению линейчатого спектра поглощения с длинноволновой стороны у края собственной полосы поглощения. В связи с этим интересно рассмотреть законы сохранения импульса применительно к образованию экситонов и процессам поглощения, обусловленным межзонными переходами. Импульс экситона связан с движением его центра масс. Если трансляционное движение центра масс экситона описывать с помощью волнового вектора K_{ex} , то энергия E_{ex} определяется равенством

$$E_{ex} = \frac{\hbar^2}{2M} K_{ex}^2 - W_{ex}^n, \quad (10.95)$$

где W_{ex}^n — энергия связи экситона в n -м возбужденном состоянии, определяемая формулой (3.10), а $M = m_e + m_h$.

Легко показать, что у экситона, образованного в результате прямого перехода, $K_{ex} \approx 0$. Действительно, мы уже видели, что согласно закону сохранения импульса при рождении электронно-дырочной пары электрон и дырка должны двигаться в *противоположных* направлениях. При этом величина скорости прямо пропорциональна волновому вектору начального состояния. В случае экситона электрон и дырка связаны и должны совместно двигаться по кристаллу в *одном и том же* направлении. Одновременно удовлетворить этим условиям можно, лишь положив $k \approx 0$, что дает $K_{ex} \approx 0$. Поэтому энергия образовавшегося экситона строго дискретна и мы получаем *линейчатый* спектр экситонного поглощения. Однако имеется конечная вероятность переходов и с *очень* малыми значениями k , приводящих к некоторому уширению линий в этом спектре.

Соотношение (10.95) получено путем интуитивного применения концепции эффективной массы. Несколько более усовершенствованный метод эффективной массы, который будет обсуждаться в разд. 11.5, применяется либо только к одним электронам, либо к одним дыркам. Дрессельхауз [56] и Эллиот [40] показали, однако, каким образом этот метод можно распространить на случай двух квазичастиц — электрона и дырки. Этот способ дает более строгую теоретическую основу для равенства (10.95). Более того, Эллиот показал, что интенсивность линий поглощения в зависимости от квантового

числа n для разрешенных (в смысле, рассмотренном в разд. 10.5) переходов пропорциональна n^{-3} , а для запрещенных переходов пропорциональна $(n^2-1)/n^5$. Для последних линия, соответствующая $n=1$, отсутствует. Некоторые аспекты этой теории рассмотрел Димок [57], а Нокс [58] дал несколько более детализированную трактовку теории экситонов, в которой процесс возбуждения электронов тесно связан с поведением кристаллической решетки.

Все эти рассуждения основаны на использовании скалярной эффективной массы. Трудности, к которым приводит учет более сложных деталей в структуре зон проводимости Ge, Si и различных соединений типа $A^{III}B^V$, обсуждали Альтарелли и Липари [59], которые показали, что таким образом можно объяснить много более тонких деталей экситонного спектра. Интенсивность поглощения в этом случае оказывается меньше, чем для соответствующего возбужденного состояния совокупности свободных атомов в отношении $(d/r_{ex})^3$, где d — постоянная решетки, а r_{ex} — радиус орбиты экситона (см. разд. 3.5). Это отношение приблизительно равно 10^{-5} . Возможно, следует отметить, что наблюдение линейчатого спектра водородоподобного поглощения не всегда является доказательством образования экситона, поскольку поглощение некоторыми типами примесей также может иметь спектр такого же вида (см. подразд. 3.4.2).

В спектрах, найденных для некоторых полупроводниковых соединений, обладающих сравнительно широкой запрещенной зоной, таких, как CdS, HgI, PbI, CdI [60], обнаружено большое число узких линий. О линейчатом спектре Cu_2O сообщали Гросс и Захарченя [61], Никитин [62] и др. Непонятной поначалу особенностью спектра Cu_2O являлось то, что здесь наблюдалось две серии линий и отсутствовала линия, соответствующая $n=1$. Теперь известно, что эти две серии линий возникают при наличии отщепленной валентной зоны и что отсутствие линии с $n=1$ объясняется тем, что переход «запрещен».

Некоторое время было непонятно также, почему для Ge и Si не были обнаружены линейчатые экситонные спектры, связанные с собственным краем поглощения. В распоряжении исследователей имелись очень чистые и совершенные монокристаллы Ge и Si, поэтому экситонное поглощение нельзя было бы спутать с линейчатыми спектрами примесей. Ситуация прояснилась, когда стало известно, что край собственного поглощения в этих материалах обусловлен непрямым переходом, что позволило объяснить отсутствие линейчатого спектра (см. ниже). Экситон, связанный с прямыми переходами из вершины валентной зоны в минимум зоны проводимости при $k=0$ (в германии этот минимум расположен на 0,15 эВ выше главного минимума), обнаружили Макфарлан и др. [29], рис. 10.9 (см. разд. 5.2). Изменение формы этой линии в зависимости от степени легирования (Sb) исследовал Рогачев [63], который показал

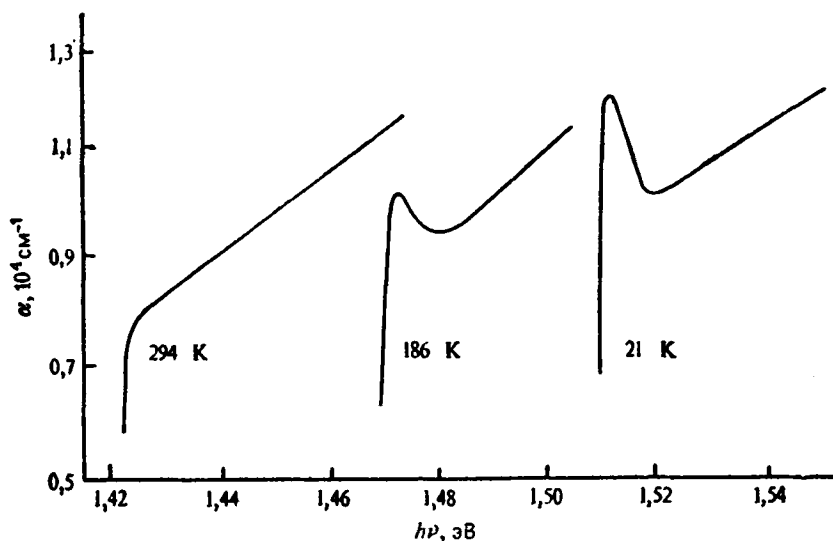


Рис. 10.14. Спектр экситонного поглощения в GaAs [64].

что эта линия может еще наблюдаться при концентрациях сурьмы до $\sim 10^{16} \text{ см}^{-3}$. GaAs, собственное поглощение которого обусловлено прямыми переходами, мог бы явиться очень удобным материалом для изучения экситонных линий в спектре поглощения. Так, Стург [64] наблюдал линию, соответствующую $n=1$, а также уширение этой линии по мере повышения температуры. Эти результаты показаны на рис. 10.14.

Экситоны могут образоваться также в результате непрямых переходов с поглощением или испусканием фона. Соответствующее экситонное поглощение может оказаться на длинноволновой стороне у края собственного поглощения, обусловленного непрямыми межзонными переходами, но только при условии, если одновременно будет испущен фотон с наибольшей энергией. Как будет видно из дальнейшего, спектральный ход кривой экситонного поглощения отличается от поглощения, при котором образуются свободные электронно-дырочные пары. Этим можно воспользоваться для отделения экситонного поглощения от поглощения, обусловленного переходами между зонами.

Теорию образования экситонов с одновременным испусканием или поглощением фононов разработали Дрессельхауз [56] и Эллиот [40]. В этой теории появляется дополнительная степень свободы в выборе начального состояния, из которого совершается непрямой переход, приводящий к образованию экситона, так как теперь импульс $\hbar\mathbf{K}_{\text{ex}}$ центра масс экситона может компенсироваться импульсом испускаемого или поглощаемого фотона. Таким образом, в

рассматриваемом случае мы уже имеем дело не с *линейчатым* спектром, а с *экситонным полосатым спектром*, четко ограниченным с длинноволновой стороны. Если E_{\min} — минимальная энергия, которая требуется для образования экситона, а E_p — энергия фонона, то для коэффициента экситонного поглощения α_{ex} получим в случае «разрешенного» перехода

$$\alpha_{\text{ex}} = D (\hbar\nu - E_{\min} \pm E_p)^{1/2} \quad (10.96)$$

и в случае запрещенного перехода

$$\alpha_{\text{ex}} = G (\hbar\nu - E_{\min} \pm E_p)^{3/2}. \quad (10.97)$$

Величины D и G приближенно можно считать постоянными. Длинноволновый край спектра поглощения экситона в n -м возбужденном состоянии определяется из условия

$$\hbar\nu = \Delta E - W_{\text{ex}}^n - E_p. \quad (10.98)$$

Для поглощения, соответствующего экситонам в самом нижнем энергетическом состоянии, мы будем опускать индекс n . Если W_{ex} — энергия связи экситона в состоянии с наименьшей энергией, то минимальное значение $\hbar\nu$, при котором еще возможен переход, удовлетворяет условию

$$\hbar\nu = \Delta E - E_p - W_{\text{ex}}. \quad (10.99)$$

Оно дает правильное значение нижнего края сплошного спектра экситонного поглощения, если под E_p понимать максимальную энергию фононов, участвующих в непрямых переходах. Область экситонного поглощения постепенно (но с изменением наклона при $\hbar\nu = \Delta E - E_p$) переходит в область собственного поглощения, обусловленного непрямыми межзонными переходами. При очень низких температурах свободные носители заряда не могут возникать в спектральном участке $\Delta E - E_p - W_{\text{ex}} < \hbar\nu < \Delta E - E_p$. Естественно, что при более высоких температурах экситоны в основном термически диссоциированы на свободные электроны и дырки. Ясно также, что свободные носители могут образоваться и при более низких температурах, однако при более высоких значениях $\hbar\nu$. Теорию не прямых переходов с образованием экситонов для Si и Ge разработали Мак-Лин и Лоудон [65], которые использовали модификацию метода эффективной массы (см. разд. 11.5).

Наиболее прямой метод исследования экситонного поглощения, связанного с непрямыми переходами, состоит в том, чтобы сравнить спектр поглощения, наблюдаемый на опыте, с тем, который имел бы место при наличии одних лишь межзонных переходов. Графически зависимость $\alpha^{1/2}$ от $\hbar\nu$ в случае межзонных переходов дает серию прямых линий, тогда как для разрешенной экситонной полосы с

нижней границей у $h\nu = E_m$ зависимость имеет вид

$$\alpha_{\text{ex}}^{1/2} \sim (h\nu - E_m)^{1/4}. \quad (10.100)$$

Такой вид спектральной зависимости легко обнаружить, так как она приводит к появлению крутого подъема на кривой поглощения. Макфарлан и др. действительно обнаружили [36] ряд подобных подъемов на кривой поглощения Ge (см. рис. 10.12), которые они отождествили с экситонным спектром поглощения, обусловленным непрямыми переходами с участием акустических фононов. При этом определенное ими значение E_{pt} совпало со значением, приведенным в разд. 10.5.3. На рис. 10.12 участки крутого роста на кривой поглощения, связанные с испусканием фонона, обозначены через E ; соответствующие крутые участки кривой, связанные с поглощением фонона, отмечены буквой A . Два менее крутых подъема кривой поглощения относятся, по мнению авторов, к случаю, когда в переходах участвуют поперечные акустические фононы. Зависимость коэффициента поглощения от $h\nu$ в этом случае такая же, как и в случае запрещенных переходов, а именно $\alpha_{\text{ex}}^{1/2} \sim (h\nu - E_m)^{1/4}$. По этой причине подъемы менее заметны на фоне кривой сплошного поглощения, обусловленного межзонными переходами. Необходимо отметить, что на опыте обнаружено небольшое остаточное поглощение (порядка $0,04 \text{ см}^{-1}$) при значениях $h\nu$, меньших значения, соответствующего длинноволновому краю поглощения в нижней экситонной полосе. Пока неизвестно, является ли это поглощение просто систематической погрешностью опыта, связанной с неполным исключением поглощения излучения свободными носителями заряда и другими механизмами, или же оно вызвано переходами более высокого порядка, сопровождающимися испусканием нескольких фононов.

В Si обнаружено [37] четыре экситонных «подъема» на кривой поглощения, в которых зависимость $\alpha^{1/2}$ от частоты подчиняется формуле (10.100) (см. рис. 10.13). В Ge значение k_{min} соответствует краю зоны в направлении $\langle 111 \rangle$, тогда как для Si установлено, что минимумы энергии расположены на осях $\langle 100 \rangle$ на некотором расстоянии от края зоны (см. разд. 13.3). В этом случае условия симметрии оказываются иными и не прямые переходы с участием как продольных, так и поперечных колебаний решетки оказываются разрешенными. К тому же правило отбора, которое в Ge исключало участие оптических фононов в процессе поглощения, перестает действовать в Si, поэтому здесь наблюдаются также и переходы, связанные с поглощением и испусканием оптических фононов. Новым доказательством того, что крутые подъемы на кривой поглощения связаны именно с экситонами, может служить и то обстоятельство, что при низких температурах они не сопровождаются параллельным ростом фотопроводимости, тогда как при комнатных темпе-

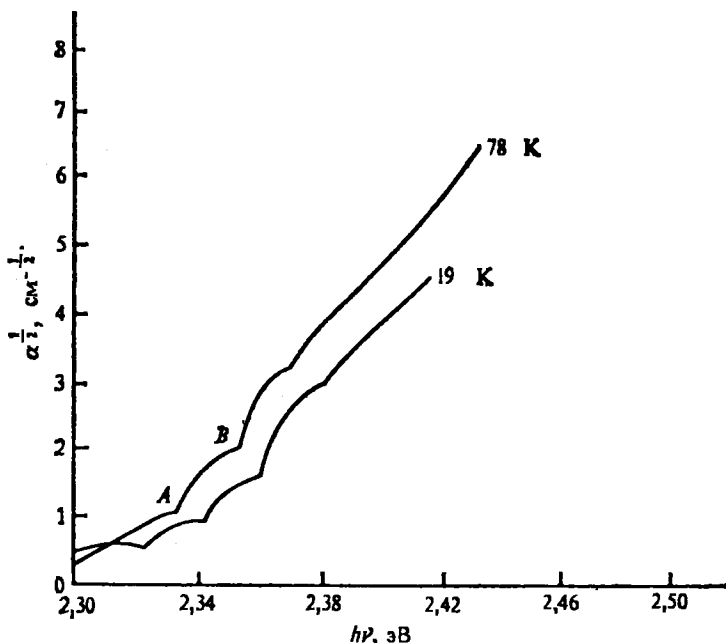


Рис. 10.15. Спектр поглощения в GaP, на котором видны экситонные полосы, возникающие в результате различных комбинаций фононов [68].

Полосы *A* и *B* связаны с испусканием LA- и соответственно TO- и LO-фононов.

ратурах ход фотопроводимости во всех деталях повторяет ход кривой поглощения.

Оценки энергии связи «непрямых» экситонов как для Si, так и для Ge дали величину порядка 5 мэВ. Новые исследования, в которых использовалась дифференциальная техника измерения поглощения, показали существование не только поглощения, соответствующего экситону с $n=1$, но также и поглощения, обусловленного экситоном с $n=2$, а также позволили получить намного более точные значения для энергии связи экситона. Для Ge Гросс и др. [66] нашли значение энергии связи экситона, равное 3,6 мэВ, в то время как для Si Шекли и Нейбур [67] получили величину 14,7 мэВ. Эти значения находятся в очень хорошем согласии с теоретическими вычислениями для Ge и Si, которые выполнили Мак-Лин и Лоудон [65].

Экситоны, связанные с непрямими переходами, наблюдались также в ряде полупроводниковых соединений. Так, например, GaP, край собственного поглощения которого обусловлен непрямими переходами, имеет спектр экситонного поглощения, состоящий из полос, а не из линий, чем существенно отличается от спектра эк-

ситонного поглощения, наблюдаемого в GaAs. Спектр поглощения, который наблюдали Гершензон и др. [68], показан на рис. 10.15. Узкие линии в спектре поглощения GaP, легированного S, обусловлены экситонами, связанными с донорами S. Такие линии наблюдал также Дин [69] для донорных примесей S, Se и Te. Позже мы обсудим излучение таких связанных экситонов и экситонных молекул (см. разд. 10.14.1 и 10.14.2).

Филлипс [70], который провел очень полный анализ экситонных спектров, наблюдаемых в большом количестве материалов, предположил, что экситоны могут образовываться и в некоторых других критических точках зонной структуры, где $\nabla_k(E_c - E_v) = 0$. Это является в то же время условием равенства скоростей электронов и дырок, что кажется разумным.

Можно ожидать, что должно иметь место поглощение, обусловленное возбуждением экситонов из их основного состояния в более высокие возбужденные состояния. Это поглощение должно происходить в очень далекой ИК-области. Используя высокие интенсивности, которые могут дать лазеры, можно возбудить достаточно большое количество экситонов, чтобы такое поглощение можно было наблюдать непосредственно. Поглощение этого вида наблюдали в чистом Ge Кононенко и др. [71]. Позже, используя интерферометрические методы, Баканан и Тимаск [72] обнаружили большое разнообразие линий поглощения в области энергий фотонов $1 \div 7$ мэВ. Они приписали эти линии экситонам, образованным не только легкими дырками в Ge, но также и тяжелыми дырками (см. разд. 2.3).

10.7. Примесное поглощение

Разнообразные электронные переходы в примесных атомах, индуцируемые падающим излучением подходящей частоты, приводят к образованию спектра примесного поглощения. Пусть E_1 — глубина примесного уровня под дном зоны проводимости, тогда при $h\nu > E_1$ получается непрерывная полоса примесного поглощения, обусловленная переходами электронов с примесных уровней в зону проводимости. Если энергия E_1 мала, то такое поглощение наблюдается лишь при очень низких температурах, когда основная часть атомов примеси оказывается неионизованной. Аналогичным образом падающее излучение может перевести электрон из валентной зоны на акцепторный уровень, расположенный вблизи этой зоны и принадлежащий неионизованному атому примеси. В разд. 3.4.1 мы видели, что примесные центры обладают рядом возбужденных состояний, например, элементы III и V групп периодической системы элементов дают в Si и Ge серию водородоподобных уровней. Возбуждение электронов примеси, т. е. перевод электрона из основного состояния в какое-либо из возбужденных состояний приводит

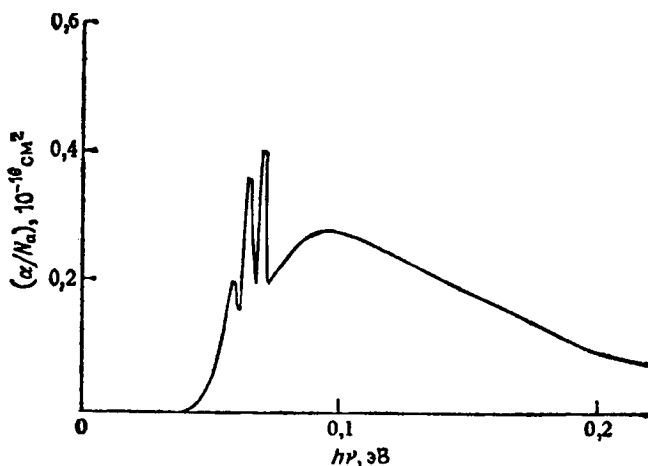


Рис. 10.16. Поглощение, обусловленное примесью Al в Si [73].

к появлению линейчатого спектра поглощения. В большинстве полупроводников донорные и акцепторные уровни размещаются, как правило, очень близко к краям зон, так что поглощение, обусловленное такими процессами, лежит в далекой ИК-области спектра; в Ge, например, основному состоянию доноров V группы соответствует уровень энергии, расположенный на глубине 0,01 эВ под дном зоны проводимости, поэтому граница ионизации должна лежать примерно около 100 мкм. Однако в Si, где основному состоянию атомов V группы соответствует уровень, который находится на 0,05 эВ ниже дна зоны проводимости, линии примесного поглощения попадают уже в более доступную для измерений область спектра. Была выполнена огромная работа по исследованию оптического поглощения примесями в Si и Ge и по сравнению экспериментальных результатов с теорией. Имеется несколько обстоятельных обзоров этих работ [43, 73—75].

На рис. 10.16 показана типичная кривая поглощения для Si с примесью Al. На этой кривой с длинноволновой стороны спектра поглощения наблюдаются три линии поглощения, обусловленные возбуждением электрона с основного уровня, а затем появляется сплошная полоса, связанная с фотоионизацией примесных центров. Хростовский и Кайзер [76] обнаружили не менее семи линий возбуждения акцепторной примеси В. Для возбужденных состояний примеси, особенно донорной примеси, положение линий находится в очень хорошем согласии с предсказаниями теории, основанной на простой водородоподобной модели. Для основного состояния предсказания теории несколько хуже согласуются с опытом, хотя бы уже потому, что теория дает одно и то же значение энергии основного

уровня независимо от природы примеси, тогда как на опыте обнаруживается довольно большой разброс этих значений в зависимости от природы атома примеси. Это и неудивительно, так как предположение о том, что кулоновский характер поля примеси сохраняется и в непосредственной близости к центру, неверно, причем такое предположение в большей мере сказывается на значении энергии в основном состоянии, где орбиты меньше, чем в случае возбужденных состояний. Для примера отметим, что акцепторный примесный уровень В в Si лежит на 0,045 эВ выше потолка валентной зоны, тогда как соответствующий уровень In находится над зоной уже на высоте 0,16 эВ.

Может иметь место примесное поглощение, соответствующее переходам электронов из валентной зоны на пустые донорные уровни, расположенные вблизи дна зоны проводимости, а также переходам электронов с заполненных акцепторных уровней в зону проводимости. Такое поглощение должно иметь место с длинноволновой стороны вблизи края собственной полосы поглощения. Примесное поглощение этого вида при небольшой концентрации примеси обычно оказывается очень слабым и слабо выражено на фоне поглощения излучения свободными носителями. Опыты по определению края собственного поглощения, как правило, проводятся на кристаллах с максимально достижимой степенью чистоты, и такого рода поглощение в них не наблюдается. В случае более глубоко лежащих донорных уровней область примесного поглощения будет достаточно хорошо отделена от области собственного поглощения, однако при этом будет меньше и вероятность ионизации доноров, а следовательно, и количество незанятых донорных уровней. Следует отметить, что для примесей с мелкими уровнями спектр поглощения будет состоять из серии линий, положение которых оказывается примерно таким же, какое ожидается для экситонного спектра. Поэтому возможно, что некоторые из спектров, приписываемые экситонам, на самом деле обусловлены примесью, обладающей водородоподобной серией уровней.

Для таких полупроводников, как Si и Ge, простую «водородоподобную» модель, как мы видели, следует несколько усовершенствовать, чтобы учесть анизотропию зоны проводимости и гофрированность изоэнергетических поверхностей валентной зоны. Кон [75] описал, каким образом это можно сделать (см. также [30], § 11.2.6). Полученные результаты он сравнил с экспериментальными данными для большого числа примесей в Si и Ge. За исключением основного состояния, для мелких примесей было получено очень хорошее согласие между теоретическими и экспериментальными результатами. Этот вопрос подробно обсуждали также Паркинсон [43] и Стоунхэм [77].

Для основных состояний примесей с мелкими уровнями, а также для примесей, основные состояния которых расположены более

глубоко в запрещенной зоне, необходим и более сложный теоретический анализ (см. разд. 11.5.1). Положение энергетических уровней таких примесей определяется главным образом с помощью электрических измерений, однако изучение спектра оптического поглощения примесей дает богатую дополнительную информацию, в особенности о возбужденных состояниях этих примесей. Вопрос о свойствах более глубоких примесей подробно рассмотрел Паркинсон [43], который привел также экспериментальные данные для большого числа примесей.

Величину коэффициента поглощения для водородоподобной модели можно легко получить, используя вероятности переходов из основного состояния электрона в возбужденные в атоме водорода. При возбуждении носителя заряда в зону проводимости или валентную зону ситуация несколько отличается от случая ионизации свободного атома водорода, поскольку зоны разрешенных энергий в полупроводнике характеризуются определенной плотностью состояний. Мосс, Барелл и Эллис [47] подробно обсудили различные подходы к этой проблеме. Для возбуждения электрона с акцепторного уровня, имеющего энергию E_a над потолком валентной зоны, в зону проводимости коэффициент поглощения α дается выражением вида

$$\alpha = AN_a (h\nu - \Delta E + E_a)^{1/2}, \quad (10.101)$$

где A — константа, N_a — концентрация акцепторов. Величина A существенно меняется в зависимости от материала и имеет значение $\sim 10^{-18}$, если α выражено в см^{-1} , N_a — в см^{-3} , а энергия — в эВ.

10.7.1. СПИНОВЫЙ РЕЗОНАНС ПРИМЕСНЫХ ЭЛЕКТРОНОВ

При наличии примесных центров, занятых одиночными электронами, в принципе можно наблюдать парамагнитный спиновый резонанс, связанный с перевертыванием спина под действием переменного магнитного поля. Вырожденные по направлению спина уровни энергии электронов расщепляются в магнитном поле каждый на два подуровня, причем величина расщепления равна $\beta g B$, где β — собственный магнитный момент электрона (магнетон Бора), равный $e\hbar/2m_0$, B — магнитная индукция, g — множитель, численно близкий к 2 (он слегка отличается от 2 из-за взаимодействия с кристаллическим полем). Разность энергий δE между расщепившимися подуровнями можно записать в виде

$$\delta E = \frac{ge\hbar B}{2m_0} = \frac{g}{2} \hbar \omega_c, \quad (10.102)$$

где ω_c — циклотронная частота для свободного электрона ($\omega_c = 2,8 \cdot 10^{10}$ Гц при $B=1$ Т). Поглощение энергии для такого типа переходов при $B=0,1$ Т попадает в сантиметровый диапазон волн; поглощение оказывается весьма слабым, однако резонансные линии поглощения являются довольно узкими и их можно наблюдать, помещая образец в объемный резонатор соответствующей частоты.

Флетчер и др. [78], а также ряд других исследователей¹⁾ наблюдали спиновый резонанс для донорных состояний в Si. Изучение спинового резонанса позволяет получить ценную информацию о примесных уровнях. В действительности линии поглощения примеси не являются одиночными, а обладают весьма интересной тонкой структурой, которая дает информацию о влиянии на примесные уровни данного центра соседних атомов кристаллической решетки и соседних атомов примеси.

Подробное обсуждение экспериментальной техники, используемой при исследовании электронного спинового резонанса в полупроводниках, дал Ланкастер [83]. Людвиг и Вудбери [84] описали применение этого метода к исследованию различных полупроводников, уделив главное внимание соединениям типа $A^{III}B^V$.

10.8. Решеточное поглощение

Наряду с поглощением, обусловленным переходами электронов в пределах энергетических зон или между зонами, и поглощением примесями существует также поглощение ИК-излучения, обусловленное возбуждением колебаний решетки, которое иногда перекрывается и несколько вуалируется уже рассмотренными нами другими видами поглощения. В разд. 8.3 мы подробно обсудили вид спектра колебаний решетки. Поглощение излучения будет иметь место, если за счет энергии поглощенного фотона возникают дополнительные фононы.

Предположим сначала, что при поглощении фотона с волновым вектором \mathbf{K} возникает один фонон с волновым вектором \mathbf{q} . Тогда по закону сохранения импульса должно выполняться равенство $\mathbf{K}=\mathbf{q}$. Как мы уже видели в связи с прямыми межзонными переходами, для фотонов ИК-излучения значение K мало, так что \mathbf{q} должно быть также малым. При этом энергия акустического фонона $\hbar\omega(\mathbf{K})$, будет очень мала, так что возникший фонон должен быть оптическим с $\mathbf{q}\approx 0$. В разд. 8.3 было показано, что частота таких фононов соответствует частоте *остаточных лучей*. Сильное поглощение на этой частоте имеет место только для кристаллов, в которых смещение атомов, обусловленное длинноволновыми оптическими колебаниями, приводит к возникновению дипольного момента и, сле-

¹⁾ См., например, [80—82].

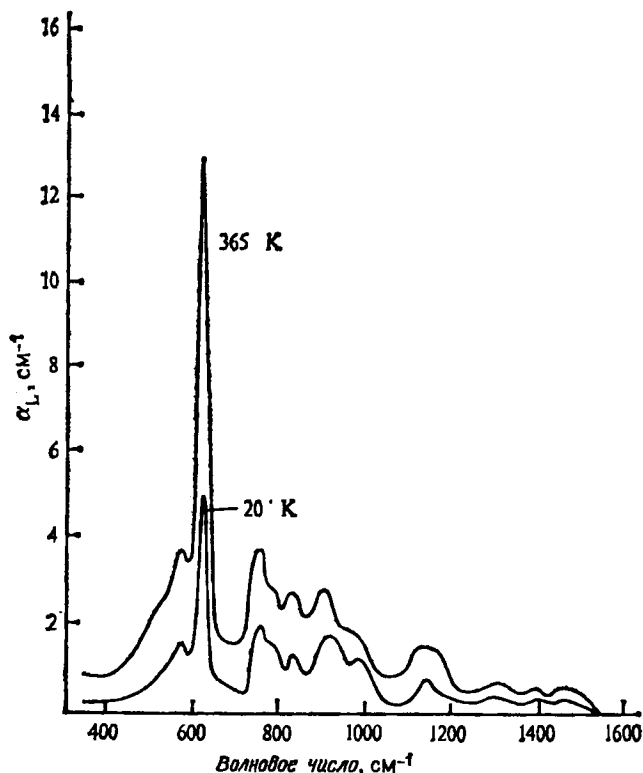


Рис. 10.17. Спектр поглощения α_L , обусловленного возбуждением колебаний кристаллической решетки в Si [86].

довательно, не должно проявляться в таких полупроводниках, как Si и Ge. Для полупроводников с частичной ионной связью следует ожидать сильного поглощения при $\omega = \omega_{\text{TO}}$, т. е. на частоте поперечных оптических колебаний, соответствующих $K=0$. Такое поглощение действительно обнаружено для всех соединений типа $A^{III}B^V$, а также для многих полярных полупроводников¹⁾. Оно соответствует поглощению в области *остаточных лучей*, которое для InSb, GaAs, GaP лежит соответственно вблизи следующих волновых чисел: 180 см^{-1} , 268 см^{-1} , 365 см^{-1} . Частоты продольных оптических колебаний при $K=0$ для этих полупроводников соответственно равны: 191 см^{-1} , 291 см^{-1} , 402 см^{-1} .

У Si и Ge поглощение, обусловленное возбуждением колебаний решетки, хотя и более слабое, но все же наблюдается, причем в

¹⁾ См., например, [85].

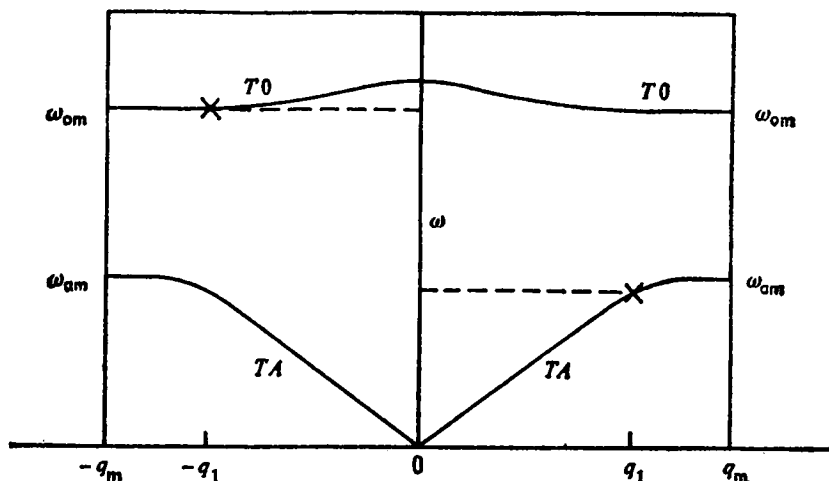


Рис. 10.18. Схема, иллюстрирующая процесс испускания двух фононов при поглощении фотона.

$\omega_{оп}$ и $\omega_{ам}$ — частоты оптического и акустического фононов с максимальным значением q_m на краю зоны.

довольно широкой спектральной области. Спектр поглощения Si показан на рис. 10.17. На спектре виден ряд отчетливых пиков, а также непрерывное поглощение в промежутках между ними. Такое поглощение должно быть обусловлено процессами, в которых участвуют по крайней мере два фонона. При наличии двух или более фононов закон сохранения импульса может выполняться и без условия малости q . Условие сохранения импульса в этом случае имеет вид $\mathbf{q}_1 + \mathbf{q}_2 = \mathbf{K}$, или, с хорошим приближением, $\mathbf{q}_1 = -\mathbf{q}_2$. Закон сохранения энергии для фотона с частотой ν запишется следующим образом:

$$\hbar\nu = \hbar\omega_1(\mathbf{q}_1) + \hbar\omega_2(\mathbf{q}_2), \quad (10.103)$$

где ω_1 и ω_2 могут относиться к различным ветвям дисперсионных кривых для решеточных колебаний. Например, ω_2 может относиться к оптическому фонону, а ω_1 — к акустическому, как показано на рис. 10.18.

Законы сохранения применительно к этому случаю обсуждали многие авторы, в том числе Джонсон [86] (см. также [30], гл. 8). Спектры поглощения, подобные наблюдаемым в Si, получены также для Ge и для полупроводников типа $A^{III}B^V$, таких, как InSb, GaAs и др., где такое поглощение наблюдается наряду с поглощением в области *остаточных лучей*. Так как в этом случае не существует каких-либо ограничений на величину импульса q рождаемых фононов, то поглощение имеет место для любых значений q

в зоне. Тем не менее наблюдаемые спектры содержат отчетливые пики. Эти пики обусловлены высокой плотностью состояний в определенных критических точках в пределах зоны. Используя дисперсионные кривые, полученные из теоретического анализа и экспериментов по рассеянию нейтронов, Джонсон и Лоудон [87] проанализировали спектры Si и Ge с этой точки зрения и получили разумное согласие с экспериментом.

Изучение инфракрасного поглощения и нейтронного рассеяния дало большое количество экспериментальных результатов для целого ряда полупроводников¹⁾ и в сильной мере стимулировало развитие теории колебаний атомов кристаллических решеток.

Хотя однофононные процессы не приводят к поглощению в чистых полупроводниках с решеткой типа алмаза, таких, как Si и Ge, подобное поглощение может происходить с участием примесей. Это поглощение изучали и подробно обсуждали Хогтон и Смит [94], которые также описали это явление в самом алмазе. Обзор работ по теории такого поглощения сделал Ньюмен [95], который развил эту теорию, рассмотрев, в частности, роль примесных пар. Он рассмотрел также роль и более сложных комплексов примесей. В этой работе он сравнил экспериментальные данные для Ge с теорией и проанализировал поглощение, обусловленное локальными колебаниями примесей в GaAs.

Проблема анализа сложных спектров поглощения, которое происходит с участием более чем двух фононов, очень сложна. Установлено, что в процессе поглощения могут участвовать три фонона, более того, наблюдались спектры поглощения, в которых, как полагают, участвуют 4 фонона. Задача несколько упрощается из-за различия температурных зависимостей некоторых из этих процессов. В нашем обсуждении мы молчаливо предполагали, что при акте поглощения должны *рождаться* два фонона, и это, безусловно, должно выполняться при низких температурах. При более высоких температурах, однако, существует заметная равновесная концентрация фононов, и в двухфононном процессе один из двух фононов может быть *поглощен* при сохранении импульса. В этом случае имеем

$$\hbar\nu = \hbar\omega_1(\mathbf{q}_1) - \hbar\omega_2(-\mathbf{q}_1). \quad (10.103a)$$

При этом должно быть $\omega_1 > \omega_2$. Как правило, ω_1 будет тогда соответствовать оптическим колебаниям, а ω_2 — акустическим.

Чтобы получить температурную зависимость поглощения, мы должны учесть очень интересный эффект, который более подробно мы рассмотрим позже в связи с работой лазеров, а именно стимулированное излучение, вероятность которого при достижении порога генерации становится равной вероятности поглощения. Вероят-

¹⁾ См., например, [88—93].

ность поглощения фононов пропорциональна числу равновесных фононов, которое в свою очередь пропорционально величине $F(\omega)$ (см. разд. 14.6):

$$F(\omega) = \left[\exp\left(\frac{\hbar\omega}{kT}\right) - 1 \right]^{-1}. \quad (10.104)$$

Как мы увидим далее, вероятность излучения фонона пропорциональна $1+F(\omega)$; единица отвечает вероятности спонтанного излучения. Полная вероятность поглощения фотона с излучением одного фонона будет тогда пропорциональна $[1+F(\omega)] - F(\omega)$ и не будет, таким образом, зависеть от температуры. Вероятность поглощения фотона с испусканием двух фононов будет пропорциональна величине

$$[1+F(\omega_1)][1+F(\omega_2)] - F(\omega_1)F(\omega_2) = 1 + F(\omega_1) + F(\omega_2),$$

в то время как для процесса поглощения фотона, в котором один фонон излучается, а другой поглощается, вероятность пропорциональна

$$[1+F_1(\omega_1)]F(\omega_2) - [1+F(\omega_2)]F(\omega_1) = F(\omega_2) - F(\omega_1).$$

Температурные зависимости для процессов с участием различных комбинаций трех фононов намного более сложны. Они были проанализированы Джонсоном [92], который таким способом идентифицировал большое число наблюдаемых фононных линий.

10.9. Фотопроводимость

Образование пар свободных электронов и дырок при поглощении излучения, вызывающего межзонные переходы, приводит к увеличению концентрации носителей заряда обоих типов. Пусть Δn и Δp — концентрации добавочных электронов и дырок соответственно. Если нет разницы в темпах захвата электронов или дырок, то $\Delta n = \Delta p$ (см. разд. 7.3), и электропроводность кристалла определяется формулой

$$\sigma = e(n_0 + \Delta n)\mu_e + e(p_0 + \Delta p)\mu_h, \quad (10.105)$$

где n_0 и p_0 — равновесные концентрации носителей заряда, определяющие равновесную электропроводность σ_0 . Таким образом, относительное изменение электропроводности равно

$$\frac{\Delta\sigma}{\sigma_0} = \frac{\sigma - \sigma_0}{\sigma_0} = \frac{\Delta n\mu_e + \Delta p\mu_h}{n_0\mu_e + p_0\mu_h}, \quad (10.106)$$

а с учетом равенства $\Delta n = \Delta p$

$$\frac{\Delta\sigma}{\sigma_0} = \frac{(1+b)\Delta p}{n_0b + p_0}, \quad (10.106a)$$

где $b = \mu_e / \mu_n$. Позже мы увидим, что при слабом освещении $\Delta p \sim I$, где I — интенсивность падающего излучения, так что $\Delta \sigma \sim I$.

Если не учитывать непрямые переходы с образованием экситонов, то длинноволновая граница собственной фотопроводимости совпадает с краем собственного поглощения. Если же учесть такие переходы, то при комнатных температурах длинноволновая граница фотопроводимости будет совпадать с длинноволновой границей экситонного поглощения, так как из-за очень малой энергии связи экситоны должны быстро распадаться на пары свободных электронов и дырок. При очень низких температурах, однако, фотопроводимость должна начинаться на частоте, которая немного выше, чем та, что соответствует длинноволновой границе экситонного поглощения. В случае диэлектриков эти соображения служили главным аргументом в пользу предположения о том, что некоторая доля поглощения обусловлена экситонами.

Так как в большинстве случаев длинноволновый край собственного поглощения в полупроводниках лежит в области энергий, меньших 1,5 эВ, то фотопроводимость обычно наблюдается в инфракрасной области спектра. Этим обусловлено использование полупроводников с неширокой запрещенной зоной, таких, как соединения свинца PbS, PbSe, PbTe и интерметаллическое соединение InSb для фотоприемников ИК-излучения [96]. Фотопроводимость, связанную с межзонными переходами, называют собственной фотопроводимостью в отличие от фотопроводимости, обусловленной примесями. Собственная фотопроводимость хорошо изучена в широком ряде веществ. Во многих случаях не было возможности проводить исследования на чистых монокристаллах, и поэтому основная часть имеющихся сведений о собственной фотопроводимости относится к пленочным образцам, полученным путем испарения в вакууме. В настоящее время известно, что процессы фотопроводимости в таких пленках чрезвычайно сложны. Собственные свойства фотопроводников PbS, PbSe, PbTe изложены автором настоящей книги в двух обзорных статьях [99]. Фотоэлектрические свойства Ge и Si описали Бурштейн, Пикус и Скляр [74], а также Мортон, Ганн и Шульц [100].

Образование в полупроводниках под действием излучения электронно-дырочных пар приводит при определенных условиях к возникновению некоторой объемной э. д. с., именуемой фото-э. д. с. Она обусловлена неоднородностями в полупроводниковых образцах, а также рядом поверхностных эффектов в них. Рассмотрение различных причин возникновения фото-э. д. с. мы пока отложим.

Фотоионизация примесей, которая обсуждалась в разд. 10.7, также приводит к фотопроводимости. Фотопроводимость, обусловленная ионизацией мелких доноров и акцепторов, наблюдается в далекой ИК-области (~ 100 мкм для примесей III и V групп в Ge), но в случае более глубоких примесей, таких, как Cu,

фотопроводимость наблюдается на более коротких длинах волн в средней ИК-области. На основе явлений собственной и примесной фотопроводимости был разработан ряд фотоприемников, перекрывающий весь спектр ИК-излучения¹⁾.

Рассмотрим теперь процессы генерации и рекомбинации носителей заряда, вызываемых освещением, для того чтобы определить величины Δn и Δp . Пусть имеется образец с прямоугольными плоскими гранями и постоянной толщиной d ; предположим, что свет падает перпендикулярно на одну из плоских граней. Допустим, далее, что излучение является монохроматическим, а его частота (или энергия кванта) достаточно велика для генерации электронно-дырочных пар. Пусть I — интенсивность этого излучения, т. е. мощность, приходящаяся на единицу площади поверхности. Если I_i — интенсивность внутри вещества, то энергия, поглощаемая в единице объема за единицу времени, будет равна αI_i , где α — коэффициент поглощения. Число электронно-дырочных пар, генерируемых в единице объема за единицу времени, окажется равным $\eta \alpha I_i / h\nu$, где η — величина, известная под названием квантового выхода. Если каждый поглощенный фотон создает *одну* электронно-дырочную пару, а поглощение, обусловленное другими механизмами, помимо межзонных переходов, пренебрежимо мало, то $\eta = 1$. Величина I_i зависит от толщины образца довольно сложным образом. Если образец настолько тонок, что $\alpha d \ll 1$, то будет иметь место интерференция между лучами, отраженными от передней и задней поверхностей, при условии, что они плоскопараллельны, и тогда количество поглощенного в образце излучения будет резко меняться с длиной волны. Подобные интерференционные эффекты рассматриваются в большинстве учебников по оптике²⁾. Если толщина образца равна нескольким длинам волн или если поверхности не являются достаточно плоскими, то интерференционные эффекты будут слабо выражены, и тогда следует суммировать *интенсивности* лучей, отраженных от различных поверхностей.

Если R_I и T_I — интенсивности соответственно отраженного и прошедшего излучения, то легко показать, что

$$R = R_s \left[1 + \frac{(1 - R_s)^2 e^{-2\alpha d}}{1 - R_s^2 e^{-2\alpha d}} \right] \quad (10.107)$$

и

$$T = \frac{(1 - R_s)^2 e^{-\alpha d}}{1 - R_s^2 e^{-2\alpha d}}, \quad (10.107a)$$

где R_s — коэффициент отражения от поверхности бесконечно толстого образца. Общее количество излучения, поглощаемого образцом за единицу времени, будет равно $(1 - T - R)I$, а средний темп

¹⁾ См., например, [101, 47].

²⁾ См., например, [96].

его поглощения при условии $\alpha d \ll 1$ равен $(1 - T - R)I/d$. Если выражения (10.107) и (10.107а) применимы, то при $\alpha d \ll 1$ имеем $(1 - T - R) \rightarrow \alpha d$, и, следовательно, излучение поглощается в объеме образца равномерно в количестве αI на единицу объема в единицу времени. Тогда темп генерации электронно-дырочных пар \mathcal{R} определяется выражением

$$\mathcal{R} = \frac{\eta \alpha I}{h\nu}, \quad (10.108)$$

причем упомянутая величина I_i совпадает в данном случае с I . Если, с другой стороны, $\alpha d \gg 1$, то непосредственно за освещенной поверхностью $I_i = I(1 - R_s)$, так как никакого сколько-нибудь заметного отражения излучения от задней поверхности не будет. В этом случае

$$I_i = I(1 - R_s)e^{-\alpha x}, \quad (10.109)$$

а темп генерации пар \mathcal{R} определяется выражением

$$\mathcal{R} = \frac{\eta \alpha I (1 - R_s) e^{-\alpha x}}{h\nu}. \quad (10.110)$$

В условиях, когда справедливы формулы (10.107) и (10.107а), общим выражением для \mathcal{R} будет

$$\mathcal{R} = \frac{\eta \alpha I (1 - R_s)}{h\nu (1 - R_s^2 e^{-2\alpha d})} \{e^{-\alpha x} + R_s \exp[-\alpha(2d - x)]\}. \quad (10.110a)$$

10.9.1. ОДНОРОДНОЕ ПОГЛОЩЕНИЕ

Рассмотрим сначала тонкий образец, для которого $\alpha d \ll 1$, так что избыточные носители заряда генерируются равномерно по всему объему. Вычислим для этого случая концентрацию избыточных носителей заряда и определим величину фототока, когда вдоль образца приложено электрическое поле \mathcal{E} . Если можно пренебречь рекомбинацией носителей заряда на поверхности, то концентрация неравновесных носителей будет везде одинакова, и тогда для концентрации избыточных дырок в полупроводнике n -типа справедливо уравнение (см. разд. 7.4)

$$\frac{d\Delta p}{dt} + \frac{\Delta p}{\tau_p} = \mathcal{R}, \quad (10.111)$$

где в данном случае \mathcal{R} — константа, определяемая выражением (10.108). Аналогичное уравнение имеет место и для избыточной концентрации электронов. Однако если времена жизни электронов и дырок одинаковы, то $\Delta n = \Delta p$, и достаточно рассмотреть одно уравнение (10.111). В стационарных условиях получим концентрацию избыточных носителей заряда Δp_0 , которая определяется

соотношением

$$\Delta p_0 = \mathcal{R} \tau_p = \frac{\eta \alpha I \tau_p}{h\nu}. \quad (10.112)$$

Оценим величину Δp_0 для типичного случая. Пусть I равно $1 \text{ мВт} \times \text{см}^{-2}$. Для $\lambda = 1 \text{ мкм}$ энергия кванта $h\nu \approx 2 \cdot 10^{-19} \text{ Дж}$, так что световой поток соответствует падению $5 \cdot 10^{15}$ квантов на квадратный сантиметр в секунду. Если положить $\eta = 1$, $\alpha = 100 \text{ см}^{-1}$ и $\tau_p = 200 \text{ мкс}$, то $\Delta p_0 = 10^{14} \text{ см}^{-3}$. Для полупроводника n -типа с $n_0 = 10^{14} \text{ см}^{-3}$ и $p_0 \ll n_0$ из (10.106) имеем $\Delta \sigma_0 / \sigma_0 = (1 + 1/b) (\Delta p_0 / n_0)$, так что если $b = 2$, то проводимость в образце при этом световом потоке, когда $\Delta p = n_0$, возрастает в 1,5 раза, и если приложенное напряжение не изменится при изменении тока, то фототок в 1,5 раза превысит темновой ток. При достаточно больших τ_p , очевидно, можно получить значительное изменение проводимости даже при средних интенсивностях светового потока; однако в дальнейшем мы будем рассматривать лишь случай относительно малого изменения проводимости, так что $\Delta \sigma_0 \ll \sigma_0$ и соответственно $\Delta p_0 \ll n_0$ или p_0 .

Прежде всего предположим, что число ловушек так мало, что $\Delta n = \Delta p$ (см. разд. 7.3). Тогда изменение проводимости описывается уравнением

$$\Delta \sigma_0 = e \Delta p_0 (1 + b) \mu_n = e \mathcal{R} \tau_p (1 + b) \mu_n, \quad (10.113)$$

а фототок i_p дается выражением

$$i_p = e \omega d \mathcal{R} \tau_p (1 + b) \mu_n \mathcal{E} = \frac{e \eta \omega d \alpha I \tau_p (1 + b) \mu_n \mathcal{E}}{h\nu}, \quad (10.114)$$

где ω — ширина образца.

Если $\Delta p \ll p$, так что τ_p постоянно, то фототок пропорционален I , в то время как при больших величинах Δp зависимость i_p от I определяется конкретным видом процесса рекомбинации. Если, например, темп рекомбинации описывается выражением типа $A n p$, как это имеет место при излучательной рекомбинации, то для стационарного случая имеем

$$\frac{\Delta p_0}{\tau_p} = A (n p - n_0 p_0).$$

Тогда Δp_0 определяется уравнением

$$\Delta p_0^2 + (n_0 + p_0) \Delta p_0 = \frac{\mathcal{R}}{A}. \quad (10.115)$$

При очень больших значениях I , когда $\Delta p_0 \gg n_0$, получим $\Delta p_0 \sim \mathcal{R}^{1/2}$ и, следовательно, $i_p \sim I^{1/2}$. Однако при малых I , когда $\Delta p_0 \ll n_0$, получим линейную зависимость, т. е. $\Delta p_0 \sim \mathcal{R}$ и, следовательно, $i_p \sim I$.

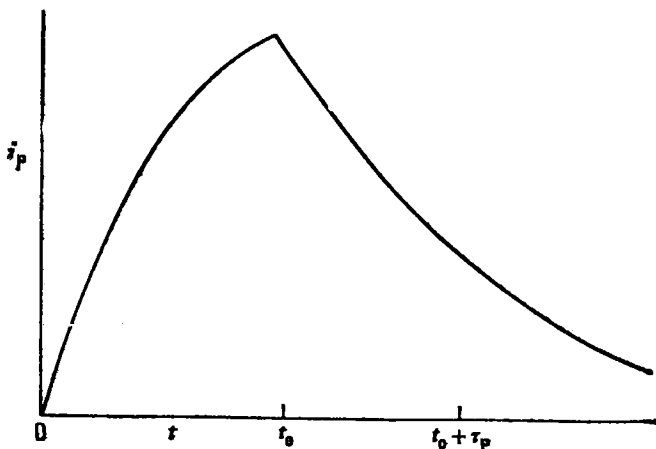


Рис. 10.19. Релаксация фотопроводимости в отсутствие ловушек.

В случае рекомбинации через ловушки можно записать (см. разд. 9.4)

$$\tau_p = \tau_0 \frac{1 + a\Delta p}{1 + c\Delta p}, \quad (10.116)$$

так что в стационарных условиях, когда $\Delta p_0 = \mathcal{R}\tau_p$,

$$(1 + c\Delta p_0) \Delta p_0 = \mathcal{R}\tau_0 (1 + a\Delta p_0). \quad (10.117)$$

Отсюда видно, что как при больших, так и при малых значениях интенсивности возбуждения величины Δp_0 , а следовательно, и i_p , пропорциональны I , но коэффициенты пропорциональности в этих крайних случаях различны.

Предположим теперь, что в момент $t=0$ имело место установившееся стационарное состояние, а затем освещение было внезапно прекращено. При достаточно малых значениях Δp , когда τ_p — константа, для последующих моментов времени t получим

$$\Delta p = \Delta p_0 \exp\left(-\frac{t}{\tau_p}\right). \quad (10.118)$$

При этом

$$i_p = i_{p_0} \exp\left(-\frac{t}{\tau_p}\right), \quad (10.119)$$

т. е. фототок убывает экспоненциально с постоянным временем τ_p .

Если в момент времени $t=0$, когда $\Delta p=0$, включен свет, интенсивность которого остается неизменной вплоть до $t=t_0$, а затем вы-

ключается при $t=t_0$, то

$$\begin{aligned} i_p &= i_{p_0} \left[1 - \exp\left(-\frac{t}{\tau_p}\right) \right] \quad (0 \leq t \leq t_0), \\ i_p &= i_{p_0} \left[1 - \exp\left(-\frac{t_0}{\tau_p}\right) \right] \exp\left(-\frac{t-t_0}{\tau_p}\right) \quad (t \geq t_0). \end{aligned} \quad (10.120)$$

Зависимость i_p от t показана на рис. 10.19. Такие кривые позволяют непосредственно определить τ_p . Однако, как будет видно из дальнейшего, при наличии в объеме полупроводника ловушек может произойти изменение формы релаксационной кривой, и если этот эффект не учесть или не исключить каким-либо образом, то полученное значение τ_p может оказаться ошибочным.

10.9.2. ВЛИЯНИЕ ЦЕНТРОВ ПРИЛИПАНИЯ

Развитая в предыдущем разделе теория фотопроводимости применима для кристаллов только с одним типом ловушек — центров рекомбинации при достаточно малых концентрациях ловушек, когда имеет место равенство $\Delta p = \Delta n$. В случае большой концентрации ловушек, когда $\Delta n \neq \Delta p$, процесс релаксации фотопроводимости не описывается экспонентой с постоянной времени τ_p (см. разд. 9.4).

Многие данные говорят о том, что в некоторых полупроводниках существуют центры захвата другого типа, которые, будучи заняты электронами, не могут дополнительно захватить и дырку, так что рекомбинации обоих типов носителей заряда на этих центрах захвата практически не происходит. Чтобы отличить такие центры захвата от обычных рекомбинационных центров, их называют центрами прилипания.

Учтем теперь возможность существования нескольких процессов с разными постоянными времени. Для определенности проанализируем поведение дырок в полупроводнике n -типа. Пусть τ_p — истинное время жизни неосновных носителей заряда, а τ_1 — среднее время до захвата дырки центром прилипания. И пусть τ_2 — среднее время, в течение которого дырка находится на центре прилипания до ее возбуждения в валентную зону. Величина Δp определяется теперь уравнением

$$\frac{d\Delta p}{dt} = \mathcal{R} - \frac{\Delta p}{\tau_p} - \frac{\Delta p}{\tau_1} + \frac{\Delta N_s}{\tau_2}, \quad (10.121)$$

где ΔN_s — концентрация избыточных дырок на центрах прилипания. Кроме того, мы имеем уравнение

$$\frac{d\Delta N_s}{dt} = \frac{\Delta p}{\tau_1} - \frac{\Delta N_s}{\tau_2}. \quad (10.122)$$

В стационарном состоянии, т. е. при $\frac{d\Delta p}{dt} = \frac{d\Delta N_s}{dt} = 0$

$$\Delta N_{s_0} = \left(\frac{\tau_2}{\tau_1}\right) \Delta p_0 \quad (10.123)$$

и

$$\Delta p_0 = \mathcal{R}\tau_p. \quad (10.124)$$

Таким образом, присутствие центров прилипания в условиях стационарного режима не сказывается на концентрации неосновных носителей заряда. Из условия электронейтральности имеем

$$\Delta n = \Delta p + \Delta N_s. \quad (10.125)$$

Тогда в стационарных условиях из соотношения (10.123) следует

$$\Delta n_0 = \Delta p_0 \left(1 + \frac{\tau_2}{\tau_1}\right). \quad (10.126)$$

Изменение проводимости σ_0 и величина стационарного фототока i_{p_0} даются соотношениями

$$\Delta\sigma_0 = e\mathcal{R}\tau_p \left(1 + b + b\frac{\tau_2}{\tau_1}\right) \mu_n \quad (10.127)$$

и

$$i_{p_0} = e w d \mathcal{R} \tau_p \left(1 + b + b\frac{\tau_2}{\tau_1}\right) \mu_n \mathcal{E}. \quad (10.128)$$

Из этих соотношений следует, что при $\tau_2 \gg \tau_1$ наличие центров прилипания приводит к весьма значительному росту фототока и что в этом случае «избыточные» дырки находятся на центрах прилипания гораздо дольше, чем в валентной зоне. Сопоставляя соотношения (10.128) и (10.114), можно при помощи выражения

$$\tau_p' = \tau_p \frac{1 + b + b\frac{\tau_2}{\tau_1}}{1 + b} \quad (10.129)$$

определить эквивалентное время жизни τ_p' . Если $\tau_2 \gg \tau_1$, то $\tau_p' \gg \tau_p$. Однако эквивалентное время жизни τ_p' не является временной характеристикой переходного процесса.

Можно получить общее решение уравнений (10.121) и (10.122), но это оказывается несколько затруднительным. Однако легко найти приближенное решение при условии, что $\tau_2 \gg \tau_1$ и $\tau_2 \gg \tau_p$. Если после достижения стационарного состояния внезапно прервать освещение, то концентрация неосновных носителей заряда начнет быстро уменьшаться с постоянной времени τ_p , так как вначале $\Delta p / \tau_1 = \Delta N_s / \tau_2$. После этого начального спада приближенно выполняется соотношение

$$\frac{d\Delta N_s}{dt} = -\frac{\Delta N_s}{\tau_2}, \quad (10.130)$$

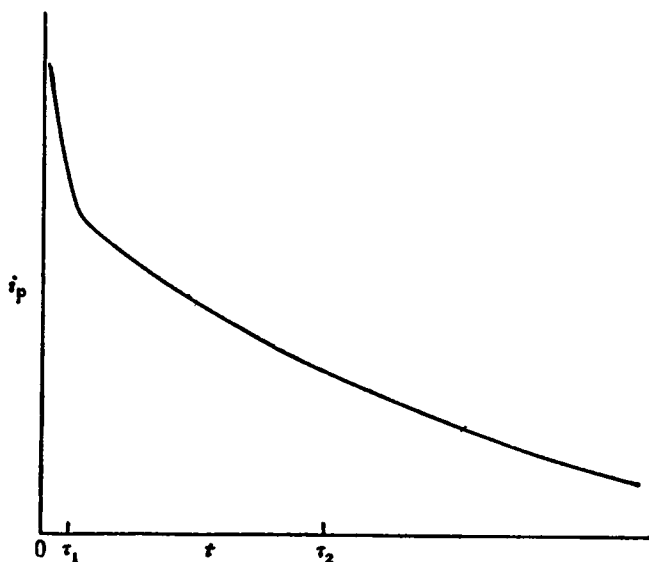


Рис. 10.20. Релаксация фотопроводимости при наличии ловушек.

так что ΔN_s уменьшается с постоянной времени τ_2 . Число избыточных носителей (электронов) при этом почти равно ΔN_s , т. е. $\Delta n \approx \Delta N_s$ и, таким образом,

$$i_{p_0} = e \omega d \mathcal{R} \tau_p b \frac{\tau_2}{\tau_1} \exp\left(-\frac{t}{\tau_2}\right). \quad (10.131)$$

Следовательно, захваченные дырки медленно с постоянной времени τ_2 , освобождаются, но пока они находятся на центрах прилипания, число избыточных *основных* носителей заряда равно числу захваченных *неосновных* носителей¹⁾. Кривая спада фототока после выключения освещения представлена на рис. 10.20. При $\tau_2 \gg \tau_1$ начальный наклон кривой, равный $(1+b) / \left(1+b \frac{\tau_2}{\tau_1}\right)$, может оказаться слишком малым, чтобы его можно было заметить. Поэтому при использовании релаксационных кривых фотопроводимости для определения τ_p необходимо проследить, чтобы вместо τ_p не измерить τ_2 .

Распределение центров захвата в полупроводнике часто оказывается гораздо более сложным, чем рассмотренное здесь. Так, возможен случай одновременного существования центров прилипания нескольких типов, причем каждому из них может соответство-

¹⁾ См., например, [99].

вать не один какой-либо уровень, а набор таких уровней. Тем не менее рассмотренная выше простейшая ситуация выявляет некоторые важные моменты, связанные с измерением времени жизни, и находит практическое применение в одном или двух важных случаях. Центры прилипания могут располагаться также и на поверхности полупроводника. В этом случае необходим особый анализ, поскольку с ними связано появление объемного заряда, играющего важную роль в процессах захвата носителей. Общий случай проблемы захвата носителей заряда и его влияния на фотопроводимость рассмотрен Роузом [102], а также Херрингом [103].

10.9.3. ВЛИЯНИЕ ПОВЕРХНОСТНОЙ РЕКОМБИНАЦИИ

До сих пор мы пренебрегали поверхностной рекомбинацией и предполагали, что величина Δp остается постоянной по всему образцу. Однако если скорость поверхностной рекомбинации s отлична от нуля, то постоянство величины Δp вблизи поверхности нарушается. Мы уже рассматривали эту задачу в разд. 9.9, где было получено решение, соответствующее случаю однородного поглощения излучения в образце. Если пренебречь влиянием центров прилипания, то единственным изменением, связанным с учетом поверхностной рекомбинации, будет замена τ_p в соотношении (10.112) на τ , где

$$\frac{1}{\tau} = \frac{1}{\tau_p} + \frac{2s}{d} \quad (10.132)$$

при условии, что толщина d значительно меньше диффузионной длины L_p (см. разд. 9.9). Отношение фототока при $s \neq 0$ к фототоку при $s=0$ равно $1/(1+2s\tau_p/d)$, и поэтому, если $s\tau_p \ll d$, поверхностная рекомбинация слабо влияет на величину фототока; с другой стороны, при $s\tau_p \gg d$ величина фототока не зависит от τ_p и определяется лишь условиями рекомбинации на поверхности. Более того, в последнем случае фототок оказывается в $d/2s\tau_p$ раз меньше, чем при $s=0$. Если же d не мало по сравнению с L_p , то, используя (9.73), получим

$$i_p = e\eta\omega d\alpha I \tau_p (1+b) \mu_n \epsilon^0 [1 - f(s)], \quad (10.133)$$

где

$$f(s) = \frac{2s\tau_p (L/d) \operatorname{sh}(d/2L_p)}{L_p \operatorname{sh}(d/2L_p) + s\tau_p \operatorname{ch}(d/2L_p)}.$$

Если $d \gg L_p$, то

$$f(s) \rightarrow \frac{2s\tau_p}{L_p + s\tau_p} \left(\frac{L_p}{d} \right), \quad (10.134)$$

так что $f(s) \rightarrow 2L_p/d$ при больших значениях s и, следовательно, как и ожидалось, при большой толщине образца ($d \gg L_p$) влияние поверхностной рекомбинации оказывается малым.

10.9.4. НЕОДНОРОДНОЕ ПОГЛОЩЕНИЕ

Теперь рассмотрим случай, когда толщина d образца велика по сравнению с величиной α^{-1} , так что падающее излучение поглощается неравномерно по толщине образца: темп поглощения излучения и, следовательно, темп генерации электронно-дырочных пар убывает по мере увеличения расстояния от освещенной поверхности. Анализ в общем случае оказывается довольно сложным, главным образом из-за влияния излучения, отраженного от неосвещенной поверхности, поэтому для простоты предположим, что $\alpha d \gg 1$. В этом случае интенсивность излучения I_i внутри образца описывается выражением (10.109). Уравнение для концентрации избыточных носителей заряда Δp в образце n -типа, если пренебречь влиянием центров прилипания, теперь имеет вид

$$D \frac{d^2 \Delta p}{dx^2} - \frac{\Delta p}{\tau_p} = - \mathcal{R} e^{-\alpha x}, \quad (10.135)$$

где $\mathcal{R} = \eta \alpha I (1 - R_s) / h\nu$. Необходимо заметить, что за исключением случая $n \gg p$, когда $D = D_n$ и $L = L_p$, коэффициент диффузии D и диффузионная длина L являются соответственно коэффициентом амбиполярной диффузии и амбиполярной диффузионной длиной (см. разд. 7.7).

Рассмотрим сначала случай толстого образца, для которого, кроме того, $d \gg L$. Соответствующим решением уравнения (10.135) будет

$$\Delta p = A e^{-x/L} - \frac{\tau_p \mathcal{R} e^{-\alpha x}}{L^2 \alpha^2 - 1}, \quad (10.136)$$

где A — постоянный коэффициент, определяемый из граничного условия при $x=0$

$$D \frac{d \Delta p}{dx} = s \Delta p. \quad (10.137)$$

Из этого условия следует

$$\frac{DA}{L} - \frac{\alpha L^2 \mathcal{R}}{\alpha^2 L^2 - 1} = -sA + \frac{s \tau_p \mathcal{R}}{\alpha^2 L^2 - 1}, \quad (10.138)$$

поэтому

$$\Delta p = \frac{\mathcal{R} \tau_p}{\alpha^2 L^2 - 1} \left(\frac{\alpha L^2 + s \tau_p}{L + s \tau_p} e^{-x/L} - e^{-\alpha x} \right). \quad (10.139)$$

Далее, интегрируя по x , найдем фототок i_p . В результате получим

$$i_p = e \mathcal{E} (1+b) \mu_h \omega \int_0^{\infty} \Delta p dx = \quad (10.140)$$

$$= \frac{e \eta \omega L I \tau_p \mu_h (1+b) (1-R_s) \mathcal{E}}{h\nu (L + s\tau_p)} \left(1 + \frac{s\tau_p}{L} \frac{1}{1+\alpha L} \right). \quad (10.141)$$

Заметим, что если $L\alpha \gg 1$, то избыточные дырки проникают значительно глубже, чем излучение, и даже при $\alpha \rightarrow \infty$ глубина проникновения их по порядку величины равна L . В последнем случае имеем $i_p = i_{p\infty}$, где

$$i_{p\infty} = \frac{e \eta \omega L I \tau_p \mu_h (1+b) (1-R_s) \mathcal{E}}{h\nu (L + s\tau_p)}. \quad (10.142)$$

Поэтому i_p можно выразить через $i_{p\infty}$ при помощи соотношения

$$\frac{i_p}{i_{p\infty}} - 1 = \frac{s\tau_p}{L} \frac{1}{1+\alpha L}. \quad (10.143)$$

Если $s\tau_p \gg L$, то при достаточно малых значениях α , когда $\alpha \approx L^{-1}$, будет иметь место сильное увеличение i_p . Это связано с тем, что при $s\tau_p \gg L$ носители заряда, генерируемые излучением на расстоянии от поверхности, меньшем L , будут рекомбинировать в основном на поверхности, но если глубина генерации существенно превышает L , то основную роль будет играть объемная рекомбинация, которой соответствует большее время жизни. Когда α становится достаточно малым, так что заметная часть излучения сможет пройти через образец, то i_p снова начинает уменьшаться, и, если $\alpha d \ll 1$, фототок i_p будет пропорционален α . В общем случае картина оказывается несколько более сложной, однако соответствующие решения можно получить из уравнения (10.135) для подходящих граничных условий при $x=0$ и $x=d$. Гаррет и Гроваль [104], а также Девор [105] получили аналитические выражения для Δp ; Гертнер [106] опубликовал серию кривых для производства численных расчетов. Типичный вид зависимости фототока от длины волны показан на рис. 10.21, где для сравнения приведена также кривая спектрального распределения поглощения α . Видно, что при малых значениях s фототок возрастает до определенной величины, после чего остается неизменным, несмотря на увеличение α . Однако при больших значениях s фототок проходит через максимум и достигает постоянной величины, меньшей, чем его соответствующая величина при $s=0$. Кривая на рис. 10.22, иллюстрирующая этот эффект, получена Гудвином [107] для InSb.

При очень больших α , когда $\alpha L \gg 1$, но d не очень велико по сравнению с L , можно получить решение уравнения (10.135) без правой части, включая ее в граничное условие. Так как теперь все дырки генерируются вблизи поверхности, то можно считать, что

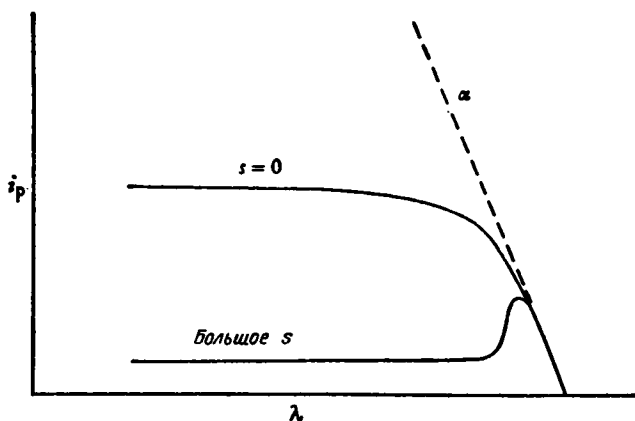


Рис. 10.21. Зависимость фототока от длины волны при большой и малой скоростях поверхностной рекомбинации s .

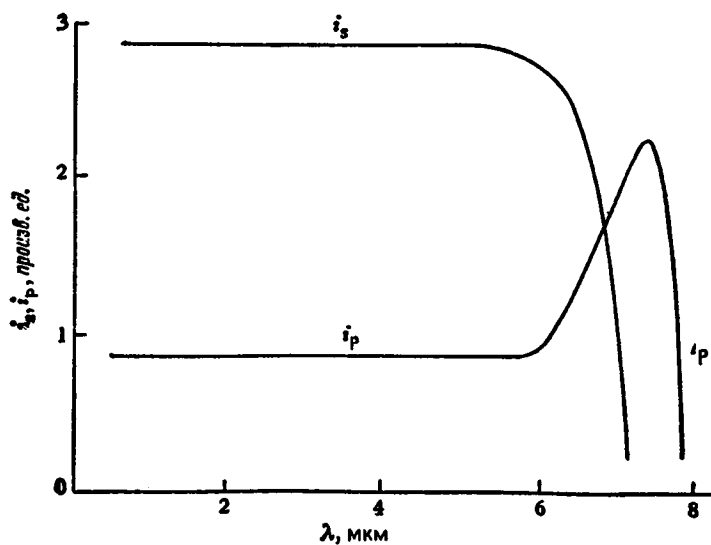


Рис. 10.22. Зависимость тока фотопроводимости и тока фотоэлектromагнитного эффекта от длины волны для InSb [107].

плотность дырочного тока при $x=0$ равна $-se\Delta p_0 + eI\eta(1-R_s)/hv$, а при $x=d$ равна $se\Delta p_d$. Соответствующее решение теперь имеет вид

$$\Delta p = A \operatorname{ch}\left(\frac{d-x}{L}\right) + B \operatorname{sh}\left(\frac{d-x}{L}\right), \quad (10.144)$$

где A и B определяются из двух граничных условий. После вычисления A и B получим

$$\Delta p = \frac{I\eta(1-R_s)\tau_p}{hv} \left[\frac{s\tau_p \operatorname{sh}\left(\frac{d-x}{L}\right) + L \operatorname{ch}\left(\frac{d-x}{L}\right)}{(L^2 + s^2\tau_p^2) \operatorname{sh}\left(\frac{d}{L}\right) + 2Ls\tau_p \operatorname{ch}\left(\frac{d}{L}\right)} \right], \quad (10.145)$$

а фототок i_p выражается соотношением

$$i_p = i_{p\infty} \frac{\left[L \operatorname{sh}\left(\frac{d}{L}\right) + s\tau_p \left(\operatorname{ch}\left(\frac{d}{L}\right) - 1 \right) \right] (L + s\tau_p)}{(L^2 + s\tau_p^2) \operatorname{sh}\left(\frac{d}{L}\right) + 2Ls\tau_p \operatorname{ch}\left(\frac{d}{L}\right)}, \quad (10.146)$$

где $i_{p\infty}$ дается выражением (10.142). При $d \ll L$ (но $\alpha d \gg 1$) имеем

$$i_p = \frac{e\eta\omega d I \tau_p \mu_h (1+b)(1-R_s)}{hv \left[2s\tau_p + d \left(1 + \frac{s^2\tau_p^2}{L^2} \right) \right]}. \quad (10.147)$$

Если $s\tau_p \ll L$ (но не $\ll d$), получим то же значение для i_p , что и в случае однородного поглощения при одинаковом темпе поглощения излучения. Иными словами, выражение (10.147) отличается от соответствующего выражения при однородном поглощении только тем, что $I\alpha d$ заменяется на $I(1-R_s)$. Если $s\tau_p$ не мало по сравнению с L , то фототок уменьшается еще больше. При этом в случае $\tau_p \gg d$ следует ожидать очень сильного уменьшения фототока по сравнению со случаем $s=0$. Эти формулы можно использовать для определения значений L и s при соответствующих условиях.

10.9.5. ФОТОПРОВОДИМОСТЬ ПРИ ДВУХФОТОННОМ ПОГЛОЩЕНИИ

До сих пор мы предполагали, что генерация свободных носителей при поглощении излучения в каждом случае обусловлена однофотонным поглощением. Однако в разд. 10.5.5 мы видели, что одновременное поглощение двух квантов излучения с частотами ν_1 и ν_2 будет приводить к появлению свободных носителей и, следовательно, к возникновению собственной фотопроводимости, при условии, что $h(\nu_1 + \nu_2) > \Delta E$. Аналогично для случая заполненного донорного уровня, расположенного на глубине E_d ниже зоны проводимости, должна иметь место примесная фотопроводимость, если

$h(\nu_1 + \nu_2) > E_d$. Подобная ситуация будет существовать также для пустого акцепторного уровня, который может принять электрон из валентной зоны. Как собственная, так и примесная фотопроводимости при двухфотонном поглощении наблюдались на опыте, причем первая — в целом ряде полупроводников¹⁾. Фотопроводимость GaAs при воздействии лазерного излучения высокой интенсивности изучали Джейрман и Ли [108], которые рассмотрели, в частности, влияние примесей.

Фотоотклик при двухфотонном поглощении часто используется для определения коэффициентов двухфотонного поглощения с помощью соотношения, подобного (10.113). Гибсон и др. [55] провели очень тщательные исследования как двухфотонной фотопроводимости, так и двухфотонного поглощения в InSb и показали, что этот метод следует использовать с большой осторожностью.

10.9.6. ВНУТРИЗОННАЯ ФОТОПРОВОДИМОСТЬ

Интересным и необычным типом фотопроводимости является фотопроводимость, которую наблюдали Гибсон и Мэггс [109] в Ge *p*-типа²⁾. Она обусловлена возбуждением дырок из зоны легких дырок в зону тяжелых. Как мы видели в разд. 10.2, такие энергетические переходы приводят к поглощению в 10-микронной области спектра. То, что при этом должен иметь место эффект фотопроводимости, с первого взгляда неочевидно, так как при таком процессе не возникает никаких дополнительных носителей заряда. Однако фотопроводимость в этом случае обусловлена различием в подвижностях легких и тяжелых дырок³⁾. Гибсон и Мэггс [109] дали теоретический анализ этого эффекта и получили очень хорошее согласие со своими экспериментальными результатами. Время релаксации внутризонной фотопроводимости очень мало, поэтому считается, что на ее основе могут быть созданы сверхбыстродействующие приемники ИК-излучения.

¹⁾ См., например, [54].

²⁾ Внутризонную фотопроводимость в германии *p*-типа при возбуждении 10-микронным излучением ранее наблюдали С. М. Рывкин и др. [Письма в ЖЭТФ. — 1969, г., т. 10, с. 470]. — *Прим. ред.*

³⁾ Механизм внутризонной фотопроводимости в германии *p*-типа оказался, однако, более сложным. Фотопроводимость чаще всего имеет отрицательный знак, то есть при облучении происходит уменьшение проводимости образца, а не увеличение, как это следовало из первоначальных простых соображений. Образующиеся при внутризонном поглощении излучения легкие дырки в стационарных условиях дают весьма малый вклад в изменение проводимости, так как они очень быстро переходят в зону тяжелых дырок, вызывая тем самым разогревание последних. Изменение подвижности горячих тяжелых дырок и дает основной вклад во внутризонную фотопроводимость (μ — фотопроводимость). — *Прим. ред.*

10.10. Диффузионная фото-э.д.с.

Когда коэффициент поглощения α не мал по сравнению с $1/d$, вследствие неоднородного распределения избыточных неосновных носителей заряда должен возникать диффузионный ток. Если, как обычно, коэффициенты диффузии электронов и дырок различны, то из условия равенства нулю составляющей полного тока в направлении нормали к освещенной поверхности следует, что должно существовать электрическое поле, перпендикулярное этой поверхности. Плотность тока j_x определяется уравнением

$$j_x = e(nb + p)\mu_n \mathcal{E}_x + e(b-1)D_n \frac{d\Delta p}{dx}. \quad (10.148)$$

Полагая $j_x=0$ и используя соотношение Эйнштейна, получаем

$$\mathcal{E}_x = - \left(\frac{kT}{e} \right) \frac{b-1}{nb+p} \frac{d\Delta p}{dx}, \quad (10.149)$$

если $\Delta p \ll n_0$. Таким образом; разность потенциалов V_t между освещенной и темной гранями равна

$$V_t = \left(\frac{kT}{e} \right) \frac{(b-1)(\Delta p_0 - \Delta p_d)}{n_0 b + p_0}. \quad (10.150)$$

При условии $\alpha d \gg 1$ и $d \gg L$ из выражения (10.145) получим

$$V_t = \left(\frac{kT}{e} \right) \frac{\eta \alpha L \tau_p (1 - R_s) (b-1)}{(n_0 b + p_0)(L + s\tau_p)(1 + \alpha L)}, \quad (10.151)$$

и если обозначить разность потенциалов при очень больших значениях α через $V_{t\infty}$, то

$$V_t = V_{t\infty} \left(\frac{\alpha L}{1 + \alpha L} \right). \quad (10.152)$$

Явление, которое мы здесь рассмотрели, известно как эффект Дембера, а разность потенциалов V_t носит название диффузионной фото-э. д. с. или фото-э. д. с. Дембера.

10.11. Фотоэлектромагнитный эффект

Если пластинка полупроводника освещается с одной стороны излучением, несколько более коротковолновым, чем излучение, соответствующее краю собственного поглощения, то из-за высокой величины коэффициента поглощения ($\alpha \sim 10^4 \text{ см}^{-1}$) электронно-дырочные пары образуются главным образом вблизи поверхности. Эти носители заряда диффундируют в глубь образца, и если теперь создано магнитное поле, направленное параллельно освещаемой поверхности, то носители заряда разного знака будут стремиться

дрейфовать к противоположным сторонам образца поперек направления магнитного поля. Это вызовет появление фото-э. д. с., величина которой в случае не очень сильных полей пропорциональна величине магнитного поля. Такое явление, известное как фотоэлектромагнитный эффект, впервые наблюдали в Cu_2O Кикоин и Носков [110] и теоретически рассмотрел Френкель [111]. Позже фотоэлектромагнитный эффект изучался в чистых монокристаллах Ge и других материалах ¹⁾.

Если параллельно освещаемой поверхности (ось z) направить магнитное поле, то в плоскости образца вдоль оси y , перпендикулярной направлению поля, потечет ток. Дырочный и электронный токи j_{hy} , j_{ey} определяются уравнениями (см. разд. 5.2.2)

$$\begin{aligned} j_{hy} &= \mu_h (e\rho\mathcal{E}_y + B j_{hx}), \\ j_{ey} &= \mu_e (en\mathcal{E}_y - B j_{ex}), \end{aligned} \quad (10.153)$$

если $\mu_h B \ll 1$, где B — магнитная индукция, а j_{hx} и j_{ex} — дырочный и электронный токи, перпендикулярные освещаемой поверхности. Как следует из разд. 10.10, $j_{ex} = -j_{hx}$ и, таким образом,

$$j_{hx} = e\rho\mu_h\mathcal{E}_x - eD_h \frac{\partial \Delta\rho}{\partial x}. \quad (10.154)$$

Подставляя из уравнения (10.149) величину \mathcal{E}_x , получаем для дырочного тока

$$j_{hx} = -eD \frac{\partial \Delta\rho}{\partial x}, \quad (10.155)$$

где D — коэффициент амбиполярной диффузии.

В некоторых ранних работах, посвященных фотоэлектромагнитному эффекту, электрическое поле в направлении оси y , обуславливающее фотоэлектромагнитную э. д. с. в разомкнутой цепи, получали из условия $j_{hy} + j_{ey} = 0$. Однако Русбрек [114] показал ²⁾, что это условие некорректно, так как при постоянном B имеем $\text{rot } \mathcal{E} = 0$. Поскольку предполагается, что $\partial \mathcal{E}_x / \partial y = 0$, то $d\mathcal{E}_y / dx$ тоже должно быть равно нулю. Для определения \mathcal{E}_y необходимо воспользоваться условием равенства нулю полного тока в направлении оси y . Это условие можно представить в виде

$$\int_0^d (j_{ey} + j_{hy}) dx = 0. \quad (10.156)$$

¹⁾ См., например, [112, 113].

²⁾ См. также [115].

Если $\Delta\rho \ll n_0$ или ρ_0 , то

$$\mathcal{E}_y = -\frac{B(b+1)}{ed(n_0b+\rho_0)} \int_0^d j_{hx} dx = \frac{BD(b+1)}{d(n_0b+\rho_0)} \int_0^d \frac{\partial \Delta\rho}{\partial x} dx = \\ = -\frac{BD(b+1)}{d(n_0b+\rho_0)} (\Delta\rho_0 - \Delta\rho_d). \quad (10.157)$$

Обычно измеряют ток короткого замыкания, обусловленный фотоэлектромагнитным эффектом. Этот ток i_s получаем, полагая в уравнениях (10.153) $\mathcal{E}_y=0$ и интегрируя их по x . Таким образом,

$$i_s = \omega\mu_h(1+b)B \int_0^d j_{hx} dx = D\omega e\mu_h(1+b)B(\Delta\rho_0 - \Delta\rho_d). \quad (10.158)$$

При условии $\alpha d \gg 1$ и $d \gg L$ можно положить $\Delta\rho_d=0$. Кроме того, из выражения (10.139) следует, что

$$\Delta\rho_0 = \frac{\mathcal{R}L\tau_p}{(\alpha L+1)(L+\sigma\tau_p)}, \quad (10.159)$$

$$\Delta\rho_0 = \frac{\eta l(1-R_s)\tau_p}{h\nu(L+\sigma\tau_p)} \left(\frac{\alpha L}{1+\alpha L} \right), \quad (10.160)$$

так что

$$i_s = \frac{\omega e\mu_h(1+b)B\eta l(1-R_s)L^2}{h\nu(L+\sigma\tau_p)} \left(\frac{\alpha L}{1+\alpha L} \right). \quad (10.161)$$

Если $i_{s\infty}$ — величина i_s при $\alpha L \gg 1$, то

$$\frac{i_s}{i_{s\infty}} = \frac{\alpha L}{1+\alpha L}. \quad (10.162)$$

Из выражений (10.141) и (10.161) можно получить формулу для отношения фотоэлектромагнитного тока при единичном магнитном поле к фототоку при единичном электрическом поле:

$$\frac{i_s/B}{i_p/\mathcal{E}} = \frac{\alpha L^2}{\left(1+\alpha L + \frac{\sigma\tau_p}{L}\right)\tau_p} \quad (10.163)$$

и

$$\frac{i_{s\infty}/B}{i_{p\infty}/\mathcal{E}} = \frac{L}{\tau_p} = \frac{D^{1/2}}{\tau_p^{1/2}}. \quad (10.164)$$

Следовательно, если из измерений подвижности известна величина коэффициента амбиполярной диффузии D , то соотношение (10.164) позволяет определить τ_p из измерений отношения $i_{s\infty}/i_{p\infty}$. Этот метод обычно используется для определения τ_p полупроводников, у которых оно слишком мало, чтобы можно было воспользоваться прямыми методами измерений. Заметим, что при малых αL значение i_s мало, так что если основная часть фотонов поглоща-

ется на глубине от освещаемой поверхности, превышающей диффузионную длину, то фотоэлектромагнитный эффект будет слабым. Фотоэлектромагнитный ток не имеет максимума вблизи края поглощения, как это свойственно току фотопроводимости при больших s . Он постепенно спадает с уменьшением α , затем начинает уменьшаться более резко, чем фототок. Спектральное распределение фотоэлектромагнитного тока i_s и фототока i_p , которое показано на рис. 10.22, измерил Гудвин [107] на образцах InSb с высокой скоростью поверхностной рекомбинации. Максимум фототока здесь хорошо выражен.

Если $\alpha L \gg 1$, но d не велико по сравнению с L , то, используя выражение (10.145), получаем

$$i_s = i_{s\infty} \frac{\left\{ s\tau_p \operatorname{sh} \left(\frac{d}{L} \right) + L \left[\operatorname{ch} \left(\frac{d}{L} \right) - 1 \right] \right\} (L + s\tau_p)}{(L^2 + s^2\tau_p^2) \operatorname{sh} \left(\frac{d}{L} \right) + 2Ls\tau_p \operatorname{ch} \left(\frac{d}{L} \right)}. \quad (10.165)$$

В этом случае

$$\frac{i_s/B}{i_p/\mathcal{E}} = \frac{L}{\tau_p} \frac{s\tau_p \operatorname{sh} \left(\frac{d}{L} \right) + L \left[\operatorname{ch} \left(\frac{d}{L} \right) - 1 \right]}{L \operatorname{sh} \left(\frac{d}{L} \right) + s\tau_p \left[\operatorname{ch} \left(\frac{d}{L} \right) - 1 \right]} \quad (10.166)$$

и, если $d \ll L$, получим интересный результат

$$\frac{i_s/B}{i_p/\mathcal{E}} = s. \quad (10.167)$$

Если $d \ll L$, то при малых величинах s фотоэлектромагнитный эффект мал. Это дает хороший метод для контроля малых значений s у тонких пластинок при условии, что $d \gg \alpha^{-1}$, а также $d \ll L$.

Теория фотоэлектромагнитного эффекта для различных условий была развита рядом авторов. Мосс, Пинчерл и Алиса М. Вудворд [112] проанализировали случай высоких интенсивностей света, когда концентрации неравновесных носителей $\Delta p \approx \Delta n$ и велики по сравнению с концентрацией основных равновесных носителей n_0 или p_0 . Пинчерл [115] дал теорию эффекта для случая сильного магнитного поля, когда $B\mu_h \gg 1$. Исчерпывающий математический анализ и критический обзор различных сделанных приближений сделал Русбрек [117]. Численные решения, приводящие к серии кривых, которые можно использовать для определения i_p и i_s для широкого круга условий и для нахождения L и τ_p по зависимости i_p и i_s от α и s , получил Гертнер [106]. Гаррет и Гроваль [104] опубликовали обзор по проблеме фотоэлектромагнитного эффекта и вывели формулы для i_s и i_p для случая, когда αd не очень велико, а также проанализировали экспериментальные данные, подтверждающие формулы при различных условиях. Хорошим примером использова-

ния фотоэлектромагнитного эффекта для определения большого числа данных, характеризующих свойства полупроводников, является исследование InSb, которое выполнил Гудвин [107].

10.12. Увлечение свободных носителей заряда фотонами

Существует еще один фотогальванический эффект, который обусловлен совершенно иным механизмом, чем те, что уже были нами описаны. Это эффект увлечения носителей заряда фотонами (photon-drag effect). Установлено, что если вдоль стержня из Ge проходит пучок излучения с длиной волны около 10 мкм (обычно это излучение CO₂-лазера с $\lambda=10,6$ мкм), то в этом стержне в продольном направлении возникает фото-э. д. с. Она обусловлена передачей импульса фотонов электронам и дыркам в полупроводнике. Если мощность излучения в пучке на единицу площади равна W , то через единицу площади за единицу времени проходит $W/h\nu$ фотонов. При поглощении одного фотона передается импульс $\hbar k = =hv/c$, где k — волновое число. Таким образом, полная величина передаваемого излучением импульса равна просто W/c на единицу площади. Поглощение в этом случае обусловлено свободными носителями, так что импульс передается электронам и дыркам. Для простоты рассмотрим, что происходит в материале явно выраженного n -типа, где дырок пренебрежимо мало. Поглощая излучение, электроны стремятся двигаться вдоль хода пучка излучения, и в стационарном состоянии этот поток должен быть сбалансирован электрическим полем \mathcal{E} , если концы стержня разомкнуты. В каждой точке стержня мы должны иметь, таким образом, равенство $\alpha W/c = =ne\mathcal{E}$, где α — коэффициент поглощения. Из-за наличия поглощения величина W изменяется вдоль стержня как $W_0 e^{-\alpha x}$, поэтому для определения фото-э. д. с. на разомкнутых концах стержня нужно провести интегрирование по длине стержня. Если пренебречь потерями на отражение, то для напряжения холостого хода получим простое выражение $V = -W_0/cen$. На практике, однако, ситуация не так проста. Этот эффект наблюдали Данишевский и др. [118], а также Гибсон, Киммит и Уолкер [119], которые учли оба типа носителей заряда и дали более сложные выражения для получаемого напряжения. Полярность фото-э. д. с. и порядок величины ее амплитуды в Ge n -типа оказались такими, как и ожидалось, но в Ge p -типа Гибсон и Уолкер [120] наблюдали смену знака при изменении температуры. Следует ожидать, что в полупроводнике сильно выраженного p -типа полярность фото-э. д. с. должна быть обратна полярности фото-э. д. с. увлечения в материале n -типа. Гибсон и Уолкер объяснили смену знака фото-э. д. с. увлечения, привлекая во внимание вырождение валентной зоны германия и про-

цессы рассеяния. Пейтел [121] наблюдал эффект увлечения носителей заряда фотонами в InAs.

Гибсон, Киммит и Уолкер [119] использовали эффект увлечения для создания приемника ИК-излучения с экстремально высоким быстродействием. Время релаксации импульса носителей заряда ограничивает временное разрешение такого приемника величиной $\sim 10^{-15}$ с, однако другие эффекты несколько ухудшают его.

10.13. Электропоглощение и электроотражение

Измерения поглощения и отражения полупроводников на частотах значительно выше края собственного поглощения выявили существование в этой области энергий еще более высокочастотных границ поглощения и нарушений непрерывности в спектрах отражения. Некоторые из этих результатов интерпретировались в представлениях зонной структуры и действительно помогли в объяснении рассчитанных зонных схем. Величины коэффициента поглощения на таких частотах очень высоки (обычно $\sim 10^5$ см $^{-1}$), поэтому измерения на ненапряженных монокристаллических образцах здесь трудноосуществимы. Естественно, что коэффициент отражения изменяется и вблизи края поглощения, однако это изменение может быть совсем слабым и его трудно обнаружить. Обзор работ, посвященных этому вопросу, содержащий также подробное обсуждение полученных результатов, сделал Кардона [122]. Обзор работ по оптическим свойствам полупроводников в видимой и ультрафиолетовой областях спектра сделал также Тауц [123].

Указанные трудности были преодолены с помощью замечательного способа, заключающегося в наблюдении *изменения* поглощения и отражения при наложении электрического поля или давления. *Изменение* отражения наблюдать легче, поэтому этот метод был широко использован для получения большого количества информации о выше- и нижележащих зонах в полупроводниках. Используя переменное электрическое поле или давление и синхронное детектирование, легко наблюдать очень малые изменения.

Изменения отражения в электрическом поле можно ожидать прежде всего из-за предсказанного Францем [124] и Келдышем [125] поглощения фотонов с энергиями, немного *меньшими* ширины запрещенной зоны, при наличии электрического поля. Электрон, возбужденный в состояние с энергией чуть ниже края зоны проводимости, может тем не менее попасть в зону проводимости с помощью процесса туннелирования (см. разд. 14.4.3), вероятность которого быстро уменьшается по мере увеличения дефицита энергии. Теоретически вычисленный коэффициент поглощения α_F для $h\nu < \Delta E$

приближенно равен

$$\alpha_F = \frac{Bhe\mathcal{E}}{\pi(512m_r)^{1/2}(\Delta E - h\nu)} \exp\left[-\frac{\pi(128m_r)^{1/2}(\Delta E - h\nu)^{3/2}}{3he\mathcal{E}}\right], \quad (10.168)$$

где \mathcal{E} — напряженность электрического поля, m_r — приведенная электронно-дырочная масса, B — константа, которая для поглощения в отсутствие поля и прямых переходов дается выражением (10.77).

Приближения, сделанные при выводе выражения (10.168), становятся неприменимы, если $(\Delta E - h\nu)$ меньше, чем $(h^2e^2\mathcal{E}^2/64m_r)^{1/2}$. Ясно, что α_F не становится бесконечным при $h\nu \rightarrow \Delta E$. Тем не менее при $h\nu \rightarrow \Delta E$ можно ожидать значительного увеличения α_F .

Выражения для α_F , полученные без этих приближений и для значений $h\nu$, превышающих ΔE , весьма сложны и содержат интегралы Эйри¹⁾. Если $h\nu > \Delta E$, коэффициент поглощения α для прямых переходов дается выражением

$$\alpha = \alpha_d [1 + F(\nu)], \quad (10.169)$$

где α_d — коэффициент поглощения при $\mathcal{E} = 0$, а $F(\nu)$ — периодическая функция, поскольку в зависимости от частоты излучения ν электрическое поле периодически либо увеличивает поглощение, либо уменьшает его. В этом обсуждении мы пренебрегли влиянием экситонов, которые, как мы знаем, изменяют поглощение вблизи его собственного края.

Обычная форма кривой поглощения показана на рис. 10.23. Можно видеть, что с низкочастотной стороны края собственного поглощения имеется крутой подъем в спектральном ходе изменения коэффициента поглощения и что на более высоких частотах изменение α является периодическим. Это типичная наблюдаемая картина, однако в деталях спектральные зависимости $\Delta\alpha$ очень разнообразны.

Для того чтобы учесть испускание или поглощение фононов при непрямых переходах, приведенные выражения следует несколько модифицировать. Пенчина [128] рассмотрел этот случай и показал, что здесь также имеет место периодическая зависимость, подобная той, что наблюдается для прямых переходов.

Влияние постоянного электрического поля на поглощение в ряде полупроводников наблюдалось, например, Келдышем, Вавиловым и Брицыным [129] в Si и Моссом [130] в GaAs. Обнаруженные изменения поглощения, в особенности с длинноволновой стороны края поглощения, хорошо согласуются с вычисленными значениями. Это явление известно как эффект Франца — Келдыша.

Мосс, Баррел и Эллис [47] указывали, однако, что, объясняя «экспоненциальные хвосты» поглощения наличием внутренних

¹⁾ См., например, [126, 127].

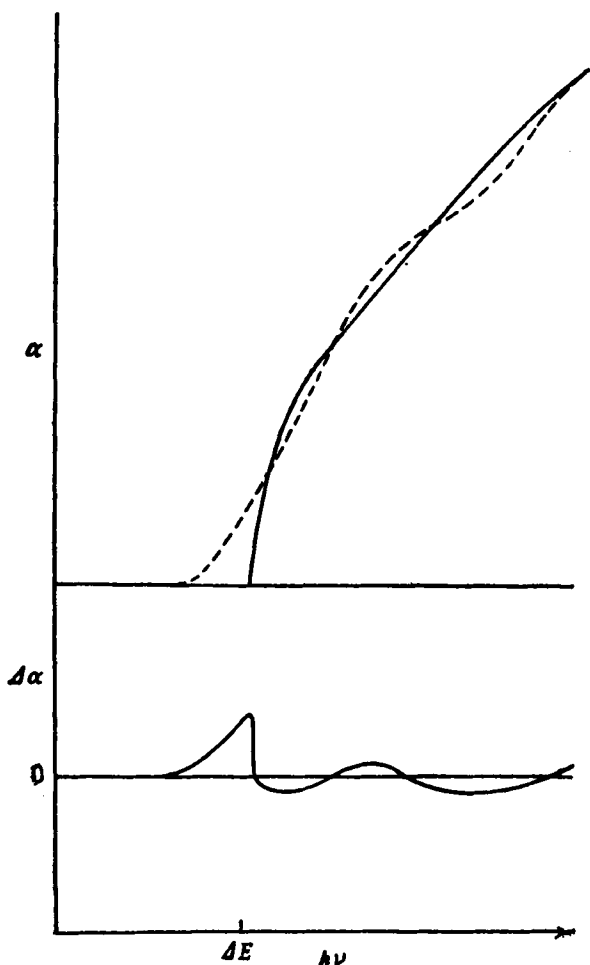


Рис. 10.23. Влияние электрического поля на поглощение вблизи края собственного поглощения, обусловленного прямыми переходами. Пунктиром показан спектр поглощения при наличии поля.

электрических полей, следует проявлять осторожность, поскольку они могут быть вызваны и другими причинами. Блосси и Хэндлер [131] использовали метод электроотражения для детального исследования края собственного поглощения в различных полупроводниках как с прямыми, так и с непрямыми переходами. Они, в частности, исследовали таким способом тонкие детали видоизменений спектра у края поглощения при наличии экситонов.

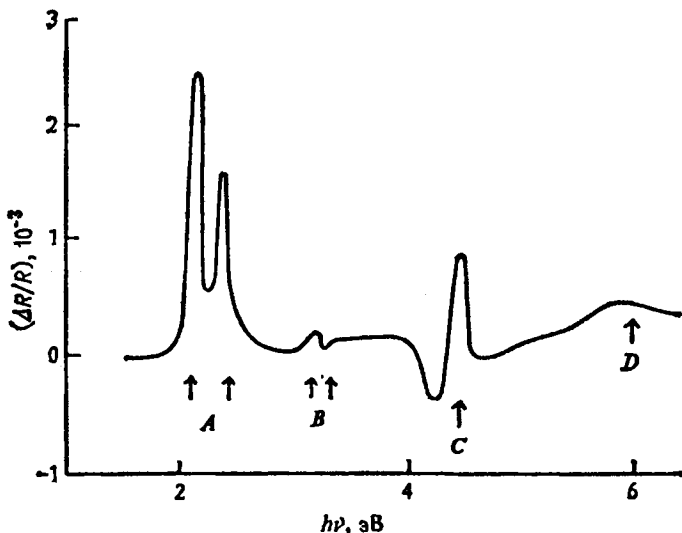


Рис. 10.24. Спектр электроотражения германия *n*-типа [133].

Еще более важным, чем изменение поглощения в электрическом поле, является соответствующее изменение отражения. Как видно из выражения (10.11), при изменении коэффициента поглощения, а следовательно, и показателя поглощения κ будет изменяться и коэффициент отражения R . Оказывается, что это изменение, как показали Аспнис и Болтка [132], намного сложнее, чем следует только из эффекта Франца — Келдыша. Сильное изменение коэффициента отражения должно иметь место также в любой критической точке, для которой $\nabla_k(E_1 - E_2) = 0$, где E_1 и E_2 относятся к двум различным зонам. По этой причине интерпретация спектров электроотражения оказалась наиболее эффективным средством, пригодным для изучения выше- и нижележащих зон в полупроводниках. Следует отметить, что $\nabla_k(E_1 - E_2) = 0$ не только в тех точках зонной структуры, для которых как $\Delta E_1(k) = 0$, так и $\Delta E_2(k) = 0$, но и в точках, для которых $E_1(k)$ и $E_2(k)$ имеют одинаковые наклоны.

Наблюдение изменений коэффициента отражения при наложении переменного электрического поля было использовано многими исследователями для получения обширной экспериментальной информации о критических точках зонной структуры для широкого класса полупроводников, в особенности для энергий в области (1—6) эВ. Спектр электроотражения германия, который получили Кардона, Шекли и Поллак [133], показан на рис. 10.24. Это довольно простой спектр. Обычно такие спектры намного сложнее, с множеством периодических изменений. Считается, что особен-

ность, обозначенная буквой *A*, соответствует переходу, обозначенному аналогично на рис. 13.1, где показана зонная структура германия. Другие особенности, обозначенные буквами *B*, *C*, *D*, вероятно, соответствуют переходам на краях зон, однако относительно их идентификации есть некоторые сомнения.

Экспериментальные методы, использованные для получения таких спектров, и связь этих спектров с зонной структурой подробно обсудил Кардона [122], который со своими сотрудниками выполнил большое число работ по спектроскопии электроотражения.

Различные виды критических точек: максимумы, минимумы, седловые точки, точки равного наклона и т. д. — все они дают свои характерные спектры электроотражения. Эти характерные особенности наблюдаемых спектров, а также дальнейшее развитие методов измерений и интерпретации рассмотрены в обзорах Серафина [135], который со своими сотрудниками провел обширные наблюдения этих спектров и дал их объяснение. Различные формы, которые могут иметь периодические спектры, перечислили и обсудили Хамакава, Хэндлер и Джо Германо [136]. Некоторые более новые достижения в технике измерений и интерпретации и их дальнейшие применения обсудили в своем обзоре Хамакава и Нишимо [137].

10.13.1. ПЬЕЗООТРАЖЕНИЕ

Тонкие детали спектра пьезоотражения несколько отличаются от тех, что наблюдаются при модуляции отражения электрическим полем, однако техника измерений и интерпретация спектров очень сходны, за исключением того, что вместо переменного электрического поля накладывается переменное давление. Поэтому в большинстве работ, на которые мы ссылались выше, обсуждаются также и спектры пьезоотражения. Прикладываемое давление приводит к малым изменениям в зонной структуре. В частности, оно модулирует разность энергий между двумя зонами и может также немного сдвигать положение критической точки в *k*-пространстве, что в свою очередь приводит к изменению отражательной способности. Этот эффект опять-таки наиболее заметен в критических точках зонной структуры; в спектрах пьезоотражения они так же ярко выражены, как и в спектрах электроотражения. Основы теории пьезоотражения, а также детали эксперимента рассмотрены в обширном обзоре Балслева [138].

10.14. Излучение полупроводников

Существует общее соотношение между излучением тела в некотором малом спектральном интервале и поглощением тела в том же самом интервале, известное как закон Кирхгофа. Если тело

сильно поглощает, то оно будет и сильно излучать, если же оно прозрачно, то излучать не будет¹⁾. Эти соображения применимы, строго говоря, только для случая термодинамического равновесия, в то время как мы будем иметь дело с явно выраженным *неравновесным* случаем, когда в объем полупроводника электрическим или оптическим путем инжектируются *избыточные* носители заряда. Однако все же следует ожидать, что полупроводник, близкий к собственному, будет слабо излучать в области длин волн, немного превышающих длину волны собственного края поглощения. В самом деле, основная часть рекомбинационного излучения, как было показано в разд. 9.2, ограничена областью длин волн, немного более коротких, чем длина волны, соответствующая краю поглощения.

10.14.1. РЕКОМБИНАЦИОННОЕ ИЗЛУЧЕНИЕ

Спектральная область рекомбинационного излучения, обусловленного свободными электронами и дырками, со стороны низких частот ограничена частотой ν , определяемой из условия $h\nu > \Delta E$. На более высоких частотах интенсивность излучения быстро спадает, причем скорость этого спада возрастает при понижении температуры (см. разд. 9.2). Как показали Хейнс, Лэкс и Флуд [139], которые провели детальное исследование спектра излучения Si, наблюдаемая спектральная форма излучения усложняется из-за наличия аннигиляции экситонов. В действительности оказывается, что при температуре 83 К примерно 5/6 рекомбинационного излучения в чистом Si обусловлено рекомбинацией через экситоны (см. разд. 9.5). При этих измерениях *избыточные* носители инжектировались в образец либо оптическим, либо электрическим способом с помощью $p-n$ -перехода. При более низких температурах излучение обусловлено главным образом аннигиляцией экситонов. Даже при 20 К оказалось возможным измерить ширину основного пика в спектре излучения, величина которой хорошо соответствует распределению энергии, ожидаемому согласно бальмановскому распределению скоростей поступательного движения экситонов при тепловом равновесии. Не является неожиданным, что при таких низких температурах энергия связи экситона в Si (0,0075 эВ) оказывается достаточной для того, чтобы в состоянии теплового равновесия значительная доля свободных электронов и дырок была связана в экситоны. Более того, это равновесие между свободными электронами, дырками и экситонами должно быстро устанавливаться.

Ситуация усложняется еще и тем, что переход в этом случае не прямой и в нем могут участвовать четыре возможных фонона TO, LO, TA, LA. Даже в случае однофононных процессов они могут

¹⁾ См., например, [96].

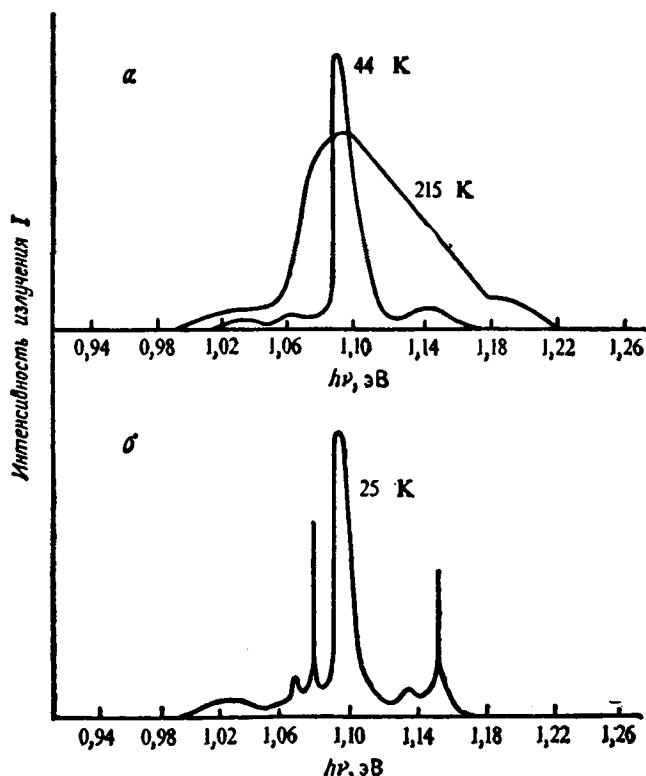


Рис. 10.25. Спектр излучения кремния.
 а — чистый Si [139]; б - Si с примесью As, $N_{As}=10^{16}$ см $^{-3}$ [140].

дать восемь линий излучения: четыре для испускания фононов и четыре для поглощения. При низкой температуре линии, соответствующие процессам с поглощением фонона, должны отсутствовать (см. разд. 10.8). При низких температурах действительно наблюдаются четыре линии, но одна из них, которая, как полагают, обусловлена испусканием ТО-фонона, преобладает, и это позволяет проанализировать форму линии (рис. 10.25,а).

10.14.2. ИЗЛУЧЕНИЕ СВЯЗАННЫХ ЭКСИТОНОВ

Эксперименты, о которых шла речь в предыдущем разделе, были выполнены на чистом Si. Когда Хейнс [140] провел исследования Si, легированного атомами As ($8 \cdot 10^{16}$ см $^{-3}$), то, кроме того излучения, о котором уже говорилось, он обнаружил ряд *очень узких* линий, две из которых были особенно сильными [см. рис. 10.25, б].

Одна из них, расположенная ближе к основной линии излучения, приписывается аннигиляции экситона, связанного с донорами As, также происходящей с испусканием ТО-фонона. Другая, расположенная на более высоких частотах, обусловлена *бесфононным* процессом, при котором изменение квазиимпульса при переходе передается всему кристаллу через донорную примесь. Поскольку *связанные* на примеси экситоны не имеют поступательной скорости, то отсутствует и связанное с этим уширение линий, и поэтому наблюдаемые пики излучения оказываются очень узкими. Найдено, что энергия связи экситона на доноре As равна 0,0065 эВ, т. е. составляет приблизительно десятую часть энергии ионизации донора.

По сравнению с элементарными полупроводниками полупроводниковые соединения обладают значительно большим разнообразием возможных энергетических состояний, и в них действительно наблюдалось очень много различных спектров излучения экситонов, связанных как на донорах, так и на акцепторах. Томас и Хопфилд [141] обнаружили, например, сложные линейчатые спектры, обусловленные экситонами, связанными на донорах в CdS и CdSe. Рейнольдс [142] интерпретировал некоторые из этих линий как результат такой аннигиляции связанного экситона, после которой доноры остаются в различных возбужденных состояниях.

Соединения типа $A^{III}B^V$, в особенности GaAs и GaP и их сплавы, также широко исследовались, главным образом из-за их прикладного значения как материалов для изготовления светодиодов. Эти соединения, легированные различными примесями, также дают большое разнообразие узколинейчатых спектров излучения, возникающих в результате аннигиляции экситонов, связанных на различных несовершенствах кристаллической решетки.

Кроме связывания экситона с одиночными донорами или акцепторами существует также ряд других процессов. Некоторые из них мы детально обсудим позже (см. разд. 14.1 и 14.2).

Действительно, все процессы поглощения, которые обсуждались в предыдущих разделах, приводят к излучению. Например, электроны могут перейти с неионизованных доноров на вакантные уровни в валентной зоне, т. е. рекомбинировать с дыркой, аналогично электроны зоны проводимости могут рекомбинировать с дыркой на неионизованном акцепторе. Все обсуждавшиеся до сих пор процессы, если их рассматривать как процессы испускания излучения, являются неэффективными, поскольку большинство актов рекомбинации осуществляется путем безызлучательных переходов.

Был обнаружен, однако, и ряд значительно более эффективных в этом смысле процессов. Если наряду с неионизованными донорами в полупроводнике имеются также и неионизованные акцепторы, то рекомбинация может осуществиться посредством обмена электроном между ними. Если не учитывать энергию кулоновского взаимодействия, то энергия, выделяемая при таком переходе, будет равна

$\Delta E - \varepsilon_d - \varepsilon_a$, где ε_d и ε_a — соответственно энергии ионизации донора и акцептора. Если, однако, учесть, что расстояние между атомами донора и акцептора конечно и равно R , то следует учитывать и энергию кулоновского взаимодействия, равную $-e^2/4\pi\epsilon_0 R$, которая будет высвобождаться при переходе. Энергия излучаемого кванта будет тогда определяться выражением

$$h\nu = \Delta E - \varepsilon_d - \varepsilon_a + \frac{e^2}{4\pi\epsilon_0 R}. \quad (10.170)$$

Поскольку электроны и дырки оказываются *связанными*, линия излучения для каждого значения R будет резкой, но будет существовать не одна, а целая серия линий, соответствующих различным группам позиций, занимаемых атомами примесей в кристалле, соответствующих расстояниям R_1, R_2, \dots и т. д. Томас и др. [143] наблюдали такие серии резких интенсивных линий для пар Si и Te в GaP, а также в ряде других полупроводников. Дин и др. [144] наблюдали спектры интенсивного излучения GaP, обусловленного, как полагают, наличием пар In—O и Cd—O. Пары этих примесей стремятся расположиться на минимально возможном расстоянии, поэтому основная часть излучения сосредоточена в сильной узкой линии.

Другим типом центра, который дает интенсивное излучение, в том числе хорошо известно красное свечение светодиодов из GaAs, является так называемый изоэлектронный центр, который образуется, когда один из атомов замещен другим, но с той же валентностью. Такие центры обсуждались в большом числе работ, в том числе в работе Уайта и др. [145], в которой спектр GaAs сравнивается со спектром InP.

Бebb и Уильямс [146] довольно подробно рассмотрели теорию излучения центров различного типа. Они же [150] сделали обзор большого количества имеющихся экспериментальных данных по излучению GaAs и сравнили их со своими результатами. Теорию вопроса, главным образом применительно к GaAs и InP, обсудил Филлипс [147]. Дин [148] составил обширный обзор работ по спектроскопии узколинейчатого излучения полупроводников.

10.14.3. ИЗЛУЧЕНИЕ СВОБОДНЫХ НОСИТЕЛЕЙ ЗАРЯДА

Тепловое излучение полупроводника обычно следует излучению идеального «черного тела». Однако, как отмечалось выше, оно будет промодулировано вариациями поглощения и, следовательно, излучательной способности. Значительная часть этого излучения будет испускаться свободными носителями заряда, а также колебаниями решетки. Последнее мы обсудим в следующем разделе. Излу-

чение свободных носителей можно увеличить, инжектируя избыточные носители. Это явление наблюдали Ульмер и Фрэнкл [149]. Они смогли, например, наблюдать увеличение интенсивности излучения в той же части спектра, где существует избыточное поглощение, обусловленное возбуждением дырок из одной валентной зоны в другую (см разд. 10.14.2).

10.14.4. ИЗЛУЧЕНИЕ КОЛЕБАНИЯМИ АТОМОВ РЕШЕТКИ

Максимум «теплового» излучения тела, имеющего комнатную температуру, оказывается вблизи 10 мкм, поэтому колебания атомов решетки, которые дают наибольший вклад в поглощение в этой области спектра и должны быть главным образом ответственны за это излучение. Как мы видели (разд. 10.14.1), колебания атомов решетки дают спектр поглощения, обладающий характерными пиками, поэтому следует ожидать, что и излучение решетки будет иметь подобный спектр. Установлено, что это действительно так. Стирволд и Поттер [116] использовали это излучение для изучения спектров колебаний решетки. Такие исследования в некоторых отношениях дополняют исследования поглощения, которые мы обсуждали в разд. 10.8, и дают дополнительную информацию о характерных частотах фононов. В частности, на частоте остаточных лучей очень хорошо выделяются полосы излучения, которые были обнаружены для ряда полупроводников типа $A^{III}B^V$. Вид спектра поглощения достаточно сложен из-за влияния сильного поглощения и соответственно высокой отражательной способности, обусловленных взаимодействием с излучением ТО- и LO-фононов (см. разд. 8.3 и 14.3). Однако из анализа излучения можно получить частоты этих фононов при $k=0$. Найденные таким способом значения очень хорошо согласуются со значениями, полученными с помощью комбинированных измерений поглощения и отражения (см. разд. 10.8).

ЛИТЕРАТУРА

1. *Fan H. Y.*, Semiconductors and Semimetals, Academic Press, 3, 1967, p. 406.
2. *Bardeen J.*, Phys. Rev., 79, 216 (1950).
3. *Fan H. Y.*, Rep. Progr. Phys., 19, 107 (1956).
4. *Stern F.*, Solid State Physics, Academic Press, 15, 1963, p. 300.
5. *Benedict T. S.*, *Shockley W.*, Phys. Rev., 89, 1152 (1953).
6. *Becker M.*, *Fan H. Y.*, Semiconducting Materials, ed. Henisch H. K., Butterworth, 1951, p. 132.
7. *Gibson A. F.*, Proc. Phys. Soc., B66, 588 (1963).
8. *Gibson A. F.*, Proc. Phys. Soc., B69, 488 (1956).
9. *Briggs H. B.*, *Fletcher R. C.*, Phys. Rev., 87, 1130 (1952); 91, 1342 (1953).
10. *Kaiser W.*, *Collins R. J.*, *Fan H. Y.*, Phys. Rev., 91, 230 (1953); 91, 1380 (1953).
11. *Kahn A. H.*, Phys. Rev., 97, 1647 (1955).

12. *Fan H. Y., Shepherd M. L., Spitzer W. G.*, Proceedings of Atlantic City Photo-conductivity Conference, John Wiley and Sons, Chapman and Hall, 1956, p. 184.
13. *Neuringer L. J., Mitkard R. C., Aggarwal R. L.*, Phys. Rev. Lett., **15**, 664 (1965); Proc. VIII th Intern. Conf. Phys. Semicond., Journ. Phys. Soc. Japan, **21**, 582 (1966).
14. *Blinovski J., Myceilski J.*, Phys. Rev., **136**, A266 (1964); **140**, A1024 (1965).
15. *Смит Р. А., Цвердлинг С., Дерматис С. Н., Зиро Дж. П.*, Труды IX Международной конференции по физике полупроводников.— Л.: Наука, 1969, с. 156.
16. *Kaczmarek E., Gortel Z. W.*, Proc. XIIth Intern. Conf. Phys. Semicond., Teubner, 1974, p. 1207.
17. *Barlow H. E. M.*, Proc. Instn Elect. Engers., **102B**, 179, 186 (1954); **104C**, 35 (1957).
18. *Smith S. D., Moss T. S., Taylor K. W.*, Phys. Chem. Solids, **11**, 131 (1959).
19. *Houghton J., Smith S. D.*, Infra-red Physics, Oxford University Press, 1966, Ch. 4.
20. *Piller H.*, Semiconductors and Semimetals, Academic Press, **8**, 1972, p. 103.
21. *Luttinger J. M., Goodman R. R.*, Phys. Rev. **100**, 673 (1955).
22. *Luttinger J. M.*, Phys. Rev., **102**, 1030 (1956).
23. *Gold L., Bullis W. M., Campbell R. A.*, Phys. Rev., **103**, 1250 (1956).
24. *Zeiger H. J., Lax B., Dexter R. N.*, Phys. Rev., **105**, 495 (1957).
25. *Lax B., Mauroides J. G.*, Solid State Physics, Academic Press, **11**, 1960, p. 261.
26. *Ландау Л.*, Zs. Phys., **64**, 629 (1930).
27. *Burstein E., Picus G. S., Gebbie H. A., Blatt F.*, Phys. Rev., **103**, 826 (1956).
28. *Zwerdling S., Keyes R. J., Foner S., Kolm H. H., Lax B.*, Phys. Rev., **104**, 1805 (1956).
29. *Macfarlane G. G., McLean T. P., Quarrington J. E., Roberts V.*, Proc. Phys. Soc., **B71**, 863 (1958).
30. *Smith R. A.*, The Wave-Mechanics of Crystalline Solids, Chapman and Hall, 1969.
31. *Bardeen J., Blatt F. J., Hall L. H.*, Proceedings of Atlantic City Photo-Conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 146.
32. *Roberts V., Quarrington J. E.*, Journ. Electron., **1**, 152 (1955).
33. *Dash W. C., Newman R.*, Phys. Rev., **99**, 1151 (1955).
34. *Zwerdling S., Roth Laura M., Lax B.*, Phys. Rev., **109**, 2207 (1958).
35. *Macfarlane G. G., Roberts V.*, Phys. Rev., **97**, 1714 (1955); **98**, 1865 (1955).
36. *Macfarlane G. G., McLean T. P., Quarrington J. E., Roberts V.*, Phys. Rev., **108**, 1377 (1957).
37. *Macfarlane G. G., McLean T. P., Quarrington J. E., Roberts V.*, Phys. Rev., **111**, 1245 (1958).
38. *Hsieh Y. C.*, Chem. Phys., **22**, 306 (1954).
39. *Brockhouse B. N., Iyengar P. K.*, Phys. Rev., **108**, 894 (1957).
40. *Elliott R. J.*, Phys. Rev., **108**, 1384 (1957).
41. *Brockhouse B. N.*, Journ. Phys. Chem. Solids, **8**, 400 (1959).
42. *Johnson F. A.*, Proc. Phys. Soc., **73**, 265 (1959).
43. *Parkinson D. H.*, Electron Structure in Solids, ed. Haidemenakis E. D. Plenum Press, 1969, p. 93, p. 102.
44. *McLean T. P.*, Progress in Semiconductors, Heywood, **5**, 1960, p. 53.
45. *Bassani F.*, Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970.
46. *Greenaway D. L., Harbeke G.*, Optical Properties and Band Structure of Semiconductors, Pergamon Press, 1968.
47. *Moss T. S., Burrell G. J., Ellis B.*, Semiconductor Optic-electronics, Butterworths, 1973.
48. *Dunke W.*, Phys. Rev., **108**, 1419 (1957).
49. *Braunstein R.*, Phys. Rev., **125**, 475 (1962).

50. Loudon R., Proc. Phys. Soc., **80**, 952 (1962).
51. Басов Н. Г., Грасюк А. З., Зубарев И. Г., Катулин В. А., Крохин О. Н., ЖЭТФ, **50**, 551 (1966).
52. Yee J. H., Journ. Phys. Chem. Solids, **33**, 693 (1972).
53. Button K. J., Lax B., Weiler Margaret H., Reine M., Phys. Rev. Lett., **17**, 1005 (1966).
54. Lee C. C., Fan H. Y., Phys. Rev., **B9**, 3502 (1974).
55. Gibson A. F., Hatch C. B., Maggs P. N. D., Tilley D. R., Walker A. C., Journ. Phys., **C9**, 3259 (1976).
56. Dresselhaus G., Phys. Rev., **106**, 76 (1957).
57. Dimmock J. O., Semiconductors and Semimetals, Academic Press, **3**, 1967, p. 259.
58. Knox R. S., Theory of Excitons, Solid State Physics, Academic Press, suppl. **5**, 1963.
59. Altarelli M., Lipari N. O., Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 811.
60. Гросс Е. Ф., Nuovo Cim., **3**, 672 (1956).
61. Гросс Е. Ф., Захарченя Б. П., Доклады АН СССР, **90**, 745 (1953).
62. Nikitine S., Journ. Phys. Radium, **17**, 817 (1956).
63. Рогачев А. А., Труды IX Международной конференции по физике полупроводников.— Л.: Наука, 1969, с. 431.
64. Sturge M. D., Phys. Rev., **127**, 768 (1962).
65. McLean T. P., Loudon R., Journ. Phys. Chem. Solids, **13**, 1 (1960).
66. Гросс Е. Ф., Сафаров В. И., Титков А. Н., Шлимак И. С., Письма в ЖЭТФ, **13**, 332 (1971).
67. Shaklee K. L., Nabory R. E., Phys. Rev. Lett., **24**, 942 (1970).
68. Gershenzon M., Thomas D. G., Deitz R. E., Proc. Vth Intern. Conf. Phys. Semicond., Inst. Phys. & Phys. Soc., London, 1962, p. 752.
69. Dean P. J., Phys. Rev., **157**, 655 (1967).
70. Phillips J. C., Phys. Rev., **136A**, 1705 (1964).
71. Кононенко В. Л., Манделъштам Т. С., Мурзин В. Н., Вавилов В. С., Proc. XIIth Intern. Conf. Phys. Semicond., Teubner, 1974, p. 152.
72. Buchanan M., Timusk T., Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 821.
73. Burstein E., Egli P. H., Advances in Electronics and Electron Physics, Academic Press, **7**, 1955, p. 1.
74. Burstein E., Picus G. S., Sclar N., Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 353.
75. Kohn W., Solid State Physics, Academic Press, **5**, 1957, p. 257.
76. Hrostowski H. J., Kaiser R. H., Bull. Amer. Phys. Soc., **2**, 66 (1955).
77. Stoneham A. M., Theory of Defects in Solids, Clarendon Press, 1975. [Имеется перевод: Стоунхэм А. Теория дефектов в твердых телах. В 2-х т.— М.: Мир, 1978.]
78. Fletcher R. C., Yager W. A., Pearson G. L., Holden A. N., Read W. T., Merritt F. R., Phys. Rev., **94**, 1392 (1954).
79. Fletcher R. C., Yager W. A., Pearson G. L., Merritt F. R., Phys. Rev., **95**, 844 (1954).
80. Honig A., Kip A. F., Phys. Rev., **95**, 1686 (1954).
81. Honig A., Phys. Rev., **96**, 234 (1954).
82. Feher G., Phys. Rev., **103**, 834 (1956).
83. Lancaster G., Electron Spin Resonance in Semiconductors, Hilger and Watts, 1966.
84. Ludwig G. W., Woodbury H. H., Solid State Physics, Academic Press, **13**, 1962, p. 223.
85. Haas M., Semiconductors and Semimetals, Academic Press, **3**, 1967, p. 3.
86. Johnson F. A., Rendiconti della Scuola Internazionale di Fisica Enrico Fermi,

- Corso XXII, Academic Press, 1963, p. 504; Progress in Semiconductors, Heywood, 9, 1965, p. 179.
87. Johnson F. A., Loudon R., Proc. Roy. Soc., A281, 274 (1964).
 88. Cochran W., Adv. Phys., 9, 387 (1960).
 89. Cochran W., The Dynamics of Atoms in Crystals, Edward Arnold, 1973.
 90. Maradudin A. A., Montroll E. W., Weiss G. H., Solid State Physics, Academic Press, suppl. 3, 1963.
 91. Spitzer W. G., Semiconductors and Semimetals, Academic Press, 3, 1967, p. 17.
 92. Johnson F. A., Proc. Roy. Soc., A310, 79, 89, 101 (1969).
 93. Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 331.
 94. Houghton J., Smith S. D., Infra-red Physics, Clarendon Press, 1966.
 95. Newman R. C., Adv. Phys., 18, 545 (1968).
 96. Smith R. A., Jones F. E., Chasmar R. P., Detection and Measurement of Infra-red Radiation, Clarendon Press, 2nd edition, 1968.
 97. Moss T. S., Photo-conductivity in the Elements, Butterworth, 1952.
 98. Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956.
 99. Smith R. A., Adv. Phys., 2, 321 (1953); Physica, 20, 910 (1954).
 100. Morton G. A., Hahn E. E., Schultz M. L., Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 556.
 101. Optical Properties of Solids, ed. Nudelman S., Mitra S. S., Plenum Press, 1969.
 102. Rose A., Progress in Semiconductors, Heywood, 2, 1957, p. 109; Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 3.
 103. Herring S., Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 81.
 104. Garreta O., Grovalet J., Progress in Semiconductors, Heywood, 1, 1956, p. 165.
 105. De Vore H. B., Phys. Rev., 102, 86 (1956).
 106. Gärtner W., Phys. Rev., 105, 823 (1957).
 107. Goodwin D. W., Report of Meeting on Semiconductors, London Physical Society, 1956, p. 137.
 108. Jayaraman S., Lee C. H., Appl. Phys. Lett., 20, 392 (1972).
 109. Gibson A. F., Maggs P. N. D., Journ. Phys., D7, 292 (1974).
 110. Кикюит И. К., Носков М. М., Phys. Zs. Sowjet., 5, 586 (1934).
 111. Френкель Я., Phys. Zs. Sowjet., 8, 185 (1935).
 112. Moss T. S., Pincherle L., Woodward Alice M., Proc. Phys. Soc., B66, 743 (1953).
 113. Aigrain P., Bulliard H., C. R. Acad. Sci., Paris, 236, 595 (1953).
 114. Roosbroeck W. Van, Bull. Amer. Phys. Soc., 30, 10 (1955).
 115. Pincherle L., Proceedings of Atlantic City Photo-conductivity Conference, Chapman and Hall, John Wiley and Sons, 1956, p. 307.
 116. Stierwald D. L., Potter R. F., Semiconductors and Semimetals, Academic Press, 3, 1967, p. 71.
 117. Roosbroeck W. Van, Phys. Rev., 101, 1713 (1956).
 118. Данишевский А. М., Кастальский А. А., Рывкин С. М., Ярошецкий И. Д., ЖЭТФ, 58, 544 (1970).
 119. Gibson A. F., Kimmit M. F., Walker A. C., Appl. Phys. Lett., 17, 75 (1970).
 120. Gibson A. F., Walker A. C., Journ. Phys., C4, 2209 (1971).
 121. Patel C. K. N., Appl. Phys. Lett., 18, 25, 274 (1971).
 122. Cardona M., Semiconductors and Semimetals, Academic Press, 3, 1967, p. 125.
 123. Tauc J., Progress in Semiconductors, Heywood, 9, 1965, p. 87.
 124. Franz W., Zs. Naturforsch., 13a, 484 (1958).
 125. Келдыш Л. В., ЖЭТФ, 33, 994 (1957).
 126. Callaway J., Phys. Rev., 130, 549 (1963).
 127. Tharmalingam K., Phys. Rev., 130, 2204 (1963).

128. *Penchina C. M.*, Phys. Rev., **A138**, 924 (1965).
129. *Келдыш Л. В., Вавилов В. С., Брицын К. И.*, Proc. Vth Intern. Conf. Phys. Semicond., Czechoslovak Acad. Sci., 1961, p. 824.
130. *Moss T. S.*, Journ. Appl. Phys., **32**, 2136 (1961).
131. *Blossey D. F., Handler P.*, Semiconductors and Semimetals, Academic Press; **9**, 1972, p. 257.
132. *Aspnes D. E., Boltka N.*, Semiconductors and Semimetals, Academic Press, **9**, 1972, p. 457.
133. *Cardona M., Shacklee K. L., Pollak F. H.*, Phys. Rev., **154**, 696 (1967).
134. *Cardona M.*, Solis State Physics, Academic Press, suppl. 11, 1969. (Имеется перевод: Кардона М. Модуляционная спектроскопия.— М.: Мир, 1972.)
135. *Seraphin B. O.*, Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 213; Semiconductors and Semimetals, Academic Press, **9**, 1972, p. 1; Optical Properties of Solids, ed. Abeles F., North Holland, 1972, p. 163.
136. *Hamakawa Y., Handler P., Germano F. A.*, Phys. Rev., **167**, 709 (1968).
137. *Hamakawa Y., Nishimo T.*, Optical Properties of Solids—New Developments, ed. Seraphin B. O., North Holland, 1976, p. 225.
138. *Balslev I.*, Semiconductors and Semimetals, Academic Press, **9**, 1972, p. 403.
139. *Haynes J. R., Lax M., Flood W. F.*, Proc. 1Vth Intern. Conf. Phys. Semicond., Journ. Phys. Chem. Solids, **8**, 392 (1959).
140. *Haynes J. R.*, Phys. Rev. Lett., **4**, 361 (1960).
141. *Thomas D. G., Hopfield J. J.*, Phys. Rev., **128**, 2135 (1962).
142. *Reynolds D. C.*, Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969, p. 110.
143. *Thomas D. G., Gershenson M., Trumbore F. A.*, Phys. Rev., **A113**, 269 (1964).
144. *Dean P. J., Henry C. H., Frosch C. J.*, Phys. Rev., **A168**, 812 (1968).
145. *White A. M., Dean P. J., Taylor L. L., Clarke R. C., Ashen D. J., Mullin J. B.*, Journ. Phys., **C5**, 1727 (1972).
146. *Bebb H. B., Williams E. W.*, Semiconductors and Semimetals, Academic Press, **8**, 1972, p. 181.
147. *Phillips J. C.*, Phys. Rev., **B1**, 1545 (1970).
148. *Dean P. J.*, Progress in Solid State Chemistry, Pergamon Press, **8**, 1973, p. 1.
149. *Ульмер Э. А., Фрэнк Д. Р.*, Труды IX Международной конференции по физике полупроводников. Л.: Наука, 1969, с. 179.
150. *Williams E. W., Bebb H. B.*, Semiconductors and Semimetals, Academic Press, **8**, 1972, p. 321.

ЗОННАЯ СТРУКТУРА И МЕТОД ЭФФЕКТИВНОЙ МАССЫ

11.1. Концепция эффективной массы

В гл. 2 мы кратко обсудили вид зависимости энергии от волнового вектора \mathbf{k} , а также вид волновых функций электронов, соответствующих состояниям с заданным значением энергии E . Из сходства зависимости энергии электрона от \mathbf{k} вблизи экстремума зоны с зависимостью $E(\mathbf{k})$ для свободного электрона мы заключили, что свободные электроны и дырки в полупроводнике имеет смысл рассматривать как частицы, обладающие эффективной массой, которая в простейшем случае может быть скалярной, но в общем случае выражается тензорной величиной. Эти массы не должны зависеть от \mathbf{k} только вблизи экстремумов зон и лишь при определенных условиях. Использование концепции эффективной массы подкрепляется еще и тем, что уравнения движения свободных электронов и дырок в полупроводнике подобны уравнениям движения классических частиц. На самом деле концепция эффективной массы имеет, как мы увидим, значительно более глубокий смысл. На это указывает уже тот факт, что для определения энергии связи примесей с мелкими уровнями и экситонов, а также для вычисления вероятности рассеяния этих «частиц» (электронов) мы применяли *квантовую теорию*. Задним числом это кажется очень нелогичным, однако, как мы увидим, на самом деле такое рассмотрение оправдано.

Прежде чем подтвердить правильность такого подхода, необходимо кратко обсудить вопрос о том, каким образом можно получить вид зависимости $E(\mathbf{k})$, используя информацию о кристаллической структуре полупроводника. Этому вопросу было посвящено большое количество публикаций, и позже мы дадим ссылки на некоторые ключевые работы. Был развит целый ряд методов расчета, каждый из которых пригоден в определенных конкретных условиях, однако полного общего метода, применимого к любым условиям, нет. Можно сразу сказать, что, хотя определить общий вид зонной структуры какого-либо конкретного материала не слишком сложно, расчет полной структуры зоны проводимости и валентной зоны,

в общем случае состоящих из ряда перекрывающихся зон, является чрезвычайно трудной проблемой. Тем не менее с течением времени достигался все больший и больший прогресс, что было обусловлено не только усовершенствованием методов расчета, но и в значительной мере повышением производительности электронных цифровых вычислительных машин.

Здесь мы дадим лишь беглый обзор используемых принципов. Более полное обсуждение этого вопроса приведено автором в книге «Wave Mechanics of Crystalline Solids» [1], в гл. 6 и 11, а также в ряде книг, на которые даны ссылки в предисловии. Конкретные методы расчета очень детально описаны также в ряде обзорных статей, на которые мы будем ссылаться ниже.

11.2. Функции Блоха и Ванье

Для идеального кристалла волновая функция электрона в состоянии, соответствующем энергии $E(\mathbf{k})$, может быть записана в виде

$$\psi(\mathbf{r}) = U_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (11.1)$$

где функция $U_{\mathbf{k}}$ обладает периодичностью кристаллической решетки.

Функции $\psi(\mathbf{r})$, известные под названием блоховских функций, являются основой для любого расчета зонной структуры. Вычисление этих функций обсуждалось многими авторами, всесторонний анализ этого вопроса дал Займан [2]. Хотя такие волновые функции описывают свободное движение электронов (и могут также применяться для дырок), комбинация этих функций позволяет учесть локальные возмущения, обусловленные несовершенствами кристалла, полями сил и т. п. Использование плоских блоховских волн для вычислений зонной структуры не приводит к быстро сходящимся приближениям, поэтому, как мы увидим далее, были разработаны различные методы, позволяющие получить лучшие приближения.

Для случая сильно локализованных силовых полей, таких, например, как поле примесного центра, понадобилось бы большое число блоховских функций, поэтому здесь удобнее пользоваться другим набором волновых функций, которые можно получить из функций Блоха. Для каждой элементарной ячейки кристалла вводится функция, которая для j -й ячейки в случае одной энергетической зоны имеет вид

$$a_j(\mathbf{r}) = N^{-1/2} \sum U_{\mathbf{k}}(\mathbf{r} - \mathbf{R}_j) \exp[i\mathbf{k} \cdot (\mathbf{r} - \mathbf{R}_j)], \quad (11.2)$$

где N — число элементарных ячеек в кристалле и \mathbf{R}_j — радиус-

вектор j -й ячейки. Из дальнейшего будет видно, что поведение функции $a_j(\mathbf{r})$ в j -й ячейке не отличается от поведения функции $a_l(\mathbf{r})$ в l -й ячейке, поэтому ее можно записать как $a(\mathbf{r} - \mathbf{R}_j)$. Эти волновые функции ввел в теорию Ванье [3], и поэтому их обычно называют функциями Ванье.

11.3. Методы расчета зонной структуры

Для случая движения в поле одномерного периодического потенциала можно получить несколько точных решений (см. [1], § 2.9). Кроме того, использование методов теории возмущений позволяет получить решения и для случая *малого* трехмерного периодического потенциала (см. [1], § 6.1, 6.2). Хотя такие решения и полезны, поскольку позволяют выявить общий характер зонной структуры и, в частности, наличие запрещенной зоны, для расчетов зонной структуры реальных кристаллов они непригодны. В то же время приближение сильной связи (см. [1], § 6.3), которое является другим предельным случаем, может быть использовано, только когда электроны настолько сильно связаны с отдельными атомами кристалла, что существует лишь очень слабое перекрытие атомных волновых функций. Это приближение оказывается полезным только для определенного, очень специфического круга задач теории полупроводников.

11.3.1. МЕТОД ЯЧЕЕК

В этом методе кристалл разделяют на элементарные ячейки, центры которых совпадают с каждым атомом кристалла, а затем вычисляют приближенные волновые функции для ячейки, используя атомные потенциалы. Основная трудность заключается в выборе граничных условий на поверхности ячейки, поэтому здесь были развиты различные приближенные методы. Если $u(\mathbf{r}')$ — волновая функция, где \mathbf{r}' — вектор с началом в центре ячейки, то волновая функция в j -й ячейке запишется в виде, который имеет требуемую блоховскую форму:

$$\psi(\mathbf{r}) = u(\mathbf{r} - \mathbf{R}_j) \exp(i\mathbf{k} \cdot \mathbf{r}). \quad (11.3)$$

Метод ячеек, первоначально предложенный Вигнером и Зейтцем [4] для монокристаллов, был существенно развит Слэтером [5] и его сотрудниками. Этот метод, правда, в сильной степени видоизмененный, Белл и др. [6] с известным успехом применили и для расчета двухатомных кристаллов, таких, как PbTe.

11.3.2. ЛИНЕЙНАЯ КОМБИНАЦИЯ АТОМНЫХ ОРБИТАЛЕЙ (ЛКАО)

Метод ЛКАО является развитием приближения сильной связи, в котором учтено влияние целого ряда атомных волновых функций, а не только одной, как в выражении (11.3). Соответствующая волновая функция имеет вид

$$\psi(\mathbf{r}) = \sum_s \alpha_s u_s(\mathbf{r} - \mathbf{R}_j) \exp(i\mathbf{k} \cdot \mathbf{R}_j). \quad (11.4)$$

Коэффициенты α_s находят вариационным методом, минимизируя энергию, значения которой являются корнями так называемого «секулярного» детерминанта.

11.3.3. МЕТОДЫ ПЛОСКИХ ВОЛН

Так как бlochовские функции обладают периодичностью решетки, их можно представить как ряды Фурье (см. [1], § 4.8), используя векторы обратной решетки \mathbf{b}_n . Тогда волновая функция (11.1) может быть записана в виде суммы плоских волн:

$$\psi(\mathbf{r}) = \sum_n A_n \exp[i(\mathbf{k} + 2\pi\mathbf{b}_n) \cdot \mathbf{r}]. \quad (11.5)$$

Чтобы найти точное решение, потребовалось бы бесконечное число таких плоских волн, но приближенное решение можно получить, используя конечное число (s) волн. Основная трудность, возникающая при использовании этого метода, заключается в том, что для сходимости волновой функции в области ее быстрого изменения, т. е. вблизи атомного остова, число этих волн должно быть довольно большим.

Были предложены два способа преодоления этой трудности и соответственно два метода расчета, которые теперь наиболее часто используются для вычисления зонной структуры. Первый из них, метод присоединенных плоских волн (ППВ), использует метод ячеек Вигнера — Зейтца *внутри* сферы радиуса R , которая чуть меньше элементарной ячейки, и плоские волны за пределами этой области со сшиванием волновых функций на границе сферы. Поскольку метод ячеек позволяет вычислить быстро изменяющуюся волновую функцию вблизи атомного остова, то для обеспечения сходимости нужно лишь небольшое число плоских волн (см. [1], § 6.9). Этот метод был предложен Слэтером [7, 8], который в дальнейшем вместе со своими сотрудниками развил его и расширил. Он успешно использовался для расчетов зонной структуры, однако главным образом применительно к металлам. Сравнительно недавно Диммок [9] сделал обзор работ, а Лоукс [10] опубликовал монографию, посвященную этому методу расчета.

Второй метод, метод ортогонализированных плоских волн (ОПВ), был впервые предложен Херрингом [11], но в дальнейшем был развит большим числом других авторов. Волновые функции в виде плоских волн, которые используются для описания валентных электронов, вообще говоря, не ортогональны более сильно локализованным волновым функциям электронов атомного остова. Метод ОПВ заключается в том, что к плоским волнам добавляется линейная комбинация атомных волновых функций, чтобы сделать их ортогональными (см. [1], § 6.10). Таким способом снова достигается быстрая сходимость с малым числом плоских волн. Этот метод был успешно использован для вычисления зонной структуры ряда полупроводников. Развитие этого метода и его применение обсуждал Вудруф [12]. Обзор более новых работ дали Коллинз, Ювема и Стукель [13].

11.3.4. ПСЕВДОПОТЕНЦИАЛ

Метод расчета зонной структуры, который первоначально предложили Филлипс и Клейнман [14] и в дальнейшем развили Коэн и Хейне [15], по существу является следствием метода ОПВ. Используя ортогонализированные плоские волны, можно видоизменить (см. [1], § 6.11) потенциал кристалла так, чтобы эффективным образом устранить его быстро меняющуюся часть вблизи атомного остова. Выражение в виде ряда Фурье в этом случае требует лишь небольшого числа значений вектора обратной решетки \mathbf{b}_n , а коэффициенты Фурье можно легко вычислить. Они подбираются таким образом, чтобы получаемые величины энергии совпадали со значениями, известными из опыта. Этот метод наиболее часто используется для определения полной зонной структуры в тех случаях, когда известно несколько значений энергии. Установлено, что этим методом можно рассчитать зонную структуру моноатомных полупроводников, подобных Si и Ge, используя всего лишь три параметра. Более важным (поскольку двухатомных полупроводников существует значительно большее количество) является то, что для материалов, подобных GaAs, использование всего шести параметров дает по существу те же результаты, что и расчет методом ОПВ. По видимому, достаточно хорошие результаты можно получить таким образом и для других сложных полупроводников, не производя полного ОПВ-расчета ¹⁾.

Обширный обзор, посвященный развитию этого метода, сделал Хейне [18], а его применение для определения зонной структуры ряда полупроводников способом подгонки к экспериментальным данным обсудили Коэн и Хейне [19].

¹⁾ См., например, [16, 17].

11.3.5. **k**-**p**-МЕТОД

Мы упомянем еще лишь об одном методе, который используется для оценки зонной структуры, о так называемом **k**-**p**-методе. В нем используется дифференциальное уравнение для величин $U_k(\mathbf{r})$, а не для полных блоховских функций (11.1). Считая значение волнового вектора \mathbf{k} малой величиной, это уравнение можно решить методами теории возмущения и получить приближенные решения для области значений \mathbf{k} вблизи $\mathbf{k}=0$. Этот способ можно применить и для области \mathbf{k} вблизи какой-либо иной критической точки $\mathbf{k}=\mathbf{k}_m$. Энергия выражается в виде (см. [1], § 6.13)

$$E_n(\mathbf{k}) = E_{n0} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{m \neq n} \frac{(\mathbf{k} \cdot \mathbf{p}_{mn})(\mathbf{k} \cdot \mathbf{p}_{nm})}{E_{n0} - E_{m0}}, \quad (11.6)$$

где E_{n0} и E_{m0} — энергии на краях зон, а \mathbf{p}_{mn} — матричный элемент импульса для $\mathbf{k}=0$, который можно получить методом ячеек. Члены суммы дают большой вклад только при $E_{m0} \approx E_{n0}$. Этот метод наиболее полезен, если энергетический зазор между двумя зонами мал, когда нужно использовать лишь один член суммы. При соответствующем выборе координатных осей в этом случае можно записать

$$E(\mathbf{k}) = \frac{\hbar^2}{2} \sum_{r=1}^3 \frac{k_r^2}{m_r}, \quad (11.7)$$

где для верхней зоны

$$\frac{1}{m_r} = \frac{1}{m_0} + \frac{2|\mathbf{p}_{12}^r|^2}{m_0^2(E_1 - E_2)}, \quad (11.8)$$

а для нижней зоны

$$\frac{1}{m_r} = \frac{1}{m_0} - \frac{2|\mathbf{p}_{12}^r|^2}{m_0^2(E_1 - E_2)}. \quad (11.8a)$$

Если $|\mathbf{p}_{12}^r|$ одно и то же для всех значений r , то можно видеть, что верхняя зона имеет минимум при $\mathbf{k}=0$, а нижняя зона максимум, если $2|\mathbf{p}_{12}^r|^2 > m_0(E_1 - E_2)$.

Кейн [20] использовал этот метод для анализа зонной структуры Si и Ge вблизи $\mathbf{k}=0$, а также и для ряда соединений типа $A^{III}B^V$. Подробное рассмотрение этого метода дал Мак-Лин [21].

Как и метод псевдопотенциала, **k**-**p**-метод приводит к интерполяционной схеме вычислений, и в таком виде его широко использовали Кардона и Поллак [22].

Если энергетический зазор ΔE мал, выражение для E в форме

(11.7) применимо только для очень малых значений k . Кейн [20] использовал $k \cdot p$ -метод, чтобы получить выражение для отклонения формы зоны от параболичности в виде

$$E = -\frac{1}{2} \Delta E + \frac{\hbar^2 k^2}{2m_0} \pm \frac{1}{2} \left[\Delta E^2 + \frac{4\hbar^2 k^2 |p_{12}|^2}{m_0^2} \right]^{1/2}. \quad (11.9)$$

Для малых значений k это выражение сводится к формулам (11.8) и (11.8а). Формула (11.9) может быть также выражена через эффективную массу для зоны проводимости

$$E = -\frac{1}{2} \Delta E + \frac{\hbar^2 k^2}{2m_0} + \frac{1}{2} \Delta E \left[1 + \frac{2\hbar^2 k^2}{\Delta E} \left(\frac{1}{m_e} - \frac{1}{m_0} \right) \right]^{1/2}. \quad (11.9a)$$

Если $m_e \ll m_0$, выражение можно упростить

$$E = -\frac{1}{2} \Delta E + \frac{1}{2} \Delta E \left(1 + \frac{2\hbar^2 k^2}{m_e \Delta E} \right)^{1/2}. \quad (11.96)$$

Формула (11.96) показывает отклонение формы зоны проводимости от параболичности.

11.4. Вычисления зонной структуры

Как указывалось выше, к настоящему времени различными методами выполнено большое число расчетов для многих полупроводников. Имеется несколько обстоятельных обзоров этих работ, которые содержат многочисленную библиографию. Обзоры Каллауэя [23], Козна и Хейне [19] и Хермана [24] достаточно полно охватывают работы по полупроводникам. В настоящее время трудно сказать, какой из методов вычисления наиболее эффективен. Метод ППВ использовался главным образом для металлов, а метод ОПВ, по-видимому, предпочтительнее для полупроводников. Как указывал Херман [24], метод псевдопотенциала очень удобен для построения полной зонной структуры, когда вычисления «из первых принципов» уже дали значения энергии для некоторого числа критических точек в зоне Бриллюэна.

Пинчерл [25] показал, что, хотя численные значения различных энергий, характеризующих структуру зон, зависят от потенциалов атомных остовов, общая форма зон может быть получена из анализа свойств симметрии материала методами теории групп. Он провел также критическое сравнение различных методов и дал очень ясное и четкое описание их специальных областей применений [26]. Филлипс [17] обсудил относительные достоинства различных методов, а также их сходство с теорией межатомных связей.

11.5. Метод эффективной массы

Определяя зонную структуру полупроводника вблизи экстремума (обычно у дна зоны проводимости и потолка валентной зоны), мы можем получить элементы тензора обратной эффективной массы. Здесь мы прежде всего предположим, что эти элементы являются постоянными величинами. Это означает, что зависимость $E(\mathbf{k})$ является квадратичной функцией компонент вектора \mathbf{k} . Рассмотрим теперь, каким образом методами квантовой механики можно описать движение частиц с таким тензором обратной эффективной массы и при каких условиях это оправдано. По-видимому, первым этот вопрос обсудил Ванье [3], используя функции, которые называются теперь его именем и которые мы обсуждали выше (разд. 11.2). Запишем блоховские волновые функции, как и ранее, в виде

$$\psi_n(\mathbf{r}) = U_{nk}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}). \quad (11.10)$$

Здесь индекс n относится к n -й энергетической зоне. Соответствующие функции Ванье тогда определяются так же, как в разд. 11.2. Можно показать в таком случае, что при наличии в кристалле некоторого дополнительного потенциала $V(\mathbf{r})$ волновая функция может быть выражена в достаточно общем виде (см. [1], § 11.2.3):

$$\psi(\mathbf{r}) = \sum_{nj} A_n F_n(\mathbf{R}_j) a_n(\mathbf{r} - \mathbf{R}_j), \quad (11.11)$$

где величины $F_n(\mathbf{R}_j)$ находятся из ряда алгебраических разностных уравнений, которые включают в себя матричные элементы потенциала с функциями Ванье. В случае когда можно ограничить суммирование по n всего одной зоной (обычно $\hbar^2 k^2 / m_0 \ll \Delta E$), мы можем при определенных условиях (здесь оказывается существенным, чтобы потенциал $V(\mathbf{r})$ не изменялся заметно на длине постоянной решетки, т. е. чтобы он был медленно меняющимся) заменить эти *разностные* уравнения *дифференциальным* уравнением для функции $F(\mathbf{r})$. Это дифференциальное уравнение имеет неожиданно простой вид. Пусть $E(\mathbf{k})$ — известная функция энергии; тогда имеем

$$E(-i\nabla) F(\mathbf{r}) + V(\mathbf{r}) F(\mathbf{r}) = E F(\mathbf{r}), \quad (11.12)$$

где E — энергия стационарного состояния, описываемого волновой функцией

$$\psi = A U_0(\mathbf{r}) F(\mathbf{r}). \quad (11.13)$$

Здесь A — нормирующий множитель, а $U_0(\mathbf{r})$ — периодическая часть блоховской функции для $k=0$. Мы видим, таким образом, что $F(\mathbf{r})$ играет роль медленно меняющейся модулирующей функции, а изменения в пределах каждой ячейки кристалла описываются функцией $U_0(\mathbf{r})$.

Если имеем скалярную эффективную массу m_e , так что $E(k) = \hbar^2 k^2 / 2m_e$, то $F(\mathbf{r})$ удовлетворяет уравнению

$$\nabla^2 F(\mathbf{r}) + \frac{2m_e}{\hbar^2} [E - V(\mathbf{r})] = 0, \quad (11.14)$$

которое является уравнением Шредингера для частицы с эффективной массой m_e .

В более общем случае, когда обратная эффективная масса $1/m_{rs}$ является тензором, функция $F(\mathbf{r})$ удовлетворяет уравнению

$$\sum_{rs} \frac{1}{m_{rs}} \frac{\partial^2 F}{\partial x_r \partial x_s} + \frac{2}{\hbar^2} [E - V(\mathbf{r})] F = 0. \quad (11.15)$$

Это уравнение можно рассматривать как уравнение Шредингера для классической частицы, гамильтониан которой имеет вид

$$\frac{1}{2} \sum_{rs} \frac{1}{m_{rs}} p_r p_s. \quad (11.16)$$

Уравнение (11.14) дает обоснование предыдущим вычислениям энергии водородоподобного примесного центра. Уравнение (11.15) является его модификацией для случая эллипсоидальных поверхностей равной энергии. Это уравнение позволяет обосновать также использование квантовомеханического подхода при анализе рассеяния электронов и дырок в полупроводниках.

При наличии нескольких минимумов выражение для волновой функции необходимо изменить. Если минимум расположен при $\mathbf{k} = \mathbf{k}_0$, то соответствующая волновая функция дается формулой

$$\psi = A U_{\mathbf{k}_0} \exp(i\mathbf{k}_0 \cdot \mathbf{r}) F(\mathbf{r}). \quad (11.16)$$

Тогда волновая функция для случая M минимумов выражается в виде

$$\psi = \sum_{r=1}^M A_r F_r(\mathbf{r}) U_{\mathbf{k}_{0r}} \exp(i\mathbf{k}_{0r} \cdot \mathbf{r}), \quad (11.17)$$

где константы A_r находятся вариационным методом. Такие волновые функции использовал Кон [27], который довольно подробно обсуждал вопрос о применении метода эффективной массы для вычислений положения энергетических уровней водородоподобных доноров, хотя он использовал другую систему базисных функций. Для вырожденных зон, как, например, в случае валентной зоны Si или Ge, необходимо внести в анализ некоторые изменения. Этот вопрос обсуждали Кон [27], а также Мендельсон и Шульц [28].

Несколько другой подход к методу эффективной массы дал Мак-Лин [29], который для исследования этого вопроса применил $\mathbf{k} \cdot \mathbf{p}$ -метод. Он получил почти такие же конечные результаты, но использованный им подход обладает тем преимуществом, что позволяет рассмотреть некоторые более сложные случаи.

11.5.1. ЭНЕРГЕТИЧЕСКИЕ УРОВНИ ПРИМЕСЕЙ И ЭКСИТОНЫ

Применение метода эффективной массы и его распространение на случай примесей с более глубокими уровнями, для которых «медленно меняющийся» потенциал уже неприменим, подробно обсудил Стоунхэм [30]. Стоунхэм и Харкер [31] выполнили расчеты для различных центров, в том числе и таких, которые состоят более, чем из одного атома. Такие расчеты предусматривают применение электронных цифровых вычислительных машин и достаточно сложны, хотя и не в такой степени, как расчеты зонной структуры.

Стоунхэм [30] сделал обширный обзор экспериментальных данных по энергетическим уровням примесей в большом количестве полупроводников и сравнил их с расчетными значениями. Этот вопрос довольно подробно рассмотрел также Миллис [32]. Очень тщательное сопоставление экспериментальных и теоретических результатов для частного случая доноров в GaAs провели Феттерман и др. [33]. Самые последние расчеты дают очень хорошее согласие теории с экспериментом, за исключением некоторых примесей с глубокими уровнями в Si и Ge, для которых существуют еще не вполне объясненные расхождения. Как мы видели, для доноров с мелкими уровнями в Si и Ge согласие между теоретическими и экспериментальными результатами очень хорошее. Исключением является их основное состояние, для которого область действия сил, обусловленных «центральной частью», является слишком малой, чтобы применить простой метод эффективной массы. Ярош [34] сумел, однако, существенно сблизить результаты теории с опытными данными, учтя изменения величин эффективной массы и диэлектрической проницаемости вблизи «центрального остова».

Чтобы вычислить энергию связи экситона, необходимо видоизменить метод эффективной массы применительно к системе, состоящей из двух частиц. Этот вопрос мы уже обсуждали в разд. 10.6.

11.6. Квантование в магнитном поле

Важным применением метода эффективной массы является квантовомеханический анализ движения электронов или дырок в магнитном поле. Магнитное поле можно рассматривать как медленно меняющееся возмущение, наложенное на периодическое поле кристалла, что является благоприятной ситуацией для применения метода эффективной массы. Возникает, однако, некоторая трудность, связанная с тем, что мы не можем представить магнитное поле скалярным потенциалом, — здесь необходима векторная функция. Но, поскольку хорошо известно, каким образом следует видоизменить волновое уравнение Шредингера, чтобы учесть магнитное поле,

кажется разумным изменить волновое уравнение для случая эффективной массы таким же образом.

Как мы уже видели условием квантования в магнитном поле является требование достаточно сильного поля, при котором обеспечивается выполнение условия $\omega_c \tau > 1$, где ω_c , как и ранее, циклотронная частота, а τ — среднее время свободного пробега. Это условие можно выразить также, как $B\mu > 1$, где B — магнитная индукция, а μ — подвижность.

Волновое уравнение для свободного электрона в магнитном поле получается простой заменой оператора импульса $\hat{\mathbf{p}} = -i\hbar\nabla$ оператором $\hat{\mathbf{p}} - e\mathbf{A}$, где \mathbf{A} — вектор-потенциал для магнитного поля $\mathbf{B} = \nabla \times \mathbf{A}$. Для электрона со скалярной эффективной массой m_e уравнение для модулирующей функции $F(\mathbf{r})$, как следует ожидать, будет иметь вид

$$\frac{1}{2m_e} (i\hbar\nabla + e\mathbf{A})^2 F(\mathbf{r}) = EF(\mathbf{r}). \quad (11.18)$$

Эта эквивалентность, которую мы использовали интуитивно, может быть обоснована, как показал Кон [35], и формальным образом. Более общее уравнение, заменяющее уравнение (11.12), должно было бы иметь вид

$$E \left(-i\nabla - \frac{e\mathbf{A}}{\hbar} \right) F(\mathbf{r}) = EF(\mathbf{r}), \quad (11.19)$$

и, как показал Кон [35], обосновать его более трудно. Для эллипсоидальных поверхностей равной энергии должно выполняться равенство

$$\sum_{r=1}^3 \frac{1}{2m_r} \left(i\hbar \frac{\partial}{\partial x_r} + e\mathbf{A}_r \right)^2 F(\mathbf{r}) = EF(\mathbf{r}). \quad (11.20)$$

Если магнитное поле направлено по оси z , то вектор-потенциал \mathbf{A} может быть выражен как $(0, Bx, 0)$, и уравнение (11.18) приводится к виду

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial z^2} + \left(\frac{\partial}{\partial y} - \frac{ieBx}{\hbar} \right)^2 F + \frac{2m_e EF}{\hbar^2} = 0, \quad (11.21)$$

т. е. к уравнению, которое было решено Ландау [36] при рассмотрении квантовой теории движения свободных электронов в магнитном поле. Решение этого уравнения и его применение к полупроводникам мы обсудим в разд. 12.5.

ЛИТЕРАТУРА

1. *Smith R. A.*, The Wave Mechanics of Crystalline Solids, Chapman and Hall, 1969.
2. *Ziman J. M.*, Solid State Physics, Academic Press, 26, 1971, p. 1. [Имеется перевод: *Займан Дж.* Вычисление блоховских функций.— М.: Мир, 1973.]
3. *Wannier G.*, Phys. Rev., 52, 191 (1937).

4. Wigner E., Seitz F., Phys. Rev., 43, 804 (1933).
5. Slater J. C., Handb. Phys., 19, 1 (1956).
6. Bell Dorothy G., Hum D. M., Pincherle L., Sciana D. W., Woodward P. M., Proc. Roy. Soc., A217, 71 (1953).
7. Slater J. C., Phys. Rev., 92, 603 (1953).
8. Slater J. C., Quantum Theory of Molecules and Solids, McGraw-Hill, 1965, vol. 2.
9. Dimmock J. O., Solid State Physics, Academic Press, 26, 1971, p. 104.
10. Loucks T. L., The Augmented Plane Wave Method, Benjamin, 1967.
11. Herring C., Phys. Rev., 57, 1169 (1940).
12. Woodruffe T. O., Solid State Physics, Academic Press, 4, 1957, p. 367.
13. Collins T. C., Ewmeta R. N., Stukel D. J., Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 81.
14. Phillips J. C., Kleinman L., Phys. Rev., 116, 287 (1959).
15. Cohen M. H., Heine V., Phys. Rev., 122, 182 (1961).
16. Phillips J. C., Solid State Physics, Academic Press, 18, 1966, p. 55.
17. Phillips J. C., Bonds and Bands in Semiconductors, Academic Press, 1973.
18. Heine V., Solid State Physics, Academic Press, 24, 1970, p. 1.
19. Cohen M. L., Heine V., Solid State Physics, Academic Press, 24, 1970, p. 38.
20. Kane E. O., Journ. Phys. Chem. Solids, 1, 245 (1956); 8, 38 (1959); Semiconductors and Semimetals, Academic Press, 1, 1966, p. 75.
21. McLean T. P., Rendiconti della Scuola Internazionale di Fisica Enrico Fermi, Corso XXII, Academic Press, 1963, p. 483.
22. Cardona M., Pollak F. H., Phys. Rev., 142, 530 (1966).
23. Callaway J., Solid State Physics, Academic Press, 7, 1958, p. 100.
24. Herman F., Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969, p. 41.
25. Pincherle L., Rendiconti della Scuola Internazionale di Fisica Enrico Fermi, Corso XXII, Academic Press, 1963, p. 1.
26. Pincherle L., Electronic Energy Bands in Solids, Macdonald, 1971.
27. Kohn W., Solid State Physics, Academic Press, 5, 1957, p. 257.
28. Mendelson K., Schultz D. R., Phys. Stat. Sol., 31, 59 (1969).
29. McLean T. P., Rendiconti della Scuola Internazionale de Fisica Enrico Fermi, Corso XXII, Academic Press, 1963, p. 479.
30. Stoneham A. M., Theory of Defects in Solids, Clarendon Press, 1975. [Имеется перевод: Стоунхэм А. Теория дефектов в твердых телах. В 2-х т.— М.: Мир, 1978.]
31. Stoneham A. M., Harker A. H., Journ. Phys., C8, 1102, 1109 (1975).
32. Milnes A. G., Deep Impurities in Semiconductors, Wiley, 1973. [Имеется перевод: Милнс А. Примеси с глубокими уровнями в полупроводниках.— М.: Мир, 1977.]
33. Fetterman H. R., et al., Phys. Rev. Lett., 26, 975 (1971).
34. Jaros M., Journ. Phys., C4, 1162 (1971).
35. Kohn W., Phys. Rev., 115, 1460 (1959).
36. Ландау Л., Zs. Phys., 64, 629 (1930).

ВЛИЯНИЕ СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ И МАГНИТНЫХ ПОЛЕЙ НА ЯВЛЕНИЯ ПЕРЕНОСА И ОПТИЧЕСКИЕ СВОЙСТВА

12.1. Изменение функции распределения

При обсуждении явлений электронного переноса в полупроводниках в гл. 5 и 6 мы негласно предполагали, что функция распределения, характеризующая заселенность разрешенных энергетических состояний электронами (или дырками), не сильно отличается от равновесной функции распределения $f_0(\mathbf{k})$ независимо от того, задается ли последняя «классическим» выражением для невырожденного случая или функцией Ферми — Дирака. То, что функция распределения должна быть немного изменена, следовало, как мы видели, из необходимости получить электрический и тепловые токи, однако это изменение сводилось к наложению малых дрейфовых скоростей на равновесную в основном функцию распределения. Это соответствует замене функции распределения $f(\mathbf{k})$ в малых членах уравнения Больцмана функцией распределения $f_0(\mathbf{k})$ (см. [1], § 10.5) и весьма похоже на влияние слабого ветра на распределение молекул по скоростям в воздухе. Малое магнитное поле создает некоторые «завихрения», но тоже существенно не нарушает равновесного распределения. В настоящей главе мы рассмотрим влияние достаточно сильных электрических и магнитных полей, для которых это условие уже не выполняется.

Прежде всего мы рассмотрим влияние электрического поля в отсутствие магнитного поля. Далее мы увидим, что при увеличении поля средняя энергия электронов увеличивается настолько, что они перестают быть в равновесии с решеткой кристалла. Сначала это отклонение можно описать с помощью «эффективной температуры» T_e электронов, которая в общем случае выше, чем температура решетки T , но с функцией распределения того же вида и с таким же дрейфовым членом. При дальнейшем увеличении поля вид функции распределения начинает отклоняться от равновесного за счет увеличения числа быстрых электронов. Если их энергия достаточна для генерации электронно-дырочных пар посредством ударной ионизации, то начинается процесс сильного возрастания концент-

рации избыточных носителей, который приводит к возникновению лавины или пробую. Как мы увидим, в зоне проводимости электроны могут возбуждаться из самого низкого минимума зоны в более высокий минимум, что может привести к очень существенным результатам.

Изменение функции распределения сильным магнитным полем в отсутствие электрического поля в некоторых отношениях может быть более значительным. Как было показано в подразд. 10.4.2, явления, связанные с квантованием в магнитном поле, ярко выражены, когда $B\mu \gg 1$ и, как мы увидим, оно резко изменяет *равновесную* функцию распределения $f_0(\mathbf{k})$, поскольку сильно изменяет спектр разрешенных энергетических уровней. Следует отметить различие между этой ситуацией и той, которая вызывается сильным электрическим полем. В последнем случае может оказаться возможным *стационарное* состояние (если нет лавинного процесса), но в в сильном поле оно далеко от равновесного. В случае сильного магнитного поля, однако, в отсутствие электрического поля может установиться равновесное состояние, поскольку в систему не вводится энергия. Как мы увидим, в этом случае новая функция распределения $f_0(\mathbf{k})$ изменяет характеристики электронного переноса даже при малых электрических полях и температурных градиентах и сильно влияет на оптические свойства. Теперь мы должны рассмотреть, как именно происходят эти изменения в функции распределения $f(\mathbf{k})$.

12.2. Обмен энергией между электронами и решеткой

Рассматривая процессы рассеяния электронов примесями и кристаллической решеткой, мы предположили, что такое рассеяние является почти упругим и что за единичный акт рассеяния происходит обмен лишь малыми порциями энергии. Мы видели, что в случае рассеяния на колебаниях решетки электрон получает или теряет порцию энергии величиной $\hbar\omega_s$, где $\omega_s/2\pi$ — частота колебания решетки, вызывающего это рассеяние. Если бы вероятности процессов рассеяния с потерей и приобретением энергии были одинаковы, то результирующий обмен энергии с решеткой должен был бы отсутствовать. На то, что это не так, указывает тот факт, что нагревание кристалла при протекании тока обусловлено электронами, которые под действием электрического поля в процессе рассеяния отдают решетке большее количество энергии, чем получают от нее. В стационарном состоянии энергия, которую получают электроны (или дырки) в электрическом поле, в точности равна потерям энергии на решетке.

Если средняя энергия электронов существенно возрастет под действием электрического поля, то это повлияет на подвижность и поэтому плотность тока уже не будет пропорциональна величине поля. Подвижность тогда следует рассматривать как функцию электрического поля. Если плотность тока j не пропорциональна электрическому полю \mathcal{E} , закон Ома нарушается. Такие отклонения от закона Ома, как мы видели, характерны для полупроводников, но обычно они обусловлены поверхностными эффектами или неоднородностями, например различного рода контактами или переходами. Эффект, о котором идет речь сейчас, является объемным эффектом и может иметь место в совершенно однородном полупроводнике.

Шокли [2] показал, что такого рода отклонения от закона Ома намного легче получить в полупроводниках, чем в металлах. В полупроводниках изменение средней энергии на величину порядка $kT(1/\epsilon_0)$ эВ при комнатной температуре) будет означать большое изменение средней энергии, которая в равновесном случае составляет $3/2 kT$, в то время как в металлах такое изменение будет трудно заметить, поскольку здесь средняя энергия электронов составляет несколько электрон-вольт.

Теперь мы должны более детально проанализировать изменение энергии электронов, обусловленное электрическим полем и рассеянием. Ограничимся на время только решеточным рассеянием, поскольку рассеяние на примесях является в сильной степени упругим процессом и в этой связи оказывается важным только при очень низких температурах. Прирост энергии δE при рассеянии с испусканием или поглощением фонона приближенно дается выражением (8.18). Для наших целей, однако, это выражение оказывается недостаточно точным, и мы должны получить выражения для случаев испускания и поглощения фононов с учетом членов с (u/v) во второй степени, где u — скорость звука в кристалле, а v — скорость электрона перед рассеянием. Пусть E_p — энергия испущенного или поглощенного фонона. Тогда вычисленная величина угла, на который рассеивается электрон, оказывается одинаковой для процессов с испусканием и поглощением фонона, если учитываются лишь члены, содержащие (u/v) в первой степени. Пусть θ' — угол рассеяния в случае поглощения фонона, а θ — угол рассеяния в случае испускания. Тогда из соотношений (8.9) и (8.17), пренебрегая членами, содержащими (u/v) в степени выше второй, можно получить

$$E_p = 2m_e v^2 \left(\frac{u}{v}\right) \sin \frac{1}{2} \theta' \left[1 + \left(\frac{u}{v}\right) \sin \frac{1}{2} \theta'\right] \quad (12.1)$$

и

$$E_p = 2m_e v^2 \left(\frac{u}{v}\right) \sin \frac{1}{2} \theta \left[1 - \left(\frac{u}{v}\right) \sin \frac{1}{2} \theta\right]. \quad (12.2)$$

Формулы (12.1) и (12.2) представляют собой лучшее приближение для $\delta E (=E_p)$, чем выражение (8.18).

Чтобы оценить среднюю величину прироста энергии при рассеянии, нам нужно знать относительные вероятности рассеяния с испусканием и поглощением фонона. Соотношение этих вероятностей, равное $(N_p + 1)/N_p$, где

$$N_p = \frac{1}{\exp(E_p/kT) - 1}, \quad (12.3)$$

является хорошо известным результатом¹⁾ квантовой теории. Нормируя вероятности p_e, p_a для испускания и поглощения так, чтобы их сумма равнялась единице, для случая $E_p \ll kT$ имеем

$$p_e = \frac{1}{2} \left(1 + \frac{E_p}{2kT} \right), \quad (12.4)$$

$$p_a = \frac{1}{2} \left(1 - \frac{E_p}{2kT} \right). \quad (12.5)$$

Вклад в потери энергии, обусловленные излучением фонона при рассеянии электрона на угол θ в телесном угле $d\omega$, равен

$$\frac{E_p}{2} \left(1 + \frac{E_p}{2kT} \right) \frac{d\omega}{4\pi}. \quad (12.6)$$

Чтобы получить подобное выражение для рассеяния на угол θ с поглощением, мы должны сначала ввести поправку, связанную с изменением угла, т. е. должны заменить E_p на $E_p \sin(\frac{1}{2}\theta)/\sin(\frac{1}{2}\theta')$ и $d\omega$ на $\sin\theta d\theta d\omega/\sin\theta' d\theta'$.

Средняя величина прироста энергии $\overline{\delta E}(\theta)$ тогда дается выражением

$$\overline{\delta E}(\theta) d\omega = \frac{d\omega}{8\pi} \left[-E_p \left(1 + \frac{E_p}{2kT} \right) + E_p \left(1 - \frac{E_p}{2kT} \right) \frac{\sin^{1/2}\theta \sin\theta d\theta}{\sin^{1/2}\theta' \sin\theta' d\theta'} \right]. \quad (12.7)$$

Из формул (12.1), (12.2) легко видеть, что

$$\sin \frac{1}{2} \theta' = \sin \frac{1}{2} \theta \left[1 - 2 \left(\frac{u}{v} \right) \sin \frac{1}{2} \theta \right], \quad (12.8)$$

так что

$$\sin \frac{1}{2} \theta' \sin \theta' d\theta' = \sin \frac{1}{2} \theta \sin \theta d\theta \left[1 - 8 \left(\frac{u}{v} \right) \sin \frac{1}{2} \theta \right]. \quad (12.9)$$

Подставляя эти выражения в формулу (12.7), а также заменяя E_p его выражением через θ и пренебрегая членами со степенями (u/v)

¹⁾ Он получается точно таким же способом, как и хорошо известный результат для фотонов, который выведен в разд. 14.6.

выше второй, имеем

$$\overline{\delta E}(\theta) d\omega = 8m_e u^2 \left(1 - \frac{m_e v^2}{4kT}\right) \sin^2 \frac{1}{2} \theta \left(\frac{d\omega}{4\pi}\right). \quad (12.10)$$

Усредняя по всем углам, получаем формулу для средней величины прироста энергии при рассеянии

$$\overline{\delta E} = 4m_e u^2 \left(1 - \frac{E}{2kT}\right), \quad (12.11)$$

где $E = \frac{1}{2} m_e v^2$ — кинетическая энергия электрона перед рассеянием. Отметим, что если $E > 2kT$, то электрон будет в среднем терять энергию.

Предположим теперь, что электроны имеют максвелловское распределение скоростей, соответствующее температуре T_e , так что их число $N(v)dv$ со скоростями в интервале от v до $(v+dv)$ дается формулой

$$N(v)dv = A \exp\left(-\frac{m_e v^2}{2kT_e}\right) v^2 dv, \quad (12.12)$$

где постоянная A является нормирующим множителем. При $T_e = T$ это соответствует равновесному распределению в отсутствие вырождения. Формулу (12.11) теперь можно использовать для усреднения прироста энергии δE по всем значениям скоростей v . Величиной, представляющей наибольший интерес, является, однако, скорость прироста энергии электронов в результате соударений. Для электрона, имеющего скорость v , средняя частота столкновений равна $\tau^{-1} = v/l$, где l — средняя длина свободного пробега, обусловленная рассеянием на колебаниях решетки. Средняя скорость изменения энергии одного электрона равна $v\overline{\delta E}/l$, а для всех электронов $(dE/dt)_s$ дается выражением

$$\begin{aligned} \left(\frac{dE}{dt}\right)_s &= \frac{4m_e u^2 l^{-1} \int_0^\infty v^3 \left(1 - \frac{m_e v^2}{4kT}\right) \exp\left(-\frac{m_e v^2}{2kT_e}\right) dv}{\int_0^\infty v^2 \exp\left(-\frac{m_e v^2}{2kT_e}\right) dv} = \\ &= \frac{8u^2 (2\pi m_e k T_e)^{1/2} (T - T_e)}{\pi l T}. \end{aligned} \quad (12.13)$$

Можно видеть, что при тепловом равновесии, когда $T = T_e$, $(dE/dt)_s$ равно нулю, т. е. в среднем энергия электронов не изменяется в результате столкновений.

Если, однако, приложено электрическое поле \mathcal{E} , то электроны приобретают энергию со скоростью $(dE/dt)_f$, которая определяется как

$$\left(\frac{dE}{dt}\right)_f = \text{сила} \times \text{средняя скорость}.$$

Если мы определим подвижность μ в поле напряженностью \mathcal{E} с помощью соотношения

$$\text{средняя скорость} = \mu \mathcal{E},$$

то получим

$$\left(\frac{dE}{dt} \right)_t = e \mu \mathcal{E}^2. \quad (12.14)$$

Для стационарного случая мы должны иметь

$$\left(\frac{dE}{dt} \right)_t + \left(\frac{dE}{dt} \right)_s = 0. \quad (12.15)$$

С помощью этого уравнения можно получить выражение для эффективной температуры T_e как функции электрического поля \mathcal{E} .

Чтобы выразить μ как функцию новой средней скорости или эффективной температуры T_e , заметим, что время релаксации по-прежнему дается формулой $\tau = l/v$. Если мы выполним усреднение по скоростям таким же образом, как в разд. 5.3, то вместо выражения (5.98) получим

$$\mu = \frac{4el}{3(2\pi m_e k T_e)^{1/2}} = \mu_0 \left(\frac{T}{T_e} \right)^{1/2}, \quad (12.16)$$

где μ_0 — подвижность в слабом поле. Подвижность μ , следовательно, имеет вид

$$\mu = \frac{B}{T T_e^{1/2}}, \quad (12.17)$$

где B та же константа, что и в формуле (5.99). Подставляя значения $(dE/dt)_s$ и $(dE/dt)_t$ в уравнение (12.15), получаем квадратное уравнение для эффективной температуры T_e электронов в виде

$$\left(\frac{T_e}{T} \right)^2 - \left(\frac{T_e}{T} \right) = \frac{3\pi}{32} \left(\frac{\mu_0 \mathcal{E}}{u} \right)^2, \quad (12.18)$$

который получил Шокли [2].

Когда дрейфовая скорость $\mu_0 \mathcal{E}$ намного меньше скорости звука u , мы видим, что T_e почти равно T , т. е. температуре, соответствующей колебаниям решетки. Для электронов в Ge с $\mu_0 = 0,39 \text{ м}^2 \cdot \text{В}^{-1} \times \text{с}^{-1}$ при $T = 300 \text{ К}$ и $u = 5,4 \cdot 10^3 \text{ м} \cdot \text{с}^{-1}$ имеем $\mathcal{E} \mu_0 \approx u$, если $\mathcal{E} = 10^4 \text{ В} \cdot \text{м}^{-1}$. Можно ожидать, следовательно, что отклонения от линейного соотношения между \mathcal{E} и $(\mu \mathcal{E})$ возникнут при довольно малых значениях поля. Решив уравнение (12.18), получим, что

$$\frac{T_e}{T} = \frac{1}{2} \left\{ 1 + \left[1 + \frac{3\pi}{8} \left(\frac{\mu_0 \mathcal{E}}{u} \right)^2 \right]^{1/2} \right\}. \quad (12.19)$$

При $\mathcal{E} \ll u/\mu_0$ формулу (12.19) можно записать в приближенном виде

$$T_e = T \left[1 + \frac{3\pi}{32} \left(\frac{\mu_0 \mathcal{E}}{u} \right)^2 \right]. \quad (12.20)$$

Из (12.20) следует, что $T_e > T$ и $T_e - T$ пропорционально \mathcal{E}^2 . Таким образом, мы видим, что эффективная температура электронов оказывается несколько выше равновесной и что это эффект второго порядка, поэтому обычное предположение о равенстве T_e и T оправдано. Подставляя значение T_e , полученное из формулы (12.16), в формулу (12.20), в том же приближении имеем

$$\mu = \mu_0 (1 - \alpha \mathcal{E}^2), \quad (12.21)$$

где $\alpha = 3\pi^2 \mu_0^2 / 64 u^2$. Таким образом, мы видим, что $(\mu - \mu_0)$ также пропорционально \mathcal{E}^2 и что при $\mathcal{E} \ll \mu_0 / u$ предположение о том, что μ не зависит от \mathcal{E} и равно μ_0 , является хорошим приближением.

Квадратичную зависимость отклонения μ от постоянной величины при возрастании \mathcal{E} от очень малых значений экспериментально наблюдали в Ge *n*-типа Артур, Гибсон и Грэнвилл [3], однако найденная ими величина α оказалась меньше, чем следует из формулы (12.21). Приведенный выше анализ основан на модели, в которой эффективная масса предполагается скалярной, тем не менее считается, что использование тензорной эффективной массы, как требуется для Ge, при обычных температурах не сильно повлияло бы на величину средней скорости обмена энергией между электронами и решеткой¹⁾. Мы должны, следовательно, предположить, что существует какой-то другой механизм рассеяния, который ускоряет обмен энергией. Этим механизмом вряд ли может быть рассеяние на примесях, поскольку оно оказывается в сильной степени упругим. Его влияние сводится главным образом к изменению зависимости τ от v , что изменяет выражение для $(dE/dt)_s$. Ганн [5] рассмотрел не только случай рассеяния на ионизированной примеси, но и случай смешанного рассеяния, в котором участвуют и ионы примеси, и решетка. В случае рассеяния только на ионах примеси значение α в формуле (12.21) отрицательно, так что подвижность при возрастании \mathcal{E} сначала должна расти, однако анализ экспериментальных условий показывает, что малые наблюдаемые величины α не могут быть обусловлены случайной компенсацией влияния решеточного и примесного рассеяния. Как показывает намного более тщательный анализ процессов рассеяния, это расхождение возникает, по всей видимости, из-за того, что не учитывается рассеяние оптическими и более высокочастотными акустическими модами колебаний. В книге «Кинетические свойства полупроводников в сильных электрических полях» [6], которая целиком посвящена этому вопросу, Эстер М. Конуэлл очень подробно исследовала эти процессы, распространив рассмотрение всех видов рассеяния, которые мы обсуждали в гл. 8, на сильные электрические поля. Она же подробно обсудила применение этих результатов к Ge и Si, а также к другим полупроводникам. Как мы

¹⁾ См., например, [4].

увидим позже, в очень сильных полях важным является влияние более высоких долин, однако оно не определяет значение α , которое дает изменение μ в умеренных полях. С несколько иной точки зрения этот вопрос рассмотрел Пейдж [7].

12.3. Горячие электроны

При дальнейшем увеличении \mathcal{E} следует ожидать значительного изменения μ относительно постоянного значения. При $\mathcal{E} \gg \mu_0/u$ формулу (12.19) можно записать в следующем приближенном виде:

$$\frac{T_e}{T} = \left(\frac{3\pi}{32}\right)^{1/2} \frac{\mu_0 \mathcal{E}}{u}. \quad (12.22)$$

Подставляя это значение в (12.16), получаем

$$\mu = \left(\frac{32}{3\pi}\right)^{1/4} (\mu_0 u)^{1/2} \mathcal{E}^{-1/2}. \quad (12.23)$$

Величина средней дрейфовой скорости $v_d = \mu \mathcal{E}$ определяется формулой

$$v_d = \left(\frac{32}{3\pi}\right)^{1/4} (\mu_0 u \mathcal{E})^{1/2}, \quad (12.24)$$

которая показывает, что v_d возрастает пропорционально $\mathcal{E}^{1/2}$, а не пропорционально \mathcal{E} , как это было при выполнении закона Ома.

Формула (12.24) неприменима для очень больших значений \mathcal{E} , так как при достаточно высоких значениях \mathcal{E} начинают играть роль другие эффекты, которые приводят к изменению этой зависимости. До сих пор мы не рассматривали рассеяние на оптических модах колебаний решетки. Для их возбуждения требуется энергия $\hbar\omega_0$, где $\omega_0/2\pi$ — частота, соответствующая оптическому колебанию с $K=0$, поэтому если величина v_d такова, что $1/2 m_e v_d^2$ сравнима с $\hbar\omega_0$, электроны смогут возбуждать оптические колебания решетки. В этом случае электрон отдает основную часть своей кинетической энергии, поэтому такой процесс будет существенно неупругим. Если мы предположим, что тепловое возбуждение оптических колебаний настолько слабо, что вероятность поглощения фонона мала по сравнению с вероятностью испускания, то увидим, что, когда энергия электрона приближается к величине $\hbar\omega_0$, его скорость, очевидно, должна уменьшаться до малого значения благодаря испусканию оптического фонона. Если τ_0 — время релаксации для этого процесса, то, пренебрегая поглощением оптических фононов, имеем

$$\left(\frac{dE}{dt}\right)_s = -\frac{\hbar\omega_0}{\tau_0}. \quad (12.25)$$

Кроме того,

$$\left(\frac{dE}{dt}\right)_f = \frac{e^2 \tau_0 \mathcal{E}^2}{m_e}, \quad (12.26)$$

и, подставляя эти величины в уравнение (12.15), получаем

$$\tau_0 = \frac{(\hbar \omega_0 m_e)^{1/2}}{e \mathcal{E}}. \quad (12.27)$$

Отсюда можно вычислить дрейфовую скорость в условиях насыщения

$$v_s = \frac{e \tau_0 \mathcal{E}}{m_e} = \left(\frac{\hbar \omega_0}{m_e}\right)^{1/2}. \quad (12.28)$$

Таким образом, плотность тока, текущего через образец, будет насыщаться при величине j_s , равной

$$j_s = ne v_s = ne \left(\frac{\hbar \omega_0}{m_e}\right)^{1/2}. \quad (12.29)$$

Теоретически этот эффект предсказал Шокли [2]. Экспериментально его наблюдал в Ge Райдер [8], а позже Ганн [9], который использовал несколько более совершенную экспериментальную технику. Ганн нашел, что плотность тока является константой для полей в пределах $(4,5 - 9) \times 10^5 \text{ В} \cdot \text{м}^{-1}$, что соответствует значению $v_s = 5,2 \cdot 10^4 \text{ м} \cdot \text{с}^{-1}$. Значения, которые получили Райдер [8], а также Артур, Гибсон и Грэнвилл [3], составили соответственно $6,2 \cdot 10^4 \text{ м} \cdot \text{с}^{-1}$ и $6,5 \cdot 10^4 \text{ м} \cdot \text{с}^{-1}$. Формула (12.28) дает $v_s = 1,7 \cdot 10^5 \text{ м} \cdot \text{с}^{-1}$, если принять $m_e = m_0/5$ и $\hbar \omega_0 = kT_0$, где $T_0 = 550 \text{ К}$. Таким образом, согласие между результатами эксперимента и простой теорией рассеяния на оптических колебаниях оказывается несколько лучшим, чем для случая слабого поля, но теоретическое значение v_s все же приблизительно вдвое выше экспериментального. Это, по-видимому, нельзя считать неожиданным, учитывая те упрощающие приближения, которые мы сделали при анализе.

Полученная Ганном форма кривой, показывающая изменение дрейфовой скорости v_d в зависимости от напряженности электрического поля \mathcal{E} , представлена на рис. 12.1. Можно видеть, что для значений \mathcal{E} , превышающих $9 \cdot 10^5 \text{ В} \cdot \text{м}^{-1}$, дрейфовая скорость начинает снова возрастать с ростом \mathcal{E} . Это означает, что некоторые электроны достигают скоростей, более высоких, чем необходимо для возбуждения оптического фонона. Этого в принципе и следовало ожидать, так как, несмотря на то что вероятность рассеяния с испусканием оптического фонона при $\frac{1}{2} m_e v^2 > \hbar \omega_0$ высока, она не равна в точности единице и существует малая вероятность того, что некоторые электроны в сильных полях достигнут более высоких энергий. Хорошим экспериментальным доказательством того, что дело обстоит именно так, является наблюдение излучения у некоторых типов $p-n$ -переходов [10, 11], в которых существуют сильные электрические поля порядка $10^7 \text{ В} \cdot \text{м}^{-1}$. Предполагается, что

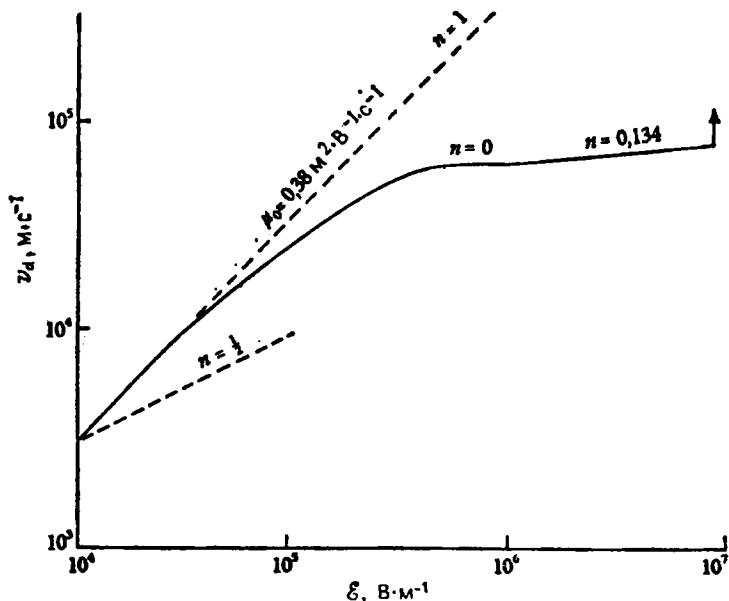


Рис. 12.1. Зависимость дрейфовой скорости v_d от напряженности электрического поля \mathcal{E} [9].

Число n показывает изменение вида зависимости $v_d \sim \mathcal{E}^n$.

это излучение возникает в результате рекомбинации быстрых электронов с дырками. Получено подтверждение о наличии излучения быстрых электронов из таких переходов и в Si [12].

Используя численный метод (метод Монте-Карло) для решения уравнения Больцмана (см. разд. 12.4), Фосетт и Пейдж [13] выполнили более тщательный расчет изменения подвижности в зависимости от напряженности электрического поля и нашли для комнатной температуры зависимость, которая очень сходна с изображенной на рис. 12.1. Как мы увидим ниже, наличие более высоких долин в направлениях $\langle 100 \rangle$ изменяет форму кривой при более низких температурах.

Если средняя скорость электронов существенно превышает тепловую скорость, соответствующую температуре решетки T , то их называют «горячими» электронами. В этом случае, как показали Фосетт и Пейдж [13], распределение электронов по скоростям сильно отличается от распределения Максвелла. Если распределение является почти максвелловским, но с повышенной температурой T_e , то электроны называют «теплыми».

Из формы кривой, показанной на рис. 12.1, видно, что при напряженности поля около $8 \cdot 10^5 \text{ В} \cdot \text{м}^{-1}$ наступает внезапное и резкое

увеличение v_d , которое наблюдается как скачкообразное возрастание тока. Оно свидетельствует о возникновении новых явлений и обусловлено резким ростом числа носителей заряда, которое до сих пор мы считали постоянным. Это возрастание вызывается ионизацией атомов кристалла и лишней раз доказывает, что некоторая часть электронов имеет энергии, существенно превышающие $\hbar\omega_0$. Так как этот процесс имеет лавинный характер, то для его возникновения необходимо лишь незначительное количество таких электронов, потому что каждый из них дает ряд вторичных, которые могут снова ускориться полем и дать еще большее количество вторичных электронов. Минимальная энергия, которой должен обладать электрон, чтобы вызвать рождение электронно-дырочной пары, несколько превышает энергию запрещенной зоны ΔE . Это следует из законов сохранения импульса и энергии. При равенстве масс электронов и дырок минимальное значение этой энергии равно $3/2 \Delta E$, что для Ge составляет величину примерно 1 эВ и намного превышает $\hbar\omega_0$.

Исчерпывающий обзор работ, посвященных взаимному обмену энергиями между электронами и решеткой, в частности для Ge и Si, сделал Пейдж [7], однако этот обзор не содержит ссылок на более поздние работы, указанные выше.

Резкий рост числа горячих электронов имеет место, когда рассеяния на оптических колебаниях недостаточно для передачи решетке энергии, приобретаемой электроном в электрическом поле, и тогда наступает состояние пробоя. Теоретически этот процесс рассмотрели Фрѐлих и Паранджап [14]. Некоторые его аспекты обсуждал также Ямашита [15]. Обзор работ по горячим электронам сделал Пѣтцль [16]. Для описания измененного распределения Фрѐлих и Паранджап ввели «смещенную» функцию Максвелла в виде

$$f(\mathbf{k}) = A \exp\left(-\frac{\hbar^2 |\mathbf{k} - \mathbf{q}|^2}{2m_e kT}\right), \quad (12.30)$$

где \mathbf{q} — вектор, параллельный направлению электрического поля, величина которого равна $m_e u_d \mathcal{E} / \hbar$, так что дрейфовая скорость выражается как $\hbar \mathbf{q} / m_e$. Считая рассеяние на оптических колебаниях основным механизмом передачи энергии решетке, они показали, что при определенной напряженности поля, величину которой они смогли вычислить, наступает пробой. Для GaAs эта величина составила, например, около $3,5 \cdot 10^5$ В·м⁻¹, что несколько ниже, чем для Ge из-за меньшей величины m_e . В этом материале, однако, еще до наступления пробоя возникают новые эффекты.

Интересно отметить, что если в функции распределения (12.30) ограничиться первым порядком по \mathcal{E} , то в точности получим функцию, которую мы использовали в разд. 7.3 [выражение (7.23)].

12.4. Эффекты междолинного переброса электронов

Если эффективная температура электронов в абсолютном минимуме зоны проводимости существенно повышается, то некоторые из них будут перебрасываться в более высокий минимум. Если предположить, что можно ввести эффективную температуру T_e , то критерием быстрого начала такого процесса будет приближение величины kT_e к $\Delta E_1 - \Delta E$, где ΔE_1 — ширина запрещенной зоны в области следующего минимума. Для полупроводников с малой величиной $\Delta E_1 - \Delta E$ такой переброс будет иметь место в относительно низких электрических полях, обычно еще до наступления лавинного процесса. В GaAs, например, следующий минимум расположен всего лишь на 0,35 эВ выше основного минимума при $k=0$. Этот эффект известен как междолинный переброс электронов. Как и ожидалось, он сильно выражен в арсениде галлия, где он оказывает заметное влияние на зависимость подвижности от электрического поля. В GaAs, который является кристаллом с частично ионной связью, рассеяние на полярных оптических колебаниях преобладает при высоких температурах и, в частности, при эффективных электронных температурах, обусловленных сильным электрическим полем. Эти температуры, как мы увидим, могут достигать нескольких тысяч градусов Кельвина при поле порядка 10^6 В·м⁻¹ и 10^4 К при 10^7 В·м⁻¹. Очевидно, что при таких полях kT_e сравнимо или выше, чем $\Delta E_1 - \Delta E$, и поэтому большое число электронов будет перебрасываться в более высокий минимум. Это не оказывало бы сильного воздействия на вольтамперную характеристику, если бы не существовало одной любопытной особенности, состоящей в том, что в этом более высоком минимуме зоны эффективная масса значительно больше и подвижность соответственно существенно меньше, чем в нижнем. Поэтому, когда большое число электронов перебрасывается в верхний минимум, следует ожидать быстрого падения тока при увеличении поля. Выражая плотность тока в виде $j = nev_d$, где v_d — некоторая дрейфовая скорость, можно получить зависимость v_d от электрического поля \mathcal{E} , подобную изображенной на рис. 12.2. Мы увидим, что v_d (и, следовательно, j) при повышении \mathcal{E} проходит через острый максимум, быстро падает и затем снова довольно медленно возрастает. Спадающая часть кривой соответствует отрицательному сопротивлению. Такой характеристики отрицательного сопротивления длительное время добивались, чтобы найти способ нейтрализации положительного сопротивления цепи и получить, таким образом, твердотельный генератор.

Если n_1 и μ_1 — соответственно концентрация и подвижность электронов в центральном минимуме, а n_2 и μ_2 — те же характеристики, но для дополнительного, более высокого минимума, то плот-

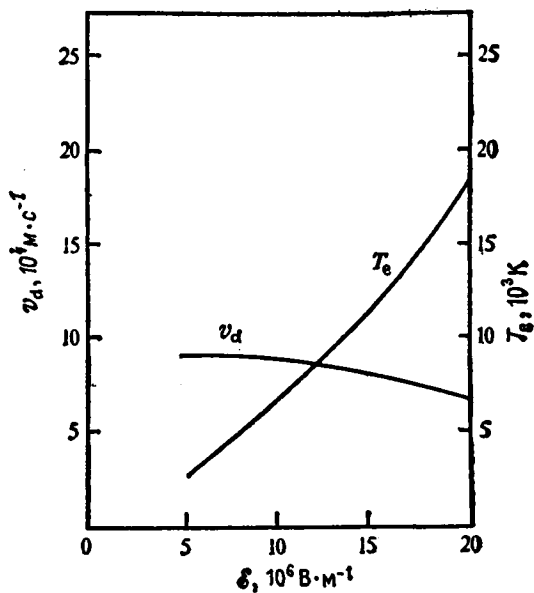
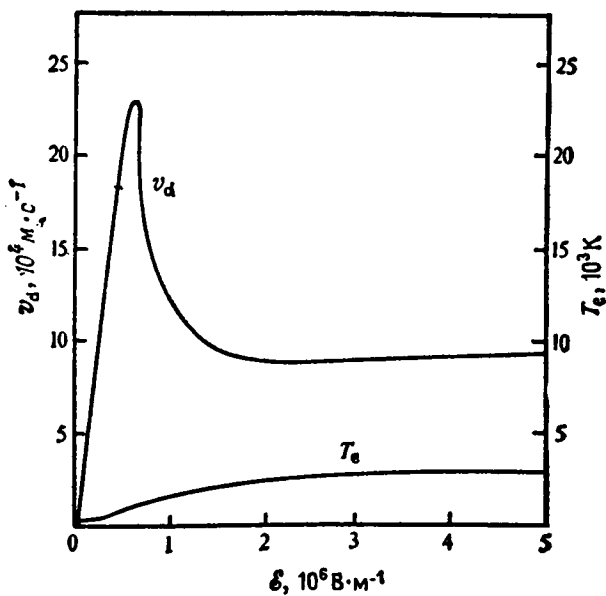


Рис. 12.2. Зависимость дрейфовой скорости v_d и эффективной температуры T_e в нижнем минимуме зоны проводимости GaAs от напряженности электрического поля E [17].

ность тока j можно записать в виде

$$j = e(n_1\mu_1 + n_2\mu_2)\mathcal{E} = nev_d, \quad (12.31)$$

где $n = n_1 + n_2$, так что

$$\frac{v_d}{\mathcal{E}} = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}. \quad (12.32)$$

Когда n_2 мало, то $v_d = \mathcal{E}\mu_1$, а когда мало n_1 , то $v_d = \mathcal{E}\mu_2$. Если μ_1 значительно выше, чем μ_2 , то, чтобы получить максимум, показанный на рис. 12.2, требуется достаточно резкое уменьшение n_1 при повышении электрического поля \mathcal{E} , иначе будет иметь место лишь небольшое уменьшение наклона кривой v_d/\mathcal{E} . Таким образом, можно ожидать, что величина напряженности поля, соответствующая максимуму, ненамного ниже критической напряженности поля пробоя, когда эффективная температура для электронов в центральной «долине» быстро растет. В GaAs это поле равно $3 \cdot 10^5$ В·м⁻¹, в то время как пробой должен наступать, как мы видели, при $\mathcal{E} = 3,5 \cdot 10^5$ В·м⁻¹. Это подтверждает и величина T_e для электронов в центральном минимуме, которая также показана на рис. 12.2.

Оказывается, что стационарное состояние, при котором отрицательное сопротивление сохраняется во всем пространстве между электродами, нестабильно. Первым это установил Ридли [18], хотя предположения, на которых основан его анализ, оказались не вполне корректными. Ганн [19] также исследовал эту нестабильность и установил, что если напряженность электрического поля, приложенного к короткому образцу из GaAs, достаточно высока, чтобы создать область отрицательного наклона на вольтамперной характеристике, то в образце устанавливаются электрические колебания, частота которых пропорциональна длине образца. На этом эффекте, который теперь называют эффектом Ганна, основана работа микроволновых генераторов, разработка которых стимулировала большое число исследований эффекта междолинного переброса электронов.

Очевидно, в состоянии, когда поддерживается генерация, через пространство между электродами с более или менее постоянной скоростью движутся некоторые области, характеризующиеся наличием в них сильного электрического поля. Эти области сильного поля, где большинство электронов находятся в минимуме зоны с низкой подвижностью, принято называть доменами. Они начинают свое движение у катода и дрейфуют со скоростью порядка максимального значения v_d , которое для GaAs приблизительно равно $2 \cdot 10^5$ м·с⁻¹. Когда один домен достигает анода, другой возникает у катода и процесс повторяется. При величине дрейфовой скорости $2 \cdot 10^5$ м·с⁻¹ для частоты генерации 10 ГГц, что соответствует длине волны 3 см, длина образца должна составлять 20 мкм.

Основную причину возникновения нестабильностей, описанных Ридли [18] и Ганном [19], нетрудно понять. Поскольку эффективная

подвижность после прохождения максимума v_d становится отрицательной, то отрицательным становится и дифференциальное удельное сопротивление ρ_d . Значение τ_0 в выражении (7.28), равное $\epsilon \rho_d$, в этих условиях будет отрицательным, и любой объемный заряд будет экспоненциально *расти*, а не исчезать. Это означает, что будет усиливаться как аккумуляция электронов в прикатодной области, так и обеднение электронами области возле анода. Детали этого процесса очень сложны и обсуждались многими авторами. Обзор работ, посвященных этому вопросу, сделал Хилсум [20].

Если одна область кристалла, в которой почти все электроны находятся в минимуме зоны с низкой подвижностью и напряженность поля в которой равна \mathcal{E}_2 , следует за областью, где большинство электронов находится в центральном минимуме и где напряженность поля равна \mathcal{E}_1 , то из условия непрерывности тока следует $\mathcal{E}_1 \mu_1 = \mathcal{E}_2 \mu_2$, или $\mathcal{E}_2 / \mathcal{E}_1 = \mu_1 / \mu_2$. Для GaAs $\mu_1 \approx 8000 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, а $\mu_2 \approx 150 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, что дает $\mathcal{E}_2 \approx 60 \mathcal{E}_1$. Такая экстремальная ситуация в действительности не реализуется, однако отсюда ясно, что области сильного поля возникнут, даже если n_2 будет несколько меньше, чем n_1 . Мы можем отметить также, что из-за более высокой эффективной массы носителей в более высоком минимуме, $0,35 m_0$ по сравнению с эффективной массой $0,067 m_0$ в центральном минимуме, и из-за того обстоятельства, что в направлении $\langle 100 \rangle$ существуют три боковых минимума, эффективная плотность состояний на более высоких уровнях будет приблизительно в 30 раз выше.

Домены будут образовываться не при любых условиях. Например, ясно, что пролетное время должно превышать время релаксации объемного заряда $\epsilon \epsilon_0 / ne \bar{\mu}_d$, где $\bar{\mu}_d$ — отрицательная дифференциальная подвижность ($\sim 3000 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$). Для GaAs это условие сводится к тому, что произведение nl должно быть больше, чем $\sim 2 \cdot 10^{11} \text{ см}^{-2}$. При длине образца $l = 20 \text{ мкм}$ это означает, что n должно превышать 10^{14} см^{-3} , а этого нетрудно достичь.

Вычисление эффективной температуры для случая многих минимумов оказывается значительно более сложным. Один из методов расчета, который был при этом использован, заключается в том, что для каждого минимума вводится эффективная температура T_s и используется «смещенное» максвелловское распределение с вектором смещения q_s . Дрейфовая скорость в каждом минимуме тогда выражается как $\hbar q_s / m_s$, где m_s — эффективная масса носителей в этом минимуме. Используя определенные законы сохранения, вычисляют эффективную температуру (в предположении, что она одинакова для всех верхних минимумов). По-видимому, первым этот метод использовал Хилсум [21], который сделал упрощающее предположение о постоянстве величин подвижности в центральной и в боковых долинах и показал, что при этом можно получить характеристику, общая форма которой совпадает с формой кривой на

рис. 12.2. Ясно, что это слишком упрощенная модель, поскольку в достаточно высоких полях, когда существенная часть электронов возбуждается в верхний минимум, подвижность должна зависеть от величины поля.

Батчер и Фосетт [22], также используя смещенные максвелловские распределения, выполнили намного более тщательный расчет, включив в рассмотрение различные механизмы рассеяния. Кривые, показанные на рис. 12.2, построены по результатам их вычислений для GaAs. Видно, что для центральной долины T_e достигает очень высокого значения $1,8 \cdot 10^4$ К при напряженности поля $2 \cdot 10^7$ В·м⁻¹. Они также выполнили вычисления для InP и CdTe и получили сходную форму зависимости v_d от \mathcal{E} .

Хотя использование смещенных максвелловских распределений позволяет получить общий вид вольтамперной характеристики, желательно все же вычислить форму кривой распределения из первых принципов (из уравнения Больцмана). Это трудная задача, но, как показали Фосетт, Бордман и Свейн [23], ее можно выполнить, используя современные вычислительные методы (метод Монте-Карло). Они применили этот метод для случая GaAs и показали, что в достаточно сильных полях смещенные максвелловские распределения дают существенно заниженные значения для количества электронов высоких энергий. Пейдж [24] применил этот метод вычислений для Ge, чтобы выяснить роль более высокого минимума в зоне проводимости. При $T=300$ К форма зависимости j (\mathcal{E}) подобна изображенной на рис. 12.1, т. е. на ней отсутствует максимум и следующий за ним участок отрицательного наклона, однако при более низких температурах отрицательный наклон все же образуется. Этот эффект, однако, отличается от того, что имеет место для GaAs главным образом из-за намного более сильного полярного рассеяния на решетке и из-за большего отношения подвижностей в последнем случае.

Конуэлл и Вессель [25] использовали иной подход, который основан на предположении о почти упругих соударениях и предусматривает решение уравнения Больцмана. Оказалось неожиданным, что результаты этого расчета не отличаются существенно от результатов, полученных с помощью метода Монте-Карло.

Этот метод обсуждался также в обзоре Конуэлл, на который мы уже ссылались. Два обзора теории зависимости дрейфовой скорости от поля сделал Батчер [17, 26].

Исследование эффектов переброса электронов, и в особенности эффекта Ганна, привели к разработке ряда новых полупроводниковых приборов. Обзор достижений в этой области недавно сделал Хилсум [27], который рассмотрел в нем также приборы, действие которых основано на эффекте лавинного пробоя. Эффект Ганна и его использование для разработки микроволновых генераторов и других электронных устройств подробно описаны в книге Хобсона [28].

12.4.1. АКУСТОЭЛЕКТРИЧЕСКИЙ ЭФФЕКТ

Наиболее сильно взаимодействие между электронами и решеткой проявляется при возбуждении колебаний решетки с помощью внешнего источника ультразвуковых волн в пьезоэлектрических материалах, где изменения давления вызывают довольно сильные электрические поля, действующие на электроны и дырки. Найн [29] наблюдал эффект изменения затухания ультразвука в кристаллах CdS, обусловленный изменением проводимости при освещении. Такие изменения, однако вызванные приложенным электрическим полем, наблюдали также Хатсон, Мак-Фи и Уайт [30]. Было найдено, что при достаточно сильном электрическом поле, когда дрейфовая скорость электронов превышает скорость распространения ультразвуковых волн, энергия электронов *передается* этим волнам и они не ослабляются, а усиливаются. Хатсон и Уайт [31] разработали теорию этого эффекта, который назвали акустоэлектрическим. В этой теории рассматривается только случай «малого сигнала». При увеличении электрического поля выше значения, при котором возникает усиление, ток стремится к насыщению, как на рис. 12.1.

Понятно, что в условиях, когда возможно усиление ультразвуковых волн, в этих волнах может быть сосредоточена большая мощность. Если эта мощность превышает некоторое определенное значение, возникают нестабильности; при этом форма распределения электронов заметно отличается от равновесной и образуются домены, сходные с теми, которые мы обсуждали в связи с эффектом Ганна (см. разд. 12.4). Эти домены являются результатом сильного взаимодействия электрического поля в ультразвуковой волне с электронами и дырками.

Было разработано большое число электронных приборов, использующих акустоэлектрический эффект. Поскольку домены распространяются с практически постоянной скоростью, близкой к скорости распространения ультразвука в кристалле, на основе этого эффекта были созданы линии задержки и другие подобные устройства. Обзор работ по созданию акустоэлектрических приборов вместе с теорией для случаев малого и большого сигналов дали Мэни и Бальберг [32].

12.5. Квантовая теория движения электронов в сильном магнитном поле

Метод эффективной массы, который мы обсуждали в разд. 11.6, очень хорошо подходит для анализа движения электронов (или дырок) в сильном внешнем магнитном поле.

В случае скалярной эффективной массы m_e уравнение для функции $F(\mathbf{r})$, которую мы ввели в разд. 11.5, можно записать в виде (см. разд. 11.6)

$$\frac{1}{2m_e} (i\hbar\nabla + e\mathbf{A})^2 F(\mathbf{r}) = EF(\mathbf{r}). \quad (12.33)$$

Если магнитное поле с индукцией B направлено вдоль оси z , вектор-потенциал \mathbf{A} можно взять в виде $(0, Bx, 0)$, так как \mathbf{B} определяется векторным произведением $\nabla \times \mathbf{A}$.

Решения этого уравнения для такого вектор-потенциала впервые получил Ландау [33] для случая свободных электронов, и эти решения можно сразу применить для электронов в полупроводниках. Функцию $F(\mathbf{r})$ можно записать в виде

$$F(\mathbf{r}) = \varphi(x') \exp(ik_y y + ik_z z), \quad (12.34)$$

заменив переменную в соответствии с равенством $x' = x - \hbar k_y / eB$.

Функция $\varphi(x')$ удовлетворяет хорошо известному уравнению для гармонического осциллятора

$$\frac{d^2\varphi}{dx'^2} + \frac{2m_e}{\hbar^2} \left(E' - \frac{1}{2} m_e \omega_c x'^2 \right) \varphi = 0, \quad (12.35)$$

где $\omega_c = eB/m_e$ — циклотронная частота, а

$$E' = E - \frac{\hbar^2 k_z^2}{2m_e}. \quad (12.36)$$

Уравнение (12.35) дает нужные решения лишь при значениях E' , равных $\hbar\omega_c(n + 1/2)$, где n — целое число. Энергия E выражается тогда формулой

$$E = \frac{\hbar^2 k_z^2}{2m_e} + \left(n + \frac{1}{2} \right) \hbar\omega_c. \quad (12.37)$$

Различные n определяют дискретные уровни энергии, которые называют уровнями Ландау. Формула для энергии этих дискретных уровней сильно отличается от формул для разрешенных значений энергии в отсутствие магнитного поля [см. (2.18)]

$$E = \frac{\hbar^2 (k_x^2 + k_y^2 + k_z^2)}{2m_e}. \quad (12.38)$$

Благодаря учету граничных условий для конечного кристалла уровни энергии, определяемые формулой (12.38), должны образовать квантованную систему близкорасположенных уровней, однако если не рассматривать вопрос о количестве уровней, этим можно пренебречь и считать, что значения энергии непрерывны. Форма энергетических зон при наличии сильного магнитного поля показана на рис. 12.3. Отметим, что кинетическая энергия электрона теперь выражена только через его движение в направлении магнитного поля и остается неизменной. Если эффективная масса электрона является

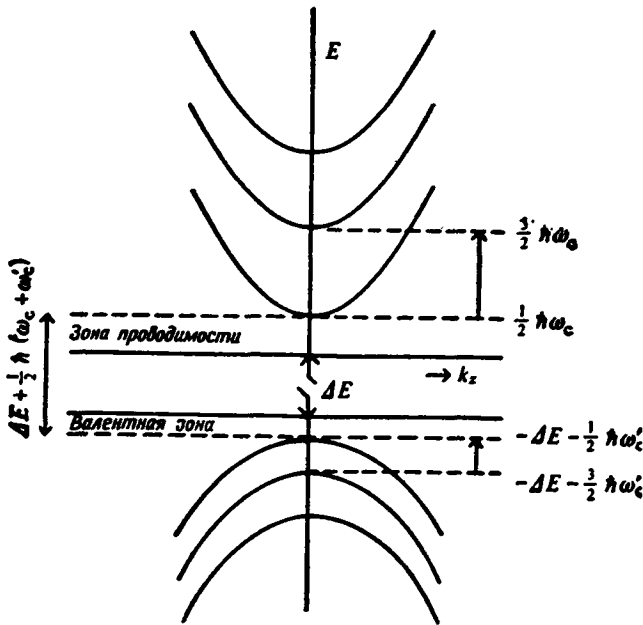


Рис. 12.3. Форма энергетических зон в сильном магнитном поле.

скалярной величиной, то движение поперек магнитного поля, т. е. в x, y -плоскости, происходит по квантованным круговым орбитам. Таким образом, энергия дается серией дискретных значений, наряду с которыми имеется кинетическая энергия движения в направлении магнитного поля. Как мы увидим ниже, одномерный характер этого распределения оказывает очень заметное влияние на плотность состояний.

На время, однако, исключим из рассмотрения кинетическую энергию и обратимся к вопросу о природе дискретных состояний. Они образуют «лестницу» равноотстоящих друг от друга уровней (лестница Ландау) с расстоянием между ними, равным $\hbar\omega_c$. Дно зоны проводимости уже не является разрешенным состоянием, наименьшая разрешенная энергия в зоне на $1/2\hbar\omega_c$ выше дна зоны в отсутствие поля. Это иллюстрируется рис. 12.4.

Энергия дырок в случае невырожденной валентной зоны дается формулой

$$E = -\Delta E - \frac{\hbar^2 k_x^2}{m_h} - \left(n' + \frac{1}{2}\right) \hbar\omega'_c, \quad (12.39)$$

где $\omega'_c = eB/m_h$. Это также иллюстрируется рис. 12.3 и 12.4.

При наличии двух соприкасающихся в экстремумах зон ситуация сильно усложняется. Этот случай рассмотрел Латинджер [34].

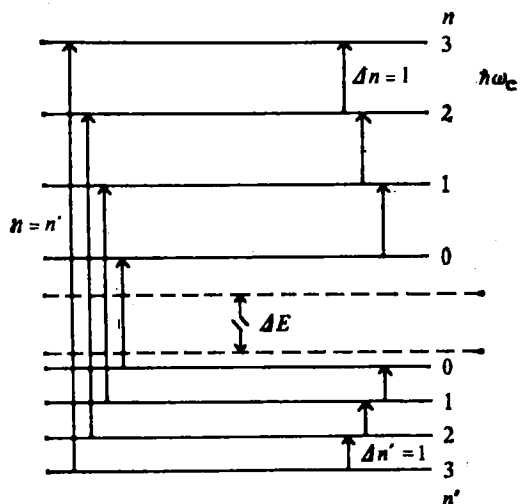


Рис. 12.4. «Лестницы» уровней Ландау для электронов в зоне проводимости и дырок в невырожденной валентной зоне.

Разрешенные переходы $\Delta n=1$ и $\Delta n'=1$ обуславливают циклотронный резонанс для электронов и дырок. Межзонные переходы разрешены при $n=n'$.

Если мы можем использовать приближение легких и тяжелых дырок, то можно ожидать, что образуются две «лестницы» с интервалами энергий $e\hbar B/m_{h1}$ и $e\hbar B/m_{h2}$. Как мы увидим, ситуация еще больше усложнится, если учесть спин электрона, что фактически даст нам целых четыре «лестницы», для которых нумерация уровней не так проста, как в отсутствие вырождения при $k=0$.

12.5.1. ПЛОТНОСТЬ СОСТОЯНИЙ В СИЛЬНЫХ МАГНИТНЫХ ПОЛЯХ

Если мы рассмотрим плотность состояний для одномерного распределения значений k_z , то найдем важное отличие от трехмерного распределения. Обозначив энергию $\hbar k_z^2/2m_e$ через E_k , найдем, что $N(E_k)$ пропорционально (см. [1], стр. 29) $E_k^{-1/2}$, а не $E^{1/2}$, как для трехмерного распределения. Это означает, что при $k_z=0$ мы имеем бесконечно высокую плотность состояний. В разд. 10.4.2 мы видели, что строго это имело бы место в отсутствие столкновений, т. е. при условии $V\mu \gg 1$. Это условие пока находится в силе, поскольку приведенный выше анализ был выполнен в предположении, что электроны и дырки движутся в кристалле свободно.

Если мы учтем кинетическую энергию электрона в направлениях x и y , то увидим, что при возрастании $V\mu$ до значений, намного пре-

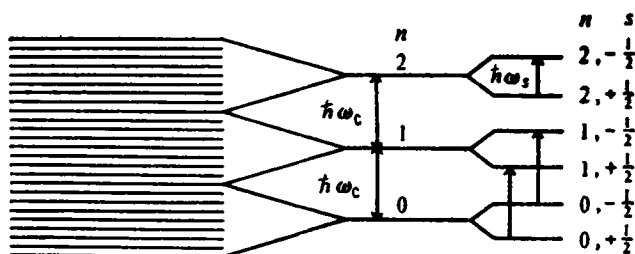


Рис. 12.5. Стягивание уровней в «лестницу» Ландау и их расщепление из-за наличия спина,

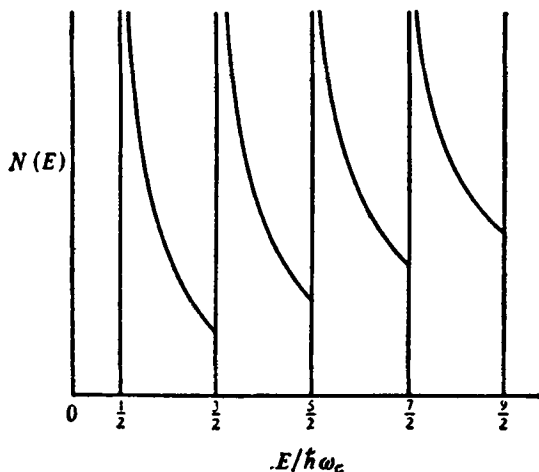


Рис. 12.6. Плотность состояний в сильном магнитном поле.

вышающих единицу, энергетические уровни, соответствующие k_x и k_y , стягиваются в ряд дискретных уровней, которые, следовательно, оказываются сильно вырожденными. Это иллюстрируется рис. 12.5. Можно оценить это вырождение и вычислить число состояний в каждом уровне Ландау. Мы убедимся при этом, что в сумме они дадут то же значение, что и раньше. Если магнитное поле мало и имеется много уровней Ландау в малом интервале δE , то мы снова получим то же число уровней, что и без магнитного поля (см. [1], § 12.3.1).

Чтобы получить выражение для зависимости плотности состояний от энергии E , мы должны провести суммирование по уровням Ландау. Для единичного объема получаем

$$N(E) = \frac{eB(2m_c)^{1/2}}{4\pi^2\hbar^2} \sum_n \left[E - \left(n + \frac{1}{2} \right) \hbar\omega_c \right]^{-1/2}. \quad (12.40)$$

Эта зависимость изображена на рис. 12.6. Нетрудно показать, что если расстояние между уровнями Ландау мало и мы заменяем эту сумму интегралом, то получим просто значение $N(E)$, которое дает формула (4.12а).

12.5.2. ЦИКЛОТРОННЫЙ РЕЗОНАНС

Если на кристалл падает излучение с энергией кванта, равной энергетическому расстоянию между уровнями Ландау, то между ними будут возбуждаться переходы. Правило отбора для гармонического осциллятора $\Delta n = \pm 1$ (см. [1], § 1.7) показывает, что могут осуществляться только переходы между соседними уровнями. (Однако из-за наличия связи возможны и намного более слабые переходы с $\Delta n = 2$, которые также наблюдаются.) Переходы с $\Delta n = \pm 1$ происходят при частоте излучения $\omega_c/2\pi$ и обусловлены просто циклотронным резонансом. Квантовая теория приводит, следовательно, к тому же результату (за исключением слабых переходов с $\Delta n = 2$), что и полуклассическая теория, которая была рассмотрена в разд. 10.4.1.

Для случая эллипсоидальных поверхностей равной энергии с эффективными массами m_1, m_2, m_3 уравнение (12.35) можно записать в более общем виде

$$\sum_{r=1}^3 \frac{1}{2m_r} \left(i\hbar \frac{\partial}{\partial x_r} + eA_r \right)^2 F(\mathbf{r}) = EF(\mathbf{r}), \quad (12.41)$$

где A_r — компоненты вектор-потенциала по трем основным осям. Если B направлено по одной из осей, скажем по оси z , это уравнение после простого преобразования можно решить, как и ранее. Получим (см. [1], § 11.4)

$$E = \frac{\hbar^2 k_z^2}{2m_3} + \frac{\hbar e B (n + 1/2)}{(m_1 m_2)^{1/2}}. \quad (12.42)$$

Круговая частота циклотронного резонанса ω_c равна, следовательно, $eB/(m_1 m_2)^{1/2}$. Аналогично, если поле направлено по осям 1 или 2, круговые частоты ω_1 и ω_2 соответственно равны $eB/(m_2 m_3)^{1/2}$ и $eB/(m_1 m_3)^{1/2}$. С помощью последующего преобразования координат можно легко показать, что для магнитного поля в направлении (l, m, n) циклотронная частота ω_c дается формулой (см. [1], § 11.4)

$$\omega_c^2 = l^2 \omega_1^2 + m^2 \omega_2^2 + n^2 \omega_3^2. \quad (12.43)$$

Это совпадает с (10.61), но теперь результат получен с помощью квантовой теории.

Если энергия не задается точно выражением (12.38), то расстояния между уровнями Ландау не будут одинаковыми, и произойдет уширение линии циклотронного резонанса. Столкновения также вы-

зывают уширение. Как мы увидим позже, изменение энергетического зазора между соседними уровнями Ландау легче наблюдать с помощью метода оптического поглощения, исследуя переходы между дырочной и электронной «лестницами» уровней. Приведенные выше уравнения для электронов в равной мере применимы и для дырок, но они не учитывают гофрирования валентной зоны, обусловленного вырождением в точке $k=0$ (см. разд. 11.6).

Взаимодействие между этим гофрированием и спином электрона дает большое количество деталей тонкой структуры валентной зоны. Латинджер и Кон [35] рассмотрели этот вопрос теоретически. Строение валентной зоны Ge экспериментально исследовали Флетчер, Ягер и Мерритт [36], а также Стиклер, Цайгер и Хеллер [37], которые провели более подробные измерения и при температуре 1,2 К разрешили большинство линий тонкой структуры, используя излучение с длиной волны 2 мм. Баттон и др., используя лазерное излучение в далекой ИК-области (см. разд. 13.4), разрешили также тонкую структуру энергетического спектра дырок в Те.

Достаточно полное обсуждение явления циклотронного резонанса с учетом факторов, определяющих ширину линии, а также влияния других эффектов, в том числе влияния спина электрона на циклотронный резонанс, провели Лэкс и Мавродес [38], а также Баттон [39].

12.6. Спиновое расщепление уровней

Помимо «стягивания» энергетических уровней в «лестницу» Ландау сильное магнитное поле приводит и к другому эффекту. До сих пор мы в значительной степени игнорировали расщепление энергетических уровней в магнитном поле, обусловленное спином электрона. Свободный электрон в магнитном поле с индукцией B имеет энергию $\pm\beta B$, где β — магнетон Бора, $\beta = \hbar^2/2m_0$. Для электронов в атоме, а также для электронов и дырок в твердом теле это выражение следует видоизменить. Обычно энергию записывают в виде $\pm^{1/2}g\beta B$, где g — число, известное как фактор расщепления Ландэ, равное 2 для свободных электронов. Из-за сильного спин-орбитального взаимодействия в полупроводниках значение g может существенно отличаться от 2 и может быть даже отрицательным.

Для случая скалярной эффективной массы m_e Лэкс, Рот и Цвердлинг [40] получили следующую формулу для величины g -фактора:

$$g = 2 \left[1 + \left(\frac{m_e - m_0}{m_e} \right) \left(\frac{\Delta E_s}{3\Delta E + 2\Delta E_s} \right) \right], \quad (12.44)$$

где, как обычно, ΔE_s — энергия спин-орбитального расщепления. Для InSb g равно -40 . Для таких полупроводников, как Ge и Si, выражение для g намного сложнее и его обсуждали Рот и Лэкс [41].

В выражение для энергии (12.37), мы должны добавить спиновый член, тогда

$$E = \frac{\hbar^2 k_z^2}{2m_e} + (n + 1/2) \frac{\hbar e B}{m_e} \pm \frac{1}{2} g \beta B, \quad (12.45)$$

так что каждый уровень Ландау оказывается двойным и мы получаем две «лестницы» уровней. Это показано на рис. 12.5. Переходы между двумя состояниями с одним значением n могут иметь место, если при этом происходит переворот спина. Это приводит к поглощению на частоте $\omega_{sp}/2\pi$, где $\omega_{sp} = g\beta B/\hbar$ — частота спинового резонанса (см. разд. 10.7.1).

12.7. Влияние магнитного поля на межзонные переходы

Теперь мы должны рассмотреть, каким образом квантование в магнитном поле влияет на межзонные переходы. Прежде всего рассмотрим простейший возможный случай прямых переходов между двумя невырожденными «сферическими» зонами, экстремумы которых расположены при $\mathbf{k}=0$. Структура уровней Ландау показана на рис. 12.4 для случая, когда эффективные массы электронов и дырок сравнимы. Установлено, что для переходов между такими зонами действует правило отбора $\Delta n=0$, так что наименьшая энергия кванта, при которой может произойти переход, уже не равна ΔE , а равна

$$\hbar\nu_0 = \Delta E + \frac{1}{2} \hbar e B \left(\frac{1}{m_e} + \frac{1}{m_h} \right), \quad (12.46)$$

что соответствует $n=0$. Нижняя граничная частота поглощения, следовательно, повышается при увеличении B . Это проявляется в виде сдвига края собственного поглощения в сторону более коротких волн при увеличении индукции магнитного поля. Такой эффект иногда называют «сдвигом зон в магнитном поле» (см. разд. 10.4.3). Для более высоких значений n , обозначив приведенную массу электронов и дырок через m_r , получим

$$\hbar\nu_n = \Delta E + \frac{(n + 1/2) \hbar e B}{m_r}. \quad (12.47)$$

Из-за теплового разброса значений кинетической энергии носителей не следует ожидать появления резких линий поглощения. Для прямых переходов без магнитного поля должно выполняться условие $\Delta k_z=0$, что соответствует $\Delta \mathbf{k}=0$, но этот разброс энергии должен еще давать и непрерывное поглощение, соответствующее различным начальным значениям k_z . Из-за очень высокого значения плотности состояний при $k_z=0$ мы, однако, должны иметь резкий максимум поглощения при $k_z=0$, в особенности при низких тем-

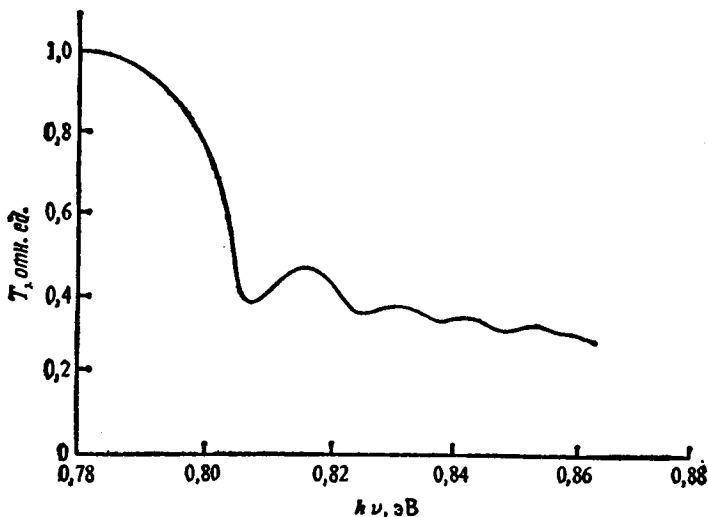


Рис. 12.7. Зависимость пропускания Ge от энергии кванта излучения в магнитном поле с индукцией $B=3,6 \text{ Т}$ [42].

пературах. Вместо линейчатого спектра поглощения был получен спектр с серией максимумов, как показано на рис. 12.7. Если построить графики зависимости положения этих максимумов от B , то мы должны получить серию прямых линий, сходящихся к значению $h\nu = \Delta E$ и имеющих наклон $(n+1/2) \hbar e B / m_0$. В точности такое поведение измеренных значений энергии и было найдено, и хотя для больших значений n и B эти линии, как правило, не являются совершенно прямыми, это дает дополнительную информацию о непараболичности зон. Результаты измерений Цвердлинга и Лэкса [42] для прямых переходов в Ge при $k=0$ показаны на рис. 12.7 и 12.8.

Такие эксперименты прямо дают точное значение ширины запрещенной зоны (0,80 эВ) при $k=0$, а также позволяют определить эффективную массу электронов (0,036 m_0) в минимуме, расположенном при $k=0$, используя известное из экспериментов по циклотронному резонансу значение эффективной массы тяжелых дырок. Очевидно, эффективную массу электронов в минимуме при $k=0$ нельзя найти из экспериментов по циклотронному резонансу, поскольку в нормальных условиях этот минимум не заселен. Аналогичные измерения были выполнены для большого числа полупроводников. Найти максимумы поглощения (или минимумы пропускания) представляется более легким делом, чем провести измерения циклотронного резонанса, и этот метод использовался для определения эффективных масс и точных значений ΔE во многих материалах с

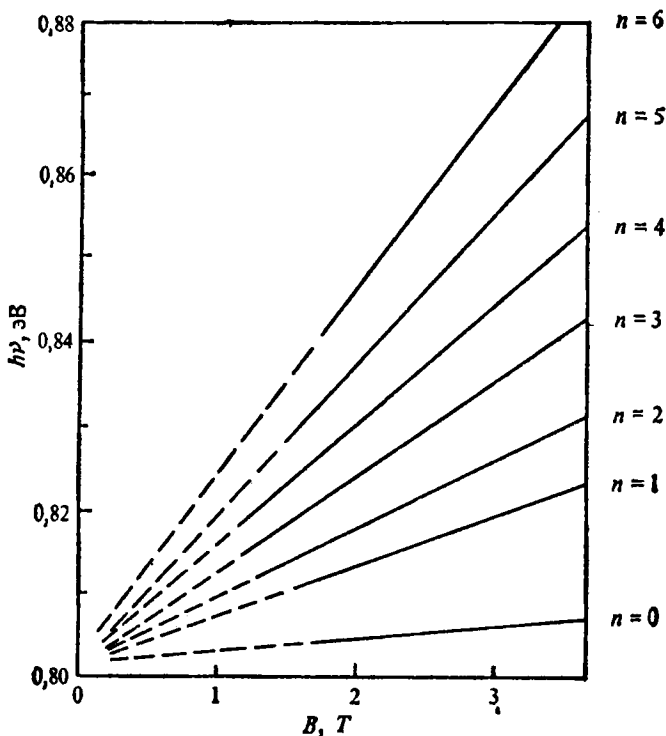


Рис. 12.8. Зависимость положения максимумов пропускания Ge от индукции магнитного поля B для различных значений индекса n уровней Ландау, между которыми осуществляется оптический переход ($\Delta n=0$) [42].

«прямыми» зонами, таких, как InSb или GaAs, даже при комнатной температуре. Для полупроводников с очень узкими зонами, таких, как InSb и сплавы, подобные $Pb_{1-x}Sn_xTe$, возникает некоторая сложность, обусловленная взаимодействием валентной зоны и зоны проводимости. Этот вопрос обсуждал Пиджен [43], который использовал этот метод даже для полуметаллов, таких, как HgTe.

Если для изучения спектров, соответствующих прямым переходам в Ge при $k=0$, использовать аппаратуру с более высоким разрешением, то обнаруживается тонкая структура этих спектров. Валентная зона Ge вырождена при $k=0$ и этим обусловлено взаимодействие между спиновыми уровнями энергии и уровнями Ландау. Это взаимодействие обсуждалось в разд. 12.5.2 в связи с циклотронным резонансом; оно проявляется также в оптических спектрах.

В отсутствие вырождения спин при переходе остается неизменным. Однако в данном случае мы имеем четыре отдельные «лестницы» Ландау для валентной зоны и две для зоны проводимости.

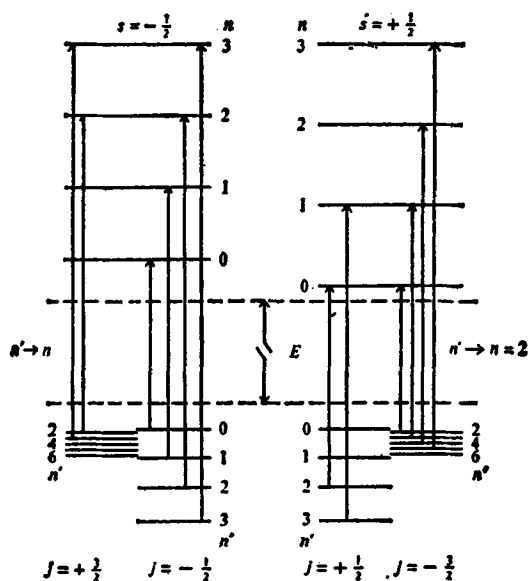


Рис. 12.9. «Лестницы» уровней Ландау для зоны проводимости и вырожденной в точке $k=0$ валентной зоны [44]. Показано расщепление, обусловленное спином, и разрешенные переходы.

Рот, Лэкс и Цвердлинг [44] подробно рассмотрели эту систему уровней и показали, что в то время как уровни зоны проводимости можно идентифицировать с помощью квантового числа Ландау n и спинового числа $s = \pm 1/2$, для валентной зоны требуются число n и магнитное квантовое число j , которое принимает значения $3/2$, $1/2$, $-1/2$, $-3/2$. «Лестницы» уровней с $j = \pm 1/2$ соответствуют зоне легких дырок и могут рассматриваться как состояния с направлением спина «вверх» и «вниз», как и для зоны проводимости. Для «лестниц» уровней с $j = \pm 3/2$ спиновые состояния смешаны с «орбитальным» угловым моментом, и их нельзя рассматривать просто как состояния со спином «вверх» или «вниз»; в этом случае существует некоторая вероятность, что спин будет иметь то или иное направление. Это обстоятельство, как мы увидим позже, приводит к некоторым важным следствиям для других полупроводников, обладающих сходными свойствами (см. разд. 14.7).

Результаты измерений спектра поглощения Ge Рот, Лэкс и Цвердлинг [44] сравнили с результатами их теоретического анализа, который показывает, что переходы между состояниями с $s = -1/2$ и состояниями с $j = -1/2$, $j = 3/2$ подчиняются правилу отбора в виде $\Delta n = 0$, а для переходов между состояниями с $s = 1/2$ и состояниями с $j = 1/2$, $j = -3/2$ действует правило отбора $\Delta n = -2$. Таким

образом они смогли объяснить все наблюдаемые линии. Схема «лестниц» уровней и разрешенные переходы показаны на рис. 12.9. Отметим, что для «лестниц» с $j = \pm 3/2$ отсутствуют состояния с $n=0$ и $n=1$.

Для непрямых переходов, таких, как переходы через минимальный энергетический зазор в Si и Ge, теория предсказывает несколько иную форму спектра из-за большей свободы в выборе энергии фонона, необходимого для сохранения импульса. Как показали Рот, Лэкс и Цвердлинг [44], в этом случае вместо максимумов на кривой поглощения будет наблюдаться серия ступенек. Такой спектр экспериментально наблюдали Цвердлинг, Лэкс, Рот и Баттон [45]. Анализ положения этих ступенек также дает информацию о ширине запрещенной зоны и зонной структуре вблизи экстремума.

Обстоятельные обзоры этих и других магнитных эффектов были даны Лэксом [38] и Мавродесом [46].

12.7.1. ЭФФЕКТ ЗЕЕМАНА В ПРИМЕСНОМ ПОГЛОЩЕНИИ

Влияние сильного магнитного поля на спектр примесного поглощения в полупроводниках очень сходно с тем, которое наблюдается в спектроскопии атомов, за исключением того, что в последних зеэмановское расщепление энергетических уровней мало по сравнению с самими значениями энергий, тогда как в примесных атомах оно является сравнимой с ними величиной и его уже нельзя рассматривать как малое возмущение. Расщепление приводит к увеличению числа уровней и числа их магнитных характеристик, а также дает большое количество новой информации об электронной структуре примесных центров. Спектры таких центров очень разнообразны и они также обсуждались Лэксом в его обзоре магнитооптических эффектов [38].

Используя фурье-спектроскопию в далекой ИК-области, Каплан [47] смог исследовать поглощение и фотопроводимость, обусловленные уровнями «водородоподобной» примеси в GaAs и InSb, и непосредственно измерить изменения положений энергетических уровней примесей в сильном магнитном поле, подтвердив результаты теоретического анализа (см. разд. 11.5.1). Обзор этого вопроса сделал также Хасегава [48].

12.8. Явления электронного переноса в сильных магнитных полях

В разд. 5.3.5 мы уже обсудили некоторые аспекты явления электронного переноса в сильном магнитном поле и нашли, что в этом случае коэффициент Холла R_H для электронов выражается про-

сто как $-1/ne$, а не $-r/ne$, где r — коэффициент, определяемый формулой (5.111). На основании этого анализа были сделаны и некоторые другие выводы, например о том, что для полупроводников типа Si при направлении поля $\langle 100 \rangle$ должно отсутствовать продольное магнитосопротивление. Это должно иметь место и для полупроводника со сферическими поверхностями равной энергии. Более того, было показано, что такой полупроводник при наличии вырождения не должен обладать и поперечным магнитосопротивлением, но оно должно иметь место в отсутствие вырождения. Были найдены выражения, определяющие предельное значение ρ_{∞} , к которому, как следует ожидать, должно стремиться удельное сопротивление в очень сильных полях. Например, для невырожденного полупроводника n -типа со сферическими поверхностями равной энергии было показано, что $\rho_{\infty}/\rho_0 = 32/9\pi$ для случая чисто решеточного рассеяния, что является весьма слабым эффектом, но если доминирует рассеяние на примесях, то $\rho_{\infty}/\rho_0 = 32/3\pi$. В более общем виде $\rho_{\infty}/\rho_0 = \langle 1/\tau \rangle \langle \tau \rangle$. Для случая смешанной проводимости выражения для ρ_{∞}/ρ_0 очень сложны и зависят от отношения n/p и отношения подвижностей b [см. выражение (5.168)]. При $n=p$, т. е. для собственного полупроводника, эффект магнитосопротивления исчезает и $\rho_{\infty} = \rho_0$.

Из всех этих выводов экспериментально подтверждено только изменение коэффициента Холла в сильном магнитном поле, для которого $\omega_c \tau \gg 1$. Более того, здесь даже найдены малые отклонения от монотонной зависимости. Они, однако, не так важны при определении концентрации носителей, и поэтому измерения R_{11} в сильном поле используются для определения n или p [или, точнее, разностной концентрации $n - p$, поскольку при смешанной проводимости в сильном поле $R_H \rightarrow [1/e(p-n)]$]. То, что большинство из этих результатов не подтверждается на опыте, едва ли можно считать неожиданным, поскольку при их выводе не учитывался эффект квантования в магнитном поле и его влияние на функцию распределения f_0 , что необходимо при $\omega_c \tau \gg 1$.

Как мы видели, для расчета эффектов электронного переноса в слабом электрическом поле в присутствии сильного магнитного поля правильной отправной точкой является использование вместо f_0 функции распределения f_{0B} , полученной с помощью выражения (12.37) для энергии. Эта функция имеет вид

$$f_{0B}^{-1} = \exp \left[\frac{E' + (n + 1/2) \hbar \omega_c - E_F}{kT} \right] + 1, \quad (12.48)$$

где $E' = \hbar^2 k_z^2 / 2m_e$ в случае скалярной эффективной массы m_e . Чтобы получить выражения для электрического или теплового токов, нужно провести суммирование по n и усреднение по k_z . Кроме того, тот вид уравнения Больцмана, на котором основан процесс усред-

нения, проведенный в разд. 5.3.5, в этих условиях неприменим, и явления переноса оказываются здесь намного более сложными.

Существует, однако, одна ситуация, когда можно избежать суммирования по n ; ее называют квантовым пределом. Если $kT \ll \ll \hbar \omega_c$, то фактически все электроны зоны проводимости будут находиться на самом нижнем уровне Ландау, и при этом нужно рассматривать только $n=0$. Тогда в отсутствие вырождения имеем

$$f_{0B} = A \exp\left(-\frac{E'}{kT}\right), \quad (12.49)$$

а число состояний на единичный интервал энергий E' равно $CE'^{-1/2}$, где A и C — постоянные, значения которых легко оценить (см. [1], § 12.3.1). Для проводимости в направлении магнитного поля можно выполнить анализ, аналогичный проведенному в разд. 5.3.1. При этом легко находим, что плотность тока j дается тем же выражением, что и ранее, но с заменой среднего времени релаксации $\langle \tau \rangle$ величиной

$$\bar{\tau} = \frac{\int_0^{\infty} \tau(E') E'^{1/2} \exp(-E'/kT) dE'}{\int_0^{\infty} E'^{1/2} \exp(-E'/kT) dE'}. \quad (12.50)$$

Мы видим, следовательно, что в этой ситуации весовая функция $E'^{1/2} \exp(-E'/kT)$ заменяется на $E'^{1/2} \exp(-E'/kT)$.

Даже если нам известно $\tau(E)$ в отсутствие магнитного поля, для определения $\tau(E')$ все же необходимо заново вычислить τ . Обычно это очень трудно сделать. Если, однако, τ является функцией скорости v , то, возможно, имеет смысл предположить, что если $\tau(E) = AE^{-s}$, то $\tau(E') = A'E'^{-s}$. При таком предположении легко находим, что

$$\frac{\rho_{\infty}}{\rho_0} = \frac{\bar{\tau}}{\langle \tau \rangle} = \frac{3-2s}{3}. \quad (12.51)$$

Для чисто решеточного рассеяния $s=1/2$ и тогда имеем $\rho_{\infty}/\rho_0 = 2/3$, в то время как при рассеянии на ионах примеси $s=-3/2$ и отношение ρ_{∞}/ρ_0 становится отрицательным. В последнем случае, однако, для малых значений E' , которые мы рассматриваем, условие $\omega_c \tau \gg 1$ не удовлетворяется. Из этого простого анализа можно сделать лишь тот вывод, что продольное магнитосопротивление в сильных полях должно существенно зависеть от содержаний примесей. Основной вывод, однако, заключается в том, что если τ не зависит от энергии, то эффект продольного магнитосопротивления будет отсутствовать, и что первоначальный анализ дает неправильный результат. Можно отметить, что при наличии только решеточного рассеяния удельное сопротивление в сильных полях *понижается*.

До сих пор мы не рассматривали проводимость в направлении, перпендикулярном направлению магнитного поля. Следует ожидать, что в квантовом пределе эта проводимость будет очень малой, так как в сильном магнитном поле электроны и дырки движутся по малым круговым орбитам. Это приводит к дополнительной трудности. Метод эффективной массы можно применять, когда «внешнее» поле слабо меняется на длине постоянной решетки. Сильное постоянное магнитное поле удовлетворяет этому условию. Но это условие означает также, что «модулирующая функция» $F(r)$ не должна сильно меняться на длине постоянной решетки. Очевидно, однако, что указанное условие перестает выполняться, когда радиус круговой орбиты электрона оказывается сравнимым с постоянной решетки кристалла. Эта ситуация определяет *верхний* предел величины магнитного поля с точки зрения применимости метода эффективной массы.

Из выражения (10.62) видно, что при $n=1$ радиус окружности, по которой движется электрон, не зависит от его эффективной массы и равен $(\hbar/eB)^{1/2}$. Для $B=10$ Т его величина составит $\sim 10^{-6}$ см, что все еще немного больше постоянной решетки, так что условие применимости, вероятно, не слишком серьезно нарушается. Этот расчет основан на «классической» квантовой теории; его можно выполнить и более точно, используя волновые функции, удовлетворяющие уравнению (12.33) (см. [1], стр. 449), однако результат будет тот же. Для материалов, подобных InSb, с малой эффективной массой электрона и, следовательно, с высокой подвижностью, достигающей значений порядка 10^6 см²·В⁻¹·с⁻¹, даже при $B=1$ Т имеем $V\mu = \omega_c \tau = 100$, т. е. условия, при которых имеет место магнитное квантование, и которые, кроме того, подходят для применения метода эффективной массы. Условие для квантового предела, когда все электроны находятся на уровне Ландау с $n=0$, может быть записано как $kT \ll \hbar\omega_c$. Для InSb $m_e \approx 0,01 m_0$ и при $B=1$ Т $\hbar\omega_c$ составляет величину $\sim 0,01$ эВ. Это означает, что при температуре $T < 12$ К условия квантового предела можно достичь, используя поле с индукцией $B=1$ Т, а при $T < 1,2$ К достаточно $B=0,1$ Т. InSb мог бы явиться, следовательно, прекрасным материалом для измерения в условиях квантового предела. Дело осложняется, однако, явлением вымораживания носителей (см. разд. 12.9).

Как мы уже указывали, строгий анализ явлений переноса и в электрических, и в тепловых полях становится очень сложным при наличии сильного магнитного поля, главным образом из-за того, что мы отходим от простого вида уравнения Больцмана. Квантовомеханическое рассмотрение этого вопроса, основанное на использовании формализма матрицы плотности, провели Аргирес [49], а также Адамс и Хальштейн [50]. Их результаты показывают, что выражение, которое мы получили для постоянной Холла в сильных полях правильно, за исключением того, что оно не содержит малых периодических членов, роль которых будет особенно

заметна при приближении к условиям вырождения; другими словами, они имеют большее значение для сильно легированных полупроводников. Это сходно с эффектом де Гааза — Ван-Альфена в металлах (см. [1], § 12.6). Несколько отличный анализ дали Кубо, Мияке и Хашитсуме [51], которые выполнили также обзор предшествующих работ по явлениям переноса в сильных магнитных полях. Еще более тщательный анализ выполнил Хайду [52], который рассмотрел также ограничения области применения разработанных ранее теорий. Были получены различные формулы, определяющие величины кинетических коэффициентов в разнообразных условиях.

Едва ли вызывает удивление малое количество измерений тепловых кинетических коэффициентов в сильных магнитных полях, поскольку такие измерения трудновыполнимы, а кроме того, и не очень важны при низких температурах, где наиболее легко реализовать условия квантового предела. Пури и Гебэлл [53] сравнили имеющиеся данные с теорией. Большее количество измерений было выполнено по электропроводности в условиях квантового предела, но, как мы уже отмечали, их результаты не всегда однозначны.

Нойрингер [54], например, сообщил о результатах измерений магнитосопротивления InSb с концентрацией носителей в пределах $(5 \cdot 10^{13} - 10^{16}) \text{ см}^{-3}$ при температуре жидкого гелия, где достигаются условия квантового предела. При увеличении поля от 0,1 до 20 Т удельное сопротивление возрастало приблизительно в 10^6 раз. Этот эффект, однако, в значительной мере обусловлен вымораживанием носителей (см. разд. 12.9). Сарьян и Нагабхасана [55], используя импульсные поля до 20 Т, наблюдали эффект насыщения как поперечного, так и продольного магнитосопротивления PbTe. Отрицательное магнитосопротивление GaAs в сильных полях наблюдали Аскенаи, Ульмет и Леотин [56]. Оно не могло быть обусловлено решеточным рассеянием, поскольку в условиях их эксперимента преобладало рассеяние на примесях. Об отрицательном магнитосопротивлении в сильных полях у Se и сильно легированного Ge сообщали также Шмарцев и др. [57]. Необычное поведение компенсированного GaAs при температурах жидкого гелия наблюдали Вул, Котельникова, Заварицкая и Воронова [58]. Они обнаружили резкое изменение проводимости, довольно похожее на изменение проводимости при переходе Мотта (см. разд. 13.14), когда происходит резкое уменьшение подвижности электронов. Считается, что понижение проводимости в полях величиной 5—15 Т вызывается также локализацией электронов в потенциальных ямах, возникающих в результате флуктуаций концентрации заряженных примесей.

Очевидно, что сжатие орбит электронов магнитным полем приводит к тому, что любая форма локализации становится более вероятной. Предстоит, однако, еще многое сделать, чтобы объяснить и полностью понять эти явления. В обзоре Бира [59], опубликованном в сборнике из серии Solid State Physics, который целиком посвя-

щен кинетическим явлениям в магнитных полях, рассмотрены некоторые экспериментальные работы, и их результаты сопоставлены с теорией. Этот вопрос подробно обсуждали также Рот и Аргирес [60].

12.9. Вымораживание носителей

В ряде полупроводников энергия связи электронов на донорах настолько мала, что даже при охлаждении до температуры жидкого гелия значительная часть доноров остается ионизованной. То же можно сказать и об акцепторах, но из-за большей, как правило, эффективной массы дырок такая ситуация является для них менее обычной, чем для электронов. Во всех халькогенидах свинца, однако, наблюдается полная ионизация как доноров, так и акцепторов. В InSb доноры с мелкими уровнями (они, правда, недостаточно надежно идентифицированы) также остаются ионизованными при очень низких температурах. Установлено, однако, что если при температуре, не превышающей нескольких градусов Кельвина, в полупроводнике создать сильное магнитное поле, то носители могут быть «заморожены» на мелких примесных центрах. Особенно поразительным является в этом смысле поведение InSb, которое описали Сладек [61] и Патли [62]. Существует два различных механизма, вызывающих вымораживание носителей. Примесный уровень донора, имеющего один слабо связанный электрон, расщепляется в магнитном поле на два, один из которых сдвигается относительно первоначального уровня энергии вверх и может попасть в зону проводимости, а другой понижается. Зазор между этим нижним уровнем и дном зоны проводимости увеличивается по мере увеличения магнитного поля и может оказаться достаточным, чтобы при низкой температуре на нем мог быть связан электрон. Величину расщепления легко вычислить, используя известное значение расщепления основного состояния атома водорода с соответствующей заменой массы свободного электрона значением эффективной массы.

Второй эффект обусловлен сжатием электронных волновых функций доноров. Его величину можно оценить, снова используя известные результаты для атома водорода с соответствующим изменением масштаба. Если концентрация примесей достаточно велика, то перекрытие волновых функций соседних атомов примеси может привести к уменьшению энергии связи до исчезающе малой величины. В результате сжатия волновых функций в магнитном поле их перекрытие может существенно уменьшиться, что будет сопровождаться соответствующим увеличением энергии связи.

Яфет, Кейес и Адамс [63] разработали теорию, которая хорошо описывает эти эффекты в случае простых «водородоподобных» примесей.

Явление вымораживания носителей не следует смешивать с магнитосопротивлением. В отличие от последнего оно является результатом изменения числа носителей, а не изменения их кинетических свойств, хотя эти два эффекта и взаимосвязаны. Недооценка эффектов вымораживания носителей является одной из основных причин неоднозначной интерпретации результатов измерений магнитосопротивления в сильных магнитных полях при низких температурах. Этот вопрос подробно обсудил Патли [64]. Ему удалось построить чувствительный фотоприемник для дальней ИК-области излучения на основе фотопроводимости InSb, обусловленной фотоионизацией электронов, связанных на примесях в присутствии магнитного поля.

ЛИТЕРАТУРА

1. *Smith R. A.*, The Wave Mechanics of Crystalline Solids, Chapman and Hall, 1969.
2. *Shockley W.*, Bell Syst. Techn. Journ., **30**, 990 (1951).
3. *Arthur J. B.*, *Gibson A. F.*, *Granville J. W.*, Journ. Electron., **2**, 145 (1956).
4. *Shibuya M.*, Phys. Rev., **99**, 1189 (1955).
5. *Gunn J. B.*, Progress in Semiconductors, Heywood, **2**, 1957, p. 213.
6. *Conwell Esther M.*, Solid State Physics, Academic Press, Suppl. 9, 1967. [Имеется перевод: Конуэлл Э. Кинетические свойства полупроводников в сильных электрических полях.— М.: Мир, 1970.]
7. *Paige E. G. S.*, Progress in Semiconductors, Heywood, **8**, 1964.
8. *Ryder E. J.*, Phys. Rev., **90**, 766 (1953).
9. *Gunn J. B.*, Journ. Electron., **2**, 87 (1956).
10. *Newman R.*, Phys. Rev., **100**, 700 (1955).
11. *Chynoweth A. G.*, *McKay K. G.*, Phys. Rev., **102**, 369 (1956).
12. *Burton J. A.*, Phys. Rev., **108**, 342 (1957).
13. *Fawcett W.*, *Paige E. G. S.*, Journ. Phys., **C4**, 1801 (1971).
14. *Fröhlich H.*, *Parandjape B. V.*, Proc. Phys. Soc., **B69**, 21 (1956).
15. *Yamashita J.*, Progress in Semiconductors, Heywood, **4**, 1960, p. 63.
16. *Pözl H.*, Electronic Materials, ed. Hannay N. B., Colombo U., Plenum Press, 1973, p. 89.
17. *Butcher P. N.*, Rept. Prog. Phys., **30**, 97 (1967).
18. *Ridley B. K.*, Solid State Commun., **1**, 88 (1963).
19. *Gunn J. B.*, I. B. M. Journ. Res. Devel., **8**, 141 (1964).
20. *Hilsum C.*, Electronic Structures in Solids, ed. Haidemenakis E. D., Plenum Press, 1969, p. 360.
21. *Hilsum C.*, Proc. IRE, New York, **50**, 185 (1962).
22. *Butcher P. N.*, *Fawcett W.*, Proc. Phys. Soc., **86**, 1205 (1965).
23. *Fawcett W.*, *Boardman A. D.*, *Swain S.*, Journ. Phys. Chem. Solids, **31**, 1963 (1970).
24. *Paige E. G. S.*, I. B. M. Journ. Res. Devel., **13**, 562 (1969).
25. *Conwell Esther M.*, *Vassel M. O.*, Phys. Rev., **166**, 797 (1968).
26. *Butcher P. N.*, Electronic Structures in Solids, ed. Haidemenakis E. D., Plenum Press, 1969, p. 366.
27. *Hilsum C.*, Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves, 1976, p. 74.
28. *Hobson G. S.*, The Gunn Effect, Clarendon Press, 1974.
29. *Nine H. D.*, Phys. Rev. Lett., **4**, 359 (1960).
30. *Hutson A. R.*, *McFee J. H.*, *White D. L.*, **7**, 237 (1961).
31. *Hutson A. R.*, *White D. L.*, Journ. Appl. Phys., **33**, 40, 2547 (1962).
32. *Many A.*, *Balberg I.*, Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969, p. 385.

33. *Ландау Л.*, *Zs. Phys.*, **64**, 629 (1950).
34. *Luttinger J. M.*, *Phys. Rev.*, **102**, 1030 (1956).
35. *Luttinger J. M.*, *Kohn W.*, *Phys. Rev.*, **97**, 869 (1955).
36. *Fletcher R. C.*, *Yager W. A.*, *Merritt F. R.*, *Phys. Rev.*, **100**, 747 (1955).
37. *Stickler J. J.*, *Zeiger H. J.*, *Heller G. S.*, *Phys. Rev.*, **127**, 1077 (1962).
38. *Lax B.*, *Mavroides J. G.*, *Rendiconti della Scuola Internazionale di Fisica Enrico Fermi, Corso XXII*, Academic Press, 1963, p. 240.
39. *Button K. J.*, *Optical Properties of Solids*, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 253.
40. *Lax B.*, *Roth Laura M.*, *Zwerdling S.*, *Proc. IVth Intern. Conf. Phys. Semicond., Journ. Phys. Chem. Solids*, **8**, 311 (1958).
41. *Roth Laura M.*, *Lax B.*, *Phys. Rev. Lett.*, **3**, 217 (1959).
42. *Zwerdling S.*, *Lax B.*, *Phys. Rev.*, **106**, 51 (1957).
43. *Pidgeon C.*, *Electronic Structure in Solids*, ed. Haidemenakis E. D., Plenum Press, 1969, p. 47.
44. *Roth Laura M.*, *Lax B.*, *Zwerdling S.*, *Phys. Rev.*, **114**, 90 (1959).
45. *Zwerdling S.*, *Lax B.*, *Roth Laura M.*, *Button K. J.*, *Phys. Rev.*, **114**, 80 (1959).
46. *Mavroides J. G.*, *Optical Properties of Solids*, ed. Abeles F., North Holland, 1972, p. 351.
47. *Kaplan R.*, *Optical Properties of Solids*, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 301.
48. *Hasegawa H.*, *Physics of Solids in High Magnetic Fields*, ed. Haidemenakis E. D., Plenum Press, 1969.
49. *Argyres P. N.*, *Phys. Rev.*, **109**, 1115 (1958).
50. *Adams E. N.*, *Halstein T. D.*, *Journ. Phys. Chem. Solids*, **10**, 254 (1959).
51. *Kubo R.*, *Miyake S. J.*, *Hashitsume N.*, *Solid State Physics*, Academic Press, **17**, 1965, p. 269.
52. *Hajdu J.*, *Electronic Structure in Solids*, ed. Haidemenakis E. D., Plenum Press, 1969, p. 305.
53. *Puri S. M.*, *Geballe T. H.*, *Semiconductors and Semimetals*, Academic Press, **1**, 1966, p. 203.
54. *Нойрингер Л. Дж.*, *Труды IX Международной конференции по физике полупроводников.*—Л.: Наука, 1969, с. 757.
55. *Suryan G.*, *Nagabhushana S.*, *Proc. XIIIth Intern. Conf. Phys. Semicond., Tipographia Marves*, 1976, p. 1121.
56. *Askenaye S.*, *Ulmet J. P.*, *Leotin J. L.*, *Solid State Commun.*, **7**, 979 (1969).
57. *Шмарцев Ю. Н.*, *Полянская Т. А.*, *Шендер Е. Ф.*, *Голубев Л. В.*, *Вуль А. Я.*, *Сайдашев И. И.*, *Proc. XIth Intern. Conf. Phys. Semicond., Polish Sci. Publ.*, 1972, p. 410.
58. *Вул Б. М.*, *Котельникова Н. В.*, *Заварицкая Э. И.*, *Воронова И. Д.*, *Proc. XIIth Intern. Conf. Phys. Semicond., Tipographia Marves*, 1976, p. 1188.
59. *Beer A. C.*, *Solid State Physics*, Academic Press, Suppl. 4, 1963.
60. *Roth Laura M.*, *Argyres P. N.*, *Semiconductors and Semimetals*, Academic Press, **1**, 1966, p. 159.
61. *Sladek R. J.*, *Journ. Phys. Chem. Solids*, **5**, 157 (1958).
62. *Putley E. H.*, *Proc. Phys. Soc.*, **76**, 802 (1960).
63. *Yafet Y.*, *Keyes R. W.*, *Adams E. N.*, *Journ. Phys. Chem. Solids*, **1**, 137 (1956).
64. *Putley E. H.*, *Semiconductors and Semimetals*, Academic Press, **1**, 1966, p. 289.

ПОЛУПРОВОДНИКОВЫЕ МАТЕРИАЛЫ

13.1. Материалы, используемые в полупроводниковых приборах и исследованиях

В этой главе мы кратко рассмотрим основные свойства небольшого числа полупроводников, которые благодаря своему научному значению и широкому применению в промышленности, изучались особенно подробно. Из элементарных полупроводников в этот класс попадают только Ge, Si, и Te. Что касается Se, то его вряд ли можно отнести к этому классу, так как его промышленное использование ограничивается в основном применением аморфного Se в технике фотокопирования, где он ведет себя скорее как диэлектрик. Изучалось также и серое олово, однако из-за сложности получения чистых образцов этот материал еще не привлек к себе такого же внимания, как другие. Основными полупроводниковыми материалами следует считать Si и Ge. Из них в фундаментальных физических исследованиях больше использовался Ge. Он имеет более низкую температуру плавления (937°C) по сравнению с Si (1420°C), поэтому его легче готовить и очищать. Германий использовался в большинстве первых твердотельных электронных приборов, например в транзисторах, однако к настоящему времени его почти полностью заменил намного более распространенный в природе Si. Сложные интегральные схемы современной электроники почти полностью изготавливают из Si. Полупроводник Te сам по себе очень интересен, поскольку в монокристаллической форме он является анизотропным проводником, однако до сих пор широкого распространения в твердотельном приборостроении не получил.

Из двойных полупроводниковых соединений известность приобрели две группы материалов. Из соединений типа $A^{III}B^{VI}$ уже давно изучался CdS благодаря использованию его в измерителях интенсивности света. В то же время PbS, PbSe и PbTe были подробно исследованы в связи с их применением в качестве материалов для детекторов инфракрасного излучения. Из соединений типа $A^{III}B^V$ наиболее интенсивно изучались InSb, GaAs и GaP, хотя все девять соединений, включающих комбинации Al, In, Ga, с одной стороны,

и Sb, As, Р — с другой, привлекали к себе некоторое внимание. Распространение получили как двойные соединения такого типа, так и их сплавы. В основном мы рассмотрим здесь двойные соединения, однако упомянем еще и несколько других. Начинают привлекать к себе внимание и более сложные полупроводники, например тройные соединения, хотя они еще не получили широкого применения в промышленности. Однако разнообразие этих материалов настолько велико, что было бы удивительно, если бы некоторые из них не оказались технологически важными. Такие соединения уже находят применение, хотя и в небольших масштабах, в нелинейной оптике, и мы кратко упомянем те из них, о свойствах которых больше всего известно.

13.2. Приготовление полупроводниковых материалов и измерение их свойств

Полупроводниковые материалы могут быть получены двумя основными способами: выращиванием из расплава и методом эпитаксиального роста. Выращивание из расплава можно проводить по так называемому методу Стокбаргера, при котором расплавленный материал очень медленно опускают через область с малым температурным перепадом, в результате чего материал затвердевает, и методу Чохральского, в котором рост кристалла начинается от затравочного кристалла, а затем этот растущий кристалл медленно вытягивают вертикально вверх из расплава. Разновидностью этого метода является метод горизонтальной зонной плавки, при котором «вытягивание» осуществляют путем перемещения «лодочки», содержащей затравочный кристалл и расплав, через узкую зону температурного перепада.

Эпитаксиальный рост представляет собой наращивание необходимого кристалла на другом кристалле путем осаждения вещества из газовой или жидкой фазы. Этим методом сравнительно легко добиться контролируемого введения примесей, однако трудно получить кристаллы больших размеров. При разработке современных электронных приборов этот метод находит все более широкое применение.

О приготовлении и очистке полупроводниковых материалов написано так много и методы так сильно различаются для разных материалов, что подробное описание этого вопроса заняло бы слишком много места. К счастью, имеется несколько замечательных книг, посвященных этой проблеме, в которых рассматривается большинство практически важных материалов (см., например, [1—6]). Известно также много разных методов определения основных свойств

полупроводниковых материалов. Некоторые из них были упомянуты выше в связи с обсуждением этих свойств. Методы можно разделить на три основные группы: теоретические вычисления, электрические и магнитные измерения и магнитооптические или оптические измерения. Здесь снова можно указать несколько книг, в которых подробно рассматриваются конкретные методики (см., например, [1, 7]). Поскольку все методики имеют свои индивидуальные особенности, объем нашей книги не позволяет нам рассмотреть их достаточно полно. Методы определения различных параметров иногда обсуждаются и обычно упоминаются в справочниках по свойствам материалов, на которые мы будем ссылаться ниже.

Значение ширины запрещенной зоны находят, вообще говоря, из сочетания экспериментальных измерений и теоретических расчетов, причем в настоящее время наиболее точные значения получаются из измерений оптического поглощения и излучения как в условиях магнитного поля, так и без него. Эффективные массы определяют методами циклотронного резонанса, фарадеевского вращения или магнитооптического поглощения. Подвижности определяют на основе измерений коэффициента Холла и электропроводности в сочетании с теорией рассеяния электронов и дырок, если, конечно, невозможно применить прямые методы. Глубины залегания примесных уровней получают путем измерений коэффициента Холла или спектра оптического поглощения и излучения, причем в настоящее время последним двум методам обычно отдают предпочтение. Времена жизни неосновных носителей находят, наблюдая времена спада сигнала от инжектированных носителей в экспериментах по фотопроводимости или в других подобных экспериментах, хотя используются также и некоторые простые непрямые методы. Для более подробного описания мы выбрали несколько полупроводников различных типов, чтобы продемонстрировать большое разнообразие свойств, которыми эти материалы обладают. Основные параметры, характеризующие свойства полупроводниковых материалов, приводятся в справочниках, в которых дано большое количество ссылок на первоисточники. Значения одних параметров, таких, как теплопроводность, постоянная решетки, плотность, уже твердо установлены, и дальнейшие исследования вряд ли могут внести в них какие-либо изменения; значения других параметров, таких, как глубина залегания примесных уровней или эффективные массы, связаны с более сложными измерениями и могут уточняться в результате современных исследований. Естественно, что для этих параметров желательно пользоваться самыми последними данными. Большинство приведенных здесь нами данных взято из работ [2, 8—10 (т. 2—7) и 11]. Много сведений о большом числе полупроводниковых материалов можно найти в трудах «материаловедческих» секций традиционных двухгодичных конференций по физике полупроводников, на которые мы часто ссылаемся.

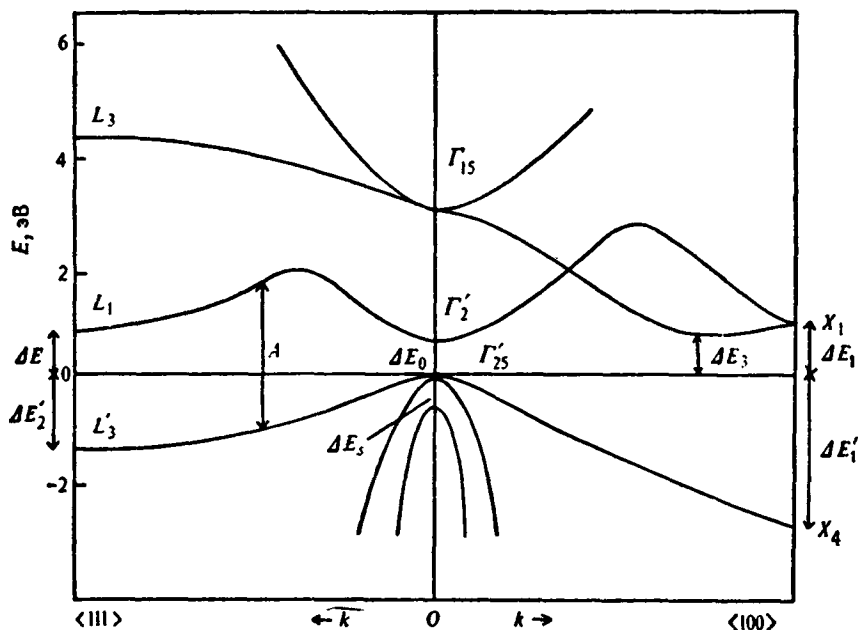


Рис. 13.1. Схематическое изображение зонной структуры Ge.

Ширины запрещенных зон (при 300 К) равны: $\Delta E = 0,665$ эВ, $\Delta E_0 = 0,805$ эВ, $\Delta E_3 = 0,28$ эВ, $\Delta E_1 = 1,2$ эВ, $\Delta E'_1 = 3,1$ эВ, $\Delta E'_2 = 1,45$ эВ, $\Delta E_5 = 0,845$ эВ. A обозначает переход, для которого $\Delta_k(E_c - E_v) = 0$ (ср. с рис. 10.24).

13.3. Кремний и германий

Материалы Si и Ge мы рассматриваем вместе, так как их свойства очень сходны, хотя между ними и есть некоторые интересные отличия. Оба являются непрямыми полупроводниками, и у того и у другого вершина валентной зоны расположена при $k=0$. В Si дно зоны проводимости расположено в направлениях $\langle 100 \rangle$ внутри зоны Бриллюэна ($0,8 k_{\max}$) и поэтому имеется шесть минимумов, в то время как в Ge дно зоны проводимости расположено в направлениях $\langle 111 \rangle$ в точках на краю зоны Бриллюэна, поэтому здесь имеется четыре минимума. Поверхности постоянной энергии и у того и у другого материала вблизи дна зоны проводимости имеют форму эллипсоидов вращения (см. разд. 2.3). У обоих полупроводников валентная зона при $k=0$ вырождена и имеет отщепленную зону, возникшую в результате спин-орбитального взаимодействия. На рис. 13.1, 13.2 схематически показаны формы зон в Ge и Si и указаны некоторые расстояния между зонами. Для Ge приведенные значения ширины запрещенных зон получены в результате эксперимента, точнее, сумма $\Delta E_1 + \Delta E'_1$ определена экспериментально,

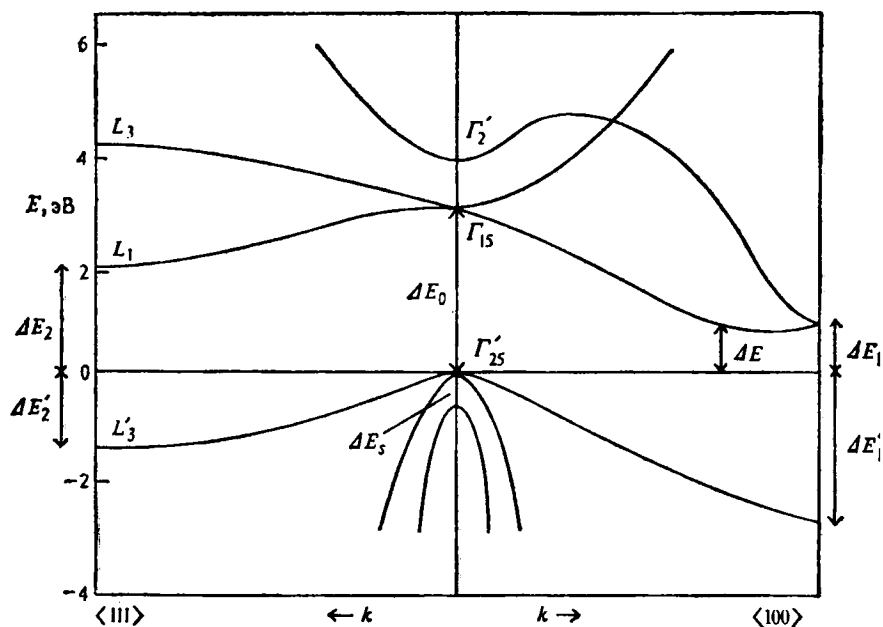


Рис. 13.2. Схематическое изображение зонной структуры Si.

Ширины запрещенных зон равны: $\Delta E = 1,12$ эВ, $\Delta E_0 = 3,4$ эВ, $\Delta E_3 = 0,035$ эВ, $\Delta E_1 = 1,2$ эВ, $\Delta E'_1 = 3,1$ эВ, $\Delta E_2 = 1,9$ эВ, $\Delta E'_2 = 2,2$ эВ.

а величины ΔE_1 и $\Delta E'_1$ получены в результате независимых вычислений методом псевдопотенциала¹⁾. Аналогично в Si суммы $\Delta E_2 + \Delta E'_2$ и $\Delta E_1 + \Delta E'_1$ определены экспериментально, однако сами эти величины получены в результате вычислений.

Продольная и поперечная эффективные массы, отнесенные к массе свободного электрона m_0 , а также отношение этих масс K_m имеют следующие значения:

	m_l/m_0	m_t/m_0	K_m
Ge	1,64	0,082	2,0
Si	0,98	0,19	5,1

Постоянные A , B , C для валентной зоны (см. уравнение (2.25)) имеют следующие значения:

	A	B	C
Ge	13,1	8,3	12,5
Si	4,0	1,1	4,1

¹⁾ См., например, [12]. Намного более подробно зонная структура Ge описана в работе [13].

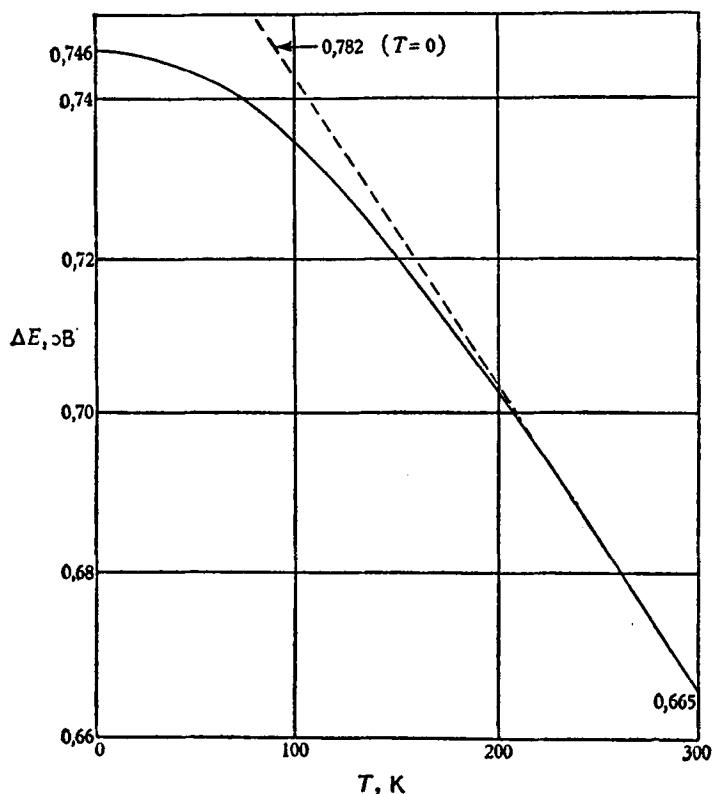


Рис. 13.3. Зависимость ширины запрещенной зоны ΔE в Ge от температуры T [14].

Для эффективных масс тяжелых m_{h1} и легких m_{h2} дырок, а также для дырок в отщепленной зоне (см. разд. 2.3) m_{h3} имеем

	m_{h1}/m_0	m_{h2}/m_0	m_{h3}/m_0
Ge	0,28	0,044	0,077
Si	0,49	0,16	0,245

На рис. 13.3 и 13.4 показано изменение ширины запрещенной зоны ΔE в зависимости от температуры для Ge и Si соответственно. Эти данные получили Макфарлан и др. [14] из измерений оптического поглощения. Можно видеть, что в широком диапазоне температур ΔE линейно зависит от T , однако при $T \rightarrow 0$ ΔE стремится к постоянной величине.

Ниже приведены подвижности электронов и дырок в очень чистых Ge и Si, величины которых обусловлены решеточным и между-

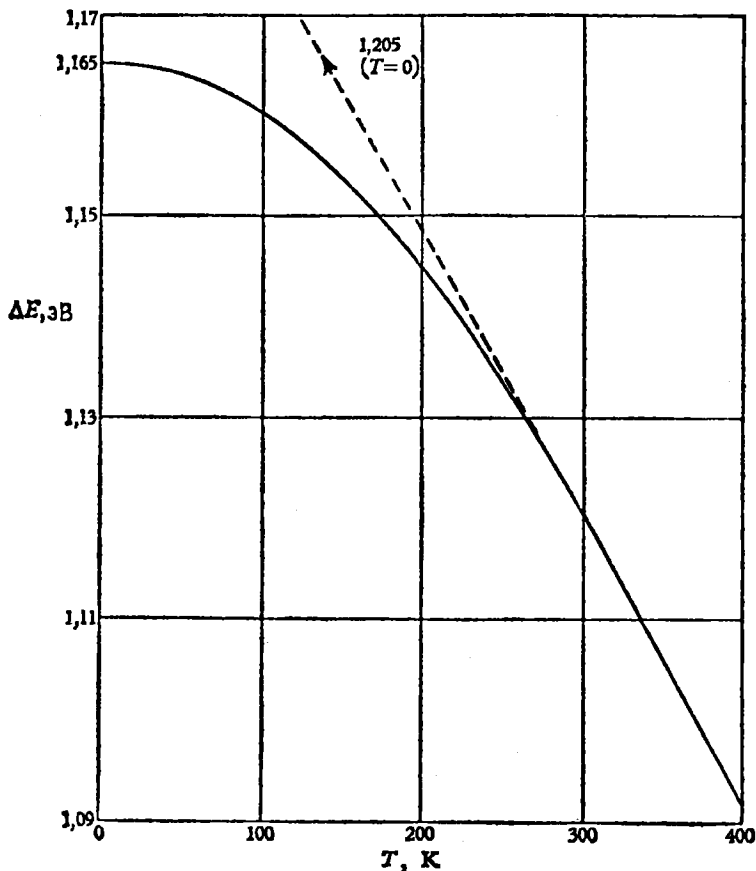


Рис. 13.4. Зависимость ширины запрещенной зоны ΔE_g в Si от температуры T [14].

линные типы рассеяния. Здесь приведены величины дрейфовых подвижностей, найденные из наблюдения дрейфа неравновесных носителей в электрическом поле, которые, однако, практически не отличаются от величин дрейфовых подвижностей, полученных из измерений коэффициента Холла и проводимости:

μ_{de} (290 K) $\text{cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$	μ_{dh} (290 K) $\text{cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$
Ge 3900 ($\sim T^{-1.6}$)	1900 ($\sim T^{-1.3}$)
Si 1350 ($\sim T^{-2.5}$)	480 ($\sim T^{-2.7}$)

Удельное сопротивление для собственного Ge $\rho_i = 47 \text{ Ом} \cdot \text{см}$, а для собственного Si $\rho_i = 2,5 \cdot 10^5 \text{ Ом} \cdot \text{см}$. Концентрации собственных но-

сителей заряда n_i равны

$$\begin{array}{ll} \text{Ge} & 2,25 \cdot 10^{13} \text{ см}^{-3} \quad (T = 290 \text{ K}) \\ \text{Si} & 7 \cdot 10^9 \text{ см}^{-3} \quad (T = 290 \text{ K}) \end{array}$$

Приготовление Ge, который при комнатной температуре был бы собственным материалом, является делом обычным, однако для Si эта задача сопряжена со значительными сложностями. Зависимость n_i от T приведена в разд. 4.2.

Ниже приведены положения энергетических уровней (в эВ) для некоторых наиболее распространенных примесей в Si и Ge. Знак «+» означает, что уровень расположен *над* вершиной валентной зоны, знак «-» означает, что уровень расположен *ниже* дна зоны проводимости (эВ).

Акцепторы	B	Al	Ga	In	Tl	
III группы						
Ge	+0,0104	+0,0102	+0,0108	+0,0112	+0,0100	
Si	+0,044	+0,069	+0,073	+0,155	+0,26	
Доноры V группы						
	P	As	Sb			
Ge	-0,0120	-0,0127	-0,0096			
Si	-0,044	-0,049	-0,039			
Другие примеси						
	Li	Cu	Ag	Au	Zn	Cd
Ge	-0,0093	+0,045	+0,13	+0,05	+0,03	+0,05
		+0,32	-0,39	+0,15	+0,09	+0,16
		-0,26	-0,029	-0,20		
Si	-0,033		-0,04			
				+0,39	+0,092	
			-0,30	+0,30		

Ge можно сплавить с Si в любых соотношениях, хотя приготовление монокристаллических образцов сплавов с соотношением концентраций Si и Ge, лежащим между 15% и 95%, затруднительно. Обзор свойств сплавов Ge/Si сделали Херман, Гликсман и Парменгер [15]. Подвижность электронов в таком материале уменьшается от значения в чистом Ge до значения $500 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ в сплаве, содержащем 15% Si, и затем по мере увеличения концентрации Si постепенно возрастает до величины, соответствующей подвижности электронов в чистом Si.

Расплавленные Si и Ge ведут себя как металлы, их удельное сопротивление растет при повышении температуры. При плавлении удельное сопротивление Ge уменьшается примерно в 20 раз, достигая величины около $6,5 \cdot 10^{-5} \text{ Ом} \cdot \text{см}$. Свойства жидкого Ge и Si изучил Доминикали [16].

13.4. Другие элементарные полупроводники

Помимо Ge и Si к элементарным полупроводникам относятся Se, Te и серое олово (α -Sn). Некоторые полупроводниковые свойства обнаруживает и В, однако подробно он не изучался. Полупроводниками могут быть и аллотропные формы других элементов, таких, как S, P и I, однако как полупроводники они изучались только в виде напыленных слоев [17].

Свойства Se описал Дженкинс ([8], стр. 49)¹⁾. Методы зонной очистки оказались для этого материала неэффективными, поэтому очистка Se явилась очень непростой задачей. Напротив, к Te может быть применен метод зонной очистки, что позволило изучать этот материал в условиях контролируемых концентраций примеси. При этом особое внимание было уделено выяснению различных механизмов рекомбинации носителей заряда [18]. Значительный теоретический интерес представляет зонная структура Te, в особенности ее сильная анизотропия.

После многих неудачных попыток к этому полупроводнику удалось, наконец, применить метод циклотронного резонанса (в основном это обусловлено появившейся теперь возможностью использовать более высокие частоты). Несмотря на то что Te всегда имеет проводимость p -типа и до сих пор легированием не удалось получить Te n -типа, циклотронный резонанс наблюдался как для термически возбужденных электронов в зоне проводимости, так и для дырок в валентной зоне. Эти эксперименты позволили определить эффективную массу электрона в зоне проводимости, она оказалась равной $0,135 m_0$. Спектр циклотронного резонанса для дырок, как оказалось, имеет сложную структуру с несколькими пиками. Это обусловлено взаимодействием спина электрона с двумя ветвями зоны, вырожденными в максимуме. Подробности этого спектра изучили Баттон и др. [19], используя длинноволновое инфракрасное излучение лазеров на HCN и D₂O длинами волн 0,337 мм и 0,1186 мм. В свете этих измерений Баттон ([20], стр. 253) изучил зонную структуру Te и обнаружил, что она находится в разумном согласии с результатами теоретических предположений.

Некоторое время назад были выполнены исследования серого олова (α -Sn) [17] и — в основном из измерений фотопроводимости и эффекта Холла — сделано заключение, что этот материал является полупроводником с запрещенной зоной шириной 0,08 эВ [21]. На основании результатов более поздних магнитооптических экспериментов теперь считается, что серое олово — это полуметалл, с шириной запрещенной зоны, равной нулю. Этот вариант зонной структуры предложили Гровс и Пол [22]. Недавние эксперименты,

¹⁾ Свойства селена подробно изучались в работах Абдуллаева и сотр. [88].—
Прим. ред.

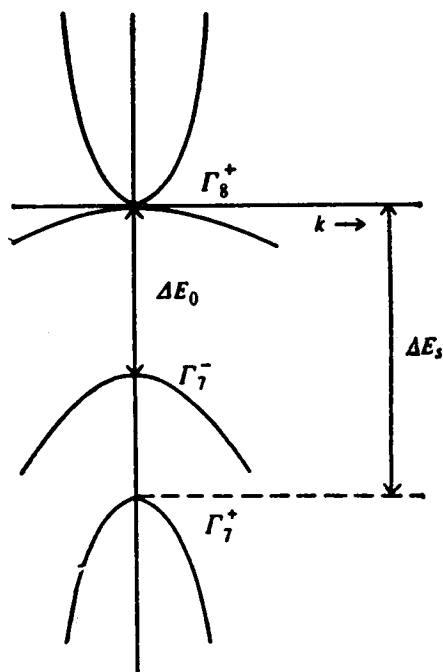


Рис. 13.5. Схематическое изображение зонной структуры α -Sn, обладающего типичной для полуметалла запрещенной энергетической зоной нулевой ширины. ΔE_s — величина спин-орбитального расщепления.

использующие методику магнитооптических измерений, также подтвердили это предположение ([23], стр. 47). В схематическом виде зонная структура α -Sn показана на рис. 13.5. Нижняя ветвь зоны Γ_8^+ , вырожденной при $k=0$, соответствует валентной зоне, а верхняя ветвь соответствует зоне проводимости. Валентная зона заполнена, однако между ней и зоной проводимости энергетический зазор отсутствует. Таким образом, серое олово является не полупроводником, а полуметаллом, (см. разд. 13.8). Пиджен ([23], стр. 47) определил энергетический зазор ΔE_0 до следующей более глубокой зоны при $k=0$, который составляет 0,413 эВ, а это слишком много, чтобы соответствовать ранее определенной ширине запрещенной зоны 0,08 эВ. К настоящему времени неясно, почему тщательные измерения, о которых сообщают Буш и Керн [21], могли дать величину энергии активации, равную 0,08 эВ. Эта величина определенно мала, но заведомо находится в пределах точности эксперимента. Не исключено, что образцы были изготовлены так, что они содержали включения примеси, с которых электроны могли быть возбуждены, минуя запрещенную зону с нулевой шириной, а это могло быть

истолковано, как наличие некоторой энергии активации. Другой возможной причиной этого несоответствия могут быть механические напряжения в кристаллах. Эвальд [24] сделал обзор работ по α -Sn и показал, что наличие механических напряжений может привести к появлению энергетического зазора между зоной проводимости и валентной зоной. Однако для получения зазора шириной в 0,08 эВ потребовалось бы очень сильное механическое напряжение.

13.5. Соединения типа $A^{III}B^V$

Свойства полупроводниковых соединений типа $A^{III}B^V$ впервые подробно изучали Уэлкер и его сотрудники. Обзор свойств этих материалов сделали Уэлкер и Вейсс [25], а также Маделунг [26]. Их свойства обсуждали также Хилсум и Роуз-Иннс [27], кроме того, с большими подробностями они обсуждены в периодическом сборнике [11]. Ульм ([8], стр. 115) довольно подробно рассмотрел свойства InSb и специально остановился на проблемах, связанных с выращиванием и очисткой кристаллов. В частности, он обсуждает вопрос о том, как избежать появления зеркальных граней при выращивании кристалла из расплава. Наличие такого эффекта (эффект граней) приводит к неравномерному распределению примесей. InSb оказался основным веществом для изучения возникновения полосатости в процессе роста кристалла, т. е. другого проявления неоднородного распределения примесей. Эти вопросы изучали Уитт и Гатос [28]. Естественно, что поскольку материалы типа $A^{III}B^V$ содержат не один, а два элемента, то приготовить их в чистом виде или с небольшим контролируемым содержанием примеси сложнее, чем элементарные полупроводники. К счастью, нет большой проблемы в получении материалов этого типа со стехиометрическим составом, так как эти соединения сами по себе образуют вещества с почти равным числом атомов каждого сорта.

Существует девять соединений типа $A^{III}B^V$, которые можно получить, комбинируя In, Ga, Al с Sb, As, P. Из всех соединений типа $A^{III}B^V$, InSb обладает наименьшей шириной запрещенной зоны, которая составляет 0,165 эВ при 300 К. Изменение ширины зоны с температурой полностью аналогично такому изменению для Si и Ge, т. е. оно линейно в широкой области температур, а вблизи абсолютного нуля ширина зоны стремится к постоянной величине (0,23 эВ). В InSb оптический переход, соответствующий минимальной ширине запрещенной зоны, является «прямым». На рис. 10.7 показан очень крутой рост поглощения вблизи фундаментального края.

Из девяти двойных полупроводниковых соединений типа $A^{III}B^V$ пять обладают «прямой» запрещенной зоной, а четыре — «непрямой». Величины минимальных значений ширины запрещенных зон ΔE

(обозначены как «П» — прямая или «Н» — непрямая) приведены в табл. 13.1. Отметим высокую подвижность электронов в InSb. При низких температурах она может превышать $10^6 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$.

Таблица 13.1

Свойства полупроводников типа $A^{III}B^V$ (при 300 К)

	ΔE , эВ	m_e/m_0	m_{h1}/m_0	μ_e , $\text{см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$	μ_h , $\text{см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$
InSb	0,18 П	0,014	0,4	78000	750
InAs	0,35 П	0,022	0,4	28000	450
GaSb	0,72 П	0,044	0,23	44000	700
InP	1,33 П	0,078	—	5200	150
GaAs	1,42 П	0,065	0,5	8900	400
AlSb	1,62 Н	—	—	200	—
AlAs	2,13 Н	—	—	—	—
GaP	2,26 Н	—	—	200	100
AlP	2,43 Н	—	—	—	—

Хорошее приближение к идеальному полупроводнику, имеющему очень малую скалярную эффективную массу электронов в зоне проводимости и не имеющему других минимумов энергии вблизи дна зоны проводимости, представляет собой InSb. Зонная структура InSb схематически показана на рис. 13.6. Обозначение различных максимумов и минимумов является общепринятым и, так же как для Si и Ge, следует из теории групп. Эта схема не совсем точна. Если выполнить вычисление формы зон в InSb вблизи $k=0$ с еще большей точностью, то обнаружится дополнительное расщепление, в результате которого по мере роста k от нуля энергия, соответствующая зоне тяжелых дырок, сначала возрастает, а уже затем начинает уменьшаться. Как показал Пиджен ([23], стр. 47), максимальная величина подъема составляет при этом всего около 1 мэВ. Хотя зона проводимости и является сферической, из-за очень малой величины эффективной массы электрона в зоне возникает отклонение зависимости $E(k)$ от квадратичной.

Зонная структура GaAs имеет очень похожую схему, однако в данном случае в зоне проводимости присутствует дополнительный минимум, который поднят всего на 0,35 эВ выше минимума, расположенного при $k=0$. Значения ширин основных запрещенных зон определены экспериментально, остальные ширины получены в результате расчетов [29].

Значения ширин запрещенных зон некоторых других из оставшихся полупроводников типа $A^{III}B^V$, а также их изменения с температурой довольно подробно приведены в справочнике ([10], т. 7).

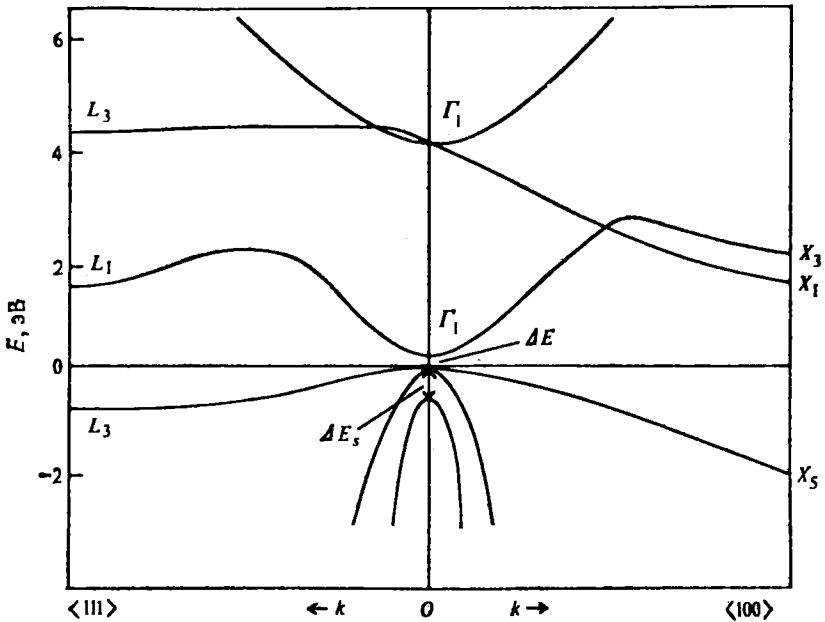


Рис. 13.6. Схематическое изображение зонной структуры InSb.

Ширины запрещенных зон (при 300 К) равны: $\Delta E = 0,18$ эВ, $\Delta E_s = 0,9$ эВ.

Все бинарные соединения могут быть сплавлены друг с другом, в результате получают тройные полупроводниковые соединения, например, такие, как $\text{GaP}_x\text{As}_{1-x}$. Изменение свойств этих материалов в зависимости от параметра x представлено в виде подробных таблиц в том же справочнике ([10], т. 7). Зонную структуру этих сплавов и их основные свойства подробно рассмотрел Лонг [30]. Например, небезынтересно заметить, что при изменении параметра x от 0 до 1 материал $\text{GaP}_x\text{As}_{1-x}$ трансформируется из полупроводника с непрямой запрещенной зоной в полупроводник с прямой запрещенной зоной, причем точке перехода соответствует значение $x = 0,44$.

Для двойных полупроводниковых соединений типа $A^{\text{III}}B^{\text{V}}$ несомненную важность имеет величина эффективного ионного заряда e^* . Эта величина является мерой того, насколько ионный или ковалентный характер имеет данная связь. Если $e^* = e$, то в данном случае связь оказывается чисто ионной, в то время как при $e^* = 0$ связь является полностью ковалентной, как в элементарных полупроводниках Si и Ge. В бинарных полупроводниковых соединениях типа $A^{\text{III}}B^{\text{V}}$ отношение e^*/e изменяется от величины 0,33 для GaSb до 0,68 для InP, тогда как для GaAs оно равно 0,5. Этот факт

указывает на то, что в GaAs связь является наполовину ионной, наполовину ковалентной (см. разд. 3.2). (Определение e^* см. в разд. 8.3, также в разд. 14.3.)

Для некоторых двойных полупроводников типа $A^{III}B^V$ необычную величину имеет такой параметр, как g -фактор (см. разд. 12.6). В InSb g -фактор имеет большое значение, равное -40 (оно слегка зависит от содержания примесей). Это означает, что спиновое расщепление в магнитном поле очень велико, поэтому InSb является подходящим материалом для использования в лазерах, перестраиваемых магнитным полем (см. разд. 14.7). В InAs величина $g = -12$, а в GaSb $g = -6$. В остальных соединениях g обычно положительно и его значение несколько меньше значения g для свободного электрона, равного 2.

Глубины залегания примесных уровней, а также природа этих уровней известны далеко не так хорошо, как в Si и Ge. Это и неудивительно ввиду намного большего числа возможных ситуаций. Например, может существовать два типа вакансий, два типа междоузельных атомов, а также всевозможные их комбинации. Замещение атомов может происходить также различными способами.

В этих соединениях элементы II группы Zn и Cd замещают атомы III группы. В InSb они создают акцепторы с энергией ионизации порядка 0,0075 эВ. Элементы VI группы S, Se, Te замещают атомы V группы и образуют доноры. Из-за малости эффективной массы электрона в InSb в соответствии с простой водородоподобной моделью эти доноры должны иметь очень малую энергию ионизации порядка 0,0007 эВ и соответственно огромный боровский радиус. За исключением случаев очень малой концентрации, они будут перекрываться, и их уровни энергии сольются с зоной проводимости. Неионизованные примеси такого типа наблюдались только в условиях очень сильного магнитного поля, которое приводит к увеличению их энергии связи (см. разд. 12.9). Cu, Ag и Au дают в InSb двойные акцепторы, в то время как замещение другими атомами III или V групп, видимо, приводит к возникновению нейтральных примесей. В GaAs эти атомы дают акцепторы с энергией связи 0,034 эВ, тогда как Ge дает донорный уровень с энергией связи 0,006 эВ. Cu и Ag ведут себя как акцепторы, однако склонны к образованию комплексов с другими несовершенствами кристаллической решетки.

Благодаря изучению излучения экситонов, связанных на примесях (см. разд. 10.14, 14.1) и дающих спектры с узкими линиями, довольно много известно о глубоких примесных уровнях в GaP (см. [31]). Например, пара близко расположенных примесных атомов, состоящая из одного донора и одного акцептора, образует двойную примесь; в частности, такой парой может быть Te (D) и Si (A). O и N образуют глубокие примесные уровни в GaAs и GaP, причем обычно в комбинации с другими примесями.

13.6. Соединения типа $A^{II}B^{VI}$

Из соединений типа A_2B_6 наиболее подробно изучались халькогениды свинца PbS , $PbSe$ и $PbTe$. Они настолько похожи друг на друга, что мы будем рассматривать их вместе. Их свойства довольно подробно описали Патли [8], стр. 71) и Скэнлон [32].

После того как первые расчеты показали, что $PbTe$ должен быть непрямым полупроводником [33], относительно зонной структуры этих полупроводников возникло много споров. В более поздних вычислениях, которые провели Конклин, Джонсон и Пратт [34], а также Лин и Клейнман [35], учитывались релятивистские поправки к волновым функциям, обусловленные наличием тяжелых атомов, и было показано, что во всех трех соединениях абсолютный минимум зоны проводимости и абсолютный максимум валентной зоны находятся на краю зоны Бриллюэна в направлениях $\langle 111 \rangle$. Поэтому в них имеется четыре максимума и четыре минимума, так что все эти материалы являются прямыми полупроводниками, однако энергетический зазор минимальной ширины расположен *не* при $k=0$. Поверхности постоянной энергии вблизи дна зоны проводимости и вблизи вершины валентной зоны имеют форму эллипсоидов вращения. Поэтому такие соединения отличаются от Si и Ge , а также от полупроводников типа $A^{III}B^V$ и образуют очень интересную и единственную в своем роде группу материалов. Все они имеют кристаллическую решетку хлористого натрия.

Зонная диаграмма для $PbTe$ схематически изображена на рис. 13.7. Намного более подробно зонную структуру описали Лин и Клейнман (см. выше). В этой же работе излагаются довольно сложные вычисления и система обозначений максимумов и минимумов согласно теории групп. В табл. 13.2 приведены ширины запре-

Таблица 13.2

Ширины запрещенных зон халькогенидов свинца

	T, K	$\Delta E, эВ$
PbS	290	0,41
	77	0,31
	4	0,29
$PbSe$	290	0,27
	77	0,17
	4	0,15
$PbTe$	290	0,32
	77	0,22
	4	0,19

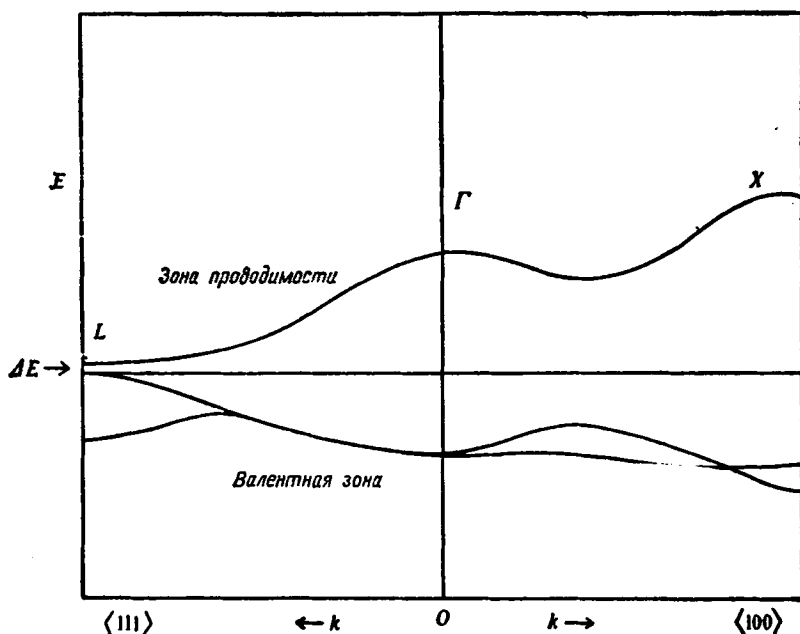


Рис. 13.7. Упрощенное схематическое изображение зонной структуры PbTe (не в масштабе).

PbS и PbSe имеют аналогичные зонные структуры.

щенных зон ΔE в этих материалах при различных температурах. Следует отметить, что в отличие от Si и Ge и соединений типа $A^{III}B^V$, ширина запрещенной зоны ΔE в этих материалах *уменьшается* при понижении температуры. Кроме того, вопреки ожиданию, PbSe имеет меньшую ширину запрещенной зоны, чем PbTe. Митчел, Палик и Земель [36] показали, что значения ширин запрещенных зон в этих материалах чувствительны к наличию механических напряжений в кристалле, и нашли величины эффективных масс на основе измерений магнитооптического поглощения. Значения, полученные ими, слегка отличаются от значений, полученных другими авторами. Мосс, Баррел и Эллис [37] проанализировали результаты различных измерений и нашли величины эффективных масс в этих материалах для температуры 77 К, которые представлены в табл. 13.3. В таблице можно видеть, что K_m велико только в PbTe, тогда как отклонение от «сферичности» в PbS совсем мало. Видно также, что параметры зоны проводимости и валентной зоны одинаковы. При 77 К это действительно почти так, однако при низких температурах возникают некоторые различия. Например, в PbTe при 4 К для зоны проводимости $K_m = 10$, а для валентной зоны $K_m = 14$.

Таблица 13.3

Эффективные массы носителей заряда в PbS, PbSe и PbTe

		m_T	m_L	K_m
PbS	Зона проводимости	0,10	0,11	1,1
	Валентная зона (дырки)	0,10	0,11	1,1
PbSe	Зона проводимости	0,05	0,08	1,6
	Валентная зона (дырки)	0,05	0,08	1,6
PbTe	Зона проводимости	0,025	0,22	8,7
	Валентная зона (дырки)	0,025	0,22	8,7

Причины возникновения аномального значения ширины запрещенной зоны у PbSe обсуждали различные авторы, однако в конце концов этот факт объяснили Берник и Клейнман [38]; они показали, что симметрия зоны проводимости в PbSe отличается от симметрии в PbS и PbTe. Величины эффективных масс, вычисленные с учетом этой поправки, находятся в лучшем согласии с экспериментальными данными, чем величины, полученные в предыдущих расчетах.

Митчел, Палик и Земель (см. выше) получили следующие значения g -фактора для этих соединений: для PbS $g=10,0$ (зона проводимости) и $g=8,5$ (валентная зона); для PbSe $g=22$ (зона проводимости) и $g=22$ (валентная зона).

Сведения о примесных уровнях в этих полупроводниках немногочисленны, главным образом из-за трудностей получения кристаллов стехиометрического состава. Избыточный атом Pb ведет себя как донор с очень неглубоким уровнем, в то время как атомы S, Se, Te в PbS, PbSe, PbTe соответственно ведут себя как акцепторы с мелким уровнем. Блоэм и Крёгер [39] довольно подробно исследовали эти вопросы и показали, что в PbS в точке максимальной температуры плавления имеется $6 \cdot 10^{11}$ см⁻³ избыточных атомов Pb. Висмут ведет себя как донор и может быть использован для компенсации материала p -типа, в то время как Ag и Cu ведут себя как акцепторы, когда они замещают атомы Pb. С другой стороны, в междоузельном состоянии атомы Ag и Cu могут вести себя как доноры.

Блоэм и Крёгер (см. выше) изучали также равновесие между вакансиями (V_c^+) и добавочными атомами Vi и показали, что при определенных условиях число вакансий может в точности сравняться с числом добавочных атомов Vi и возникнет ситуация, известная под названием компенсации валентности.

Все три соединения имеют большие величины показателя преломления. В ближнем инфракрасном диапазоне они равны 4,1 для PbS, 4,85 для PbSe и 5,65 для PbTe. Они довольно медленно растут при понижении частоты, однако имеются сообщения о необычайно сильном увеличении показателя преломления при очень низких частотах, особенно для PbTe. Однако точные его значения пока нельзя считать до конца установленными. Кохрэн [40] дает для PbTe величину 17,6. Свойства этой интересной группы полупроводников подробно обсуждает в своем обзоре Далвей [41].

Группа полупроводников CdS, CdSe и CdTe изучалась не в таком объеме, как халькогениды свинца. Величины ΔE в них равны 2,4, 1,75 и 1,51 эВ соответственно. CdS использовался в качестве типичного полярного полупроводника для исследования свойств как анионных, так и катионных вакансий и их связи с концентрацией свободных носителей каждого типа. Физико-химические свойства этих материалов, а также свойства соединений на основе свинца подробно изучали Крёгер, Винк и ван ден Бумгард [42]. CdS широко применялся и для исследования экситонных спектров (см. разд. 10.6). В результате энергетические уровни некоторых примесей известны очень точно, однако однозначно идентифицировать дефекты удается не всегда.

CdS является также одним из основных материалов для изучения электроакустического эффекта (см. подразд. 12.4.1). Использование этого материала в акустических линиях задержки, а также его широкое использование в измерителях интенсивности света послужило стимулом для изучения этого материала с точки зрения полупроводниковых свойств. Изучался и ряд других полупроводниковых соединений, включающих S, Se и Te, хотя и намного менее подробно. Например, Bi_2Te_3 привлек к себе значительное внимание в качестве термоэлектрического материала. Его свойства описал Голдсמיד ([8], стр. 165). Обзор свойств соединений CdSb и ZnSb дал Блэкуэлл ([8], стр. 199).

Обзор свойств и методов приготовления некоторых полярных полупроводников, и особенно халькогенидов свинца, сделал Скэнлон [32]. Он рассмотрел также возникновение вакансий и отклонений от стехиометричности. Более современный обзор свойств полупроводниковых соединений типа $A^{II}B^{VI}$ можно найти в сборнике под редакцией Томаса [44].

13.7. Тройные и четверные соединения

Все рассмотренные выше соединения были двойными, за исключением смешанных кристаллов, которые на самом деле являются просто сплавами. Некоторое время рассматривалась возможность использования тройных или даже четверных полупроводниковых

соединений, однако в этом направлении был достигнут лишь незначительный прогресс, в основном из-за сложностей в приготовлении и очистке этих соединений. Поскольку поддерживать стехиометрическое равновесие трудно даже при наличии только двух элементов, не вызывает удивления тот факт, что при наличии трех элементов задача еще более усложняется. Хан и др. [45], а также Гудмэн и Дуглас [46] обсуждали свойства таких соединений и описали принципы, на которых основано получение еще более сложных полупроводниковых соединений. Здесь так же, как и в случае двойных соединений, используются представления о компенсации валентностей и образовании замкнутых восьмизлектронных оболочек.

Было приготовлено несколько соединений этого типа и показано, что они являются полупроводниками, однако подробной информации об их свойствах пока почти нет. Примерами таких соединений могут быть CuFeS_2 , CuInSe_2 и AgInSe_2 . Первое принадлежит к классу так называемых халькопиритов, в которых уже давно было обнаружено явление выпрямления переменного тока в точечном контакте с металлом. Кроме того, Гудмэн и Дуглас нашли, что это вещество прозрачно в инфракрасном диапазоне при длине волны более 2 мкм, так что предположительно оно имеет запрещенную зону шириной около 0,5 эВ; его кристаллическая структура очень близка к структуре алмаза. В этой структуре атом Cu одновалентен, а атом Fe трехвалентен, так что возникают четыре электрона, которые вместе с двенадцатью электронами от S_2 образуют две замкнутые восьмизлектронные оболочки. Четверные соединения типа $\text{Cu}_2\text{FeSnS}_4$, которые называют станнитами, также должны проявлять полупроводниковые свойства, и по аналогии Гудмэн и Дуглас предположили, что соединения типа $\text{Cu}_2\text{CdSnTe}_4$ также должны быть полупроводниками; в число новых соединений, которые приготовили и исследовали Гудмэн и Дуглас, входят CuInTe_2 , CuInSe_2 и CuAlTe_2 . Они обладают сильно выраженными выпрямляющими свойствами, имеют оптические запрещенные зоны шириной около 0,9 эВ и электронные подвижности порядка $300 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$. Попытки получить монокристаллы этих веществ пока успехом не увенчались.

Очевидно, если рассматривать тройные и четверные соединения, число возможных полупроводниковых соединений становится очень большим. Моозер и Пирсон [47] выдвинули некоторые предложения по расширению принципа скомпенсированной валентности, о котором упоминалось выше.

Некоторые тройные соединения были исследованы после обнаружения у них нелинейных оптических свойств, причем их выбор основывался на принципе скомпенсированной валентности. Наибольшее внимание из них привлекли к себе халькопириты. Они принимают две основные формы, которые можно обозначать как 245_2 и 136_2 (строчные цифры обозначают группу периодической системы

элементов, а индексы — число атомов в соединении). Первую форму можно рассматривать как расширение класса соединений типа $A^{III}B^V$, вторую — как расширение класса соединений типа $A^{II}B^{VI}$. Приведем примеры соединений типа 245_2 :

$CdGeAs_2$	$\Delta E = 0,55$ эВ,
$CdSnP_2$	$\Delta E = 1,15$ эВ,
$ZnGeP_2$	$\Delta E = 2,2$ эВ,

и примеры соединений типа 136_2

$Ag\ In\ Se_2$	$\Delta E = 1,2$ эВ,
$Ag\ In\ S_2$	$\Delta E = 2,0$ эВ,
$Ag\ Ga\ S_2$	$\Delta E = 2,7$ эВ,
$Cu\ Al\ S_2$	$\Delta E = 3,5$ эВ.

Быстро появилась обширная литература, посвященная этим материалам. Свойства некоторых из них рассмотрел Хилсум ([2], стр. 69). Р. К. Смит [48] обсудил использование этих материалов в оптоэлектронике, дал список величин ΔE для некоторых халькопиритов того и другого типа и сведения о том, является ли запрещенная зона в них прямой или не прямой. Кроме того, он рассмотрел их оптические свойства и, в частности, использование их нелинейных характеристик в целях оптического смещения частот.

Интересной особенностью этих материалов является то, что многие из них имеют прямые запрещенные зоны и поэтому могут оказаться полезными в области люминесцентных применений и для создания лазеров (см. разд. 14.6). И действительно, в некоторых из них лазерный эффект был обнаружен.

Проблемы приготовления кристаллов, которые при первых исследованиях этих материалов казались непреодолимыми, к настоящему времени частично удалось разрешить, и в некоторых лабораториях уже получены большие монокристаллы с довольно хорошими оптическими свойствами. Методики их приготовления, используемые в настоящее время, описал Фейгельсон [49]. В частности, он рассмотрел выращивание кристаллов $CdGeAs_2$, $AgGaSe_2$ и $AgGaS_2$ для оптических применений. Выращивание кристалла $AgGaS_2$ высокого качества описали также Маттис и др. [50].

В своей книге [51], полностью посвященной тройным полупроводникам — халькопиритам, Шей и Уэрник рассмотрели известные свойства большого числа соединений. Об электрических свойствах этих соединений известно намного меньше, чем об оптических, однако объем информации по этому вопросу в настоящее время быстро возрастает. Многие соединения типа 245_2 при комнатной температуре имеют подвижности электронов или дырок порядка $100\text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ или менее, однако некоторые, например $CdGeAs_2$, имеют значительно более высокие подвижности (порядка $1500\text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$). Подвижности в соединениях типа 136_2 оказы-

ваются намного ниже, примерно ($1 - 20$) $\text{см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$, за исключением AgInSe_2 , в котором $\mu_c \approx 750 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$. Электрические свойства, полученные на основе измерений эффекта Холла, проводимости и других измерений, подробно описали Шэй и Уэрник (см. выше). Этому классу полупроводников посвящен целый том журнала *Journal de Physique* (148—501).

Зонная структура этих соединений имеет некоторые интересные отличия от зонной структуры соответствующих полупроводниковых соединений типа $A^{III}B^V$ и $A^{II}B^{VI}$. В основном эта разница обусловлена тем, что тетраэдрическая связь здесь слегка нарушена и что элементарная ячейка в два раза больше элементарной ячейки двойных соединений. Использование давления для усиления отклонений от тетраэдрической симметрии позволило прояснить довольно сложную зонную структуру этих материалов. Для определения некоторых критических точек в зонной структуре с успехом использовался также метод спектроскопии электроотражения. Эти методы рассмотрели Бендориус и др. [52] и Шилейка [53]. О результатах вычислений зонной структуры сообщали различные авторы, в том числе Полыгалов, Поплавной и Ратнер [54], а также Пейсман и др. [55]. Хотя эти результаты и не полностью совпадают с экспериментальными данными, тем не менее они оказали большую помощь в выяснении характера зонной структуры между критическими точками, так что теперь для некоторых из этих соединений мы располагаем очень хорошей зонной схемой.

13.8. Очень узкозонные полупроводники

Много лет велись поиски полупроводников с запрещенной зоной шириной порядка или менее $0,1$ эВ. До 1959 г. единственным таким материалом был α -Sn. Теперь оказывается, что его запрещенная зона может не быть «собственной» запрещенной зоной (см. разд. 13.4). Поиски подталкивались потребностью в инфракрасных детекторах, чувствительных к излучению с длиной волны более 10 мкм. Единственное простое полупроводниковое соединение, ширина запрещенной зоны которого в какой-то мере близка к значению $0,1$ эВ, — это InSb (ширина его запрещенной зоны при комнатной температуре равна $0,18$ эВ). Существуют, однако, полуметаллы, имеющие нулевую или очень малую «отрицательную» ширину запрещенной зоны, которые можно сплавить с полупроводниками, имеющими запрещенную зону несколько более широкую, чем $0,1$ эВ. Лоусон и его сотрудники предположили, что если сплавить такой полупроводник с небольшим количеством полуметалла, то ширина запрещенной зоны полупроводника должна уменьшиться. Это подтвердилось на опыте и было продемонстрировано [56] на примере сплава полуметалла HgTe с полупроводником CdTe , обладающим

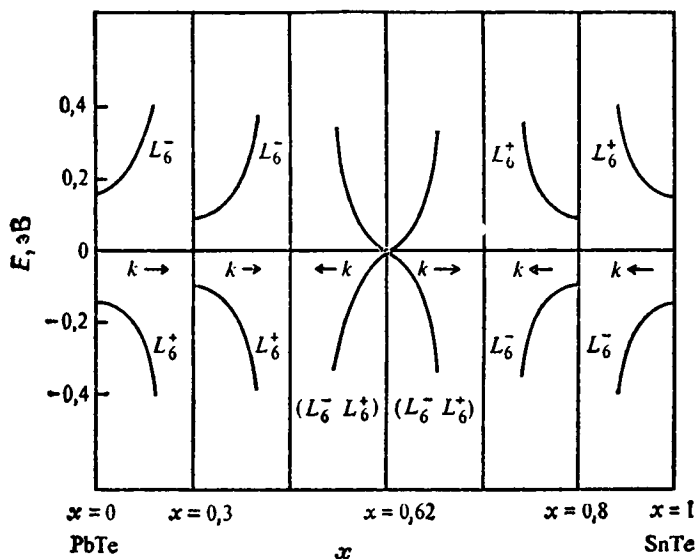


Рис. 13.8. Иллюстрация обращения зон L_6^- и L_6^+ в $Pb_{1-x}Sn_xTe$ при изменении x от 0 до 1 [60].

запрещенной зоной шириной $\Delta E = 1,5$ эВ. Изучение этого интересного сплава, так же как и разработка приборов на его основе, сдерживались сложностями в приготовлении монокристаллов высокого качества с однородным составом. Бартлетт, Динс и Эллен [57] сумели получить более совершенные кристаллы, которые были использованы не только для изучения этого полупроводникового материала, но и в качестве основы для создания инфракрасных детекторов [58] для медицинских и других применений, требующих наблюдения тел при температуре, при которой максимум излучения лежит в диапазоне длин волн около 10 мкм. Хотя в HgTe нулевая запрещенная зона расположена при $k=0$, при увеличении k валентная зона слегка приподнимается, так что возникает перекрытие с зоной проводимости, или, иначе говоря, возникает запрещенная зона с «отрицательной» шириной. При добавлении CdTe зона проводимости в HgTe приподнимается, однако появление настоящей запрещенной зоны не происходит, пока содержание CdTe не достигнет примерно 10%.

Как показали Шмит и Штельцер [59], ширина запрещенной зоны в $Hg_xCd_{1-x}Te$ изменяется почти линейно от $-0,1$ до $1,5$ эВ при изменении x от 0 до 1.

Проводились разнообразные измерения эффективной массы носителей заряда при различных концентрациях Hg. Обзор этих работ сделали Мосс, Баррелл и Эллис ([37], стр. 398).

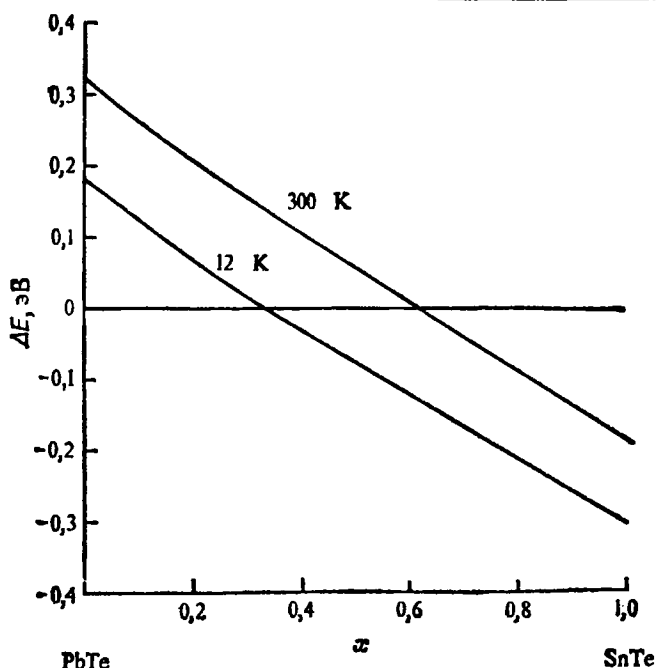


Рис. 13.9. Зависимость ширины запрещенной энергетической зоны от состава соединения $Pb_{1-x}Sn_xTe$ [60].

В результате исследований полупроводников группы $PbTe$ и $SnTe$ (см. разд. 13.6) Диммок, Мелнгейлис и Штраусс [60] предложили другой метод получения полупроводников с очень малой и контролируемой шириной запрещенной зоны. Они заметили, что зоны в $PbTe$ и $SnTe$ очень похожи друг на друга, однако расположение наинизшей зоны проводимости и наивысшей валентной зоны в одном материале *перевернуто* относительно соответствующего расположения в другом материале. Это проиллюстрировано на рис. 13.8, на котором показана упрощенная схема зон с использованием обычных обозначений согласно теории групп. Можно видеть, что зона L_6^- является в $PbTe$ зоной проводимости, а в $SnTe$ — валентной зоной. Аналогично зона L_6^+ является в $PbTe$ валентной зоной, а в $SnTe$ — зоной проводимости. Было обнаружено, что эти два материала сплавляются друг с другом, и поэтому следует ожидать, что при некотором составе сплава валентная зона и зона проводимости сомкнутся, затем они поменяются ролями и при дальнейшем изменении состава сплава начнут снова расходиться. При критическом составе, который соответствует 62% $SnTe$ при комнатной температуре и 34% при 4,2 К, ширина запрещенной зоны равна нулю. Это проиллюстрировано на рис. 13.9.

Как мы уже видели, в чистом $PbTe$ при уменьшении температуры ширина запрещенной зоны уменьшается. В $SnTe$, как и в большинстве других полупроводников, при понижении температуры ширина запрещенной зоны увеличивается. Это согласуется с зависимостью, представленной на рис. 13.9.

Такой материал может быть выращен в виде хороших монокристаллов, и путем тщательного контроля за содержанием Te можно добиться получения материала как n -, так и p -типа. Как показали Мелнгейлис и Харман [61], подвижности как электронов, так и дырок в этом материале довольно высоки. В материале, содержащем 17% $SnTe$, μ_e и μ_h при 77 К превышают $30\,000\text{ см}^2 \cdot V^{-1} \times \times c^{-1}$. На основе измерений циклотронного резонанса Эллис и Мосс [62] показали, что эффективные массы электронов m_{e^*} и дырок m_{h^*} в этом материале довольно малы и равны примерно $0,014 m_0$, причем $K_m = 7$. Эти величины можно получить теоретически, исследуя зонную структуру вблизи L -точки. Такие вычисления с помощью $k \cdot p$ -метода (см. подразд. 11.3.5) выполнил Диммок [63], стр. 319). Использование этого материала для создания инфракрасных детекторов описали Мосс, Баррелл и Эллис ([37], стр. 390).

Вследствие малости эффективной массы электрона в этих материалах, несмотря на то что зона проводимости еще может оставаться «сферической», зависимость E от k будет сильно отличаться от параболической. Влияние таких отклонений на явления переноса подробно обсудил Завадский [64]. Общий обзор свойств этих интересных новых материалов сделан в сборнике [63], гл. 6—8.

13.9. Оксидные полупроводники

Некоторые окислы металлов ведут себя как полупроводники с относительно широкими запрещенными зонами. Из них лучше всего были изучены Cu_2O и ZnO , если не брать в расчет окислы переходных металлов, такие, как NiO , свойства которых представляют интереснейшие теоретические проблемы и, следовательно, привлекают к себе больше внимания. Обычно они имеют незаполненные d -оболочки, поэтому следовало бы ожидать, что они будут вести себя как металлы. Некоторые окислы, в особенности ReO_3 , именно так себя и ведут, показывая непрерывный рост сопротивления при повышении температуры, причем сама величина удельного сопротивления окисла всего в десять раз больше, чем у благородных металлов. С другой стороны, чистый NiO ведет себя как диэлектрик с шириной запрещенной зоны $\Delta E = 3,7$ эВ. Гораздо удивительнее поведение таких окислов, как VO_2 и V_2O_3 , которые при высоких температурах имеют металлическую проводимость, однако при понижении до некоторой критической температуры их проводимость

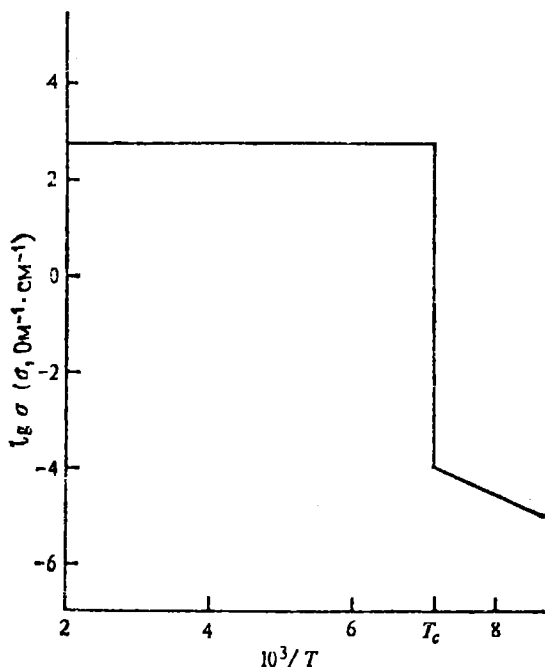


Рис. 13.10. Зависимость проводимости σ от температуры T (К) для окисла переходного металла, в котором имеет место переход металл — полупроводник. Указанные на рисунке величины приблизительно соответствуют характеристикам V_2O_5 .

резко падает более чем в 10^6 раз. При низких температурах они ведут себя подобно полупроводникам (рис. 13.10).

Совершенно ясно, что столь необычное поведение полностью противоречит теории, основанной на блоховских волновых функциях в кристалле. Можно себе представить, что за счет перекрытия зон материал с заполненными зонами окажется металлом, однако не так просто понять, каким образом материал с частично заполненной зоной (в нашем случае это d -зона) может вести себя как полупроводник или даже как превосходный диэлектрик (например, MnO).

Было предпринято несколько попыток разрешить этот парадокс. Мотт [65] предположил, что в случае слабого перекрытия между соседними атомами следует пользоваться не блоховскими, а локализованными волновыми функциями, так как обычное описание не учитывает эффекты корреляции между электронами. Если этими эффектами пренебречь, то твердый водород должен стать металлом, и действительно, такая фаза была предсказана для твердого водорода в условиях высоких давлений, когда атомы находятся близко

друг к другу. При нормальных давлениях корреляционные эффекты заставляют электроны и атомы спариваться с образованием молекул H_2 , так что твердый H_2 является молекулярным кристаллом. Мотт предположил, что аналогичные корреляционные эффекты приводят к локализации электронов в d -зонах в окислах типа V_2O_3 , так что для участия этих электронов в проводимости требуется некоторая энергия активации. Однако при критической температуре это состояние нарушается и должен произойти резкий переход к металлическому состоянию. Такой переход обычно называют переходом Мотта. Доказательство того, что он действительно имеет место в окислах переходных металлов, в настоящее время вызывает сомнения, так как появляются другие возможности для объяснения такого резкого перехода.

Было обнаружено, что как в V_2O_3 , так и в VO_2 происходит изменение кристаллической структуры; соответствующее ему изменение зонной структуры может явиться причиной для перехода от металлических свойств к полупроводниковым. Вопрос о том, является ли это изменение кристаллической структуры следствием перехода Мотта, остается до сих пор открытым. В обоих веществах в результате изменения кристаллической структуры должна возникнуть элементарная ячейка вдвое большего размера и, следовательно, зона Бриллюэна вдвое меньшего размера. Кроме того, как показал Адлер и Брукс [66], d -зона при этом расщепляется.

Это проиллюстрировано на рис. 13.11. Расщепленная d -зона соответствует низкотемпературной ситуации, переход является таким, который обычно называют фазовым переходом первого рода, он происходит резко, в очень узком температурном диапазоне. Это обычно указывает на то, что при температурах выше температуры перехода d -зона заполнена наполовину, так что уровень Ферми расположен вблизи середины зоны, и материал должен демонстрировать металлическую проводимость. При температурах ниже температуры перехода возникнет запрещенная зона конечной ширины, причем нижняя зона окажется полностью заполненной, так что материал должен вести себя как полупроводник. Более того, следует ожидать, что температура перехода должна зависеть от давления. Влияние давления на проводимость и магнитные свойства VO_2 и V_2O_3 изучали Фейнлиб и Пол [67].

Схема зонной структуры, изображенная на рис. 13.11, очень сильно упрощена. Каждая ветвь расщепленной d -зоны — двойная, верхняя и нижняя зоны отстоят друг от друга недалеко, но не настолько близко, чтобы объяснить наблюдаемую на опыте величину энергии активации. Адлер и др. [68] тщательно сравнили имеющиеся экспериментальные данные с теорией Адлера и Брукса (см. выше) и обнаружили, что они находятся в хорошем согласии с предположениями, основанными на гипотезе изменения кристаллической структуры. Менее вероятно, что этот переход является след-

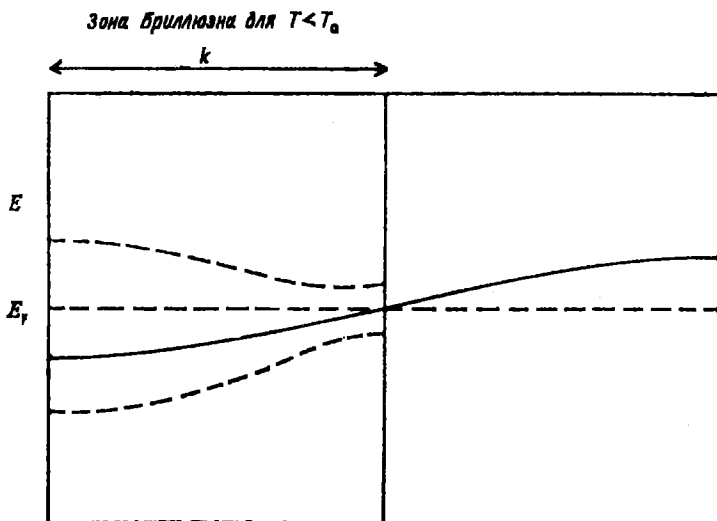


Рис. 13.11. Иллюстрация перестройки запрещенной энергетической зоны в результате уменьшения зоны Бриллюна в процессе изменения кристаллической структуры при $T = T_c$.

ствием магнитного упорядочения. Эти проблемы, а также проблемы, связанные с поведением NiO, довольно подробно рассмотрел Фейнлиб ([23], стр. 231).

Эта теория не объясняет полупроводниковых свойств NiO. Такое объяснение почти наверняка потребовало бы вычисления зонной структуры с учетом не только d -зоны, но и ковалентной связи между d -электронами и электронами ионов O^{2-} . Такие попытки сделали Фейнлиб и Адлер [69]. Обзор магнитных свойств и явлений проводимости дал Адлер [70]. Другие обзоры сделали также Доних [71] и Остин и Гэмбл [72].

Соединения $SrTiO_3$ и $BaTiO_3$, которые в обычном состоянии являются диэлектриками, в результате соответствующего легирования могут получить полупроводниковые свойства. Как отметил Фредерикс ([23], стр. 259), их зонная структура в некотором смысле аналогична зонной структуре окислов переходных металлов.

13.10. Тугоплавкие полупроводники

В то время как большинство рассмотренных нами полупроводников удовлетворительно работает в электронных приборах при комнатной температуре, полупроводники с запрещенной зоной ме-

нее 1 эВ при повышении температуры теряют свойства, придаваемые им путем введения контролируемых количеств примесей. Существует ряд применений, для которых существенно, чтобы материал сохранял желаемые свойства при намного более высоких температурах, а это означает, что он должен иметь значительно более широкую запрещенную зону и, кроме того, высокую температуру плавления. Поэтому велись поиски материалов, которые не только обладали бы этими свойствами, но и могли бы быть приготовлены с проводимостью как n -типа, так и p -типа, что позволило бы сформировать p — n -переходы и т. д. Такие материалы обычно называют тугоплавкими полупроводниками.

Большинство работ в этом направлении было выполнено с SiC, который, как одно время считалось, может заменить Si при работе в области высоких температур. Кроме того, в этом материале было обнаружено сильное излучение света, что позволило бы использовать его в оптоэлектронных устройствах. В этом смысле результаты большого числа работ принесли разочарование. Отчасти сложности возникают из-за того, что SiC может существовать в нескольких аллотропных формах. Кроме того, очень сложной задачей оказалось контролируемое введение примесей. Из измерений положения края фундаментального поглощения SiC Чойк и Патрик [73] определили ширину запрещенной зоны, которая оказалась равной $\Delta E = 2,8$ эВ.

В последнее время привлекли внимание соединения типа $A^{III}B^V$ с более широкими запрещенными зонами, такие, как AlP, AlAs, InP (см. разд. 13.5). Некоторые соединения азота с элементами V группы можно отнести к материалам этого типа, например AlN, GaN и InN. К тугоплавким полупроводникам можно отнести также и соединения бора, такие, как BN, BP и BAs. Предстоит еще сделать массу исследований, чтобы свойства этих соединений были поняты так же хорошо, как свойства соединений типа $A^{III}B^V$. Кроме того, хотя первые работы по контролируемому введению примесей в таких материалах уже выполнены, еще многое предстоит сделать, прежде чем легирование тугоплавких полупроводников будет освоено в такой же степени, как для соединений типа $A^{III}B^V$ с меньшей шириной запрещенной зоны ΔE .

Свойства тугоплавких полупроводников подробно рассмотрели Шмарцев, Валов и Боршевский [74].

13.11. Сверхпроводящие полупроводники

Имеется ряд сообщений о наблюдении в некоторых полупроводниках эффекта сверхпроводимости при температурах менее 1 К. Однако в некоторых случаях такая интерпретация экспериментальных результатов кажется сомнительной. Например, в разное время

появились сообщения о наблюдении сверхпроводимости во всех трех халькогенидах свинца PbS, PbSe и PbTe, если все они были сильно нестехиометричными и имели проводимость *n*-типа, т. е. обладали высокой концентрацией электронов. В настоящее время считается, что обнаруженная на опыте сверхпроводимость на самом деле связана с наличием в материале нитевидных вкраплений чистого свинца.

Хейн и др. [75] обнаружили сверхпроводимость в SnTe и довольно надежно — в GeTe. И в том и в другом материале также имеются отклонения от стехиометричности, в результате которых может возникнуть концентрация свободных носителей порядка 10^{21} см⁻³.

Однако наиболее изученный сверхпроводящий полупроводник — это окисел SrTiO₃, в котором сверхпроводимость обнаружена при наличии кислородных дефектов и при концентрациях носителей в диапазоне $5 \cdot 10^{18}$ — 10^{21} см⁻³. Были исследованы и некоторые другие смешанные окислы, такие, как WO₃, однако, для того чтобы они стали сверхпроводниками, в них необходимо добавить несколько процентов щелочного металла. Эти и другие материалы обсудил Фредерикс ([23], стр. 270).

Некоторые полупроводники становятся сверхпроводящими в условиях высоких давлений. В этом случае они ведут себя как металлы, а не как полупроводники ([23], стр. 270), например InSb, InTe и элементарный полупроводник Te.

Теории сверхпроводимости в приложении к полупроводникам были разработаны Гуревичем, Ларкиным и Фирсовым [77] и Коэн [78]. Первая из них относится в основном к полярным полупроводникам, в то время как вторая является более общей и находится в хорошем согласии с экспериментальными наблюдениями, в особенности это относится к зависимости температуры перехода от концентрации носителей.

13.12. Магнитные полупроводники

Некоторые материалы, интерес к которым возник в основном из-за их магнитных свойств, являются полупроводниками. Большинство таких материалов — это окислы или халькогениды, и мы уже обращались к ним в разд. 13.9. Например, NiO, CoO и FeO — это антиферромагнетики и в то же время полупроводники с низкой подвижностью носителей заряда (см. разд. 13.14). Различные гранаты и шпинели — это ферримагнетики, обладающие полупроводниковыми свойствами, с низкими величинами подвижности, однако не такими низкими, как в антиферромагнитных окислах (10^{-2} см²·В⁻¹·с⁻¹). FeCr₂S₄ и Cr₂S₃ — это полупроводники, и опять же с очень низкой подвижностью. Более удивительны некоторые

ферромагнитные материалы, обладающие полупроводниковыми свойствами, из которых наиболее изучены EuO , EuS , EuSe и CdCr_2Se_4 .

Во всех этих материалах присутствуют либо атомы переходных металлов с $3d$ -электронами, либо атомы редкоземельных металлов с $4f$ -электронами. Эти материалы довольно подробно рассмотрел Хаас [12], стр. 169). Кроме того, к этому вопросу обращались Остин и Элвелл [79], которые обсудили также трудности, встречающиеся при очистке и приготовлении монокристаллов высокого качества из этих материалов. Малые включения ферромагнитных примесей, например Fe , в некоторых материалах могут дать значительные по величине эффекты и скрыть собственные свойства этих материалов.

13.13. Органические полупроводники

Большой интерес вызывают некоторые органические вещества, которые обнаруживают полупроводниковые свойства. Из органических полупроводников наиболее подробно исследованию подвергся антрацен, кроме того, некоторое внимание привлекли и фталоцианиты. Работы по этим материалам сдерживались сложностями приготовления соединений с чистотой, хотя бы отдаленно напоминающей степень очистки, достигнутую для неорганических полупроводников. Органические соединения склонны смешиваться с малыми количествами родственных им соединений, которые, однако, могут иметь очень непохожие свойства. Другая сложность состоит в приготовлении кристаллов высокого качества и достаточно большого размера, так как эти материалы обычно оказываются мягкими и очень хрупкими. Если не принимать специальных мер предосторожности, то кристаллы часто оказываются подвергнутыми сильным остаточным механическим напряжениям.

Несмотря на все это, о полупроводниковых свойствах некоторых таких материалов известно уже очень многое. В основном информация была получена в результате изучения температурной зависимости электропроводности таких материалов, а также в результате измерений их оптического поглощения, для которого не требуется кристаллов большого размера. Свойства органических полупроводников довольно подробно рассмотрели Инокучи и Аканто [80], а также Гутман и Лионс [81]. Их свойства в условиях давления описал Дрикамер [82].

13.14. Полупроводники с низкой подвижностью носителей заряда

Мы уже кратко рассматривали ряд примеров, когда подвижность электронов и дырок в полупроводниках становится низкой или даже очень низкой: это имеет место в случае примесной проводимости при очень низких температурах в некоторых окислах, рассмотренных в разд. 13.9, и в большинстве органических полупроводников. Было обнаружено, что в окислах подвижность имеет величину всего 10^{-8} — 10^{-6} см²·В⁻¹·с⁻¹, однако такие окислы являются исключением, и их следует относить к классу диэлектриков, а не к классу полупроводников. Чаще встречаются подвижности порядка 1 — 10^{-2} см²·В⁻¹·с⁻¹. В аморфных материалах подвижности обычно не превышают эту же величину, однако обсуждение жидких и аморфных полупроводников мы отложим до гл. 15.

Изучая кристаллические полупроводники с низкой подвижностью носителей заряда, важно понять, что теория, основанная на представлениях о блоховских волнах в кристалле, к этим материалам действительно неприменима и что поэтому следует пользоваться представлениями, развитыми в связи с объяснением явления прыжковой проводимости, которые кратко описаны в разд. 5.4 и к которым мы позже вернемся при рассмотрении свойств аморфных полупроводников (см. разд. 15.1, 15.2).

Нетрудно понять, почему теория энергетических зон, которая позволила успешно объяснить свойства полупроводников с высокой подвижностью, оказывается неприменимой, если величина подвижности становится менее 1 см²·В⁻¹·с⁻¹. Если применить соотношение $\mu = e\tau/m_e$, которое не связано с представлениями о кристаллической структуре или с зонной теорией, то при $m_e = m_0$ и $\mu = 1$ см²·В⁻¹·с⁻¹ мы получим $\tau = 1,6 \cdot 10^{-16}$ с. Поскольку состояние электрона будет неопределенно в течение времени τ , то в соответствии с принципом Гейзенберга неопределенность его энергии можно выразить в виде $\delta E \sim \hbar/\tau$. Для нашего значения τ получаем, что $\delta E \approx 1$ эВ. Неопределенность в энергии такой величины делает само понятие зонной структуры бессмысленным. Более того, теряет смысл и величина средней длины свободного пробега (см. разд. 15.1).

Полупроводники с низкой подвижностью носителей заряда и проблемы, связанные с определением их свойств, рассмотрели Иоффе и Регель [83]. Фрёлых и Сивелл [84] обсудили ограничения применимости зонной теории, особенно для случая узких зон с большой эффективной массой. Клингер [85] рассмотрел явления переноса в полупроводниках с низкой подвижностью, включая явление прыжковой проводимости. Более подробное обсуждение прыжковых процессов и соответствующие библиографические ссылки можно найти в разд. 5.4, а также в разделах, посвященных аморфным полупроводникам (разд. 15.1, 15.2).

13.15. Другие полупроводники

Из огромного числа материалов, которые обычно называют полупроводниками, мы обсудили только малую часть. Многие, хотя и не все материалы такого типа — это относительно сложные соединения, обычно окислы или халькогениды. Так как число двойных соединений ограничено, то они были исследованы более полно. В то же время лишь очень малое их число было изучено хотя бы приблизительно в такой же степени, как материалы, рассмотренные в настоящей книге. Как мы уже видели (разд. 13.7), началось изучение тройных полупроводниковых соединений, и здесь пока еще имеется достаточно большой выбор объектов для исследований. С учетом четверных соединений и соединений более высокого порядка, число возможных вариантов становится просто огромным. Некоторые из них уже были изучены, однако весьма поверхностно, и будет удивительно, если среди них не найдется одного-двух материалов с очень специфическими свойствами. Однако следует отдавать себе отчет, что чем из большего числа компонентов состоит соединение, тем сложнее его будет приготовить и довести до состояния, требуемого в полупроводниковой технологии.

Существуют таблицы полупроводниковых соединений, которые содержат известные сведения о них, в основном ширину запрещенной зоны (см. ссылки к разд. 13.1; см. также работу [86]).

Довольно необычную группу полупроводников представляют материалы, в которых время жизни неосновных носителей заряда τ_n (или τ_p) меньше, чем время диэлектрической релаксации τ_0 (см. разд. 7.3). Свойства этих материалов, называемых обычно релаксационными полупроводниками, обсуждали Ван Русбек и Кэйси [87]. Как мы видели в разд. 7.3, τ_0 обычно имеет величину порядка 10^{-12} с. Соотношение $\tau_n \gg \tau_0$ приводит к состоянию квазинейтральности. Если это условие не выполняется, то в полупроводнике может возникать пространственный заряд и соответственно явление истощения основных носителей (см. разд. 7.3).

После появления первой работы, посвященной этому вопросу, новых экспериментальных исследований почти не последовало. Совсем недавно к этой проблеме вернулись Попеску и Хениш [43], которые рассмотрели, что будет происходить в таком материале, если инжектировать в него неосновные носители. Они рассмотрели также роль ловушек, на которые могут быть захвачены неосновные носители, и влияние этих ловушек на скорость рекомбинации.

ЛИТЕРАТУРА

1. Kane P. F., Larrabee G. B., Characterization of Semiconducting Materials McGraw-Hill, 1975.
2. Electronic Materials, ed. Hannay N. B., Colombo U., Plenum Press, 1973.

3. *Brooks M. S., Kennedy J. K.*, Ultra-purification of Semiconductors, Macmillan, 1962.
4. *Parker R. L.*, Crystal growth mechanisms, Solid State Physics, Academic Press, **25**, 152 (1970).
5. The Art and Science of Growing Crystals, ed. Gilman J. J., Wiley, 1963.
6. Crystal Growth, ed. Bardsley W., Heurle D. J. T., Mullin J. B., North Holland, 1973.
7. Runyan W. R., Semiconductor Measurements and Instrumentation, McGraw-Hill, 1975.
8. Materials used in Semiconductor Devices, ed. Hogarth C. A., Wiley-Interscience, 1965.
9. Selected Constants Relative to Semiconductors, Pergamon Press, 1961.
10. Handbook of Electronic Materials, IFI-Plenum, 1971—1972.
11. Semiconductors and Semimetals, ed. Willardson R. K., Beer A. C., Academic Press, 1966—68, vol. 1—4.
12. *Brust D., Phillips J. C., Bassani F.*, Phys. Rev. Lett., **9**, 94 (1962).
13. *Brust D., Cohen M. L., Phillips J. C.* Phys. Rev. Lett., **9**, 389 (1962).
14. *Macfarlane G. G. et al.*, Phys. Rev. **108**, 1377 (1957); **111**, 1245 (1958).
15. *Herman F., Glicksman M., Parmenter R. A.*, Progress in Semiconductors, Heywood, **2**, 1. (1957).
16. *Dominicali C. A.*, Journ. Appl. Phys., **28**, 749 (1957).
17. *Moss T. S.*, Photoconductivity of the Elements, Butterworths, 1952.
18. *Blakemore J. S.*, Semiconductor Statistics, Pergamon Press, 1962. [Имеется перевод: Блекмор Дж. Статистика электронов и дырок в полупроводниках.— М.: Мир, 1964.]
19. *Button K. J. et al.*, Phys. Rev. Lett., **23**, 14 (1969).
20. Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970, p. 253.
21. *Busch G., Kern R.*, Solid State Physics, Academic Press, **11**, 1 (1960).
22. *Groves S., Paul W.*, Proc. VIIth Int. Conf. on Phys. of Semiconductors, Dunod, 1964, p. 41.
23. Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969.
24. *Ewald A. W.*, Helv. Phys. Act., **41**, 795 (1968).
25. *Welker H., Weiss H.*, Solid State Physics, **3**, 1 (1956).
26. *Madelung O.*, Physics of III—V Compounds, Wiley, 1964.
27. *Hilsum C., Rose-Innes A. C.*, Semiconducting III—V Compounds, Pergamon Press, 1961.
28. *Witt A. F., Gatos H. C.*, Journ. Electrochem. Soc., **113**, 808 (1966).
29. *Cohen M. L., Bergstresser T. K.*, Phys. Rev., **141**, 789 (1966).
30. *Long D.*, Semiconductors and Semimetals, Academic Press, **1**, 143 (1966).
31. *Gershenson M.*, Semiconductors and Semimetals, Academic Press, **2**, 289 (1966).
32. *Scanlon W. W.*, Solid State Physics, Academic Press, **9**, 83 (1959).
33. *Bell Dorothy G. et al.*, Proc. Roy. Soc. **A217**, 71 (1953).
34. *Conklin J. B., Johnson L. E., Pratt G. W.*, Phys. Rev., **137A**, 1282 (1965).
35. *Lin P. J., Kleinman L.*, Phys. Rev., **142**, 478 (1966).
36. *Mitchell D. L., Palik E. D., Zemel J. N.*, Proc. VIth Int. Conf. on Phys. of Semiconductors, Dunod, 1964, p. 325.
37. *Moss T. S., Burrell G. J., Ellis B.*, Semiconductor Opto-electronics, Butterworths, 1973.
38. *Bernick R. L., Kleinman L.*, Solid State Commun., **8**, 569 (1970).
39. *Bloem J., Kröger F. A.*, Zs. Phys. Chem., **7**, 1 (1956).
40. *Cochran W.*, Phys. Lett., **13**, 193 (1964).
41. *Dalven R.*, Solid State Physics, Academic Press, **28**, 179 (1973).
42. *Kröger F. A., Vink H. J., van den Boomgard*, Zs. Phys. Chem., **203**, 1 (1954).
43. *Popescu C., Henisch H. K.*, Phys. Rev., **B11**, 1563 (1975); Journ. Phys. Chem. Solids, **37**, 47 (1975); Phys. Rev., **B14**, 517 (1976).
44. II—VI Semiconducting Compounds, ed. Thomas D. G., Benjamin, 1967.

45. Hahn H. et al., Zs. Anorg. Allg. Chem., 271, 153 (1953).
46. Goodman C. H. L., Douglas R. W., Physica, 20, 1107 (1954).
47. Mooser E., Pearson W. B., Report of Meeting on Semiconductors, April 1956, London Physical Society, p. 65.
48. Smith R. C., Journ. de Phys., 36, C3-89 (1975).
49. Feigelson R. S., Journ. de Phys., 36, C3-57 (1975).
50. Matthes M. et al., Journ. de Phys., 36, C3-105 (1975).
51. Shay J. L., Wernick J. H., Ternary Chalcopyrite Semiconductors — Growth, Electronic Properties and Applications, Pergamon Press, 1975.
52. Bendorius R. A. et al., Proc. XIth Conf. on Phys. of Semiconductors, Polish Sci. Publ., 1972, p. 838.
53. Shileika A., Surf. Sci., 37, 730 (1973).
54. Полягалов Ю. И., Поплавной А. С., Ратнер А. М., Journ. de Phys. Suppl., C3-129 (1975).
55. Paseman L. et al., Phys. Stat. Solidi, 77, 527 (1976).
56. Lawson W. D. et al., Journ. Phys. Chem. Solids, 9, 325 (1959).
57. Bartlett B. E., Deans J., Ellen P. C., Journ. Math. Sci., 4, 266 (1969).
58. Bartlett B. E. et al., Infrared Journ., 9, 35 (1969).
59. Schmitt J. L., Stelzer E. L., Journ. Appl. Phys., 40, 4865 (1969).
60. Dimmock J. O., Melngailis L., Strauss A. J., Phys. Rev. Lett., 16, 1193 (1966).
61. Melngailis L., Harman T. C., Appl. Phys. Lett., 13, 180 (1968).
62. Ellis B., Moss T. S., Phys. Stat. Solidi, 41, 531 (1970).
63. Dimmock J. O., в сб.: Physics of Semimetals and Narrow-Gap Semiconductors, ed. Carter D. L., Bates R. T., Pergamon Press, 1971, p. 319.
64. Zawadzki W., Adv. Phys., 23, 435 (1974).
65. Mott N. F., Can. Journ. Phys., 34, 1356 (1956); Phil. Mag., 6, 287 (1961).
66. Adler D., Brooks H., Phys. Rev., 155, 826 (1967).
67. Feinleib J., Paul W., Phys. Rev., 155, 841 (1967).
68. Adler D. et al., Phys. Rev., 155, 851 (1967).
69. Feinleib J., Adler D., Phys. Rev. Lett., 21, 1010 (1968).
70. Adler D., Solid State Physics, Academic Press, 21, 1 (1968).
71. Doniach S., Advan. Phys., 18, 819 (1969).
72. Austin I. G., Gamble R., в сб.: Conduction in Low-Mobility Materials, ed. Klein N., Tannhäuser D. S., Pollak M., Taylor and Francis, 1971, p. 17.
73. Шоуке В. Ж., Патрик Л., Phys. Rev., 105, 1721 (1957).
74. Шмарцев Ю. В., Валов Ю. А., Борщевский А. С. Жаростойкие алмазоподобные полупроводники./Под ред. Горюновой Н. А., Наследова Д. А.— М.: Металлургия, 1964.
75. Hein R. A., et al. Phys. Rev. Lett., 12, 320 (1964).
76. Schooley J. F. et al., Phys. Rev. Lett., 14, 305 (1965).
77. Гуревич В. Л., Ларкин А. И., Фирсов Ю. А., ФТТ, 4, 185 (1962).
78. Cohen M. L., Phys. Rev. A, 134, 442 (1964).
79. Austin I. G., Elwell D., Contemp. Phys., 11, 455 (1970).
80. Inokuchi H., Akanto H., Solid State Physics, Academic Press, 12, 93 (1961).
81. Gutman F., Lyons L. E., Organic Semiconductors, Wiley, 1967.
82. Drikkamer H. G., Solid State Physics, Academic Press, 17, 1 (1965).
83. Иоффе А. Ф., Регель А. П., Journ. Phys. Chem. Solids, 8, 6 (1959); Prog. in Semiconductors, Heywood, 4, 237 (1960).
84. Fröhlich H., Sewell G. L., Proc. Phys. Soc., 74, 643 (1959).
85. Klitinger M. J., Proc. Vth Int. Conf. on Phys. of Semiconductors, Inst. of Phys. Lond., 1962, p. 205.
86. Pamplin B., New Scientist, 64, 789 (1974).
87. Van Roosbroeck W., Casey H. C., Proc. Xth Int. Conf. on Phys. of Semiconductors, USAEC Publications, 1970, p. 832.
- 88*. Абдуллаев Г. Б. Физические процессы, происходящие в селене и селеновых приборах.— Баку, 1959; Абдуллаев Г. В., Абдинов Д. Ш. Физика селена.— Баку, изд-во «Эльм», 1975.

НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ ВОПРОСЫ ФИЗИКИ ПОЛУПРОВОДНИКОВ

14.1. Экситонные молекулы

Связывание одиночных экситонов на различных примесных центрах уже обсуждалось в разд. 10.6 и 10.14. Самым удивительным свойством этих комплексов является излучательная аннигиляция связанных экситонов, которая дает спектр излучения в форме интенсивных узких линий. Хотя еще в 1958 г. Ламперт [1] предсказал, что пара экситонов может соединиться с образованием экситонной молекулы, весьма схожей с H_2 , экспериментально этот факт был обнаружен лишь в 1966 г., когда Хейнс [2]¹⁾ наблюдал излучение такой молекулы в спектре оптически возбужденного Si. В Ge излучение, обусловленное экситонными молекулами, позже обнаружили Бенуа и ля Гийом, Сальван и Вус ([3], стр. 516). Аналогичное излучение было обнаружено и в ряде полупроводников типа $A^{III}B^{VI}$, таких, как CdS и CdSe, а также в материалах типа CuBr. Эти работы и последующие теоретические исследования изложены в обзоре Ханамуры ([4], стр. 81).

Путем сравнения с молекулой H_2 Хейнс оценил энергию связи экситонной молекулы относительно распада на два экситона и нашел, что она должна быть равна ~ 2 мэВ. На первый взгляд могло бы показаться, что энергия связи молекулы W_M и энергия связи одиночного экситона W_{ex} находятся в простой зависимости $W_M/W_{ex} = W_{H_2}/W_H = 0,33$. Однако вследствие близости эффективных масс дырки и электрона задача вычисления энергии связи экситонной молекулы не совсем аналогична задаче о молекуле H_2 . Тем не менее Акимото и Ханамура [5] показали, что такая молекула будет стабильной при всех возможных величинах отношения масс элект-

¹⁾ В настоящее время связь излучения, которое наблюдал Хейнс, с экситонными молекулами вызывает сомнения. Предполагают, что это излучение скорее относится к конденсированной фазе экситонов (см. разд. 14.2). То же самое можно сказать об излучении, обнаруженном в германии. Однако излучение в CuBr и подобных ему материалах, по-видимому, связано именно с экситонными молекулами.

рона и дырки. Для многодолинных полупроводников, таких, как Ge и Si, расчет энергии связи очень сложен. Такие вычисления выполнили Акимото [6] и другие авторы. Вопрос об энергии связи экситонной молекулы исследовал Ханамура ([4], стр. 81). Оказывается, что энергия связи молекулы пропорциональна энергии связи экситона, но коэффициент пропорциональности изменяется от 0,01 до 0,03 в зависимости от отношения m_e/m_h , причем он имеет максимальное значение для наименьших величин отношения (как и следовало ожидать, из аналогии с молекулой водорода), а также при наиболее сильной анизотропии. Таким образом, оказывается, что биэкситон энергетически метастабилен. К вопросу о существовании других, более стабильных групп экситонов мы вернемся несколько позже. Оказывается, что в таких материалах, как CuCl и CdS, условия в большей степени, чем это имеет место для Ge и Si, предпочтительны для образования одиночных молекул, а не их групп. Вероятно, поэтому экситонную молекулу легче было обнаружить именно в таких материалах.

Вэнг и Киттель [7] предположили, что в Ge и Si из-за их многодолинной структуры должны существовать метастабильные молекулы Ex_1 и Ex_{1a} , и они должны быть более стабильными, чем молекулы Ex_2 . Однако до настоящего времени нет убедительного экспериментального доказательства существования этих молекул. С другой стороны, как мы увидим в разд. 14.2, объединяться могут намного больше группы экситонов, образуя капли макроскопического размера.

Имеются факты, указывающие на то, что биэкситон может быть связан на нейтральной примеси, такой, например, как N в GaP. Мерц, Фолкнер и Дин [8] наблюдали узкие линии в спектре фотолюминесценции GaP, которые, как они считают, обусловлены излучательной рекомбинацией биэкситонов, связанных на примесях N.

14.2. Конденсация экситонов в электронно-дырочные капли

Так как экситоны — это частицы с нулевым спином (электрон и дырка имеют антипараллельные спины — аналогично основному состоянию атома водорода), то можно ожидать, что при низкой температуре они сконденсируются в концентрированную фазу, подобную жидкости (бозе-конденсация). Это было бы аналогично конденсации водорода в жидкий H_2 . Есть некоторые свидетельства (Наката и др. [9]), что такая конденсация может происходить в сильнополярных материалах, подобных CuCl и CuBr. В таких материалах, как Si и Ge, оказывается, однако, что состояние с наименьшей энергией возникает, когда одиночные экситоны распадаются и образуются плотная электронно-дырочная плазма. Это было бы аналогично

жидкому металлическому водороду, если бы он существовал. В действительности существует критическая температура T_c , ниже которой конденсация молекул трансформируется в конденсацию «металлического» типа, что является одним из видов перехода Мотта, обсуждаемого в разд. 9.13.

До сих пор такой переход обнаружен не был, несмотря на многочисленные исследования при температуре жидкого гелия, которая по оценкам явно ниже T_c . В 1968 г. Келдыш [10] предположил, что экситоны должны конденсироваться в «металлическую» фазу. С тех пор эта проблема изучалась многими экспериментаторами и оказалась одной из наиболее интенсивно развивающихся областей исследования в физике твердого тела последних лет.

При вычислении энергии связи на один экситон предполагается, что «металлическое» состояние находится в равновесии с «газом» свободных экситонов. Некоторые исследователи, например Бринкман и др. [11], а также Комбеско и Нозье [12], пытались получить зависимость энергии связи от среднего расстояния r между частицами в такой плазме. Основным результатом этих исследований явилось то, что в Ge энергия связи на один экситон имеет четко выраженный максимум при величине r , равной r_m , при этом энергия связи равна 1,7 мэВ (вместо 3,6 мэВ (см. разд. 10.6) для свободного экситона). Интересно отметить, что для простого полупроводника с равными и изотропными эффективными массами электрона и дырки, вычисления дают только очень неглубокий максимум энергии связи. Величина r_m соответствует плотности экситонов n_0 , равной $1,8 \cdot 10^{17} \text{ см}^{-3}$. Вус и Бенуа а ля Гийом ([4], стр. 143) обсудили эти расчеты, а также провели некоторые другие вычисления.

В большинстве экспериментальных работ, посвященных исследованию «металлической» фазы, необходимая концентрация экситонов в полупроводнике создавалась путем облучения образца светом лазера, однако в некоторых экспериментах применялась инжекция носителей заряда из $p^+ - n$ -перехода. Первое доказательство существования электронно-дырочных капель (ЭДК) появилось в результате наблюдения дополнительных уширенных линий с длинноволновой стороны от линии излучения свободного экситона в Ge и Si Покровским и Свистуновой [13]. Главным доказательством того, что эта фаза имеет «металлический» характер, является форма линии, которая соответствует концентрации n_0 порядка $2 \cdot 10^{17} \text{ см}^{-3}$.

Доказательство существования ЭДК и оценка их размера были получены очень простым методом. Если капли создаются в области n -типа вблизи $p - n$ -перехода, то какая-нибудь одна из них, достигнув самого перехода, разрушится в его сильном электрическом поле. В этом случае возникнет сильный импульс тока, который выделится на фоне тока, обусловленного фото-э. д. с. $p - n$ -перехода. Эксперимент такого типа был впервые успешно проделан Асининым и др. [14] и неоднократно повторялся при различных условиях не-

которыми другими авторами (Вус и Бенуа а ля Гийом ([4], стр. 143). Из этих экспериментов следует, что радиусы капель в Ge имеют величины примерно 1—10 мкм. Эффект наблюдается при температуре ~ 2 К и уже при температуре 15 К отсутствует. Это связано с тем, что при более высокой температуре капли испаряются и превращаются в газ свободных экситонов. Этот результат подтверждается также спектроскопическими измерениями.

Замечательный эксперимент провели Покровский и Свистунова [15], Багаев и др. [17], а также некоторые другие исследователи. В этом эксперименте для обнаружения ЭДК и измерения их радиусов использовалось рэлеевское рассеяние излучения с длиной волны 3,39 мкм от He—Ne-лазера. Полученные величины радиусов лежат в области 2—8 мкм. Багаев и др. [16] обнаружили, что радиус капли *возрастает* при увеличении температуры, изменяясь от 4 мкм при 2,6 К до 10 мкм при 3,2 К.

Исследуя излучение капель и поглощение на каплях в дальнем инфракрасном диапазоне длин волн, Вавилов, Заяц и Мурзин ([3], стр. 509), считая, что максимум поглощения связан с излучением с энергией кванта $h\nu=8,7$ мэВ, получили оценку $n_0=2 \cdot 10^{17}$ см $^{-3}$, что находится в хорошем согласии с результатами других измерений.

Хотя в большинстве экспериментов наблюдались капли размером примерно 1—10 мкм, появилось несколько сообщений и о каплях очень большого размера (см., например, работу Фелдмана [17]). Возникли, однако, сомнения, не являются ли эти почти сферические образования скоплениями большого числа капель намного меньшего размера. Если Ge подвергнуть неоднородному механическому сжатию, то можно создать область кристалла, в которой конденсация будет более предпочтительна по сравнению с остальным объемом кристалла. Этим методом Джеффрис, Вольф и Маркиевич ([18], стр. 879) создали капли радиусом 0,3 мкм с концентрацией $n_0=0,7 \times 10^{17}$ см $^{-3}$. Эксперименты и упомянутые выше расчеты показали, что при приложении механического сжатия энергия связи уменьшается. Багаев, Галкина и Гоголин ([3], стр. 500) обнаружили уменьшение энергии связи на основании сдвига максимума спектра излучения, присущего конденсированной фазе экситонов. Они также обнаружили, что в сильном магнитном поле порядка 10 Т линия излучения расщепляется, что обусловлено спиновым расщеплением уровней энергии электронов и дырок.

Из литературы известны различные оценки времени жизни капля. После прекращения возбуждения они рекомбинируют за время ~ 40 мкс. Джеффрис, Вольф и Маркиевич ([18], стр. 879) обнаружили, что капли большого размера, для которых концентрация носителей имеет меньшую величину, характеризуются временем спада сигнала излучения порядка 200 мкс. Видимо, небольшие капли могут диффундировать, однако в опубликованных данных по дли-

не диффузии капель имеется очень большой разброс. Как бы то ни было, оказывается, что при определенных условиях капли могут перемещаться на расстояния порядка 1 мм. Алексеев, Багаев и Галкина [19] показали, что неоднородное поле механического напряжения усиливает диффузию капель.

Большинство экспериментальных работ по исследованию ЭДК было проведено на Ge, однако некоторые из них проделаны на сплавах Ge—Si. Теоретически и экспериментально показано, что энергия связи в них имеет меньшую величину. Анизотропия эффективной массы электрона и наличие «непрямой» запрещенной зоны в Ge делают этот материал особенно благоприятным для образования ЭДК. Видимо, в полупроводниках с прямой запрещенной зоной условия для образования ЭДК менее благоприятны, однако Лекени и Шах ([18], стр. 394) сообщили об обнаружении ЭДК в CdS и CdSe. Энергия связи в CdS имеет удивительно большую величину 13 эВ, в то время как в CdSe она меньше 2 эВ.

14.3. Поляроны и поляритоны

Поляроны и поляритоны хотя и имеют похожие названия, но по сути своей это совершенно различные образования. Поляроны — это свободные электроны или дырки, передвигающиеся в материале, поляризованном вследствие их присутствия, и переносящие эту поляризацию по мере своего движения. Поляритоны — это возбуждения, которые возникают в процессе взаимодействия света либо инфракрасного излучения с другими возбуждениями, например с фононами или экситонами. Сначала кратко рассмотрим поляроны.

В полярной среде электрон будет отталкивать отрицательные ионы и притягивать положительные, поляризуя таким образом материал вокруг себя. В полярном кристалле степень этой поляризации, обусловленной кулоновским полем электрона, может быть достаточно велика. При движении электрон будет нести эту поляризацию за собой, что в свою очередь будет влиять на эффективную массу самого электрона. Теорию этого эффекта разработали Фрелих, Пельцер и Зинау [20] еще в 1950 г. Позже Фрелих [21] ее модернизировал ¹⁾.

Если параметр взаимодействия α , определяемый уравнением (8.28), мал, то модифицированная эффективная масса m_e^* электрона дается выражением

$$m_e^* = m_e \left(1 + \frac{\alpha}{6} \right).$$

¹⁾ Основы теории поляронов были развиты еще ранее в работах Пекара [84]. — Прим. перев.

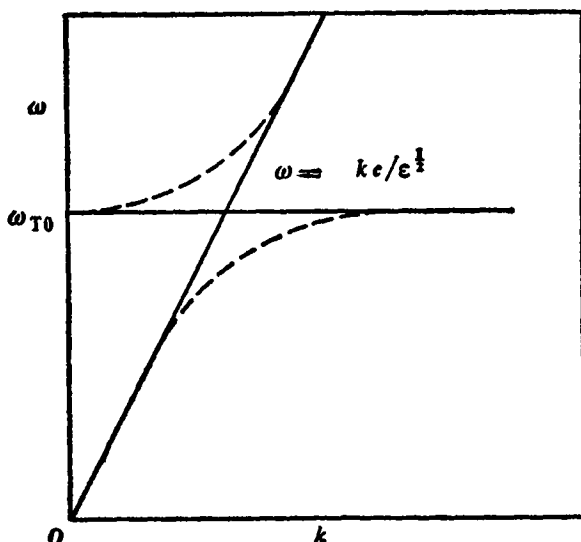


Рис. 14.1. Иллюстрация взаимодействия фотона и ТО-фонона с образованием поляртона.

Пунктиром показана дисперсионная кривая для поляртона.

В полупроводниках типа $A^{III}B^V$ параметр α достаточно мал, так что эффективная масса изменяется слабо. В PbS $\alpha=0,28$ и опять же эффект пренебрежимо мал, хотя в этом случае теория и не вполне справедлива. Для сильно полярных кристаллов α имеет еще большую величину, поэтому возникает потребность в более сложной теории. Полярон это есть проявление взаимодействия электрона с оптическим фононом и с этой точки зрения его рассматривали Харпер, Ходби и Стрэдлинг [22]. Целый сборник статей [23] посвящен обсуждению свойств полярнов не только в галогенидах щелочных металлов, для которых понятие полярона и было впервые развито, но также и в полярных полупроводниках.

Теперь кратко рассмотрим поляритоны. Для начала рассмотрим взаимодействие фотонов с ТО-модой колебаний решетки. Как мы уже видели, энергия $\hbar\omega$ для ТО-моды вблизи $k=0$ слабо зависит от величины k и поэтому на рис. 14.1 ω представлена в виде горизонтальной линии. С другой стороны, для фотона ω равна $kc/\epsilon^{1/2}$ (здесь мы используем k в качестве волнового вектора как фотона, так и фонона). Кривые дисперсии фотона и фонона пересекаются при $k=\epsilon^{1/2}\omega_{TO}/c$. Если фотон и фонон могут взаимодействовать между собой, то согласно квантовой механике, эти кривые на самом деле не пересекаются, а ведут себя так, как показано пунктиром на рис. 14.1. Для малых и больших величин фотон и фонон ведут

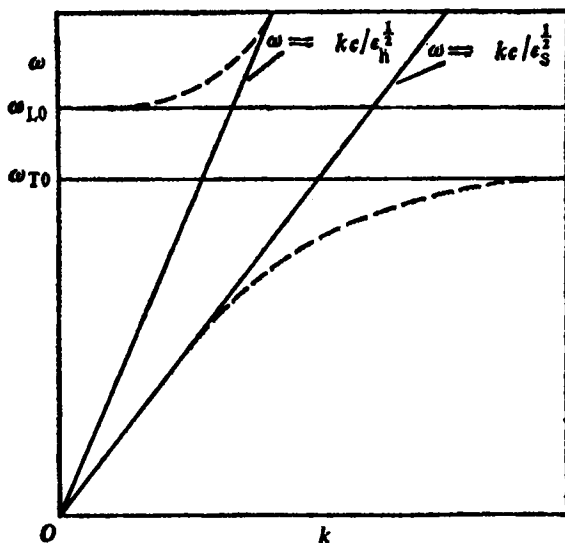


Рис. 14.2. Иллюстрация взаимодействия фотона с ТО- и LO-фононами в полярном кристалле.

Пунктиром показана дисперсионная кривая для поляритона.

себя независимо друг от друга, но вблизи $k = \epsilon^{1/2} \omega_{\text{ТО}}/c$ они сильно взаимодействуют и образуют то, что мы называем поляритонном.

В действительности рис. 14.1 представляет собой очень сильное упрощение ситуации, реально встречающейся в сложных полярных полупроводниках, так как при $k=0$ существуют два различных оптических фонона, ТО- и LO-фононы. Ситуация, которая обычно имеет место в кристаллах с ионной связью, проиллюстрирована на рис. 14.2, на котором вследствие соотношения Лиддана — Закса — Теллера (см. разд. 8.3) $\omega_{\text{LO}}^2 = \omega_{\text{ТО}}^2 \epsilon_h / \epsilon_s$, где ϵ_h и ϵ_s — это высоко- и низкочастотные диэлектрические проницаемости. При малых величинах k имеем фотон и LO-фонон, которые почти не взаимодействуют. При больших величинах k имеем фотон и ТО-фонон, и в этом случае взаимодействие опять же мало.

Однако если $kc/\epsilon^{1/2}$ близко к $\omega_{\text{ТО}}$ или к ω_{LO} , то имеет место сильное взаимодействие и, как следствие, две различные частоты поляритонов для данной величины k .

Таким образом поляритон отражает изменение показателя преломления в интервале между частотами $\omega_{\text{ТО}}$ и ω_{LO} , которое приводит к сильному отражению и поглощению, связанному с частотой остаточных лучей, что обсуждалось в разд. 8.3. Используя понятие поляритона, Миллз и Бурштейн [24] дали простой и изящный вывод соотношения Лиддана — Закса — Теллера, а также нашли его

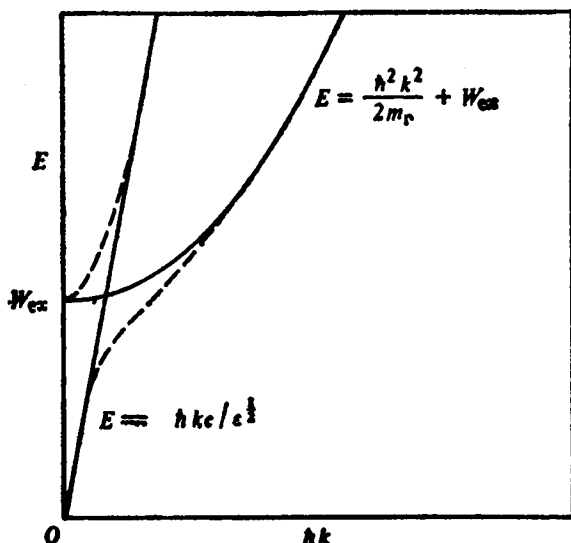


Рис. 14.3. Иллюстрация взаимодействия фотона и свободного экситона с образованием поляритона.

Пунктиром показана кривая дисперсии для поляритона.

связь со степенью ионности кристалла. Они получили дисперсионное соотношение для поляритона в виде

$$\omega^2 = \frac{1}{2} \left(\frac{c^2 k^2}{\epsilon_h} + \omega_{\text{TO}}^2 + \omega_p^2 \right) \pm \frac{1}{2} \left[\left(\frac{c^2 k^2}{\epsilon_h} - \omega_{\text{TO}} - \omega_p \right)^2 + \frac{4c k^2 \omega_p^2}{\epsilon_h} \right]^{1/2}, \quad (14.1)$$

при этом

$$\omega_{\text{LO}}^2 = \omega_{\text{TO}}^2 + \frac{\omega_p^2}{\epsilon_h} = \left(\frac{\epsilon_s}{\epsilon_h} \right) \omega_{\text{TO}}^2, \quad (14.2)$$

где ω_p — ионно-плазменная частота, равная $e^* (N/M_i \epsilon_0 \epsilon_s)^{1/2}$ (ср. ур. (10.45)), e^* — эффективный заряд иона, M_i — приведенная масса, а N — концентрация ионов.

Взаимодействия фотонов и фононов с образованием поляритонов, а также взаимодействие с магнитными возбуждениями довольно подробно рассмотрели Миллз и Бурштейн [24].

Другим видом возбуждения, который взаимодействует с фотоном, является экситон. Движение центра масс экситона описывается параболической кривой дисперсии, как показано на рис. 14.3. Пунктирные кривые отвечают дисперсионному соотношению для поляритона. Распад таких поляритонов обсуждал Хопфилд [25], стр. 77). Нижняя часть дисперсионной кривой проявляется в виде излучения с длиной волны, превышающей низкочастотную границу

излучения свободного экситона. Тэйт и др. [26] наблюдали это излучение в CdS при 12К.

Между фотонами и возбуждениями кристалла возможно множество других взаимодействий, включая плазменные колебания (плазмоны) и поверхностные волны. Многие из них были идентифицированы при помощи рассеяния света типа комбинационного рассеяния, при котором энергия поляритона либо вычитается из энергии рассеянного светового кванта, либо складывается с ней (см., например, разд. 14.7, [27]). Вопрос подробно обсуждается в серии статей, опубликованных в сборнике [28].

Существуют и другие взаимодействия подобного рода. Одно из них аналогично поляритону, однако включает в себя электромагнитное излучение (обычно в микроволновой области спектра), свободные электроны или дырки и оптические фононы. Это взаимодействие между циклотронным резонансом и оптическими фононами. При приближении частоты циклотронного резонанса к частоте длинноволнового оптического фонона между ними возникает сильное взаимодействие. Этот эффект проявляется в виде расщепления линии циклотронного поглощения, а также в виде резкого уширения этой линии. Простое теоретическое описание этого интересного явления и обзор экспериментальных результатов сделал Уоллис [29].

Взаимодействие между фононами и плазмонами также изучалось с помощью метода комбинационного рассеяния света. Это интересное взаимодействие описали Макуортер ([30], стр. 350), а также Платцман и Вольф [31].

14.4. Сильнолегированные полупроводники

В течение многих лет основные усилия в приготовлении полупроводников для экспериментальных исследований и использования в электронных и оптических приборах были направлены, помимо выращивания кристаллов, на устранение нежелательных примесей и однородное введение необходимых. Обычно количества последних очень малы, и поэтому были развиты специальные методики их введения, такие, как первоначальное приготовление сильнолегированного материала (лигатуры), а затем использование лигатуры в качестве примесной добавки. Обычно использовались концентрации доноров и акцепторов, намного меньшие величин N_c и N_v , определенных в разд. 4.2, но все еще большие по сравнению с концентрацией n_i собственных носителей при комнатной температуре. Существенно большие концентрации использовались в основном для формирования n^+ и p^+ -переходов (см. разд. 7.10). Явления электронного переноса в таких сильнолегированных областях не приводили к новым проблемам, и поэтому они подробно не изучались. Однако с разработкой диодов Есаки (см. ниже) и полупро-

водниковых лазеров сильнолегированный материал стал важной и активной частью прибора. Поэтому в последние годы значительно больше внимания уделялось оптическим свойствам и явлениям переноса в сильнолегированных материалах, т. е. в материалах, для которых $p \geq N_v$ или $n \geq N_c$. В итоге несколько книг было посвящено только этим материалам (см. ниже).

Для таких сильнолегированных материалов мы больше не можем использовать простые экспоненциальные формулы для концентраций n и p , приведенные в разд. 4.3 [(4.29) и (4.30)], а должны воспользоваться выражениями, которые можно получить аналогичным способом, но на основании функции распределения Ферми — Дирака. Как это делается, показано в разд. 4.2, уравнение (4.19). Мы должны также модифицировать выражения для средних значений таких величин, как $\langle \tau \rangle$, что ведет к значительно более сложным выражениям для коэффициентов переноса, в особенности для коэффициентов термомагнитных эффектов. Мы не будем выписывать выражений для этих коэффициентов переноса. Полный анализ, приводящий к их выводу, представлен в нескольких книгах и обзорных статьях, в частности в работах Бонч-Бруевича [32] и Фистуля [33]. Мы просто рассмотрим одно-два новых явления, которые были обнаружены в результате исследований свойств сильнолегированных полупроводников.

14.4.1. ЭКРАНИРОВАНИЕ СВОБОДНЫМИ НОСИТЕЛЯМИ

Одним непосредственным результатом возрастания концентраций электронов и дырок должно быть изменение энергетических уровней, соответствующих донорным и акцепторным примесным центрам. Это относится в особенности к примесям «водородоподобного» типа, которые мы обсуждали в подразд. 3.4.1. Для примесей, энергетические уровни которых лежат далеко от зоны проводимости и валентной зоны, силы, действующие на «лишний» электрон или дырку, имеют меньший радиус действия и поэтому меньше изменяются под действием свободных электронов, присутствие которых обусловлено другими примесями. Для примесей «водородоподобного» типа кулоновское поле экранируется другими свободными электронами, и вместо прежнего выражения для потенциальной энергии $-e^2/4\pi\epsilon_0 r$ имеем

$$V(r) = -\frac{e^2}{4\pi\epsilon_0 r} \exp\left(-\frac{r}{r_0}\right), \quad (14.3)$$

где r_0 — так называемый дебаевский радиус, определяемый выражением

$$r_0^2 = \frac{\epsilon\epsilon_0 kT}{ne^2}. \quad (14.4)$$

Этот результат мы уже получали в связи с рассеянием носителей заряда на ионизованной примеси (см, разд. 8.9). В поле такого потенциала может быть лишь небольшое число связанных состояний. Их может и не быть вовсе, если r_0 достаточно мало, так что примесь больше не сможет связать электрон. Тогда все электроны станут свободными и образуют примесную зону. Условия возникновения такого явления обсуждали Пинчерль [34], а также Бонч-Бруевич [32]. Экспериментальные исследования, включающие использование высокой концентрации носителей, инжектированных в слаболегированные полупроводники, были проведены Гласко и Мироновым [35], а также Бонч-Бруевичем [36]. Они подтверждают теорию экранирования, но показывают также, что в случае очень сильного легирования возникают новые проблемы. Однако общий результат состоит в том, что для примесей, подобных донорам V группы в Ge (которые при значениях n , меньших 10^{18} см $^{-3}$, имеют энергии ионизации порядка 0,01 эВ), с увеличением n энергия ионизации постепенно уменьшается, стремясь к нулю, когда n приближается к N_c (около 10^{19} см $^{-3}$). Свойства Ge в случае сильного легирования подробно описал Панков [37].

14.4.2. ЭФФЕКТ БУРШТЕЙНА — МОССА

Другим интересным физическим эффектом, возникающим в результате сильного легирования, является изменение положения края фундаментального поглощения при сильном увеличении концентраций электронов или дырок. Бурштейн [38] и Мосс [39] заметили, что край поглощения в InSb n -типа особенно чувствителен к легированию и сдвигается в сторону более коротких длин волн при увеличении степени легирования. Объяснение этого явления было независимо дано Бурштейном и Моссом и эффект обычно называют эффектом Бурштейна — Мосса. Из-за чрезвычайной малости эффективной массы электрона в InSb уровни зоны проводимости быстро заполняются с увеличением n , сдвигая уровень Ферми выше дна зоны проводимости. Это проиллюстрировано на рис. 14.4. Заполненные состояния ниже уровня Ферми запрещены для переходов из вершины валентной зоны, поэтому переходом с наименьшей энергией будет переход не на дно зоны проводимости, как для слаболегированных материалов, а на уровень Ферми. Так как последний поднимается с увеличением n , то край полосы поглощения сдвигается в сторону более коротких длин волн. В InSb p -типа из-за большей массы дырок этот эффект выражен намного слабее.

Эффект наблюдался в нескольких полупроводниках, включая Ge n -типа. Для n -Ge подробную теорию дали Панков и Эгрен [40]. Они же провели ее сравнение с экспериментом. Для величин n , превышающих 10^{19} см $^{-3}$, в области края полосы поглощения было также обнаружено увеличение наклона кривой поглощения, кото-

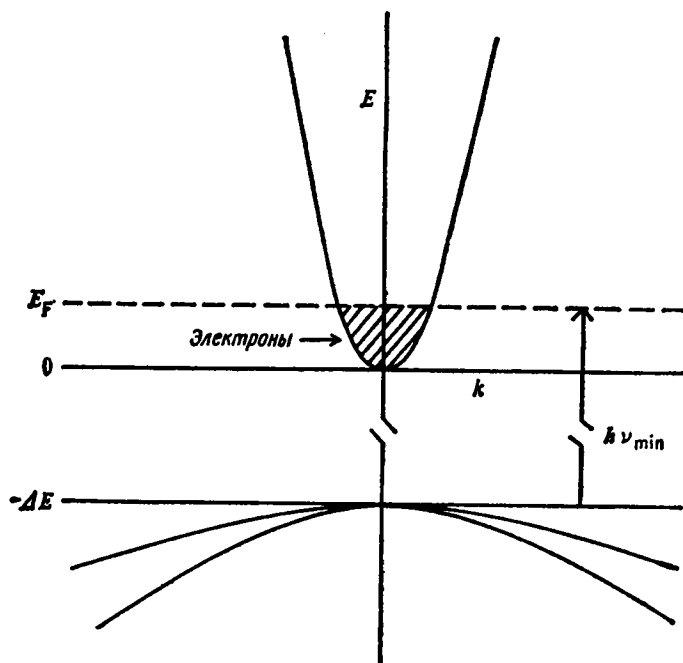


Рис. 14.4. Влияние сильного легирования InSb на положение низкочастотного края полосы фундаментального поглощения.

рое, как предполагают, обусловлено особенностями процессов рассеяния в условиях таких высоких концентраций носителей. При этом предположении в полупроводниках с непрямыми переходами, таких, как Ge, закон сохранения энергии может выполняться и без участия фононов.

14.4.3. ТУННЕЛИРОВАНИЕ В СИЛЬНОЛЕГИРОВАННЫХ ПОЛУПРОВОДНИКАХ

Одним из наиболее интересных новых эффектов, обусловленных сильным легированием, является заметное изменение формы вольт-амперной характеристики $p-n$ -перехода, когда обе области, и n -типа, и p -типа, сильно легированы. Этот эффект впервые обнаружил Есаки [41]. Диоды с такими $p-n$ -переходами обычно называют диодами Есаки. Вольт-амперная характеристика такого перехода для прямого направления показана на рис. 14.5. Точке O , т. е. отсутствию напряжения смещения, соответствует зонная схема $p-n$ -перехода, показанная на рис. 14.6. Ее следует сравнить с

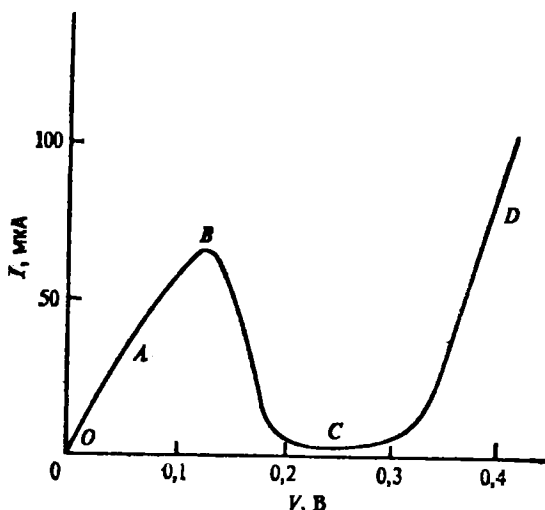


Рис. 14.5. Типичная вольтамперная характеристика (I/V) в прямом направлении для $p-n$ -перехода в сильнолегированном Si (диод Есаки).

Участки, обозначенные O, A, B, C, D , соответствуют рис. 14.6, $o, a, б, в, г$ соответственно.

рис. 7.8 для слаболегированного $p-n$ -перехода. Можно видеть, что разность энергий между E_c и E_v в n - и p -частях вдали от перехода в данном случае оказывается *больше*, чем ширина запрещенной зоны, из-за вырождения в n - и p -частях. Здесь подразумевается, что возбуждение носителей через потенциальный барьер при приложении напряжения будет меньше, чем в слаболегированном переходе. Переход носителей между n - и p -частями имеет, однако, место из-за туннелирования через область перехода при условии, что *имеются* разрешенные уровни с одинаковыми энергиями [42]. В равновесных условиях большинство занятых электронных состояний (т. е. свободных от дырок) в области p -типа лежит напротив занятых электронных состояний в области n -типа, и между ними будет иметь место только небольшое число актов обмена электронами. Однако, как только будет приложено напряжение, некоторые занятые электронные состояния в n -области окажутся напротив пустых электронных (занятых дырочных) состояний в p -области, что проиллюстрировано на рис. 14.6, a . Это ведет к плавному увеличению прямого тока, как показано на участке A на рис. 14.5. Такое увеличение продолжается до тех пор, пока энергетические состояния, заполненные дырками, в p -области не перекроются полностью с состояниями, заполненными электронами, в n -области, как показано на рис. 14.6, $б$. Дальнейшее увеличение напряжения V ведет к уменьшению этого перекрытия, как показано на рис. 14.6, $в$. Ток I

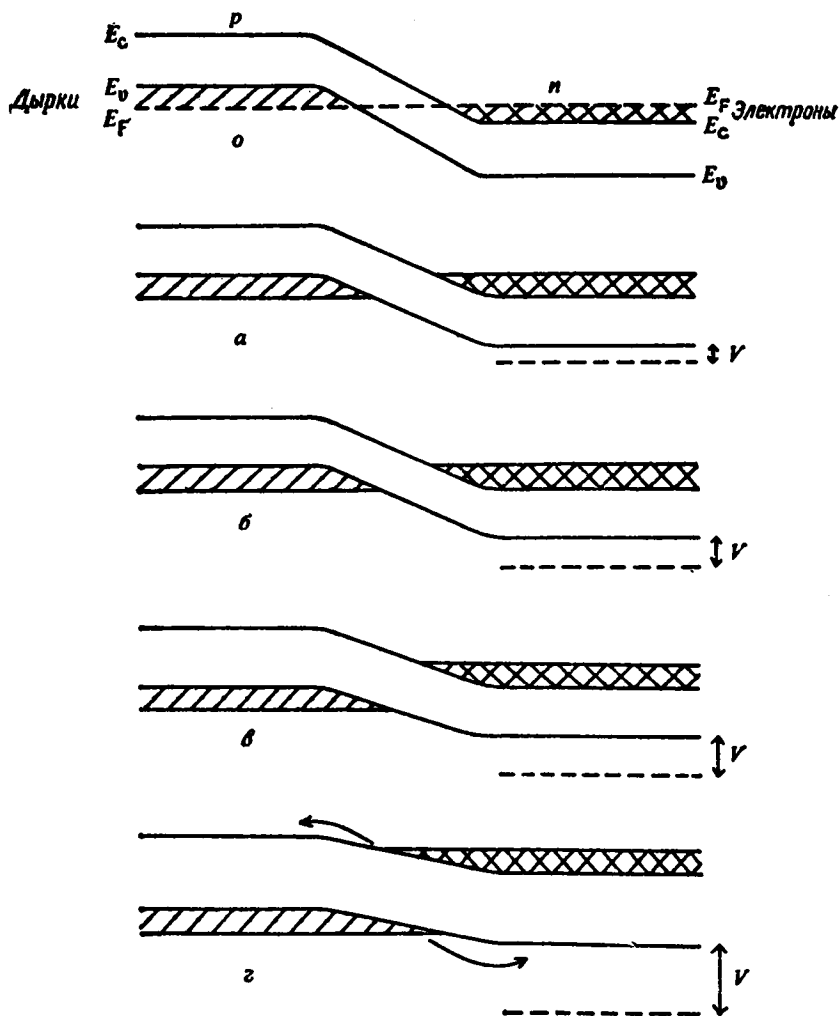


Рис. 14.6. Изменение зонной схемы в сильнолегированном $p-n$ -переходе при приложении к нему напряжения V в прямом направлении.

достигает максимума в точке B на рис. 14.5, а затем быстро падает. Именно этот отрицательный наклон вольтамперной характеристики и придает диоду Есаки его практическую важность. Так как концентрация электронов и дырок возле уровня Ферми при комнатной температуре меняется с энергией не резко, то имеется небольшое перекрытие их занятых состояний, приводящее к некоторому току

на участке C . Величина тока проходит через плоский минимум и остается малой до тех пор, пока напряжение не возрастет настолько, чтобы установился нормальный механизм теплового возбуждения носителей в p — n -переходе через потенциальный барьер, уменьшающийся с увеличением прямого напряжения, как продемонстрировано на рис. 14.6, z . Это обуславливает быстрое возрастание тока, как показано на участке D рис. 14.5.

Туннельные диоды изучались на основе многих полупроводников помимо Ge и Si , например на $GaSb$ и $GaAs$. Все их характеристики весьма схожи, кроме положения максимума в точке B и минимума в точке C на рис. 14.5.

Теорию процесса туннелирования создал Келдыш [43] и развили Такеути и Фунада [44]. Эти теории дают количественное описание простой модели, приведенной выше, а также учитывают более тонкие вопросы, опущенные в нашем описании явления.

Так как процесс туннелирования слабо зависит от температуры, а положение уровня Ферми в случае сильного вырождения также нечувствительно к изменениям температуры, то вольтамперные характеристики вплоть до области минимума при низких температурах будут очень похожи одна на другую. Однако в этой области наблюдается новое явление. Вместо гладкой характеристики, показанной на рис. 14.5, были обнаружены неожиданные увеличения туннельного тока при напряжениях, с которыми можно связать различные возбуждения кристалла, в особенности фононы. Исследования этого эффекта привели к появлению нового раздела в физике полупроводников — *туннельной спектроскопии*, которую мы рассмотрим ниже (см. подразд. 14.4.4). Явление туннелирования в p — n -переходах обсуждается в работе Есаки ([30], стр. 1), а также в серии статей, посвященных различным аспектам данной проблемы (см. сборник «Туннельные явления в твердых телах» [45]).

14.4.4. ЭЛЕКТРОННАЯ ТУННЕЛЬНАЯ СПЕКТРОСКОПИЯ

Есаки и Мияхара [46], Холоньяк и др. [47], изучая характеристики кремниевого туннельного диода, обнаружили на вольтамперной характеристике четыре излома, которые по энергии почти точно соответствовали изломам в спектре поглощения Si , вызванным участием фононов в непрямых переходах (см. разд. 10.5.3). Это явление было интерпретировано как туннельный эффект, происходящий с участием фононов. Неудивительно, что в полупроводнике с непрямой запрещенной зоной такое участие фононов будет способствовать туннельному процессу, так как в процессе перехода квазимпульс электрона должен меняться от значения в максимуме валентной зоны к значению в минимуме зоны проводимости.

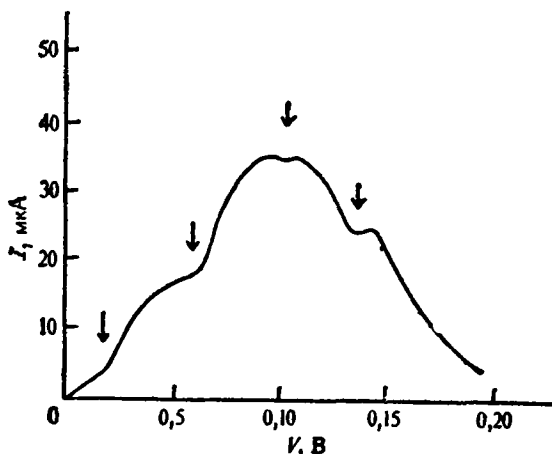


Рис. 14.7. Изломы на кривой I/V , обусловленные туннельным процессом, происходящим с участием фононов, для туннельного диода при температуре жидкого гелия.

В результате этих исследований и появился предмет туннельной спектроскопии. При непосредственном наблюдении *наклона* вольт-амперной характеристики, т. е. при измерении dI/dV в зависимости от V , неоднородности на кривой проявляются более четко, поэтому в настоящее время обычно используется именно эта дифференциальная методика. Для выявления еще более тонкой структуры непосредственно регистрируется даже вторая производная вольт-амперной характеристики: d^2I/dV^2 .

Типичная кривая зависимости I от V для туннельного диода при температурах жидкого гелия приведена на рис. 14.7. Его полезно сравнить с рис. 14.5, на котором представлена характеристика диода при комнатной температуре.

Теорию туннельного процесса, стимулированного фононами, разрабатывали Келдыш [48], Кейн [49] и Клейнман [50] ([45], стр. 181). Эту проблему обсудил Дюк [51], который рассмотрел более широкий класс твердых тел, включая металлы. Обнаружено, что вольт-амперные характеристики контактов металл — полупроводник и контактов металл — окисел — полупроводник также имеют неоднородности. Такие контакты широко используются для изучения фононов и других возбуждений. В число последних входят уровни Ландау (см. разд. 12.5.1), которые дают о себе знать при работе контакта в сильном магнитном поле. Этого следует ожидать из-за резкого изменения плотности состояний в магнитном поле. Плазменные колебания (см. разд. 10.3) также проявляются как неоднородности на кривой зависимости тока от напряжения, и даже некоторые призмесные уровни, вызывающие неожиданные изменения туннельного

тока ([30], стр. 1). Недавно Вольф [52] подробно рассмотрел этот вопрос, сделав обзор теоретических и экспериментальных результатов для широкого класса материалов.

14.5. Эффекты, связанные с высоким давлением

О зонной структуре твердых тел много сведений дала методика, использующая высокие гидростатические давления. Особенно эффективной она оказалась для изучения полупроводников, так как их электронные свойства очень сильно зависят от зонной структуры и, в частности, от ширины запрещенной зоны. Поскольку последняя очень чувствительна к величине межатомного расстояния, следует ожидать, что ширина запрещенной зоны существенно меняется при приложении внешнего давления. На первый взгляд может показаться, что ширины запрещенных зон должны возрастать при увеличении давления, так как уменьшение межатомного расстояния в кристалле обычно означает усиление межатомной связи. Однако оказалось, что это не так, и в кристаллах, имеющих несколько запрещенных зон, обусловленных несколькими минимумами зоны проводимости, одни зоны при увеличении давления становятся шире, а другие уже. В полупроводниках с прямой запрещенной зоной ΔE обычно увеличивается при увеличении давления, однако есть несколько исключений, таких, как Te и халькогениды свинца. Различное поведение минимумов при $k=(000)$ и в направлениях $\langle 111 \rangle$ и $\langle 100 \rangle$ в кристаллах, имеющих структуру алмаза и цинковой обманки, поставило ряд вопросов, которые предстоит разрешить, сравнивая соединения типа $A^{III}B^V$, Si и Ge .

Перед тем как все это обсуждать, мы должны указать на небольшое различие в единицах измерения давления, используемых в литературе, которая в настоящее время очень обширна. Наиболее часто используются единицы $10^3 \text{ кг} \cdot \text{см}^{-2}$ и кбар, которые связаны соотношением $1 \text{ кбар} \approx 1,02 \cdot 10^3 \text{ кг} \cdot \text{см}^{-2}$. Если не требуется измерять давления с высокой точностью, то эти единицы практически одинаковы. (Заметим также, что $1 \text{ кбар} = 987 \text{ ат}$.)

Вполне заметные эффекты можно наблюдать уже при относительно умеренных давлениях, скажем до 15 кбар, но для наблюдения более значительных эффектов необходимы давления вплоть до 500 кбар. Довольно однородное гидростатическое давление вплоть до величины в 100 кбар получить можно, но выше этого предела имеется некоторая неопределенность. Методы получения и измерения различных давлений подробно описаны в нескольких книгах, посвященных этому вопросу [53].

Используя давления до 12 кбар, вполне доступные на относительно простом оборудовании, Кейес [54] показал, что электропро-

водность InSb p -типа ($N_a \approx 10^{17} \text{ см}^{-3}$) быстро падает с увеличением давления при температуре 85°C , в то время как при 0°C она остается почти постоянной. Это показывает, что при 85°C и нулевом давлении проводимость является собственной, так что изменение ее большей частью обусловлено изменением ΔE при изменении давления, тогда как при 0°C проводимость оказывается примесной и обусловленной в основном наличием ионизованных акцепторов. Такого рода исследования дали много информации о механизмах проводимости в различных полупроводниках. Изменение деформационных потенциалов (см. разд. 8.5) при изменении давления помогло выявить сложные зависимости между ними. Основы теории эффектов, обусловленных давлением, а также обзор экспериментальных работ, посвященных главным образом явлениям переноса при наличии давления, дал Кейес [55]. Подробные исследования Ge и Si описаны в обзорной статье Пола и Брукса [56].

Видимо, наиболее важными параметрами, которые следует изучать в условиях давления, являются ширина запрещенной зоны ΔE и различные другие ширины зон в зонной структуре. Их можно определять методом оптического поглощения, который не требует применения электрических контактов; в условиях высокого давления изготовление их часто является сложной задачей. Пол с сотрудниками, а также Дрикамер с сотрудниками выполнили много работ по этому вопросу, в основном в области более высоких давлений. Пол и Варшауер [53], стр. 179) сделали обзор этих работ. Исследования в области очень высоких давлений описал Дрикамер [57].

Как определили Сликхаус и Дрикамер [58], при увеличении давления ширина непрямоугольной запрещенной зоны в Ge увеличивается с крутизной $7,5 \cdot 10^{-3} \text{ эВ} \cdot \text{кбар}^{-1}$, в то время как в Si эта величина уменьшается с крутизной $2 \cdot 10^{-3} \text{ эВ} \cdot \text{кбар}^{-1}$. Однако в Ge при давлении около 50 кбар крутизна увеличения становится равной нулю и при более высоких давлениях стремится к такой же величине, как в Si, а именно к $-2 \cdot 10^{-3} \text{ эВ} \cdot \text{кбар}^{-1}$. Это проиллюстрировано на рис. 14.8. В дальнейших экспериментах с соединениями $A^{III}B^V$ было установлено, что такое поведение обусловлено тем фактом, что ширина запрещенной зоны в точке (111) увеличивается при увеличении давления, а вблизи точки (100) уменьшается. Оказывается также, что, за немногими исключениями, крутизна изменения ширины запрещенной зоны с давлением почти одинакова в Ge, Si и в полупроводниках типа $A^{III}B^V$. Это показывает, что в Ge край зоны в точке (100) (X -точка) при некотором давлении лежит ниже края зоны в точке (111) (L -точка), а при большем давлении Ge имеет зонную структуру, сходную с зонной структурой Si. Для давления, соответствующего этому переходу, Пол и Брукс [56] дают величину 35 кбар, которая хорошо согласуется с результатами измерений, проведенных Сликхаусом и Дрикамером [58].

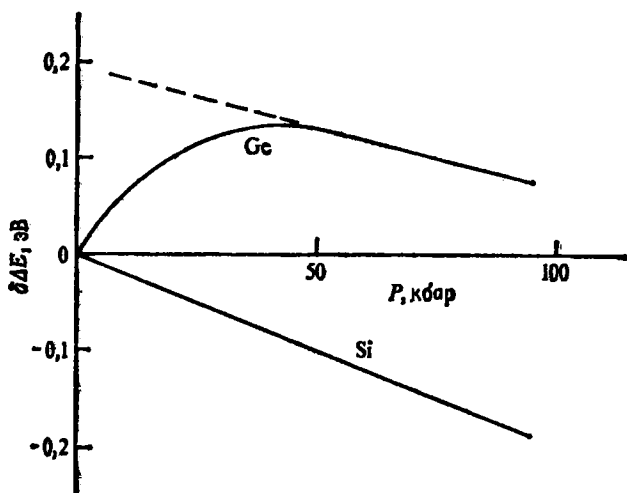


Рис. 14.8. Изменение ширины запрещенной зоны ΔE в зависимости от давления P для Si и Ge [58].

Было найдено, что при увеличении давления ширина прямой запрещенной зоны в Ge увеличивается с крутизной $15 \cdot 10^{-3}$ эВ \times кбар $^{-1}$. Близкая по величине крутизна была найдена для некоторых прямозонных полупроводников типа $A^{III}B^V$, в частности для InSb и GaAs, для которых она приблизительно равна $12 \cdot 10^{-3}$ эВ \times кбар $^{-1}$. Эдвардс и Дрикамер [59] наблюдали в GaAs интересную перестройку от прямой к непрямои запрещенной зоне. Для малых давлений крутизна имеет упомянутую выше величину $12 \cdot 10^{-3}$ эВ \times кбар $^{-1}$, но при давлении 18 кбар она меняется на $7,3 \cdot 10^{-3}$ эВ \times кбар $^{-1}$, что характерно для минимума в точке (111). Отсюда видно, что минимум при $k=(000)$ поднялся выше, чем минимум при $k=(111)$, и при более высоких давлениях запрещенная зона должна быть непрямои. При значительно более высоком давлении, около 50 кбар, крутизна увеличения становится равной нулю, а затем отрицательной, показывая тем самым, что наиболее низкими теперь являются минимумы при $k=(100)$. Аналогичные эффекты наблюдались и для ряда других соединений типа $A^{III}B^V$.

Во многих полупроводниках изучалось влияние давления на энергию ионизации примесей с мелкими уровнями. Эффект обычно очень мал, наблюдаемые крутизны изменения имеют порядок нескольких единиц $\times 10^{-5}$ эВ \cdot кбар $^{-1}$. Для некоторых примесей с глубоко лежащими уровнями наблюдаемые крутизны изменения имеют порядок 10^{-3} эВ \cdot кбар $^{-1}$. Этот вопрос обсуждался в обзоре Пола и Варшауера ([53], стр. 179).

Когда используемые давления превышают 100 кбар, наблюдаются намного более сильные эффекты. Например, Миномура и Дрикамер [60] наблюдали как в Si, так и в Ge неожиданное уменьшение сопротивления по величине более чем на 5 порядков. В Si это происходит приблизительно при 190 кбар, а в Ge приблизительно при 120 кбар. При более высоких давлениях их сопротивление возрастает с температурой, так же как у металлов. При переходной температуре как в Si, так и в Ge Джемисон [61] наблюдал изменение кристаллической структуры. Структура изменяется от решетки алмаза к тетрагональной, которая сходна с решеткой металлического олова. Таким образом, это изменение аналогично переходу от серого олова к белому олову и является переходом от полуметалла к металлу (см. разд. 13.8).

Самара и Дрикамер [62] наблюдали аналогичные переходы в некоторых полупроводниках типа $A^{III}B^V$, а также в некоторых полупроводниках типа $A^{IV}B^{VI}$, таких, как ZnTe. Среди соединений типа $A^{III}B^V$ наибольшим давлением, соответствующим переходу, характеризуется GaAs, который не переходит в металлическое состояние, пока давление не достигнет 250 кбар.

При изменении давления меняются многие свойства полупроводников, в частности подвижность носителей заряда. Натан, Пол и Брукс [63] изучали влияние давления на междолинное рассеяние (см. разд. 8.7). Они обнаружили, что в Ge при комнатной температуре при изменении давления от нуля до 30 кбар подвижность уменьшается в 5 раз, причем это изменение не было линейным; авторы нашли правильное объяснение характера изменения подвижности с давлением.

Изменения некоторых других свойств при изменении давления, включая показатель преломления, обсуждались в обзорах Кейеса, Пола и Брукса, Пола и Варшауера, а также Дрикамера, на которые мы уже ссылались выше.

14.6. Лазерный эффект в полупроводниках

Обсуждая рекомбинационное излучение в разд. 9.2, мы рассматривали только спонтанное испускание излучения при переходе с верхнего уровня на нижний, при этом считалось, что темп испускания не зависит от локальной интенсивности излучения той же частоты. Однако еще в 1917 г. Эйнштейн [64] указал, что следует принимать во внимание и другой процесс, а именно вынужденное испускание, при котором темп испускания пропорционален плотности энергии излучения в единичном частотном интервале. Если плотность энергии не слишком высока, то в видимой и ближней инфракрасной области спектра этим процессом обычно можно пренебречь. Однако он имеет принципиальное значение, если условия та-

ковы, что интенсивность излучения может расти, приводя к лазерному эффекту, т. е. к резонансной генерации излучения. В настоящее время лазеры хорошо известны как источники интенсивного когерентного, почти монохроматического излучения в широком спектральном диапазоне. Их физические основы и практическая разработка были описаны в большом количестве статей и в целом ряде книг. Нас будет интересовать лазерный эффект только в полупроводниках, и, так же как в случае электронных приборов на их основе, мы воздержимся от описания развития лазерной техники и сосредоточим свое внимание на физических основах лазерного эффекта в полупроводниках. Физические принципы работы лазеров рассмотрены в нескольких книгах [65, 66].

Мы уже встречались с явлением вынужденного испускания в случае испускания фононов. Тогда мы использовали тот факт, что поглощение фононов пропорционально их плотности N_p , а испускание фононов пропорционально $(1+N_p)$, где единица соответствует спонтанному испусканию, а N_p — вынужденному (см. разд. 8.7, 12.2). На основании термодинамических рассуждений Эйнштейн [64] показал, что вероятность вынужденного испускания равна вероятности поглощения и ввел коэффициент B для их описания и коэффициент A для описания спонтанного испускания. Эти коэффициенты называют коэффициентами Эйнштейна, их величины можно получить при помощи простого вычисления.

Для начала рассмотрим два определенных уровня с энергиями E_1 и E_2 , такими, что $E_2 > E_1$. Этим уровням соответствует излучение с угловой частотой ω , причем $\hbar\omega = E_2 - E_1$. Предположим, что N_1 систем находятся на уровне 1, а N_2 систем находятся на уровне 2, так что выполняется соотношение $N_2/N_1 = \exp(-\hbar\omega/kT)$. Если $\mathcal{W}(\omega)$ — плотность энергии в единичном интервале значений угловой частоты, то из определения коэффициентов Эйнштейна имеем

$$\frac{dN_2}{dt} = -AN_2 - BN_2\mathcal{W}(\omega) + B'N_1\mathcal{W}(\omega). \quad (14.5)$$

В условиях термодинамического равновесия $dN_2/dt=0$, тогда

$$\mathcal{W}(\omega) = \frac{AN_2}{B'N_1 - BN_2} = \quad (14.6)$$

$$= \frac{A}{B' \exp\left(\frac{\hbar\omega}{kT}\right) - B}. \quad (14.6a)$$

В условиях равновесия эта формула должна сводиться к формуле Планка для интенсивности излучения (см. разд. 9.2), но выраженной в единицах частоты ω . Таким образом, в условиях равновесия должно выполняться равенство $B=B'$, т. е. вероятность спонтанного испускания должна быть равна вероятности поглощения для

каждой системы в отдельности. Кроме того,

$$\frac{A}{B\overline{W}(\omega)} = \exp\left(\frac{\hbar\omega}{kT}\right) - 1 = N_p^{-1}, \quad (14.7)$$

где N_p — среднее число заполнения для фотонов частоты ω . Точно такое же рассуждение применимо к фононам с тем же результатом. Величину A можно получить из формулы Планка, но в дальнейшем она нам не потребуется.

Приведенные выше соотношения справедливы в условиях теплового равновесия, однако понятие *вероятности* можно применять и в случае неравновесного излучения. Эти соотношения можно строго вывести на основании квантовой теории излучения (см., например, работу Лоудона [65]). Некоторые трудности возникают, если $\overline{W}(\omega)$ надо применить для описания монохроматического излучения, для которого плотность в единичном интервале частот обращается в бесконечность. Эти трудности, однако, можно преодолеть с помощью квантовой теории излучения.

Теперь рассмотрим ситуацию, при которой отношение N_2/N_1 не соответствует тепловому равновесию, а $\overline{W}(\omega)$ не соответствует плотности энергии теплового излучения. Если на время пренебречь спонтанным испусканием, которое вносит потери в систему, то темп поглощения излучения будет пропорционален $B(N_1 - N_2)$. На основании этого можно ввести коэффициент поглощения в виде

$$\overline{\alpha} = \alpha \frac{N_1 - N_2}{N}, \quad (14.8)$$

где $N = N_1 + N_2$. Это выражение сводится к обычному коэффициенту поглощения в случае, когда $N_2 \ll N_1$, что обычно имеет место на оптических частотах, за исключением условий сильного внешнего возбуждения. Однако, если каким-либо образом N_2 может быть сделано больше, чем N_1 , коэффициент поглощения станет отрицательным и вместо потерь возникнет *усиление*. Если это усиление превышает потери, обусловленные спонтанным излучением и другими причинами, то по мере прохождения луча через такую среду интенсивность излучения будет расти. В этом и состоит лазерный эффект. Процесс создания условия $N_2 > N_1$, т. е. инверсной заселенности уровней при помощи внешних средств, часто называют «накачкой».

Если в полупроводнике создано большое количество электронно-дырочных пар путем инжекции через контакты, либо иным способом, например с помощью внешнего излучения, то в нем образуется естественная система с инверсной заселенностью. Верхние состояния соответствуют электронам в зоне проводимости, а нижние состояния — дыркам вблизи вершины валентной зоны, так что нижние состояния свободны от электронов. Неудивительно, что при надлежащих условиях в таком полупроводнике может возникнуть лазерный эффект. Конечно, реальная система намного сложнее рас-

смотренной простой двухуровневой схемы, однако основные принципы остаются неизменными. Вводя плотность состояний электронов N_e и плотность состояний дырок N_h , можно выразить темп генерации G для электронных уровней с $E \simeq E_2$ и дырочных уровней с $E \simeq E_1$ в виде (см. разд. 9.2)

$$G = BN_e(E) N_h(E) P_e(E_1) P_h(E_2) W(\omega), \quad (14.9)$$

где $P_e(E)$ и $P_h(E)$ — вероятности заполнения, определенные в разд. 4.1. Темп рекомбинации R дается формулой

$$R = BN_e N_h P_h(E_1) P_e(E_2) W(\omega) + AN_e N_h P_h(E_1) P_e(E_2), \quad (14.10)$$

причем в (14.9) и (14.10) коэффициент B один и тот же. В стационарном случае темп генерации должен быть равен темпу рекомбинации. Приравнявая (14.9) и (14.10), получаем

$$BW(\omega) \left[\exp \frac{\hbar\omega}{kT} - 1 \right] = A, \quad (14.11)$$

что совпадает с уравнением (14.7). Опять же если мы выразим темп генерации и темп рекомбинации в единицах вероятности на одну электронно-дырочную пару, то найдем, что отношение вероятностей вынужденного и спонтанного излучения дается формулой (14.7).

До сих пор мы не упоминали очень важного свойства вынужденного излучения, а именно, что оно имеет направление, поляризацию и частоту, такие же, как у вынуждающего его излучения, и, таким образом, когерентно с ним. Излучение же, возникающее в результате спонтанного испускания, имеет случайное направление и поляризацию.

При больших отклонениях от теплового равновесия понятие *вероятностей* соответствующих переходов остается еще применимым, поэтому соображения, приводящие к возникновению отрицательных величин коэффициента поглощения, т. е. положительного коэффициента *усиления*, также остаются еще верными. Сначала рассмотрим полупроводник, в котором созданы концентрации электронов и дырок, соответствующие сильному вырождению, т. е. полупроводник, у которого зона проводимости заполнена до квазиуровня Ферми E_{Fe} , а валентная зона оказывается пустой выше уровня $E_{Fh} = -\Delta E - E'_{Fh}$.

Такая ситуация проиллюстрирована на рис. 14.9. На рисунке видно, что поглощаться могут только фотоны с энергией большей, чем $\Delta E + E_{Fe} + E'_{Fh}$. Это напоминает эффект Бурштейна — Мосса с той только разницей, что в формуле для эффекта Бурштейна — Мосса отсутствует член E_{Fh} , так как там мы имеем дело либо с зоной проводимости, заполненной до уровня E_F , либо с валентной зоной, пустой выше уровня E_F (см. подразд. 14.4.2). В отличие от поглощения испускание может происходить на частотах, лежащих в ин-

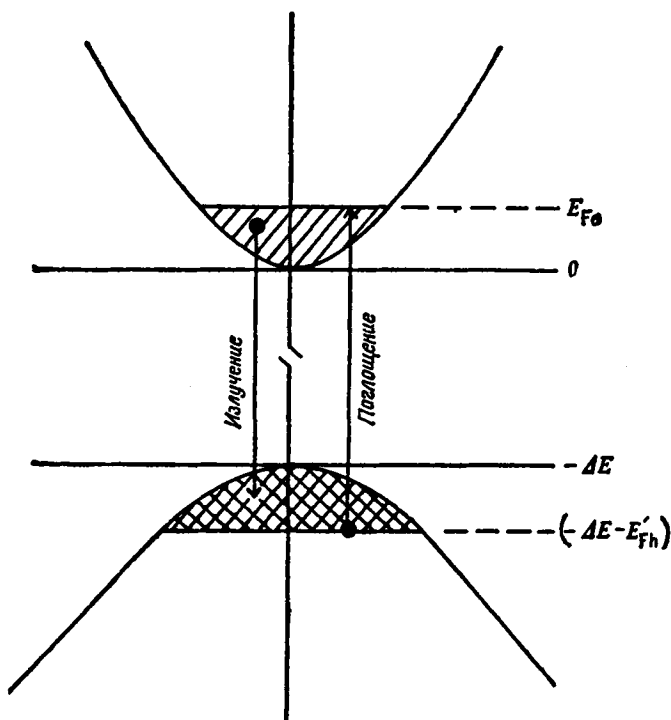


Рис. 14.9. Поглощение и вынужденное испускание из области с высокой концентрацией электронов и дырок, созданных путем инжекции.

тервале $\Delta E < h\nu < \Delta E + E_{Fe} + E'_{fh}$. Так как в данной ситуации нижний уровень пуст, то это означает полную инверсию заселенности уровней, что должно приводить к значительному усилению излучения g (отрицательное значение α). Реально достигнуты довольно большие значения g , до 100 см^{-1} и более, так что полупроводники являются подходящими материалами для получения лазерного эффекта. Если на такую систему не наложены дополнительные ограничения, то ее спектр излучения будет все же довольно широким. Так же как и во всех лазерных системах с положительным усилением в широком диапазоне частот, для увеличения эффективной длины пробега излучения в материале и для ограничения частотного диапазона лазерной генерации применяются оптические резонаторы. У полупроводникового лазера оптический резонатор часто представляет собой монокристалл с двумя параллельно сколотыми противоположными гранями. Хотя лазерный эффект удалось наблюдать и в однородном материале при сильном возбуждении с помощью другого лазера, несомненно, более распространенным

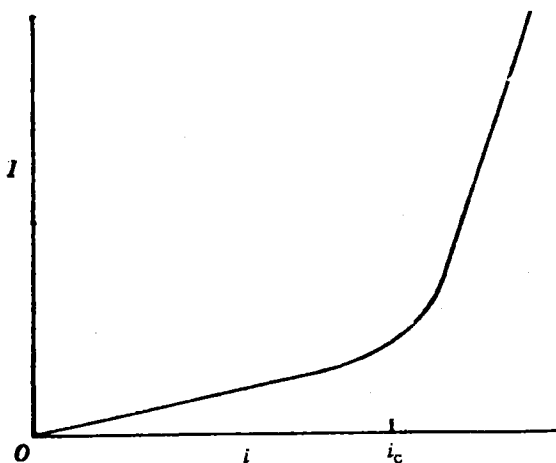


Рис. 14.10. Иллюстрация резкого возрастания интенсивности I выходного излучения при достижении критической величины i_c инжекционного тока через лазерный диод.

методом возбуждения является инжекция носителей заряда из p - n -перехода. В этом случае активная область состоит из узкой области перехода, в которой присутствуют и электроны, и дырки, ограниченной двумя параллельно сколотыми гранями. В более современных устройствах используются гетеропереходы. Как уже обсуждалось в разд. 7.14, последние имеют преимущества, заключающиеся в меньшем поглощении излучения свободными носителями и в возможности расположить активную область между областями материала с меньшим показателем преломления и, таким образом, заключить излучение в резонаторе.

Предположим, что резонатор имеет длину L , а коэффициенты отражения на гранях равны R_1 и R_2 . Если α — потери на единице длины, вызванные посторонним поглощением и потерями излучения, а g — усиление за счет вынужденного испускания, то излучение, имевшее на входе в резонатор интенсивность I_0 , прошедшее через резонатор и вернувшееся в ту же точку, будет иметь интенсивность

$$I = I_0 R_1 R_2 \exp(2gL - 2\alpha L). \quad (14.12)$$

Следовательно, условие для возникновения лазерного эффекта имеет вид

$$g > \alpha - \frac{1}{2L} \ln(R_1 R_2). \quad (14.13)$$

Если через лазерный диод начать пропускать ток, то при малых величинах тока излучение будет обусловлено спонтанным испус-

канием рекомбинационного излучения. На таком принципе основана работа светоизлучающих диодов (СД). Если затем ток через лазерный диод превысит некоторую величину, то интенсивность излучения диода резко возрастет. Это проиллюстрировано на рис. 14.10. Одновременно наблюдается сужение спектра излучения, что также указывает на начало лазерного эффекта.

Как мы видели, излучательная рекомбинация — это относительно малоэффективный процесс, поэтому только малая часть энергии, затраченной на инжекцию электронов, выделяется в виде излучения. Остальная часть выделяется в виде тепла и тем самым ограничивает инжекционный ток, если, конечно, не используется импульсный режим накачки. Именно импульсный режим использовался во всех ранних работах.

Мы обсуждали лазерный эффект на переходах типа зона — зона, однако ясно, что он может происходить и с участием примесных центров. Некоторые из них, дающие сильное излучение, вызванное рекомбинацией связанных экситонов, мы уже обсуждали в разд. 10.14. В этом случае зона проводимости в лазерном процессе аналогична широкой полосе состояний, в которую происходит накачка электронов в газовом лазере перед тем, как они попадут в верхнее активное состояние.

14.6.1. ЛАЗЕРНЫЕ МАТЕРИАЛЫ

В первых работах по полупроводниковым лазерам, которые выполнили Р. Н. Холл и др. [67], Натан и др. [68], а также Квист и др. [69], в качестве активного материала использовался GaAs в основном потому, что он имеет прямую запрещенную зону, которая, как мы видели, ведет к более высокому темпу рекомбинации, чем в случае непрямых переходов, происходящих с участием фононов. Арсенид галлия применялся еще и потому, что существовала разработанная методика изготовления светодиодов из этого материала. В экспериментах использовался импульсный режим работы при 77 К. Считалось, что халькогениды свинца, которые имеют *прямые* запрещенные зоны вблизи краев зоны Бриллюэна в направлениях $\langle 111 \rangle$, также хорошо подходят для изучения лазерного эффекта. Авторы первых работ, посвященных исследованию этих материалов, особенно Гурвиц, Калава и Редикер [70], использовали накачку электронным лучом, однако было обнаружено, что возможен также и диодный режим работы. Для возбуждения лазерного эффекта в GaAs, CdTe, CdSe и CdS Басов и др. ([25], стр. 277) использовали оптическую накачку рубиновым лазером.

Более поздние технические разработки показали, что полупроводниковые лазеры могут быть сделаны на основе различных материалов, включая халькогениды свинца и их сплавы, такие, как $Pb_{1-x}Cd_xS$, и различные сплавы соединений типа $A^{III}B^V$. Были ис-

пользованы также узкозонные материалы $Pb_{1-x}Sn_xTe$ и $Pb_{1-x}Sn_xSe$ (см. разд. 13.8). На основе этих материалов были созданы полупроводниковые лазеры, перекрывающие по частям спектральный диапазон от 0,6 мкм до 32 мкм. Эти лазеры можно перестраивать по длине волны путем изменения тока инжекции, каждый в своем узком спектральном диапазоне. При этом изменяется температура лазера, а следовательно, и ширина запрещенной зоны и в свою очередь частота излучения лазера. Лазеры этого типа, а также их применение в молекулярной спектроскопии подробно описали Мурадян ([71], стр. 75), а также Мелнгайлис и Мурадян ([72], стр. 1).

Лазеры на основе GaAs с гетеропереходами $Ga_{1-x}Al_xAs$ в настоящее время могут работать в непрерывном режиме при комнатной температуре, лазеры же на солях свинца для этого следует охлаждать до 77 К. Однако лазеры этого типа обычно работают в импульсных режимах. Для перестройки излучения лазеров используются и другие различные методы, включая применение гидростатического давления и магнитного поля, тем не менее перестройка путем изменения тока остается, несомненно, самой простой. Мурадян ([71], стр. 75) при работе лазера в непрерывном режиме с выходной мощностью порядка нескольких мВт наблюдал очень узкие линии излучения, шириной около 50 кГц.

Одни из первых обзоров по разработке полупроводниковых лазеров сделали Квиссер [73] и Пилькун [74].

14.7. Рассеяние излучения, обусловленное переворачиванием спина, и лазеры на этой основе

Другое новое явление в полупроводниках, которое нашло себе применение в лазерах, — это одна из форм комбинационного рассеяния света, известная как «переворот спина» (spin-flip phenomenon). Явление заключается в рассеянии света электронами зоны проводимости, сопровождающееся переворачиванием спина электрона. Если электрон находится в магнитном поле, то его спиновая энергия $g\beta V$ (см. разд. 12.6) либо отбирается от энергии рассеянного кванта, либо добавляется к ней, что приводит соответственно к уменьшению или увеличению частоты излучения на величину $g\beta V/h$. На языке комбинационного рассеяния в первом случае возникает стоксов сдвиг частоты, во втором случае — антистоксов. Оказывается, что переходы такого типа запрещены для электронов в свободном пространстве, но не запрещены для электронов в зоне проводимости в полупроводнике, имеющем вырожденную валентную зону.

Другие виды комбинационного рассеяния на электронах в полупроводнике, находящемся в магнитном поле, возникают из-за

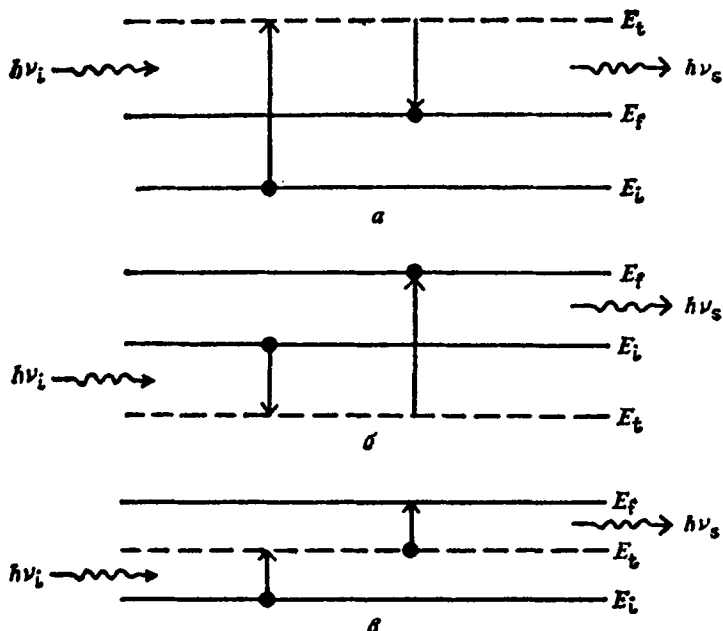


Рис. 14.11. Комбинационное рассеяние, представленное в виде двухступенчатого процесса, включающего промежуточное состояние ($h\nu_s = h\nu_i - E_f + E_t$).

а) $E_t > E_f > E_i$; б) $E_t < E_i < E_f$; в) $E_i < E_t < E_f$.

переходов между уровнями Ландау в зоне проводимости. Комбинационное рассеяние происходит путем двухступенчатого перехода, в некотором смысле сходного с непрямыми межзонными переходами, хотя в данном случае фононы в нем не участвуют. Падающее излучение вызывает переход из исходного состояния i в промежуточное состояние t , а затем из состояния t в конечное состояние f . Промежуточное состояние занято в течение столь короткого отрезка времени, что при этом не требуется выполнения закона сохранения энергии, точно так же как для промежуточного состояния в межзонных переходах, которые обсуждались в подразд. 10.5.3. Энергия будет сохраняться в целом, если сравнивать энергию квантов падающего и рассеянного излучения с энергией начального и конечного состояний. Промежуточное состояние может быть выше начального или конечного состояний, быть ниже их и даже может лежать между ними. Эти три ситуации показаны на рис. 14.11. Промежуточное состояние должно быть реальным состоянием, на которое может произойти разрешенный переход. Если $E_i < E_t$, то частота рассеянного излучения ν_s будет меньше частоты

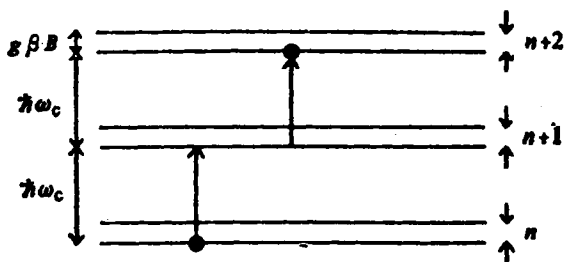
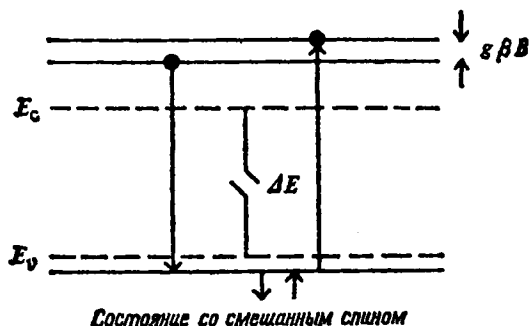


Рис. 14.12. Рассеяние при переходе с $\Delta n=2$ через промежуточный уровень Ландау без изменения направления спина ($h\nu_s = h\nu_i - 2\hbar\omega_c$).



Состояние со смешанным спином

Рис. 14.13. Рассеяние с переворотом спина через промежуточное состояние со смешанным спином, находящееся в валентной зоне ($h\nu_s = h\nu_i - g\beta B$).

падающего излучения ν_i , так как

$$h\nu_s = h\nu_i - (E_f - E_i). \quad (14.14)$$

В этом случае будет иметь место стоксов сдвиг частоты. Если $E_i > E_f$, то сдвиг будет антистоксов. Для уровней Ландау в зоне проводимости Вольф [75] теоретически рассмотрел ситуацию, при которой состояние i соответствует уровню Ландау (n), состояние f — уровню Ландау ($n+2$), а промежуточным состоянием является уровень ($n+1$). Это проиллюстрировано на рис. 14.12. Оказывается, что разрешено только рассеяние, для которого $\Delta n=2$. При таких переходах изменения направления спина запрещены.

Более того, если разность энергий уровней Ландау ($n+2$) и ($n+1$) в точности совпадает с разностью энергий для уровней ($n+1$) и (n), то вероятность такого перехода равна нулю. Это должно выполняться для электронов в свободном пространстве и для полупроводников, для которых E квадратично зависит от компонент вектора k . В таких полупроводниках комбинационное рассеяние этого типа не имеет места. Слашер, Пейтел и Флери [76] наблюдали это явление в InSb, в котором его существование обусловлено отклоне-

нием от параболичности зоны проводимости в InSb из-за малости эффективной массы электронов (см. подразд. 11.3.5). Против ожидания авторам удалось также наблюдать переход с $\Delta n=1$.

Яфет [77] рассмотрел ситуацию, при которой начальное состояние — это уровень Ландау $n=0$ со спином, направленным «вверх» или «вниз», конечное состояние — это уровень $n=0$ с противоположным направлением спина, а промежуточное состояние находится на вершине валентной зоны (см. рис. 14.13). В этом случае переход разрешен для состояния со смешанным спином с $n=1$, как это имеет место в InSb. Такой переход в InSb наблюдали Слашер, Пейтел и Флери [76], причем оказалось, что этот переход намного острее, чем переход с $\Delta n=2$. Теорию переходов с переворотом спина расширили Деннис и др. [78], подчеркнув важную роль промежуточного состояния со смешанным спином. Если электрон возбудить в это состояние, то первоначальное направление его спина теряется и имеется большая вероятность того, что он перевозбудится в состояние с $n=0$ в зоне проводимости со спином, направленным либо вверх, либо вниз. Рассеяние света, сопровождающееся переворотом спина электрона, происходит именно при тех переходах; при которых электрон возбуждается с *изменением* спина. Оказалось, что InSb — это единственный в своем роде материал, и он особенно подходит для изучения этого типа рассеяния, которое настолько сильно, что *рассеянное* излучение можно было бы легко использовать для создания лазера.

В таких лазерах входное излучение частоты ν_1 можно рассматривать как излучение накачки, а рассеянное излучение и его прирост в результате вынужденного испускания можно рассматривать как выходное излучение. С помощью фильтров его можно легко отделить от излучения накачки. Опять же InSb является особенно предпочтительным материалом из-за очень большой величины g -фактора (~ 40) (см. разд. 13.5), что приводит к намного большему частотному сдвигу, чем в материалах с нормальной величиной g -фактора. Лазерный эффект, основанный на рассеянии излучения, связанном с переворотом спина, впервые наблюдали Пейтел и Шоу [79], используя поперечное рассеяние. Вскоре после них лазерный эффект наблюдали Оллвуд и др. [80], а также Мурадян и др. [81], которые использовали продольное рассеяние. Хотя при продольном рассеянии возникают дополнительные трудности по отделению лазерного излучения от излучения накачки, однако интенсивность лазерного излучения больше, чем при поперечном рассеянии.

Благодаря тому что частотный сдвиг излучения пропорционален величине магнитной индукции B , лазеры, основанные на переворачивании спина, представляют непрерывно перестраиваемый по частоте источник излучения и поэтому привлекают к себе значительное внимание.

В качестве источников излучения накачки при этом обычно использовались CO_2 -лазеры с длиной волны излучения около 10,6 мкм и CO -лазеры с длиной волны излучения около 5,3 мкм. CO -лазер обладает тем преимуществом, что его длина волны излучения находится почти в резонансе с переходом между начальным и промежуточным состояниями, что значительно увеличивает интенсивность комбинационного рассеяния. Были сделаны попытки использовать в лазерах, основанных на перевороте спина, ряд других материалов помимо InSb , однако ни один из них, кроме, возможно, $\text{Hg}_x\text{Cd}_{1-x}\text{Te}$ (см. разд. 13.8), не давал такого интенсивного рассеяния, как InSb .

Необходимость в получении непрерывно перестраиваемого узкополосного источника в инфракрасном диапазоне спектра стимулировала проведение большого количества работ, посвященных исследованию явления рассеяния света с переворотом спина, уже вскоре после открытия этого явления. Быстро появилась обширная литература, посвященная этому вопросу. Обзор последовавших теоретических работ выполнил Скотт ([72], стр. 123). С. Д. Смит ([71], стр. 13) и Мурадян ([71], стр. 82) рассмотрели физические проблемы, связанные с развитием таких лазеров и их применением в спектроскопии.

14.8. Комбинационное рассеяние на колебаниях атомов решетки

Рассеяние света на фононах, отвечающих колебаниям атомов решетки, совершенно аналогично обычному комбинационному рассеянию на нормальных колебаниях молекул. Экспериментальные методы здесь намного сложнее, однако задача сильно упрощается из-за наличия лазеров с их мощным монохроматическим излучением. Эти методы подробно описал Мурадян [82]. Теорию комбинационного рассеяния на колебаниях атомов решетки создал Лоудон [83], который рассмотрел также экспериментальные результаты и привел информацию о частотах фононов, полученную этим способом. Оказалось, что эта информация является ценным дополнением к той, которая получена методами абсорбционной спектроскопии и рассеяния нейтронов. В частности, тот факт, что некоторые из правил отбора, действующих при поглощении, отличаются от правил отбора, определяющих этот тип рассеяния, позволил более определенно идентифицировать некоторые из фононов.

ЛИТЕРАТУРА

1. *Lampert M. A.*, Phys. Rev. Lett., 1, 450 (1958).
2. *Haynes J. R.*, Phys. Rev. Lett., 17, 860 (1966).
3. Proc. Xth Intern. Conf. on Phys. of Semiconductors, U. S. Atomic Energy Commission, 1970.

4. Optical Properties of Solids — New Developments, ed. Seraphin B. O., North Holland, 1976.
5. Akimoto O., Hanamura E., Journ. Phys. Soc. Japan, 33, 1537 (1972); Solid State Comm., 10, 253 (1962); 11, xiii (1962).
6. Akimoto O., Journ. Phys. Soc. Japan, 35, 973 (1973).
7. Wang J. S., Kittel C., Phys. Lett., 42A, 189 (1972).
8. Merz J. L., Faulkner R. A., Dean P. J., Phys. Rev., 188, 1228 (1969).
9. Nakata N. et al., Journ. Phys. Soc. Japan., 38, 903 (1975).
10. Келдыш Л. В., Труды IX Международной конференции по физике полупроводников (Москва).— Л.: Наука, 1968, с. 1303.
11. Brinkman W. F. et al., Phys. Rev. Lett., 28, 1961 (1972); Phys. Rev., B7, 1508 (1973); B8, 1570 (1973).
12. Combescot M., Nozières P., Journ. Phys. C: Solid State Phys., 5, 2369 (1972).
13. Покровский Я. Е., Свицтунова К. И., Письма ЖЭТФ, 9, 435 (1969).
14. Аснин В. М., Рогачев А. А., Саблина Н. И., Письма ЖЭТФ, 11, 162 (1970).
15. Покровский Я. Е., Свицтунова К. И., Письма ЖЭТФ, 13, 297 (1971).
16. Багаев В. С., Пенин Н. А., Сибельдин Н. Н., Цветков В. А., ФТТ, 15, 3269 (1973).
17. Feldman B. J., Phys. Rev. Lett., 33, 359 (1974).
18. Proc. XIIIth Intern. Conf. on Phys. of Semiconductors, Tipographia Marves, 1976.
19. Алексеев А. С., Багаев В. С., Галкина Т. И., ЖЭТФ, 63, 1020 (1972).
20. Fröhlich H., Pelzer H., Zienau S., Phil. Mag., 41, 221 (1950).
21. Fröhlich H., в сб.: Polarons and Excitons, ed. Kuper C. G., Whitfield G. D., Oliver and Boyd, 1962, p. 1.
22. Harper P. G., Hudby J. W., Stradling R., Rept. Prog. Phys., 36, 1 (1973).
23. Polarons in Ionic Crystals and Polar Semiconductors, ed. Devreese J. T., North Holland, 1972.
24. Mills D. L., Burstein E., Rept. Prog. Phys., 37, 817 (1974).
25. Proc. VIIIth Intern. Conf. on Phys. of Semiconductors, Journ. Phys. Soc. Japan, Suppl. 21, 77 (1966).
26. Tait W. C. et al., в сб.: II—VI Semiconducting Compounds, ed. Thomas D. G., Benjamin, 1967, p. 370.
27. Henry C. H., Hopfield J. J., Phys. Rev. Lett., 15, 964 (1965).
28. Polaritons, ed. Burstein E., de Martini F., Pergamon Press, 1974.
29. Wallis R. F., в сб.: Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970.
30. Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969.
31. Platzman P. M., Wolff P. A., Solid State Physics, Academic Press, 1973, Suppl. 1.13.
32. Бонч-Бруевич В. Л., Вопросы электронной теории сильнолегированных полупроводников, в сб.: Физика твердого тела.— М.: Наука, 1965, с. 129—236.
33. Фистуль В. И., в сб.: Semiconductors and Semimetals, Academic Press, 1, 101 (1966); Сильнолегированные полупроводники.— М.: Наука, 1967.
34. Pincherle L., Proc. Phys. Soc. A, 64, 663 (1951).
35. Гласко В. Б., Миронов А. Г., ФТТ, 4, 336 (1962).
36. Бонч-Бруевич В. Л., ФТТ, 3, 768 (1961).
37. Pankove J. I., Progress in Semiconductors, Heywood, 9, 47 (1965).
38. Burstein E., Phys. Rev., 93, 632 (1954).
39. Moss T. S., Proc. Phys. Soc. B, 67, 775 (1954).
40. Pankove J. I., Aigrain P., Phys. Rev., 126, 956 (1962).
41. Esaki L., Phys. Rev., 109, 603 (1958).
42. Smith R. A., Wave Mechanics of Crystalline Solids, 2nd ed., Chapman and Hall, 1969, §§ 2.10, 6.5.
43. Келдыш Л. В., ЖЭТФ, 33, 994 (1957).
44. Takeuti T., Funada H., Journ. Phys. Soc. Japan, 20, 1854 (1965).
45. Tunneling Phenomena in Solids, ed. Burstein E., Lundquist S., Plenum Press,

1969. [Имеется перевод: Туннельные явления в твердых телах, под ред. Бурштейна Э., Лундквиста С.— М.: Мир, 1973.]
46. *Esaki L., Miyahara Y.*, Solid State Electron., 1 13 (1960).
 47. *Holonyak N. et al.*, Phys. Rev. Lett., 3, 167 (1959).
 48. *Келдыш Л. В.*, ЖЭТФ, 34, 1138 (1958).
 49. *Kane E. O.*, Journ. Appl. Phys., 32, 83 (1961).
 50. *Kleinman L.*, Phys. Rev., 140, A637 (1965).
 51. *Duke C. B.*, Solid State Physics, Academic Press, 1969, Suppl. 10.
 52. *Wolf E. L.*, Solid State Physics, Academic Press, 30, 2 (1975).
 53. Solids under Pressure, ed. Paul W., Warschauer, McGraw-Hill, 1963.
 54. *Keyes R. W.*, Phys. Rev., 99, 490 (1955).
 55. *Keyes R. W.*, Solid State Physics, Academic Press, 11, 149 (1960).
 56. *Paul W., Brooks H.*, Progress in Semiconductors, Heywood, 7, 135 (1963).
 57. *Drickamer H. G.*, Solid State Physics, Academic Press, 17, 1 (1965).
 58. *Slykhouse T. E., Drickamer H. G.*, Journ. Phys. Chem. Solids, 7, 210 (1958).
 59. *Edwards A. L., Drickamer H. G.*, Phys. Rev., 122, 1149 (1961).
 60. *Minomura S., Drickamer H. G.*, Journ. Phys. Chem. Solids, 23, 451 (1963).
 61. *Jamieson J. C.*, Science, 139, 762 (1963).
 62. *Samara G. A., Drickamer H. G.*, Journ. Phys. Chem. Solids, 23, 457 (1963).
 63. *Nathan M. I., Paul W., Brooks H.*, Phys. Rev. 124, 391 (1961).
 64. *Einstein A.*, Phys. Zs., 18, 121 (1917).
 65. *Loudon R.*, The Quantum Theory of Light, Clarendon Press, 1973. [Имеется перевод: Луддон Р. Квантовая теория света.— М.: Мир, 1976.]
 66. *Sargent M., Scully M. O., Lamb W. E.*, Laser Physics, Addison-Wesley, 1974.
 67. *Hall R. N. et al.*, Phys. Rev. Lett., 9, 366 (1962).
 68. *Nathan M. I. et al.*, Appl. Phys. Lett., 1, 62 (1962).
 69. *Quist T. M. et al.*, Appl. Phys. Lett., 1, 91 (1962).
 70. *Hurwitz C. E., Calawa A. R., Rediker R. H.*, Journ. Quantum Electron. 1, 102 (1965).
 71. Very High Resolution Spectroscopy, ed. Smith R. A., Academic Press, 1976.
 72. Laser Applications to Optics and Spectroscopy, ed. Jacobs S. F. et al., Addison-Wesley, vol. 2, 1975.
 73. *Queisser H. J.*, в сб.: Electronic Materials, ed. Hannay N. B., Colombo U., Plenum Press, 1973, p. 41.
 74. *Pilkuhn M. H.*, Phys. Stat. Solidi, 25, 9 (1968).
 75. *Wolff P. A.*, Phys. Rev. Lett., 16, 225 (1966).
 76. *Slusher R. E., Patel C. K. N., Fleury P. R.*, Phys. Rev. Lett., 18, 77 (1967).
 77. *Yafet Y.*, Phys. Rev., 152, 858 (1966).
 78. *Dennis R. B., Pidgeon C. R., Wherret B. S., Wood R. A.*, Proc. Roy. Soc., A331, 203 (1972).
 79. *Patel C. K. N., Shaw E. D.*, Phys. Rev. Lett., 24, 251 (1970).
 80. *Allwood R. L. et al.*, Journ. Phys. C: Solid State Phys., 3, L186 (1970).
 81. *Mooradian A., Brueck S. R. J., Blum F. A.*, Appl. Phys. Lett., 17 481 (1970).
 82. *Mooradian A.*, Science, 169, 20 (1964).
 83. *Loudon R.*, Adv. Phys., 13, 423 (1964).
 - 84*. *Пекар С. И.*, Исследования по электронной теории ионных кристаллов.— М.: Гостехиздат, 1951.

АМОРФНЫЕ ПОЛУПРОВОДНИКИ

15.1. Новые теоретические представления

Долгое время считалось, что теоретические представления, на которых основано обсуждение кристаллических полупроводников, к аморфным полупроводникам неприменимы. Например, в 1960 г. Иоффе и Регель [1] особое внимание уделяли тому факту, что зонная теория, основанная на понятии блоховских волн, в случае подвижностей менее $10 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ вообще неприменима, а при подвижностях менее $100 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ дает плохие результаты (см. разд. 13.14). Все это, несомненно, относится к исследованным до сих пор аморфным полупроводникам. Мы уже обращали внимание на тот факт, что даже в кристаллических полупроводниках с низкой подвижностью носителей заряда проводимость может иметь иной характер и ее обычно называют «прыжковой» или «диффузионной» проводимостью. Выше мы уже рассмотрели основные идеи, используемые при описании этого типа проводимости. Для общего описания явлений переноса в аморфных полупроводниках некоторые представления еще можно использовать, однако при этом возникают и совершенно новые проблемы.

Некоторые из возражений, выдвинутых Иоффе и Регелем против теоретических представлений, на основе которых с таким успехом было дано подробное описание физических процессов, определяющих свойства полупроводников с высокой подвижностью, ныне едва ли можно считать справедливыми. Видимо, трудно поддержать главное возражение Иоффе и Регеля, что представления об энергетических зонах применимы не ко *всем* полупроводниковым материалам. Из экспериментальных и теоретических работ, выполненных за последние шестнадцать лет, стало ясно, что кристаллические полупроводники с подвижностями носителей заряда выше, скажем, $100 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ являются совершенно другими материалами по сравнению с полупроводниками с низкой подвижностью и с теми же материалами, но в аморфном состоянии. В последних значения подвижности *намного меньше*, чем в кристаллическом состоянии. На-

пример, подвижность электронов в аморфном Ge при комнатной температуре составляет $0,15 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ по сравнению с величиной $4000 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ в хорошем кристаллическом Ge. Наряду с объяснением низких значений подвижностей, найденных в стеклообразных материалах, таких, как As_2S_3 , новая теория должна объяснить, почему подвижность материалов типа Ge при переходе их в аморфное состояние падает более, чем на четыре порядка величины. Теория также должна объяснить, почему некоторые полупроводниковые материалы (речь не идет о Ge) сохраняют некоторые из своих полупроводниковых свойств при плавлении ¹⁾.

Если представить себе весь диапазон применимости зонной теории, основанной на понятии блоховских волн, то вряд ли стоит критиковать эту теорию за то, что она не может объяснить перечисленных выше фактов. В течение последних шестнадцати лет теория продолжала давать все более точное описание различных и тонких явлений, происходящих в полупроводниках с высокой подвижностью и предсказывать новые явления, которые вслед за тем получали экспериментальное подтверждение. Кроме того, зонная теория создала прочную основу для расчетов электронных и оптоэлектронных приборов. Чего же еще можно желать от теории? Тот факт, что без существенных переделок теория смогла объяснить основные свойства металлов, возможно, мог навести на мысль, что она должна быть применима вообще ко всем материалам. Не исключено, что когда-нибудь такая общая теория и появится, но на сегодняшний день мы ею не располагаем.

В этих условиях теоретики попытались разработать иной подход и ввести новые представления, которые приблизили бы нас к пониманию основных свойств аморфных полупроводников, в том числе жидких и стеклообразных. На этом пути уже многое сделано, однако следует признать, что в теории имеется много неопределенностей и вопросов, которые еще не разрешены. Об аморфных полупроводниках появилось много литературы; к настоящему времени им посвящено почти столько же обзоров и монографий, сколько кристаллическим материалам. По этой причине мы сочли уместным ограничиться кратким рассмотрением вопроса. Иначе нам пришлось бы значительно расширить настоящую книгу, что невозможно и нежелательно. Мы постараемся выявить основные различия между кристаллическими и некристаллическими полупроводниками и особо остановиться на свойствах, наличие которых оправдывает причисление последних к классу материалов, именуемых полупроводниками.

Как мы уже видели в разд. 1.3, факт уменьшения электропроводности при повышении температуры не всегда может быть истол-

¹⁾ Подробное объяснение свойств жидких полупроводников можно найти в статье Эндерби ([2], стр. 361).

кован как однозначное доказательство того, что материал *не является* полупроводником. Большинство аморфных полупроводников имеет, однако, отрицательный температурный коэффициент сопротивления. Более отчетливым признаком является наличие запрещенной энергетической зоны, которая обычно проявляется в экспериментах по измерению оптического поглощения. Существование запрещенной энергетической зоны для аморфных материалов также остается фундаментальным свойством, однако, как мы увидим ниже, понятие о типе этой запрещенной зоны следует пересмотреть. В разд. 3.2 в качестве иного подхода к физическому описанию полупроводника мы рассмотрели представления, основанные на понятии химической связи. В разд. 3.3 мы видели, что с их помощью можно дать лучшее определение того, что мы называем полупроводником.

Представление о средней длине свободного пробега нельзя перенести на аморфные материалы, поскольку эта длина зависит от дальнего порядка в расположении атомов. В то же время ясно, что к аморфным материалам можно с успехом применить идеи, основанные на понятии химической связи. Картина химических связей большей частью определяется взаимодействием между ближайшими соседями, взаимодействие с более далекими соседями учитывается только как возмущение. Остается справедливым положение о том, что полупроводник — это материал с «насыщенными», т. е. парными электронными связями. Для того чтобы создать положительную дырку, т. е. недостаток одного электрона возле некоторого атома, и избыточный электрон, т. е. лишний электрон возле другого атома, в полупроводнике необходима энергия немногим более 1 эВ. Энергия связи проявится в виде увеличения поглощения в инфракрасном диапазоне спектра. В аморфном материале картина окружения атома случайным образом меняется от атома к атому, поэтому, как и следовало ожидать, край полосы фундаментального поглощения будет выражен не так резко, как в монокристалле.

15.2. Электронные состояния

Нет нужды далеко искать причину, благодаря которой в аморфном материале блоховский подход к описанию электронных состояний требует существенных переделок. Рассеяние на беспорядочно расположенных атомах делает бессмысленной непрерывную фазу блоховской волны. Тем не менее Андерсон [3], Мотт [4], Коэн [5] и их сотрудники теоретически показали, что могут существовать так называемые «расширенные» состояния, т. е. состояния, волновая функция которых простирается на область макроскопических размеров. Кроме них существуют также связанные состояния, волновая функция которых локализована в области с размерами, не-

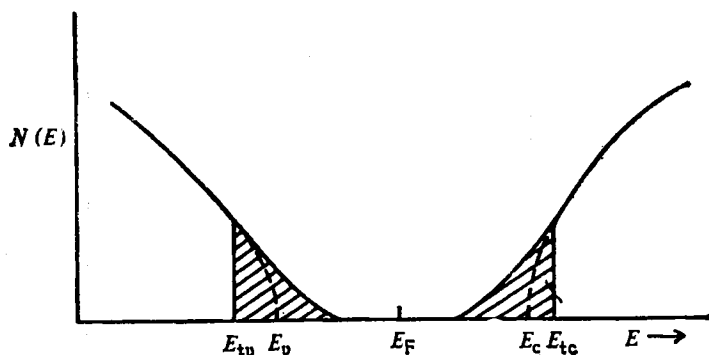


Рис. 15.1. Хвосты зон, состоящие из локализованных состояний (заштрихованные области).

намного превышающими межатомное расстояние. Более того, при данной энергии эти два типа состояний сосуществовать не могут. Поэтому между ними имеется четкое различие. Было показано, что расширенные состояния могут образовывать энергетические зоны, очень похожие на энергетические зоны в кристалле, и что уровни энергии, соответствующие локализованным состояниям, расположены вблизи краев зон, расширяя таким образом зоны в область энергий, в которой в кристаллическом материале разрешенных уровней энергии нет. Это проиллюстрировано на рис. 15.1. Энергии, обозначенные через E_v и E_c , соответствуют краям зон в кристаллическом материале. Энергии, обозначенные через E_{tv} и E_{tc} , соответствуют разделению между расширенными и локализованными состояниями. Вблизи этих значений энергии в кристаллическом материале располагались бы вершина валентной зоны и дно зоны проводимости соответственно. Области, занятые локализованными состояниями, заштрихованы.

Было показано, что в случае полупроводников IV группы в отсутствие разорванных связей (обычно их называют «болтающимися» связями) в аморфном состоянии будет существовать запрещенная энергетическая зона [6—8]. Однако в реальном материале это идеальное условие, по-видимому, не выполняется и хвосты зон могут перекрываться таким образом, что во всем промежутке между E_c и E_v существуют уровни энергии локализованных состояний, как показано на рис. 15.2, а (см. также рис. 15.3). Мотт, Дэвис и Стрит [9] обсуждали природу этих состояний и, в частности, их отношение к болтающимся связям. С этой точки зрения энергетическая зона исчезает. Однако электропроводность по локализованным состояниям запрещена, и подвижность носителей заряда, находящихся в этих состояниях, равна нулю. Поэтому вместо энергетической запрещенной зоны мы можем говорить о запрещенной зоне

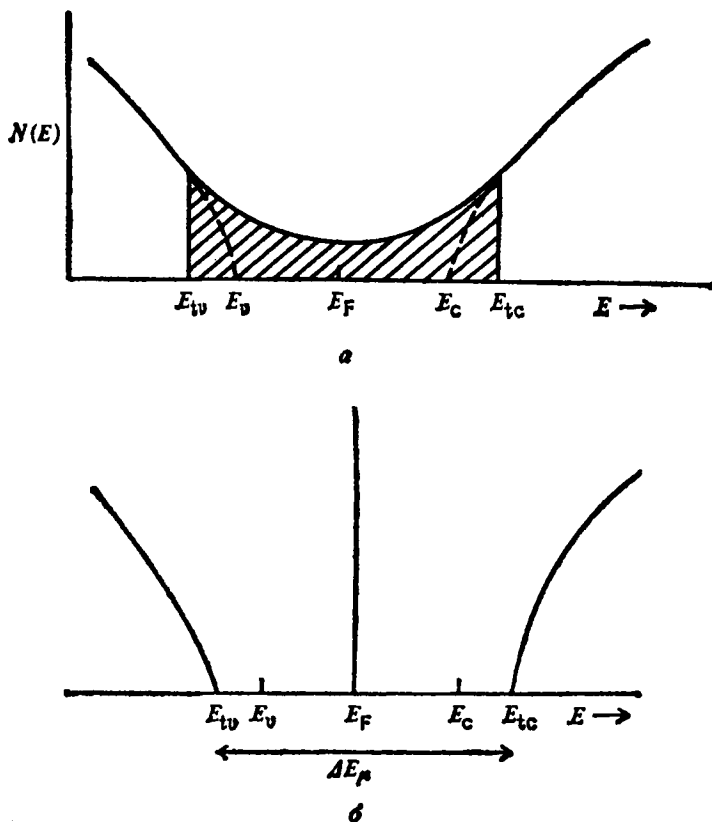


Рис. 15.2. Схема изменения зон. *a* — перекрытие хвостов локализованных состояний, приводящее к исчезновению запрещенной зоны; *б* — запрещенная зона подвижностей ΔE_μ .

подвижностей ΔE_μ , что проиллюстрировано на рис. 15.2, *б*. Ниже мы увидим, что величина ΔE_μ несколько превышает ΔE — ширину запрещенной энергетической зоны в кристаллическом материале. Этим проблемам посвящено много теоретических работ. Мотт и Дэвис [10] сделали обзор этих работ, и дали обстоятельное изложение проблемы, а также обсудили интерпретацию большого количества экспериментальных данных. Экспериментальные и теоретические работы были недавно рассмотрены в серии статей различных авторов [2, 11, 12], а также в нескольких монографиях [13].

С помощью запрещенной зоны подвижностей можно объяснить существование энергии активации, которая обычно проявляется в процессах проводимости. Обычно электропроводность σ как функ-

ция температуры следует эмпирическому закону $\sigma = A \exp(-W/kT)$, хотя вместо одной экспоненты здесь присутствует сумма нескольких экспонент, каждая со своей энергией активации. Для возникновения электропроводности электроны должны быть возбуждены с локализованных состояний вблизи вершины валентной зоны в расширенные состояния в зоне проводимости. Так как плотность состояний в запрещенной зоне подвижностей мала по сравнению с плотностью состояний в области крутого подъема выше уровня раздела локализованных и расширенных состояний, то отсюда становится понятным существование оптической запрещенной зоны. Однако край полосы поглощения не должен быть таким же резким, как в кристаллическом материале, а локализованные состояния должны дать «хвост» с длинноволновой стороны от края полосы поглощения. Как мы увидим ниже, именно такая ситуация реально имеет место.

Теперь мы должны выяснить, что следовало бы считать экспериментальным доказательством существования локализованных состояний в запрещенной зоне подвижностей. В какой-то степени его можно получить из измерения температурной зависимости электропроводности и коэффициента Холла, однако более прямым доказательством было бы измерение фотоэлектрической эмиссии под действием света, а также использование эффекта поля (см. разд. 7.11.1). Подобные измерения для *a*-Si мы рассмотрим в разд. 15.4, а сейчас только отметим, что существование таких состояний можно считать хорошо доказанным и что можно даже определить плотность этих состояний.

Перед тем как обсуждать результаты измерений на конкретных материалах, рассмотрим, какими способами определяют основные параметры. Электропроводность, термо-э. д. с. и оптическое поглощение получаются в результате прямых измерений. Коэффициент Холла измерять сложно обычно из-за высокого удельного сопротивления образцов. Более того, здесь имеются серьезные теоретические трудности, заключающиеся в том, что существует много теоретических доказательств того, что коэффициент Холла R_H несколько меньше, чем $1/ne$ или $1/pe$ в примесном материале *n*- или *p*-типа. Можно, конечно, получить холловскую подвижность μ_H , найдя произведение σR_H , однако интерпретация этой величины и ее связь с дрейфовой подвижностью μ_d не столь очевидна, как в кристаллическом материале с высокой подвижностью. К счастью, дрейфовую подвижность можно получить независимо, с помощью методики, которую разработал Спир [12]. На поверхность тонкого образца оптически или электронной бомбардировкой в течение очень короткого промежутка времени инжектируют порцию электронов. При помощи быстродействующего осциллографа наблюдают приход этих электронов на противоположную сторону образца. Зная толщину образца, по времени задержки получают величину μ_d .

Обычно за приходом основного импульса следует «хвост», который говорит о том, что по мере продвижения через материал носители испытывают захват ловушками.

15.3. Колебания атомов решетки

Несмотря на то что в аморфном материале зонная структура для решеточных колебаний отсутствует, существует, однако, некоторая аналогия с кристаллическими модами. Для длинноволновых акустических колебаний нерегулярности в расположении атомов должны усредниться и зависимость ω/k будет определяться упругими постоянными. Изменение последних будет проявляться просто как изменение акустических колебаний, и поэтому можно по-прежнему говорить о длинноволновых акустических фононах. А длинноволновые оптические фононы опять же в основном определяются ближайшим атомным окружением. Действительно, в своей оригинальной работе Раман рассматривал только моды колебаний локального молекулярного окружения. Для сложных веществ, таких, как As_2S_3 , примитивная ячейка которых представляет собой молекулярную группу, это является хорошим приближением и, как и следовало ожидать, спектры комбинационного рассеяния и спектры решеточного поглощения, обусловленного оптическими модами, для аморфной и кристаллической форм одного и того же материала отличаются мало. С другой стороны, в материалах, подобных Si и Ge, тетраэдрическая группа атомов связана с соседними намного сильнее. В этом случае двухфононные спектры аналогичны спектрам кристаллического материала, однако в них отсутствует тонкая структура колебательного спектра, характерная для кристалла. Так как в аморфном материале не действует закон сохранения квазиимпульса, то в нем может происходить однофононное поглощение.

15.4. Аморфные Si и Ge

Из-за того что аморфные полупроводники IV группы Si и Ge обладают открытой структурой, их свойства сильно отличаются от свойств стеклообразных аморфных полупроводников, таких, как As_2S_3 , которым было посвящено множество экспериментальных работ. Более того, Si и Ge могут быть получены в очень чистом виде, а их свойства в кристаллическом состоянии хорошо изучены и поняты. Широко изучались свойства этих материалов и в аморфном состоянии, что дало прекрасную возможность провести сравнение свойств двух форм одного и того же вещества. Поэтому перед тем, как обсуждать менее простые для понимания аморфные полупроводниковые соединения, мы для начала рассмотрим Si и Ge.

Аморфная форма Si и Ge образована группами, состоящими из четырех атомов, объединенных в тетраэдры. Основное отличие аморфной формы от кристаллической заключается в том, что в первой из них тетраэдры ориентированы относительно друг друга случайным образом. Тот факт, что взаимное расположение ближайших соседей в аморфном состоянии остается неизменным, играет очень важную роль, обуславливая не слишком сильное различие в свойствах двух форм, как это поначалу предполагали Июффе и Регель (см. разд. 15.1). При плавлении это тетраэдрическое взаимное расположение исчезает. Возникает намного более плотная упаковка, при которой число ближайших соседей равно восьми, и Si и Ge становятся металлами, так же как при очень высоких давлениях (см. разд. 14.5).

Существуют две точки зрения на проблему случайного взаимного расположения атомов. В одной рассматривают все возможные взаимные ориентации, в другой рассматриваются только две ориентации, характерные для структуры алмаза и для структуры вюрцита. Эти две структуры рассмотрел Григоровичи [12], стр. 45) и показал, что как шестиугольные кольца атомов, характерные для кристаллической структуры, так и пятиугольные кольца атомов, существующие в аморфной структуре, способны использовать все пары связывающих (валентных) электронов. При этом нет болтающихся связей, если, конечно, они не появляются в результате тепловых, оптических или каких-нибудь еще воздействий или не являются результатами структурных дефектов.

Для получения аморфных пленок толщиной порядка нескольких микрон используются различные методы, включая термическое испарение, катодное распыление, осаждение путем разложения газа электрическим разрядом. До последнего времени считалось, что одним из очень сильных различий между аморфными и кристаллическими Si и Ge является тот факт, что, как мы уже видели, кристаллическая форма очень чувствительна к введению весьма малых количеств примесей элементов III и V групп, в то время как аморфные Si и Ge нечувствительны к введению примесей, так что для заметного изменения их электропроводности требуется концентрация примесей порядка 10^{20} см⁻³. Считалось, что это обусловлено тем, что избыточный или недостающий электрон может быть соответственно захвачен или восполнен благодаря изменению связей в структурном дефекте или в других связях. Незаполненные локализованные состояния могут быть присоединены и к примесным состояниям, если уровень Ферми в нелегированном материале расположен вблизи середины запрещенной зоны подвижностей (см. рис. 15.2). В своей недавней работе Спир [24] показал, однако, что наличие многих состояний в запрещенной зоне зависит от метода приготовления образца и что можно приготовить тонкие пленки с меньшим числом состояний в запрещенной зоне, чем это было найдено раньше. В ре-

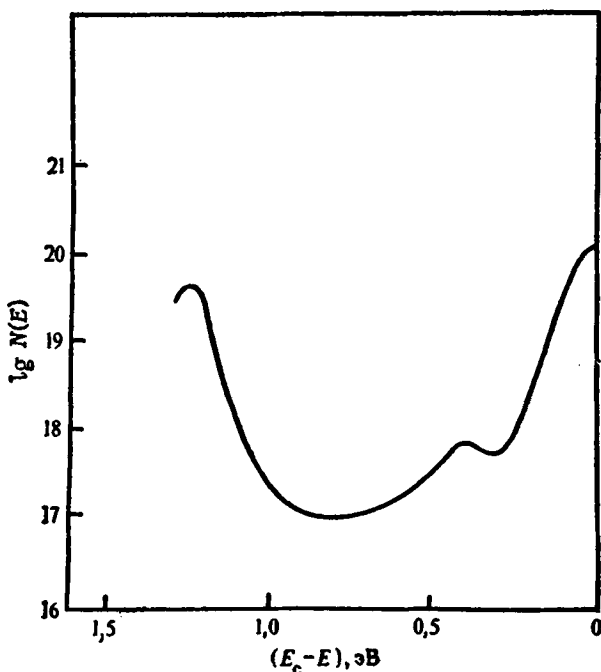


Рис. 15.3. Плотность состояний $N(E)$ $\text{см}^{-3} \cdot \text{эВ}^{-1}$ в запрещенной зоне подвижностей в a-Si [13].

зультате у него появилась возможность влиять на электропроводность, вводя в материал примеси как n -, так и p -типа в концентрациях, намного меньших, чем обычно. Более того, ему удалось осуществить p — n -переходы, что явилось огромным шагом вперед в технологии аморфных полупроводников, и открыть таким образом возможность использования их в солнечных батареях большой площади.

Пленки были получены методом осаждения в тлеющем разряде в атмосфере силана или германа, а для введения примесей использовалось несколько миллионных долей фосгена или диборана. Этими методами удалось повысить электропроводность образцов от величины порядка 10^{-12} $\text{Ом}^{-1} \cdot \text{см}^{-1}$, соответствующей собственной электропроводности, до примерно 10^{-2} $\text{Ом}^{-1} \cdot \text{см}^{-1}$ для материала как n -, так и p -типа. Подвижности носителей заряда в a-Si имеют величины порядка $0,1$ $\text{см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ и слабо зависят от легирования, однако уменьшаются при понижении температуры.

Спир и Комбер [14] в своих измерениях плотности состояний в запрещенной зоне подвижностей методом эффекта поля (см. разд. 7.11.1) показали, что плотности состояний лежат около $10^{17} \text{см}^{-3} \times$

$\times \text{эВ}^{-1}$, в то время как обычно считалось, что в аморфных пленках Si и Ge эта величина составляет около $10^{19} \text{ см}^{-3} \cdot \text{эВ}^{-1}$. Мадан, Комбер и Спир [15] приводят подробности метода вычисления $N(E)$ из измерений по эффекту поля. Результаты измерений на пленках, специально подготовленных для легирования, приведены на рис. 15.3. Зависимость плотности состояний N от энергии E имеет интересную структуру и оказывается более сложной, чем та, которая получена в результате наложения двух «хвостов» и показана на рис. 15.2.

15.4.1. ОПТИЧЕСКИЕ СВОЙСТВА a-Si И a-Ge

Как измерили Донован и др. [16], в широком спектральном диапазоне, например от 0,5 до 1,5 эВ оптическое поглощение a-Si и a-Ge не очень сильно отличается от оптического поглощения в кристаллических материалах, однако не обладает структурой, свойственной последним. Крутой рост поглощения вблизи края полосы фундаментального поглощения германия начинается при меньшей энергии квантов, примерно при 0,55 эВ, коэффициент поглощения быстро растет до величины около 10^9 см^{-1} при 0,6 эВ и затем продолжает расти более медленно до величины примерно 10^8 см^{-1} при 1,5 эВ. Спектры отражения также сходны, однако отсутствуют пики, связанные с критическими точками зонной структуры, подобные тем, что показаны на рис. 10.24. Этого и следовало ожидать, так как зонная структура, характерная для кристаллических тел, не пригодна для описания аморфных. Спайсер и Донован [17] провели измерения поглощения на a-Ge вплоть до 9 эВ и показали, что в области 5—9 эВ различие между a-Ge и кристаллическим Ge невелико, однако поглощение в a-Ge немного сильнее. При исследовании фундаментального поглощения a-Ge и a-Si с использованием аппаратуры с высоким спектральным разрешением не проявляется никакой тонкой структуры, обусловленной переходами с участием фононов или экситонами, вместо нее можно видеть постепенное увеличение поглощения, начинающееся при различных величинах $h\nu$, которые примерно на 0,1 эВ меньше, чем в монокристаллическом материале. Этот сдвиг и другие характеристики сильно меняются от образца к образцу. Это указывает на то, что материалы, приготовленные термическим испарением или катодным распылением, не представляют собой действительно собственный материал, а содержат множество дефектов в виде полостей или других нарушений идеально разупорядоченной решетки. Пленки, которые приготовили Спир и его сотрудники, видимо, намного ближе стоят к «собственно» аморфному материалу, чем пленки, приготовленные другими методами, и поэтому в них может проявиться форма края полосы поглощения, наиболее характерная для аморфного материала.

Во всех аморфных материалах длинноволновый хвост оптического поглощения имеет почти экспоненциальную форму. Поглоще-

ние тоже зависит от метода приготовления и почти наверняка обусловлено электронами, появляющимися из состояний в запрещенной зоне подвижностей (хотя были предложены и другие объяснения [2], стр. 159).

Форма этого хвоста приблизительно подчиняется правилу, которое называют правилом Урбаха [18]. Правило гласит, что коэффициент поглощения α выражается формулой $\alpha = A \exp\{A(h\nu - E_0)/kT\}$, где A , a и E_0 — константы.

15.5. Другие элементарные аморфные полупроводники

Проводились эксперименты и с различными другими аморфными элементарными полупроводниками, в том числе с Se, Te и В. Из них только Te удалось очистить примерно до состояния, достигнутого для Si и Ge. Так как кристаллический Te сильно анизотропен и в кристаллическом состоянии имеет ярко выраженную c -ось, то он с трудом образует аморфные слои. Но даже в этом случае удалось выполнить довольно подробное сравнение его аморфной и кристаллической форм (2), стр. 159). Сложность приготовления аморфного Se заключается не только в его очистке, но и в том, что этот элемент существует в двух кристаллических разновидностях, моноклинной и триклинной, а также в смешанном состоянии. Из-за того, что этот материал получил широкое распространение в фотокопировальных машинах, он привлекает к себе значительное внимание. Его оптическое поглощение неоднократно измерялось, однако оно, по-видимому, очень сильно зависит от метода приготовления. Вероятно, спектральная кривая оптического поглощения аморфного Se лежит приблизительно между кривыми поглощения для двух его форм ([11], стр. 245). Изучались также сплавы Se и Te.

Дрейфовые подвижности Se измерялись в нескольких лабораториях, при этом получены примерно одинаковые результаты. Значения подвижностей электронов и дырок при комнатной температуре, которые сообщают Долежалек и Спир [19], равны соответственно $6 \cdot 10^{-3} \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ и $0,14 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$. Они экспоненциально падают при уменьшении температуры, что указывает на наличие некоего активационного процесса. С другой стороны, холловская подвижность при комнатной температуре имеет приблизительно такую же величину, но зависит от температуры существенно слабее, что указывает на ограничение дрейфовой подвижности процессами прилипания на ловушках. Фритцше ([2], стр. 211) рассмотрел эту проблему, а также сравнил экспериментальные данные, полученные в разных лабораториях. Данные по чистому Te очень скудны. Сплавление Se с 1% Te уменьшает подвижность дырок при комнатной температуре примерно в десять раз. Оптические свойства аморфных

пленок Se и Te очень похожи на оптические свойства аморфных пленок Si и Ge, у них точно так же край полосы поглощения сдвинут в сторону более низких частот по отношению к краю полосы поглощения в кристаллическом материале. Ситуация сильно осложняется из-за существования двух кристаллических форм Se и из-за различного поведения Te при облучении его светом, поляризованным параллельно или перпендикулярно оси c . Этот вопрос довольно подробно рассмотрел Тауц ([2], стр. 159).

15.6. Некоторые аморфные полупроводниковые соединения

Наряду с аморфными пленками полупроводников IV группы были получены и изучались аморфные пленки некоторых соединений типа $A^{III}B^V$, включая InSb, GaAs и GaSb [20]. Так же как и в случае Si и Ge, проводимость этих пленок падала при уменьшении температуры намного быстрее, чем в кристаллическом материале, причем степень падения зависела от температуры отжига пленок после их осаждения. По-видимому, получить собственную аморфную пленку из этих материалов еще сложнее, чем из Si и Ge. О свойствах соединений типа $A^{III}B^VI$ в аморфном состоянии известно очень мало.

15.7. Стеклообразные полупроводники

Было изучено огромное число стеклообразных материалов, проявляющих полупроводниковые свойства (см. общие ссылки к разд. 15.2). Одними из наиболее простых представителей этой группы материалов являются халькогенидные стекла, такие, как As_2S_3 , As_2Se_3 и As_2Te_3 . Из-за их относительно простой структуры значительная часть фундаментальных исследований полупроводниковых стекол была проведена именно на них, в особенности на As_2S_3 . Позже к себе привлекли внимание намного более сложные стекла, такие, как $Te_{a_1}Ge_{b_1}A_4$, где A — это обычно элемент V группы периодической системы элементов. Интерес к этим материалам был вызван их применениями в переключателях и запоминающих устройствах (см. разд. 15.7.2). Более простые стекла, однако, дают нам лучшую возможность разобраться в сложных физических проблемах, которые возникают при изучении этого класса материалов, поэтому мы ограничимся в своем обсуждении только этими материалами.

Опираясь на выводы из разд. 15.2, можно заключить, что эти материалы отличаются от полупроводников IV группы намного меньшим количеством состояний в запрещенной зоне подвижностей. Действительно, вполне может оказаться, что локализованные со-

стояния располагаются в небольших хвостах вблизи краев зон расширенных состояний, и картина скорее напоминает рис. 15.1, а не рис. 15.2. Если это на самом деле так, то в таком материале будет существовать настоящая запрещенная энергетическая зона, которая будет лишь ненамного уже запрещенной зоны подвижностей, которую по-прежнему можно определить как зону между двумя энергетическими уровнями, разделяющими локализованные и расширенные состояния.

Конечно, в запрещенной зоне будет существовать некоторое число состояний, но, как предполагают, они обусловлены «болтающимися» связями, возникшими в результате наличия микрополостей и других нарушений идеального беспорядка, либо примесями. Так же как и в случае элементарных полупроводников, важно быть уверенным, что исследуемый материал не является поликристаллическим. Это не всегда просто выяснить, так как многие из этих стекол имеют склонность к кристаллизации.

Электропроводность полупроводников этого типа можно описать экспоненциальным законом $\sigma = \sigma_0 \exp(-E_g/2kT)$ при изменении σ в пределах 4—6 порядков величины. Если отвлечься от отклонений от этого закона, которые наблюдаются при низких температурах, становится ясно, что для возникновения проводимости требуется термическая активация. Величина E_g приблизительно равна ширине зоны, полученной с помощью оптических измерений, однако в точности с ней не совпадает. Эти результаты укладываются в предложенную модель и показывают, что уровень Ферми в собственном материале расположен вблизи середины запрещенной зоны (из-за того, что в знаменателе показателя экспоненты стоит величина $2kT$).

Значения электропроводности при комнатной температуре сильно отличаются друг от друга, но обычно имеют величину порядка $10^{-5} \text{ Ом}^{-1} \cdot \text{см}^{-1}$ и уменьшаются примерно до $10^{-8} \text{ Ом}^{-1} \cdot \text{см}^{-1}$ при 150 К. Значения подвижностей (обычно это μ_n) при комнатной температуре также довольно различны, но лежат в пределах $0,1—0,01 \text{ см}^2 \times \text{В}^{-1} \cdot \text{с}^{-1}$ и не так быстро уменьшаются при понижении температуры (менее чем в десять раз между 300 и 150 К). Очевидно, эти материалы не очень чувствительны к введению химических примесей. Электропроводность и подвижность носителей заряда в сплавах обычно оказываются ниже, чем в исходных материалах. Электрические свойства стеклообразных полупроводников рассмотрел Фритцше ([2], стр. 211).

15.7.1. ОПТИЧЕСКИЕ СВОЙСТВА

Широко изучались также оптические свойства аморфных материалов ([2], стр. 159). Свойства стеклообразных полупроводников более воспроизводимы и меньше изменяются от образца к образцу, чем это имеет место для элементарных аморфных полупро-

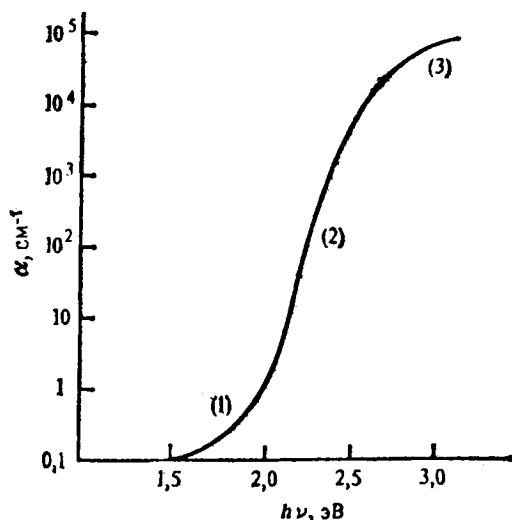


Рис. 15.4. Коэффициент поглощения α в а- As_2S_3 [21].

водников. Почти наверняка это обусловлено трудностями приготовления бездефектных пленок аморфных элементарных полупроводников. На рис. 15.4 представлена спектральная кривая поглощения в As_2S_3 , которую получили Косек и Тауц [21]. Подобная форма спектральной кривой поглощения характерна для многих стеклообразных полупроводников. Спектр поглощения можно условно разделить на три части: 1) длинноволновый хвост, форма которого иногда описывается правилом Урбаха (см. разд. 15.4.1); 2) экспоненциальная часть, которая перекрывает интервал изменения величины коэффициента поглощения в пределах четырех-пяти порядков, и для которой наклон почти не зависит от температуры (за исключением самых высоких температур); 3) область сильного поглощения, которую можно описать степенным законом $\alpha = (h\nu - E_g^0)^r / h\nu$. Из него может быть получена «оптическая» запрещенная зона E_g^0 . Параметр r обычно равен 2, хотя из опыта известны и другие значения, в том числе 3. Если построить зависимость $(h\nu\alpha)^{1/2}$ от $h\nu$ для области сильного поглощения в As_2S_3 , как это сделано на рис. 15.5, то мы получим прямую линию, для которой $r=2$. Это аналогично изменению, наблюдаемому в случае непрямого перехода в кристаллическом полупроводнике.

Интерпретация этих трех участков кривой поглощения с точки зрения теоретических представлений, описанных в разд. 15.2, рассмотрел Тауц ([2], стр. 159). Считается, что поглощение на участке (3) обусловлено переходами из расширенных состояний возле вершины валентной зоны на расширенные состояния в зоне проводимо-

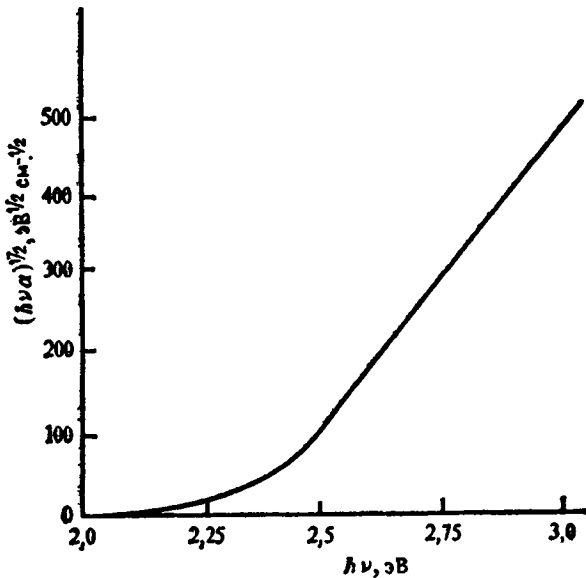


Рис. 15.5. Область сильного поглощения (см. рис. 15.4), представленная в виде зависимости $(hv\alpha)^{1/2}$ от hv .

сти. Так как закон сохранения квазиимпульса $\hbar k$ в нашем случае теряет силу, то интегрирование по плотности состояний следует производить так же, как в случае не прямых переходов в кристалле (см. разд. 10.5.3). Так как плотности состояний $N_c(E)$, $N_v(E)$ пропорциональны $(E - E_c)^{1/2}$ и $(E_v - E)^{1/2}$, то для коэффициента поглощения получается квадратичная зависимость от частоты, так же как и для непрямого перехода. Отклонение величины r от 2 указывало бы либо на иную форму зависимости $N(E)$, либо на наличие оптически запрещенного перехода. Экстраполируя линейный участок кривой на рис. 15.4, можно получить ширину «оптически» запрещенной зоны E_g^o , которая соответствует разности величин E_c и E_v , показанных на рис. 15.1.

Предполагается, что участок (2) обусловлен переходами из расширенных состояний вблизи вершины валентной зоны на локализованные состояния вблизи дна зоны проводимости и точно так же переходами из локализованных состояний вблизи вершины валентной зоны на расширенные состояния в зоне проводимости. Если плотность состояний в хвосте описывается формулой $N(E) = B \exp(-E/E_t)$, где B и E_t — константы, то мы получаем соотношение $\alpha = C \exp(-hv/E_t)$ при условии, что E_t одна и та же для обоих хвостов, иначе α будет выражаться через сумму двух экспонент. Тауц [12], стр. 159) проанализировал спектры поглощения во множе-

стве стеклообразных материалов и нашел, что этот закон справедлив в широком интервале значений коэффициента поглощения. Для количественного объяснения наблюдаемой величины коэффициента поглощения требуется, чтобы плотность состояний была порядка $10^{20} \text{ см}^{-3} \cdot \text{эВ}^{-1}$.

Интерпретация длинноволнового хвоста [участок (1)] намного сложнее. Делались попытки приписать это поглощение переходам между состояниями, расположенными в хвостах зоны проводимости и валентной зоны, однако, как показал Тауц ([2], стр. 159), эти попытки наталкиваются на некоторые серьезные трудности. Более того, было обнаружено, что вид кривой на этом участке в отличие от двух других участков очень сильно зависит от качества приготовления материала. Поэтому почти наверняка он определяется примесями или дефектами.

В разд. 15.3 мы уже кратко останавливались на спектрах поглощения этих материалов, обусловленных колебаниями атомов решетки. В основном они дают информацию о колебательных уровнях молекулярных групп, из которых состоит стекло. Всего несколько характерных частот можно приписать колебаниям определенных химических связей. Например, спектры поглощения в области *остаточных лучей*, хотя и не такие узкие, как в кристаллах, однако находятся на том же месте [22].

15.7.2. ТОКОВЫЕ НЕУСТОЙЧИВОСТИ И ЭФФЕКТ ПЕРЕКЛЮЧЕНИЯ

В стеклообразных полупроводниках описанного выше типа было обнаружено новое явление, представляющее собой особый случай токовой неустойчивости. Оно наблюдалось в различных материалах, в том числе в халькогенидных стеклах, однако более заметно выражено в более сложных стеклах, электропроводность которых при комнатных температурах лежит в интервале 10^{-8} — $10^{-6} \text{ Ом}^{-1} \cdot \text{см}^{-1}$. Если электрическое поле в образце (толщиной около 1 мкм) превысит примерно $10^6 \text{ В} \cdot \text{м}^{-1}$, то электропроводность образца резко возрастет до величины порядка 1 — $10 \text{ Ом}^{-1} \cdot \text{см}^{-1}$. Это явление нельзя рассматривать как электрический пробой в обычном понимании, так как при уменьшении напряжения ток будет линейно уменьшаться, пока не достигнет некоторой критической величины, после чего восстановится нормальная вольтамперная характеристика образца (рис. 15.6). Более того, эффект имеет место и при обратной полярности приложенного напряжения. Именно из-за наличия этого эффекта переключения тока необычайно возрос интерес к стеклообразным полупроводникам. Это явление было открыто Овшинским [23] и дало мощный толчок разработке приборов, поскольку на его основе можно создать переключатели, запоминаю-

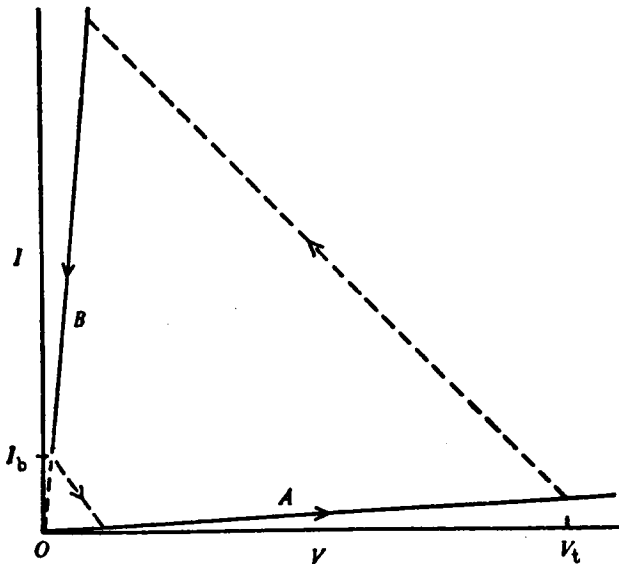


Рис. 15.6. Иллюстрация эффекта переключения вольт-амперной характеристики I/U с высокого сопротивления (ветвь A) на низкое сопротивление (ветвь B) при критическом напряжении V_k и обратного переключения при критическом токе I_b .

шие устройства и т. д., которые имеют очень малые времена переключения порядка 10^{-10} с. Хотя явление было обнаружено во множестве материалов, для его применений используются стекла (большая часть сплавы халькогенидов). В их число входят такие стекла, как $Te_{50}As_{30}Si_{10}Ge_{10}$, и другие не менее сложные соединения. Фритцше ([11], стр. 425; [2], стр. 211) сделал обзор применений этого эффекта, а также обсудил его физическую природу. Для объяснения эффекта переключения тока были сделаны различные предположения, однако до сих пор нет полного согласия относительно деталей механизма этого эффекта. Возможно, что за него ответственны сразу несколько различных процессов, таких, как локальный нагрев, вызывающий плавление материала или изменение его кристаллической структуры. Природу этого необычного явления, применение которого опередило развитие теории, еще предстоит выяснить.

15.8. Заключение

Из нашего краткого рассмотрения аморфных полупроводников должно быть ясно, что хотя уже многое достигнуто в понимании основных физических процессов, которые придают этим материалам

особенные свойства, остается немало серьезных теоретических и практических задач, которые предстоит разрешить, прежде чем мы поймем свойства аморфных полупроводников так же, как кристаллических, обладающих высокой подвижностью носителей заряда. Теории предстоит еще ответить на важные вопросы и разрешить некоторые проблемы, связанные с состояниями электронов в разупорядоченных системах. С практической точки зрения существуют трудности очистки и приготовления образцов, которые еще предстоит преодолеть, прежде чем можно быть уверенным, что изучаемый нами материал действительно отражает свойства данного вещества. Прогресс, достигнутый с $a\text{-Si}$ и $a\text{-Ge}$, показывает, что каждый шаг вперед в технологии приводит к новому развитию как фундаментальной науки, так и прикладной (см. разд. 15.4). Эти проблемы представляют сложную задачу, но вместе с тем открывают новое поле деятельности для выдвижения новых идей и дают надежду на появление интересных и необычных результатов.

Маловероятно, что аморфные материалы в ближайшем будущем заменят монокристаллы в обычных применениях в микроэлектронике и оптоэлектронике, однако, как мы видели, аморфные материалы нашли себе новые применения. Практический и физический интерес к этой проблеме не оставляет сомнений в том, что быстрый прогресс в понимании свойств этих материалов будет продолжаться.

ЛИТЕРАТУРА

1. Progress in Semiconductors, Heywood, 4, 237 (1960).
2. Amorphous and Liquid Semiconductors, ed. J. Tauc, Plenum Press, 1974.
3. Anderson P. W., Phys. Rev., 109, 1492 (1958).
4. Mott N. F., Adv. Phys., 16, 49 (1967).
5. Cohen M. H., Fritzsche H., Ovshinsky S. R., Phys. Rev. Lett., 22, 1065 (1969).
6. Weaire D., Journ. Non-Cryst. Solids, 6, 181 (1971).
7. Thorpe M. F., Weaire D., Phys. Rev., B4, 3518 (1971).
8. Heine V., Journ. Phys. C: Sol. State Phys., 4, 1221 (1971).
9. Mott N. F., Davis A. E., Street R., Phil. Mag., 32, 961 (1975).
10. Mott N. F., Davis E. A., Electronic Processes in Non-crystalline Materials, Clarendon Press, Oxford, 1971. [Имеется перевод: Мотт Н., Дэвис Э. Электронные процессы в некристаллических веществах. — М.: Мир, 1974.]
11. Electronic and Structural Properties of Amorphous Semiconductors, ed. P. G. Le Comber, J. Mort, Academic Press, 1973.
12. Spear W. E., Adv. Phys., 23, 523 (1974).
13. Губанов А. И. Квантово-электронная теория аморфных полупроводников. — М.—Л.: изд-во АН СССР, 1963.
14. Spear W. E., Le Comber P. G., Phil. Mag., 33, 935 (1976).
15. Madan A., Le Comber P. G., Spear W. E., Journ. Non-Cryst. Solids, 20, 259 (1976).

16. *Donovan T. M. et al.*, Phys. Rev., **B2**, 397 (1970).
17. *Spicer W. E., Donovan T. M.*, Journ. Non-Cryst. Solids, **2**, 66 (1970).
18. *Urbach F.*, Phys. Rev., **92**, 1324 (1953).
19. *Dolezalek F. K., Spear W. E.*, Journ. Non-Cryst. Solids, **4**, 97 (1970).
20. *Stuke J.*, Proc. Int. Conf. on Low-Mobility Matls, Eilat, 1971, p. 193.
21. *Kosek F., Tauc J.*, Czech. Journ. Phys., **B20**, 94 (1970).
22. *Mitchell D. L., Bishop S. G., Taylor P. C.*, Journ Non-Cryst. Solids, **10**, 231 (1972).
23. *Ovshinsky S. R.*, Phys. Rev. Lett., **21**, 1450 (1968).
24. Proc. XIIIth Intern. Conf. on Phys. of Semiconductors, Tipographia Marves, 1976, p. 515.

ЛИТЕРАТУРА

Ниже приведен краткий и очень неполный список изданий, посвященных физике полупроводников. При составлении этого перечня мы постарались перекрыть все вопросы, затронутые в книге. На большинство из этих изданий мы уже ссылались в тексте.

1. Серии и периодические издания

Progress in Semiconductors, eds. Gibson A. F., Aigrain P., Burgess R. E. и позже, Kröger F. A., Heywood, 1956—67; Semiconductors and Semimetals, eds. Willardson R. K., Beer A. C., Academic Press, 1966; Solid State Physics, eds. Seitz F., Turnbull D. и позже, Ehrenreich H., Academic Press, 1955; см. также труды международных конференций по физике полупроводников, которые проводятся каждые два года, начиная с 1952 г.

2. Теоретические основы физики твердого тела

Ziman J. M., Electrons and Phonons, Oxford University Press, 1960. (Имеется перевод: Займан Дж. Электроны и фононы.— М.: ИЛ, 1962); Ziman J. M., Principles of the Theory of Solids, Cambridge University Press, 2nd edition, 1972. (Имеется перевод: Займан Дж. Принципы теории твердого тела.— М.: Мир, 1974.) Jones W., March N. H., Theoretical Solid State Physics, Wiley Interscience, 1972; Haug A., Theoretical Solid State Physics, Pergamon Press, 1972; Slater J. C., Insulators, Semiconductors and Metals, McGraw-Hill, 1967 (Имеется перевод: Слатер Дж. Диэлектрики, полупроводники, металлы.— М.: Мир, 1969.); Smith R. A., Wave Mechanics of Crystalline Solids, Chapman and Hall, 2nd edition, 1969.

3. Общие вопросы

Иоффе А. Ф. Физика полупроводников.— М.-Л.: изд-во АН СССР, 1957; Adler R. B., Smith A. C., Longini R. L., Introduction to Semiconductor Physics, Wiley, 1964; Moll J. L., Physics of Semiconductors, McGraw-Hill, 1964; McKelvey J. P., Solid State and Semiconductor Physics, Harper and Row, 1966; Wang S., Solid State Electronics, McGraw-Hill, 1966; Blatt F. J., Physics of Electronic Conduction in Solids, McGraw-Hill, 1968; Sze S. M., Physics of Semiconductor Devices, Wiley Interscience, 1969; Wolf H. F., Semiconductors, Wiley Interscience, 1971; Elliott R. J., Gibson A. F., An Introduction to Solid State Physics and its Applications, Macmillan, 1974.

4. Примеси и несовершенства в кристаллах

Milnes A. G., Deep Impurities in Semiconductors, Wiley, 1973; Lattice Defects in Semiconductors, ed. Huntley F. A., Inst. of Phys. Series, 1974; Point Defects in Solids, eds. Crawford J. H., Slifkin L. M., Plenum Press, 1975; *Stoneham A. M.*, Theory of Defects in Solids, Clarendon Press, 1975. (Имеется перевод: *Стоунхэм А. М.* Теория дефектов в твердых телах.— М.: Мир, 1978.)

5. Ионная имплантация и радиационные повреждения

Mayer J. W., *Eriksson L.*, *Davies J. A.*, Ion Implantation in Semiconductors, Academic Press, 1970; Radiation Effects in Semiconductors, ed. Corbett J. W., Watkins G. D., Gordon and Breach, 1971; *Dearnaley G. et al.*, Ion Implantation, North Holland, 1973; *Cartex G.*, *Grant W. A.*, Ion Implantation in Semiconductors, Edward Arnold, 1976.

6. Распределение электронов по энергетическим уровням

Blakemore J. S., Semiconductor Statistics, Pergamon Press, 1962. (Имеется перевод: *Блекмор Дж.* Статистика электронов и дырок в полупроводниках.— М.: Мир, 1964.)

7. Характеристики явлений переноса

Electronic Structure in Solids, ed. Haidemenakis E. D., Plenum Press, 1969; Сборник трудов I и II совещаний по термоэлектричеству / Под ред. Кутасова В. А.— М.-Л., 1963; *Many A.*, *Goldstein Y.*, *Grover N. B.*, Semiconductor Surfaces, North Holland, 1965; *Frankl D. R.*, Electronic Properties of Semiconductor Surfaces, Pergamon Press, 1967; Metal — Semiconductor Contacts, ed. Pepper M., Inst. of Phys. Conf. Series, 1974; Physics of Solids in High Magnetic Fields, ed. Haidemenakis E. D., Plenum Press, 1969; *Бонч-Бруевич В. Л.* Вопросы электронной теории сильнолегированных полупроводников, в сб. Физика твердого тела.— М.: Наука, 1965, стр. 129—236; *Фистуль В. И.* Сильнолегированные полупроводники.— М.: Наука, 1967; Tunneling Phenomena in Solids, ed. Burstein E., Lundquist S., Plenum Press, 1969. (Имеется перевод: Туннельные явления в твердых телах/Под ред. Бурштейна Э., Лундквиста С.— М.: Мир, 1973; см. также приложения к Solid State Physics (§ 1).

8. Оптические свойства

Houghton J., *Smith S. D.*, Infra-red Physics, Oxford University Press, 1966; *Greenaway D. L.*, *Harbeke G.*, Optical Properties and Band Structure of Semiconductors, Pergamon Press, 1968; Optical Properties of Solids, ed. Haidemenakis E. D., Gordon and Breach, 1970; *Pankove J. I.*, Optical Processes in Semiconductors, Dover Publ., 1975 (Имеется перевод: Панков Ж. Оптические процессы в полупроводниках.— М.: Мир, 1973.) Optical Properties of Solids, ed. Abeles. F. North Holland, 1972; *Moss T. S.*, *Burrell G. J.*, *Ellis B.*, Semiconductor Opto-electronics, Butterworths, 1973; Optical Properties of Solids — New Developments, ed. Seraphin B. O., North Holland, 1976.

9. Структура энергетических зон

Pincherle L., Electronic Energy Bands in Solids, Macdonald, 1971; *Phillips J. C.*, Bonds and Bands in Semiconductors, Academic Press, 1973; см. также Slater J. C. (§ 2).

10. Приготовление материалов

Brooks M. S., Kennedy J. K., Ultra-Purification of Semiconductors, Macmillan, 1962; The Art and Science of Growing Crystals, ed. Gilman J. J., Wiley, 1963; Crystal Growth, ed. Bardsley W., Heurle D. J. T., Mullin J. B., North Holland, 1973—; Electronic Materials, ed. Hannay N. B., Colombo U., Plenum Press, 1973; *Kane P. F., Larrabee G. B.*, Characterization of Semiconducting Materials, McGraw-Hill, 1975; *Runyan W. R.*, Semiconductor Measurement and Instrumentation, McGraw-Hill, 1975.

11. Свойства материалов

Materials used in Semiconductor Devices, ed. Hogarth C. A., Wiley Interscience, 1965; II—VI Semiconducting Compounds, ed. Thomas D. G., Benjamin, 1967; Semiconductors and Semimetals (§ 1); Handbook of Electronic Materials, v. 2—7, IFI Plenum, 1971—1972; Electronic Materials (§ 10); Physics of Semimetals and Narrow-Gap Semiconductors, eds. Carter D. L., Bates R. T., Pergamon Press, 1971; *Шмарцев Ю. В., Валов Ю. А., Борщевский А. С.* Жаростойкие алмазоподобные полупроводники/Под ред. Горюновой Н. А., Наследова Д. А.—М.: Металлургия, 1964; *Gutman F., Lyons L. E.*, Organic Semiconductors, Wiley, 1967.

12. Специальные вопросы

Polarons and Excitons, eds. Kuper C. G., Whitfield G. D., Oliver and Boyd, 1962; Polarons in Ionic Crystals and Polar Semiconductors, ed. Devreese J. T.; North Holland, 1972; Polaritons, eds. Burstein E., de Martini F., Pergamon Press, 1974; The Physics and Technology of Semiconductor Light Emitters and Detectors, ed. Froya A., North Holland, 1973; *Sargent M., Scully M. O., Lamb W. E.*, Laser Physics, Addison-Wesley, 1974; Laser Applications to Optics and Spectroscopy, eds. Jacobs S. F. et al., Addison-Wesley, 1975; Very High Resolution Spectroscopy, ed. Smith R. A., Academic Press, 1976; Solids Under Pressure, eds. Paul W., Warshauer D. M., McGraw-Hill, 1963.

13. Аморфные полупроводники и полупроводники с низкой подвижностью носителей заряда

Губанов А. И., Квантово-электронная теория аморфных полупроводников.— М.—Л.: Изд-во АН СССР (Ленинградское отделение), 1963; *Adler D.*, Amorphous Semiconductors, Butterworths, 1971; *Mott N. F., Davis E. A.*, Electronic Processes in Non-Crystalline Materials, Oxford University Press, 1971; Electronic and Structural Properties of Amorphous Semiconductors, eds. Le Comber P. G., Mort J., Academic Press, 1973; Amorphous and Liquid Semiconductors, ed. Tauc J., Plenum Press, 1974; см. также Труды международных конференций по аморфным полупроводникам, которые проводятся каждые два года.

ДОПОЛНИТЕЛЬНЫЙ СПИСОК ЛИТЕРАТУРЫ

Общие руководства

- Бонч-Бруевич В. Л., Калашников С. Г.* Физика полупроводников.— М.: Наука, 1978.
Зеегер К. Физика полупроводников.— М.: Мир, 1977.
Киттель Ч. Введение в физику твердого тела.— М.: Мир, 1978.
Ашкрофт Н., Мермин Н. Физика твердого тела, т. 1 и 2.— М.: Мир, 1979.

Теория полупроводников, теория твердого тела

- Ансельм А. И.* Введение в теорию полупроводников.— М.: Наука, 1978.
Бир Г. Л., Пикус Г. Е. Симметрия и деформационные эффекты в полупроводниках.— М.: Наука, 1972.
Займан Дж. Вычисление блоховских функций.— М.: Мир, 1973.
Маделунг О. Теория твердого тела.— М.: Наука, 1980.
Цидильковский И. М. Электроны и дырки в полупроводниках.— М.: Наука, 1972.
Шкловский Б. И., Эфрос А. Л. Электронные свойства легированных полупроводников.— М.: Наука, 1979.

Примеси и дефекты в полупроводниках

- Емцов В. В., Машовец Т. В.* Примеси и точечные дефекты в полупроводниках.— М.: Радио и связь, 1981.

Явления электронного переноса

- Аскеров Б. М.* Кинетические эффекты в полупроводниках.— Л.: Наука, 1970.

Плазма и электрическая неустойчивость в полупроводниках

- Басс Ф. Г., Гуревич Ю. Г.* Горячие электроны и сильные электромагнитные волны в плазме полупроводников и газового разряда.— М.: Наука, 1975.
Бонч-Бруевич В. Л., Звягин И. П., Миронов А. Г. Доменная электрическая неустойчивость в полупроводниках.— М.: Наука, 1972.
Владимиров В. В., Волков А. Ф., Мейлихов Е. З. Плазма полупроводников.— М.: Атомиздат, 1979.
Денис В., Пожела Ю. Горячие электроны.— Вильнюс: Минтис, 1971.
Левинштейн М. Е., Пожела Ю. К., Шур М. С. Эффект Ганна.— М.: Советское радио, 1975.
Пожела Ю. К. Плазма и токовые неустойчивости в полупроводниках.— М.: Наука, 1977.
Райс Т., Хенсел Дж., Филлипс Т., Томас Г. Электронно-дырочная жидкость в полупроводниках.— М.: Мир, 1980.

Оптические и фотоэлектрические свойства полупроводников

- Акимов И. А., Черкасов Ю. А., Черкашин М. И. Сенсibilизированный фотоэффект.— М.: Наука, 1980.
 Аут И., Генцов Д., Герман К. Фотоэлектрические явления.— М.: Мир, 1979.
 Кардона М. Модуляционная спектроскопия.— М.: Мир, 1972.
 Уханов Ю. И. Оптические свойства полупроводников.— М.: Наука, 1977.
 Фридкин В. М. Фотосегнетоэлектрики.— М.: Наука, 1979.

Контактные явления

- Ламперт М., Марк П. Инжекционные токи в твердых телах.— М.: Мир, 1973.
 Литовченко В. Г., Горбань А. П. Основы физики микроэлектронных систем металл — диэлектрик — полупроводник.— Киев: Наукова думка, 1978.
 Свойства структур металл — диэлектрик — полупроводник. Под ред. Ржанова А. В.— М.: Наука, 1976.
 Стриха В. И. Теоретические основы работы контакта металл — полупроводник.— Киев: Наукова думка, 1974.

Поверхностные явления в полупроводниках

- Волькенштейн Ф. Ф. Физикохимия поверхности полупроводников.— М.: Наука, 1973.
 Дэвисон С., Левин Дж. Поверхностные (таммовские) состояния.— М.: Мир, 1973.
 Киселев В. Ф. Поверхностные явления в полупроводниках и диэлектриках.— М.: Наука, 1970.
 Моррисон С. Химическая физика поверхности твердого тела.— М.: Мир, 1980.
 Ржанов А. В. Электронные процессы на поверхности полупроводников.— М.: Наука, 1971.

Гальваномагнитные явления в полупроводниках

- Вайс Г. Физика гальваномагнитных полупроводниковых приборов и их применение.— М.: Энергия, 1974.
 Кучис Е. В. Методы исследования эффекта Холла.— М.: Советское радио, 1974.

Магнитные полупроводники

- Нагаев Э. Л. Физика магнитных полупроводников.— М.: Наука, 1979.

Аморфные полупроводники

- Иоффе А. Ф., Регель А. Р. Некристаллические, аморфные и жидкие полупроводники. Избранные труды А. Ф. Иоффе, т. 2.— М.: Наука, 1975.

Полупроводниковые приборы

- Берг А., Дин П. Светодиоды.— М.: Мир, 1979.
 Богданкевич О. В., Дарзник С. А., Елисеев П. Г. Полупроводниковые лазеры.— М.: Наука, 1976.
 Васильев А. М., Ландсман А. П. Полупроводниковые фотопреобразователи.— М.: Советское радио, 1971.
 Пасынков В. В., Чиркин Л. К., Шинков А. Д. Полупроводниковые приборы.— М.: Высшая школа, 1981.
 Пауль Р. Транзисторы.— М.: Советское радио, 1973.
 Секен К., Томпсет М. Приборы с переносом заряда.— М.: Мир, 1978.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Аккумуляция носителей заряда 231—232, 250
Акцептор 35, 82, 83
— неионизованный, равновесная концентрация 116
Алмаз 32
Антимонид индия 468—471
— — аморфный 536
— — вымораживание носителей 455, 456
— — детекторы ИК-излучения 378, 456
— — константа взаимодействия α 284
— — лазеры на перевороте спина 521
— — поглощение двухфотонное 362
— — — собственное 349, 350
— — сдвиг зон в магнитном поле 333, 334
Антирацен 687
Арсенид галлия 468—471
— — аморфный 536
— — константа взаимодействия α 284
— — лазеры 518
— — магнитосопротивление 454
— — светодиоды 404
— — экситоны 365, 404
— — электропоглощение 398
— — эффекты сильного электрического поля 433—438
Барьер потенциальный 33, 242, 250, 253, 258, 265
— Шоттки 266. См. также Контакт металл — полупроводник
Бизкситон 492—493
Вакансия 65, 66, 86—92
— анионная 87
— двойная (дивакансия) 92
— катионная 87
Валентность 31, 74, 78
— компенсированная 32, 71, 476
Вектор волновой 41, 44
— обратной решетки 414
Вымораживание дефектов 66
Внедрение ионное 92
Возбуждение тепловое 34—36, 54
Волны звуковые 272, 273
— плоские 41
Вращение Фарадея 340
Время жизни 224
— — в Ge 303, 304
— — в Te 303
— — измерения 460
— — неосновных носителей 210
— — в нитевидных образцах 320—324
— релаксации 125, 126
— — зависимость от энергии 145—149
— — легких дырок 282
— — не зависящее от энергии 127
— — объемного заряда 207, 210, 437, 489
— — тензор 149, 269
— — усредненное 142—144
Вымораживание носителей 455, 456
— — в InSb 456
Выпрямитель 21
— с гетеропереходом 266
— кремниевый 21, 246
— меднозакисный (купроксный) 21
— с $p-n$ -переходом 243, 259
— селеновый 21
Выпрямление 16, 21, 22, 43, 259, 266
Выращивание кристаллов 459
Вырождение 101, 105, 106, 110, 197
— спиновое примесного уровня 102
Галенит 20, 22
Галлий — алюминий — мышьяк 518
Галогениды щелочных металлов 20, 71, 89
Германий, кремний 16, 18, 22, 23, 50, 53, 69, 461—465, 531—535
— — аморфные 526, 531—535
— — изготовление 533
— — коэффициент поглощения 534, 535
— — $p-n$ -переходы 533
— — плотность состояний 533, 534
— — поглощение собственное 534
— — подвижность носителей 525, 526, 533
— — — уровень Ферми 532
— — — эффект поля 534
— — время жизни 311, 313, 314
— — капли электронно-дырочные 493—496
— — картина связей 76, 77
— — магнитосопротивление 165, 170, 454
— — массы эффективные 447, 462, 463
— — отщепление спин-орбитальное 335
— — переходы не прямые 358, 359
— — — прямые 350, 351
— — поглощение в далекой ИК-области 336, 337
— — — в магнитном поле 447—450
— — — двухфотонное 362, 363
— — — решеточное 374—376
— — подвижность носителей 127, 434
— — — в сильном магнитном поле 428, 429
— — — при высоком давлении 511
— — применение в технике 21—24
— — примесь водорода 86
— — примесь водородоподобная 84, 85
— — пробой лавинный 433
— — провондаемость диэлектрическая 84
— — распад, свойства 465
— — резонанс циклотронный 342, 445
— — — электронно-спиновый 373

- — рекомбинация 313, 314
- — сильнолегированные 506
- — скорость дрейфовая (в сильном поле) 431, 438
- — — поверхностной рекомбинации 319
- — собственные, концентрация носителей 108, 112
- — — удельное сопротивление 464
- — спектр электроотражения 400
- — сплав 465
- — структура зонная 461, 462
- — — кристаллическая 75, 511
- — — темп излучательной рекомбинации 303, 304
- — температура плавления 458
- — теплопроводность 180
- — уровни примесей 465
- — условия вырождения 114
- — фононы 358
- — фотопроводимость внутризонная 391
- — ширина запрещенной зоны 76, 447
- — — — зависимость от температуры 463, 464
- — — экситон 96, 97
- — — «непрямой» 367, 368
- — — «прямой» 351, 364, 365
- — — энергия связи 368, 494
- — — эффект *Бурштейна — Мосса* 502
- — — увлечения 396
- — — эффекты междолинного переброса электронов 438
- Гетеропереход 263—266, 518
- Границы зерен 70
- Движение носителей (под действием внешних полей сил) 55—59
- Детекторы ИК-излучения 23, 378, 379, 391, 397, 456, 478, 481
- кристаллические 22
- Дефекты 64—69, 91—92
- «вмороженные» 66
- *Френкеля* 66
- *Шоттки* 66
- Диаграмма энергетическая 29, 43, 60, 80, 82, 204
- — гетероперехода 264, 265
- — $p-n$ -перехода 249
- — $n^+ - n$ -перехода 236, 242
- Диод *Есаки* (туннельный) 500, 503—506
- Дислокации 67—71
- рассеяние 295, 296
- рекомбинация 315, 316
- стенки 69
- Диффузия носителей 202, 203
- — в примесных материалах 210—215
- — в собственных материалах 225—229
- — в электрическом поле 212—215
- — — общее решение 218—221
- — облака носителей 220—222, 261, 262
- Длина (радиус) *Дебая* 289, 290, 501
- диффузии 212, 228
- — амбиполярный 228
- — по полю 219, 228
- — против поля 219, 228
- — свободного пробега, средняя 124, 145—147, 280, 488
- Домены (сильное поле) 436, 437
- акустические 439
- Донор 35, 79—81
- неионизованный, равновесная концентрация 116
- Дрейф носителей заряда 123, 124
- облака неосновных носителей 220—222, 226, 261, 262
- Дырки 34, 54, 55, 57, 73, 110
- время релаксации 282
- дрейфовая скорость в магнитном поле 136
- концентрация 35, 109—115
- легкие и тяжелые 50, 282, 369, 442
- связанные 81
- эффективная масса 55, 57—58
- Емкость барьерного слоя 241
- диффузионная 248
- Закон дисперсии для свободных электронов 41
- — для колебаний решетки 272, 273
- — для поляритонов 497—499
- *Кирхгофа* (для излучения) 401, 402
- *Ома* (отклонения) 425, 428—438
- *Планка* (для излучения) 301
- Заряд иона эффективный 275, 470, 499
- — объемный 207
- — время релаксации 207, 499
- — нестабильность 437
- Заселенность инверсная 513, 515
- Зона 27, 42, 44, 45
- *Бриллюэна* 44, 45, 48—50, 53, 484
- валентная 34, 47, 50
- заполненная 30, 32, 33, 45, 60
- запрещенная 27, 31, 32, 47
- — зависимость от давления 509, 510
- — от состава 470, 478—480
- — от температуры 463, 464, 468, 472, 473, 481
- — измерения 107, 447, 448, 460
- — «оптическая» 538, 539
- — подвижностей 528—530, 533
- примесная 171—173
- проводимости 34, 47
- Зоны вырожденные 50—53, 282, 335
- — — — — в сильных магнитных полях 441, 442
- — — — — поглощение свободными носителями 335
- — — — — рассеяние носителей 282
- — — — — эффективные массы 53
- — — — — гофрированные 170
- — — — — незаполненные 30
- — — — — «сферические» 47
- — — — — форма 45—54
- — — — — расчет 413—417
- — — — — «хвосты» 528—530, 537, 540
- — — — — «эллипсоидальные» 48, 49
- Излучение 302, 304, 401, 402
- быстрых электронов 431, 432
- вынужденное 512, 513
- изолэктронного центра 405
- колебаниями решетки 406
- рекомбинационное 302—304, 402, 403
- свободных носителей 405, 406
- спонтанное 512, 513
- экситонов 403—405
- Имплантиция ионная см. Введение ионное
- Импульс фонона 278
- фотона 345
- Инжекция носителей 37, 215, 219, 220, 233, 244, 260
- — оптическая 206
- Ионизация ударная 245, 423, 424, 433
- Истоциение примесное 111
- Кадмий — ртуть — теллур (КРТ) 478, 479, 522
- Капли электронно-дырочные 493—496
- Карбид кремния 485

- Квазиимпульс 42, 44, 56
 Квазиуровень Ферми 208, 209, 514
 Квантование в магнитном поле 342, 343, 420, 421, 439—455
 Колебания нулевые 271
 — — — рассеяние 278, 279
 — — — решетки 269—275, 531, 540
 — — — излучение 406
 — — — квантовая теория 271
 — — — комбинационное рассеяние 522
 — — — обмен энергией с электронами 424—430
 — — — поглощение 373—377
 — — — уровни энергии 271
 Компенсация валентностей 32, 71, 476
 — примесей 113, 114, 117—119, 171—173
 — — коэффициент 171
 Конденсация экситонов 493—496
 Константа взаимодействия (оптические колебания) 284, 496
 Константы оптические 326—329
 Контакт инжектирующий 214, 259, 264
 — коллекторный 245
 — металл — окисел — полупроводник (МОП) 259
 — металл — полупроводник 256—261
 — неинжектирующий 250, 259
 — омический 229, 259
 Концентрация носителей 34, 37, 38, 99—121
 — — — в примесных полупроводниках 109—121
 — — — измерения 133
 — — — при низких температурах 118—120
 — — — собственная 102—109
 Коэффициент диффузии 202, 204, 324
 — — — амбиполярной 226, 227, 387, 393
 — — — инъекции 212, 244, 259, 260, 264
 — — — Нернста 194, 197
 — — — отражения 327, 400
 — — — изменение в электрическом поле 397, 400
 — — — поглощения 327, 397, 513, 534, 535
 — — — Риги — Ледюка 196, 198
 — — — Томсона 181
 — — — Холла 133, 149—156
 — — — измерения 132
 — — — многодолинные полупроводники 154, 156
 — — — прычские полупроводники 133, 150
 — — — сильное поле 141, 151
 — — — смешанная проводимость 136, 140, 151—153
 — — — Эттингсгаузена 192, 197
 Кремний см. Германий и кремний
 Кристалл ионный 71—73

 Лазер 24, 263, 466, 471, 511—522
 — материалы 517, 518
 — на перевороте спина 518—522
 — резонатор 515, 516
 Лестница Ландау 441—443, 449, 450
 Ловушки 227, 307—314
 — влияние на фотопроводимость 382—386
 — в процессе рекомбинации 300, 307—314, 318, 319
 — — — в Ge 311
 — — — поверхностные 256, 318, 319

 Магнетон Бора 372
 Магнитосопротивление 141, 142, 156—170
 — в сильных магнитных полях 452, 454
 — коэффициент 142, 157
 — отрицательное 452, 454
 — поверхности постоянной энергии сферические 156—161
 — — — — эллипсоидальные 161—170
 Максимумы и минимумы энергетических зон 46
 — — эквивалентные 48, 49, 107, 153
 Масса эффективная «дрейфовая» 130, 115
 — — — плотности состояний 168
 Материалы полупроводниковые 458—460, 489, 535—537
 — — — для изготовления лазеров 517, 518
 — — — изготовление 459
 — — — измерения 460. См. также Полу-проводники и Соединения
 Матрица плотности 453
 Метод Кельвина 254
 — к-р 416, 417
 — ЛКАО 414
 — Стокбаргера 459
 — Чотральски-го 459
 — эффективной массы 418—420
 — ячеек 413
 Методы плоских волн 414—415
 Минимум вышерасположенный 424, 434, 469
 — — в Ge 438
 Модель Блоха 26, 27
 — Зоммерфельда 26, 27
 Молекула экситонная 492, 493

 Напряжение Зинера 245
 Непарабolicность 416, 417, 447
 Несовершенства (в кристалле) 64—71
 Носители заряда неосновные 38, 112, 210
 — — — время жизни 210
 — — — длина диффузионная 212, 219, 224
 — — — основные 38, 112

 Обмен энергиями между электронами и решеткой 123, 278, 424—430
 Окислы металлов 481—484
 Олово серое 466—468
 Орбита боровская 85
 Отжиг 66, 92, 536

 Пакет волновой 42
 Пары электронно-дырочные 34, 73, 76, 215, 377
 — — генерация 301—304, 377, 513
 — — рекомбинация 209
 Переход Мотта 454, 483, 494
 — полупроводниковый, $p-n$ 233—241, 251, 503—506
 — — — вольтамперная характеристика 217
 — — — емкость барьерного слоя 241
 — — — — диффузионная 248
 — — — коэффициент инжекции 244
 — — — объемный заряд 237—240
 — — — проводимость дифференциальная 247
 — — — свойства на высоких частотах 247, 248
 — — — тонкий 240, 241
 — — — широкий 241
 — — — электростатический потенциал 236, 237
 — — — энергетическая схема 236
 — — — $p+n, p+n-p$ 240, 244
 — — — $p+r, p+n-p$ 249, 251
 Переходы фазовые 511
 — энергетические непряные 345, 352—360
 — — — прямые 345—350
 — — — запрещенные 348, 366
 — — — разрешенные 346, 348, 349, 366
 — — — фазовые 511
 Плазмон 300

- Оже 304—307
- поверхностная 256, 317, 386
- — измерения 395
- — скорость 318
- через доноры и акцепторы 316, 317
- через ловушки 307—314
- через экситоны 314, 315
- Решетка кристаллическая идеальная 26, 40, 64, 269, 270
- типа алмаза 75
- Ртуть — кадмий — теллур *см.* Кадмий — ртуть — теллур
- Светодиоды 24, 405, 517
- Свинец — олово — теллур (COT) 479, 480, 481, 518
- Связи «болтающиеся» 528, 532, 537
- Связь гомеополлярная 74—76
- ионная (полярная) 71—73
- — смешанная 77, 78
- Сдвиг зон в магнитном поле 343, 344, 441
- Селен 21, 466
- аморфный 535, 536
- соединения 472—477
- Сечение рассеяния интегральное 124
- — для проводимости 125
- Скорость групповая фонона 272
- — электрона 42, 44
- дрейфовая 123—127, 261, 262, 430
- — в магнитном поле 131, 136
- — насыщения 431, 432
- утечки 232, 251
- Слой инверсный (поверхностно-барьерный) 254, 259
- оксидный 256, 259
- Соединения азота 485
- алюминия, арсенид и фосфид 485. *См. также* Полупроводники $A^{III}B^V$ и Соединения тройные
- бериллия, селенид, сульфид и теллурид 78
- бора, арсенид, нитрид и фосфид 485
- европия 487
- пода, нитриды кадмия, свинца и ртути 96, 364
- кадмия, селенид, сульфид и теллурид 475, 496. *См. также* Кадмий — ртуть — теллур
- Соединения тройные
- меди, бромид и хлорид 492, 493, нитрид 18
- — оксиды 18, 21, 96, 364, 481. *См. также* Соединения четверные
- мышьяка, селенид, сульфид и теллурид 536, 538—540. *См. также* Арсенид галлия, Арсенид индия и Полупроводники $A^{III}B^V$
- свинца, селенид, сульфид и теллурид 23, 78, 87, 89, 90, 284, 285, 378, 472—475. *См. также* Свинец — олово — теллур
- серы 20, 472—477
- теллура 53, 180, 471—481, 486. *См. также* Кадмий — ртуть — теллур и Свинец — олово — теллур
- тройные 475—481
- хрома 486, 487
- четверные 475—477
- Соотношение Бриджмена 195
- Лиддана — Закса — Теллера 275, 498
- Эйнштейна 204
- Сопровождение отрицательное 436, 437
- удельное Ge и Si 464
- Состав стехиометрический 87, 91, 476
- Спектр излучения 403
- поглощения в далекой ИК-области 336, 337
- — примесного 370
- — решеточного 374
- — собственного 349, 359
- — экситонного 96, 97, 351, 365, 368
- Спектроскопия туннельная 506—508
- Станиты 476
- Стенки дислокационные 69
- Стекла халькогенидные 536—539
- Структура зонная 45—54
- — вычисления 413—417
- — измерения 400, 401
- кристаллическая 72
- — типа алмаза 75
- — типа цинковой обманки 77
- Теллур 16, 18, 303, 305, 466, 486
- аморфный 535, 536
- излучательная рекомбинация 303
- рекомбинация Оже 303, 305
- циклотронный резонанс 445, 466
- электропроводность 129
- Темп генерации 209, 302, 305, 380, 514
- рекомбинации 302, 305, 318, 514
- Температура эффективная электронов 423, 427, 428, 432, 437
- Теория Блоха 26, 27, 41, 412
- Вильсона 24, 25, 27, 31, 33
- Зоммерфельда 25—28, 31
- зонной структуры 413—417
- Теплопроводность 176, 179, 191
- германия 180
- Термо-э. д. с. 16, 17, 19, 181—185, 187
- Техника модуляционная 397, 401
- Ток насыщения 244, 247, 258
- Точки критические (зонной структуры) 369, 400, 401
- Туннелирование 171, 259, 397, 503—508
- с участием фононов 506, 507
- Увлечение носителей фононами 186, 187
- — фотонами 396, 397
- Угол Холла 131, 133, 136, 150
- Упорядочение дислокаций 69—71
- Уравнения Больцмана 142, 150, 175, 432, 433
- непрерывности 207, 211
- Праассона 207, 223, 237, 239, 253, 289
- Шокли (для p — n -перехода) 243
- Шредингера 40, 41, 419, 421
- Уровень Ферми 101
- — в примесных полупроводниках 111, 114, 118
- — в собственных полупроводниках 105, 109
- — при низких температурах 117—121
- Уровни энергетические примесей 80, 82, 84, 86, 371, 465, 471
- — Ландау 440
- — Тамма 252
- Фононы 277—279, 300
- испускание и поглощение 377, 426
- равновесная плотность 377, 426, 512
- Фосфид галлия 468—471
- — поглощение решеточное 374
- — экситонное 368
- — экситоны 314, 315, 368, 369, 405, 493
- Фотопроводимость 16, 17, 20, 37, 377—380
- в InSb при наличии магнитного поля 456
- влияние ловушек 383—386
- влияние поверхностной рекомбинации 386, 387
- внутризонная 391
- двухфотонная 362, 390, 391
- Фталоцианиты 487
- Функции Блоха 41, 412, 418

- Ванье 412, 418
- Функция весовая 143, 452
- волновая 41
- распределения 99, 205, 423
- — для дырок 102
- — классическая 101
- — Ферми — Дирака 100

- Халькопириты 476—478
- Характеристики вольтамперные контакта
- металл — полупроводник 258
- — p — n -перехода 246, 247

- Центр изоэлектронный 405
- Центры захвата *см.* Ловушки
- прилипания 383—386

- Частота ионно-плазменная 499
- остаточных лучей 174, 284, 360, 373—375, 406, 498
- плазменная 338
- циклотронная 59, 137, 340, 341
- Число квантовое 85, 93, 363, 364
- — магнитное 343, 440, 443, 447
- — правила отбора 442, 444, 449
- — спиновое 443, 446, 449

- Ширина запрещенной зоны *см.* Зона запрещенная

- Э. д. с. Холла 133, 134
- Экранирование свободными носителями 289, 501, 502
- Экситон 92—97, 348, 351, 328—334, 399, 493
- возбужденные состояния 93, 94, 369
- поглощение 96, 348, 351
- связанный 403—405
- энергия связи 93, 368
- Эксклюзия носителей 230, 232
- Эксперимент *Бенедикта* и *Шокли* 332
- *Хейнса* и *Шокли* 221, 261, 262
- Экстракция носителей 216, 232
- Электронейтральность 207, 211, 489
- Электроны «горячие» и «теплые» 432
- Электропроводность 127—130
- германия и кремния 464
- по примесной зоне 171—173
- примесная 35, 36
- собственная 34, 35
- Энергия активации 34, 529
- кинетическая носителей заряда 41, 61, 62
- — — в магнитном поле 343, 441, 446
- Эффект акустоэлектрический 439
- *Бурштейна* — *Мосса* 502, 514
- *Ганна* 436, 438
- *Дембера* 392
- *Земмана* 450
- *Нернста* 188, 194, 197
- поля 255, 256, 319, 530
- *Ризи* — *Ледюка* 188, 196, 198
- фотогальванический 17, 378, 396
- — поперечный (эффект *Дембера*) 592
- фотоэлектромагнитный 392
- — использование для измерения скорости поверхности для измерения скорости рекомбинации 395
- *Франца* — *Келдыша* 397, 400
- *Холла* 17, 18, 131
- — адиабатический 187
- — в аморфных полупроводниках 530
- — в очень сильных полях 141, 156, 454
- — изотермический 188
- *Эттингсгаузена* 187, 192, 197
- Эффекты высокочастотные 335, 337, 338
- при высоком давлении 508—511

- Ячейка элементарная 41, 44, 483
- — *Вигнера* — *Зейтца* 413

ОГЛАВЛЕНИЕ

Предисловие редактора перевода	5
Предисловие	8
Из предисловия автора к первому изданию	11
1. ПРОСТЕЙШИЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ	15
1.1. Ранние исследования, посвященные полупроводникам	15
1.2. Применение полупроводников	21
1.3. Элементарная теория полупроводников	24
1.4. Способы управления концентрацией носителей	37
Литература	39
2. УРОВНИ ЭНЕРГИИ В КРИСТАЛЛИЧЕСКИХ ТВЕРДЫХ ТЕЛАХ	40
2.1. Волновая механика свободных электронов	40
2.2. Движение в пространстве с периодическим потенциалом	41
2.3. Форма энергетических зон	45
2.4. Положительные дырки	54
2.5. Движение электронов и дырок в кристалле под действием внешних силовых полей	55
2.6. Схема энергетических уровней	59
2.7. Влияние нарушений периодичности кристаллической решетки на движение электронов и дырок в кристалле	62
Литература	63
3. ПРИМЕСИ И НЕСОВЕРШЕНСТВА В КРИСТАЛЛАХ	64
3.1. Типы несовершенств решетки	64
3.2. Химическая связь в полупроводниках	71
3.3. Другое определение полупроводников	78
3.4. Примеси замещения в полупроводниках	79
3.5. Экситоны	92
Литература	97
4. РАВНОВЕСНЫЕ КОНЦЕНТРАЦИИ НОСИТЕЛЕЙ ЗАРЯДА	99
4.1. Распределение электронов по энергетическим уровням	99
4.2. Собственные полупроводники	102
4.3. Примесные полупроводники	109
Литература	121

5. ЯВЛЕНИЯ ЭЛЕКТРОННОГО ПЕРЕНОСА	122
5.1. Рассеяние электронов на дефектах в кристалле. Время релаксации	122
5.2. Время релаксации, не зависящее от энергии	127
5.3. Время релаксации, зависящее от энергии	142
5.4. Электропроводность при очень низких температурах	171
Литература	173
6. ЭЛЕКТРОННЫЕ ТЕПЛОВЫЕ ЯВЛЕНИЯ	175
6.1. Теплопроводность	175
6.2. Термоэлектродвижущая сила	180
6.3. Терромагнитные явления	187
6.4. Терромагнитные явления в условиях сильного вырождения	197
6.5. Случай сильных магнитных полей	198
6.6. Относительные величины гальваномагнитных и терромагнитных эффектов	199
Литература	200
7. ДИФФУЗИЯ ЭЛЕКТРОНОВ И ДЫРОК	202
7.1. Неоднородные полупроводники	202
7.2. Соотношение Эйнштейна	203
7.3. Отклонения от теплового равновесия	206
7.4. Электронно-дырочная рекомбинация	209
7.5. Диффузия и проводимость в примесных полупроводниках ($n \gg p$ или $p \gg n$)	210
7.6. Дрейф облака неосновных носителей заряда в электрическом поле	221
7.7. Полупроводники с проводимостью, близкой к собственной	222
7.8. Сопоставление различных контактных явлений	232
7.9. p - n -переход	233
7.10. n^+ - n - и p^+ - p -переходы	249
7.11. Поверхностные свойства полупроводников	251
7.12. Контакты металл—полупроводник	256
7.13. Дрейфовая подвижность электронов и дырок	261
7.14. Гетеропереходы	263
Литература	266
8. РАССЕЯНИЕ ЭЛЕКТРОНОВ И ДЫРОК	268
8.1. Изменение состояния	268
8.2. Механизмы рассеяния	268
8.3. Рассеяние на колебаниях атомов решетки	269
8.4. Фононы	275
8.5. Рассеяние на колебаниях решетки	279
8.6. Рассеяние на оптических колебаниях в полярных кристаллах	284
8.7. Междолинное рассеяние	285
8.8. Межэлектронное рассеяние	286
8.9. Рассеяние на ионизованных примесях	287
8.10. Рассеяние на нейтральных примесях	291
8.11. Рассеяние на дислокациях	295
8.12. Вклад различных видов рассеяния в величину подвижности	296
Литература	297
9. РЕКОМБИНАЦИЯ ЭЛЕКТРОНОВ И ДЫРОК	299
9.1. Механизмы рекомбинации	299
9.2. Излучательная рекомбинация	301
9.3. Рекомбинация Оже	304

9.4. Рекомбинация через ловушки	307
9.5. Рекомбинация через экситоны	314
9.6. Рекомбинация на дислокациях	315
9.7. Рекомбинация через доноры или акцепторы при низких температурах	316
9.8. Поверхностная рекомбинация	317
9.9. Среднее время жизни в нитевидных образцах и тонких полосках	320
Литература	324
10. ОПТИЧЕСКИЕ И ВЫСОКОЧАСТОТНЫЕ ЯВЛЕНИЯ	326
10.1. Оптические характеристики полупроводников	326
10.2. Поглощение свободными носителями заряда	329
10.3. Плазменный резонанс	337
10.4. Высокочастотные эффекты в магнитном поле	338
10.5. Собственное поглощение	344
10.6. Экситонное поглощение	363
10.7. Примесное поглощение	369
10.8. Решеточное поглощение	373
10.9. Фотопроводимость	377
10.10. Диффузионная фото-э.д.с.	392
10.11. Фотоэлектромагнитный эффект	392
10.12. Увлечение свободных носителей заряда фотонами	396
10.13. Электропоглощение и электроотражение	397
10.14. Излучение полупроводников	401
Литература	406
11. ЗОННАЯ СТРУКТУРА И МЕТОД ЭФФЕКТИВНОЙ МАССЫ	411
11.1. Концепция эффективной массы	411
11.2. Функции Блоха и Ванье	412
11.3. Методы расчета зонной структуры	413
11.4. Вычисления зонной структуры	417
11.5. Метод эффективной массы	418
11.6. Квантование в магнитном поле	420
Литература	421
12. ВЛИЯНИЕ СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ И МАГНИТНЫХ ПОЛЕЙ НА ЯВЛЕНИЯ ПЕРЕНОСА И ОПТИЧЕСКИЕ СВОЙСТВА	423
12.1. Изменение функции распределения	423
12.2. Обмен энергией между электронами и решеткой	424
12.3. Горячие электроны	430
12.4. Эффекты междолинного переброса электронов	434
12.5. Квантовая теория движения электронов в сильном магнитном поле	439
12.6. Спиновое расщепление уровней	445
12.7. Влияние магнитного поля на межзонные переходы	446
12.8. Явления электронного переноса в сильных магнитных полях	450
12.9. Вымораживание носителей	455
Литература	456
13. ПОЛУПРОВОДНИКОВЫЕ МАТЕРИАЛЫ	458
13.1. Материалы, используемые в полупроводниковых приборах и исследованиях	458
13.2. Приготовление полупроводниковых материалов и измерение их свойств	459
13.3. Кремний и германий	461
13.4. Другие элементарные полупроводники	466

13.5. Соединения типа $AIII\text{BV}$	468
13.6. Соединения типа $AIV\text{VI}$	472
13.7. Тройные и четверные соединения	475
13.8. Очень узкозонные полупроводники	478
13.9. Оксидные полупроводники	481
13.10. Тугоплавкие полупроводники	484
13.11. Сверхпроводящие полупроводники	485
13.12. Магнитные полупроводники	486
13.13. Органические полупроводники	487
13.14. Полупроводники с низкой подвижностью носителей заряда	488
13.15. Другие полупроводники	489
Литература	489

14. НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ ВОПРОСЫ ФИЗИКИ ПОЛУПРОВОДНИКОВ

14.1. Экситонные молекулы	492
14.2. Конденсация экситонов в электронно-дырочные капли	493
14.3. Полярны и поляритоны	496
14.4. Сильнолегированные полупроводники	500
14.5. Эффекты, связанные с высоким давлением	508
14.6. Лазерный эффект в полупроводниках	511
14.7. Рассеяние излучения, обусловленное переворачиванием спина, п лазеры на этой основе	518
14.8. Комбинационное рассеяние на колебаниях атомов решетки	522
Литература	522

15. АМОРФНЫЕ ПОЛУПРОВОДНИКИ

15.1. Новые теоретические представления	525
15.2. Электронные состояния	527
15.3. Колебания атомов решетки	531
15.4. Аморфные Si и Ge	531
15.5. Другие элементарные аморфные полупроводники	535
15.6. Некоторые аморфные полупроводниковые соединения	536
15.7. Стеклообразные полупроводники	536
15.8. Заключение	541
Литература	542
Литература	544
Дополнительный список литературы	547
Предметный указатель	549

УВАЖАЕМЫЙ ЧИТАТЕЛИ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, изд-во «Мир».

Роберт Аллан Смит

ПОЛУПРОВОДНИКИ

Научный редактор А. И. Власенко. Мл. научн. редакторы В. И. Аксенова, В. Н. Цлаф.
Художник Н. И. Василевская. Художественный редактор Л. Е. Безрученков. Технический
редактор Е. В. Ящук. Корректор С. А. Денисова

ИБ № 2479

Сдано в набор 03.12.81. Подписано к печати 12.04.82. Формат 60×90^{1/16}. Бумага
газетная. Гарнитура литературная. Печать высокая. Объем 17,50 бум. л. Усл. печ. л.
35,00. Усл. кр.-отт. 35,00. Уч.-изд. л. 34,10. Изд. № 2/1183. Тираж 30 000 экз.
Заказ № 3533. Цена 2 р. 70 к.

ИЗДАТЕЛЬСТВО «МИР». Москва, 1-й Рижский пер., 2

Ордена Октябрьской Революции и ордена Трудового Красного Знамени Первая Образцовая
типография имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28