Ф. Ф. ВОЛЬКЕНШТЕЙН

ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

ОГИЗ
ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ТЕХНИКО-ТЕОРЕТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1947 ЛЕНИНГРАД

Ф. Ф. ВОЛЬКЕНШТЕЙН

ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

ОГИЗ
ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ТЕХНИКО-ТЕОРЕТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1947 ЛЕНИНГРАД

АННОТАЦИЯ

В монографии рассматриваются с некоторой единой точки зрения разнообразные электронные процессы, развивающиеся в реальных кристаллах. В книге даётся сводка экспериментального материала, хотя в основном книга носит теоретический характер. Она предназначена для физиков-экспериментаторов и инженеров, занимающихся электрическими свойствами твёрдых тел.

Редактор *К. П. Гуров.* Техн. редактор *М. С. Бондарев.*
Подписано к печати 21/V 1947. 22 печ. л. 20,45 уч.-изд. л. 38 400 тип. зн. в печ. л. А04537. Тираж 7 000 экз. Цена книги 12 р. 25 к. Переплёт 2 р. Заказ № 7023.

1-я Образцовая тип. треста „Полиграфкнига" Огиза при Совете Министров СССР. Москва, Валовая, 28.

ОГЛАВЛЕНИЕ

Введение 7

Глава I. Идеальные и реальные кристаллические решётки 15

§ 1. Характеристика реальной решётки 15
§ 2. Энергия решётки 26
§ 3. Электропроводность решётки 38
§ 4. Энергетический спектр решётки (частный случай) . 46
§ 5. Энергетический спектр решётки (общий случай) . . 53

Глава II. **Электронные уровни в идеальном кристалле** . . . 64

§ 1. Теория свободных электронов 64
§ 2. Учёт периодического поля решётки (гомеополярная решётка) 83
§ 3. Учёт периодического поля решётки (гетерополярная решётка) 110
§ 4. Учёт взаимодействия электронов 122

Глава III. **Электронные уровни в реальном кристалле** . . . 132

§ 1. Постановка задачи 132
§ 2. Гомеополярная решётка с пустыми узлами 141
§ 3. Гетерополярная решётка с пустыми узлами 156
§ 4. Решётка с междуузельными атомами 168

Глава IV. **Влияние температуры и примесей на электропроводность полупроводников** 179

§ 1. Обзор экспериментальных и теоретических работ . 179
§ 2. Общая теория электропроводности реальных гетерополярных решёток 199
§ 3. Сравнение теории с экспериментом 224

Глава V. Влияние сильных электрических полей на электропроводность полупроводников 241

§ 1. Сводка экспериментальных данных 241
§ 2. Влияние электрического поля на поведение электрона в решётке 248
§ 3. Деформация энергетических зон под влиянием электрического поля 261
§ 4. Ионизирующее действие поля 278
§ 5. Влияние поля на термическую ионизацию 299

Глава VI. Влияние освещения на электропроводность полупроводников 313

§ 1. Нормальный фотоэлектрический эффект 313
§ 2. Фотоэлектрически неактивное поглощение света . . 330
§ 3. Аномальный фотоэлектрический эффект 340

Эта книга в основной своей части была написана в Ленинграде в 1941—42 г., когда город, скованный кольцом вражеской блокады, подвергавшийся непрерывным бомбардировкам и артиллерийским обстрелам, лишённый освещения, отопления и умирающий от голода, героически сопротивлялся врагу. Памяти ученых, погибших в Ленинграде в эту трагическую зиму 1941—42 г., автор посвящает эту книгу.

ВВЕДЕНИЕ

А. Основные закономерности. Все кристаллические тела в отношении их электрических свойств могут быть разделены, как известно, на три класса: 1) металлы, 2) полупроводники, 3) диэлектрики. Полупроводники представляют собой наиболее обширный класс кристаллов. Большинство тел, с которыми мы встречаемся в природе, являются именно полупроводниками.

Кристаллы этих трёх классов различаются прежде всего чисто количественным признаком: величиной электропроводности σ. Мы имеем:

1) для металлов $\sigma = 10^4 - 10^5$ $ом^{-1}$ $см^{-1}$,
2) для полупроводников $\sigma = 10^2 - 10^{-10}$ $ом^{-1}$ $см^{-1}$,
3) для диэлектриков $\sigma = 10^{-14} - 10^{-15}$ $ом^{-1}$ $см^{-1}$.

Таким образом, к классу полупроводников принадлежат кристаллы с весьма разнообразными значениями σ, отличающимися в миллионы раз.

Различие между металлами, полупроводниками и диэлектриками не только количественное, но и качественное. Это значит, что на воздействие одних и тех же физических факторов кристаллы этих трёх различных классов реагируют различным образом.

Полупроводники, занимая промежуточное положение между металлами и диэлектриками, в некоторых отношениях похожи на металлы, в других отношениях они похожи на диэлектрики, но в то же время обладают целым рядом совершенно специфических свойств. Характерной особенностью полупроводников является то, что все их электрические свойства чрезвычайно неустойчивы. Иначе говоря, эти свойства резко изменяются от внешних воздействий. Электропроводность полупроводников, являющаяся предметом этой книги, также принадлежит к числу таких неустойчивых свойств. Она обнаруживает настолько острую чувствительность к различным факторам,

что вообще возникает вопрос, можем ли мы в отношении полупроводников считать электропроводность константой материала, как мы это имеем право делать в отношении металлов и диэлектриков. К главнейшим факторам, от которых зависит электропроводность полупроводников, относятся: 1) примеси, 2) температура, 3) электрическое поле, 4) освещение. Под влиянием этих факторов величина электропроводности варьируется в чрезвычайно широких пределах. В одном и том же полупроводнике, но при различных условиях, мы получаем для электропроводности значения, отличающиеся часто в миллионы раз.

Рассмотрим вкратце роль каждого из перечисленных выше факторов.

1. Термин «примеси» в отношении полупроводников не следует понимать слишком буквально. Наличие «примеси» не обязательно должно означать присутствие в кристаллической решётке чужеродных атомов, введённых извне. В отношении решёток, изображающихся химической формулой M_mR_r, где M — символ металла, а R — символ металлоида (большинство полупроводников представляет собой соединения именно такого типа), под «примесью» обычно понимают всякое отклонение от стехиометрического соотношения. В формуле M_mR_r такого рода отклонения могут быть отражены, если индексы m и r считать не целыми, а дробными числами, хотя и близкими к целым.

Стехиометрические нарушения могут быть двоякого рода: мы можем иметь в полупроводнике стехиометрический избыток металла или стехиометрический избыток металлоида. В первом случае полупроводник, как правило, обладает электронной проводимостью; во втором случае — дырочной.

Величина электропроводности очень чувствительна к тому, насколько сильно нарушено в полупроводнике стехиометрическое соотношение. При прочих равных условиях электропроводность тем больше, чем значительнее стехиометрические нарушения. Полупроводники представляют собой обычно кристаллы с нарушенным стехиометрическим соотношением, что и является чрезвычайно важной особенностью полупроводников. Именно это обстоятельство предопределяет все электрические, оптические и фотоэлектрические свойства полупроводников.

2. Зависимость электропроводности полупроводников от температуры для разных полупроводников различна. Наиболее

типичные кривые изображены схематически на рис. 1. На этом же рисунке выписаны названия полупроводников, которые являются наиболее характерными представителями того или иного типа. Как видно из рис. 1, зависимость электропроводности от температуры у полупроводников имеет в одних случаях металлический характер (рис. 1, *а*), а в других случаях тот же характер, что и у диэлектриков (рис. 1, *б*).

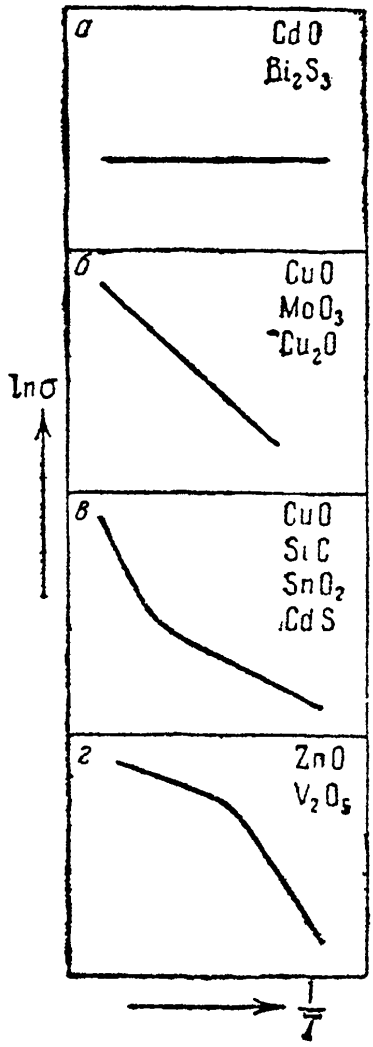


Рис. 1. Зависимость электропроводности от температуры для различных полупроводников.

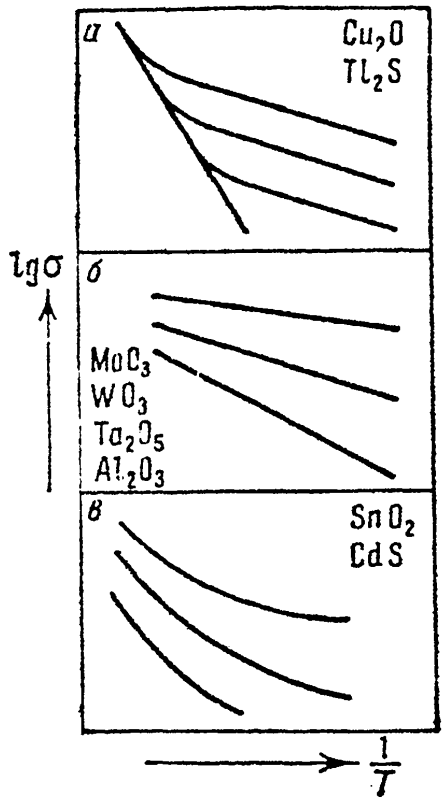


Рис. 2. Температурный ход электропроводности для образцов с разной концентрацией примесей.

Изменения в стехиометрическом составе не только меняют величину проводимости, но изменяют также и её температурный ход. Системы кривых, соответствующих образцам, отличающимся друг от друга различной концентрацией примесей, изображены на схематическом рис. 2. По мере увеличения примесей кривая сдвигается вверх и, вообще говоря, делается более пологой.

Заметим, что различные типы температурных кривых определяются не столько природой полупроводника, сколько условиями эксперимента. Один и тот же полупроводник может характеризоваться тем или иным температурным ходом электропроводности в зависимости от исследуемой области температур и от количества содержащихся в нём примесей.

3. Зависимость электропроводности полупроводников от внешнего электрического поля обнаруживается лишь при достаточно сильных полях. Ниже некоторого критического напряжения соблюдается закон Ома, т. е. электропроводность не зависит от поля. При переходе через это критическое напряжение закон Ома сменяется законом Пуля, и мы получаем быстрое (экспоненциальное) возрастание проводимости с полем. Соответствующие кривые схематически изображены на рис. 3. Различные кривые относятся к одному и тому же образцу, но исследованному при различных температурах. При повышении температуры кривая сдвигается вверх. В то же время при повышении температуры возрастание проводимости с полем делается более медленным.

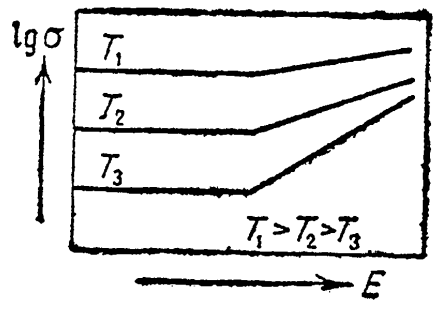


Рис. 3. Зависимость электропроводности от напряжённости поля (для одного и того же полупроводника при разных температурах).

4. Электропроводность полупроводников, как правило, очень чувствительна к освещению. Поглощение света обычно вызывает значительное увеличение проводимости, причём под влиянием освещения, вообще говоря, изменяется характер зависимости электропроводности от тех или иных факторов, так что световая проводимость обнаруживает иные закономерности, чем проводимость в темноте при прочих равных условиях.

Однако, наряду с этим, так называемым нормальным внутренним фотоэффектом, в ряде случаев поглощение света не вызывает сколько-нибудь заметного увеличения проводимости (фотоэлектрически неактивное поглощение света), а иногда даже приводит к уменьшению проводимости (аномальный фотоэффект).

Осуществление того или иного из этих возможных случаев определяется не только природой полупроводника и не только природой света (его интенсивностью и спектральным составом).

но зависит от ряда условий: от концентрации примесей, содержащихся в полупроводнике, от температуры, от напряжённости электрического поля, наложенного на полупроводник. Так, при повышении температуры и при увеличении поля нормальный фотоэффект в ряде случаев сменяется аномальным эффектом.

В. Металлы, полупроводники и диэлектрики. Теория металлов имеет своей задачей объяснение макроскопических свойств металлов с точки зрения строения кристалла вообще. Подобная же задача ставится перед теорией полупроводников. Обе теории, имея перед собой одну и ту же конечную цель, вынуждены, однако, исходить из разных позиций.

Теория металлов, как известно, базируется на представлении об идеальной кристаллической решётке. Это — такая решётка, структура которой обладает строгой пространственной периодичностью. В теории металлов обычно пренебрегают нарушениями этой периодичности, вызванными местными повреждениями. Такая идеализация характерна для теории металлов. Все основные свойства металлов могут быть получены на основе представлений об идеальной решётке.

С совершенно иным положением дел мы встречаемся, когда пытаемся построить теорию полупроводников. Если здесь оперировать с идеальной решёткой, то все основные свойства полупроводников остаются за рамками теории. Всякая попытка построить теорию полупроводников на базе таких представлений заведомо обречена на неудачу.

Действительно, полупроводники, как правило, представляют собой гетерополярные кристаллы с нарушенным стехиометрическим соотношением. Характером и степенью этого нарушения обусловливаются все специфические свойства полупроводников. Таким образом, полупроводники заведомо являются реальными кристаллами, ибо нельзя себе представить стехиометрических нарушений в рамках идеальной решётки. Если в теории металлов мы игнорируем всякого рода дефекты, присутствующие в решётке, то учёт этих дефектов и выяснение их роли являются центральным вопросом в теории полупроводников, поскольку не периодическая структура решётки, а именно локальные нарушения этой периодичности являются ответственными за все особенности полупроводников.

Таким образом, теория полупроводников базируется на представлении о реальной кристаллической решётке. Степень

отклонения реальной решётки от идеальной является решающей характеристикой полупроводника.

Следует ожидать, что все макроскопические свойства полупроводников могут быть извлечены из анализа представлений о реальной кристаллической решётке, подобно тому как из представлений об идеальной решётке извлекаются, в конечном счёте, все макроскопические свойства металлов.

Заметим, что, наряду с полупроводниками, мы имеем другой класс кристаллов, которые не являются полупроводниками, но которые, как и полупроводники, обладают гетерополярными решётками и в которых, как и в полупроводниках, стехиометрическое соотношение, как правило, не соблюдается. Это — щёлочно-галоидные кристаллы, являющиеся типичными диэлектриками. Хотя щёлочно-галоидные кристаллы по своим макроскопическим свойствам существенно отличаются от полупроводников, тем не менее между этими двумя классами кристаллов нет принципиального различия.

Это подтверждается тем фактом, что спектры поглощения полупроводников очень похожи на спектры поглощения щёлочно-галоидных кристаллов, будучи лишь сдвинуты по сравнению с последними в красную сторону спектра. Это означает, что как щёлочно-галоидным кристаллам, так и кристаллам полупроводников соответствует одна и та же схема энергетических уровней, и различие между этими двумя классами кристаллов сводится лишь к различию в масштабе, принятому на этой схеме. Все весьма специфические оптические свойства щёлочно-галоидных кристаллов, в настоящее время тщательно изученные, строго говоря, должны повторяться и в полупроводниках, но должны быть при этом перенесены из ультрафиолетовой и видимой частей спектра в красную и инфракрасную область, плохо доступную экспериментальному исследованию. С другой стороны, все электрические свойства полупроводников, теоретически говоря, должны наблюдаться и в щёлочно-галоидных кристаллах, но при значительно более высоких температурах, лежащих за пределами экспериментальных возможностей.

Можно думать, что как электрические свойства полупроводников, так и оптические свойства щёлочно-галоидных кристаллов могут быть извлечены из одной и той же энергетической схемы. Таким образом, теория полупроводников и теория щёлочно-галоидных кристаллов представляют собой по

существу всего лишь две разные стороны одной и той же теории. Существенные качественные различия между этими двумя классами кристаллов в конечном счёте имеют чисто количественное происхождение.

С. Содержание настоящей книги. В то время как металлы, с одной стороны, и диэлектрики (например, в лице своего типичного представителя — каменной соли), с другой стороны, являются предметом экспериментального исследования уже в течение многих десятилетий, полупроводники привлекают к себе внимание физиков сравнительно с недавних пор. Изучение полупроводников представляет двоякий интерес. Оно имеет несомненный физический интерес, так как полупроводники, как уже было сказано, составляют обширный класс кристаллов, характеризующихся рядом весьма специфических свойств. Изучение полупроводников имеет также практический интерес, поскольку в настоящее время полупроводники нашли себе применение в различных отраслях техники.

Сейчас мы имеем уже несколько сотен экспериментальных работ, посвящённых электрическим свойствам полупроводников. Эти работы предоставляют в распоряжение теоретика достаточно богатый экспериментальный материал. Однако теоретические работы в этой области ещё весьма немногочисленны. Отдельные экспериментальные факты и закономерности имеют ту или иную теоретическую интерпретацию. Но до сих пор ещё не построена общая теория полупроводников, подобная электронной теории металлов или кинетической теории газов, которая охватывала бы с некой единой точки зрения всё многообразие их макроскопических свойств. Среди работ по полупроводникам важное место занимают работы академика А. Ф. Иоффе и его школы. Нашими представлениями о полупроводниках мы в значительной степени обязаны этой школе.

В настоящее время существует несколько русских и иностранных экспериментальных обзоров по полупроводникам (правда, уже устаревших). Не существует, однако, систематического и доступного для экспериментаторов изложения теоретических вопросов, относящихся к этой области. Настоящая книга имеет своею целью восполнить этот пробел. Она распадается на две части:

К первой части относятся три первых главы. В этих главах развиваются представления о реальных кристаллических

решётках, частным случаем которых являются решётки, обладающие стехиометрическими нарушениями. Первая глава, носящая характер введения, построена в рамках чисто классических модельных представлений. В ней рассматривается структура реальной решётки и некоторые свойства такой решётки. Во второй главе излагается «зональная теория» твёрдого тела и некоторые связанные с этим вопросы теории металлов. Третья глава посвящена так называемым «локальным уровням», возникающим в энергетическом спектре реальной решётки.

Ко второй части относятся три последующие главы. В них с точки зрения общих представлений, развитых в предыдущих главах, рассматриваются макроскопические свойства полупроводников. В четвёртой главе исследуется влияние температуры и примесей на электропроводность полупроводников; в пятой главе — влияние электрического поля; в шестой главе — влияние освещения. Таким образом, книга ограничивается явлениями электропроводности и притом явлениями, протекающими в объёме. Поверхностные явления и явления на контактах в ней не рассматриваются.

Каждой из глав предпосылается сводка экспериментальных фактов. Однако книга не претендует на то, чтобы быть исчерпывающим экспериментальным обзором. Она носит теоретический характер, хотя и предназначена для физиков-экспериментаторов и инженеров, занимающихся электрическими свойствами твёрдых тел. Предполагается, что читатель знаком с элементами квантовой механики.

ГЛАВА I

ИДЕАЛЬНЫЕ И РЕАЛЬНЫЕ КРИСТАЛЛИЧЕСКИЕ РЕШЁТКИ

§ 1. Характеристика реальной решётки

А. Различные типы дефектов в решётке. Все макроскопические свойства кристаллов могут быть разделены на два класса. К одному классу принадлежат все те свойства, которые определяются периодической структурой решётки и для которых учёт дефектов, неизбежно присутствующих во всякой реальной решётке, играет роль незначительной поправки. К другому классу относятся все те свойства, которые, наоборот, определяются именно локальными нарушениями в периодической структуре решётки, для которых, следовательно, учёт дефектов имеет принципиальное значение. Свойства, принадлежащие к этому второму классу, обычно называют *структурно-чувствительными* свойствами. Электропроводность полупроводников относится именно к таким свойствам. Сюда относятся также оптические и фотоэлектрические свойства щёлочно-галоидных кристаллов, активность кристаллических катализаторов и многие другие свойства кристаллов.

Дефекты, присутствующие в реальной решётке, могут быть двух сортов: дефекты *макроскопического* и дефекты *микроскопического* характера.

К макроскопическим дефектам относятся трещины, коллоидные частицы постороннего вещества, наблюдающиеся, например, в окрашенных щёлочно-галоидных кристаллах, всякого рода макроскопические вкрапления и пр.

Что касается микроскопических дефектов, то отметим здесь следующие основные типы дефектов этого сорта:

1) Пустой узел, образованный в результате удаления из идеальной решётки атома или иона.

2) Собственный атом, или ион решётки, расположенный в междуузелье.

3) Ион в гетерополярной решётке, находящийся в нормальном положении, но несущий на себе аномальный заряд.

4) Чужеродный атом, помещённый в междуузелье.

5) Чужеродный атом, помещённый в узел, т. е. замещающий собственный атом решётки.

На рис. 4 схематически изображены эти пять типов микродефектов. Первые два из них можно назвать *механическими* дефектами. Такие дефекты в случае гомеополярной решётки не изменяют её химического состава, а в случае

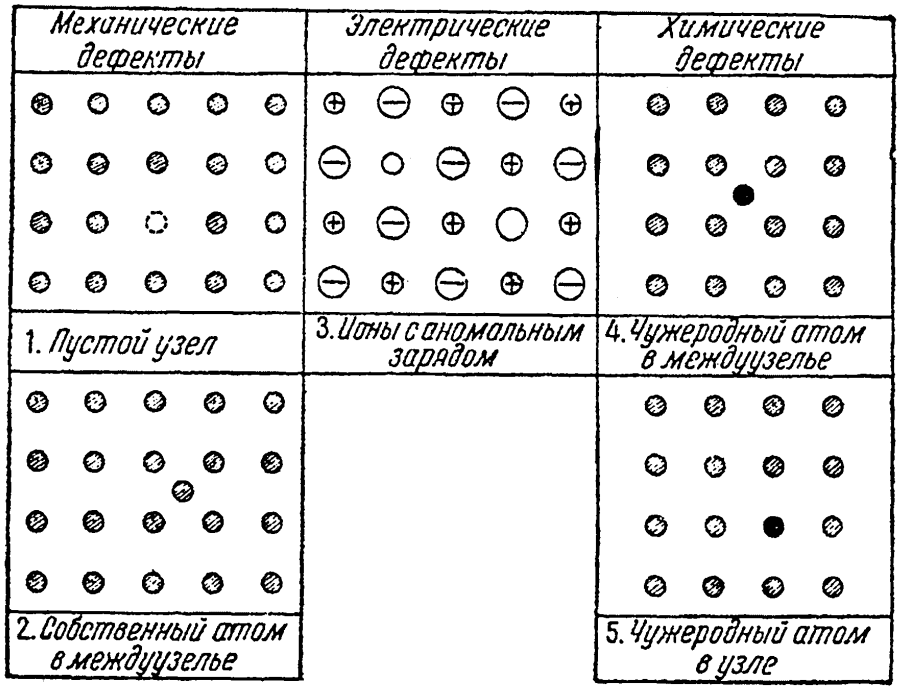


Рис. 4. Основные типы микродефектов.

гетерополярной решётки могут приводить к тому или иному нарушению стехиометрического соотношения. Дефекты третьего типа могут быть названы *электрическими* дефектами. Они также не влияют на химическую формулу решётки. Наконец, два последних типа микродефектов можно назвать *химическими* дефектами. Это есть «примесь» в узком смысле этого слова, искажающая химический состав кристалла.

Химические дефекты вводятся в решётку извне. Их возникновение обусловлено той или иной обработкой образца.

Механические дефекты могут возникнуть не только путём введения их извне, но и в результате нагревания решётки. Наконец, электрические дефекты не только могут быть введены извне или созданы теплом; они могут возникать в решётке также под влиянием освещения.

В дальнейшем мы ограничимся в основном рассмотрением решёток, обладающих механическими микродефектами. Мы оставим в стороне решётки с химическими примесями. Мы будем игнорировать также наличие макроскопических дефектов.

Ограничимся рассмотрением решёток типа M_mR_r. Структурными элементами такой решётки являются положительные ионы металла с зарядом $+p$ и отрицательные ионы металлоида с зарядом $-q$, где

$$\frac{q}{p}=\frac{m}{r}$$

(здесь p и q — абсолютная величина заряда, выраженная в единицах заряда электрона. Таким образом, p и q суть целые положительные числа). Мы имеем в нашей решётке два сорта узлов: *металлические узлы*, в которых помещаются ионы металла, и *металлоидные узлы*, в которых сидят ионы металлоида.

Если *все* металлические узлы решётки замещены металлическими ионами, а металлоидные узлы — металлоидными ионами и если при этом не существует ионов, расположенных *вне* этих узлов, то такую решётку мы будем называть *идеальной решёткой*.

Реальные решётки, с которыми мы будем иметь дело, отличаются от идеальной двумя чертами:

Во-первых, в реальной решётке не все узлы замещены ионами: в ней имеется некоторое число *пустых* металлических узлов и некоторое число *пустых* металлоидных узлов.

Во-вторых, в реальной решётке не все ионы помещаются в узлах: в ней имеется некоторое число *междуузельных* металлических ионов и некоторое число *междуузельных* металлоидных ионов. Эти междуузельные ионы мы будем называть *иррегулярными* ионами в отличие от *регулярных* ионов, помещающихся в предназначенных для них узлах.

Подавляющее большинство полупроводников обладает решётками именно такого типа.

В. Степень отклонения реальной решётки от идеальной и степень нарушения стехиометрического соотношения. При описании реальной кристаллической решётки мы будем пользоваться следующими обозначениями. Пусть мы имеем на единицу объёма:

X_M, X_R — число пустых узлов (соответственно: металлических и металлоидных);

Y_M, Y_R — число иррегулярных ионов (соответственно: металлических и металлоидных);

Z_M, Z_R — общее число узлов (соответственно: металлических и металлоидных);

N_M, N_R — общее число ионов (соответственно: металлических и металлоидных).

Эти числа следующим образом связаны между собой:

$$\begin{cases} N_M = Z_M - X_M + Y_M \\ N_R = Z_R - X_R + Y_R \end{cases} \tag{1,1}$$

причём

$$\begin{cases} X_M \ll Z_M, & Y_M \ll Z_M \\ X_R \ll Z_R, & Y_R \ll Z_R \end{cases} \tag{1,2}$$

Кроме того,

$$\frac{Z_M}{Z_R} = \frac{m}{r} = \frac{q}{p}. \tag{1,3}$$

Реальная решётка отличается от идеальной в той или иной степени в зависимости от того, каково соотношение между числами

$$X_M, X_R, Y_M, Y_R, Z_M, Z_R.$$

Соотношение между этими числами характеризует также и то, насколько сильно нарушено в нашей решётке стехиометрическое соотношение.

Степень отклонения реальной решётки от идеальной, а также *степень отклонения от стехиометрического соотношения* представляют собой две независимых характеристики, свойственные каждой решётке. Эти характеристики обозначим соответственно через μ и ν и определим их следующим образом:

$$\mu = \left(\frac{Y_M}{Z_M} + \frac{X_R}{Z_R}\right) + \left(\frac{X_M}{Z_M} + \frac{Y_R}{Z_R}\right), \tag{1,4}$$

$$\nu = \left(\frac{Y_M}{Z_M} + \frac{X_R}{Z_R}\right) - \left(\frac{X_M}{Z_M} + \frac{Y_R}{Z_R}\right), \tag{1,5}$$

где, согласно (1,2):

$$\begin{cases} 0 \leqslant \mu \ll 1, \\ 0 \leqslant |\nu| \ll 1. \end{cases}$$

Рассмотрим подробнее каждую из этих характеристик:

1) Реальная решётка получается из идеальной, коль скоро в этой последней нарушается свойственный ей порядок. Каждый пустой узел и каждый междуузельный ион представляют собой нарушение порядка. Абсолютное число таких нарушений выражается через

$$X_M,\ X_R,\ Y_M,\ Y_R.$$

Степень беспорядка может быть охарактеризована относительными числами:

$$\frac{X_M}{Z_M},\ \frac{X_R}{Z_R},\ \frac{Y_M}{Z_M},\ \frac{Y_R}{Z_R}.$$

Сумма μ всех этих чисел

$$\mu = \frac{X_M}{Z_M} + \frac{X_R}{Z_R} + \frac{Y_M}{Z_M} + \frac{Y_R}{Z_R}$$

показывает, насколько сильно нарушен порядок в нашей решётке.

2) Стехиометрическим соотношением называется, как известно, условие

$$\frac{N_M}{N_R} = \frac{m}{r}, \tag{1,6}$$

которое, на основании (1,3), можно переписать ещё так:

$$\frac{N_M}{Z_M} - \frac{N_R}{Z_R} = 0. \tag{1,7}$$

Это условие заведомо выполняется в случае идеальной решётки, для которой

$$X_M = Y_M = 0 \text{ и, следовательно, } N_M = Z_M,$$
$$X_R = Y_R = 0 \text{ и, следовательно, } N_R = Z_R.$$

В случае реальной решётки условие (1,7) может выполняться или же может быть нарушено. Разность

$$\nu = \frac{N_M}{Z_M} - \frac{N_R}{Z_R}.$$

характеризует степень отклонения от стехиометрического со-

отношения (1,7). Подставляя сюда (1,1), мы получаем для ν выражение (1,5).

В зависимости от того, будет ли ν положительным или отрицательным, мы будем иметь решётку со стехиометрическим избытком или недостатком металла. Избыток металла может быть осуществлён за счёт введения в решётку дополнительных ионов металла или за счёт удаления из неё ионов металлоида. Стехиометрический недостаток металла может быть осуществлён за счёт удаления ионов металла или за счёт введения ионов металлоида. Имеем:

Для решёток с соблюдённым стехиометрическим соотношением: $\frac{N_M}{N_R}=\frac{m}{r}$ и, следовательно, $\nu=0$	Для решёток с избытком металла (недостатком металлоида): $\frac{N_M}{N_R}>\frac{m}{r}$ и, следовательно, $\nu>0$	Для решёток с недостатком металла (избытком металлоида): $\frac{N_M}{N_R}<\frac{m}{r}$ и, следовательно, $\nu<0$

Рассмотрим два крайних случая: (1) решётки, содержащие пустые узлы, но не содержащие междуузельных ионов, для которых, следовательно, $Y_M=Y_R=0$; (2) решётки, содержащие междуузельные ионы, но не содержащие пустых узлов, для которых, следовательно, $X_M=X_R=0$.

Заметим, что решётки первого и второго типов в стехиометрическом отношении эквивалентны друг другу, если между числами X_M, X_R, характеризующими решётку первого типа, и числами Y_M, Y_R, характеризующими решётку второго типа, выполняются соотношения:

$$\begin{cases} \frac{Y_M}{m}=\frac{X_R}{r}, \\ \frac{Y_R}{r}=\frac{X_M}{m}. \end{cases} \tag{1,8}$$

Действительно, из (1,4) и (1,5) имеем:

$$\begin{cases} \frac{Y_M}{m}+\frac{X_R}{r}=\frac{1}{2}Z(\mu+\nu), \\ \frac{X_M}{m}+\frac{Y_R}{r}=\frac{1}{2}Z(\mu-\nu), \end{cases} \tag{1,9}$$

где

$$Z = \frac{Z_M}{m} = \frac{Z_R}{r}$$

— общее число молекул $M_m R_r$ в единице объёма идеальной решётки. Из (1,9) вытекает условие (1,8).

Параметры μ и ν играют важную роль. Ими определяются все основные свойства реальных решёток. Мы убедимся в дальнейшем, что параметрами μ и ν определяется температурный ход электропроводности в полупроводниках: варьируя эти параметры, получаем различные типы температурных кривых. Мы убедимся также, что от этих параметров зависит фотоэлектрическая активность полупроводников: придавая этим параметрам различные значения, можем получить положительный или отрицательный фотоэффект или вообще отсутствие фотоэффекта. Различные полупроводники (более того, различные образцы одного и того же полупроводника), вообще говоря, отличаются значениями μ и ν. Именно поэтому электрические и фотоэлектрические свойства различных образцов оказываются различными. Факт — хорошо известный экспериментаторам.

С. Тепловые нарушения в решётке. Значения параметров μ и ν для каждой данной решётки определяются прежде всего биографией решётки, т. е. зависят от того, каким способом был изготовлен исследуемый образец и каким воздействиям он подвергался на продолжении всей своей предшествующей жизни. Значение параметра μ, кроме того, зависит ещё от температуры. Действительно, каждое из чисел X_M, X_R, Y_M, Y_R, которые входят в μ и в ν, может быть представлено как сумма двух членов, из которых первый не зависит от температуры, а второй является функцией температуры:

$$\left.\begin{aligned} X_M &= X_M^0 + Z_M e^{-\frac{U_M}{kT}}, \\ X_R &= X_R^0 + Z_R e^{-\frac{U_R}{kT}}, \\ Y_M &= Y_M^0 + Z_M e^{-\frac{U_M}{kT}}, \\ Y_R &= Y_R^0 + Z_R e^{-\frac{U_R}{kT}}. \end{aligned}\right\} \qquad (1,10)$$

Здесь члены

$$X_M^0, X_R^0, Y_M^0, Y_R^0$$

характеризуют число повреждений в решётке, которыми она обладает уже при нуле температуры, т. е. число повреждений, данных решётке, так сказать, от рождения. Эти числа зависят от технологии образца. Члены

$$Z_M e^{-\frac{U_M}{kT}}, \quad Z_R e^{-\frac{U_R}{kT}}$$

характеризуют число повреждений, возникающих в решётке под действием температуры. Действительно, при нагревании регулярные ионы диссоциируют из узлов в междуузелья, образуя, таким образом, иррегулярные ионы и пустые узлы. Здесь

U_M — работа диссоциации металлического иона,

U_R — работа диссоциации металлоидного иона.

Подставляя (1,10) в (1,4) и в (1,5), мы получаем:

$$\left\{\begin{aligned} \mu &= \left(\frac{Y_M^0}{Z_M}+\frac{X_R}{Z_R}\right)+\left(\frac{X_M^0}{Z_M}+\frac{Y_R^0}{Z_R}\right)+2\left(e^{-\frac{U_M}{kT}}+e^{-\frac{U_R}{kT}}\right)= \\ &= \mu_0+2\left(e^{-\frac{U_M}{kT}}+e^{-\frac{U_R}{kT}}\right). \qquad (1,11) \\ \nu &= \left(\frac{Y_M^0}{Z_M}+\frac{X_R^0}{Z_R}\right)-\left(\frac{X_M^0}{Z_M}+\frac{Y_R^0}{Z_R}\right)=\nu_0. \end{aligned}\right.$$

Таким образом, нагревание вызывает дополнительный беспорядок в решётке. При повышении температуры степень отклонения реальной решётки от идеальной возрастает, а степень отклонения от стехиометрического соотношения остаётся, конечно, неизменной (если при этом не происходит испарения ионов в атмосферу).

Мы в дальнейшем будем пренебрегать тепловыми нарушениями. Иначе говоря, мы будем считать, что нарушения, вызванные нагреванием, малы по сравнению с теми нарушениями, которые присутствуют в решётке с самого начала:

$$\left.\begin{aligned} X_M &\approx X_M^0, \quad Y_M \approx Y_M^0, \\ X_R &\approx X_R^0, \quad Y_R \approx Y_R^0. \end{aligned}\right\}$$

Таким образом, будем исходить из предположения, что

$$\mu_0 \gg 2\left(e^{-\frac{U_M}{kT}} + e^{-\frac{U_R}{kT}}\right). \qquad (1,12)$$

Убедимся в том, что для всех полупроводников, с которыми мы фактически имеем дело, условие (1,12) действительно соблюдается.

Количество «примеси», содержащейся в образцах, подвергаемых экспериментальному исследованию, как правило, значительно превышает 10^{12} атомов на кубический сантиметр. Поэтому, полагая в (1,11):

$$\begin{cases} X_M^0 \approx X_R^0 \approx Y_M^0 \approx Y_R^0 \gg 10^{12} см^{-3}, \\ Z_M \approx Z_R \approx 10^{22} см^{-3}, \end{cases}$$

получаем

$$\mu_0 \gg 10^{-10}. \qquad (1,13)$$

С другой стороны, все экспериментальные данные по электропроводности полупроводников относятся к температурам, не достигающим 1 000° К. Поэтому, полагая в (1,11):

$$\begin{cases} T < 10^3 \text{K}, \\ U_R \approx U_M \approx 2\,\text{eV}, \end{cases}$$

получаем

$$e^{-\frac{U_M}{kT}} + e^{-\frac{U_R}{kT}} < 10^{-10}. \qquad (1,14)$$

Из сравнения (1,13) с (1,14) вытекает справедливость условия (1,12).

Таким образом, тепловые нарушения начинают сказываться при значительно меньших концентрациях «примеси» или при значительно более высоких температурах, чем те концентрации и те температуры, с которыми мы имеем дело в опытах с полупроводниками.

Если в полупроводниках число «тепловых» нарушений может считаться малым по сравнению с числом «врожденных» повреждений, то в щёлочно-галоидных кристаллах мы имеем дело, повидимому, с противоположным случаем. Об этом свидетельствует хорошо известный экспериментальный факт: экспоненциальное возрастание числа центров окрашивания с температурой. При построении теории щёлочно-галоидных

кристаллов, которые обладают, повидимому, более упорядоченными решётками, чем решётки полупроводников, условие (1,12) уже нельзя считать приемлемым.

D. Компенсирующие электроны и дырки. Идеальная решётка превращается в реальную в результате удаления регулярных ионов или введения иррегулярных ионов. Заметим, что эта операция обязательно должна сопровождаться введением или удалением из решётки электронов. Этот второй процесс должен сопутствовать первому для того, чтобы электрический заряд, образующийся вследствие введения или удаления ионов, был бы скомпенсирован, т. е. для того, чтобы была сохранена электрическая нейтральность решётки. Вместо того чтобы говорить об *удалении электрона*, можно говорить о *введении дырки*, и наоборот. Наличие дырки означает отсутствие электрона. Как мы убедимся в дальнейшем, такой способ выражаться оказывается весьма удобным.

Итак, введение или удаление металлического иона должно сопровождаться введением p электронов или p дырок соответственно. Введение или удаление металлоидного иона должно сопровождаться, соответственно, введением q дырок или q электронов. Решётка, характеризующаяся заданными значениями чисел X_M, X_R, Y_M, Y_R, должна содержать в себе компенсирующий заряд, составляющийся из n_1 электронов и n_2 дырок, где

$$\begin{cases} n_1 = pY_M + qX_R, \\ n_2 = pX_M + qY_R. \end{cases} \tag{1,15}$$

Эти электроны и дырки, присутствующие во всякой реальной решётке для соблюдения электрического баланса, вообще говоря, не обязаны (как мы увидим ниже) рекомбинировать друг с другом. В некоторых решётках такая рекомбинация энергетически выгодна, в других — невыгодна, т. е. приводит к повышению общей энергии системы. Эти электроны и дырки играют весьма важную роль. Они составляют тот резерв, за счёт которого при надлежащих условиях возникает электропроводность решётки.

Общий запас электронов и дырок зависит от того, насколько сильно наша решётка отличается от идеальной (т. е. зависит от параметра μ). Преобладание же электронов или преобладание дырок в этом общем запасе выражено тем более резко, чем сильнее нарушено стехиометрическое соотношение

(т. е. зависит от параметра ν). Действительно, на основании (1,3), (1,4) и (1,5):

$$\begin{cases} n_1 + n_2 = Q\mu, \\ n_1 - n_2 = Q\nu, \end{cases}$$

где $Q = pZ_M = qZ_R$ — абсолютная величина заряда всех положительных или всех отрицательных ионов идеальной решётки. Мы видим, что в решётках с соблюдённым стехиометрическим соотношением ($\nu = 0$) электронов столько же, сколько и дырок (т. е. $n_1 = n_2$); в решётках со стехиометрическим избытком металла ($\nu > 0$) мы имеем избыток электронов (т. е. $n_1 > n_2$); в решётках со стехиометрическим недостатком металла ($\nu < 0$) мы имеем недостаток электронов, т. е. избыток дырок (т. е. $n_1 < n_2$).

Компенсирующие электроны и дырки, введённые в решётку и рассаженные по её ионам, приводят к образованию *аномальных ионов*, т. е. ионов с аномальным зарядом. Так, если все n_1 электронов и n_2 дырок размещены по одному на каждый ион, то мы будем иметь в нашей решётке наряду с пустыми узлами и междуузельными ионами, ещё n_1 аномальных ионов с избыточным отрицательным зарядом и n_2 аномальных ионов с избыточным положительным зарядом, например, n_1 металлических ионов с зарядом $+p-1$ и n_2 металлоидных ионов с зарядом $-q+1$. Таким образом, гетерополярная решётка, обладающая *механическими* повреждениями, неизбежно должна содержать в себе также и *электрические* повреждения, обеспечивающие её электрическую нейтральность.

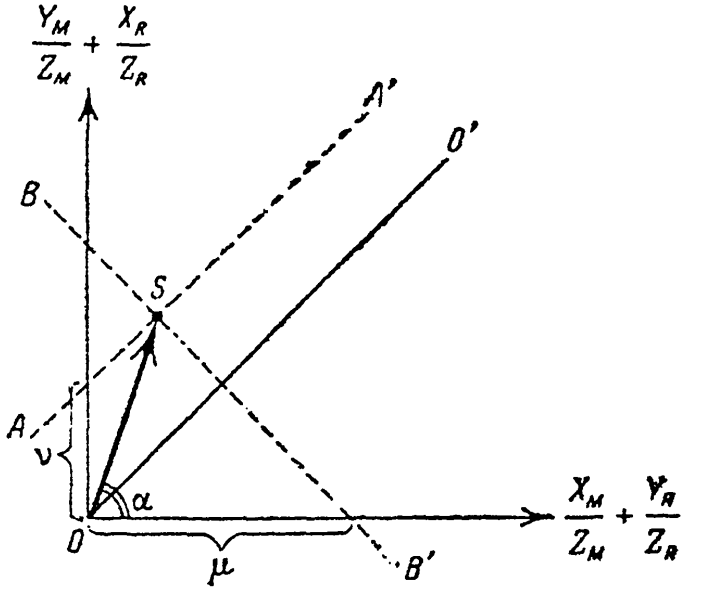


Рис. 5. Диаграмма состояний реальной решётки.

Е. Диаграмма состояний реальной решётки. Различные состояния решётки, отличающиеся значениями параметров μ и ν, могут быть изображены на диаграмме состояний. Такая диаграмма состояний представлена на рис. 5.

По осям откладываются величины

$$\left(\frac{Y_M}{Z_M}+\frac{X_R}{Z_R}\right) \text{ и } \left(\frac{X_M}{Z_M}+\frac{Y_R}{Z_R}\right),$$

характеризующие степень беспорядка. Каждая точка внутри квадрата изображает состояние решётки. Точке O соответствует идеальная решётка. Точкам, расположенным на диагонали OO', соответствуют решётки с соблюдённым стехиометрическим соотношением. Точки, лежащие над этой диагональю, изображают решётки с избытком металла, а точки, лежащие под диагональю, — решётки с недостатком металла. Предположим, что наша решётка изображается на диаграмме точкой S. Значения параметров μ и ν могут быть определены из диаграммы так, как это показано на рисунке. Угол α, который составляет радиус-вектор OS с осью абсцисс, даёт нам соотношение между числом электронов и дырок, содержащихся в нашей решётке. Действительно,

$$\operatorname{tg}\alpha=\frac{\frac{Y_M}{Z_M}+\frac{X_R}{Z_R}}{\frac{X_M}{Z_M}+\frac{Y_R}{Z_R}}=\frac{pY_M+qX_R}{pX_M+qY_R}=\frac{n_1}{n_2}.$$

Если решётка подвергается воздействиям, в результате которых изменяется её стехиометрический состав, т. е. изменяется ν, а μ при этом остаётся неизменным, то точка S передвигается по прямой BB'. Если, наоборот, ν остаётся постоянным, а μ изменяется, то точка S перемещается по прямой AA'. Таким образом, при повышении температуры точка S перемещается вверх по прямой AA'.

§ 2. Энергия решётки

А. Энергия идеальной решётки. Мы будем называть энергией решётки ту работу, которую надо затратить для того, чтобы образовать нашу решётку из изолированных нейтральных *атомов*, находящихся в нормальном состоянии. Энергию последних мы принимаем, следовательно, за нуль. Нам будет удобно определять энергию именно таким образом в отличие от обычного определения, согласно которому под энергией решётки понимают ту работу, которая требуется для построения нашей решётки из *ионов*, разведённых первоначально на бесконечность. Таким образом, в отличие

от обычного определения, в наше определение энергии включается энергия ионизации. Обозначим энергию решётки (на единицу объёма) через W. Если ионы считать точечными, то энергия W складывается из двух членов: из кулоновской энергии взаимодействия между ионами и из энергии ионизации, т. е. энергии превращения нейтральных атомов в соответствующие ионы.

Примем следующее обозначение. Пусть $J(Q', Q)$ — энергия, необходимая для того чтобы ион с зарядом Q превратить в ион с зарядом Q', где Q и Q' — положительные или отрицательные числа. Очевидно, если $Q' > Q$, то наша операция представляет собой удаление электронов из иона; если же $Q' < Q$, то, наоборот, это значит, что мы вводим электроны в наш ион. Работа $J(Q', Q)$ может быть как положительной, так и отрицательной, т. е. изменение заряда иона с Q на Q' может сопровождаться затратой энергии или выделением энергии. В частности, энергия $J(+1,0)$ называется *ионизационным потенциалом* атома ($Q=0$, $Q'=+1$), а энергия $J(-1,0)$ называется *энергией сродства* атома к электрону ($Q=0$, $Q'=-1$). Если $J(-1,0)>0$, то атом называется электроположительным. Если же $J(-1,0)<0$, то атом называется электроотрицательным; про такой атом говорят, что он обладает сродством к электрону. Заметим, что функция $J(Q', Q)$ обладает следующими свойствами:

$$\begin{aligned} &1)\ J(Q', Q) = -J(Q, Q'), \\ &2)\ J(Q', Q) = J(Q', Q'') + J(Q'', Q). \end{aligned} \tag{2,1}$$

Пользуясь принятым обозначением, мы можем написать:

$$W = \frac{1}{V} \sum_g \left\{ J_g(Q_g, 0) + \frac{1}{2} Q_g e \Phi_g \right\}, \tag{2,2}$$

где g — номер иона;

Q_g — заряд g-го иона, выраженный в единицах заряда электрона ($e = 4{,}78 \cdot 10^{-10}$ *CGSE*); так что число Q_g есть *целое* положительное или отрицательное число;

Φ_g — потенциал, созданный всеми ионами (за исключением g-го) в точке, где находится g-ый ион;

e — абсолютная величина заряда электрона.

Очевидно:

$$\begin{cases} \sum\limits_g Q_g = 0, \\ \Phi_g = \sum\limits_{g'} \dfrac{Q_{g'}}{R_{gg'}} e, \end{cases} \tag{2,3}$$

где $R_{gg'}$ — расстояние между g-ым и g'-ым ионами.

Выражение для энергии W может быть записано ещё в такой форме:

$$W = \frac{1}{V} \sum_g J_g(Q_g, 0) - \frac{Q_0^2 e^2}{a} ZM, \tag{2,4}$$

где:

Q_0 — наименьший заряд иона (т. е. наименьшее по абсолютной величине из чисел Q_g);

a — наикратчайшее расстояние между двумя ионами (т. е. наименьшее из чисел $R_{gg'}$);

Z — число молекул в единице объёма;

M — коэффициент Маделунга ($M > 0$).

Если мы пожелаем учесть конечные размеры ионов, то это может быть сделано путём введения в (2,2) или в (2,4) дополнительных поправочных членов, выражающих собой, во-первых, энергию сил отталкивания, обусловленную взаимодействием электронных оболочек, и, во-вторых, энергию деформации (поляризации) отдельных ионов. Мы будем пренебрегать этими поправочными членами. Мы имеем право это делать при рассмотрении целого ряда вопросов.

Мы ограничимся в дальнейшем рассмотрением решёток типа $M_m R_r$, состоящих лишь из двух сортов ионов, и притом координационных решёток, т. е. таких, у которых все ионы одного и того же сорта имеют одно и то же окружение. Координационная решётка, очевидно, при параллельном переносе начала координат из одного металлического иона в любой другой металлический ион, или из одного металлоидного иона в любой другой металлоидный ион, остаётся подобной сама себе. Для таких решёток мы имеем:

$$\left.\begin{aligned} Q_g &= \begin{cases} = Q_M = +p \\ = Q_R = -q \end{cases} \\ \Phi_g &= \begin{cases} = \Phi_M \\ = \Phi_R \end{cases} \\ J_g(Q_g, 0) &= \begin{cases} = J_M(+p, 0) \\ = J_R(-q, 0) \end{cases} \end{aligned}\right\}, \tag{2,5}$$

причём, так как

$$\frac{q}{p}=\frac{m}{r},$$

то

$$\begin{cases} q=\eta m, \\ p=\eta r, \end{cases}$$

где η — целое положительное число ($\eta=0, 1, 2, \ldots$), характеризующее зарядность наших ионов.

Для таких решёток выражение (2,2) может быть переписано так:

$$\boxed{W=Z\left\{mJ_M(+\eta r,0)+rJ_R(-\eta m,0)+\frac{1}{2}\eta mre(\Phi_M-\Phi_R)\right\}} \quad (2,6)$$

Из сравнения (2,6) с (2,4) мы имеем:

$$\Phi_R-\Phi_M=2k\frac{\eta eM}{a}, \quad (2,7)$$

где

$$k=\begin{cases} =\frac{m}{r}\ldots, & \text{если } m<r, \\ =\frac{r}{m}\ldots, & \text{если } r<m. \end{cases}$$

В частности, для решёток типа MR (как, например, NaCl, CuO, PbS), для которых $m=r=1$, имеем:

$$\Phi_R=-\Phi_M=\frac{\eta eM}{a}. \quad (2,8)$$

В. Структура решётки. Для того чтобы решётка находилась в равновесии, её энергия W должна быть минимальна. Наличие минимума у функции W обеспечивается добавочными членами, которыми мы пренебрегали и которые отсутствуют в формулах (2,2), (2,4) и (2,6). Эти добавочные члены выражают энергию сил отталкивания. Условием минимума W фиксируется структура решётки.

В частности, из условия минимума может быть определён 1) параметр a (если задано η и все прочие параметры),

2) параметр η (если задано a и все прочие параметры).

Рассмотрим подробнее обе эти задачи.

1. Параметр a характеризует степень сближения ионов. На рис. 6 кривая ABC изображает энергию W решётки как функцию $\frac{1}{a}$. Кривая ABC есть результат сложения двух монотонных кривых OO' и AA'. Кривая OO' представляет собой энергию сил отталкивания. Прямая AA' соответствует кулоновской энергии и энергии ионизации. Эта часть энергии выражается формулой (2,6). Результирующая кривая ABC обладает минимумом. Этим минимумом и фиксируется значение a.

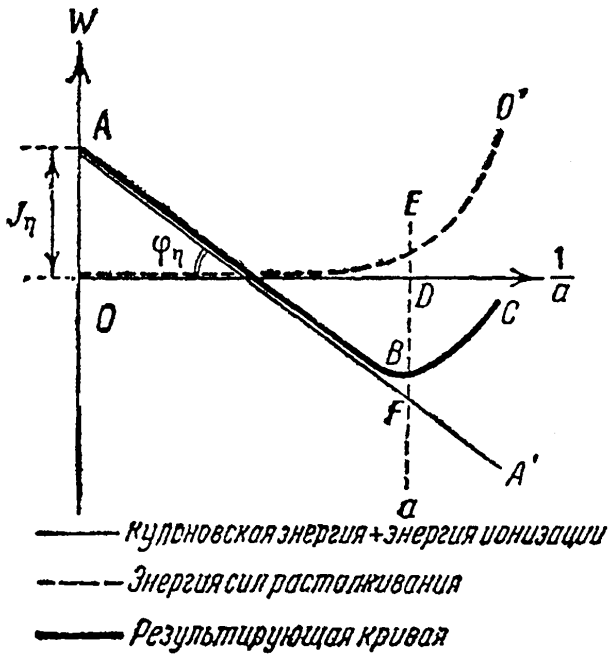


Рис. 6. Энергия идеальной решётки как функция $1/a$.

Из рис. 6, между прочим, видно, что при вычислении энергии решётки, той частью энергии, которая обусловливается силами отталкивания, можно с полным правом пренебрегать до тех пор, пока $DE = FB \ll DF$.

На рис. 6 приняты следующие обозначения:

$$J_\eta = Z\{mJ_M(+\eta r, 0) + rJ_R(-\eta m, 0)\},$$

$$\operatorname{tg}\varphi_\eta = \begin{cases} = Z\cdot\eta^2 e^2 M m^2 \ldots, & \text{если } m < r, \\ = Z\cdot\eta^2 e^2 M r^2 \ldots, & \text{если } r < m. \end{cases}$$

2. Параметр η определяет числа p и q, т. е. заряды тех ионов, из которых построена решётка. Таким образом, числа p и q строго фиксированы, коль скоро задана химическая природа решётки и её геометрия. На рис. 7 изображена энергия W как функция $\frac{1}{a}$ для разных η (в пренебрежении энергией сил отталкивания). Мы имеем систему пересекающихся прямых; каждому значению η соответствует своя прямая. Энергия решётки изображается огибающей ломаной кривой (жирная кривая на рисунке). Точки перелома соответствуют

тем значениям a, при которых меняется структура решётки, т. е. изменяются заряды составляющих её ионов. Из рис. 7 видно, как по заданному значению a однозначно определяется значение η. (Обозначения на рис. 7 те же, что и на рис. 6.)

Вычисление показывает, что решётка NaCl построена из ионов Na^+ и Cl^- (т. е. $\eta = 1$), а решётка ZnO — из ионов Zn^{++} и O^{--} (т. е. $\eta = 2$). Если, однако, решётку NaCl подвергнуть всестороннему растяжению, а решётку ZnO — всестороннему сжатию (так, чтобы при этом решётки оставались подобными самим себе), то решётки испытают изменения в структуре:

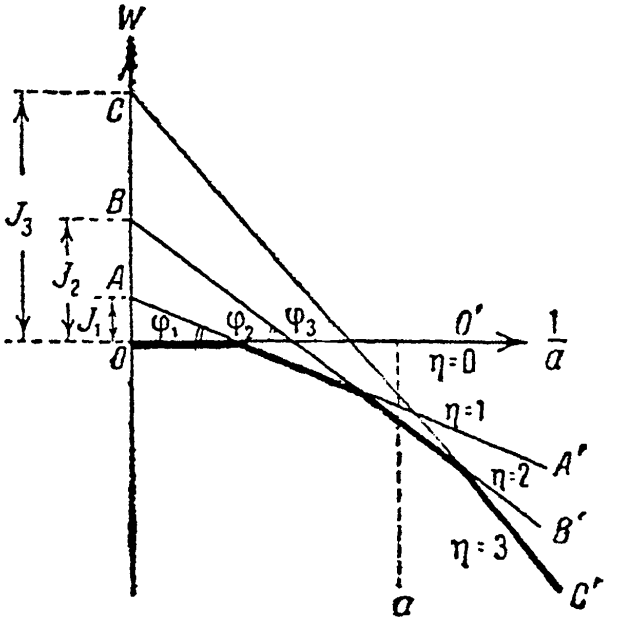


Рис. 7. Энергия идеальной решётки как функция $1/a$ при разных η.
$OO' \ldots \eta = 0$,
$AA' \ldots \eta = 1$,
$BB' \ldots \eta = 2$,
$CC' \ldots \eta = 3$.

$$Na^+ + Cl^- \longrightarrow Na^{++} + Cl^{--} \ldots \text{ (при уменьшении } a),$$
$$Zn^{++} + O^{--} \longrightarrow Zn^+ + O^- \ldots \text{ (при увеличении } a).$$

C. Электроны и дырки в идеальной решётке. Представим себе, что в нашу идеальную решётку $M_m R_r$ мы вводим n электронов и столько же дырок. Разместим эти электроны и дырки по ионам решётки (по одному на каждый ион). Эту операцию можно рассматривать как пересаживание электронов внутри решётки с одних ионов на другие. Такое пересаживание можно осуществить, например, освещая решётку светом соответствующей частоты. В результате нашей операции каждый ион решётки приобретает дополнительный заряд ΔQ_g, где очевидно:

$$\Delta Q_g = \begin{cases} +1, \\ -1, \\ 0, \end{cases}$$

причём

$$\sum_g \Delta Q_g = 0.$$

Ионы, для которых $\Delta Q_g \neq 0$, мы будем называть *аномальными* ионами в отличие от *нормальных* ионов, для которых $\Delta Q_g = 0$. Мы имеем два сорта аномальных ионов: n ионов с избыточным положительным зарядом и столько же ионов с избыточным отрицательным зарядом.

Создание в идеальной решётке n пар аномальных ионов представляет собою операцию, требующую для своего осуществления некоторой затраты энергии. Обозначим через W энергию, необходимую для этой операции. На основании (2,1), (2,2) и (2,3) мы будем иметь:

$$W = \sum_g \left\{ J_g(Q_g + \Delta Q_g, Q_g) + \Delta Q_g e \left(\Phi_g + \frac{1}{2}\varphi_g \right) \right\}, \quad (2,9)$$

где φ_g — дополнительный потенциал в точке g, созданный аномальными ионами:

$$\varphi_g = \sum_{g'} \frac{\Delta Q_{g'}}{R_{gg'}} e.$$

В формуле (2,9) члены

$$\frac{1}{2}\sum_g \Delta Q_g e \varphi_g = \frac{1}{2}\sum_g \sum_{g'} \frac{\Delta Q_g \Delta Q_{g'}}{R_{gg'}} e^2 \quad (2,10)$$

выражают собой дополнительную энергию взаимодействия между аномальными ионами, возникающую вследствие наличия на них избыточного заряда. Из (2,10) мы видим, что аномальные ионы с противоположными дополнительными зарядами притягиваются друг к другу ($\Delta Q_g \Delta Q_{g'} = -1$), в то время как аномальные ионы с одноимёнными дополнительными зарядами друг от друга отталкиваются ($\Delta Q_g \Delta Q_{g'} = +1$). Здесь правильнее было бы сказать, что притягиваются или отталкиваются не аномальные ионы, а аномальные *состояния* ионов.

Если аномальные ионы разведены достаточно далеко друг от друга так, что их взаимодействием можно пренебречь, то в выражении (2,9) следует положить $\varphi_g = 0$. В этом случае аномальные состояния могут свободно блуждать по решётке, передаваясь от одного иона к соседнему одноимённому иону. Такие невзаимодействующие аномальные ионы мы будем называть *свободными* аномальными ионами. Ими обеспечивается проводимость в идеальной решётке.

Из общей формулы (2,9) может быть получено выражение для энергии $W_{gg'}$, затрачиваемой при пересаживании электрона в идеальной решётке с g-го иона на g'-й ион. Мы имеем:

$$W_{gg'} = J_g(Q_g + 1, Q_g) + J_{g'}(Q_{g'} - 1, Q_{g'}) + \\ + e(\Phi_g - \Phi_{g'}) - \frac{e^2}{R_{gg'}}. \qquad (2,11)$$

Это есть энергия образования одной пары аномальных ионов в идеальной решётке. Каждый из двух ионов, участвующих в реакции (g-й ион, отдающий электрон, и g'-й ион, принимающий на себя электрон), может быть по своей природе как металлическим, так и металлоидным. Таким образом, возможны четыре комбинации:

$$\begin{aligned} &W_{gg'} = W_{RM},\ W_{gg'} = W_{MM},\ W_{gg'} = W_{RR},\ W_{gg'} = W_{MR}, \\ &W_{RM} = J_R(-q+1, -q) + J_M(+p-1, +p) + A - B, \\ &W_{MM} = J_M(+p+1, +p) + J_M(+p-1, +p) - B, \\ &W_{RR} = J_R(-q+1, -q) + J_R(-q-1, -q) - B, \\ &W_{MR} = J_M(+p+1, +p) + J_R(-q-1, -q) - A - B, \end{aligned} \qquad (2,12)$$

где $$A = 2\alpha\eta M\frac{e}{a}, \quad B = \frac{e^2}{R},$$

и где R — расстояние между соответствующими ионами. Числа

$$W_{RM},\ W_{MM},\ W_{RR},\ W_{MR}$$

можно, следовательно, трактовать как энергетические эквиваленты следующих четырёх возможных реакций:

$$\left.\begin{aligned} &\text{I)}\ \ M^{+p} + R^{-q} = M^{+p-1} + R^{-q+1} - W_{RM}, \\ &\text{II)}\ \ M^{+p} + M^{+p} = M^{+p-1} + M^{+p+1} - W_{MM}, \\ &\text{III)}\ \ R^{-q} + R^{-q} = R^{-q-1} + R^{-q+1} - W_{RR}, \\ &\text{IV)}\ \ R^{-q} + M^{+p} = R^{-q-1} + M^{+p+1} - W_{RM}. \end{aligned}\right\} \qquad (2,13)$$

Заметим, что эти формулы остаются справедливыми лишь до тех пор, пока мы пренебрегаем деформацией решётки, возникающей вокруг аномальных ионов. В действительности, при введении в решётку электрона или дырки, нормальная конфигурация ионов несколько нарушается. Кроме того, каждый аномальный ион вызывает поляризацию окружающих ионов. Эти эффекты мы здесь не учитываем.

D. Электроны и дырки в реальной решётке. Обратимся теперь к реальной решётке, содержащей в качестве дефектов пустые узлы или междуузельные ионы. Перенумеруем все узлы решётки. Перенумеруем также дефекты. Пусть

$$\begin{cases} g, g' \text{ — номера узлов,} \\ h, h' \text{ — номера дефектов,} \\ Q_g \text{ — заряд } g\text{-го иона,} \\ Q_h \text{ — заряд } h\text{-го дефекта.} \end{cases}$$

Очевидно:

$$\left.\begin{aligned} Q_g &= \begin{cases} +p & \text{в случае металлического иона,} \\ -q & \text{в случае металлоидного иона.} \end{cases} \\ Q_h &= \begin{cases} +p & \text{в случае металлического междуузельного иона,} \\ -p & \text{в случае металлического пустого узла,} \\ -q & \text{в случае металлоидного междуузельного иона,} \\ +q & \text{в случае металлоидного пустого узла.} \end{cases} \end{aligned}\right\} \quad (2,14)$$

Представим себе теперь, что в нашу решётку мы вводим n_1 электронов и n_2 дырок и тем или иным способом рассаживаем их по ионам решётки. В результате этой операции каждый ион приобретает дополнительный заряд ΔQ_g, где ΔQ_g может быть положительным или отрицательным, или равным нулю. Электрическая нейтральность решётки выражается условием:

$$n_2 - n_1 = \sum_{g \neq h} \Delta Q_g = -\sum_h Q_h. \quad (2,15)$$

Вычислим энергию W, затрачиваемую на нашу операцию. Пренебрегая деформацией решётки, мы будем иметь:

$$W = \sum_g \left\{ J_g(Q_g + \Delta Q_g, Q_g) + \Delta Q_g e \left(\Phi_g + \frac{1}{2}\varphi_g + \vartheta_g\right)\right\}, \quad (2,16)$$

где

$$\begin{cases} \Phi_g \text{ — потенциал в точке } g \text{ в случае идеальной решётки,} \\ \varphi_g \text{ — дополнительный потенциал, созданный аномальными ионами,} \\ \vartheta_g \text{ — потенциал, созданный дефектами,} \end{cases}$$

причём

$$\begin{cases} \Phi_g = \sum\limits_{g'} \dfrac{Q_{g'}}{R_{gg'}} e, \\ \varphi_g = \sum\limits_{g'} \dfrac{\Delta Q_{g'}}{R_{gg'}} e, \\ \vartheta_g = \sum\limits_{h} \dfrac{Q_h}{R_{gh}} e. \end{cases} \tag{2,17}$$

В формуле (2,16) члены

$$\frac{1}{2}\sum_g \Delta Q_g e\varphi_g = \frac{1}{2}\sum_g\sum_{g'} \frac{\Delta Q_g \Delta Q_{g'}}{R_{gg'}} e^2$$

выражают энергию взаимодействия аномальных ионов между собой, а члены

$$\sum_g \Delta Q_g e\vartheta_g = \sum_g\sum_h \frac{\Delta Q_g Q_h}{R_{gh}} e^2 \tag{2,18}$$

— энергию взаимодействия аномальных ионов с дефектами решётки. Это взаимодействие, как это видно из (2,18), может иметь характер притяжения (если $\Delta Q_g Q_h < 0$) или отталкивания (если $\Delta Q_g Q_h > 0$). В соответствии с этим все дефекты, присутствующие в решётке, можно разделить на два сорта: дефекты, для которых $Q_h > 0$, и дефекты, для которых $Q_h < 0$. Мы назовём их, соответственно, *электроположительными* и *электроотрицательными* дефектами. К электроположительным дефектам относятся, как это видно из (2,14), металлоидные пустые узлы и металлические междуузельные ионы. Электроотрицательными дефектами являются металлические пустые узлы и металлоидные междуузельные ионы. Мы видим из (2,18), что аномальные ионы с избыточным отрицательным зарядом притягиваются к электроположительным повреждениям и отталкиваются от электроотрицательных повреждений. Аномальные ионы с избыточным положительным зарядом, наоборот, отталкиваются от электроположительных и притягиваются к электроотрицательным повреждениям. Таким образом, дефекты в решётке играют роль ловушек, локализующих около себя соответствующие аномальные ионы.

В случае идеальной решётки $\vartheta_g = 0$, и формула (2,16) превращается в формулу (2,9). Кроме того, в случае идеаль-

ной решётки

$$\sum \Delta Q_g = 0$$

и, следовательно, согласно (2,15):

$$n_1 = n_2.$$

Из общей формулы (2,16) может быть получен ряд выражений, которые нам пригодятся в дальнейшем. Выпишем их здесь. В частности, для энергии W_{1g} и $W_{2g'}$, затрачиваемой при введении в нашу реальную решётку соответственно электрона и дырки (не взаимодействующих между собой), мы имеем, согласно (2,16):

$$\begin{cases} W_{1g} = J_g(Q_g - 1, Q_g) - e(\Phi_g + \vartheta_g), \\ W_{2g'} = J_{g'}(Q_{g'} + 1, Q_{g'}) + e(\Phi_{g'} + \vartheta_{g'}). \end{cases} \tag{2,19}$$

Предположим, что электрон локализован около h-го (электроположительного) дефекта, а дырка — около h'-го (электроотрицательного) дефекта. Предположим, иначе говоря, что g-й ион, принявший на себя электрон, находится вблизи h-го дефекта, а g'-й ион, лишившийся своего электрона, расположен по соседству с h'-м дефектом.

Если концентрация дефектов в решётке достаточно мала, так что расстояние между соседними дефектами достаточно велико, то можно считать, что каждый из аномальных ионов (g-й и g'-й) взаимодействует практически лишь с тем дефектом, около которого он локализован. В этом случае, согласно (2,17):

$$\begin{cases} e\vartheta_g = \dfrac{Q_h}{R_{gh}} e^2, \\ e\vartheta_{g'} = \dfrac{Q_{h'}}{R_{gh'}} e^2 \end{cases}$$

(где $Q_h > 0$, $Q_{h'} < 0$), и выражения (2,19) принимают вид:

$$\begin{cases} W_{1g} = J_g(Q_g - 1, Q_g) - e\Phi_g - \dfrac{Q_h}{R_{gh}} e^2, \\ W_{2g'} = J_{g'}(Q_{g'} + 1, Q_{g'}) + e\Phi_{g'} + \dfrac{Q_{h'}}{R_{gh'}} e^2. \end{cases} \tag{2,20}$$

В том случае, когда в качестве дефектов мы имеем дело с пустыми узлами, выражения (2,20) могут быть переписаны

так (поскольку g-й ион, так же, как и g'-й ион, по своей природе может быть как металлическим, так и металлоидным):

$$W_{1g}=\begin{cases} =J_M(+p-1,+p)-e\Phi_M-\frac{q}{a}e^2=W'_{1M}, \\ =J_R(-q-1,-q)-e\Phi_R-\frac{q}{d}e^2=W'_{1R}, \end{cases}$$
$$W_{2g}=\begin{cases} =J_M(+p+1,+p)+e\Phi_M-\frac{p}{d}e^2=W'_{2M}, \\ =J_R(-q+1,-q)+e\Phi_R-\frac{p}{a}e^2=W'_{2R}, \end{cases} \quad (2,21)$$

где a — наикратчайшее расстояние между двумя разноимёнными узлами решётки, d — наикратчайшее расстояние между двумя одноимёнными узлами (связь между a и d различна в зависимости от кристаллографической структуры решётки).

Из двух уровней W'_{1M} и W'_{1R} *нормальному* состоянию электрона соответствует наинизший из них; *нормальному* состоянию дырки соответствует наинизший из двух уровней W'_{2M} и W'_{2R}. Таким образом, электрон, введённый в нашу решётку с пустыми узлами, располагается около пустого металлоидного узла на ближайшем к нему металлическом или металлоидном ионе (в зависимости от того, что является энергетически более выгодным), а дырка оказывается локализованной около пустого металлического узла на ближайшем металлическом или металлоидном ионе (опять-таки в зависимости от того, что для неё выгоднее).

Состояния (2,21) являются *связанными* состояниями для электрона и дырки. *Свободное* состояние мы получим, выведя электрон и дырку из сферы действия дефектов, т. е. полагая в (2,20) $R_{gh}=R_{g'h'}=\infty$. Для свободного состояния, следовательно, мы будем иметь вместо (2,21):

$$W_{1g}=\begin{cases} =J_M(+p-1,+p)-e\Phi_M=W_{1M}, \\ =J_R(-q-1,-q)-e\Phi_R=W_{1R}, \end{cases}$$
$$W_{2g'}=\begin{cases} =J_M(+p+1,+p)+e\Phi_M=W_{2M}, \\ =J_R(-q+1,-q)+e\Phi_R=W_{2R}. \end{cases} \quad (2,22)$$

Числа

$$W_{1M},\ W_{1R},\ W_{2M},\ W_{2R}$$

можно трактовать как энергетические эквиваленты следующих реакций:

$$\left.\begin{array}{ll} \text{а)} & M^{+p} + e = M^{+p-1} - W_{1M}, \\ \text{б)} & R^{-q} + e = R^{-q-1} - W_{1R}, \\ \text{в)} & M^{+p} - e = M^{+p+1} - W_{2M}, \\ \text{г)} & R^{-q} - e = R^{-q+1} - W_{2R}. \end{array}\right\} \qquad (2,23)$$

Первые две из этих реакций представляют собой реакции *введения* электрона (в идеальную решётку); две последние являются реакциями присоединения дырки, т. е. реакциями *удаления* электрона (из идеальной решётки).

§ 3. Электропроводность решётки

А. Различные составляющие электропроводности. Электропроводность кристаллической решётки по своей природе может быть *ионной* и *электронной*. Необходимым и достаточным критерием ионной проводимости служит эффект Фарадея. Достаточным, но не необходимым критерием электронной проводимости является эффект Холла. В зависимости от того, какие частицы служат носителями тока, ионная проводимость может быть: катионной, анионной или смешанной. Электронная проводимость может быть: чисто электронной, чисто дырочной или смешанной.

Электропроводность σ определяется числом n_i «свободных» зарядов того или иного сорта и их подвижностью v_i. «Свободными» мы называем такие заряды, которые способны перемещаться по решётке независимо друг от друга и без затраты энергии. Строго говоря, в электропроводности смешаны в той или иной пропорции все составляющие:

$$\sigma = \sum_i e_i v_i n_i .$$

Соотношение между различными составляющими электропроводности различно в зависимости от условий. Внешними воздействиями на решётку можно увеличивать роль той или иной из составляющих. Обычно, одна какая-нибудь из составляющих существенно преобладает, так что остальными можно пренебречь. Заметим, что подвижность ионов очень мала по сравнению с подвижностью электронов. Поэтому преобладание электронной составляющей над ионной наступает даже

тогда, когда число свободных электронов ещё очень мало по сравнению с числом свободных ионов.

B. Ионная проводимость. До тех пор пока решётка остаётся идеальной, ионная проводимость невозможна. Действительно, единственно возможный способ перемещения ионов в идеальной решётке, не нарушающий её идеальности, заключается в обмене местами между одноимёнными ионами. Однако такой обмен местами не может дать тока.

Для того чтобы сделать возможной ионную проводимость, необходимо предварительно нарушить идеальность решётки, переведя некоторое количество ионов из узлов в междуузелья. Междуузельные ионы способны перемещаться внутри решётки и служат, таким образом, носителями тока. С другой стороны, пустые узлы также способны странствовать по решётке. Их также можно рассматривать как носителей тока. Соответственно этому мы различаем два механизма ионной проводимости: 1) механизм Шоттки — носителями тока являются пустые узлы; 2) механизм Френкеля — носителями тока являются междуузельные ионы.

Эти два сорта носителей, вообще говоря, обладают разными подвижностями. Число носителей обоих сортов в реальной решётке не обязано быть одинаковым. Действительно, реальная решётка уже с самого начала, т. е. при абсолютном нуле температуры, может обладать некоторым запасом пустых узлов или междуузельных ионов. Если же при абсолютном нуле решётка была идеальной, то эта идеальность нарушается при нагревании. При нагревании происходит тепловая диссоциация ионов из узлов в междуузелья. В этом случае число междуузельных ионов в точности равно числу пустых узлов. Таким образом, всякая решётка при температуре, отличной от нуля, обладает тем или иным запасом носителей тока. Этот запас зависит как от биографии решётки, так и от температуры.

Строго говоря, когда мы имеем дело с ионной проводимостью, одновременно действуют оба механизма: механизм Френкеля и механизм Шоттки. Однако в одних случаях руководящая роль принадлежит механизму Френкеля, в других, наоборот, — механизму Шоттки. Эти два механизма схематически иллюстрированы на рис. 8.

C. Условия возникновения электронной (дырочной) проводимости в идеальной решётке. Идеальная решётка типа M_mR_r в нормальном состоянии построена из металлических

ионов с зарядом $+p$ и металлоидных ионов с зарядом $-q$. Такие ионы мы будем называть нормальными ионами. В решётке, состоящей из нормальных ионов, электронная проводимость невозможна. Для того чтобы сделать возможной электронную проводимость, решётку надо предварительно возбудить. Иначе говоря, на некоторых ионах решётки необходимо создать аномальный заряд. Аномальные ионы, т. е. ионы с аномальным зарядом, могут быть получены в результате следующих операций:

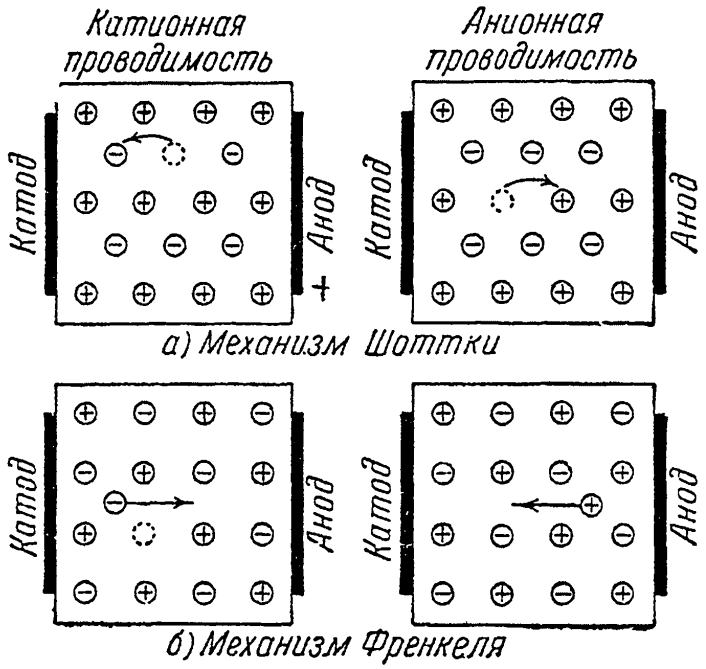


Рис. 8. Механизмы ионной проводимости.

1. В результате введения в решётку посторонних электронов и рассаживания их по ионам. Решётка при этом заряжается отрицательно. (Опыты Тартаковского по бомбардировке кристалла электронным пучком.)

2. В результате удаления из решётки принадлежащих ей электронов. Иными словами, в результате введения в решётку посторонних дырок и рассаживания их по ионам. В этом случае решётка заряжается положительно.

3. Наконец, в результате пересаживания электронов внутри решётки с иона на ион. В этом случае аномальные ионы создаются за счёт внутренних ресурсов решётки. Решётка при этом остаётся электрически нейтральной. Заметим, что такое пересаживание электронов можно описывать как введение в решётку одновременно электронов и дырок в равном количестве.

Существование в решётке аномальных ионов является условием, необходимым, но недостаточным для возникновения электронной проводимости.

Действительно, если аномальные ионы противоположных знаков находятся в непосредственном соседстве друг с другом,

то они оказываются связанными между собой. Такая группа аномальных ионов, в целом электрически нейтральная, представляет собой то, что Я. И. Френкель назвал экситоном. Экситон может перемещаться по решётке. Это значит, что вся группа аномальных состояний, остающихся связанными между собой, перемещается по решётке как одно целое. Это не может дать тока.

Чтобы сделать проводимость возможной, аномальные ионы необходимо развести настолько, чтобы можно было пренебрегать взаимодействием между ними, т. е. чтобы можно было поведение каждого аномального иона рассматривать независимо от остальных. Каждое аномальное состояние способно при этом перемещаться по решётке без затраты энергии, передаваясь с одного иона на соседний одноимённый ион. В идеальной решётке аномальные состояния странствуют беспрепятственно. В реальной решётке они могут застревать у мест повреждений. Перемещение аномального состояния происходит за счёт перескакивания электрона с иона на ион. Эти электроны, способные путешествовать по решётке без затраты энергии, перепрыгивая с иона на ион, и служат носителями тока.

В некоторых случаях переход электрона с одного иона на другой удобно описывать как переход дырки в противоположном направлении. Таким образом, различие между электронной и дырочной проводимостью не столько физическое, сколько терминологическое. Если мы имеем дело с аномальными ионами, обладающими избыточным отрицательным зарядом, то мы говорим об электронной проводимости. Если же мы имеем дело с аномальными ионами, обладающими избыточным положительным зарядом, то в этом случае удобнее говорить о дырочной проводимости.

D. Механизмы электронной (дырочной) проводимости в идеальной решётке. В зависимости от того, какие ионы (металлические или металлоидные) служат передатчиками электрона или дырки, мы можем различать два механизма электронной и два механизма дырочной проводимости. В одних случаях электрону энергетически выгодно путешествовать по металлическим ионам; в других случаях — по металлоидным. То же самое можно сказать о дырке. Эти возможные механизмы схематически иллюстрированы на рис. 9. Они могут быть изображены также следующими символическими формулами:

Электронная проводимость:

$$\text{а)}\ M^{+p-1} + M^{+p} \longrightarrow M^{+p} + M^{+p-1},$$
$$\text{б)}\ R^{-q-1} + R^{-q} \longrightarrow R^{-q} + R^{-q-1}.$$

Дырочная проводимость:

$$\text{в)}\ R^{-q+1} + R^{-q} \longrightarrow R^{-q} + R^{-q+1},$$
$$\text{г)}\ M^{+p+1} + M^{+p} \longrightarrow M^{+p} + M^{+p+1}. \qquad (3,1)$$

Каждый из двух электронных механизмов может, вообще говоря, комбинировать с любым из двух дырочных механизмов, так что мы получаем следующие четыре возможных комбинации:

I) *ав*, III) *бв*,
II) *аг*, IV) *бг*. (3, 2)

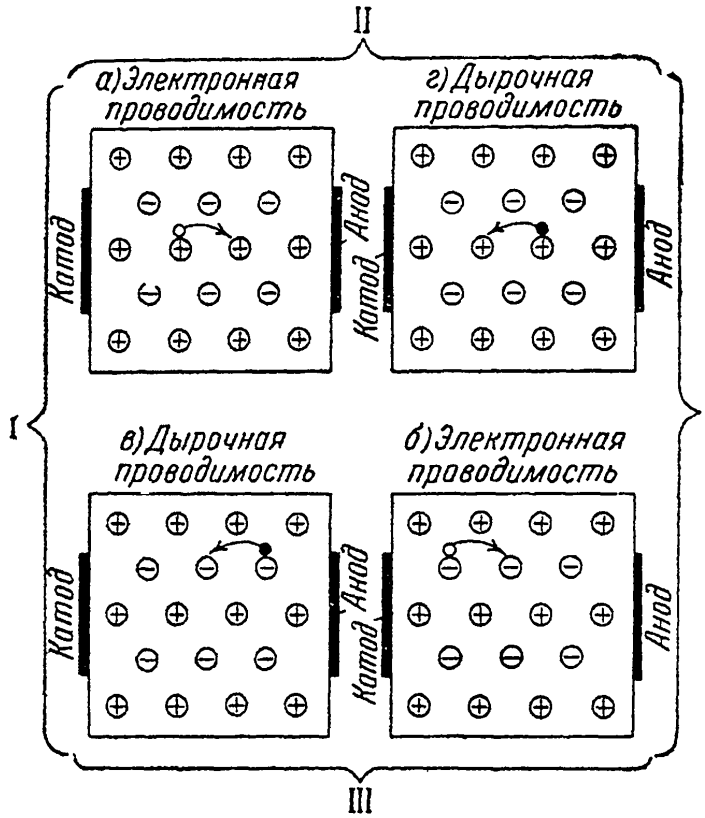


Рис. 9. Механизмы электронной проводимости.

На рис. 9 эти комбинации отмечены фигурными скобками. Эти четыре комбинации соответствуют четырём реакциям (2,13).

В действительности осуществляться будет тот тип проводимости, который в данной решётке и при данных условиях энергетически наиболее выгоден, т. е. тот, который соответствует энергетически наиболее выгодной из четырёх возможных реакций (2,13). Вопрос о том, какая из этих четырёх реакций наивыгоднейшая, может быть решён при помощи рис. 10. На этом рисунке энергия $W_{gg'}$ изображена как функция $\frac{1}{a}$ согласно формулам (2,12), в которых положено $R = \infty$. Условие $R = \infty$ означает, что аномальные ионы считаются независимыми, что является критерием проводимости.

На рис. 10 мы имеем:

$$AA' = W_{RM}, \quad CC' = W_{RR},$$
$$BB' = W_{MM}, \quad DD' = W_{MR}.$$

Рис. 10, *а* относится к случаю, когда $W_{MM} < W_{RR}$, а рисунок 10, *б* — к случаю, когда $W_{RR} < W_{MM}$. Можно показать, что прямые BB' и CC' заведомо лежат над осью абсцисс, а точка E'

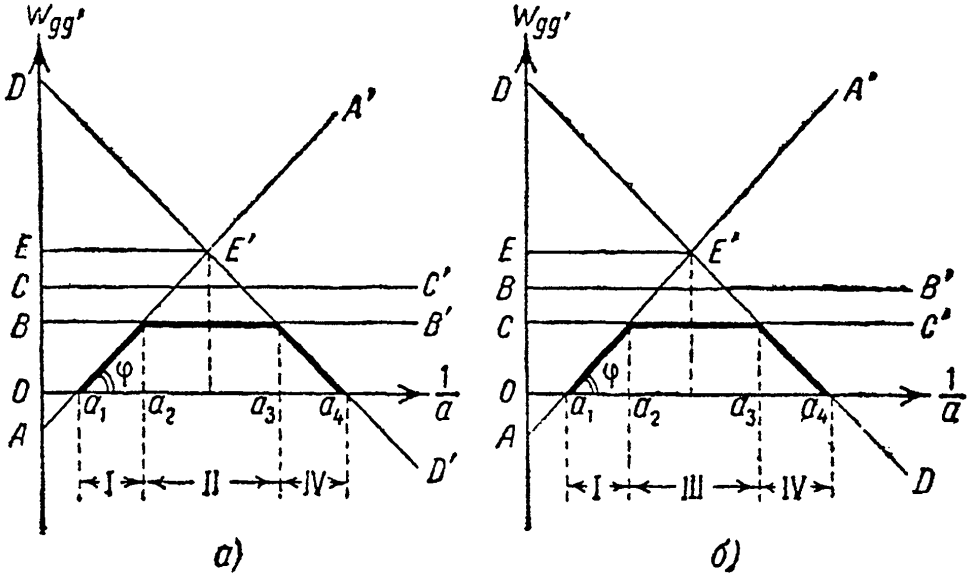


Рис. 10. Различные типы проводимости в идеальной решётке как функции $1/a$.

$$\begin{cases} OA = I_M(+p-1, +p) + I_R(-q+1, -q), \\ OB = I_M(+p-1, +p) + I_M(+p+1, +p), \\ OC = I_R(-q-1, -q) + I_R(-q+1, -q), \\ OD = I_R(-q-1, -q) + I_M(+p+1, +p). \end{cases} \qquad \operatorname{tg}\varphi = 2\alpha\eta e^2 M$$

(пересечение прямых AA' и DD') лежит заведомо над прямой BB' или CC' (как это и изображено на рисунке). Прямые AA', BB', CC', DD' соответствуют четырём возможным типам проводимости, перечисленным в (3, 2). В действительности, как мы уже говорили, осуществляется тот механизм проводимости, который при данных условиях является энергетически наиболее выгодным (жирная ломаная кривая на рис. 10). Таким образом, при прочих равных условиях механизм проводимости фиксируется значением параметра a. Мы имеем тот или иной механизм в зависимости от того, в какое место на оси абсцисс попадает данное a. Если изменять a, т. е. изменять масштаб, подвергая решётку всестороннему сжатию или растяжению, так, однако, чтобы решётка при этом оставалась подобной

сама себе, то один механизм проводимости будет сменяться другим в последовательности, изображённой на рис. 11.

Мы видим из рис. 10, что при $a > a_1$ и при $a < a_4$ энергия возбуждения $W_{gg'}$ оказывается отрицательной. В этих областях превращение нормальных ионов в аномальные должно происходить самопроизвольно. Таким образом, решётка заданной структуры, состоящая из металлических ионов с зарядом $+p$ и из металлоидных ионов с зарядом $-q$ (т. е. характеризующаяся заданным значением η), оказывается устойчивой лишь в ограниченном интервале значений a. Мы возвращаемся к вопросу, который мы исследовали в § 2-В. Точки a_1 и a_4 на рис. 10 соответствуют точкам излома на рис. 7.

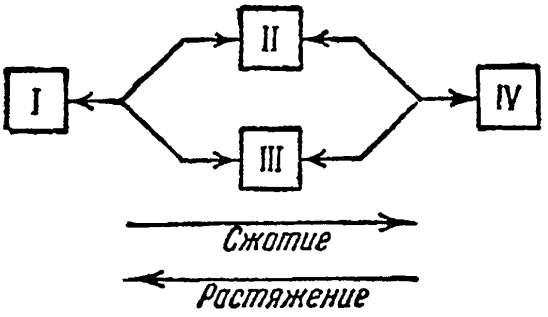


Рис. 11. Смена механизмов проводимости при всестороннем растяжении и сжатии решётки.

В заключение заметим, что вычисления, проделанные для ряда решёток, показывают, что, например, решётка ZnO обладает проводимостью типа I, а решётка CuO — проводимостью типа II. Если решётку ZnO подвергнуть всестороннему сжатию, а решётку CuO — всестороннему растяжению, то решётки обменяются механизмами проводимости.

Е. Проводимость в реальной решётке. Обратимся теперь к реальной решётке. И здесь, как и в случае идеальной решётки, проводимость обусловливается наличием свободных аномальных ионов. Перенос тока сводится к передаче аномального состояния с одного иона на другой одноимённый ион. Электронная проводимость осуществляется аномальными ионами с избыточным отрицательным зарядом, дырочная проводимость — аномальными ионами с избыточным положительным зарядом. Как и в идеальной решётке, мы имеем два возможных механизма электронной проводимости и два возможных механизма дырочной проводимости, соответственно тому, какие ионы (металлические или металлоидные) служат передатчиками электронов и дырок.

Однако реальная решётка в отношении электропроводности отличается от идеальной двумя чертами.

Реальная решётка, в отличие от идеальной, содержит повреждения. Эти повреждения (пустые узлы или междуузель-

ные ионы) служат центрами притяжения для аномальных ионов. Реальную решётку можно представлять себе как такую решётку, в которой расставлены ловушки, захватывающие аномальные ионы. Эти ловушки играют связывающую роль: аномальный ион (или, выражаясь точнее, аномальное состояние), попав в сферу действия такой ловушки, оказывается локализованным. Наличие таких центров прилипания — это первая черта, отличающая реальную решётку от идеальной.

Вторая отличительная черта состоит в том, что реальная решётка даже в нормальном состоянии может содержать в себе некоторый запас аномальных ионов. Этим она отличается от идеальной решётки, в которой аномальные ионы могут возникать лишь в результате возбуждения. Наличие аномальных ионов обусловлено теми компенсирующими электронами и дырками, которые должны быть введены в реальную решётку для обеспечения её электрической нейтральности.

В нормальном состоянии аномальные ионы локализованы в реальной решётке около мест повреждений. Таким образом, реальная решётка, как и идеальная, в нормальном состоянии не обладает электронной проводимостью, но не потому, что в ней нет аномальных ионов, а потому, что эти аномальные ионы находятся в связанном состоянии. Проводимость в реальной решётке может быть возбуждена не только в результате создания новых свободных аномальных ионов, как это имеет место для идеальной решётки, но и в результате перевода уже существующих аномальных ионов из связанного состояния в свободное. Этот второй процесс требует меньших энергетических затрат, чем первый.

Аномальный ион, локализованный около дефекта, «нейтрализует» электрический заряд этого дефекта. Дефект вместе с прилипшим к нему аномальным ионом представляет собой электрически нейтральное образование. Перемещение этих образований по решётке не может дать тока. Таким образом, дефекты связывают аномальные ионы и этим исключают электронную проводимость. Аномальные ионы в свою очередь связывают дефекты и тем самым исключают возможность ионной проводимости.

Ионная и электронная проводимости возникают в решётке одновременно, коль скоро дефект и аномальный ион отведены друг от друга настолько, чтобы можно было пренебречь взаимодействием между ними.

§ 4. Энергетический спектр решётки (частный случай)

А. Электронные и дырочные уровни. Мы рассмотрим здесь реальные решётки наиболее простого типа. Пусть мы имеем решётку с пустыми узлами, причём ограничимся решётками, построенными из одновалентных ионов, для которых, следовательно, $p = q = 1$. Примером таких решёток могут служить:

NaCl, KJ, AgBr, AgF, CsCl, CuCl и др.

Пусть наша решётка содержит (в единице объёма) X_M пустых металлических и X_R пустых металлоидных узлов. Электрическая нейтральность решётки может быть сохранена путём введения в неё X_R электронов и X_M дырок, т. е. путём создания в ней X_R аномальных ионов с дополнительным отрицательным зарядом и X_M аномальных ионов с дополнительным положительным зарядом. В нормальном состоянии, т. е. в состоянии с минимальной энергией, все аномальные ионы с избыточным отрицательным зарядом локализованы около пустых металлоидных узлов, а все аномальные ионы с избыточным положительным зарядом — около пустых металлических узлов. На рис. 12 изображены возможные нормальные расположения аномальных ионов (иначе говоря, электрона и дырки) около пустого узла. Из двух возможностей *а* и *б* (см. рис. 12) электрон выбирает ту, которая для него энергетически более выгодна. В свою очередь, дырка из двух возможных расположений *в* и *г* выбирает наивыгоднейшее. Каждый из двух

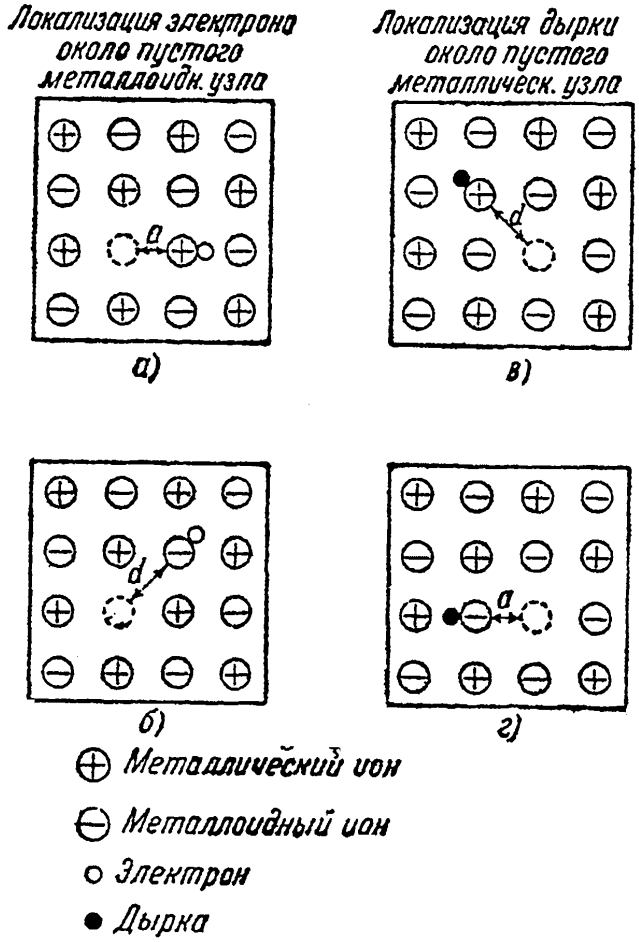


Рис. 12. Нормальное размещение электронов и дырок в реальной решётке.

рисунков в левом столбце может сочетаться с любым из двух рисунков в правом столбце. Какая из четырёх возможных комбинаций

$$ав,\ аг,\ бв,\ бг$$

будет осуществляться в действительности (т. е. будет наивыгоднейшей) — это зависит от природы решётки.

Мы будем предполагать, что концентрация пустых узлов мала, так что они расположены достаточно далеко друг от друга, чтобы можно было пренебречь взаимодействием между ними. В этом случае можно также пренебречь взаимодействием аномальных ионов между собой, а также взаимодействием каждого аномального иона со всеми пустыми узлами решётки, за исключением того пустого узла, около которого данный аномальный ион локализован.

В этом случае электроны и дырки, вводимые в решётку, могут считаться независимыми между собой. Энергия W_{1g}, необходимая для введения электрона, и энергия W_{2g}, затрачиваемая при введении дырки, выражаются формулами (2, 20). В случае электронов и дырок, *локализовінных* около пустых узлов, формулы (2, 20) превращаются в формулы (2,21), в которых, однако, мы должны положить $p = q = 1$. В случае *свободных* электронов и дырок (т. е. выведенных из сферы действия пустых узлов) формулы (2,20) превращаются в формулы (2,22).

Возвращаясь к формулам (2, 21) и (2,22), мы примем следующие обозначения. Пусть

$$\left.\begin{aligned} W_1' &\text{ — наименьшее из двух чисел } W_{1M}' \text{ и } W_{1R}' \text{ в (2,21),}\\ W_2' &\text{ — наименьшее из двух чисел } W_{2M}' \text{ и } W_{2R}' \text{ в (2,21),}\\ W_1 &\text{ — наименьшее из двух чисел } W_{1M} \text{ и } W_{1R} \text{ в (2,22),}\\ W_2 &\text{ — наименьшее из двух чисел } W_{2M} \text{ и } W_{2R} \text{ в (2,22).} \end{aligned}\right\} \quad (4,1)$$

Электронные уровни W_1' и W_1 изображены на рис. 13, *а*. Это — такие уровни, на которые может быть посажен вводимый в решётку электрон. Дырочные уровни W_2' и W_2 изображены отдельно на рис. 13, *б*. Это — такие уровни, на которые может быть посажена дырка, т. е. с которых может быть снят электрон, удаляемый нами из решётки. На рис. 13, *б*, в отличие от рис. 13, *а*, энергия откладывается по оси ординат не снизу вверх, а сверху вниз, поскольку

введение дырки эквивалентно удалению электрона и наоборот. На рис. 13, *в* энергетические спектры электрона и дырки соединены вместе. Энергия электрона и дырки, удалённых за пределы решётки, принята за нуль (ось OO').

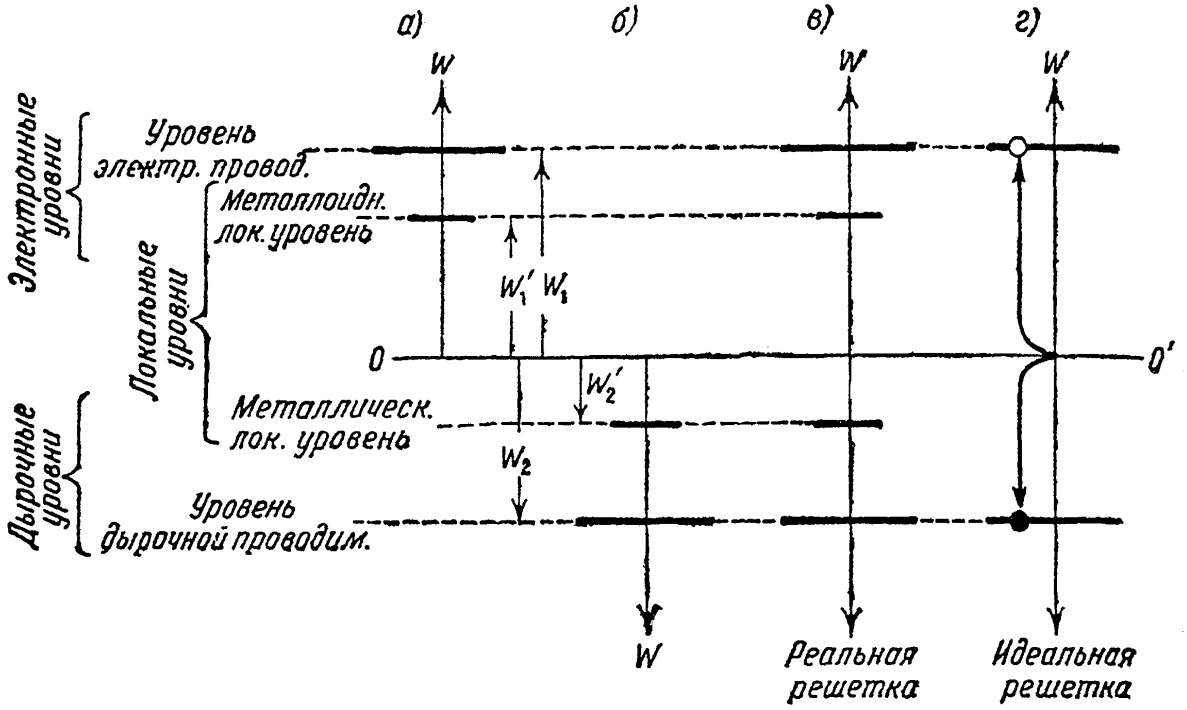


Рис. 13. «Нормальные уровни» и «уровни проводимости» для электрона и дырки.

Уровень W_1' мы будем называть *металлоидным* локальным уровнем. Электрон, посаженный на этот уровень, локализован около пустого металлоидного узла. Уровень W_2' мы будем называть *металлическим* локальным уровнем. Дырка, сидящая на этом уровне, локализована около пустого металлического узла. Уровни W_1' и W_2' соответствуют нормальному состоянию электрона и дырки. Это состояние — непроводящее. Действительно, пустой металлоидный узел и компенсирующий электрон, локализованный около него, связаны между собой и могут передвигаться по решётке лишь как одно целое. Это не даст тока. То же можно сказать про пустой металлический узел с локализованной около него дыркой. Пустой металлоидный узел в совокупности с электроном и пустой металлический узел в совокупности с дыркой представляют собой электрически нейтральные образования. Энергии W_1' и W_2' можно трактовать как энергии «нейтрализации» соответственного пустого металлоидного и пустого металли-

ческого узлов. Локальные уровни W_1' и W_2' являются специфической особенностью реальной решётки. В идеальной решётке эти уровни отсутствуют, и в энергетическом спектре остаются лишь уровни W_1 и W_2, как это и изображено на рис. 13, *г*.

Уровень W_1 мы будем называть уровнем электронной проводимости. Электрон, поднятый на этот уровень, является свободным электроном, преодолевшим взаимодействие с пустым металлоидным узлом. Уровень W_2 мы будем называть уровнем дырочной проводимости. Дырка, находясь на этом уровне, является свободной дыркой, вышедшей из сферы притяжения пустого металлического узла. Для того чтобы создать в решётке проводимость, надо посадить электрон на уровень W_1, сняв его с какого-либо более низкого уровня, или посадить дырку на уровень W_2, т. е. удалить электрон с этого уровня, подняв его на более высокий уровень, или же совершить то и другое одновременно. Последняя операция может быть осуществлена путём перевода электрона с уровня W_2 на уровень W_1; в результате на уровне W_2 появляется дырка, а на уровне W_1 — электрон, что, кстати, свидетельствует о появлении в решётке двух свободных (невзаимодействующих) аномальных ионов, из которых один несёт на себе избыточный электрон, а другой — избыточную дырку (например, нейтральные атомы Na и Cl в решётке $NaCl = Na^+ + Cl^-$). Энергия, затрачиваемая на эту операцию, которую можно рассматривать как одновременное введение в решётку электрона и дырки (см. рис. 13, *г*), будет, очевидно, равна $W_1 + W_2$.

Заметим, что на энергетической диаграмме рисунка 13 перемещение электрона снизу вверх (т. е. с более низких на более высокие уровни) связано с затратой энергии, а перемещение электрона сверху вниз — с выделением энергии. Для дырки — наоборот. Таким образом, электрон всегда стремится упасть вниз, т. е. перейти с более высокого на более низкий свободный уровень; дырка, наоборот, имеет тенденцию всплыть наверх. Поэтому, если на рис. 13 электрон помещается над дыркой, то таким электрону и дырке энергетически выгодно рекомбинировать друг с другом.

Заметим, что металлоидный уровень W_1' может лежать над осью OO' (как это изображено на рис. 13) или под этой осью, т. е. мы можем иметь $W_1' > 0$ или $W_1' < 0$. В послед-

нем случае введение электрона в решётку сопровождается не затратой, а выделением энергии, и можно говорить о «сродстве» электрона к решётке, подобно тому как говорят о сродстве электрона к атому.

В то же время металлический W_2' уровень также может лежать под осью OO' (как это изображено на рис. 13) или над этой осью, т. е. мы можем иметь $W_2' > 0$ или $W_2' < 0$. В последнем случае можно говорить о «сродстве» дырки к решётке. Действительно, если $W_2' < 0$, т. е. если металлический уровень помещается над осью OO', то введение дырки сопровождается выделением, а не затратой энергии.

Наконец, металлоидный уровень W_1' может быть расположен *под* металлическим уровнем W_2' или *над* ним (на рис. 13 изображён именно этот последний случай), т. е. мы можем иметь:

$$A)\ W_1' + W_2' < 0 \text{ или } B)\ W_1' + W_2' > 0. \qquad (4,2)$$

Будем различать эти два случая. В решётках, для которых осуществляется случай А, введение электрона и дырки является энергетически выгодным. Это может иметь место, когда решётка обладает сродством к электрону или к дырке. Для решёток, соответствующих случаю В, эта операция невыгодна [1]).

В. Условия рекомбинации электронов и дырок. Рассмотрим теперь схему электронных и дырочных уровней применительно ко всей нашей решётке в целом. Для нашей решётки, в которой имеется X_R пустых металлоидных узлов и X_M пустых металлических узлов, мы будем иметь в энергетическом спектре X_R металлоидных уровней и X_M металлических уровней. Выражаясь таким образом, мы хотим сказать, что каждый пустой металлоидный узел является ловушкой для электрона и таких независимых друг от друга ловушек имеется X_R штук, а каждый пустой металлический узел является ловушкой для дырки и таких, также независимых между собой ловушек имеется X_M штук. В нашу решётку мы вводим X_R электронов, которые рассаживаются по металлоидным уровням, и X_M дырок, которые рассаживаются по металлическим уровням. В случае решёток типа B эти электроны и

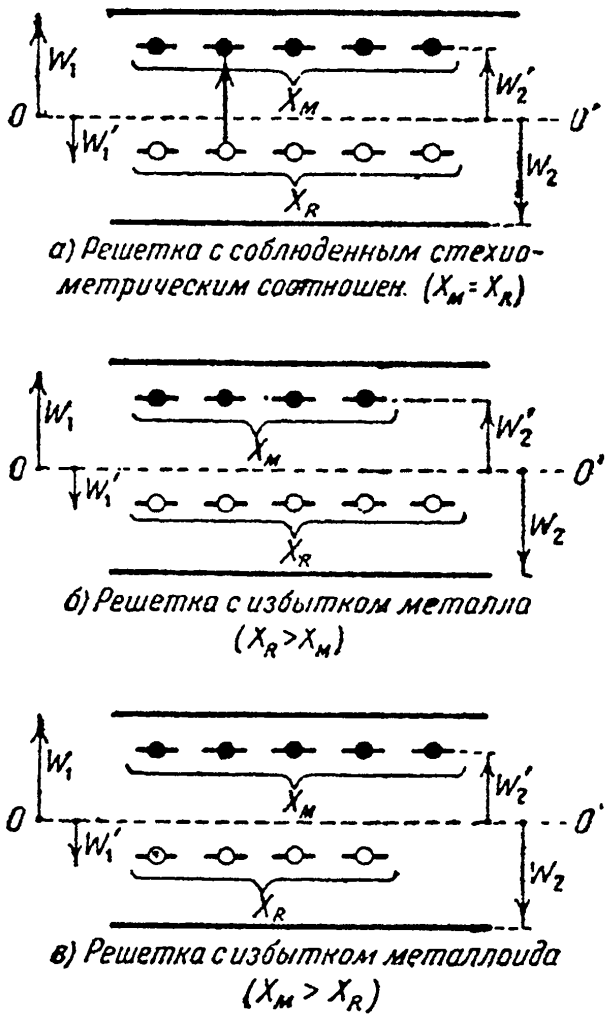
[1]) Заметим, что в случае идеальной решётки операция введения электрона и дырки всегда невыгодна, ибо мы имеем всегда:

$$W_1 + W_2 > 0.$$

дырки имеют тенденцию рекомбинировать друг с другом, ибо такая рекомбинация приводит в этом случае к понижению общей энергии системы. В случае же решёток типа A такая рекомбинация оказывается энергетически невыгодной. Это может быть пояснено с помощью энергетических схем, представленных на рис. 14 и 15.

A) Рис. 14 относится к случаю A, т. е. к тому случаю, когда металлоидные уровни в энергетическом спектре лежат *под* металлическими.

Рекомбинация электрона с дыркой на языке нашей энергетической схемы означает переход электрона с уровня W_1' на уровень W_2', как это показано на рис. 14,*а* жирной стрелкой. Такой переход связан с затратой энергии и является, следовательно, энергетически невыгодным.



а) Решетка с соблюденным стехиометрическим соотношен. ($X_M = X_R$)

б) Решетка с избытком металла ($X_R > X_M$)

в) Решетка с избытком металлоида ($X_M > X_R$)

Рис. 14. Схема электронных (дырочных) уровней для решётки «типа A» ($W_1' + W_2' < 0$).

Заметим, что в решётках типа A, несмотря на наличие пустых узлов, ионная проводимость не может осуществляться. Действительно, пустые узлы в такой решётке нейтрализованы компенсирующими зарядами и не могут принимать участия в проводимости. Они могут передвигаться по решётке лишь вместе с прилипшими к ним аномальными ионами, т. е. лишь в качестве электрически нейтральных образований. Для того чтобы металлоидные пустые узлы сделать способными к переносу тока, надо снять электроны с уровней W_1', а для того чтобы металлические пустые узлы сделать способными к переносу тока, надо снять дырки с уровней W_2'.

B) Рис. 15 относится к случаю *B*, т. е. к тому случаю, когда металлоидные уровни расположены над металлическими.

Рекомбинация электрона с дыркой, как и в случае рис. 14, выражается переходом электрона с уровня W_1' на уровень W_2'. Однако теперь этот переход, изображённый в левой части рис. 15 стрелками, направленными вниз, является энергетически выгод-

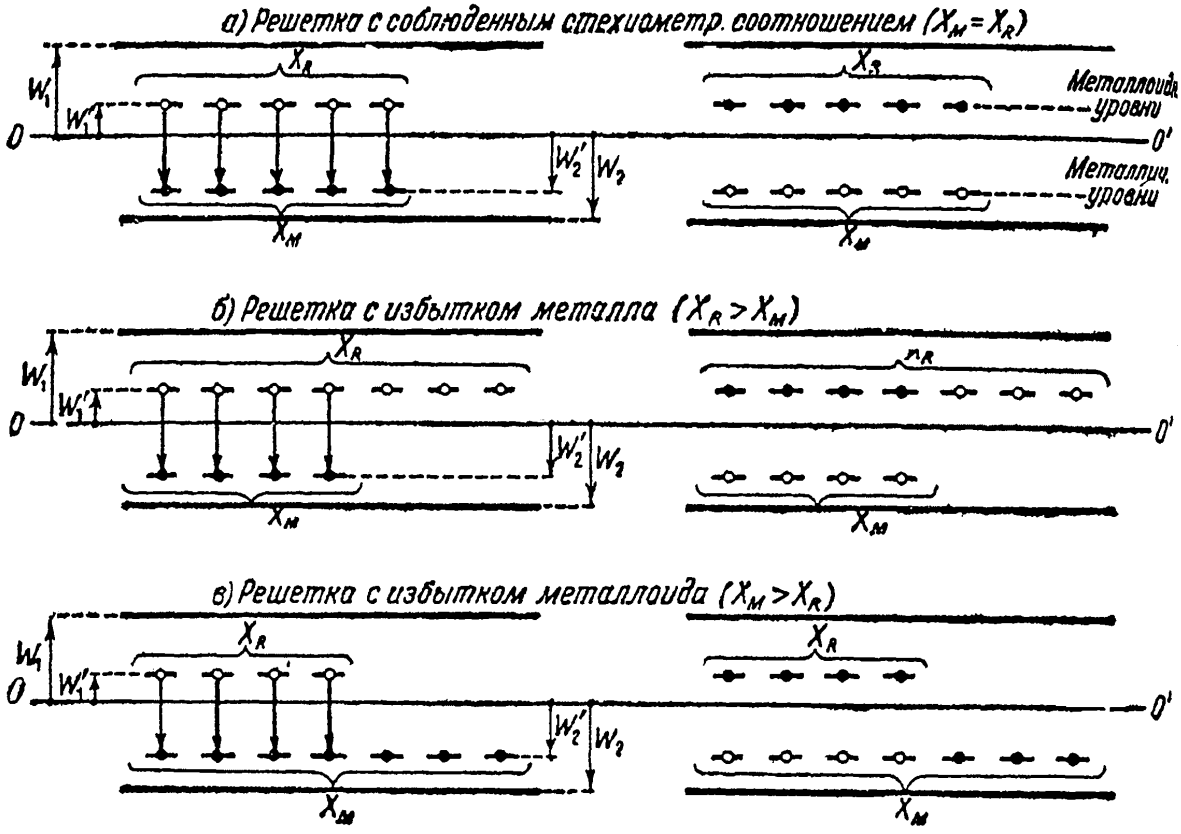


Рис. 15. Схема электронных (дырочных) уровней для решётки «типа *B*» ($W_1' + W_2' > 0$).

ным. В решётке, предоставленной самой себе и изолированной от воздействия каких-либо возбуждающих агентов, электроны покинут уровни W_1' и перейдут на уровни W_2', в результате чего энергетические схемы, изображённые в левой части рис. 15, должны быть заменены энергетическими схемами, изображёнными в правой части рисунка. Нейтрализованной остаётся лишь та часть пустых узлов, которая определяет собой отклонение от стехиометрического соотношения. Действительно при стехиометрическом избытке металла, как видно из рис. 15, *б* (правая часть), на уровнях W_1' остались лишь те электроны, которым нехватило мест на уровнях W_2'; при стехиометрическом избытке металлоида, как это видно из рис. 15, *в* (пра-

вая часть), среди уровней W_2' лишь те остались незанятыми, для которых нехватило электронов.

Заметим, что в этом случае наша решётка будет обладать ионной проводимостью, причём число носителей тока (металлических и металлоидных пустых узлов) будет определяться степенью отклонения нашей реальной решётки от идеальной, но не степенью отклонения от стехиометрического соотношения, т. е. значением параметра μ, но не значением параметра ν.

Условия (4,2), являющиеся критериями двух разных типов решёток, могут быть легко расшифрованы, если в (4,2) подставить (2,21), приняв во внимание обозначения (4,1). Заметим, однако, что формулы (2,21) остаются правильными лишь до тех пор, пока мы пренебрегаем деформацией решётки. В действительности же введение электрона или дырки всегда сопровождается некоторым сдвигом ионов из их нормальных положений равновесия. Деформация даёт некоторый выигрыш энергии, так что учёт этой деформации приводит к некоторому понижению электронных уровней W_1' и повышению дырочных уровней W_2' на нашей энергетической диаграмме. При учёте деформации условие $W_1' + W_2' < 0$ следует рассматривать как условие *достаточное,* но *не необходимое* для того, чтобы решётка могла быть отнесена к типу A, а условие $W_1' + W_2' > 0$ как условие *необходимое*, но *недостаточное* для того, чтобы решётка принадлежала к типу B.

Заметим в заключение, что вычисления, сделанные в рамках наших приближений (пренебрежение деформацией решётки), показывают, что, в частности, все щёлочно-галоидные решётки относятся к типу B, что, впрочем, и следовало ожидать.

§ 5. Энергетический спектр решётки (общий случай)

А. Условия рекомбинации электронов и дырок. Результаты предыдущего параграфа могут быть в качественном отношении перенесены на общий случай произвольной реальной решётки. Пусть мы имеем решётку типа M_mR_r, построенную из металлических ионов с зарядом $+p$ и металлоидных ионов с зарядом $-q$. Мы рассмотрим здесь реальную решётку с пустыми узлами и одновременно с этим — другой тип реальной решётки: решётку с междуузельными ионами.

Пусть мы имеем на единицу объёма:

$\begin{cases} N_1 — \text{общее число электроположительных дефектов,} \\ N_2 — \text{общее число электроотрицательных дефектов.} \end{cases}$

Очевидно, согласно (1,9):

$$\left.\begin{aligned} N_1 &= \begin{cases} = X_R = \frac{1}{2} Zr(\mu + \nu), \\ = Y_M = \frac{1}{2} Zm(\mu + \nu), \end{cases} \\ N_2 &= \begin{cases} = X_M = \frac{1}{2} Zm(\mu - \nu), \\ = Y_R = \frac{1}{2} Zr(\mu - \nu), \end{cases} \end{aligned}\right\} \tag{5,1}$$

где Z — число молекул M_mR_r в единице объёма идеальной решётки.

Для соблюдения электрической нейтральности мы вводим в нашу решётку n_1 электронов и n_2 дырок, где, согласно (1,15):

$$\left.\begin{aligned} n_1 = \alpha N_1 &= \begin{cases} = qX_R, \\ = pY_M, \end{cases} \\ n_2 = \beta N_2 &= \begin{cases} = pX_M, \\ = qY_R. \end{cases} \end{aligned}\right\} \tag{5,2}$$

Здесь приняты обозначения:

$$\alpha = \begin{cases} = q\ldots & \text{для решётки с пустыми узлами,} \\ = p\ldots & \text{для решётки с междуузельными ионами,} \end{cases}$$

$$\beta = \begin{cases} = p\ldots & \text{для решётки с пустыми узлами,} \\ = q\ldots & \text{для решётки с междуузельными ионами.} \end{cases}$$

Эти электроны и дырки могут полностью или частично рекомбинировать друг с другом или же могут отказываться от такой рекомбинации, если она энергетически невыгодна. Теоретически возможны оба случая в зависимости от природы решётки.

Предположим сперва, что рекомбинация отсутствует. Тогда в нормальном состоянии все n_1 электронов размещаются вокруг электроположительных дефектов (по α электронов около каждого дефекта). Все n_2 дырок размещаются вокруг электроотрицательных дефектов (по β дырок около каждого дефекта).

Степень связанности различных электронов, локализованных около электроположительного повреждения, различна. То же

можно сказать о дырках, локализованных около электроотрицательного повреждения.

Обозначим через W'_{11} энергию, которая выделится, если оторвать первый электрон от дефекта и увести его за пределы решётки; через W'_{12} — энергию, выделяющуюся при отрыве следующего электрона (после того как удалён первый); через W'_{1i} — энергию, выделяющуюся при отрыве i-го электрона (если предварительно удалены все $i-1$ электронов). Соответствующие энергии для дырки обозначим через W'_{2i}. Эти энергии являются аналогами ионизационных потенциалов атома. Очевидно:

$$\left.\begin{aligned} W'_{11} > W'_{12} > W'_{13} > \ldots > W'_{1\alpha} \\ W'_{21} > W'_{22} > W'_{23} > \ldots > W'_{2\beta} \end{aligned}\right\}. \qquad (5,3)$$

Будем различать следующие три случая:

$$\left.\begin{aligned} &A)\ W'_{11} + W'_{21} < 0 \text{ и, следовательно, } W'_{1\alpha} + W'_{2\beta} < 0 \\ &B)\ W'_{1\alpha} + W'_{2\beta} > 0 \text{ и, следовательно, } W'_{11} + W'_{21} > 0 \\ &C)\ W'_{11} + W'_{21} > 0, \text{ но } W'_{1\alpha} + W'_{2\beta} < 0 \end{aligned}\right\}. \qquad (5,4)$$

В первом случае электроны и дырки не рекомбинируют (этот случай соответствует случаю A § 4); во втором случае реком-

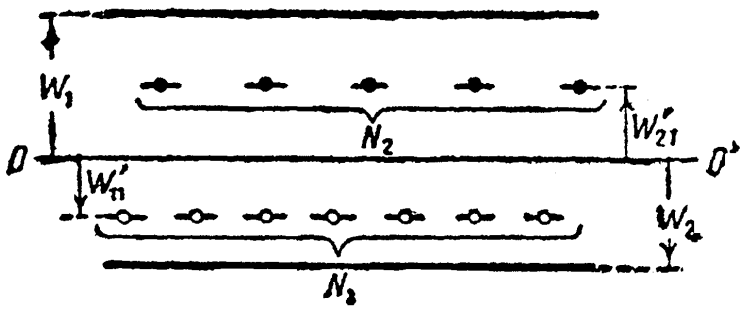


Рис. 16. Схема уровней для решётки «типа A».

бинация происходит полностью (этот случай соответствует случаю B § 4); в третьем случае имеет место лишь частичная рекомбинация. Рассмотрим подробнее два первых случая.

A) В нормальном состоянии все электроположительные и электроотрицательные повреждения полностью нейтрализованы. Энергетический спектр для решёток этого типа изображён на рис. 16, который является повторением рисунка 14. Здесь в качестве локальных уровней изображены уровни W'_{11} и W'_{21}

В нормальном состоянии каждый из электроположительных уровней W'_{11} занят электроном, а каждый из электроотрицательных уровней W'_{21} занят дыркой. Уровни W_1 и W_2 представляют собой, как и в предыдущем параграфе, уровни электронной и дырочной проводимости соответственно.

Схемой, изображённой на рис. 16, можно пользоваться лишь до тех пор, пока все или *почти все* уровни W'_{11} заняты электронами и все или *почти все* уровни W'_{21} свободны, т. е. лишь до тех пор, пока число возбуждённых электронов мало по сравнению с общим числом N_1 электроположительных повреждений, а число возбуждённых дырок мало по сравнению с общим числом N_2 электроотрицательных повреждений. Действительно, пока это имеет место, можно считать, что в возбуждении участвуют лишь наиболее слабо связанные электроны и дырки (по одному от каждого повреждения), в то время как более «глубокие», т. е. более прочно связанные, электроны и дырки вовсе не принимают участия в игре. В этом случае можно пренебрегать наличием в энергетическом спектре более глубоких уровней W'_{1i} и W'_{2i} (где $i > 1$), которые на рис. 16 вовсе не изображены.

B) Обратимся теперь к случаю *B*. Электроны рекомбинируют с дырками. В результате электронные уровни W'_{1i} оказываются лишёнными своих электронов, а дырочные уровни W'_{2i} оказываются заполненными электронами.

Энергетический спектр для решёток этого типа изображён на рис. 17. Рисунок 17, как мы видим, является повторением рисунка 15. На рис. 17, в отличие от рис. 16, в качестве локальных уровней изображены уровни $W'_{1\alpha}$ и $W'_{2\beta}$ вместо уровней W'_{11} и W'_{21} на рис. 16.

При соблюдённом стехиометрическом соотношении мы имеем:

$$n_1 = n_2 = 0;$$

все уровни $W_{1\alpha}$ свободны от электронов, а все уровни $W'_{2\beta}$ свободны от дырок (рис. 17, *a*). Электроположительные и электроотрицательные повреждения лишены нейтрализующих электронных и дырочных оболочек и их заряды остаются нескомпенсированными.

При стехиометрическом избытке металла мы имеем в соответствии с (1,16):

$$n_1 = Q\nu, \quad n_2 = 0,$$

где $Q = pZ_M = qZ_R$ — абсолютная величина заряда всех положительных или всех отрицательных ионов решётки (на 1 $см^3$). Среди уровней $W'_{1\alpha}$ мы имеем n_1 занятых уровней и $N_1 - n_1$ свободных (рис. 17, *б*). Иначе говоря, среди общего числа N_1 электроположительных повреждений мы имеем n_1 повреждений, на которых локализовано по одному электрону, т. е. положительный заряд которых частично скомпенсирован, и $N_1 - n_1$ повреждений, вовсе лишённых электронных оболочек.

При стехиометрическом избытке металлоида мы имеем:

$$n_1 = 0, \quad n_2 = -Q\nu.$$

Среди уровней $W'_{2\beta}$ заняты электронами $N_2 - n_2$ уровней; остальные n_2 уровней свободны (т. е. заняты дырками) (рис. 17, *в*). Иначе говоря, среди общего числа N_2 электроотрицательных повреждений мы имеем n_2 повреждений, на которых локализовано по одной дырке, т. е. отрицательный заряд которых частично скомпенсирован, и $N_2 - n_2$ повреждений, вовсе лишённых дырочных оболочек.

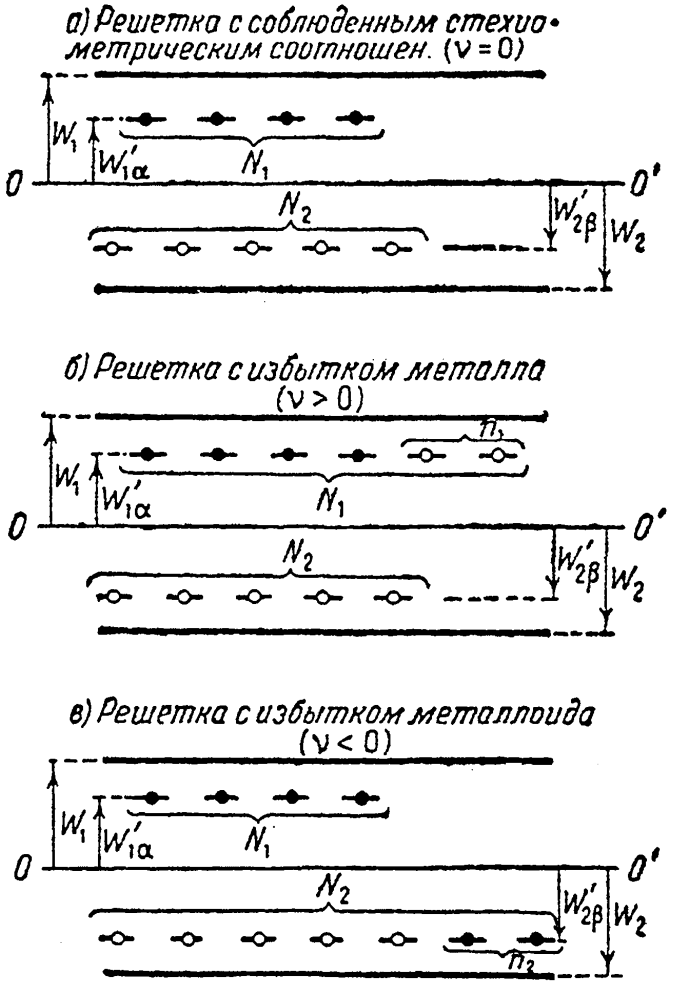


Рис. 17. Схема уровней для решётки «типа *B*».

Таким образом, в нашей решётке даже при нарушенном стехиометрическом соотношении, вообще говоря, отсутствуют повреждения, полностью нейтрализованные. Избыточные электроны или дырки, вместо того чтобы концентрироваться

целыми группами около тех или иных повреждений и обеспечивать тем самым полную нейтрализацию этих повреждений, размещаются по одному около каждого повреждения, что является энергетически заведомо более выгодным.

Схема уровней, изображённая на рис. 17, остаётся справедливой лишь до тех пор, пока число электронов, сидящих на уровнях $W'_{1\alpha}$, мало по сравнению с общим числом этих уровней, а число дырок, сидящих на уровнях $W'_{2\beta}$, мало по сравнению с общим числом этих уровней, т. е. до тех пор, пока $n_1 \ll N_1$ и $n_2 \ll N_2$, что означает $|\nu| \ll \mu$. В противном случае начинают заполняться более высокие уровни W'_{1i} (где $i < \alpha$) и W_{2i} (где $i < \beta$), которые на нашем рисунке 17 вовсе не изображены.

В. Учёт взаимодействия дефектов. Рассмотрим ещё один вопрос, представляющий интерес. Заметим, что положение локальных уровней на наших энергетических схемах (рис. 16 и 17), вообще говоря, зависит от концентрации повреждений. Эта зависимость проявляется при не слишком малых концентрациях, когда повреждения расположены в решётке достаточно близко друг к другу, так что их взаимодействием уже нельзя пренебрегать, как мы это делали, например, в предыдущем параграфе. Посмотрим, как отражается на положении локальных уровней учёт этого взаимодействия. Результаты оказываются различными, в зависимости от того, имеем ли мы дело с решётками типа *A* или с решётками типа *B*.

В решётках типа *A* в нормальном состоянии повреждения полностью «нейтрализованы» локализованными около них аномальными ионами. Таким образом, в таких решётках каждое повреждение можно рассматривать как электрически нейтральное образование. Электростатическое взаимодействие между такими повреждениями практически отсутствует.

Иначе обстоит дело в решётках типа *B*. В таких решётках дефекты или вовсе не нейтрализованы (случай соблюдённого стехиометрического соотношения) или же нейтрализованы лишь частично (случай нарушенной стехиометрии). Поэтому в таких решётках дефекты (при достаточной их концентрации) вступают в электростатическое взаимодействие друг с другом.

Рассмотрим решётки этого типа. Сделаем два предположения.

1) Во-первых, предположим, что число аномальных ионов в нашей решётке мало по сравнению с числом повреждений, т. е. $n_1 \ll N_1$ и $n_2 \ll N_2$. Тогда, учитывая взаимодействие повреждений друг с другом и с аномальными ионами, мы можем пренебрегать взаимодействием аномальных ионов между собой.

2) Во-вторых, сделаем какое-либо предположение о расположении повреждений. Предположим, что повреждения распределены равномерно по всему объёму, причём (пренебрегая отклонением от стехиометрического соотношения) будем считать «решётку», составленную из этих повреждений, увеличенной копией нашей действительной решётки. Так, например, в решётке типа NaCl электроположительные и электроотрицательные повреждения предполагаются расположенными в шахматном порядке. Если в качестве повреждений фигурируют междуузельные ионы, то представление о такой «решётке», построенной из повреждений, может быть получено, если из нашей решётки удалить все регулярные ионы, оставив иррегулярные ионы на своих местах. Если же в качестве повреждений мы имеем дело с пустыми узлами, то представление о «решётке», составленной из пустых узлов, мы получим, если все те ионы, которые мы удаляем из идеальной решётки для создания пустых узлов, мы оставим на своих местах, а удалим, наоборот, все остальные ионы. Мы предполагаем, что решётка, оставшаяся в результате такой операции, представляет собой нашу идеальную решётку в увеличенном масштабе. Это предположение является не таким искусственным, как это может показаться на первый взгляд, ибо именно такое распределение повреждений, при заданной их концентрации, является энергетически наиболее выгодным.

Первое из двух сделанных предположений позволяет нам воспользоваться формулами (2,19). Пусть избыточный электрон, введённый в решётку, в нормальном состоянии сидит на g-м ионе, находящемся вблизи h-го (электроположительного) повреждения, так что g-й ион является аномальным ионом с дополнительным отрицательным зарядом. Энергию, затрачиваемую для создания такого иона, мы обозначим (в согласии с § 2 D) через W_{1g}. Очевидно: $W_{1g} = W'_{1\alpha}$. Предположим далее, что дырка в нормальном состоянии находится на g'-м ионе, расположенном около h'-го (электроотрицательного) повреждения. Таким образом, g'-й ион, лишённый своего электрона, является аномальным ионом с дополнительным положительным

зарядом. Энергию, необходимую для образования такого иона, обозначим через $W_{2g'}$. Очевидно: $W_{2g'} = W'_{2\beta}$. Пользуясь обозначениями § 2, мы имеем, согласно (2,19):

$$\begin{cases} W_{1g} = W'_{1\alpha} = J_g(Q_g - 1, Q_g) - e(\Phi_g + \vartheta_g), \\ W_{2g'} = W'_{2\beta} = J_{g'}(Q_{g'} + 1, Q_{g'}) + e(\Phi_g + \vartheta_{g'}). \end{cases} \tag{5,5}$$

Второе из двух сделанных нами предположений позволяет установить следующие соотношения. Обозначая через ϑ_h потенциал, созданный всеми дефектами (за исключением h-го) в точке, где находится h-й дефект, и вспоминая (2,7) и (2,8), мы можем написать:

$$\left.\begin{aligned} \Phi_g &= \begin{cases} = \Phi_M \ldots, & \text{если } g\text{-й узел металлический,} \\ = \Phi_R \ldots, & \text{если } g\text{-й узел металлоидный,} \end{cases} \\ \vartheta_h &= \begin{cases} = \vartheta_1 \ldots, & \text{если } h\text{-й дефект электроположительный,} \\ = \vartheta_2 \ldots, & \text{если } h\text{-й дефект электроотрицательный,} \end{cases} \end{aligned}\right\} \tag{5,6}$$

причём

$$\begin{cases} \Phi_R - \Phi_M = 2k\dfrac{\eta e M}{a}, \\ \vartheta_2 - \vartheta_1 = 2k\dfrac{\eta e M}{L}. \end{cases} \tag{5,7}$$

В частности, для решёток типа MR, для которых $m = r = 1$:

$$\begin{cases} \Phi_R = -\Phi_M = \dfrac{\eta e}{a} M, \\ \vartheta_2 = -\vartheta_1 = \dfrac{\eta e}{L} M. \end{cases} \tag{5,8}$$

Здесь

M — коэффициент Маделунга для нашей решётки,
a — наикратчайшее расстояние между двумя узлами,
L — наикратчайшее расстояние между двумя дефектами.

Остальные обозначения те же, что и в § 2.

Параметр L непосредственно связан с параметром μ, характеризующим степень отклонения нашей реальной решётки от идеальной. Действительно:

$$\frac{a^3}{L^3} = \frac{N_1 + N_2}{Z_M + Z_R},$$

где, пренебрегая стехиометрическими нарушениями, т. е. полагая $|\nu| \ll \mu$, мы имеем, согласно (5,1):

$$\begin{cases} N_1 + N_2 = \dfrac{1}{2} Z(m + r)\mu, \\ Z_M + Z_R = Z(m + r). \end{cases}$$

Следовательно:

$$L = a\left(\frac{2}{\mu}\right)^{1/3}. \tag{5,9}$$

Считая, что расстояние R_{gh}, а также $R_{g'h'}$ между дефектом и локализованным около него аномальным ионом мало по сравнению с расстоянием между двумя соседними дефектами, мы можем положить в (5,5):

$$\begin{cases} \vartheta_g \approx \vartheta_1 + \dfrac{\alpha}{R_{gh}} e, \\ \vartheta_{g'} \approx \vartheta_2 - \dfrac{\beta}{R_{g'h'}} e, \end{cases}$$

так что формулы (5,5) примут вид:

$$\begin{cases} W'_{1\alpha} = J_g(Q_g - 1, Q_g) - e\Phi_g - e\vartheta_1 - \dfrac{\alpha}{R_{gh}} e^2, \\ W'_{2\beta} = J_{g'}(Q_{g'} + 1, Q_{g'}) + e\Phi_{g'} + e\vartheta_2 - \dfrac{\beta}{R_{g'h'}} e^2. \end{cases} \tag{5,10}$$

Расстояние между локальными уровнями $W'_{1\alpha}$ и $W'_{2\beta}$ на рис. 17 оказывается функцией от μ. Действительно, на основании (5,7), (5,9) и (5,10):

$$W'_{1\alpha} + W'_{2\beta} = J_g(Q_g - 1, Q_g) + J_{g'}(Q_{g'} + 1, Q_{g'}) - \\ - e(\Phi_g - \Phi_{g'}) - e^2\left(\frac{\alpha}{R_{gh}} + \frac{\beta}{R_{g'h'}}\right) + 2e^2 k \frac{\eta M}{a}\left(\frac{\mu}{2}\right)^{1/3}. \tag{5,11}$$

Мы видим, что по мере возрастания μ локальные уровни раздвигаются. При достаточно малом μ, т. е. при достаточно малой концентрации повреждений, можно пренебречь взаимодействием между ними, полагая в (5,10) $\vartheta_1 = \vartheta_2 = 0$. В этом случае последний член в (5,11) исчезает.

Что касается уровней проводимости W_1 и W_2, то их положение в энергетическом спектре рис. 17 не зависит от μ. Действительно, эти уровни соответствуют электрону и дырке, выведенным на такие эквипотенциальные поверхности, на которых потенциал дефектов практически равен нулю. На этих поверхностях лежат траектории свободных электронов и дырок, участвующих в проводимости. Таким образом, если g-й и g'-й аномальные ионы являются свободными, то в формулах (2,19) следует положить $\vartheta_g = \vartheta_{g'} = 0$, и мы будем иметь вместо (5,5):

$$\begin{cases} W_{1g} = W_1 = J_g(Q_g - 1, Q_g) - e\Phi_g, \\ W_{2g'} = W_2 = J_{g'}(Q_{g'} + 1, Q_{g'}) + e\Phi_{g'}. \end{cases}$$

На рис. 18 изображено положение уровней в энергетическом спектре, как функция μ, для частного случая решёток

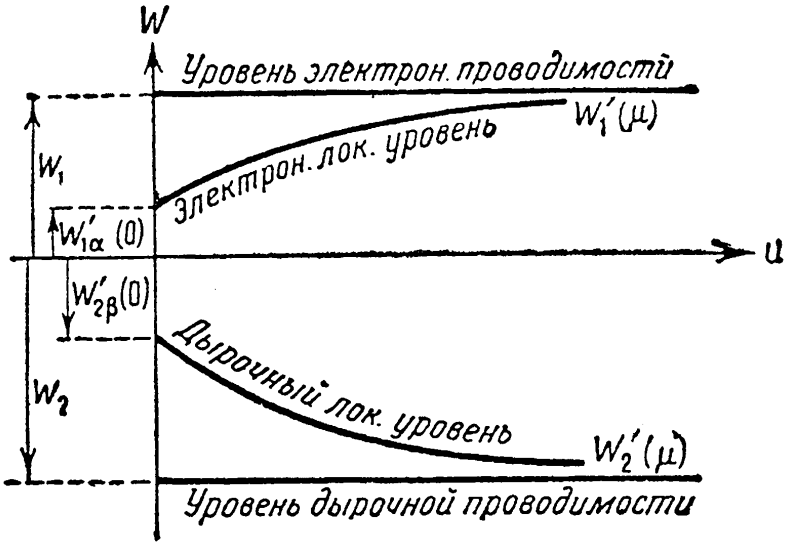


Рис. 18. Положение уровней в энергетическом спектре как функция μ.

типа MR. Для таких решёток мы имеем, согласно (5,8), (5,9) и (5,10):

$$\begin{cases} W_{1\alpha} = J_g(Q_g - 1, Q_g) - e\Phi_g - \dfrac{\alpha}{R_{gh}} e^2 + \dfrac{\eta e^2 M}{a}\left(\dfrac{\mu}{2}\right)^{1/3}. \\ W_{2\beta} = J_{g'}(Q_{g'} + 1, Q_{g'}) + e\Phi_{g'} - \dfrac{\beta}{R_{g'h'}} e^2 + \dfrac{\eta e^2 M}{a}\left(\dfrac{\mu}{2}\right)^{1/3}. \end{cases} \quad (5,12)$$

По мере возрастания μ, т. е. по мере того как увеличивается концентрация пустых узлов, электронные уровни $W'_{1\alpha}$ ползут вверх, а дырочные уровни $W'_{2\beta}$ опускаются вниз. Чем больше μ, тем труднее вводить в нашу решётку электроны и дырки, но тем легче связанные электроны и дырки переводить в свободные состояния.

В заключение заметим, что схемы уровней (рис. 14, 15, 16, 17), построенные в этой главе на базе чисто модельных представлений, будут вновь получены в главе III, в которой поведение электрона и дырки в реальной кристаллической решётке рассматриваются квантомеханически. Мы получим качественно ту же картину энергетического спектра, с той лишь разницей, что уровни проводимости окажутся размытыми в более или менее широкие энергетические полосы (зоны).

Заметим, наконец, что в рассмотренной схеме уровней собственно уже заключена вся теория электропроводности полупроводников. Все разнообразные электрические и фотоэлектрические свойства полупроводников могут быть извлечены из этой схемы (см. главы IV, V, VI).

Литература к главе I

[1] Chemical reactions involving solids (a general discussion), *Trans. of Far Soc.*, **34**, 822—883, 1938.
[2] Борн и Гепперт-Мейер, Теория твёрдого тела, ОНТИ, 1938.
3] Frenkel, *Sow. Phys.*, **12**, 1, 2, 1936.
4] Schottky, *Zsf. Phys. Chem.*, **29**, 335, 1935.

ГЛАВА II

ЭЛЕКТРОННЫЕ УРОВНИ В ИДЕАЛЬНОМ КРИСТАЛЛЕ

§ 1. Теория свободных электронов

А. Исходные предположения. Центральной задачей теории атома является построение энергетического спектра. Из энергетического спектра вытекают все основные свойства атома. То же относится к молекуле. Она отличается от атома, в конечном счёте, лишь одной единственной чертой: наличием нескольких силовых центров вместо одного. То же самое можно сказать о кристаллической решётке, которая есть не что иное, как одна громадная молекула. Здесь мы имеем целую систему силовых центров, расположенных в определённом порядке. Энергетический спектр кристаллической решётки это есть ключ к пониманию всех основных её свойств. Из анализа энергетического спектра должны вытекать, как следствия, все электрические, фотоэлектрические и оптические свойства решётки.

Кристаллическая решётка представляет собой систему взаимодействующих электронов, движущихся в поле положительных ядер, которые в первом приближении можно считать закреплёнными в некоторых жёстко зафиксированных положениях равновесия. Перенумеруем все электроны: пусть i или j — номера электронов. Перенумеруем также все ядра: пусть g — номера ядер. Поведение системы описывается уравнением Шредингера:

$$H\Psi = W\Psi, \tag{1,1}$$

где Ψ — волновая функция, зависящая от пространственных и спиновых координат всех электронов:

$$\Psi = \Psi\,(s_1 x_1 y_1 z_1, s_2 x_2 y_2 z_2, \ldots).$$

Если пренебречь магнитным взаимодействием электронов (что допустимо) и сохранить лишь электростатическое взаимодействие, то оператор H будет иметь вид:

$$H = -\frac{h^2}{8\pi^2 m}\sum_i \Delta_i + \sum_i \sum_g U_{ig} + \frac{1}{2}\sum_i \sum_j P_{ij}, \qquad (1,2)$$

где

$U_{ig} = U_g(x_i y_i z_i)$ — энергия взаимодействия i-го электрона с g-м ядром;

$P_{ij} = P(x_i y_i z_i, x_j y_j z_j)$ — энергия взаимодействия i-го электрона с j-м электроном.

Уравнение (1,1) не может быть решено точно. Обычно вводят те или иные приближения, в результате которых уравнение (1,1) упрощается. Затем отыскиваются решения такого упрощённого уравнения.

Первое упрощение, которое мы сделаем, касается учёта взаимодействия электронов. Это взаимодействие мы учтём путём введения некоторого эффективного поля, в котором каждый из электронов будем считать движущимся независимо от остальных. Иначе говоря, в выражении (1,2) мы положим

$$\frac{1}{2}\sum_j P_{ij} = S_i,$$

где $S_i = S(x_i y_i z_i)$ — так называемый самосогласованный потенциал (см. § 4 С).

В этом случае оператор энергии (1,2) принимает вид:

$$H = \sum_i H_i, \qquad (1,3)$$

где, согласно (1,2):

$$H_i = -\frac{h^2}{8\pi^2 m}\Delta_i + V(x_i y_i z_i), \qquad (1,4)$$

$$V(x_i y_i z_i) = \sum_g U_g(x_i y_i z_i) + S(x_i y_i z_i). \qquad (1,5)$$

В этом случае волновая функция Ψ может быть представлена в следующем мультипликативном виде:

$$\Psi(s_1 x_1 y_1 z_1,\ s_2 x_2 y_2 z_2, \ldots) = \prod_i \phi_i(s_i x_i y_i z_i), \qquad (1,6)$$

т. е. может быть представлена как произведение *индивидуальных* функций, относящихся к отдельным электронам. Такое представление означает, что каждому электрону может быть

приписано своё индивидуальное состояние ψ_i и своя индивидуальная энергия W_i, иначе говоря, наша многоэлектронная задача превращается в задачу одноэлектронную. Действительно, в результате подстановки выражений (1,3), (1,4) и (1,6) в уравнение (1,1), это уравнение распадается на ряд уравнений

$$H_i\psi_i = W_i\psi_i, \quad i = 1, 2, 3, \ldots, \tag{1,7}$$

где

$$\sum_i W_i = W.$$

Из этих уравнений определяются индивидуальные функции ψ_i. Уравнения (1,7) представляют собой уравнения Шредингера для каждого отдельного электрона, причём, как легко заметить, одинаковые для всех электронов. Действительно, опуская индекс у координат, мы можем написать вместо (1,7):

$$H\psi_i = W_i\psi_i, \tag{1,8}$$

где теперь через H обозначен следующий оператор:

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V(xyz). \tag{1,9}$$

Таким образом, индивидуальные функции $\psi_i(s, x, y, z)$, приписываемые различным электронам, можно рассматривать как различные решения одного и того же уравнения Шредингера (1,9).

Заметим, что в случае бесконечной идеальной решётки потенциал $V(xyz)$, входящий в (1,9), обладает важным свойством: он является периодической функцией от координат:

$$V(x, y, z) = V(x - n_1 a, y - n_2 a, z - n_3 a),$$

где n_1, n_2, n_3 — любые целые числа.

Дифференциальное уравнение типа (1,8) с периодическим потенциалом принадлежит к числу дифференциальных уравнений, достаточно изученных в математике. С уравнением (1,8) оперирует *теория Блоха*, которая будет нами рассмотрена в следующем параграфе. Здесь же мы сделаем ещё один шаг упрощения, заменяя в (1,9) периодический потенциал $V(xyz)$ некоторым *средним* его значением, постоянным по всему объёму кристалла. Тогда уравнение (1,8) превращается в совсем простое уравнение (уравнение Шредингера с постоянным потенциалом $V = \text{const}$). Из уравнения такого типа исходит *теория Зоммерфельда*, рассмотрением которой мы и ограничимся в настоящем параграфе.

Итак, в результате последовательных упрощений, мы приходим к представлению о кристалле как о семействе электронов, движущихся независимо друг от друга в некотором усреднённом поле электронов и положительных ионов решётки. Мы возвращаемся, таким образом, к картине, с которой оперировала старая теория металлов Друде. Теория Зоммерфельда и теория Друде исходят из одной и той же физической картины: свободный электронный газ, запертый в потенциальном ящике с плоским дном. Однако теория Зоммерфельда отличается от теории Друде следующими двумя существенными чертами:

1. В теории Друде движение каждого отдельного электрона подчинялось законам классической механики; в теории Зоммерфельда поведение электрона подчиняется квантовой механике.

2. В теории Друде к коллективу электронов применялась классическая статистика; в теории Зоммерфельда классическая статистика Максвелла заменяется квантовой статистикой Ферми-Дирка.

Эти реформы сыграли важную роль. С их помощью теория металлов была выведена из того тупика, в который её привела теория Друде.

В. Квантовомеханическое описание поведения электрона (бесконечный кристалл). Обратимся к уравнению (1,8), которое перепишем так:

$$H\psi = W\psi, \qquad (1,10)$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V.$$

Рассмотрим сперва бесконечный кристалл. Будем считать потенциал V постоянным во всём пространстве. Не нарушая общности, мы можем положить $V=0$. Решения уравнения (1,10) имеют вид:

$$\psi(xyz) = Ce^{\frac{i}{a}(\lambda_1 x + \lambda_2 y + \lambda_3 z)}, \qquad (1,11)$$

где λ_1, λ_2, λ_3 — произвольные вещественные числа, пропорциональные импульсам p_1, p_2, p_3:

$$\lambda_i = \frac{2\pi a}{h} p_i. \qquad (1,12)$$

Заметим, что любая линейная комбинация из функции (1,11), отличающихся друг от друга знаками при λ_1, λ_2, λ_3, будет также решением уравнения (1,10). Выражение (1,11) можно переписать в более компактном виде так:

$$\boxed{\psi(xyz) = Ce^{\frac{i}{a}\vec{\Lambda}\vec{R}},} \tag{1,13}$$

где $\vec{R}$ — вектор с компонентами x, y, z (радиус-вектор), а $\vec{\Lambda}$ — вектор с компонентами λ_1, λ_2, λ_3. Очевидно:

$$\Lambda^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \frac{4\pi^2 a^2}{h^2} p^2 = \frac{8\pi^2 a^2 m}{h^2} W \tag{1,14}$$

и

$$\vec{\Lambda} = \frac{2\pi a}{h} \vec{p}.$$

Выражение (1,14) может быть получено в результате подстановки (1,13) в уравнение (1,10). Таким образом, функция (1,11) или (1,13) соответствует состоянию электрона с определённым импульсом $\vec{p}$ и определённой энергией W. В переводе на волновой язык это означает, что мы имеем дело с плоской волной:

$$\begin{cases} \text{длина волны: } \dfrac{h}{p}, \\ \text{частота: } \dfrac{W}{h}. \end{cases}$$

Потребуем, чтобы функция $\psi(xyz)$ была *периодической* функцией от координат с периодом L, т. е.

$$\psi(x, y, z) = \psi(x + L, y + L, z + L). \tag{1,15}$$

Условия (1,15) (так называемые *условия Кармана-Борна*) играют роль граничных условий. В нашем случае бесконечного кристалла эти условия, накладываемые на функцию ψ, не являются обязательными, но весьма удобны при решении ряда вопросов. Подставляя (1,11) в (1,15), мы получаем:

$$\boxed{\begin{array}{ll} \lambda_1 = \dfrac{2\pi a}{L} j_1, & p_1 = \dfrac{h}{L} j_1 \\ \lambda_2 = \dfrac{2\pi a}{L} j_2, & p_2 = \dfrac{h}{L} j_2 \\ \lambda_3 = \dfrac{2\pi a}{L} j_3, & p_3 = \dfrac{h}{L} j_3 \end{array}} \tag{1,16}$$

где j_1, j_2, j_3 — произвольные *целые* числа.

Мы видим, что условия Кармана-Борна превращают параметры λ_1, λ_2, λ_3, а вместе с тем и импульсы p_1, p_2, p_3 из непрерывных в дискретные. Теперь электрон может обладать не любыми значениями импульса, а лишь теми значениями, которые удовлетворяют условиям (1,16). Пространство импульсов, изображённое на рис. 19, распадается на ячейки объёма $\Delta\gamma = \frac{h^3}{L^3}$. Электрон может находиться лишь в тех точках этого пространства, которые соответствуют узлам «решётки», изображённой на рис. 19.

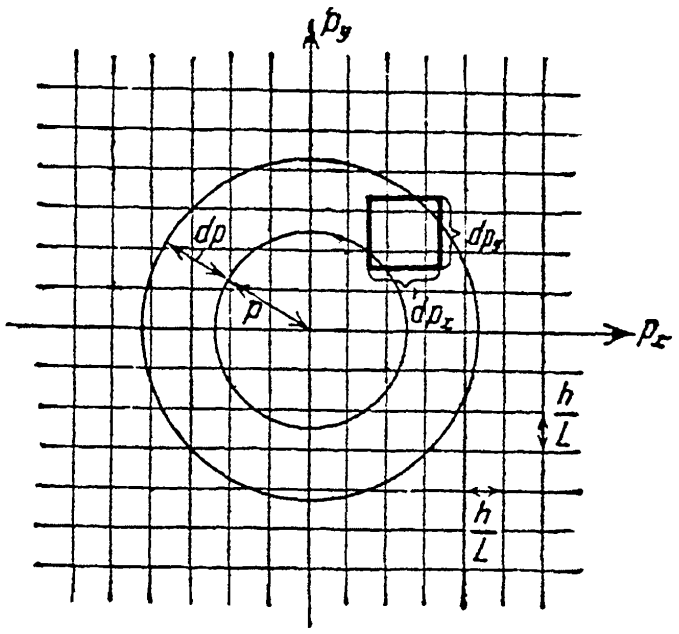


Рис. 19. Пространство импульсов.

Вместе с тем делается дискретным также энергетический спектр. Действительно, из (1,14) и (1,16) мы имеем:

$$W = \frac{h^2}{2mL^2}(j_1^2 + j_2^2 + j_3^2). \qquad (1,17)$$

Целочисленные переменные j_1, j_2, j_3 играют роль *квантовых чисел*. Зависимость p_i и W от j_i изображена на рис. 20 и 21. На оси W рис. 21 изображена система энергетических уровней электрона.

Заметим, что состояние (1,11):

$$\psi_{j_1 j_2 j_3}(xyz) = Ce^{\frac{2\pi i}{L}(j_1 x + j_2 y + j_3 z)}, \qquad (1,18)$$

соответствующее энергии (1,17), является вырожденным. Действительно, каждому значению энергии соответствует несколько состояний, поскольку одно и то же W может быть осуществлено различным набором значений j_1, j_2, j_3, в чём легко убедиться из (1,17). В одномерной задаче мы имеем двукратное вырождение, поскольку энергия W_j является чётной функцией j.

в то время как функция ψ_j этим свойством не обладает. При переходе к трёхмерному случаю кратность Z вырождения увеличивается, причём, чем выше уровень W, тем сильнее он вырожден, так что $Z = Z(W)$.

Подсчитаем число состояний $Z(W)\,dW$, соответствующих энергиям, лежащим в интервале от W до $W + dW$. Иначе говоря, нас интересует число ячеек целочисленной «решётки», изображённой на рис. 19, попадающих в шаровой слой радиуса

$$p = \sqrt{2mW}$$

и толщины

$$dp = \frac{1}{2}\sqrt{\frac{2m}{W}}\,dW.$$

Очевидно:

$$\boxed{Z(W)\,dW = \frac{4\pi p^2 dp}{\Delta\gamma} = 2\pi\frac{L^3}{h^3}(2m)^{3/2}\sqrt{W}\,dW.} \qquad (1,19)$$

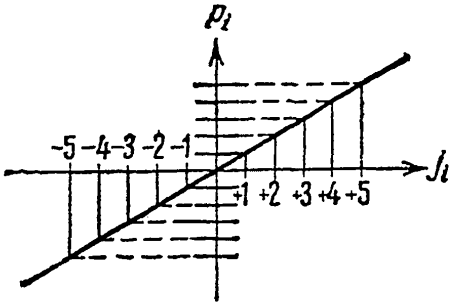


Рис 20. Возможные значения импульса.

На рис. 22 изображена кривая Z как функция W. Заметим, что при учёте спина, наличие которого мы здесь игнорировали, кратность вырождения увеличится вдвое.

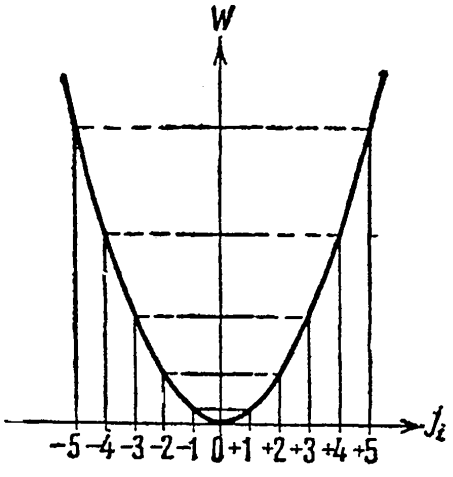


Рис. 21. Возможные значения энергии.

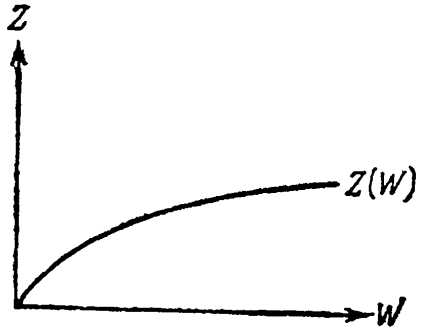


Рис. 22. Кратность вырождения Z как функция W.

С. Квантовомеханическое описание поведения электрона (ограниченный кристалл). Обратимся теперь к ограниченному кристаллу. Представим себе кристалл в виде куба

с рёбрами длины L. Систему координат расположим так, как это изображено на рис. 23. В уравнении (1,10) потенциал имеет теперь два разных значения: внутри и вне кристалла. Положим $V=0$ *внутри* кристалла; *вне* кристалла пусть мы имеем $V=\text{const}$ (причём $V>0$). На границах кристалла мы имеем скачок потенциала. Кристалл, таким образом, представляет собой потенциальный ящик с вертикальными стенками и плоским дном.

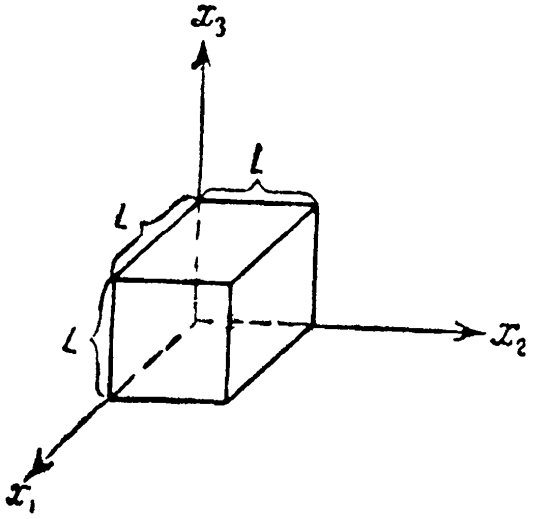


Рис. 23.

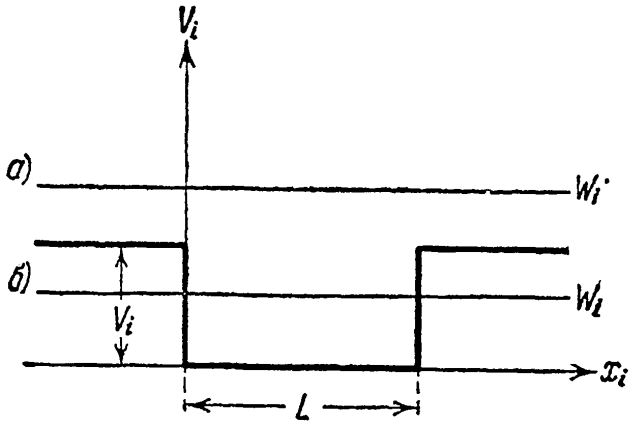


Рис. 24. Потенциальный ящик.

Обращаясь к уравнению Шредингера (1, 10), обозначим для удобства координаты электрона x, y, z через x_1, x_2, x_3. Решение уравнения (1, 10) можно искать в таком виде:

$$\psi(x_1 x_2 x_3)=\psi_1(x_1)\,\psi_2(x_2)\,\psi_3(x_3). \qquad (1{,}20)$$

Уравнение (1, 10) при этом распадается на три уравнения:

$$H_i\psi_i=W_i\psi_i,\ i=1,2,3, \qquad (1{,}21)$$

где

$$H_i=-\frac{h^2}{8\pi^2 m}\frac{d^2}{dx_i^2}+V_i(x_i),$$

причём

$$\begin{cases} W_1+W_2+W_3=W, \\ V_1(x_1)+V_2(x_2)+V_3(x_3)=V(x_1 x_2 x_3). \end{cases} \qquad (1{,}22)$$

Ход потенциала $V_i(x_i)$ (где $i=1,2,3$), изображён на рис. 24:

$$V_i(x_i)=\begin{cases} =0\ldots\ldots & \text{при } 0\leqslant x_i\leqslant L, \\ =V_i=\text{const}\ldots & \text{при } x_i<0 \text{ и } x>L. \end{cases}$$

При решении уравнения (1,21) следует различать два случая (см. рис. 24):

$$\text{а)}\ W_i > V_i,$$
$$\text{б)}\ W_i < V_i.$$

Первый случай не представляет для нас интереса. Это — случай свободной частицы, движущейся *над* потенциальной ямой. Мы ограничимся рассмотрением второго случая. Это — случай связанной частицы, запертой *внутри* потенциальной ямы.

Скачкообразный ход потенциала $V_i(x_i)$ приводит к тому, что функция $\varphi_i(x_i)$ разрывается на три части. Мы имеем:

$$\varphi_i(x_i) = \begin{cases} = A_i e^{\lambda_i' \frac{x_i}{a}} = \psi_{i1}(x_i) \ldots\ldots\ldots\ldots & x_i \leqslant 0, \\ = B_i e^{i\lambda_i \frac{x_i}{a}} + C_i e^{-i\lambda_i \frac{x_i}{a}} = \psi_{i2}(x_i) \ldots & 0 \leqslant x_i \leqslant L, \\ = D_i e^{-\lambda_i' \frac{x_i}{a}} = \psi_{i3}(x_i) \ldots\ldots\ldots\ldots & x_i \geqslant L, \end{cases} \quad (1,23)$$

где

$$\begin{cases} \lambda_i = \dfrac{2\pi a}{h} \sqrt{2mW_i} = \dfrac{2\pi a}{h} p_i, \\ \lambda_i' = \dfrac{2\pi a}{h} \sqrt{2m(V_i - W_i)}. \end{cases} \quad (1,24)$$

Из условия непрерывности функции $\varphi_i(x_i)$ и её градиента в точках разрыва потенциала (т. е. в точках $x_i = 0$ и $x_i = L$) получаем соотношения:

$$\left.\begin{aligned} A_i &= B_i + C_i, \\ \lambda_i' A_i &= i\lambda_i (B_i - C_i), \\ B_i e^{i\lambda_i \frac{L}{a}} + C_i e^{-i\lambda_i \frac{L}{a}} &= D_i e^{-\lambda_i' \frac{L}{a}}, \\ i\lambda_i \left(B_i \cdot e^{i\lambda_i \frac{L}{a}} - C_i e^{-i\lambda_i \frac{L}{a}}\right) &= -\lambda_i' D_i e^{-\lambda_i' \frac{L}{a}}, \end{aligned}\right\} \quad (1,25)$$

т. е. получаем систему из четырёх линейных однородных уравнений с четырьмя неизвестными: A_i, B_i, C_i, D_i. Детерминант Δ_i системы имеет вид:

$$\Delta_i = e^{-\lambda_i' \frac{L}{a}} \left\{ (\lambda_i' - i\lambda_i)^2 e^{-i\lambda_i \frac{L}{a}} - (\lambda_i' + i\lambda_i) e^{i\lambda_i \frac{L}{a}} \right\}.$$

Для того чтобы система (1,25) имела решения, её детерминант Δ_i должен равняться нулю. Это даёт характеристиче-

ское уравнение:

$$(\lambda_i' - i\lambda_i)^2 e^{-i\lambda_i \frac{L}{a}} - (\lambda_i' + i\lambda_i)^2 e^{i\lambda_i \frac{L}{a}} = 0.$$

После простых преобразований получаем:

$$2\lambda_i\lambda_i' \cos \lambda_i \frac{L}{a} - (\lambda_i^2 - \lambda_i'^2) \sin \lambda_i \frac{L}{a} = 0,$$

откуда

$$\operatorname{tg} \lambda_i \frac{L}{a} = \frac{2\lambda_i\lambda_i}{\lambda_i^2 - \lambda_i'^2} = \frac{2}{\frac{\lambda_i}{\lambda_i'} - \frac{\lambda_i'}{\lambda_i}}. \tag{1,26}$$

Вспоминая (1,24), мы видим, что все величины, входящие в (1,26), фиксированы по условиям задачи, за исключением W_i; параметр W_i представляет собой единственный параметр, которым мы можем распоряжаться. Однако, уравнение (1,26) накладывает определённые ограничения на возможные значения этого параметра.

Из (1,25) мы, между прочим, получаем следующую связь между коэффициентами B_i и C_i:

$$\frac{B_i}{C_i} = \frac{i\lambda_i + \lambda_i'}{i\lambda_i - \lambda_i'} = \frac{i\frac{\lambda_i}{\lambda_i'} + 1}{i\frac{\lambda_i}{\lambda_i'} - 1}. \tag{1,27}$$

Рассмотрим предельный случай: положим, что

$$\lambda_i' = \infty. \tag{1,28}$$

Это значит, что мы рассматриваем потенциальную яму с бесконечно высокими стенками ($V_i = \infty$). Физически это означает, что уровень W_i лежит достаточно глубоко внутри ямы (см. рис. 24). В этом случае правая часть уравнения (1,26) обращается в нуль:

$$\operatorname{tg} \lambda_i \frac{L}{a} = 0.$$

Это даёт:

$$\lambda_i = \frac{\pi a}{L} j_i, \tag{1,29}$$

где $j_i = 0, \pm 1, \pm 2, \ldots$ Правая часть уравнения (1,27) обращается в -1, откуда

$$B_i = -C_i. \tag{1,30}$$

Подставляя (1,28) и (1,30) в (1,23), а также (1,29) в (1,24), мы получим:

$$\begin{cases} \psi_i(x_i) = 0 \ldots\ldots\ldots\ldots\ldots\ldots X_i \leqslant 0, \\ \psi_i(x_i) = -C_i \sin \pi j_i \frac{x_i}{L} \ldots\ldots\ldots 0 \leqslant X_i \leqslant L, \\ \psi_i(x_i) = 0 \ldots\ldots\ldots\ldots\ldots\ldots X_i \geqslant L. \end{cases}$$

$$W_i = \frac{h^2}{8mL^2} j_i^2.$$

Теперь легко построить собственные функции и собственные значения исходного уравнения (1,21). На основании (1,20) и (1,22) мы будем иметь:

$$\begin{cases} \psi(x_1 x_2 x_3) = C \sin \frac{\pi j_1 x_1}{L} \sin \frac{\pi j_2 x_2}{L} \sin \frac{\pi j_3 x_3}{L} \text{ внутри кристалла,} \\ \psi(x_1 x_2 x_3) = 0 \ldots\ldots \text{ вне кристалла,} \end{cases} \tag{1,31}$$

где множитель $C = -C_1 C_2 C_3$ определяется из условия нормировки;

$$W = \frac{h^2}{8mL^2}(j_1^2 + j_2^2 + j_3^2). \tag{1,32}$$

Сравнивая (1,32) с (1,17), мы замечаем, что в случае ограниченного кристалла с бесконечно большим скачком потенциала на границах мы получаем дискретный энергетический спектр, подобный тому, который получался в неограниченном кристалле при наложении на него условий Кармана-Борна. Что касается поведения электрона, то в неограниченном кристалле оно описывалось *бегущей* волной (1,18), в то время как в ограниченном кристалле мы получаем *стоячую* волну (1,31) с узлами на границах кристалла.

D. Различие между статистикой Ферми и статистикой Максвелла. Мы рассмотрели выше поведение отдельного электрона в кристалле. В действительности кристалл содержит в себе целое множество электронов. Исследование поведения всей такой системы электронов в целом представляет собой задачу статистики.

Квантовая статистика Ферми-Дирака, которой подчиняется любая система электронов, отличается от классической статистики Максвелла-Больтцмана следующими двумя чертами:

1) Статистика Ферми-Дирака, в отличие от классической статистики, имеет дело с *тождественными* частицами, т. е. с частицами, принципиально неразличимыми. Тождественность частиц означает, что состояния системы, отличающиеся перестановкой частиц, не являются различными состояниями; все такие состояния представляют собой физически одно и то же состояние. Это обстоятельство, учитываемое при выводе функции распределения, отражается на окончательных формулах.

Неразличимость электронов накладывает определённые ограничения на функцию

$$\Psi(q_1, q_2, q_3, \ldots, q_N), \tag{1,33}$$

описывающую поведение всей нашей системы электронов и являющейся решением уравнения Шредингера (1,1). (Здесь мы для сокращённости обозначили через q_i всю четвёрку координат s_i, x_i, y_i, z_i, i-го электрона.) Действительно, модуль функции Ψ не должен изменяться при перестановке двух любых электронов, т. е. двух любых четвёрок координат q_i и q_k:

$$|\Psi(q_i, q_k)|^2 = |\Psi(q_k, q_i)|^2; \tag{1,34}$$

отсюда вытекает, что

$$\Psi(q_i, q_k) = e^{i\varphi}\Psi(q_k, q_i),$$

где φ — любое вещественное число. Если $\varphi = n\pi$, (где n — целое) (1,34), то функция $\Psi(q_i, q_k)$ называется *симметричной* (при n чётном) или *антисимметричной* (при n нечётном) относительно перестановки координат q_k и q_i. Заметим, что если функция Ψ симметрична или антисимметрична относительно перестановки i-го и k-го электронов, то, как это можно показать, она будет сохранять свой тип симметрии (т. е. будет оставаться соответственно симметричной или антисимметричной) при перестановке двух любых других электронов. Заметим, кроме того, что если функция Ψ (1,33) удовлетворяет уравнению Шредингера (1,1), то этому уравнению будет также удовлетворять любая другая функция Ψ, отличающаяся от данной перестановкой координат. Любая линейная комбинация таких функций Ψ также будет являться решением уравнения

Шредингера (1,1). Таким образом, состояние (1,33) является вырожденным (*перестановочное вырождение*).

2) Другая отличительная черта статистики Ферми-Дирака состоит в том, что частицы, с которыми она имеет дело, подчинены *принципу Паули.*

Принцип Паули, в своей самой общей формулировке, заключается в требовании, чтобы из всех возможных типов решений уравнения Шредингера были сохранены лишь *антисимметричные* решения. Все остальные решения должны быть отброшены как не имеющие физического смысла.

Заметим, что функция Ψ, имеющая мультипликативный вид

$$\Psi(q_1, q_2, \ldots, q_N) = \prod_{i=1}^{N} \psi_i(q_i), \tag{1,35}$$

не обладает требуемым свойством антисимметричности. Зато линейная комбинация, построенная из мультипликативных функций типа (1,35), отличающихся друг от друга всеми возможными перестановками координат, может быть сделана антисимметричной при соответствующем подборе коэффициентов. Такая антисимметризованная линейная комбинация из функций (1,35), удовлетворяющая, конечно, тому же уравнению Шредингера, что и функция (1,35), может быть записана в виде детерминанта:

$$\Psi(q_1, q_2, \ldots, q_N) = C \begin{vmatrix} \psi_1(q_1), \psi_2(q_1), \ldots, \psi_N(q_1), \\ \psi_1(q_2), \psi_2(q_2), \ldots, \psi_N(q_2), \\ \ldots\ldots\ldots, \ldots\ldots \\ \psi_1(q_N), \psi_2(q_N), \ldots, \psi_N(q_N). \end{vmatrix} \tag{1,36}$$

Действительно, функция Ψ (1,36) при перестановке двух любых электронов меняет знак, поскольку такая перестановка означает перестановку двух строк в детерминанте. Кроме того, мы замечаем, что функция Ψ (1,36) обращается в нуль, если в нашем семействе электронов два каких-нибудь электрона (например, i-й и k-й) находятся в одинаковом состоянии, так что $\psi_i = \psi_k$. Действительно, в этом случае i-й и k-й столбцы детерминанта оказываются одинаковыми, и детерминант, следовательно, исчезает.

Отсюда вытекает другая, частная формулировка принципа Паули: в частном случае, когда каждой частице может быть

приписано своё индивидуальное состояние, принцип Паули сводится к требованию, чтобы все частицы нашей системы находились *в различных* индивидуальных состояниях. Если возвратиться к импульсному пространству, изображённому на рис. 19, то это будет означать (при игнорировании спина), что в одной и той же точке этого пространства не может помещаться одновременно более одного электрона, подобно тому как не могут находиться несколько электронов в одной и той же точке обычного координатного пространства.

Е. Функция распределения в статистике Ферми. В статистической физике поведение системы частиц описывается с помощью так называемой *функции распределения*. Вычисление этой функции является конечной целью всякой статистической теории. Обозначим функцию распределения через $F(W)$ и определим её следующим образом:

$$F(W)=\frac{N(W)}{Z(W)}, \tag{1,37}$$

где

$N(W)\,dW$ — число частиц, обладающих энергией, лежащей в интервале от W до $W+dW$;

$Z(W)\,dW$ — число состояний (ячеек импульсного пространства), приходящихся на тот же интервал энергии.

Отношение (1,37) показывает, какая доля состояний, содержащихся в данном интервале энергии, занята электронами. Иначе говоря, функция $F(W)$ характеризует степень заполнения электронами шарового слоя, изображённого на рис. 19.

Для $Z(W)$ мы имеем, согласно (1,19):

$$Z(W)=\frac{G}{\Delta\gamma}4\pi m\sqrt{2mW}, \tag{1,38}$$

где $\Delta\gamma$ — объём ячейки импульсного пространства, а G — максимальное число частиц, которые могут поместиться в каждой ячейке. Формула (1,38) относится как к классической статистике, так и к статистике Ферми. В классической статистике объём ячейки $\Delta\gamma$ не фиксируется; отношение $\frac{G}{\Delta\gamma}$ выпадает из окончательных формул классической статистики. В статистике Ферми (в применении к электронному газу):

$$\Delta\gamma=\frac{h^3}{L^3};\quad G=2. \tag{1,39}$$

Функция распределения $F(W)$ имеет следующий вид:

$$F(W)=\begin{cases} =e^{-\frac{W-\varepsilon}{kT}} \ldots\ldots \text{в статистике Максвелла,} & (1,40)\\ =\dfrac{1}{e^{\frac{W-\varepsilon}{kT}}+1} \ldots \text{в статистике Ферми.} & (1,41)\end{cases}$$

Различный вид функции распределения для обеих статистик является следствием их принципиально различного подхода к задаче, о чём мы говорили выше.

Параметр ε, фигурирующий в формулах (1,40) и (1,41), определяется из естественного условия:

$$\int_0^\infty N(W)\,dW=\int_0^\infty F(W)\,Z(W)\,dW=N, \qquad (1,42)$$

где N — общее число частиц в нашем семействе.

В случае классической статистики, подставляя (1,38) и (1,40) в (1,42), мы получаем в результате интегрирования:

$$\varepsilon=kT\ln\frac{\Delta\gamma}{G}\frac{N}{(2\pi mkT)^{3/2}}, \qquad (1,43)$$

и, следовательно, в случае классической статистики функция распределения может быть переписана так:

$$F(W)=\frac{\Delta\gamma}{G}\cdot\frac{N}{(2\pi mkT)^{3/2}}e^{-\frac{W}{kT}}. \qquad (1,44)$$

Отсюда на основании (1,37) и (1,38) мы получаем известное максвелловское *распределение частиц по энергиям:*

$$N(W)\,dW=\frac{2\pi N}{(\pi kT)^{3/2}}\sqrt{W}e^{-\frac{W}{kT}}\,dW. \qquad (1,45)$$

Величины $\Delta\gamma$ и G выпадают из этой формулы.

В случае квантовой статистики, подставляя (1,38) и (1,41) в (1,42), мы получаем интеграл, который, вообще говоря, в конечном виде не берётся. Приходится довольствоваться приближённым вычислением этого интеграла для двух крайних предположений о параметре ε:

1) так называемый *невырожденный случай*, когда $\frac{\varepsilon}{kT}$ велико и отрицательно, т. е. когда $-\varepsilon \gg kT$;

2) так называемый *вырожденный* случай, когда $\frac{\varepsilon}{kT}$ велико и положительно, т. е. когда $\varepsilon \gg kT$.

Для этих двух случаев мы получаем из (1,42) в результате интегрирования:

$$\varepsilon = \begin{cases} \approx kT \ln \frac{h^3}{2L^3} \cdot \frac{N}{(2\pi mkT)^{3/2}} \cdots\cdots\cdots\cdots \text{при} -\frac{\varepsilon}{kT} \gg 1; & (1,46) \\ \approx \frac{h^2}{2mL^2} \left(\frac{3N}{8\pi}\right)^{2/3} \cdots\cdots\cdots\cdots\cdots\cdots \text{при} \frac{\varepsilon}{kT} \gg 1. & (1,47) \end{cases}$$

Рассмотрим подробнее каждый из этих двух случаев:

1) В *невырожденном случае* можно пренебречь единицей в знаменателе выражения (1,41). Функция Ферми превращается в функцию Максвелла (1,44), в которой $\Delta\gamma$ и G имеют значения (1,39), соответствующие электронному газу. Распределение электронов по энергиям получается в точности максвелловское [формула (1,45)].

Условие $-\varepsilon \gg kT$ на основании (1,46) может быть расшифровано так:

$$\frac{h^3}{2L^3} \cdot \frac{N}{(2\pi mkT)^{3/2}} \ll 1. \qquad (1,48)$$

Это условие является критерием невырожденного случая, т. е. критерием превращения статистики Ферми в статистику Максвелла. Это условие осуществляется при: а) достаточно больших m, б) достаточно высоких T, в) достаточно малых N.

В случае обыкновенного материального газа, а также в случае электронного газа в полупроводнике условие (1,48) соблюдается уже при обычных температурах. Для материального газа это имеет место вследствие малого N и большого m; для электронного газа в полупроводнике это имеет место благодаря малости N, несмотря на малость m. В обоих случаях мы можем пользоваться статистикой Максвелла (поскольку статистика Ферми приводит здесь практически к тем же результатам, что и статистика Максвелла).

В случае электронного газа в металле, обладающего значительной плотностью N, условие (1,48) осуществляется лишь при очень высоких температурах, лежащих далеко за пределами тех температур, с которыми имеет дело экспериментатор. В

этом случае мы должны пользоваться статистикой Ферми, которая приводит здесь к существенно иным результатам, чем статистика Максвелла.

2) Обратимся к *вырожденному случаю*, который всегда наступает при достаточно низких температурах. Рассмотрим поведение функции распределения $F(W)$ при $T=0$ и при $T>0$.

При $T=0$ мы имеем, согласно (1,41):

$$F(W)=\begin{cases}=1\ldots\text{для } W<\varepsilon,\\ =0\ldots\text{для } W>\varepsilon.\end{cases}$$

На рис. 25 изображён ход функции $F(W)$ и $N(W)$ для этого случая (жирная кривая). Энергию $W=\varepsilon$, при которой про-

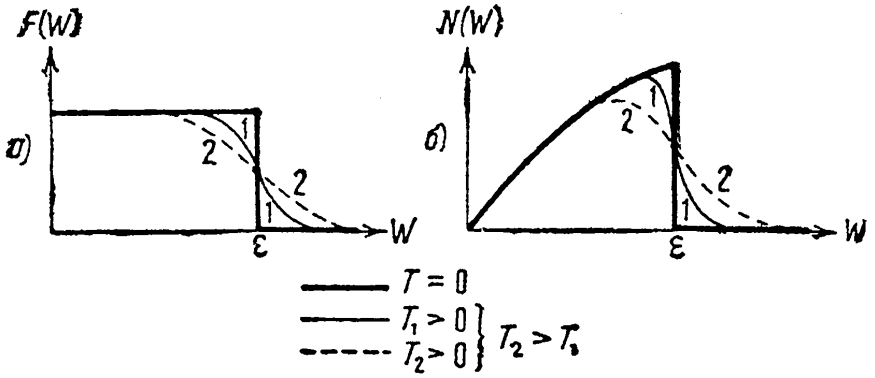


Рис. 25. *а* — функция распределения в статистике Ферми; *б* — распределение электронов по энергиям в статистике Ферми.

исходит обрыв кривой, мы будем называть *уровнем Ферми*, или *уровнем химического потенциала*. Энергии ε в импульсном пространстве на рис. 19 соответствует сфера с радиусом

$$p_\varepsilon=\sqrt{2m\varepsilon}. \qquad (1,49)$$

Мы видим, что при нуле температуры все ячейки импульсного пространства, находящиеся *внутри* этой сферы, заполнены электронами, а все ячейки *вне* этой сферы свободны, как это показано на рис. 26, *а*. Отсюда, между прочим, следует, что

$$\frac{\frac{4}{3}\pi p_\varepsilon^3}{\Delta\gamma}=\frac{N}{G}.$$

Подставляя сюда (1,49) и (1,39), мы вновь приходим к формуле (1,47) для ε.

При $T>0$ функции $F(W)$ и $N(W)$ имеют вид, изображённый на рис. 25 тонкими кривыми (кривые 1 и 2, соответственно для температур T_1 и T_2, где $T_2>T_1>0$). Кривые $F(W)$, соответствующие разным T, имеют общую точку пе-

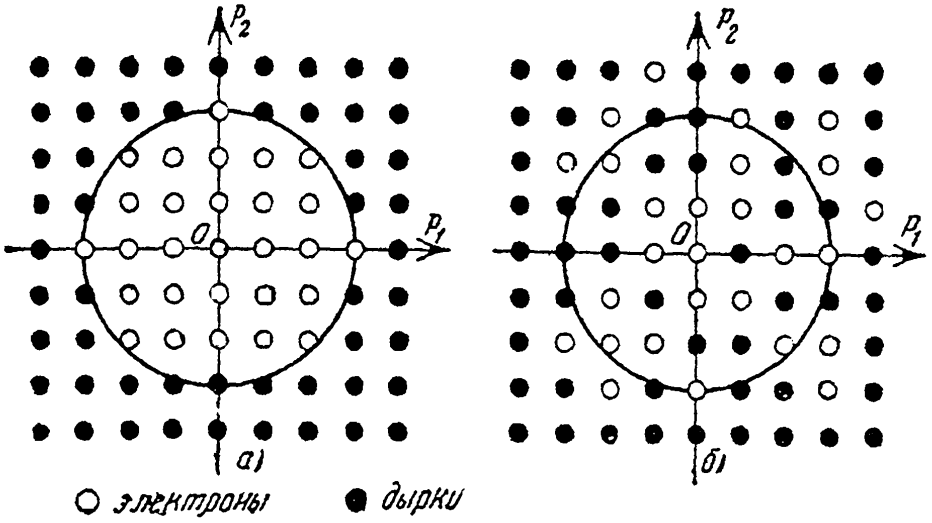


Рис. 26. Распределение электронов в пространстве импульсов.

ресечения, лежащую при $W=\varepsilon$. Действительно, как это видно из (1,41), каково бы ни было T, мы имеем всегда $F(\varepsilon)=\frac{1}{2}$. По обе стороны от этой точки кривая $F(W)$ симметрична, т. е.

$$F(\varepsilon+\Delta W)=1-F(\varepsilon-\Delta W),$$

что также видно из (1,41). Таким образом, при повышении температуры плотная упаковка электронов по ячейкам фазового пространства, изображённая на рис. 26, *а*, разрыхляется: внутри сферы радиуса p_ε появляются свободные места (дырки) за счёт электронов, выброшенных за пределы этой сферы. Мы получаем картину, изображённую на рис. 26, *б*. При этом, однако, симметрия в распределении электронов по импульсам, характерная для $T=0$ (как это видно из рис. 26, *а*), сохраняется. Каждому электрону с импульсом p может быть приведён в соответствие электрон с импульсом $-p$. Поэтому нагревание само по себе (при отсутствии электрического поля) не вызывает тока. Если, наложив на кристалл поле, мы наблюдаем электрический ток, то это свидетельствует о том, что симметрическое распределение электронов по импульсам

нарушено: вся картина, изображённая на рис. 26, *б* или 26, *а*, перестаёт быть симметричной относительно точки O. Электрическое поле, как и температура, представляет собой фактор, вызывающий перераспределение электронов в импульсном пространстве.

В заключение рассмотрим ещё один вопрос. Вычислим полную энергию $\overline{W}$, которой обладает электронный газ при некоторой заданной температуре. Очевидно:

$$\overline{W}=\int_0^\infty W\cdot N(W,T)\,dW=\int_0^\infty WF(W,T)\,Z(W)\,dW. \quad (1,50)$$

Вычисление легко провести для двух крайних случаев (невырожденный и вырожденный случай). Мы имеем:

$$\overline{W}=\begin{cases}=\frac{3}{2}NkT\ldots\ldots\ldots\ldots\ldots\ldots\text{ невырожденный случай,}\\ =\frac{2}{5}\cdot\frac{h^2}{L^2}\cdot\frac{\pi}{m}\left(\frac{3N}{8\pi}\right)^{2/3}=W_0\ldots\text{ вырожденный случай.}\end{cases}$$

Кривая $\overline{W}=\overline{W}(T)$ изображена на рис. 27 (жирная кривая). Точка T_0, определяемая из условия

$$\frac{3}{2}NkT_0=W_0,$$

делит шкалу температур на две части. Мы имеем:

вырожденный газ при $T\ll T_0$,
невырожденный газ при $T\gg T_0$.

Наклон кривой $\overline{W}(T)$ на рис. 27 характеризует собой теплоёмкость c электронного газа. Действительно, по определению:

$$c=\frac{d\overline{W}}{dT}.$$

Рассматривая рис. 27, мы можем сделать сразу два важных вывода, являющихся характерными следствиями применения к электронному газу статистики Ферми:

а) при нуле температуры энергия электронного газа не равна нулю;

б) теплоёмкость электронного газа (в области вырождения) практически равна нулю: $c \approx 0$.

Этими двумя обстоятельствами электронная теория металлов Зоммерфельда отличается от теории Друде.

Теория Зоммерфельда, построенная на квантовой механике и квантовой статистике, явилась значительным шагом вперёд по сравнению с классической теорией Друде. Однако она обладает рядом недостатков, являющихся следствием тех грубых приближений, которые лежат в её основе. Отметим здесь два из этих недостатков.

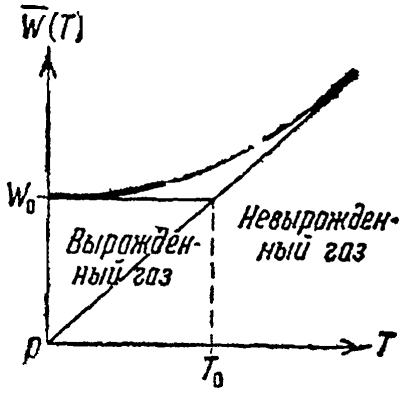


Рис. 27. Полная энергия $\overline{W}$ электронного газа как функция температуры T в статистике Ферми.

1) Оставаясь в рамках теории Зоммерфельда, нельзя построить оптики металлов. Действительно, свободный электронный газ не может поглощать света: его взаимодействие со световыми квантами сводится к Комптон-эффекту.

2) Оставаясь в рамках теории Зоммерфельда, нельзя объяснить различия между металлическими и неметаллическими кристаллами.

Эти недостатки исчезают при переходе к теории, учитывающей периодический потенциал решётки.

§ 2. Учёт периодического поля решётки (гомеополярная решётка)

А. Некоторые общие свойства уравнения Шредингера с периодическим потенциалом. Обратимся к уравнению Шредингера (1,8), описывающему поведение электрона в периодическом поле решётки (при учёте взаимодействия между электронами методом самосогласованного потенциала). Запишем это уравнение так:

$$H\psi = W\psi, \qquad (2,1)$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V(x, y, z).$$

Ограничимся рассмотрением бесконечной кубической решётки, построенной из одинаковых атомов. Пусть a — постоянная

решётки. Потенциал $V(x, y, z)$ является периодической функцией координат с периодом a:

$$V(x,y,z) = V(x - n_1 a, \quad y - n_2 a, \quad z - n_3 a), \tag{2,2}$$

где

$$n_1, n_2, n_3 = 0, \pm 1, \pm 2, \ldots$$

Дифференциальное уравнение типа (2,1) было исследовано Хиллом [1]. Затем некоторые его свойства применительно к теории металлов были рассмотрены Джонесом и Зинером [2]. Мы приведём здесь (без выводов) основные формулы, на которые в дальнейшем нам придётся ссылаться.

Собственные функции и собственные значения уравнения (2,1) могут быть представлены в таком виде:

$$\psi(x,y,z) = e^{\frac{i}{a}(\lambda_1 x + \lambda_2 y + \lambda_3 z)} u(\lambda_1, \lambda_2, \lambda_3; \; x, y, z), \tag{2,3}$$

$$W = W(\lambda_1, \lambda_2, \lambda_3). \tag{2,4}$$

Функция $u(\lambda_1, \lambda_2, \lambda_3; \; x, y, z)$ является периодической функцией координат с тем же периодом, что и потенциал, т. е.

$$u(\lambda_1, \lambda_2, \lambda_3; \; x, y, z) = u(\lambda_1, \lambda_2, \lambda_3; \; x - n_1 a, \quad y - n_2 a, \; z - n_3 a). \tag{2,5}$$

Точный вид этой функции определяется видом функции $V(x, y, z)$. При этом фиксируется также вид функции $W(\lambda_1, \lambda_2, \lambda_3)$.

Параметры $\lambda_1, \lambda_2, \lambda_3$, фигурирующие в (2,3) и (2,4), представляют собой *произвольные* вещественные числа. Их можно рассматривать как компоненты некоторого вектора $\vec{\Lambda}$. Это позволяет переписать выражение (2,3) в следующем более компактном виде:

$$\boxed{\psi(x,y,z) = e^{\frac{i}{a}\vec{\Lambda}\vec{R}} u_{\vec{\Lambda}}(x,y,z).} \tag{2,6}$$

Каждой тройке значений $\lambda_1, \lambda_2, \lambda_3$ соответствуют, как это видно из (2,3) и (2,4), своя собственная функция $\psi_{\vec{\Lambda}}$ и своё собственное значение $W_{\vec{\Lambda}}$.

Электрону, находящемуся в состоянии $\lambda_1, \lambda_2, \lambda_3$, может быть приписан импульс $\vec{p}(\lambda_1, \lambda_2, \lambda_3)$. Это есть *среднее* значение импульса, соответствующее групповой скорости волнового пакета. Импульс $\vec{p}$ как функция параметров $\lambda_1, \lambda_2, \lambda_3$ может быть вычислен, если известна энергия W как функция $\lambda_1, \lambda_2, \lambda_3$. Можно показать различными способами, что связь между $\vec{p}(\lambda_1, \lambda_2, \lambda_3)$ и $W(\lambda_1, \lambda_2, \lambda_3)$ имеет вид

$$\vec{p} = \frac{2\pi a m}{h} \operatorname{grad}_{\vec{\Lambda}} W. \tag{2,7}$$

Значок $\vec{\Lambda}$ при градиенте обозначает, что здесь имеется в виду градиент не в обычном координатном пространстве x, y, z, а в пространстве чисел $\lambda_1, \lambda_2, \lambda_3$, так что составляющие вектора $\vec{p}$ имеют вид:

$$p_l = \frac{2\pi a m}{h} \frac{\partial W}{\partial \lambda_l}, \tag{2,8}$$

где $l = 1, 2, 3$.

Заметим, что условие вещественности параметров $\lambda_1, \lambda_2, \lambda_3$ приводит к представлению о *разрешённых* и *запрещённых* значениях энергии W. «Разрешёнными» мы будем считать те вещественные значения W, которые получаются из (2,4) при вещественных значениях $\lambda_1, \lambda_2, \lambda_3$. Те же (хотя бы и вещественные) значения W, которые, однако, соответствуют комплексным $\lambda_1, \lambda_2, \lambda_3$, мы будем называть «запрещёнными». Мы увидим в дальнейшем, что «разрешённые» значения W образуют в энергетическом спектре полосы (так называемые *зоны*), разделённые «запрещёнными» участками.

Из (2,3) видно, что для «разрешённых» значений энергии модуль волновой функции периодичен с периодом решётки. Физически это означает, что все атомы решётки равноправны. «Запрещённым» же значениям энергии соответствуют волновые функции, обращающиеся в бесконечность на бесконечности. Такие волновые функции не имеют физического смысла.

Заметим далее, что на функцию (2,3) может быть наложено условие Кармана-Борна, которым мы уже пользовались в предыдущем параграфе:

$$\psi(x, y, z) = \psi(x + L, y + L, z + L).$$

Из (2,3) видно, что это условие делает непрерывные параметры $\lambda_1, \lambda_2, \lambda_3$ дискретными, так же, как это имело место в теории свободных электронов:

$$\lambda_i = \frac{2\pi a}{L} j_i,$$

где $j_i = 0, \pm 1, \pm 2, \ldots$ Вместе с тем оказываются дискретными импульс $\vec{p}$ и энергия W:

$$\begin{cases} \vec{p} = \vec{p}(j_1, j_2, j_3), \\ W = W(j_1, j_2, j_3). \end{cases}$$

Заметим, наконец, что волновая функция (2,6) отличается от волновой функции (1,13)

$$\Psi(x, y, z) = Ce^{\frac{i}{a}\vec{\Lambda}\vec{R}}, \tag{2,9}$$

с которой мы имели дело в теории свободных электронов, наличием модулирующего множителя $u_{\vec{\Lambda}}(x, y, z)$. Если в уравнении (2,1) положить $V(x, y, z) = \text{const}$, то $u_{\vec{\Lambda}}(x, y, z) = \text{const}$, и функция (2,6) превращается в (2,9). При этом, как мы знаем, функция (2,4) принимает простой вид:

$$W = \frac{h^2}{8\pi^2 a^2 m}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2),$$

откуда, согласно (2,8)

$$p_i = \frac{h}{2\pi a}\lambda_i,$$

что находится в соответствии с (1,12).

Решение уравнения (2,1) в форме (2,3) или (2,6) не может нас удовлетворить: оно не позволяет судить о структуре энергетического спектра. В теории металлов пользуются другими методами решения уравнения (2,1), связанными с физической природой задачи.

В. Метод Кронига-Пенни. В методе Кронига-Пенни [3] непрерывная периодическая функция $V(x, y, z)$ заменяется зубчатообразной разрывной кривой, изображённой на рис. 28, прямоугольные потенциальные ямы чередуются с прямоуголь-

ными барьерами. При такой апроксимации периодический характер потенциала сохраняется. Такая апроксимация позволяет довести решение уравнения Шредингера (2,1) до конца.

Ограничимся рассмотрением одномерной модели. Пусть

$$\begin{cases} a & \text{— ширина ямы,} \\ b & \text{— ширина барьера,} \\ V_0 & \text{— высота барьера.} \end{cases}$$

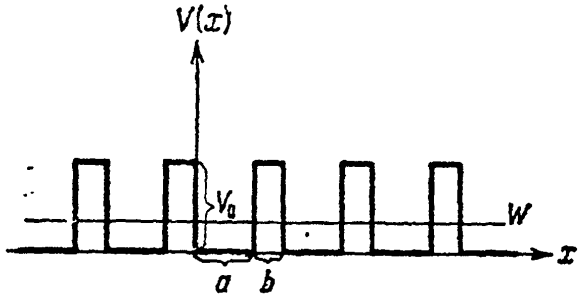


Рис. 28. Периодический потенциал решётки в модели Кронига-Пенни.

Выбирая начало координат так, как это изображено на рис. 28, мы будем иметь:

$$V(x)=\begin{cases} =0 \ \ldots \text{ при } nc \leqslant x \leqslant nc+a, \\ =V_0 \ldots \text{ при } nc+a \leqslant x \leqslant (n+1)c, \end{cases} \tag{2,10}$$

где $c=a+b$; n — любое целое число:

$$n=0, \pm 1, \pm 2, \ldots.$$

Предположим, что энергия W электрона лежит ниже вершин барьеров, т. е. $W<V_0$, как это изображено на рис. 28. Тогда решения уравнения Шредингера (2,1) будут иметь вид:

$$\psi(x)=\begin{cases} =A_n e^{i\alpha x}+B_n e^{-i\alpha x} \ldots \text{ при } nc \leqslant x \leqslant nc+a, \\ =C_n e^{\beta x}+D_n e^{-\beta x} \ldots \text{ при } nc+a \leqslant x \leqslant (n+1)c, \end{cases} \tag{2,10}$$

где

$$\begin{cases} \alpha=\dfrac{2\pi}{h}\sqrt{2mW}, \\ \beta=\dfrac{2\pi}{h}\sqrt{2m(V_0-W)} \end{cases} \tag{2,11}$$

Условие непрерывности волновой функции и её производной на границах барьеров даёт:

$$\begin{cases} C_{n-1}e^{\beta nc}+D_{n-1}e^{-\beta nc}=A_n e^{i\alpha nc}+B_n e^{-i\alpha nc}, \\ \beta C_{n-1}e^{\beta nc}-\beta D_{n-1}e^{-\beta nc}=i\alpha A_n e^{i\alpha nc}-i\alpha B_n e^{-i\alpha nc}, \\ C_n e^{\beta(nc+a)}+D_n e^{-\beta(nc+a)}=A_n e^{i\alpha(nc+a)}+B_n e^{-i\alpha(nc+a)}, \\ \beta C_n e^{\beta(nc+a)}-\beta D_n e^{-\beta(nc+a)}=i\alpha A_n e^{i\alpha(nc+a)}-i\alpha B_n e^{-i\alpha(nc+a)}, \end{cases} \tag{2,12}$$

Между коэффициентами C_n, D_n и C_{n-1}, D_{n-1} существует следующая связь:

$$\begin{cases} C_{n-1} = C_n e^{-i\lambda}, \\ D_{n-1} = D_n e^{-i\lambda}, \end{cases} \tag{2,13}$$

где λ — произвольный вещественный параметр. Действительно, используя общее свойство (2,3) волновой функции, мы можем написать:

$$\psi(x-c) = \psi(x)\, e^{-i\lambda}.$$

Отсюда вытекает (2,13). Подставляя (2,13) в (2,12), мы получаем систему из четырёх уравнений с четырьмя неизвестными: A_n, B_n, C_n, D_n. Условие совместности этих уравнений имеет вид:

$$\cos \alpha a \operatorname{ch} \beta b + \frac{\beta^2 - \alpha^2}{2\alpha\beta} \sin \alpha a \operatorname{sh} \beta b = \cos \lambda. \tag{2,14}$$

Это выражение существенно упрощается, если перейти к пределу, считая, что потенциальные барьеры становятся *бесконечно тонкими* ($b \to 0$), но в то же время *бесконечно высокими* ($V_0 \to \infty$ или $\beta \to \infty$), и при этом так, что площадь bV_0, характеризующая прозрачность барьера, остаётся неизменной. Такой переход к пределу является приёмом, характерным для метода Кронига-Пенни. Полагая в (2,14):

$$b \to 0, \quad \beta \to \infty, \quad \beta^2 \frac{ab}{2} \to P,$$

и, кроме того, принимая обозначение

$$\mu = \alpha a = \frac{2\pi a}{h} \sqrt{2mW}, \tag{2,15}$$

мы будем иметь

$$\boxed{P \frac{\sin \mu}{\mu} + \cos \mu = \cos \lambda.} \tag{2,16}$$

Параметр

$$P = \lim \frac{\beta^2 ab}{2} \tag{2,17}$$

характеризует собой степень *прозрачности* барьеров по отношению к электронам, или, иными словами, степень *связанности* электронов внутри атома. Значение P может быть различно у различных кристаллов. Если $P = 0$, то перед нами

совершенно свободный электронный газ, такой, с которым имеет дело теория Зоммерфельда; в этом случае потенциальные барьеры сглажены, их роль сведена к нулю. Если же $P=\infty$, то барьеры совершенно непроницаемы; электроны в этом случае заперты в отдельных ячейках решётки, и их движения ограничены узкими пределами.

Левая часть уравнения (2,16) представляет собой, как это видно на основании (2,15), функцию от W, а правая часть является функцией от λ. Таким образом, уравнение (2,16) даёт нам в неявном виде W как функцию λ. Трансцендентное уравнение (2,16) может быть решено относительно W в следующих трёх случаях:

а) случай абсолютно свободных электронов: $P=0$,

б) случай абсолютно связанных электронов: $P=\infty$,

в) случай сильно связанных электронов: $P\gg 1$.

Рассмотрим эти три случая.

а) *Абсолютно свободные электроны.* Полагая в (2,16) $P=0$, мы получаем

$$\mu=\lambda,$$

откуда, согласно (2,15):

$$W=\frac{h^2}{8\pi^2a^2m}\lambda^2.$$

Если на функцию $\psi(x)$ наложено условие Кармана-Борна, то

$$\lambda=\frac{2\pi a}{L}j$$

и, следовательно:

$$\boxed{W=\frac{h^2}{2mL^2}j^2,} \qquad (2,18)$$

где

$$j=0,\pm 1,\pm 2,\ldots$$

Зависимость W от j изображена на рис. 29 тонкой параболой. На том же рисунке изображена система энергетических уровней (тонкие горизонтальные чёрточки на вертикальной оси).

б) *Абсолютно связанные электроны.* Выражение (2,16) можно переписать так:

$$\frac{\sin\mu}{\mu}+\frac{\cos\mu}{P}=\frac{\cos\lambda}{P}.$$

Полагая здесь $P=\infty$, мы получаем

$$\frac{\sin\mu}{\mu}=0,$$

откуда

$$\mu=\pi j, \tag{2,19}$$

где

$$j=\pm 1, \pm 2, \ldots$$

Подставляя (2,19) в (2,15):

$$\boxed{W=\frac{h^2}{8ma^2}j^2.} \tag{2,20}$$

Зависимость W от j изображена на рис. 29 жирной параболой. Система энергетических уравнений изображена жирными гори-

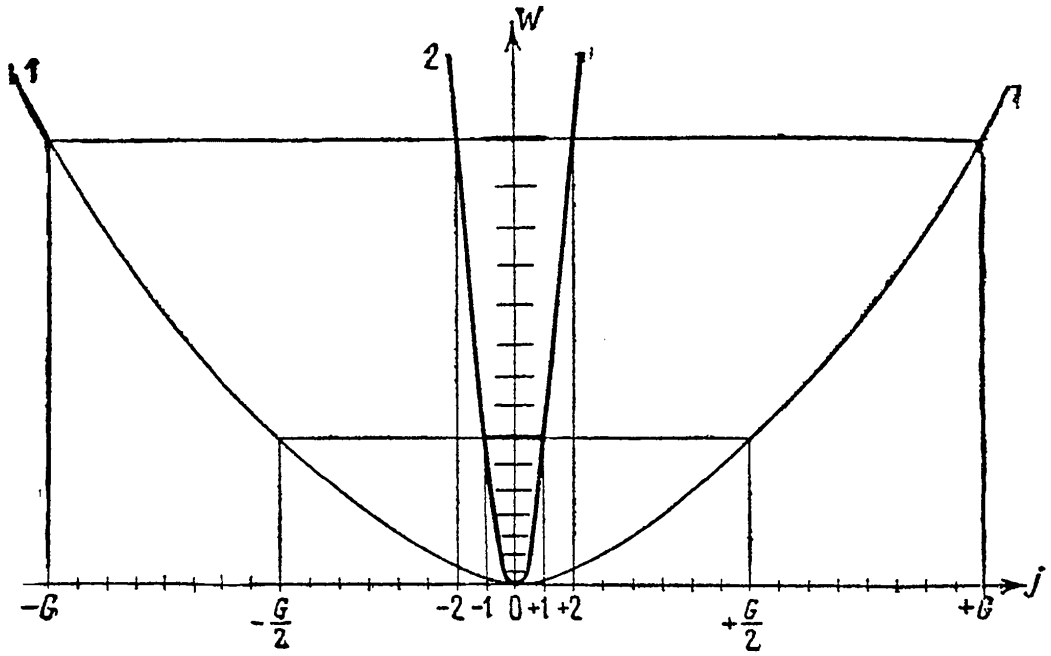


Рис. 29. Система уравнений для абсолютно свободных (кривая *1*) и абсолютно связанных (кривая *2*) электронов.

(Принято обозначение: $G=\frac{L}{a}$.)

зонтальными чёрточками на вертикальной оси. Сравнивая формулы (2,20) и (2,18), мы замечаем, что при переходе от абсолютно связанных к абсолютно свободным электронам число уровней в энергетическом спектре увеличивается в $\frac{L}{2a}$ раз.

в) *Сильно связанные электроны.* Рассмотрим область значений μ, близких к точке $\mu = k\pi$, где $k = 1, 2, 3, \ldots$, т. е. положим в (2,16):

$$\mu = k\pi + \Delta\mu,$$

где

$$\Delta\mu \ll 1. \tag{2,21}$$

Разложим левую часть выражения (2,16) в ряд по степеням $\Delta\mu$ и ограничимся первым членом разложения. Мы будем иметь:

$$(-1)^k \left(1 + \Delta\mu \cdot \frac{P}{k\pi}\right) = \cos\lambda,$$

откуда

$$\Delta\mu = \mu - k\pi = \frac{k\pi}{P}\left\{(-1)^k \cos\lambda - 1\right\}. \tag{2,22}$$

Очевидно, эта формула справедлива лишь при $P \gg 1$, ибо только в этом случае $\Delta\mu \ll 1$ при любом λ, т. е. выполняется исходное предположение (2,21). Вспоминая (2,15), получаем из (2,22):

$$W = \frac{h^2k^2}{8ma^2}\left\{1 - \frac{1}{P} + (-1)^k \frac{\cos\lambda}{P}\right\}^2. \tag{2,23}$$

При $P = \infty$ эта формула превращается в формулу (2,20), что и должно быть. Так как у нас $P \gg 1$, то формулу (2,23) можно переписать ещё так:

$$\boxed{W_k(\lambda) = A_k + (-1)^k B_k \cos(\lambda).} \tag{2,24}$$

Здесь приняты обозначения:

$$\begin{cases} A_k = \dfrac{h^2k^2}{8ma^2}\left(1 - \dfrac{1}{P}\right)^2 \approx \dfrac{h^2k^2}{8ma^2}, \\ B_k = \dfrac{h^2k^2}{8ma^2} \cdot \dfrac{2}{P}\left(1 - \dfrac{1}{P}\right) \approx \dfrac{h^2k^2}{8ma^2} \cdot \dfrac{2}{P}. \end{cases} \tag{2,25}$$

При наложении условий Кармана-Борна мы имеем:

$$W_k(j) = A_k + (-1)^k B_k \cos 2\pi \frac{a}{L} j, \tag{2,26}$$

где

$$\begin{cases} k = 1, 2, 3, \ldots \\ j = 0; \pm 1, \pm 2, \ldots \end{cases}$$

Зависимость W от j при разных k, согласно формуле (2,26), изображена на рис. 30, *а* (синусоидальные кривые). На этом же рисунке вновь нанесена парабола, изображающая зависимости W от j для случая абсолютно свободных электронов. Мы видим, что при переходе от абсолютно свободных

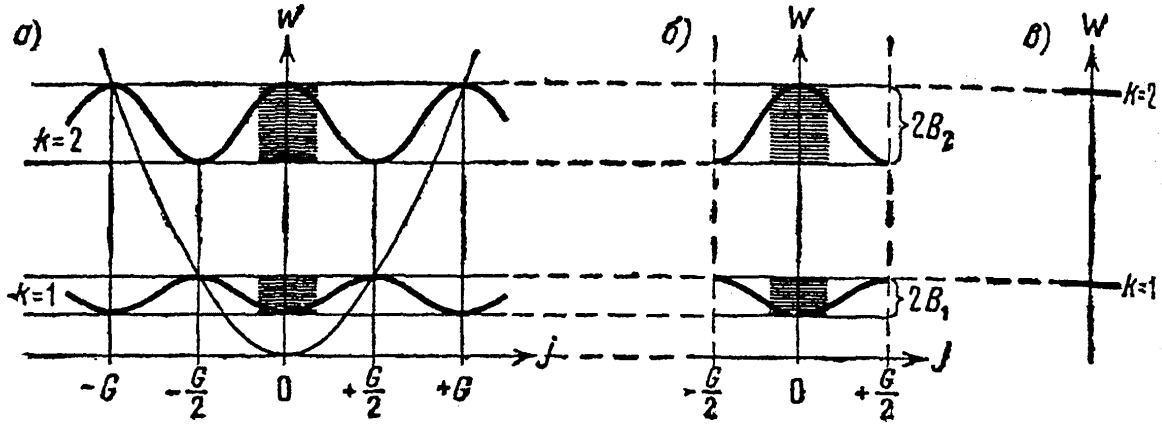


Рис. 30. Переход от свободных электронов к сильно связанным и от сильно связанных к абсолютно связанным.

к сильно связанным электронам энергетические уровни стягиваются в отдельные *зоны*, разделённые запрещёнными участками. Число k можно рассматривать как номер зоны, а число j (абсолютную величину j) как номер уровня внутри зоны. Так как j в (2,26) входит под знаком косинуса, то

$$W_k(j) = W_k(j+G),$$

где $G = \frac{L}{a}$ — число атомов решётки внутри карман-борновского периода L. Следовательно, можно ограничиться значениями j, лежащими в интервале длиной G, полагая, например (см. рис. 30, *б*):

$$j = 0, \pm 1, \pm 2, \ldots, \pm\frac{G}{2}.$$

Мы имеем для k-ой зоны на основании (2,23):

$$\begin{cases} \text{верхняя граница зоны:} & W_k^{\max} = \dfrac{h^2k^2}{8ma^2}; \\ \text{нижняя граница зоны:} & W_k^{\min} = \dfrac{h^2k^2}{8ma^2}\left(1-\dfrac{2}{P}\right)^2; \\ \text{ширина зоны:} & W_k^{\max} - W_k^{\min} = \dfrac{h^2k^2}{8ma^2}\dfrac{4}{P}. \end{cases}$$

Мы видим, что зона тем шире, чем больше k (т. е. чем выше лежит зона) и чем меньше P (т. е. чем свободнее электроны). При $P=\infty$ каждая зона вырождается в дискретный уровень, как это показано на рис. 30, *в*.

Отметим в заключение, что во все наши формулы входят только два параметра, характеризующие природу кристалла: постоянная решётки a и степень связанности электронов P. Химическая индивидуальность атомов, из которых построена решётка, и характерная для них система энергетических термов не находят своего отражения в энергетическом спектре кристалла. В этом заключается характерный недостаток модели Кронига-Пенни.

С. Метод Блоха. В методе Блоха [4] задача о поведении электрона в решётке ставится следующим образом. Мы исходим из изолированного атома. Поведение электрона внутри такого атома считается известным. Задача состоит в том, чтобы определить, что происходит с электроном и с его системой энергетических уровней под влиянием остальных атомов решётки. Действие этих последних рассматривается как слабое возмущение.

Будем предполагать, что мы имеем дело с неограниченной (трёхмерной) кубической решёткой. Перенумеруем атомы решётки. Номер атома будем характеризовать вектором $\vec{G}$ с целочисленными компонентами g_1, g_2, g_3. Числа g_1, g_2, g_3 определяют координаты ядра $\vec{G}$-го атома:

$$x_{\vec{G}}=ag_1,\quad y_{\vec{G}}=ag_2,\quad z_{\vec{G}}=ag_3,$$

где a — постоянная решётки.

Рассмотрим $\vec{G}$-й атом решётки, считая его изолированным от остальных. Положим, что все *внутренние* электроны атома вместе с ядром образуют положительно заряженный атомный остов, рассматриваемый как целое. В поле $U_{\vec{G}}(xyz)$ этого атомного остова движется каждый из *валентных* электронов независимо от остальных (строго говоря, такой метод приемлем лишь в случае одновалентных атомов). Пусть поведение валентного электрона описывается функцией $\varphi^{i}_{\vec{G}}(xyz)$. Эта функция удовлетворяет, очевидно, такому уравнению Шре-

дингера:

$$H_{\vec{G}}\,\varphi^{i}_{\vec{G}} = W_i \varphi^{i}_{\vec{G}}, \tag{2,27}$$

где

$$H_{\vec{G}} = -\frac{h^2}{8\pi^2 m}\Delta + U_{\vec{G}}(x, y, z). \tag{2,28}$$

Очевидно:

$$\begin{cases} U_{\vec{G}}(x, y, z) = U(x - g_1 a,\ y - g_2 a,\ z - g_3 a), \\ \varphi^{i}_{\vec{G}}(x, y, z) = \varphi^{i}(x - g_1 a,\ y - g_2 a,\ z - g_3 a). \end{cases} \tag{2,29}$$

Значок i представляет собой номер собственной функции и соответствующего ей собственного значения. Возможность вырождения мы учтём, если будем считать, что среди чисел $W^0, W^1, W^2, \ldots$ могут быть равные. Так, например, если считать, что $W^1 = W^2 = W^3$, то это будет означать, что функции $\varphi^1, \varphi^2, \varphi^3$ являются вырожденными (трёхкратное вырождение). Функции $\varphi^{i}_{\vec{G}}$ будем считать ортогональными и нормированными:

$$\varphi^{i}_{\vec{G}}\,\varphi^{i'}_{\vec{G}}\,d\tau = \begin{cases} = 1, & \text{если } i' = i, \\ = 0, & \text{если } i' \neq i. \end{cases}$$

Обратимся теперь к кристаллической решётке. Пусть поведение электрона в решётке описывается функцией $\psi(x, y, z)$. Эта функция удовлетворяет уравнению Шредингера (2,1):

$$H\psi = W\psi, \tag{2,30}$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V(x, y, z).$$

Считая, что коллективизированы лишь валентные электроны атомов, мы можем написать:

$$V(x, y, z) = \sum_{\vec{G}} U_{\vec{G}}(x, y, z) + S(xyz). \tag{2,31}$$

Здесь суммирование по значку $\vec{G}$ означает тройную сумму по всем значениям g_1, g_2, g_3. Таким образом, периодический потенциал V складывается из потенциалов положительных ионов (атомных остовов), расположенных в узлах решётки, и самосогласованного потенциала S.

Наша задача состоит в том, чтобы найти собственные функции и собственные значения уравнения (2,30), считая известными собственные функции и собственные значения уравнения (2,27). Эта задача рассматривалась, как задача на теорию возмущений, Блохом [4] (для s-состояния), Вильсоном [5] (для p-состояния) и Блохинцевым [6] (для d-состояния). Мы рассмотрим её как задачу на применение вариационного принципа, что, конечно, эквивалентно по существу теории возмущений.

Будем искать функцию ψ в виде линейной комбинации атомных функций $\varphi^{i}_{\vec{G}}$:

$$\boxed{\psi(x, y, z) = \sum_i \sum_{\vec{G}} a^{i}_{\vec{G}} \varphi^{i}_{\vec{G}}(x, y, z).} \qquad (2,32)$$

Здесь суммирование по значку i должно охватывать состояния с *близкими* значениями W_i, т. е. такие, для которых

$$|W_i - W_i'| \leqslant |W - W'|.$$

В частном случае, когда функция φ^i невырождена и когда терм W_i достаточно удалён от своих соседей, суммирование по значку i в (2,32) можно вовсе опустить.

Поскольку функция ψ задаётся нами в виде (2,32), задача сводится к отысканию коэффициентов $a^{i}_{\vec{G}}$. Эти коэффициенты должны быть подобраны так, чтобы функция (2,32) наилучшим возможным образом удовлетворяла уравнению (2,30). Для этого необходимо, чтобы интеграл

$$J = \int \bar{\psi} (H - W) \psi d\tau \qquad (2,33)$$

обращался в минимум. В (2,33) интегрирование производится по тому объёму, внутри которого ψ нормирована. Условие минимума интеграла J выражается системой линейных однородных уравнений:

$$\frac{\partial J}{\partial \bar{a}^{i}_{\vec{G}}} = 0.$$

Перепишем эту систему в раскрытом виде, полагая, как это следует из (2,27) и (2,30):

$$H = H_{\vec{G}} + V - U_{\vec{G}}.$$

Мы будем иметь:

$$J=\sum_{i}\sum_{i'}\sum_{\vec{G}}\sum_{\vec{G}'}\bar{a}^{i}_{\vec{G}}\,a^{i'}_{\vec{G}'}\times$$

$$\times\left\{(W_{i'}-W)\int\bar{\varphi}^{i}_{\vec{G}}\varphi^{i'}_{\vec{G}'}\,d\tau+\int\bar{\varphi}^{i}_{\vec{G}}(V-U_{\vec{G}})\,\varphi^{i'}_{\vec{G}},\,d\tau\right\}.$$

Принимая для сокращённости следующие обозначения:

$$A^{i\ i'}_{\vec{G}\ \vec{G}'}=\int\bar{\varphi}^{i}_{\vec{G}}\varphi^{i'}_{\vec{G}'}\,d\tau,$$

$$B^{i\ i'}_{\vec{G}\ \vec{G}'}=\int\bar{\varphi}^{i}_{\vec{G}}(V-U_{\vec{G}})\,\varphi^{i'}_{\vec{G}'}\,d\tau, \qquad (2,34)$$

получим

$$\boxed{\frac{\partial J}{\partial\bar{a}^{i}_{\vec{G}}}=\sum_{i'}\sum_{\vec{G}}a^{i'}_{\vec{G}'}\left\{(W_{i'}-W)A^{i\ i'}_{\vec{G}\ \vec{G}'}+B^{i\ i'}_{\vec{G}\ \vec{G}}\right\}=0.} \qquad (2,35)$$

Мы получаем бесконечную систему линейных однородных уравнений с бесконечным числом неизвестных. Неизвестными здесь служат коэффициенты $a^{i'}_{\vec{G}'}$.

Мы рассмотрим два частных случая, когда решения этой системы могут быть получены весьма просто.

а) Случай, когда исходное состояние атома является *s*-состоянием, т. е. невырождено. При этом будем предполагать, что наш *s*-терм достаточно удалён от соседних термов.

б) Случай, когда исходное состояние атома является *p*-состоянием. Опять-таки будем при этом предполагать, что наш *p*-терм достаточно удалён от соседних термов.

а) *s-состояние.* В этом случае суммирование по значку i в выражении (2,32) можно вовсе опустить. Функция $\varphi(x, y, z)$, будучи *s*-функцией, обладает сферической симметрией, т. е.

$$\varphi(x, y, z)=f(R).$$

Учитывая это обстоятельство, а также считая, что функции двух соседних атомов слабо перекрываются, мы можем

написать:

$$A_{\vec{G}\,\vec{G}} = 1\ldots, \text{ если } g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3,$$

$$A_{\vec{G}\,\vec{G}} = 0\ldots \text{ во всех остальных случаях.}$$

$$B_{\vec{G}\,\vec{G}} = \alpha^s\ldots, \text{ если } g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3,$$

$$B_{\vec{G}\,\vec{G}} = \beta^s\ldots, \text{ если } \begin{cases} g_1' = g_1 \pm 1,\ g_2' = g_2, g_3' = g_3 \\ g_1' = g_1,\ g_2' = g_2 \pm 1,\ g_3 = g_3 \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3 \pm 1, \end{cases}$$

$$B_{\vec{G}\,\vec{G}} = 0\ldots \text{ во всех остальных случаях.} \qquad (2{,}36)$$

Таким образом (2,35) перепишется так:

$$a_{g_1 g_2 g_3}(W^s - W + \alpha^s) + \\ + \beta^s(a_{g_1+1,\,g_2,\,g_3} + a_{g_1-1,\,g_2,\,g_3} + a_{g_1,\,g_2+1,\,g_3} + \\ + a_{g_1,\,g_2-1,\,g_3} + a_{g_1,\,g_2,\,g_3+1} + a_{g_1,\,g_2,\,g_3-1}) = 0. \qquad (2{,}37)$$

Решение этой системы будем искать в виде:

$$\boxed{a_{\vec{G}} = Ce^{i(\lambda_1 g_1 + \lambda_2 g_2 + \lambda_3 g_3)} = Ce^{i\vec{\Lambda}\,\vec{G}}.} \qquad (2.38)$$

Здесь λ_1, λ_2, λ_3 — любые вещественные или комплексные числа, которые можно рассматривать как компоненты вектора $\vec{\Lambda}$.

Подставляя (2,12) в (2,11), получим:

$$Ce^{i\vec{\Lambda}\,\vec{G}}\left\{W^s - W + \alpha^s + 2\beta^s(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3)\right\} = 0,$$

откуда

$$\boxed{W = W^s + \alpha^s + 2\beta^s(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3).} \qquad (2{,}39)$$

Подставляя (2,38) в (2,32), получим:

$$\boxed{\psi(x, y, z) = C\sum_{\vec{G}} e^{i\vec{\Lambda}\,\vec{G}}\varphi_{\vec{G}}(x, y, z).} \qquad (2{,}40)$$

Подставляя (2.39) в (2,8), получим:

$$\boxed{p_i = -\frac{4\pi a m}{h}\beta^s \sin\lambda_i,} \qquad (2{,}40\text{a})$$

где $i = 1, 2, 3$.

Числа λ_1, λ_2, λ_3 можно рассматривать как квантовые числа: они вполне определяют собственную функцию (2,40), соответствующее ей собственное значение (2,39) и средний импульс электрона (2,40 а). Заметим, что в нашем случае бесконечной решётки числа λ_1, λ_2, λ_3 обязательно должны быть вещественными, ибо в противном случае функция (2,40) обращалась бы в бесконечность на бесконечности. Зависимость W и p_l от λ_l изображена на рис. 31, *а*. Так как функция (2,40) является

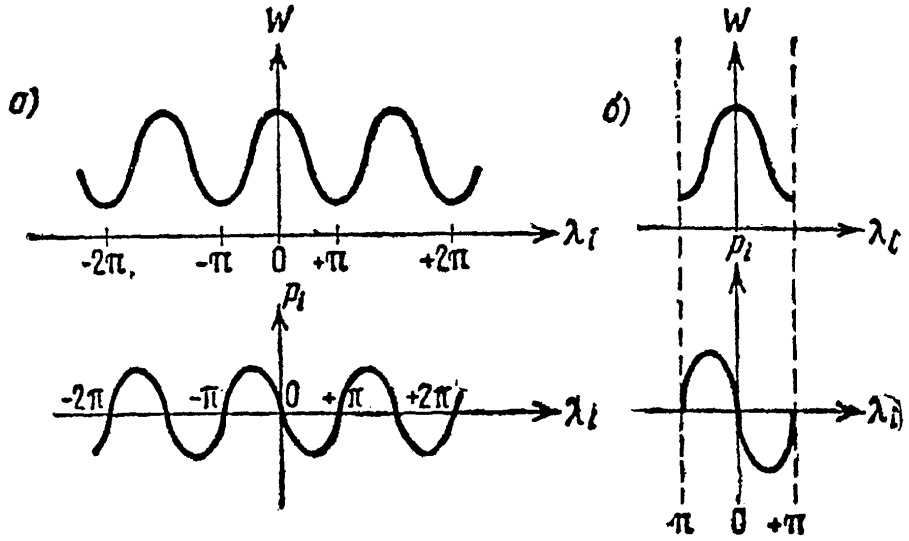


Рис. 31. Зависимость энергии W и импульса p_l от квантового числа λ_l.

периодической функцией от λ_1, λ_2, λ_3 с периодом 2π, то можно ограничиться значениями λ_l, лежащими в интервале шириной 2π, полагая, например,

$$-\pi \leqslant \lambda_l \leqslant +\pi$$

(см. рис. 31, *б*).

Заметим, что на функцию ψ можно наложить условие периодичности (условие Кармана-Борна). В этом случае параметры λ_1, λ_2, λ_3 делаются дискретными:

$$\lambda_l = \frac{2\pi a}{L} j_l,$$

где

$$j_l = 0, \pm 1, \pm 2, \ldots, \pm \frac{N}{2};$$

здесь $N = \frac{L}{a}$.

Заметим, что собственные значения (2,40), вообще говоря, вырождены: каждому собственному значению $W(\lambda_1, \lambda_2, \lambda_3)$ соот-

ветствует несколько собственных функций $\psi(\lambda_1, \lambda_2, \lambda_3, x, y, z)$. Это вырождение проистекает от двух причин: во-первых, от того, что $W(\lambda_1, \lambda_2, \lambda_3)$ является *чётной функцией* от $\lambda_1, \lambda_2, \lambda_3$, а во-вторых, от того, что $W(\lambda_1, \lambda_2, \lambda_3)$ является *симметричной* функцией относительно перестановки параметров $\lambda_1, \lambda_2, \lambda_3$, в то время как функция $\psi(\lambda_1, \lambda_2, \lambda_3; x, y, z)$ этими свойствами не обладает.

Мы видим из (2,39), что при переходе от изолированного атома к решётке терм W^s сдвигается на величину α^s и размывается в полосу (зону), как это показано на рис. 32. Мы имеем:

$$\begin{cases} \text{верхняя граница зоны} & W_{\max} = W^s + \alpha^s + 6\beta^s, \\ \text{нижняя граница зоны} & W_{\min} = W^s + \alpha^s - 6\beta^s, \\ \text{ширина зоны} & W_{\max} - W_{\min} = 12\beta^s. \end{cases}$$

Зона тем шире, чем больше обменный интеграл β^s, т. е. чем сильнее перекрываются функции $\varphi_{\vec{G}}$ и $\varphi_{\vec{G}'}$ двух соседних атомов. Это перекрытие тем больше, чем теснее расположены атомы в решётке (чем меньше a), а при заданном a — чем выше лежит уровень W^s. Отметим также, что это перекрытие волновых функций является источником нестрогости решения, так как мы использовали условие ортогональности функций φ_G, заведомо не выполняющееся для перекрывающихся функций.

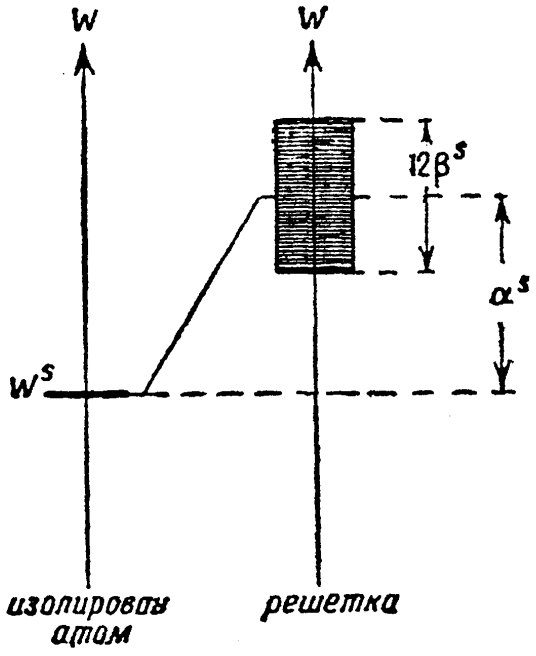


Рис. 32. Происхождение энергетической зоны.

Если параметры $\lambda_1, \lambda_2, \lambda_3$ непрерывны, то зона сплошная; если же эти параметры дискретны, то зона распадается на систему тесно расположенных дискретных уровней. Тройка чисел j_1, j_2, j_3 может служить номером уровня внутри зоны. Общее число состояний, содержащихся в зоне, равно числу атомов, заключённых внутри куба с ребром L.

Около верхней границы зоны (когда λ_i близки к нулю) и около нижней границы (когда λ_i близки к $\pm\pi$) косинусы в (2,39) можно разложить в ряд и ограничиться первыми членами разложения. Тогда выражение (2,39) примет такой вид:

$$W = W^s + \alpha^s \pm 6\beta^s \mp \beta^s (\lambda_1^2 + \lambda_2^2 + \lambda_3^2),$$

где верхний знак соответствует разложению около верхней границы, а нижний знак — разложению около нижней границы зоны. Сравнивая эту формулу с соответственной формулой (1,14) для свободных электронов:

$$W = \frac{h^2}{8\pi^2 a^2 m} (\lambda_1^2 + \lambda_2^2 + \lambda_3^2),$$

мы замечаем, что электрон, находящийся на дне зоны или поднятый под потолок зоны, может трактоваться как свободный электрон зоммерфельдовской теории с той лишь разницей, что ему следует приписать некоторую *эффективную массу* m^* вместо истинной массы m:

$$m^* = \mp \frac{h^2}{8\pi^2 a^2 \beta^s}.$$

Абсолютная величина этой эффективной массы тем меньше, чем шире зона. В частности, эффективная масса совпадает с истинной, когда ширина зоны равна 5 eV (при $a = 3 \cdot 10^{-8}$ *см*). Следует особо отметить, что эффективная масса электрона положительна на дне зоны, но отрицательна под потолком зоны. Это значит, что под потолком зоны электрон ведёт себя весьма своеобразным образом: он ускоряется в сторону, противоположную действующей на него силе.

б) *p-состояние.* p-состояние является трёхкратно вырожденным. Значок i в выражении (2,35) может, следовательно, принимать три значения ($i = 1, 2, 3$), причём

$$W^1 = W^2 = W^3 \equiv W^p$$

и

$$\begin{cases} \varphi^1(x, y, z) = x f(R), \\ \varphi^2(x, y, z) = y f(R), \\ \varphi^3(x, y, z) = z f(R), \end{cases}$$

где функции φ^1, φ^2, φ^3 соответствуют трём возможным значениям магнитного квантового числа m ($m = 0, +1, -1$).

Считая, что функции двух соседних атомов слабо перекрываются, мы можем написать:

$$A^{\,i\;i'}_{\vec{G}\,\vec{G}'} = \begin{cases} = 1 \ldots\ldots, \text{ если} \begin{cases} i' = i, \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3. \end{cases} \\ = 0 \ldots\ldots \text{ во всех остальных случаях.} \end{cases}$$

$$B^{\,i\;i'}_{\vec{G}\,\vec{G}'} = \begin{cases} = \alpha^p \ldots, \text{ если} \begin{cases} i' = i, \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3. \end{cases} \\ = \beta^p \ldots, \text{ если} \begin{cases} \begin{cases} i' = i = 1, \\ g_1' = g_1 \pm 1,\ g_2' = g_2,\ g_3' = g_3. \end{cases} \\ \begin{cases} i' = i = 2, \\ g_1' = g_1,\ g_2' = g_2 \pm 1,\ g_3' = g_3. \end{cases} \\ \begin{cases} i' = i = 3, \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3 \pm 1. \end{cases} \end{cases} \\ = \gamma^p \ldots, \text{ если} \begin{cases} \begin{cases} i' = i = 1, \\ g_1' = g_1,\ g_2' = g_2 \pm 1,\ g_3' = g_3, \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3 \pm 1. \end{cases} \\ \begin{cases} i' = i = 2, \\ g_1' = g_1 \pm 1,\ g_2' = g_2,\ g_3' = g_3, \\ g_1' = g_1,\ g_2' = g_2,\ g_3' = g_3 \pm 1. \end{cases} \\ \begin{cases} i' = i = 3, \\ g_1' = g_1 \pm 1,\ g_2' = g_2,\ g_3' = g_3, \\ g_1' = g_1,\ g_2' = g_2 \pm 1,\ g_3' = g_3. \end{cases} \end{cases} \\ = 0 \ldots \text{ во всех остальных случаях.} \end{cases}$$

Таким образом, (2,35) перепишется так:

$$\begin{aligned} &a^1_{g_1g_2g_3}(W^p - W + \alpha^p) + \beta^p(a_{g_1+1,\,g_2,\,g_3} + a^1_{g_1-1,\,g_2,\,g_3}) + \\ &+ \gamma^p\,(a^1_{g_1,\,g_2+1,\,g_3} + a^1_{g_1,\,g_2-1,\,g_3} + a^1_{g_1,\,g_2,\,g_3+1} + a^1_{g_1,\,g_2,\,g_3-1}) = 0, \\ &a^2_{g_1g_2g_3}(W^p - W + \alpha^p) + \beta^p\,(a^2_{g_1,\,g_2+1,\,g_3} + a^{\,2}_{g_1,\,g_2-1,\,g_3}) + \\ &+ \gamma^p\,(a^2_{g_1+1,\,g_2,\,g_3} + a^2_{g_1-1,\,g_2,\,g_3} + a^2_{g_1,\,g_2,\,g_3+1} + a^2_{g_1,\,g_2,\,g_3-1}) = 0, \\ &a^3_{g_1g_2g_3}(W^p - W + \alpha^p) + \beta^p\,(a^3_{g_1,\,g_2,\,g_3+1} + a^3_{g_1,\,g_2,\,g_3-1}) + \\ &+ \gamma^p(a^3_{g_1+1,\,g_2,\,g_3} + a^3_{g_1-1,\,g_2,\,g_3} + a^3_{g_1,\,g_2+1,\,g_3} + a^3_{g_1,\,g_2-1,\,g_3}) = 0. \quad (2,42) \end{aligned}$$

Решения этой системы будем искать в следующем виде:

$$\boxed{\begin{aligned} a^{1}_{\vec{G}} &= C_1 e^{i\vec{\Delta}\vec{G}}, \\ a^{2}_{\vec{G}} &= C_2 e^{i\vec{\Delta}\vec{G}}, \\ a^{3}_{\vec{G}} &= C_3 e^{i\vec{\Delta}\vec{G}}. \end{aligned}} \qquad (2,43)$$

Подставляя (2,43) в (2,42), получаем:

$$C_1 e^{i\vec{\Delta}\vec{G}}\left\{W^p - W + \alpha^p + 2\beta^p\cos\lambda_1 + 2\gamma^p(\cos\lambda_2 + \cos\lambda_3)\right\} = 0,$$

$$C_2 e^{i\vec{\Delta}\vec{G}}\left\{W^p - W + \alpha^p + 2\beta^p\cos\lambda_2 + 2\gamma^p(\cos\lambda_1 + \cos\lambda_3)\right\} = 0,$$

$$C_3 e^{i\vec{\Delta}\vec{G}}\left\{W^p - W + \alpha^p + 2\beta^p\cos\lambda_3 + 2\gamma^p(\cos\lambda_1 + \cos\lambda_2)\right\} = 0.$$

Здесь возможны следующие три случая:

1) $C_1 \neq 0$, $C_2 = C_3 = 0$. В этом случае мы имеем:

$$\boxed{\begin{aligned} W &= W^p + \alpha^p + 2\beta^p\cos\lambda_1 + 2\gamma^p(\cos\lambda_2 + \cos\lambda_3), \\ \psi(xyz) &= C_1\sum_{\vec{G}} e^{i\vec{\Delta}\vec{G}}\varphi^{1}_{\vec{G}}(xyz). \end{aligned}} \qquad (2,44а)$$

2) $C_2 \neq 0$, $C_1 = C_3 = 0$. В этом случае мы имеем:

$$\boxed{\begin{aligned} W &= W^p + \alpha^p + 2\beta^p\cos\lambda_2 + 2\gamma^p(\cos\lambda_1 + \cos\lambda_3), \\ \psi(xyz) &= C_2\sum_{\vec{G}} e^{i\vec{\Delta}\vec{G}}\varphi^{2}_{\vec{G}}(xyz). \end{aligned}} \qquad (2,44б)$$

3) $C_3 \neq 0$, $C_1 = C_2 = 0$. В этом случае мы имеем:

$$\boxed{\begin{aligned} W &= W^p + \alpha^p + 2\beta^p\cos\lambda_3 + 2\gamma^p(\cos\lambda_1 + \cos\lambda_2), \\ \varphi(xyz) &= C_3\sum_{\vec{G}} e^{i\vec{\Delta}\vec{G}}\varphi^{3}_{\vec{G}}(xyz). \end{aligned}} \qquad (2,44в)$$

В энергетическом спектре мы получаем зону, которую можно рассматривать как совокупность трёх зон [см. (2,44а, б, в)].

наложенных друг на друга. Нижние и верхние границы этих зон совпадают. Действительно:

$$\begin{cases} \text{верхняя граница зоны:} & W_{\max} = W^p + \alpha^p + 2\beta^p + 4\gamma^p, \\ \text{нижняя граница зоны:} & W_{\min} = W^p + \alpha^p - 2\beta^p - 4\gamma^p, \\ \text{ширина зоны:} & W_{\max} - W_{\min} = 4\beta^p + 8\gamma^p. \end{cases}$$

Наложением условий Кармана-Борна параметры $\lambda_1, \lambda_2, \lambda_3$ могут быть сделаны дискретными. При этом каждая зона распадается на систему дискретных уровней. Общее число состояний равно утроенному числу атомов, содержащихся в кубе с ребром L.

Напомним в заключение, что s- и p-состояния можно рассматривать раздельно только в том случае, когда расстояние между s- и p-термами велико по сравнению с шириной соответствующих зон. Если это условие не имеет места, то s- и p-состояния должны рассматриваться совместно. В этом случае в выражении (2,35) значок i имеет четыре значения ($i = 0, 1, 2, 3$). Вместо двух не связанных между собой систем уравнений (2,37) и (2,42) мы получаем одну общую систему из четырёх уравнений: зоны s и p вступают во взаимодействие друг с другом, в результате чего структура зон искажается. Этот случай был рассмотрен Блохинцевым [6].

Картина энергетического спектра схематически изображена на рис. 33.

D. Проводники и изоляторы. С точки зрения классических теорий различие между проводниками и изолятором сводилось к наличию или отсутствию в кристалле так называемых *свободных* электронов, т. е. таких, которые обладают способностью беспрепятственно передвигаться по всему объёму кристалла. Различие между *свободными* и *связанными* электронами (в классическом понимании этих терминов) может быть пояснено с помощью рис. 34. На этом рисунке изображён ход потенциала $V(x, y, z)$, т. е. то периодическое поле решётки, в котором движется электрон: потенциальные ямы чередуются с потенциальными барьерами. Полная энергия W электрона изображается на рис. 34 горизонтальным уровнем. Этот уровень может лежать *над* вершинами потенциальных барьеров (уровень AA') или *под* ними (уровень BB'). Соответственно этому следует различать два случая.

Если энергия электрона изображается уровнем AA', то мы имеем дело со *свободным* электроном. Такой электрон беспре-

пятственно переходит от атома к атому *поверх* разделяющих их барьеров.

Если же энергия электрона изображается уровнем BB', расположенным ниже вершин барьеров, то мы имеем дело со *связанным* электроном. С классической точки зрения такой электрон вынужден навсегда оставаться внутри своего атома; действительно, чтобы перейти в соседний атом, электрон должен пройти через запрещённую область (заштрихованную на чертеже), внутри которой его полная энергия меньше потенциальной, т. е. кинетическая энергия отрицательна. Атомы оказываются разделёнными стенками, совершенно непроницаемыми для классического электрона.

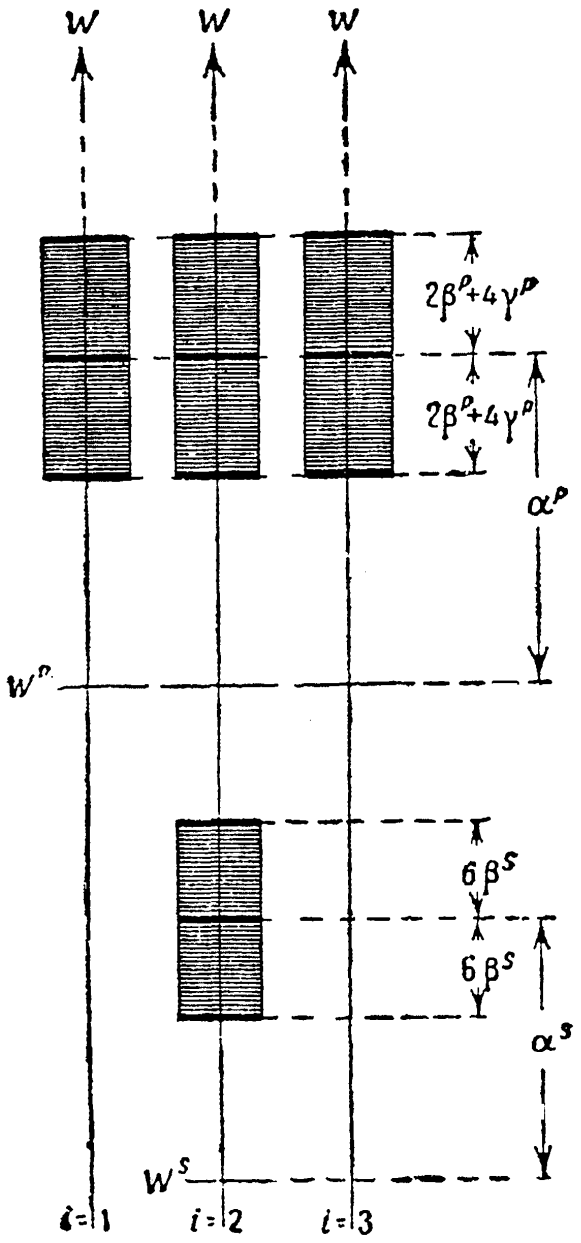


Рис. 33. Энергетический спектр электрона в решётке (W^s и W^p — соответствующие термы изолированного атома).

В квантовой теории резкое различие между свободными и связанными электронами стирается, поскольку квантовая механика позволяет электрону переходить из атома в атом не только *поверх* барьера, но и *сквозь* барьер. Если в классической теории вероятность такого перехода могла быть равна единице (свободный электрон) или нулю (связанный электрон), то теперь эта вероятность может принимать все промежуточные значения. Она тем больше, чем меньше площадь, заштрихованная на чертеже.

Таким образом, с точки зрения квантовой теории все электроны кристалла являются «свободными» в том смысле, что все они могут с той или иной степенью лёгкости пере-

двигаться по кристаллу. Дав электронам известную свободу передвижения, квантовая механика наложила, однако, другие ограничения на их движение. Различие между проводником и изолятором связано именно с этими новыми ограничениями.

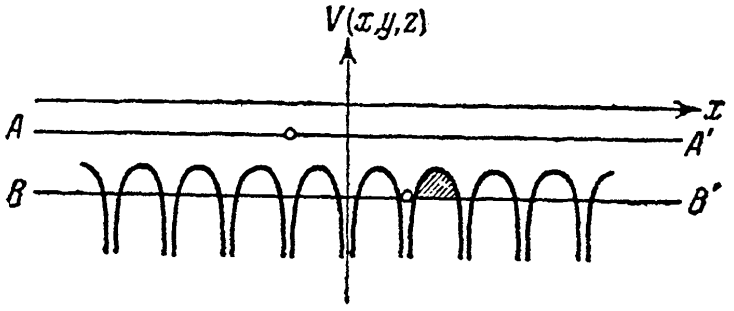


Рис. 34. Свободный и связанный электроны в кристалле.

Следует заметить, что способность электронов к передвижению является условием необходимым, но ещё недостаточным для того, чтобы решётка проводила. Действительно, если электроны движутся беспорядочно по всем направлениям, мы ещё не обнаруживаем электрического тока. Для осуществления проводимости необходимо, чтобы внешнее электрическое поле было бы в состоянии до известной степени регулировать это беспорядочное движение, вызывая необходимую асимметрию в распределении электронов по импульсам.

Это второе условие является необходимым дополнением к первому. В классических теориях *первое* из этих двух условий являлось критерием проводимости; в квантовой теории таким критерием является *второе* условие. Действительно, законы, управляющие движением электронов, таковы, что нарушение симметрии в распределении по импульсам не во всяком кристалле может быть осуществлено. Поэтому не всякий кристалл является проводником, хотя всякий кристалл содержит в себе электроны, свободно странствующие по решётке. Посмотрим, каково происхождение этого нового ограничительного закона, характерного для квантовой теории.

Рассмотрим электронный газ в кристалле. Он составляется из валентных электронов отдельных атомов. Все электроны нашего коллектива должны быть распределены по состояниям, характеризуемым тройками квантовых чисел j_1, j_2, j_3. При этом в одном и том же состоянии может находиться не более двух электронов (при учёте спина). Электрону в состоянии j_1, j_2, j_3 может быть приписана энергия $W(j_1, j_2, j_3)$, т. е. определённый уровень внутри энергетической зоны, и среднее значение импульса $\vec{p}(j_1, j_2, j_3)$.

Энергетическая зона содержит в себе ограниченное число состояний (уровней). Таким образом, она обладает определённой «ёмкостью», т. е. способна вместить в себя лишь ограниченное число электронов. При абсолютном нуле температуры электроны размещаются по состояниям (или, иначе говоря, по уровням зоны) максимально плотным образом, начиная заполнение зоны с наиболее низких её уровней, что

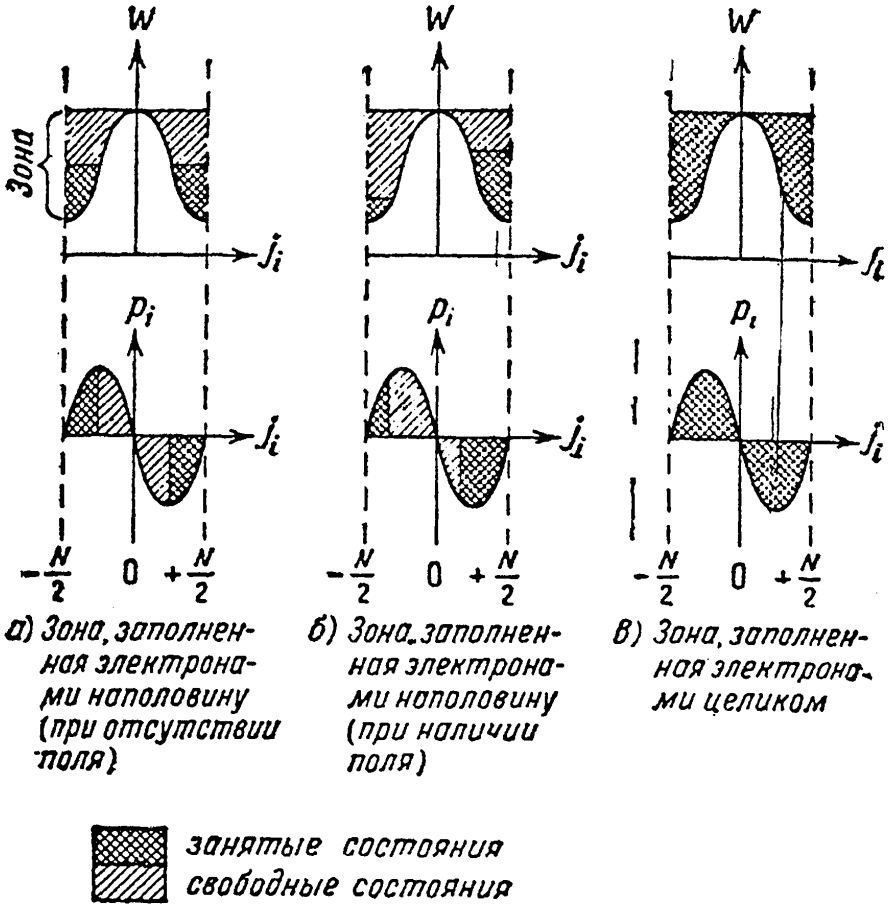


Рис. 35. Размещение электронов по состояниям j.

соответствует минимальному значению полной энергии системы. При этом зона может оказаться заполненной электронами частично или целиком. Следует различать эти два случая.

1. Представим себе сперва, что общее число коллективизированных электронов меньше, чем число состояний внутри зоны. Тогда при нуле температуры они расположатся на дне зоны, так что верхняя часть зоны останется свободной. Это иллюстрируется рис. 35, *а*, на котором изображены W и p_i как функции j_i, причём область значений j_i, занятая электронами, заштрихована двойной штриховой. Рис. 35 является повторением рис. 31, *б*; он относится, следовательно, к случаю, когда

зона, принимающая в себя электроны, является s-зоной. Мы видим, что распределение электронов по импульсам p_l, отвечающее максимально плотной упаковке, симметрично; ток в кристалле равен нулю, ибо каждому электрону с импульсом p_l может быть приведён в соответствие электрон с импульсом $p_{\bar{l}}$.

При наложении электрического поля электроны перераспределяются между уровнями так, что число электронов с положительными значениями j_l превышает число электронов с отрицательными значениями j_l (или наоборот). Это значит, что в одну сторону движется больше электронов, чем в противоположную. Рис. 35, *а* должен быть заменён теперь рис. 35, *б*, иллюстрирующим распределение электронов по состояниям j_l при наличии внешнего поля. Роль поля сводится к тому, что оно освобождает некоторые занятые уровни и заполняет кое-какие пустые, вводя характерную асимметрию, видную из сравнения рис. 35, *а* и 35, *б*.

Такое перемещение электронов с одних уровней на другие осуществимо, однако, лишь в том случае, когда по соседству с занятыми уровнями имеются пустые уровни. Это как раз имеет место в рассматриваемом случае, когда зона заполнена электронами лишь частично.

2. Представим себе теперь, что число электронов в решётке как раз равно числу состояний, содержащихся в зоне. В этом случае при самой плотной упаковке электроны заполняют все уровни зоны. Это иллюстрируется рис. 35, *в*, на котором заштрихована теперь *вся* область значений j_l. Для рассматриваемого случая существенно следующее дополнительное обстоятельство: соприкасается ли наша заполненная зона с вышележащей пустой зоной или отделена от неё запрещённым участком?

а) Если зона, заполненная электронами, соприкасается с пустой зоной, то эту пустую зону можно рассматривать как продолжение заполненной. В этом случае перераспределение электронов по импульсам может быть осуществлено за счёт уровней этой верхней зоны. Действительно, две такие соприкасающиеся зоны, из которых нижняя заполнена электронами, а верхняя свободна, можно рассматривать как одну зону, заполненную лишь наполовину.

б) Если же зона, заполненная электронами, отделена от вышележащей пустой зоны запрещённым участком, то в этом

случае перераспределение электронов по импульсам в пределах данной зоны оказывается невозможным. Единственно, что можно делать, это переставлять электроны один на место другого, но таким путём никогда нельзя получить желаемую асимметрию в распределении электронов по положительным и отрицательным значениям j_i (см. рис. 35, *в*). Внешнее поле оказывается не в силах как бы то ни было влиять на движение электронов, находящихся в данной зоне, и, следовательно, решётка оказывается нечувствительной к воздействию внешнего поля.

Мы приходим, таким образом, к следующим признакам, отличающим *проводник* от *изолятора* (см. рис. 36):

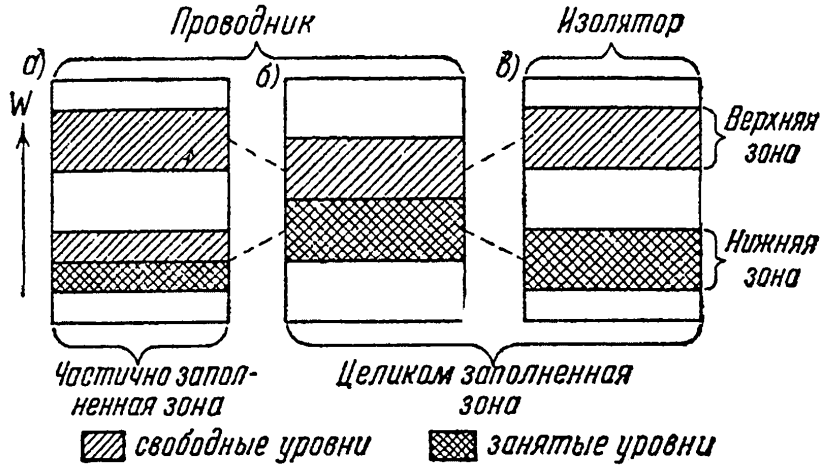


Рис. 36. Различие между проводником и изолятором.

I. Мы имеем дело с проводником, если зона, содержащая электроны, содержит также достаточное количество незанятых уровней или если она соприкасается с вышележащей пустой зоной.

II. Мы имеем дело с изолятором, если зона заполнена электронами целиком и при этом отделена запрещённым участком от вышележащей пустой зоны.

Заметим, что в результате внешних воздействий на решётку в изоляторе может быть искусственно создана некоторая проводимость. Посмотрим, каким образом это может быть осуществлено.

С точки зрения классической теории, для этого необходимо связанные электроны перевести в свободное состояние, т. е. поднять электроны с уровней BB' на рис. 34 на уровни AA', расположенные поверх потенциальных барьеров. Этот

процесс всегда связан с затратой энергии, как это видно из рис. 34.

С точки зрения квантовой теории, мы имеем дело с аналогичным процессом. Для создания в изоляторе проводимости необходимо перевести часть электронов из нижней заполненной зоны (так называемой *нормальной* зоны) в верхнюю пустую зону (так называемую зону *проводимости*). Этот процесс и здесь связан с затратой энергии, как это видно из рис. 36, *в*. Электроны, попавшие в зону проводимости,

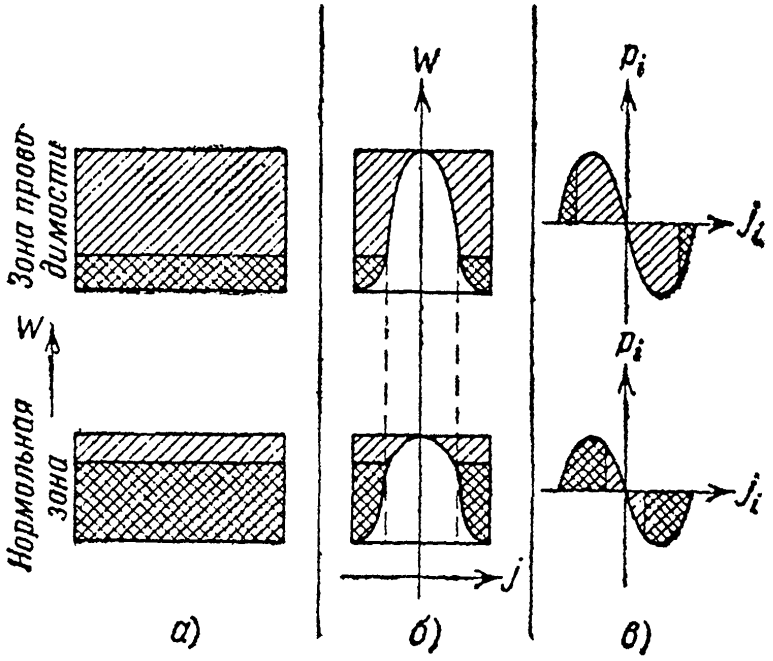


Рис. 37. Заполнение электронами нормальной зоны и зоны проводимости (при возбуждении).

естественно в полной мере приобретают способность переносить ток. При этом в нормальной зоне освобождаются некоторые уровни, и поэтому электроны, оставшиеся в нормальной зоне, тоже получают известную свободу. Число незаполненных мест (*дырок*) в нормальной зоне в точности равно числу электронов в зоне проводимости. Это иллюстрируется рис. 37.

Ток в таком возбуждённом кристалле складывается из *электронного* тока, обусловленного электронами, находящимися в зоне проводимости, и так называемого *дырочного* тока, обусловленного пустыми местами, содержащимися в нормальной зоне. Очевидно, ток будет тем больше, чем больше электронов переброшено в зону проводимости и чем больше, следовательно, дырок в нормальной зоне.

Внешние факторы, вызывающие проводимость в изоляторе, т. е. переводящие электроны из связанного состояния в свободное (с точки зрения классической картины) или поднимающие их из нормальной зоны в зону проводимости (с точки зрения квантовой картины), назовем *ионизирующими агентами*.

Примером такого агента может служить свет. В этом случае перевод электрона в зону проводимости происходит за счёт энергии светового кванта.

Другим примером является температура. При нагревании максимально плотная упаковка электронов внутри нормальной зоны начинает разрыхляться: электроны выбрасываются из неё в зону проводимости. Здесь этот процесс происходит за счёт тепловой энергии. Число электронов, поднятых температурой из нормальной зоны в зону проводимости, тем больше, чем выше температура и чем уже запрещённый участок между зонами. Кристаллы с достаточно широким запрещённым участком остаются идеальными изоляторами даже при максимальных экспериментально достижимых температурах. Кристаллы с достаточно узким запрещённым участком, так называемые *полупроводники*, обнаруживают заметную проводимость уже при комнатных температурах.

Таким образом, различие между изоляторами, полупроводниками и проводниками (металлами) имеет чисто количественное происхождение. Изолятор представляет собой один крайний случай полупроводника, в то время как металл является другим крайним случаем.

§ 3. Учёт периодического поля решётки (гетерополярная решётка)

В предыдущем параграфе мы имели дело с гомеополярной решёткой, построенной из одинаковых атомов. Однако подавляющее большинство полупроводников обладает гетерополярными решётками типа M_mR_r. Это — решётки, построенные из ионов двух сортов: положительных металлических и отрицательных металлоидных ионов. Задача о поведении электрона в такой решётке рассматривалась Слэтером [7] и Шокли [8].

В случае гетерополярной решётки потенциал $V(x, y, z)$, фигурирующий в уравнении Шредингера (1,9) или (2,1), сохраняет, конечно, свойство периодичности. Поэтому мы заранее можем ожидать, что энергетический спектр электрона в слу-

чае гетерополярной решётки будет иметь зональную структуру, как это имело место в гомеополярной решётке, поскольку такая структура не связана с какими-либо специальными предположениями о виде функции $V(x, y, z)$.

Однако представление об энергетических зонах в применении к гетерополярной решётке следует вновь проанализировать. Необходимость этого связана в конечном счёте с тем, что в случае гетерополярной решётки периодический потенциал $V(x, y, z)$ допускает различные интерпретации. Здесь возможны два подхода к задаче: можно идти по пути последовательного применения метода Блоха; возможен также и другой метод, более совершенный с точки зрения учёта взаимодействия электронов. Рассмотрим оба метода.

А. Распространение метода Блоха на случай гетерополярной решётки. Оставаясь в рамках метода Блоха, мы должны рассматривать нашу решётку состоящей из системы положительных ионов и семейства коллективизированных электронов. Коллективизированные электроны составляются из валентных электронов как металлических, так и металлоидных атомов. Среди положительных ионов мы имеем ионы двух сортов: положительные ионы металла и положительные ионы металлоида, получающиеся из соответствующих атомов в результате удаления с них всех валентных электронов. Электронный газ, с которым мы имеем здесь дело, обладает большой плотностью: на каждую кристаллическую ячейку приходится по несколько электронов. Каждый электрон движется в поле металлических и металлоидных положительных ионов, которые предполагаются неподвижными, и в самосогласованном поле электронов.

В качестве примера рассмотрим решётку CsCl. Решётка состоит из положительных ионов цезия с зарядом $+1$ и положительных ионов хлора с зарядом $+5$. Электронный газ составляется из валентных электронов как цезия, так и хлора: каждый атом цезия отдаёт в общественное пользование по одному электрону, а каждый атом хлора — по пять электронов. На каждую ячейку приходится, таким образом, по шесть коллективизированных электронов.

При таком подходе к задаче в энергетическом спектре электрона мы получаем систему зон, разделённых, вообще говоря, запрещёнными участками. Среди зон, фигурирующих в спектре, следует различать зоны двух типов.

а) Зоны, происшедшие от термов металлоидного атома. Такие зоны характеризуются тем, что модуль волновой функции для них, будучи периодичен с периодом, равным постоянной решётки, имеет максимумы в тех местах, где находятся атомы металлоида, и минимумы там, где находятся атомы металла.

б) Зоны, происшедшие от термов металлического атома. У таких зон модуль волновой функции, обладая той же периодичностью, даёт максимумы в металлических узлах решётки, а минимумы — в металлоидных узлах.

В энергетическом спектре перемешаны зоны того и другого типа. В частности, нормальная зона в одних кристаллах может быть зоной металлоидного типа (например, в кристалле NaCl), а в других — зоной металлического типа (например, в кристалле Cu_2O). В зависимости от того, к какому из этих двух типов принадлежит нормальная зона, мы приходим к двум разным представлениям о механизме дырочной проводимости. То же относится к зоне проводимости: она также может быть металлической или металлоидной. В зависимости от этого мы имеем два разных механизма электронной проводимости. Это будет пояснено в дальнейшем.

Заметим ещё, что в ионных решётках валентные электроны металлических атомов размещаются не в предназначенной для них металлической зоне, а в нижележащей металлоидной зоне, соответствующей нормальному терму металлоидного атома. Это утверждение есть не что иное, как переведённое на энергетический язык констатирование того факта, что атомы металла отдали свои валентные электроны атомам металлоида.

Поясним сказанное на двух конкретных примерах.

В качестве первого примера рассмотрим кристалл NaCl (или любой щёлочно-галоидный кристалл). В атоме натрия мы имеем следующее распределение электронов:

$$(1s)^2\,(2s)^2\,(2p)^6\,(3s)^1.$$

В атоме хлора:

$$(1s)^2\,(2s)^2\,(2p)^6\,(3s)^2\,(3p)^5.$$

В кристалле NaCl зона $3p$Cl лежит непосредственно под зоной $3s$Na (ширина запрещённого участка ~ 7eV). В зоне $3s$Na следовало бы разместить N электронов, где N — число молекул NaCl в единице объёма (по одному с каждого атома натрия). В зоне $3p$Cl размещаются $5N$ электронов (по пять

электронов с каждого атома хлора); при этом в этой зоне остаётся ещё N свободных мест. Естественно, что эти места заполняются валентными электронами натрия, перешедшими в зону $3p$Cl из более высокой зоны $3s$Na (см. рис. 38, *а*). Это приводит сразу к двум результатам: кристалл NaCl оказывается ионным кристаллом (потому что валентные электроны натрия размещаются в хлорной зоне); кристалл NaCl оказывается изолятором (потому что зона $3s$Na совершенно пустая, а зона $3p$Cl заполнена электронами до отказа). Зона $3s$Na играет роль зоны проводимости, а зона $3p$Cl представляет собой нормальную зону.

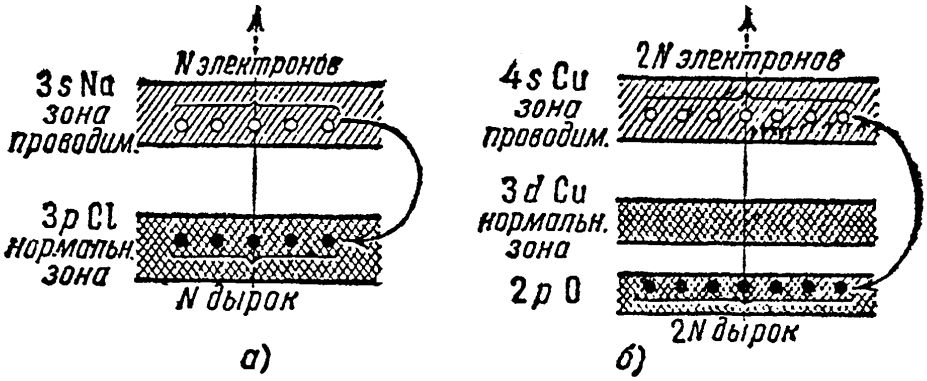


Рис. 38. Энергетический спектр электрона в гетерополярной решётке (метод Блоха). *а*) решётка NaCl, *б*) решётка Cu_2O.

Представим себе теперь, что в нормальной зоне образовалась дырка. Это значит, что в нашей решётке, построенной из ионов Na^+ и Cl^-, появился один нейтральный атом Cl. Дырка свободно странствует по нормальной зоне; это значит, что состояние нейтральности хлорного атома блуждает по кристаллу, передаваясь от одного иона хлора к другому, в то время как сами ионы остаются неподвижными. Создать дырку в нормальной зоне — это значит нейтрализовать ион хлора. С другой стороны, электрон в зоне проводимости свидетельствует о существовании нейтрального атома натрия. Находясь в зоне проводимости, электрон свободно путешествует по кристаллу, а вместе с ним путешествует состояние нейтральности атома натрия. Создать электрон в зоне проводимости — это значит нейтрализовать ион натрия. Если наложить электрическое поле, то нейтральное состояние атома Na движется к аноду и исчезает на нём, т. е. электрон уходит на анод по зоне проводимости (электронная проводимость): нейтральное

состояние атома Cl движется к катоду и исчезает на нём, т. е. дырка уходит на катод по нормальной зоне (дырочная проводимость).

В качестве второго примера рассмотрим кристалл Cu_2O. В атоме меди мы имеем следующее распределение электронов:

$$(1s)^2\,(2s)^2\,(2p)^6\,(3s)^2\,(3p)^6\,(3d)^{10}\,(4s)^1.$$

В атоме кислорода:

$$(1s)^2\,(2s)^2\,(2p)^4.$$

В кристалле Cu_2O непосредственно под зоной $4s$Cu лежит зона $3d$Cu (ширина запрещённого участка $\sim 1{,}5$eV); зона $2p$O лежит ещё ниже. Зона $4s$Cu содержит $2\cdot 2\,N = 4N$ мест, где N — число молекул Cu_2O в единице объёма; в ней следовало бы разместить $2N$ электронов (по одному с каждого атома Cu). Зона $3d$Cu содержит $10\cdot 2N = 20$ мест. В ней размещаются $20N$ электронов (по десять электронов с каждого атома Cu). Таким образом, эта зона целиком заполнена. Зона $2p$O содержит $6N$ мест; в ней размещаются $4N$ электронов (по четыре электрона с каждого атома O). В этой зоне остаётся $2N$ свободных мест, которые заполняются электронами, взятыми из зоны $4s$Cu (см. рис. 38, *б*). Таким образом, зона $2p$O оказывается заполненной до отказа, а зона $4s$Cu — совершенно пустой. Зона $4s$ Cu представляет собой зону проводимости; нормальной зоной является зона $3d$Cu.

Представим себе, что в нормальной зоне образовалась дырка. Это значит, что в нашей решётке, состоящей из ионов Cu^+ и O^{--}, появился ион Cu^{++}. С другой стороны, электрон, внесённый в зону проводимости, свидетельствует о появлении в решётке нейтрального атома Cu. Таким образом, если в кристалле Cu_2O течёт дырочный ток, то это значит, что состояние Cu^{++} движется по кристаллу в сторону катода, передаваясь от одного иона Cu^+ к соседнему.

Итак, мы видим, что в таких кристаллах, как NaCl, в которых нормальная зона является зоной металлоидного типа, дырочный ток представляет собой нечто физически совсем иное, чем, например, в кристалле Cu_2O, у которого нормальной зоной служит зона металлического типа.

В общем случае решётки типа

$$M_mR_r = mM^{+p} + rR^{-q},$$

состоящей из положительных металлических ионов с зарядом $+p$

и отрицательных металлоидных ионов с зарядом $-q$, мы получаем четыре возможных механизма проводимости, которые могут быть сведены в следующую таблицу.

Зона проводим.	Нормальная зона	Механизм электронн. проводимости	Механизм дырочной проводимости
Металлич.	Металлоидн.	$M^{+p-1} + M^{+p} \to \to M^{+p} + M^{+p-1}$	$R^{-q+1} + R^{-q} \to \to R^{-q} + R^{-q+1}$
Металлич.	Металлич.	$M^{+p-1} + M^{+p} \to \to M^{+p} + M^{+p-1}$	$M^{+p+1} + M^{+p} \to \to M^{+p} + M^{+p+1}$
Металлоидн.	Металлоидн.	$R^{-q-1} + R^{-q} \to \to R^{-q} + R^{-q-1}$	$R^{-q+1} + R^{-q} \to \to R^{-q} + R^{-q+1}$
Металлоидн.	Металлич.	$R^{-q-1} + R^{-q} \to \to R^{-q} + R^{-q-1}$	$M^{+p+1} + M^{+p} \to \to M^{+p} + M^{+p+1}$

Заметим в заключение, что разделение всех зон на два типа (*металлоидные* и *металлические* зоны), строго говоря, остаётся справедливым лишь до тех пор, пока зоны противоположных типов можно считать невзаимодействующими между собой, т. е. лишь до тех пор, пока соответствующие термы изолированных атомов достаточно удалены друг от друга.

В. Другой метод. Обратимся теперь к иному методу, которым мы и будем пользоваться в дальнейшем, более совершенному в отношении учёта взаимодействия электронов, чем метод Блоха. Будем представлять себе решётку составленной из положительных и отрицательных ионов. В поле этих ионов движутся коллективизированные электроны. Семейство коллективизированных электронов составляется из случайных электронов, внесённых в решётку извне или сорванных с отдельных ионов решётки. Плотность нашего электронного газа ничтожно мала по сравнению с плотностью электронного газа, с которым мы имели дело в методе Блоха. Поэтому взаимодействием электронов с полным правом можно совсем пренебречь. По этой же причине можно пренебречь искажающим влиянием аномальных ионов, присутствующих в решётке. Действительно, число таких аномальных ионов, отдавших свои внешние электроны в общественное пользование, не может превышать числа коллективизированных электронов.

Рассмотрим для определённости решётку типа CsCl, элементарной ячейкой которой является центрированный куб (см. рис. 39, *а*). Перенумеруем ячейки нашей решётки. Пусть $\vec{G}$ — номер ячейки. Каждая ячейка содержит два противоположных иона M^+ и R^- (см. рис. 39, *б*). Оба иона, принадлежащие

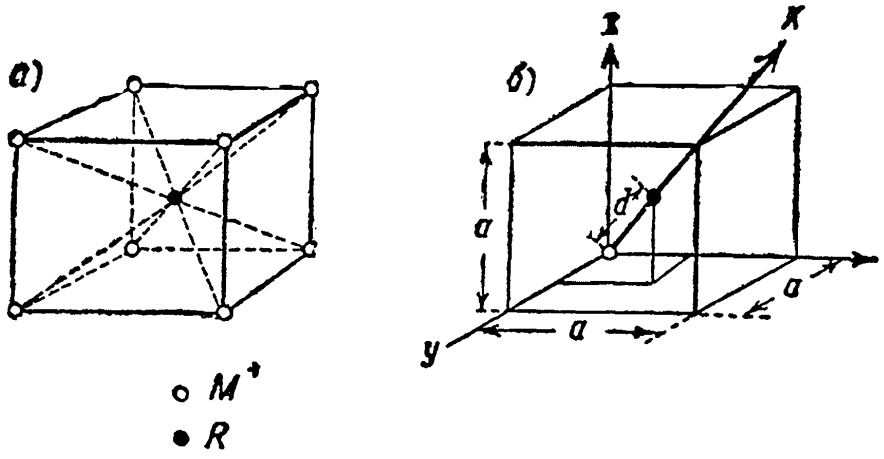


Рис. 39. *а*) — решётка типа CsCl, *б*) — элементарная ячейка типа CsCl.

одной и той же ячейке, имеют, следовательно, один и тот же номер $\vec{G}$. Если через $x\overset{M}{\vec{G}}, y\overset{M}{\vec{G}}, z\overset{M}{\vec{G}}$ обозначить координаты ядра $\vec{G}$-го металлического иона, а через $x\overset{R}{\vec{G}}, y\overset{R}{\vec{G}}, z\overset{R}{\vec{G}}$ — координаты ядра $\vec{G}$-го металлоидного иона, то

$$\left.\begin{aligned} x\overset{M}{\vec{G}} &= g_1 a, & y\overset{M}{\vec{G}} &= g_2 a, & z\overset{M}{\vec{G}} &= g_3 a, \\ x\overset{R}{\vec{G}} &= \left(g_1+\frac{1}{2}\right)a, & y\overset{R}{\vec{G}} &= \left(g_2+\frac{1}{2}\right)a, & z\overset{R}{\vec{G}} &= \left(g_3+\frac{1}{2}\right)a, \end{aligned}\right\}$$

где a — наикратчайшее расстояние между двумя одноимёнными ионами. Если через d обозначить наикратчайшее расстояние между двумя разноимёнными ионами, то, очевидно, для решётки типа CsCl между a и d существует соотношение

$$d=\frac{\sqrt{3}}{2}a.$$

В нашем случае потенциал $V(x, y, z)$ в уравнении Шредингера (2,1) имеет вид

$$V(xyz)=\sum_{\vec{a}} U\overset{M}{\vec{a}}(xyz)+\sum_{\vec{a}} U\overset{R}{\vec{a}}(xyz), \qquad (3,1)$$

причём

$$U\overset{M}{\underset{G}{\rightarrow}}(xyz) = U^M(x - g_1a,\ y - g_2a,\ z - g_3a),$$

$$U\overset{R}{\underset{G}{\rightarrow}}(xyz) = \qquad (3,2)$$

$$= U^R\left(x - g_1a - \frac{1}{2}a,\ y - g_2a - \frac{1}{2}a,\ z - g_3a - \frac{1}{2}a\right).$$

Рассматриваемый метод отличается от метода Блоха видом потенциалов U^M и U^R и отсутствием самосогласованного потен-

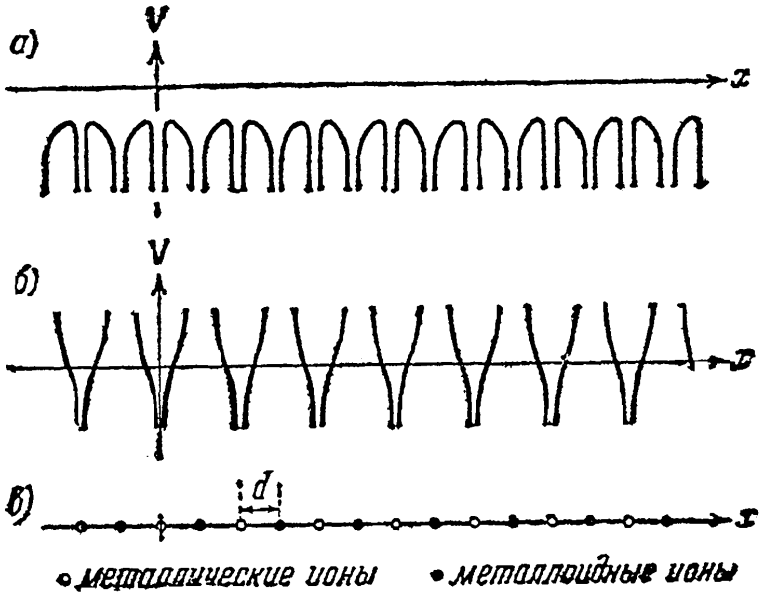


Рис. 40. Ход потенциала в идеальной гетерополярной решётке.

циала. С точки зрения метода Блоха, мы должны считать (для CsCl):

$$\left.\begin{aligned} U^M(xyz) &= -\frac{e^2}{R}, \\ U^R(xyz) &= -\frac{5e^2}{R}. \end{aligned}\right\} \qquad (3,3)$$

Ход потенциала изображён на рис. 40, *а*. С точки зрения рассматриваемого метода, мы имеем:

$$\left.\begin{aligned} U^M(xyz) &= -\frac{e^2}{R}, \\ U^R(xyz) &= +\frac{e^2}{R}. \end{aligned}\right\} \qquad (3,4)$$

Ход потенциала изображён на рис. 40, *б*. Выражения (3,3) и (3,4), конечно, справедливы лишь при достаточно больших R,

где

$$R = \sqrt{x^2 + y^2 + z^2},$$

т. е. до тех пор, пока ионы можно считать точечными.

Волновую функцию ψ будем искать, как и в предыдущем параграфе (§ 2, С), в виде:

$$\psi(xyz) = \sum_i \sum_{\vec{G}} a^i_{\vec{G}} \varphi^i_{\vec{G}}(xyz), \tag{3,5}$$

где теперь $\varphi^i_{\vec{G}}$ представляют собой волновые функции $\vec{G}$-го изолированного *металлического* атома, т. е. удовлетворяют уравнению:

$$H_{\vec{G}} \varphi^i_{\vec{G}} = W_i \varphi^i_{\vec{G}}, \tag{3,6}$$

где

$$H_{\vec{G}} = -\frac{h^2}{8\pi^2 m}\Delta + U^{\text{м}}_{\vec{G}}(xyz).$$

Очевидно:

$$\varphi^i_{\vec{G}}(x, y, z) = \varphi^i(x - g_1 a, y - g_2 a, z - g_3 a).$$

Мы можем повторить все рассуждения предыдущего параграфа (§ 2, С). Система уравнений, служащая для определения коэффициентов $a^i_{\vec{G}}$ в (3,5), имеет вид (2,35):

$$\sum_i \sum_{\vec{G}'} a^{i'}_{\vec{G}'} \left\{ (W_{i'} - W) A^{ii'}_{\vec{G}\,\vec{G}'} + B^{ii'}_{\vec{G}\,\vec{G}'} \right\} = 0, \tag{3,7}$$

где теперь

$$A^{ii'}_{\vec{G}\,\vec{G}'} = \int \overline{\varphi}^i_{\vec{G}} \varphi^{i'}_{\vec{G}'}\, d\tau;$$

$$B^{ii'}_{\vec{G}\,\vec{G}'} = \int \overline{\varphi}^i_{\vec{G}} (V - U^{\text{м}}_{\vec{G}'}) \varphi^{i'}_{\vec{G}'}\, d\tau.$$

Потенциал

$$V_{\vec{G}}(xyz) = V(xyz) - U^{\text{м}}_{\vec{G}}(xyz) \tag{3,8}$$

представляет собой потенциал в точке xyz, созданный всеми ионами решётки, за исключением $\vec{G}$-го (металлического) иона. Если положить в (3,8):

$$x = g_1 a, \quad y = g_2 a, \quad z = g_3 a,$$

мы получим потенциал решётки в $\vec{G}$-ом металлическом узле. Очевидно, если ионы считать точечными:

$$V_{\vec{G}}(g_1 a, g_2 a, g_3 a) = \frac{e^2}{d} M, \tag{3,9}$$

где M — потенциал Маделунга (для решёток типа CsCl: $M = 1{,}7627$).

Как и в предыдущем параграфе, мы имеем:

$$a^i_{\vec{G}} = C_i e^{i\vec{\lambda}\vec{G}}. \tag{3,10}$$

Для изолированного s-состояния значок i имеет одно значение ($i = 0$). Для изолированного p-состояния значок i принимает три значения ($i = 1, 2, 3$), причём среди трёх коэффициентов C_1, C_2, C_3 в (3, 10) лишь один может быть отличен от нуля. s-терм превращается в s-зону:

$$W(\lambda_1, \lambda_2, \lambda_3) = W^s + \alpha^s + 2\beta^s(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3). \tag{3,11}$$

p-терм порождает три p-зоны, накладывающиеся друг на друга:

$$W_1(\lambda_1, \lambda_2, \lambda_3) = W_2(\lambda_2, \lambda_1, \lambda_3) = W_3(\lambda_3, \lambda_2, \lambda_1),$$

где

$$W_1(\lambda_1, \lambda_2, \lambda_3) = W^p + \alpha^p + 2\beta^p \cos\lambda_1 + 2\gamma^p(\cos\lambda_2 + \cos\lambda_3). \tag{3,12}$$

Параметры α^s, β^s и α^p, β^p, γ^p, фигурирующие в (3,11) и в (3,12), определяются, согласно (2,36) и (2,41). Параметр

$$\alpha^i = \int \overline{\varphi^i_{\vec{G}}} V_{\vec{G}} \varphi^i_{\vec{G}}\, d\tau$$

имеет простой физический смысл: это есть потенциальная энергия электрона, принадлежащего $\vec{G}$-ому металлическому атому, в поле всех остальных ионов (за исключением данного $\vec{G}$-го иона). Электрон здесь мы представляем себе размазанным с плотностью $\overline{\varphi^i_{\vec{G}}}\varphi^i_{\vec{G}}$. Что касается параметров β^s, β^p, γ^p, то они не имеют простого физического смысла: это — «обменные» интегралы специфически квантовомеханического происхождения.

Если электрон, принадлежащий $\vec{G}$-ому металлическому атому, считать точечным, сосредоточенным в $\vec{G}$-ом металлическом узле,

т. е. если положить

$$\overline{\varphi}_{\vec{G}}\,\varphi_{\vec{G}} = \begin{cases} =1 \text{ в точке } x=g_1 a,\ y=g_2 a,\ z=g_3 a, \\ =0 \text{ вне этой точки} \end{cases}$$

и если также и ионы считать точечными, то квантовомеханические параметры α^s, β^s и α^p, β^p, γ^p превращаются в свои классические аналоги:

$$\left.\begin{matrix}\alpha^s \\ \alpha^p\end{matrix}\right\} \to \frac{e^2}{d} M,$$

$$\left.\begin{matrix}\beta^s \\ \beta^p \\ \gamma^p\end{matrix}\right\} \to 0.$$

Итак, в рассматриваемом методе свободный электрон с самого начала приписывается к металлическим ионам. Это обстоятельство содержится уже в самом виде волновой функции ψ [см. (3,5)]. Мы полўчаем для электрона в качестве энергетического спектра систему зон, соответствующих термам изолированного металлического атома (см. правую часть рис. 41). Наинизшая из этих зон является зоной проводимости. Таким образом, наличие в решётке свободного электрона означает, что один из металлических ионов решётки нейтрализован, причём этот нейтральный атом металла может быть в нормальном или в каком-либо из возбуждённых состояний, в зависимости от того, помещается ли наш электрон в зоне проводимости или в какой-либо из вышележащих электронных зон. Тот факт, что волновая функция электрона (3,5) составляется как линей-

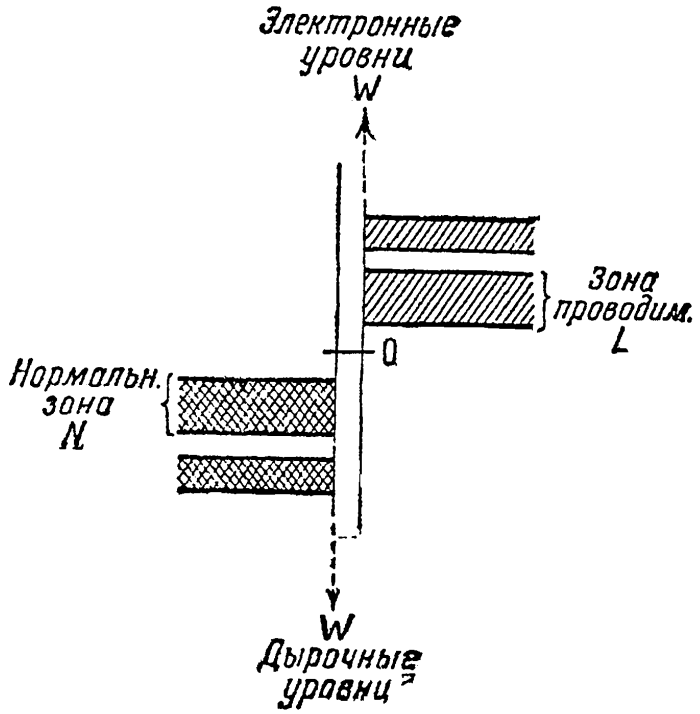


Рис. 41. Энергетический спектр электрона (дырки) в идеальной гетерополярной решётке.

ная комбинация из отдельных атомных функций, означает, что наш нейтральный металлический атом не локализован в решётке (электрон может сидеть на любом из металлических ионов).

Всё, что сказано здесь об электронах, может быть повторено по отношению к дыркам. Мы можем говорить по аналогии о «дырочном» газе, приписывая заранее дырки к отрицательным металлоидным ионам решётки. Для дырок мы получим энергетический спектр, представляющий собой систему зон, соответствующих термам изолированного металлоидного атома (см. левую часть рис. 41). «Электронный» и «дырочный» спектры могут быть совмещены на одном рисунке, если при этом электронные и дырочные уровни откладывать в противоположных направлениях, как это и сделано на рис. 41. (Действительно, введение дырки эквивалентно удалению электрона.) Наивысшая из дырочных зон является нормальной зоной. Таким образом, наличие в решётке свободной дырки означает, что один из металлоидных ионов нейтрализован, причём этот нейтральный атом металлоида может быть в нормальном или в каком-либо из возбуждённых состояний, в зависимости от того, находится ли наша дырка в нормальной зоне или в какой-либо из нижележащих дырочных зон.

Дырочные зоны можно рассматривать как зоны, «заполненные» электронами, т. е. как такие зоны, с которых можно черпать электроны. Наоборот, электронные зоны в этом смысле являются «пустыми», т. е. такими зонами, в которые можно вводить электроны. До тех пор, пока дырочные зоны не содержат дырок, а электронные не содержат электронов, наша решётка является идеальным изолятором. Для того чтобы создать в решётке проводимость, надо снабдить электронами зоны, изображённые в правой части рисунка 41 (электронная проводимость), или снабдить дырками зоны, изображённые в левой части рис. 41 (дырочная проводимость), или же осуществить и то и другое одновременно (смешанная проводимость). Механизм электронной и дырочной проводимости может быть изображён символическими формулами:

$$\left.\begin{array}{l} M + M^+ \rightarrow M^+ + M \ldots\ldots \text{ электронная проводимость,} \\ R + R^- \rightarrow R^- + R \ldots\ldots \text{ дырочная проводимость.} \end{array}\right\}$$

Такое представление о механизме проводимости скрыто уже в самой постановке нашей задачи.

Можно быть уверенным, что в энергетическом спектре наивысшая из дырочных зон (зона N на рис. 41) лежит ниже, чем наинизшая из электронных зон (зона L на рис. 41), т. е. что зона проводимости располагается над нормальной зоной. Действительно, если бы это было не так, то электроны из зоны N должны были бы переселиться в зону L, а это означало бы, что наше исходное представление о решётке, построенной из ионов M^+ и R^-, неправильно. Это означало бы, что такая структура решётки является энергетически невыгодной.

§ 4. Учёт взаимодействия электронов

А. Постановка задачи. Возвратимся к исходному уравнению Шредингера (1,1), описывающему поведение электронов в кристаллической решётке:

$$H\Psi = W\Psi, \tag{4,1}$$

где

$$H = \frac{h^2}{8\pi^2 m}\sum_i \Delta_i + \sum_i\sum_g U_{ig} + \frac{1}{2}\sum_i\sum_j P_{ij}. \tag{4,2}$$

Здесь члены U_{ig} соответствуют энергии взаимодействия электронов с ядрами, а члены P_{ij} выражают энергию взаимодействия электронов друг с другом.

Уравнение (4,1) не может быть решено точно. Это проистекает благодаря наличию в выражении (4,2) членов P_{ij}. Приходится довольствоваться приближённым решением. При этом речь может идти лишь о том, чтобы это приближение по возможности было бы наилучшим. В этом параграфе мы рассмотрим различные приближённые методы решения уравнения (4,1) и положение среди них метода самосогласованного поля, которым мы пользовались в § 2.

Обычно в качестве приближённого решения уравнения (4,1) исходят из мультипликативной функции

$$\Psi(s_1x_1y_1z_1;\ s_2x_2y_2z_2, \ldots) = \prod_i \psi^i(s_ix_iy_iz_i). \tag{4,3}$$

Представление волновой функции в виде (4,3) означает превращение нашей многоэлектронной задачи в задачу одноэлектронную, т. е. означает пренебрежение взаимодействием между электронами. Мультипликативная функция Ψ является вырожденной. Это вырождение — двойственного происхождения: оно составляется из *перестановочного* и *спинового* вырождения.

Перестановочное вырождение является следствием полной тождественности электронов. Действительно, переставляя в выражении (4,3) четвёрки координат s_i, x_i, y_i, z_i одну на место другой, мы получим волновые функции, которые должны удовлетворять уравнению Шредингера так же хорошо, как и исходная функция (4,3). Линейная комбинация из мультипликативных функций Ψ, отличающихся друг от друга перестановкой электронов, имеет вид:

$$\Phi = \sum_P a_P P \Psi, \tag{4,4}$$

где P — оператор перестановки, действующий на координаты электронов (как пространственные, так и спиновые), т. е. переставляющий нижние значки в (4,3). Суммирование в (4,4) производится по всем возможным перестановкам, так что сумма (4,4) состоит из $N!$ слагаемых, где N — общее число электронов.

Коэффициенты a_p в (4,4) можно подобрать так, чтобы функция Φ была антисимметричной относительно перестановки электронов. Для этого следует положить:

$$a_P = \begin{cases} = +C\ldots, & \text{если перестановка } P \text{ чётная,} \\ = -C\ldots, & \text{если перестановка } P \text{ нечётная.} \end{cases} \tag{4,5}$$

В этом случае функция (4,4) может быть записана в виде детерминанта (1,36). Обычно пользуются именно такой антисимметризованной волновой функцией Φ. Этим стирается индивидуальность электронов (см. § 1, D).

Функция Φ уже не обладает перестановочным вырождением, но ещё обладает спиновым. Переставляя в выражении (4,4) спины s_i один на место другого, мы будем получать новые функции, удовлетворяющие тому же уравнению Шредингера (4,1), что и исходная функция Φ, соответствующая заданному размещению спинов. Линейная комбинация из функций Φ, отличающихся перестановкой спинов, имеет вид

$$\chi = \sum_R b_R R \Phi, \tag{4,6}$$

где R — оператор перестановки, действующий на спины электронов, т. е. перестанавливающий неодинаковые спины между собой; b_R — коэффициенты, которыми мы вправе распоряжаться и которые следует подобрать наиболее выгодным образом (см. ниже). Суммирование в (4,6) производится по всем перестановкам неодинаковых спинов; если среди N спинов нам задано p

положительных и $n = N - p$ отрицательных, то сумма в выражении (4,6) содержит $\frac{N!}{n!\,p!}$ слагаемых. В функции χ уже использовано не только перестановочное, но и спиновое вырождение.

Мы будем в дальнейшем называть:

функции ψ^i — *индивидуальными* функциями,
функции Ψ — *мультипликативными* функциями,
функции Φ — *антисимметричными* функциями,
функцию χ — полной *коллективной* функцией.

Задача о решении уравнения (4,1) распадается на две задачи:

1) Задача о нахождении наилучших индивидуальных функций ψ^i, т. е. такой выбор этих функций, при котором антисимметричная функция Φ удовлетворяет уравнению Шредингера (4,1) наилучшим возможным образом.

2) Задача о нахождении наилучшей коллективной функции χ, при заданных индивидуальных функциях, т. е. такой выбор коэффициентов b_R в (4,6), при которых функция χ (построенная из заданных индивидуальных функций) наилучшим образом удовлетворяет уравнению (4,1).

В. Метод Хартри-Фока. Рассмотрим первую из этих двух задач. Она может быть решена с помощью вариационного принципа. Фок [9], пользуясь этим принципом, показал, что наилучшими индивидуальными функциями являются функции ψ^i, удовлетворяющие следующей системе дифференциальных уравнений:

$$\{H^0 + S(xyz) - W_i\}\,\psi^i(sxyz) - \sum_j A_{ji}(xyz)\psi^j(sxyz) = 0, \quad (4,7)$$

где

$$H^0 = -\frac{h^2}{8\pi^2 m}\Delta + \sum_g U_g(xyz),$$

$$A_{ji} = \sum_{s'} \int \overline{\psi}^j(s'x'y'z')\,P(xyz;\ x'y'z')\,\psi^i(s'x'y'z')\,d\tau', \quad (4,8)$$

$$S = \sum_j A_{jj} = \sum_j \sum_{s'} \int \overline{\psi}^j(s'x'y'z')\,P(xyz;\ x'y'z')\,\psi^j(s'x'y'z')\,d\tau'.$$

Здесь $U_g(xyz)$ представляет собой энергию взаимодействия g-го ядра с электроном, находящимся в точке x, y, z; $P(xyz;\ x'y'z')$ — энергию взаимодействия между двумя электронами, из которых один находится в точке x, y, z, а другой — в точке x', y', z'. Члены A_{ji}, входящие в уравнения (4,7), мы бу-

дем называть *обменными* потенциалами, а член S — *самосогласованным* потенциалом. Обменные потенциалы не имеют наглядного физического смысла; физический же смысл самосогласованного потенциала весьма прост; самосогласованный потенциал $S(xyz)$ выражает энергию взаимодействия электрона, находящегося в точке x, y, z, с размазанным зарядом всех электронов нашей системы, в том числе и со своим собственным размазанным зарядом.

Уравнения Фока (4,7) собственно и предназначены служить для определения индивидуальных функций. Обладая функциями ψ^i, полученными из уравнений Фока, мы можем быть уверены, что хотя с их помощью мы и не можем удовлетворить уравнению Шредингера (4,1) вполне точно, но зато мы ему удовлетворяем лучше, чем при помощи каких-либо других индивидуальных функций.

Однако уравнениями Фока практически никогда не пользуются для отыскания индивидуальных функций, ибо решение этих уравнений связано с большими математическими трудностями. Обычно идут по пути упрощения этих уравнений. На этом пути можно различать два последовательных шага.

Первый шаг состоит в пренебрежении обменными потенциалами, т. е. в вычёркивании из уравнений (4,7) членов A_{jl}, для которых $j \neq i$.

Уравнения Фока превращаются при этом в так называемые уравнения Хартри [10]:

$$\{H^0 + S(xyz) - A_{ll}(xyz) - W_i\}\,\psi^i(s, xyz) = 0. \qquad (4,9)$$

Эти уравнения представляют собой уравнения Шредингера для каждого отдельного электрона, хотя и неодинаковые для различных электронов (благодаря наличию члена A_{ll}). В этом — преимущество уравнений (4,9) по сравнению с уравнениями (4,7).

Благодаря наличию члена A_{ll} уничтожается действие электрона самого на себя.

Физический смысл того приближения, которое мы делаем, переходя от уравнений Фока к уравнениям Хартри, весьма прост. Оперирование с уравнениями Хартри означает, что взаимодействие данного электрона со всеми остальными заменяется его взаимодействием с некоторым внешним эффективным полем, созданным усреднённым зарядом этих электронов.

Заметим, что уравнение (4,9) было получено Хартри не из уравнений Фока, а независимо от Фока и раньше его. Зада-

ча Хартри, так же как и задача Фока, состояла в отыскании наиболее выгодных индивидуальных функций ψ^i. Хартри, как и Фок, решал эту задачу с помощью вариационного принципа. Однако Хартри, в отличие от Фока, исходил из мультипликативной функции Ψ, в то время как Фок пользовался антисимметризованной функцией Φ, что, конечно, правильнее. Вследствие этого дифференциальные уравнения для определения индивидуальных функций у Хартри и у Фока оказались разными [ср. (4,7) и (4,9)].

С. Метод самосогласованного поля. Для практического решения уравнений Хартри в них пренебрегают членом A_{ii}. Этот шаг не только не имеет достаточного обоснования, но и содержит в себе вполне очевидную непоследовательность. Действительно, пренебрегая членом A_{ii} в (4,9), мы, тем не менее, сохраняем этот член внутри самосогласованного потенциала S, как это видно из (4,8). Уравнения Хартри превращаются при этом в такие уравнения:

$$\{H^0 + S(x, y, z) - W_i\}\, \psi^i(s, x, y, z) = 0. \tag{4,10}$$

Считая, что спиновая и пространственная части волновой функции разделяются, т. е.

$$\psi^i(s; x, y, z) = u^i(s)\, \psi^i(x, y, z),$$

мы можем переписать (4,10) так:

$$H\psi^i(xyz) = W_i\, \psi^i(xyz), \tag{4,11}$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V(x, y, z) + S(x, y, z);$$

$$V(x, y, z) = \sum_g U_g(x, y, z)$$

— периодический потенциал решётки.

Уравнения (4,11) представляют собой уравнения Шредингера для каждого отдельного электрона, и притом одинаковые для всех электронов, так что индивидуальные функции ψ^i, приписываемые различным электронам, можно рассматривать как различные решения одного и того же уравнения Шредингера (4,11). В этом заключается преимущество уравнений (4,11) по сравнению с уравнениями (4,9). Конечно, уравнение (4,11), из которого исходят в методе самосогласованного поля, является результатом коренной порчи уравнений Фока. Индиви-

дуальные функции ψ^i, получаемые из этого уравнения, ни в коем случае не могут считаться наилучшими индивидуальными функциями.

Уравнение (4,11) решается, вообще говоря, методом последовательных приближений. Действительно, для решения уравнения (4,11) надо знать самосогласованный потенциал $S(xyz)$, а этот потенциал, в свою очередь, вычисляется уже на основании готовых решений ψ^i уравнения (4,11), ибо по определению (4,8):

$$S(x, y, z) = \sum_j e^2 \int \frac{|\psi^j(x' y' z')|^2 dx' dy' dz'}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}}. \quad (4,12)$$

Чтобы выйти из этого порочного круга, поступают следующим образом. Сперва функция $S(xyz)$ задаётся произвольно. Находятся решения уравнения (4,11), соответствующие такому произвольно заданному виду функции $S(xyz)$. Из полученных ψ^i строится функция $S(xyz)$ по формуле (4,12). После подстановки полученной функции $S(xyz)$ в уравнение (4,12) вновь ищутся решения этого уравнения и т. д. Так поступают до тех пор, пока не будут получаться такие функции $\psi^i(xyz)$, что уравнение (4,11) при $S(xyz)$, вычисленном с помощью этих функций, будет давать в качестве решений те же самые функции $\psi^i(xyz)$.

Заметим, что самосогласованный потенциал $S(xyz)$ обладает важным свойством: если потенциал решётки $V(xyz)$ является периодической функцией координат с периодом a, то и самосогласованный потенциал $S(xyz)$ также будет периодической функцией координат с тем же периодом a. Убедимся в этом. Будем искать решения уравнения (4,11), полагая сперва $S = \text{const}$. Тогда уравнение (4,11) будет представлять собой уравнение с периодическим потенциалом. Решения такого уравнения, как мы знаем (см. § 2 А), могут быть представлены в следующем виде:

$$\psi^j = (x, y, z) = e^{\frac{i}{a}\vec{\Delta}\vec{R}} u^j_{\vec{\Delta}}(x, y, z),$$

где

$$u^j_{\vec{\Delta}}(x + n_1 a, y + n_2 a, z + n_3 a) = u^j_{\vec{\Delta}}(x, y, z).$$

Поэтому, согласно (4,12):

$$S(x, y, z) = e^2 \sum_j \int \frac{|u^j_{\vec{\Lambda}}(x', y', z')|^2 dx' dy' dz'}{\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}}. \quad (4,13)$$

Такой самосогласованный потенциал уже обладает свойством периодичности. Действительно:

$$S(x+n_1a, y+n_2a, z+n_3a) =$$

$$= e^2 \sum_j \int \frac{|u^j_{\vec{\Lambda}}(x', y', z')|^2 dx' dy' dz'}{\sqrt{(x-x'+n_1a)^2+(y-y'+n_2a)^2+(z-z'+n_3a)^2}} =$$

$$= e^2 \sum_j \int \frac{|u^j_{\vec{\Lambda}}(x''+n_1a, y''+n_2a, z''+n_3a)|^2 dx'' dy'' dz''}{\sqrt{(x-x'')^2+(y-y'')^2+(z-z'')^2}} = S(x, y, z).$$

(Здесь мы пользовались вспомогательными переменными: $x'' = x' - n_1a$; $y'' = y' - n_2a$; $z'' = z' - n_3a$.) Если теперь подставить (4,13) в (4,11), то мы опять получим уравнение с периодическим потенциалом. Поэтому при последовательном уточнении $S(xyz)$ свойство периодичности будет сохраняться.

D. Метод Гейзенберга. Рассмотрим ещё один метод решения многоэлектронной проблемы. Этот метод, предложенный Гейзенбергом [11], применяется в некоторых задачах теории металлов. Он отличается специфическим выбором индивидуальных функций.

Припишем каждому электрону своё ядро, причём положим, что на каждое ядро (или на каждый атомный остов приходится по одному электрону. Таким образом, мы ограничиваемся решётками, построенными из одновалентных атомов. Возвратимся к уравнениям Фока (4,7), которые служат для определения индивидуальных функций, причём на этот раз выключим как взаимодействие электронов друг с другом, так и взаимодействие каждого данного электрона со всеми чужими ядрами. Тогда уравнения (4,7) превратятся в такие:

$$H_i\psi^i(s, xyz) = W_i\psi^i(s, xyz), \quad (4,14)$$

где

$$H_i = -\frac{h^2}{8\pi^2 m}\Delta + U_i(xyz). \quad (4,15)$$

Здесь i — номер электрона и номер соответствующего ядра.

Из уравнений (4,14) определяются индивидуальные функции в методе Гейзенберга. Каждому электрону соответствует своё уравнение. Легко заметить, что уравнения (4,14) представляют собой уравнения Шредингера для изолированных атомов. Действительно, заменяя значок i значком g и вспоминая, что (в случае кубической решётки):

$$U_g(x, y, z) = U(x - g_1 a, y - g_2 a, z - g_3 a),$$

мы будем иметь

$$\left.\begin{aligned} \psi^g(s, xyz) &= u_g(s)\,\varphi_g(xyz) = \\ &= u_g(s)\,\varphi(x - g_1 a, y - g_2 a, z - g_3 a), \\ W_g &= W^0. \end{aligned}\right\} \qquad (4,16)$$

Уравнения (4,14) при этом можно переписать так:

$$H_g \varphi_g = W^0 \varphi_g,$$

где

$$H_g = -\frac{h^2}{8\pi^2 m}\Delta + U_g(xyz).$$

Итак, в методе Гейзенберга в качестве индивидуальных функций пользуются *атомными* функциями (4,16). Иначе говоря, кристаллическая решётка в нулевом приближении рассматривается как система изолированных (т. е. невзаимодействующих) нейтральных атомов так, как будто расстояние между атомами достаточно велико. В нулевом приближении каждый электрон движется независимо от других в поле своего ядра; он как бы не замечает существования остальных электронов и всех ядер, за исключением своего собственного. Этим модель Гейзенберга отличается от модели Блоха, в которой взаимодействие каждого данного электрона со всеми чужими ядрами учитывается с самого начала.

Из индивидуальных функций (4,16) можно построить мультипликативную функцию (4,3), а затем полную коллективную функцию (4,6). Коэффициенты b_R в (4,6) остаются произвольными до тех пор, пока атомы достаточно раздвинуты, т. е. пока не включено взаимодействие. Но как только мы сдвинем атомы, т. е. включим взаимодействие, коэффициенты b_R сразу фиксируются. В гейзенберговском методе учёт взаимодействия скрывается именно в этих коэффициентах. Задача об отыскании коэффициентов b_R была решена до конца, и притом вполне точно, Бете [12] для случая линейной решётки и s-со-

стояния. Решение задачи в общем случае связано с большими математическими трудностями.

Мы видим, что в методе Гейзенберга поведение электронов в решётке рассматривается как многоэлектронная задача в отличие от метода самосогласованного поля, в котором эта задача сводится к одноэлектронной.

Заметим, что благодаря своей математической сложности метод Гейзенберга не получил широкого распространения в теории металлов. Заметим, кроме того, что этот метод не применим к теории электропроводности. Действительно, как можно показать, коллективная функция χ, построенная из индивидуальных функций Гейзенберга (из атомных функций), характеризуется отсутствием тока. Это связано с тем, что гейзенберговская коллективная функция χ содержит в себе лишь *неполярные* состояния, т. е. только такие состояния, которые соответствуют размещению электронов по одному на каждое ядро (на каждый атомный остов). Заметим, что если коллективную функцию χ строить не из гейзенберговских, а из блоховских индивидуальных функций (см. § 2, С):

$$\psi^i = C \sum_{\vec{G}} e^{i\vec{\Lambda}_i\vec{G}} \varphi_{\vec{G}},$$

то такая коллективная функция будет включать в себя, наряду с неполярными состояниями, также различные *полярные* состояния, характеризующиеся таким размещением электронов, когда на одни атомы посажены лишние электроны за счёт других атомов. При этом, как это можно показать [13], неполярная часть блоховской коллективной функции в точности совпадает с коллективной функцией Гейзенберга.

Литература к главе II

А. Общие руководства

1. Mott and Jones, The theory of the properties of Metals and Alloys, Oxford, 1936.
2. Mott and Gurney, Electronic processes in ionic crystals, Oxford, 1940.
3. Seitz, The modern theory of solids, Oxford, 1941.
4. Бете и Зоммерфельд, Электронная теория металлов, ОНТИ, 1938.
5. Вильсон, Квантовая теория металлов, ГТТИ, 1941.

6. Нардгейм, Теория металлического состояния, УФН, т. XV, вып. 5, 6, 7, 8, 1935.
7. Brillouin, La méthode du champ self-consistent, Paris, 1933.
8. Fröhlich, Electronentheorie der Metalle, 1936.

В. Цитированные оригинальные статьи

[1] Hill, *Acta Mat.*, **8**, 27, 1886.
[2] Jones and *Zener*, *Proc. of R. Soc.*, A, **144**, 101, 1934.
[3] Kronig and Penney, *Proc. of R. Soc.*, A, **130**, 499, 1931.
[4] Bloch, *Zs. f. Phys.*, **52**, 555, 1928.
[5] Wilson, *Proc. of R. Soc.*, A, **133**, 458, 1931.
[6] Blochinzew, *Sow. Phys.*, **5**, 316, 1934.
[7] Slater and Shockley, *Phys. Rev.*, **50**, 705, 1936.
[8] Shockley, *Phys. Rev.*, **50**, 754, 1938.
[9] Fock, *Zs. f. Ph.*, **61**, 126, 1930; **62**, 765, 1930.
[10] Hartree, *Proc. Cambr. Phil, Soc.*, **24**, 89, 111, 1928.
[11] Heisenberg, *Zs. f. Ph.*, **49**, 619, 1928.
[12] Bethe, *Zs. f. Ph.*, **71**, 205, 1931.
[13] Волькенштейн, *Уч. Зап. ГПИ им. Герцена*, Л., 1940.

ГЛАВА III

ЭЛЕКТРОННЫЕ УРОВНИ В РЕАЛЬНОМ КРИСТАЛЛЕ

§ 1. Постановка задачи

А. Происхождение локальных уровней. Переход от идеальной решётки к реальной находит своё отражение в структуре энергетического спектра. Всякий дефект, представляя собой нарушение периодической структуры решётки, приводит к возникновению в энергетическом спектре дополнительных дискретных уровней, расположенных в запрещённом участке между энергетическими зонами. Эти так называемые *локальные* уровни играют очень важную роль. Ими обусловливаются все специфические особенности полупроводников, с одной стороны, и щёлочно-галоидных кристаллов, с другой.

Локальные уровни, возникающие во всякой реальной решётке, характеризуются волновыми функциями с резко выраженными максимумами около мест повреждения и быстро спадающими по мере удаления от повреждения. Электрон, попавший на такой уровень, оказывается, таким образом, локализованным в более или менее узкой области пространства. Локальные уровни в спектре играют роль ловушек, способных захватывать электроны.

Заметим, однако, что такой захват электронов не всегда связан с дефектами в решётке. Застревание электрона может иметь место и в идеальной решётке, будучи вызвано не органическими нарушениями, присутствующими в решётке с самого начала, а теми деформациями, которые возникают в решётке благодаря самому факту наличия в ней свободного электрона. Действительно, образование свободного электрона сводится к созданию в решётке иона с аномальным зарядом (см. гл. I): вокруг такого аномального иона решётка, есте-

ственно, оказывается деформированной (в гл. I мы пренебрегали этой деформацией и эффектами, от неё проистекающими). Перемещение аномального состояния с одного иона на другой должно сопровождаться соответствующим перемещением всей области деформации. Таким образом, свободный электрон в процессе движения сам создаёт себе ловушки, в которых он может застрять, поскольку деформация решётки, возникающая вокруг свободного электрона, является фактором, локализующим его. Ловушки такого происхождения были рассмотрены Френкелем [1], а также Ландау [2].

В каждый данный момент число электронов, законсервированных в такого рода ловушках, ничтожно мало по сравнению с числом электронов, законсервированных на обычных уровнях прилипания, т. е. около органических повреждений решётки (пустые узлы, междуузельные атомы). Действительно, число электронов, застрявших в ловушках Френкеля, составляет лишь некоторую долю от общего числа свободных электронов, а число свободных электронов, в свою очередь, составляет лишь ничтожную часть от общего числа электронов, запасённых на локальных уровнях. Этим запасом, иначе говоря, числом органических повреждений, присутствующих в решётке, регулируются все основные свойства полупроводников, о чём свидетельствует резкая зависимость всех свойств от малейших изменений в стехиометрическом составе. Можно считать, что прилипания, обусловленные механизмом Френкеля, играют в полупроводниках второстепенную роль. Мы в дальнейшем будем игнорировать такого рода прилипания, пренебрегая вместе с тем деформациями, возникающими в решётке вокруг аномальных ионов.

Задача, стоящая перед нами, может быть сформулирована следующим образом. Рассмотрим уравнение Шредингера, описывающее поведение электрона в решётке

$$H\psi = W\psi,$$

где

$$H = -\frac{\hbar^2}{8\pi^2 m}\Delta + V(x, y, z).$$

Можно видеть, что в случае идеальной решётки потенциал $V(x\,y\,z)$ обладает важным свойством: он является *периодической* функцией от координат. В случае реальной решётки

потенциал $V(xyz)$ уже не является строго периодическим. Переход от идеальной решётки к решётке реальной (в частности, к решётке с нарушенным стехиометрическим соотношением) выражается в добавлении к строго периодическому потенциалу некоторой поправки $V'(xyz)$. Этот добавочный возмущающий потенциал $V'(xyz)$ может считаться практически отличным от нуля лишь в непосредственной близости от повреждения. Вид этого потенциала $V'(xyz)$ зависит от природы повреждений в решётке. Задача состоит в выяснении того, как появление возмущающего потенциала $V'(xyz)$ сказывается на структуре энергетического спектра.

Мы рассмотрим в настоящей главе решётки с микроскопическими повреждениями двух сортов:

а) решётки, из которых удалена часть атомов (решётки с пустыми узлами);

б) решётки, в которые введены избыточные атомы (решётки с междуузельными атомами).

При этом мы ограничимся ради математической простоты рассмотрением одномерных моделей таких решёток. Результаты в качественном отношении сохраняют свою силу и при переходе к трёхмерному случаю.

Предварительно, в качестве простейшего примера реальной решётки, рассмотрим решётку, ограниченную с одной стороны. Наличие ограничивающей поверхности уже само по себе является нарушением периодической структуры решётки. В энергетическом спектре появляются так называемые *поверхностные уровни* (уровни Тамма), представляющие собой простейший пример локальных уровней. Задача впервые была рассмотрена Таммом [3], а затем рядом других авторов. [4], [5], [6]. Тамм ограничился одномерным случаем и притом оперировал со специальной моделью Кронига-Пенни. Мы также рассмотрим одномерный случай, но не будем ограничивать себя моделью Кронига-Пенни.

В. Поверхностные уровни Тамма. Рассмотрим цепочку, состоящую из совершенно одинаковых атомов (одномерная модель гомеополярной решётки). Пусть атомы расположены на расстоянии a друг от друга. Будем считать цепочку неограниченной справа и обрывающейся слева. Перенумеруем атомы. Пусть g — номер атома. Мы будем считать, что g может принимать следующие значения:

$$g = 1, 2, 3, \ldots$$

Уравнение Шредингера для электрона, принадлежащего такой решётке, имеет вид:

$$H\psi = W\psi, \tag{1,1}$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V^*(xyz).$$

Здесь потенциал V^* складывается из потенциалов всех ионов U_g и самосогласованного потенциала S всех электронов:

$$V^*(xyz) = \sum_{g=1}^{\infty} U_g(xyz) + S(xyz),$$

причём

$$U_g(x, y, z) = U(x - ga, y, z). \tag{1,2}$$

Заметим, что потенциал V^* не является периодической функцией от координаты x. Он может считаться таковой лишь

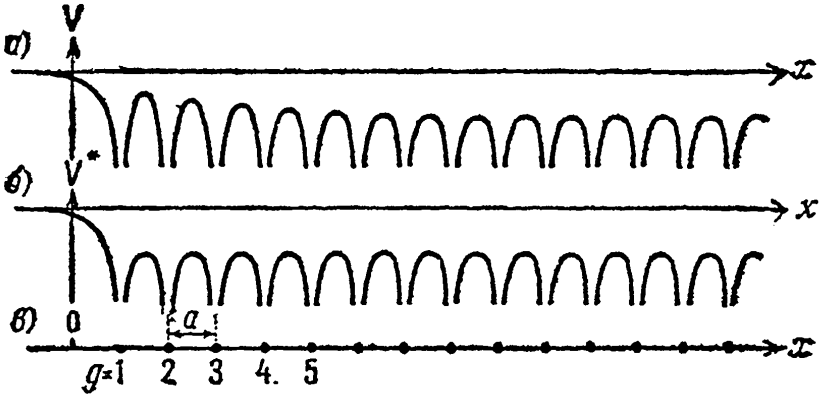


Рис. 42. Ход потенциала в гомеополярной решётке, ограниченной с одной стороны.

приближенно, при достаточно больших x (рис. 42, *а*). Однако мы можем положить:

$$V^*(xyz) = V(xyz) - V'(xyz), \tag{1,3}$$

где V — периодический потенциал, соответствующий случаю идеальной неограниченной решётки, а V' — добавка, нарушающая периодичность и вызванная наличием границы. Очевидно,

$$V'(xyz) \to 0 \text{ при } x \to \infty.$$

Мы будем считать, что эта добавка V' практически отлична от нуля лишь в области первого атома ($g = 1$), так что уже второй и все последующие атомы ($g = 2, 3, 4, \ldots$) не чув-

ствуют наличия границы. Иначе говоря, мы будем считать (в этом состоит наша апроксимация), что

$$\left.\begin{aligned} V^* &\approx V \ldots\ldots\ldots\ldots \text{ при } x > a, \\ V^* &= V - V' \ldots\ldots \text{ при } x \leqslant a. \end{aligned}\right\} \tag{1,4}$$

Это значит, что истинный ход потенциала, изображённый на рис. 42, *а*, мы заменяем апроксимированным ходом, изображённым на рис. 42, *б*.

Будем искать решения уравнения (1,1) в следующем виде:

$$\boxed{\psi(x, y, z) = \sum_{g=1}^{\infty} a_g \varphi_g(x, y, z),} \tag{1,5}$$

где φ_g — индивидуальная функция g-го атома, если его рассматривать изолированно. Функция φ_g удовлетворяет следующему уравнению:

$$H_g \varphi_g = W_0 \varphi_g,$$

где

$$H_g = -\frac{h^2}{8\pi^2 m}\Delta + U_g(x, y, z). \tag{1,6}$$

Мы будем предполагать для простоты, что состояние φ не вырождено, т. е. что функция φ есть s-функция. На основании (1,2) мы имеем:

$$\varphi_g(x, y, z) = \varphi(x - ga, y, z).$$

Задача состоит в отыскании коэффициентов a_g. Составим вариационный интеграл:

$$J = \int \overline{\psi}(H - W)\psi\, d\tau.$$

Подставляя сюда (1,5) и полагая, на основании (1,1), (1,3) и (1,6):

$$H = H_g + V^* - U_g,$$

мы получаем:

$$J = \sum_{g=1}^{\infty}\sum_{g'=1}^{\infty} \overline{a}_g a_{g'} \left\{(W_0 - W)\int \overline{\varphi}_g \varphi_{g'}\, d\tau + \int \overline{\varphi}_g (V^* - U_{g'})\varphi_{g'}\, d\tau\right\}.$$

Будем искать минимум этого интеграла. Принимая для сокращённости следующие обозначения:

$$A_{gg'} = \int \bar{\varphi}_g \varphi_{g'}\, d\tau,$$
$$B_{gg'} = \int \bar{\varphi}_g (V^* - U_{g'})\, \varphi_{g'}\, d\tau, \qquad (1,7)$$

получим такую систему уравнений:

$$\frac{dJ}{d\bar{a}_g} = \sum_{g'=1}^{\infty} a_{g'} \{(W_0 - W) A_{gg'} + B_{gg'}\} = 0. \qquad (1,8)$$

Рассмотрим интегралы $A_{gg'}$ и $B_{gg'}$, фигурирующие в этих уравнениях. Так как функции φ_g обладают сферической симметрией и слабо перекрываются, а также на основании нашего предположения (1,4) можно приближённо считать:

$$\left.\begin{aligned} A_{gg'}&\begin{cases} = 1 \ldots\ldots, & \text{если } g' = g \\ = 0 \ldots\ldots, & \text{если } g' \neq g, \end{cases} \\ B_{gg'}&\begin{cases} = \alpha \ldots\ldots, & \text{если } g' = g = 2, 3, 4, \ldots \\ = \alpha' \ldots\ldots, & \text{если } g' = g = 1 \\ = \beta \ldots\ldots, & \text{если } \begin{cases} g' = g + 1 = 2, 3, 4, 5, \ldots \\ g' = g - 1 = 1, 2, 3, 4, \ldots \end{cases} \\ = 0 \ldots\ldots & \text{во всех остальных случаях.} \end{cases} \end{aligned}\right\} \qquad (1,9)$$

Наша система (1,8) принимает такой вид:

$$\left.\begin{aligned} &\text{для } g = 2, 3, 4, \ldots \\ &\qquad a_g (W_0 - W + \alpha) + \beta (a_{g-1} + a_{g+1}) = 0, \\ &\text{для } g = 1 \\ &\qquad a_1 (W_0 - W + \alpha') + a_2 \beta = 0. \end{aligned}\right\} \qquad (1,10)$$

Решение этой системы будем искать в следующем виде:

$$\boxed{a_g = A e^{i\lambda g} + B e^{-i\lambda g}.} \qquad (1,11)$$

Подставляя (1,11) в (1,10):

$$\begin{aligned} &(A e^{i\lambda g} + B e^{-i\lambda g})(W_0 - W + \alpha + 2\beta \cos\lambda) = 0, \\ &A e^{i\lambda} (W_0 - W + \alpha' + \beta e^{i\lambda}) + \\ &\qquad + B e^{-i\lambda} (W_0 - W + \alpha' + \beta e^{-i\lambda}) = 0. \end{aligned} \qquad (1,12)$$

При решении этой системы следует различать три случая:

а) $A \neq 0$, $B \neq 0$. Этот случай приводит к представлению о зонах.

б) $A \neq 0$, $B = 0$ } Эти случаи приводят к представлению
в) $A = 0$, $B \neq 0$ } о локальных уровнях.

Рассмотрим все три случая.

Случай а. Будем искать решения системы (1,12), полагая $A \neq 0$, $B \neq 0$. Первое из уравнений (1,12) даёт:

$$\boxed{W = W_0 + \alpha + \beta \cos \lambda.} \qquad (1,13)$$

Подставляя (1,13) во второе уравнение (1,12):

$$Ae^{i\lambda}(\alpha' - \alpha - \beta e^{-i\lambda}) + Be^{-i\lambda}(\alpha' - \alpha - \beta e^{i\lambda}) = 0. \qquad (1,14)$$

Итак, коэффициенты A и B в (1,11) оказываются связанными, в то время как в неограниченной решётке они оставались свободными и независимыми. На основании (1,14) выражение (1,11) принимает вид:

$$\boxed{a_g = C\{\beta \sin \lambda g + (\alpha' - \alpha) \sin \lambda (g-1)\},} \qquad (1,15)$$

где

$$C = \frac{2iA}{(\alpha' - \alpha)\, e^{-i\lambda} - \beta} = \frac{-2iB}{(\alpha' - \alpha)\, e^{i\lambda} - \beta}.$$

Параметр λ может быть каким угодно, но обязательно вещественным. Если бы λ было мнимым или комплексным, то функция ψ обращалась бы на бесконечности в бесконечность, что непосредственно видно из (1,15). Итак, волновым функциям (1,15) (λ — вещественное!) соответствуют уровни, располагающиеся [согласно (1,13)] внутри границ:

$$\begin{cases} W_{\max} = W_0 + \alpha + 2\beta, \\ W_{\min} = W_0 + \alpha - 2\beta, \end{cases}$$

и образующие энергетическую зону (см. рис. 43).

Случаи б и в. Предположим теперь, что один из коэффициентов, A или B, в (1,11) равен нулю. Положим для определённости $A \neq 0$, $B = 0$ (случай б). В этом случае λ не обязано быть вещественным для того, чтобы функция ψ оставалась всюду конечной. Этим наша ограниченная решётка отличается от решётки неограниченной. Действительно, как

мы сейчас убедимся, конечность ψ обеспечивается теперь в том случае, когда

$$\lambda = r\pi + i\lambda'', \quad (1,16)$$

где r — целое число, и

$$\lambda'' > 0.$$

Первое уравнение (1,12) даёт попрежнему:

$$W = W_0 + \alpha + 2\beta\cos\lambda. \quad (1,17)$$

Второе уравнение (1,12) в нашем случае даёт:

$$W_0 - W + \alpha' + \beta e^{i\lambda} = 0$$

или, подставляя (1,13) в (1,17):

$$\alpha' - \alpha - \beta e^{-i\lambda} = 0,$$

откуда

$$e^{-i\lambda} = \frac{\alpha' - \alpha}{\beta}. \quad (1,18)$$

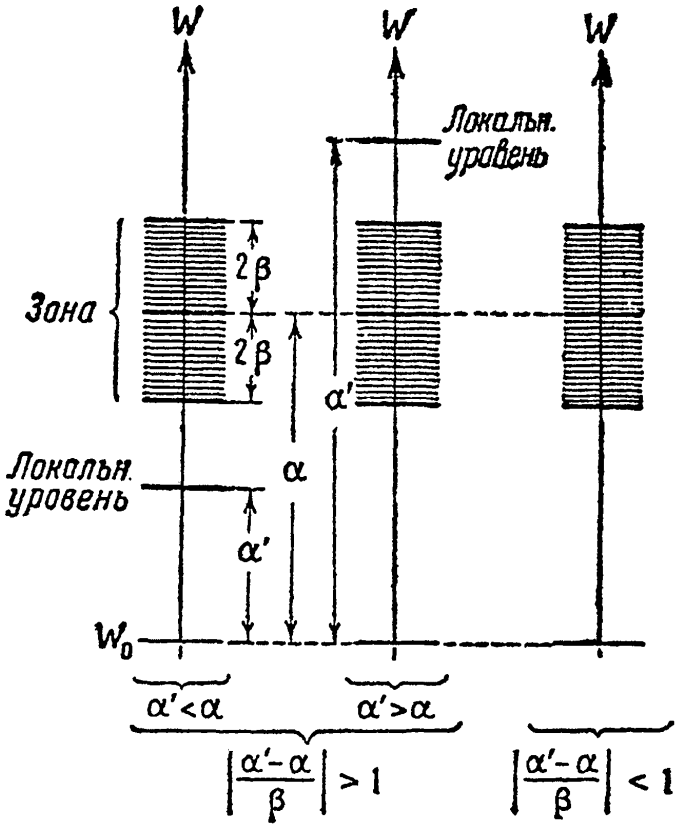


Рис. 43. Положение поверхностного уровня в энергетическом спектре (гомеополярная решётка, ограниченная с одной стороны).

Так как α, α', β суть вещественные числа, то уравнению (1,18) можно удовлетворить только в том случае, если λ имеет вид (1,16):

$$(-1)^r e^{\lambda''} = \frac{\alpha' - \alpha}{\beta},$$

причём

$$\begin{cases} \text{если } \dfrac{\alpha' - \alpha}{\beta} > 0, \text{ то } r \text{ должно быть чётным;} \\ \text{если } \dfrac{\alpha' - \alpha}{\beta} < 0, \text{ то } r \text{ должно быть нечётным.} \end{cases}$$

Волновая функция имеет вид:

$$\boxed{\psi = \sum_{g=1}^{\infty} A e^{i\lambda g}\varphi_g = A\sum_{g=1}^{\infty}\left|\frac{\beta}{\alpha' - \alpha}\right|^g \varphi_g} \quad (1,19)$$

Конечность волновой функции обеспечивается лишь в том случае, когда

$$\left|\frac{\beta}{\alpha'-\alpha}\right|<1, \text{ т. е. } \lambda''>0,$$

что и было отмечено вначале. Волновой функции (1,19) соответствует локальный уровень, положение которого в энергетическом спектре может быть определено, если подставить (1,16) в (1,13):

$$W=W_0+\alpha+2\beta\cos(r\pi+i\lambda'')=W_0+\alpha+2\beta(-1)^r\operatorname{ch}\lambda''.$$

Полагая

$$e^{\lambda''}=\left|\frac{\alpha'-\alpha}{\beta}\right|\gg 1,$$

будем иметь:

$$W=W_0+\alpha+2\beta(-1)^r\left|\frac{\alpha'-\alpha}{\beta}\right|.$$

Если

$$\beta>0 \text{ и } \begin{cases}\alpha'>\alpha, \text{ то } r \text{ чётное},\\ \alpha'<\alpha, \text{ то } r \text{ нечётное},\end{cases}$$

если же

$$\beta<0 \text{ и } \begin{cases}\alpha'>\alpha, \text{ то } r \text{ нечётное},\\ \alpha'<\alpha, \text{ то } r \text{ чётное}.\end{cases}$$

Поэтому

$$\boxed{W=W_0+\alpha'.} \qquad (1,20)$$

Все наши рассуждения можно повторить для случая $A=0$ и $B\neq 0$ (случай в). В этом случае конечность ψ обеспечивается при

$$\lambda=r\pi-i\lambda'',$$

где, как и раньше, r целое и $\lambda''>0$. Мы приходим к той же собственной функции (1,19) и к тому же собственному значению (1,20).

Локальный уровень (1,20) лежит

под зоной, если $\alpha'<\alpha$;

над зоной, если $\alpha'>\alpha$.

Если же

$$\left|\frac{\alpha'-\alpha}{\beta}\right|<1,$$

то локальный уровень совсем отсутствует (иначе выражаясь, лежит внутри зоны). В этом последнем случае оба коэффициента A и B в (1,11) мы обязаны считать отличными от нуля.

Мы рассмотрели здесь одномерную модель гомеополярной решётки, ограниченной с одной стороны. Наши результаты могут быть легко распространены на случай гетерополярной решётки, одномерная модель которой представляет собой цепочку, составленную из ионов двух сортов, чередующихся между собой.

§ 2. Гомеополярная решётка с пустыми узлами

А. Постановка задачи. Рассмотрим бесконечную цепочку, составленную из одинаковых атомов, расположенных на одинаковых расстояниях друг от друга. Пусть a — расстояние между соседними атомами. Такая цепочка представляет собой одномерную модель гомеополярной решётки.

Предположим теперь, что в нашей цепочке периодически повторяются пустые узлы. Пусть каждый N-й атом отсутствует. Таким образом, между двумя пустыми узлами заключено $N-1$ атомов (см. рис. 44, *г*).

Совместим начало координат с одним из пустых узлов и перенумеруем атомы так, как это показано на рис. 44, *г*. Пусть g — номер атома. Если x_g есть координата ядра g-го атома, то, очевидно,

$$x_g = ga.$$

В некоторых случаях нам будет удобнее пользоваться другим способом нумерации атомов, заменяя значок g парой значков k, n. Пустые узлы разбивают нашу цепочку на отдельные области; k есть номер области, а n — номер атома внутри области:

$$g = 0, \pm 1, \pm 2, \pm 3, \ldots$$
$$k = 0, \pm 1, \pm 2, \pm 3, \ldots$$
$$n = 0, 1, 2, \ldots, N-1.$$

Очевидно,

$$g = kN + n.$$

Пустые узлы имеют номера $g = kN$, т. е. для них $n = 0$.

Ход потенциала $V^*(xyz)$ изображён на рис. 44, *а*. Потенциал складывается из двух частей:

$$V^*(xyz) = V(xyz) - V'(xyz),$$

где V — потенциал идеальной решётки, а потенциал V' можно рассматривать как некоторую поправку. Очевидно, потенциал V является периодической функцией от x с периодом a,

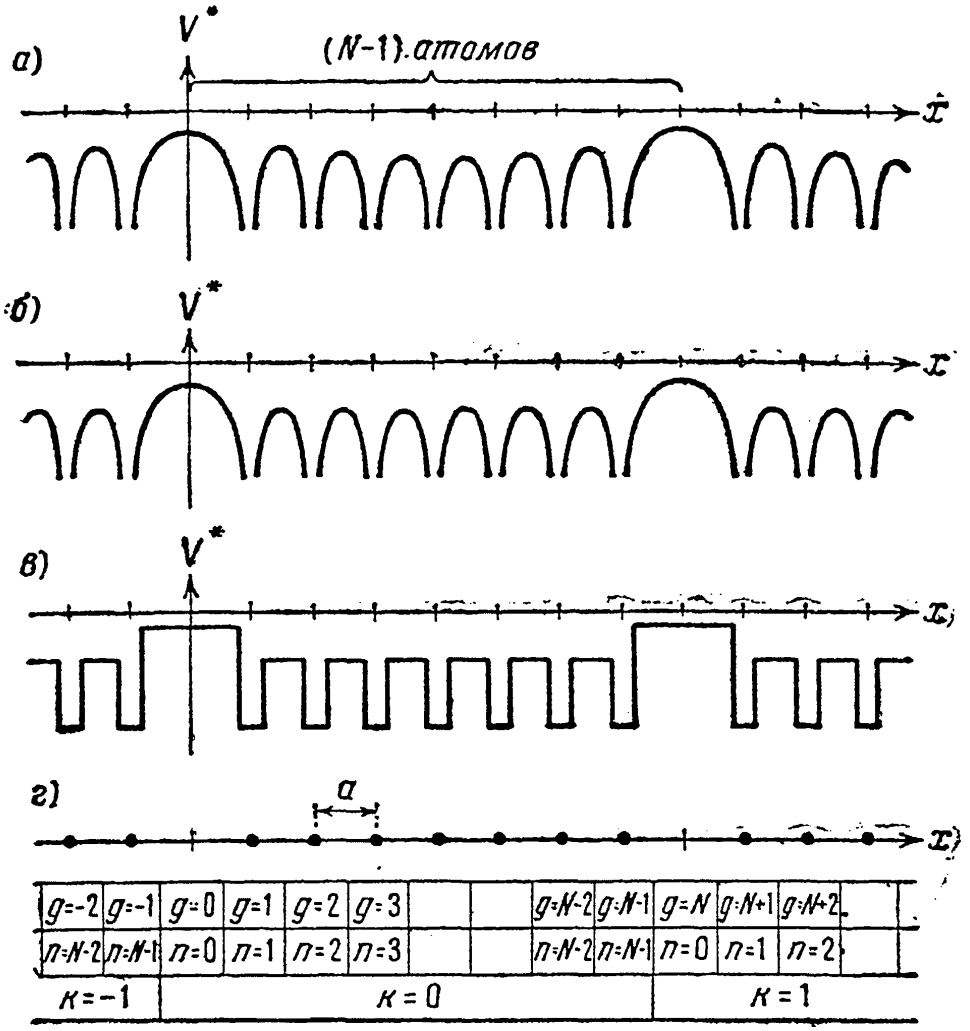


Рис. 44. Ход потенциала в гомеополярной решётке с пустыми узлами.

а потенциал V' — периодической функцией от x с периодом $L = Na$:

$$V(x, y, z) = V(x - a, y, z),$$
$$V'(x, y, z) = V'(x - L, y, z).$$

Потенциал V' отличен от нуля практически лишь вблизи точек $x = kN$, т. е. вблизи пустых узлов, и быстро спадает по мере удаления от этих точек. Мы будем считать, что

$$\begin{aligned} V(xyz) &\neq 0 \ldots \text{ при } (kN-1)\,a \leqslant x \leqslant (kN+1)\,a, \\ V(xyz) &= 0 \ldots \text{ при } (kN+1)\,a < x < (k+1)\,Na - a, \end{aligned} \qquad (2,1)$$

т. е., что V' практически отличен от нуля лишь на расстоянии a от повреждения. Таким образом, наличие повреждения чувствуют на себе лишь непосредственно соседние к повреждению атомы. Таким образом, истинный ход потенциала, изображённый на рис. 44, *а*, мы заменяем апроксимированным ходом, изображённым на рис. 44, *б*. Заметим, что Соколов [4], рассматривая эту же задачу, идёт ещё дальше, пользуясь апроксимацией, изображённой на рис. 44, *в* (модель Кронига-Пенни).

Рассмотрим g-й атом ($g \neq kN$), считая его изолированным. Уравнение Шредингера для электрона имеет вид:

$$H_g \varphi_g = W_0 \varphi_g,$$

где

$$H_g = -\frac{h^2}{8\pi^2 m}\Delta + U_g,$$

$$U_g(x, y, z) = U(x - ga, y, z)$$

$$\varphi_g(x, y, z) = \varphi(x - ga, y, z).$$

Мы будем считать для простоты, что функция φ является s-функцией, т. е. не вырождена.

Рассмотрим теперь решётку в целом:

$$H\psi = W\psi,$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V^*,$$

$$V^*(x, y, z) = \sum_g{}' U_g(x, y, z) + S.$$

Здесь S — самосогласованный потенциал электронов. Штрих при знаке суммы означает, что суммирование проводится по всем значениям g, за исключением значений $g = kN$. Иначе говоря, мы суммируем по фактически существующим атомам, а не по узлам. Волновую функцию ψ будем искать в следующем виде:

$$\psi(x, y, z) = \sum_g{}' a_g \varphi_g(x, y, z).$$

Задача сводится к отысканию коэффициентов a_g.

В. Собственные функции и собственные значения. Напишем вариационный интеграл:

$$J=\int \bar{\psi}(H-W)\psi\, d\tau=$$
$$=\sum_g{}'\sum_{g'}{}' \bar{a}_g a_{g'} \int \bar{\varphi}_g (H_{g'}-W+V^*-U_g)\varphi_{g'} d\tau=$$
$$=\sum_g{}'\sum_{g'}{}' \bar{a}_g a_{g'} \{(W_0-W)A_{gg'}+B_{gg'}\}. \qquad (2,2)$$

Здесь приняты обозначения:

$$A_{gg'}=\int \bar{\varphi}_g \varphi_{g'} d\tau, \quad B_{gg'}=\int \bar{\varphi}_g (V^*-U_{g'})\varphi_{g'}\, d\tau. \qquad (2,3)$$

Система уравнений, служащая для определения коэффициентов a_g, имеет вид:

$$\frac{\partial J}{\partial \bar{a}_g}=\sum_{g'}{}' a_{g'}\{(W_0-W)A_{gg'}+B_{gg'}\}=0. \qquad (2,4)$$

Мы получаем бесконечную систему линейных однородных уравнений ($g=0, \pm 1, \pm 2, \ldots$). Считая функции φ_g слабо перекрывающимися, а также пользуясь тем обстоятельством, что функции φ_g обладают сферической симметрией, кроме того, используя нашу апроксимацию (2,1) относительно потенциала V^*, мы можем положить:

$$A_{gg'}=\begin{cases}=1\ldots, & \text{если } g'=g\\ =0\ldots, & \text{если } g'\neq g,\end{cases}$$

$$B_{gg'}=\begin{cases}=\alpha\ldots, & \text{если } g'=g\neq kN,\ kN\pm 1\\ =\alpha'\ldots, & \text{если } g'=g=kN\pm 1\\ =\beta\ldots, & \text{если } \begin{cases} g'=g+1\neq kN,\ kN+1\\ g'=g-1\neq kN, kN-1,\end{cases}\\ =0\ldots & \text{во всех остальных случаях.}\end{cases} \qquad (2,5)$$

Заметим, что

$$\alpha>\alpha'>0>\beta. \qquad (2,6)$$

Если заменить значок g значками k и n, так что $a_g=a_{k,n}$

то систему (2,4) можно переписать так:

$$\begin{cases} \text{для } n=2, 3, 4, \ldots, N-3, N-2 \ (k \text{ — любое}) \\ \quad \text{а) } (W_0 - W + \alpha)\, a_{k,n} + \beta (a_{k,n+1} + a_{k,n-2}) = 0, \\ \text{для } n=1 \ (k \text{ — любое}) \\ \quad \text{б) } (W_0 - W + \alpha')\, a_{k,1} + \beta a_{k,2} = 0, \\ \text{для } n=N-1 \ (k \text{ — любое}) \\ \quad \text{в) } (W_0 - W + \alpha')\, a_{k,N-1} + \beta a_{k,N-2} = 0. \end{cases} \tag{2,7}$$

Из этих уравнений определяются интересующие нас коэффициенты $a_{k,n}$. Условие совместности этих уравнений даёт нам собственные значения.

Для дальнейшего удобно на функцию ψ наложить условия Кармана-Борна, сделав её периодической с периодом L:

$$\psi(x, y, z) = \psi(x + L, y, z). \tag{2,8}$$

Из условия периодичности (2,8) следует:

$$a_{k,n} = a_n.$$

Коэффициенты a_n будем искать в виде:

$$a_n = Ae^{i\lambda n} + Be^{-i\lambda n}, \tag{2,9}$$

где λ — любое, вообще говоря, комплексное число; A и B — произвольные константы. Симметрический характер нашей задачи заставляет наложить на коэффициенты a_n ещё следующее дополнительное условие:

$$|a_n|^2 = |a_{N-n}|^2,$$

откуда вытекает

$$a_n = e^{i\mu} a_{N-n}. \tag{2,10}$$

Подставляя (2,9) в (2,10), получаем

$$A = Be^{-i\lambda N + i\mu}, \quad B = Ae^{i\lambda N + i\mu}.$$

Таким образом, мы должны положить

$$e^{i\mu} = e^{-i\mu} = \pm 1,$$

т. е.

$$\mu = \pi s,$$

где s — целое число. Условие (2,10) приводит к тому, что константы A и B оказываются связанными. Благодаря условию (2,10) уравнения (2,7б) и (2,7в) сливаются в одно уравнение.

Итак, принимая следующее обозначение:

$$C = 2Ae^{+i\lambda\frac{N}{2}} = (-1)^s 2Be^{-i\lambda\frac{N}{2}},$$

мы имеем:

$$a_n = \begin{cases} = C\cos\lambda\left(\frac{N}{2}-n\right)\ldots, & \text{если } s \text{ чётное} \quad (2,11а) \\ = -iC\sin\lambda\left(\frac{N}{2}-n\right)\ldots, & \text{если } s \text{ нечётное.} \quad (2,11б) \end{cases}$$

Подставляя (2,11а) и (2,11б) в уравнение (2,7а):

$\cos\lambda\left(\frac{N}{2}-n\right)\cdot(W_0 - W + \alpha + 2\beta\cos\lambda) = 0\ldots$, если s чётное,

$\sin\lambda\left(\frac{N}{2}-n\right)\cdot(W_0 - W + \alpha + 2\beta\cos\lambda) = 0\ldots$, если s нечётное,

откуда

$$W = W_0 + \alpha + 2\beta\cos\lambda. \qquad (2,12)$$

Подставляя (2,11а) или (2,11б) и (2,12) в уравнение (2,7б, в):

$$C\left\{(\alpha'-\alpha)\cos\lambda\left(\frac{N}{2}-1\right) - \beta\cos\lambda\frac{N}{2}\right\} = 0,$$

$$-iC\left\{(\alpha'-\alpha)\sin\lambda\left(\frac{N}{2}-1\right) - \beta\sin\lambda\frac{N}{2}\right\} = 0,$$

откуда

$(\alpha'-\alpha)\cos\lambda\left(\frac{N}{2}-1\right) - \beta\cos\lambda\frac{N}{2} = 0\ldots$, если s чётное, (2,13а)

$(\alpha'-\alpha)\sin\lambda\left(\frac{N}{2}-1\right) - \beta\sin\lambda\frac{N}{2} = 0\ldots$, если s нечётное. (2,13б)

Для идеальной решётки условие (2,13а) или (2,13б) отсутствует. Поэтому в случае идеальной решётки параметр λ может быть произвольным. В нашем случае λ может принимать лишь вполне определённые значения, удовлетворяющие уравнениям (5,13а) или (5,13б). «Возможные» значения λ суть корни этих уравнений. Мы получим полный набор собственных функций, подставляя в (2,11а) корни уравнения (2,13а), а в (2,11б) корни уравнения (2,13б). Соответствующие собственные значения определяются из формулы (2,12).

Корни уравнений (2,13а) и (2,13б), вообще говоря, могут быть

комплексными: $\lambda = \lambda' + i\lambda''$
чисто вещественными: $\lambda = \lambda'$
чисто мнимыми: $\lambda = \pi r + i\lambda''$

(здесь λ' и λ'' — существенно вещественные, r — целое число). Нас, однако, могут интересовать только чисто вещественные или чисто мнимые корни, ибо в случае комплексного λ энергия W оказывается комплексной. Действительно, полагая в (5,12) $\lambda = \lambda' + i\lambda''$:

$$W = W_0 + \alpha + 2\beta\,(\cos\lambda' \operatorname{ch}\lambda'' - i\sin\lambda' \operatorname{sh}\lambda'').$$

Мы видим, что W остаётся вещественной лишь в следующих случаях:

если $\lambda'' = 0$. В этом случае $W = W_0 + \alpha + 2\beta\cos\lambda'$,
если $\lambda' = \pi r$. В этом случае $W = W_0 + \alpha + (-1)^r 2\beta \operatorname{ch}\lambda''$.

Вещественные λ приводят к представлению о зоне. Мнимые λ приводят к представлению о локальных уровнях. Мы имеем

при вещественных λ: $W_0 + \alpha - 2\beta \leqslant W \leqslant W_0 + \alpha + 2\beta$,
при мнимых λ: $W \leqslant W^0 + \alpha - 2\beta$ и $W \geqslant W_0 + \alpha + 2\beta$.

В нашей задаче $\beta < 0$. Поэтому локальные уровни, для которых r чётное, лежат *под* зоной; локальные уровни, для которых r нечётное, лежат *над* зоной.

С. Энергетические зоны. Обратимся к отысканию *вещественных* корней уравнений (2,13а) и (2,13б). Мы ограничимся рассмотрением двух предельных случаев:

$$\text{а)}\ \frac{\alpha' - \alpha}{\beta} \gg 1,$$

$$\text{б)}\ \frac{\alpha' - \alpha}{\beta} \ll 1.$$

Случай а. Перепишем (2,13а) и (2,13б) так:

$$\begin{cases} \cos\lambda\left(\frac{N}{2} - 1\right) = \frac{\beta}{\alpha' - \alpha}\cos\lambda\frac{N}{2} \ldots & s \text{ чётное}, \quad (2,14а) \\ \sin\lambda\left(\frac{N}{2} - 1\right) = \frac{\beta}{\alpha' - \alpha}\sin\lambda\frac{N}{2} \ldots & s \text{ нечётное}. \quad (2,14б) \end{cases}$$

Так как

$$\left|\cos\lambda\frac{N}{2}\right| \leqslant 1 \quad \text{и} \quad \left|\sin\lambda\frac{N}{2}\right| \leqslant 1,$$

то

$$\left|\cos\lambda\left(\frac{N}{2}-1\right)\right|\ll 1 \quad \text{и} \quad \left|\sin\lambda\left(\frac{N}{2}-1\right)\right|\ll 1.$$

Поэтому можно положить

$$\lambda\left(\frac{N}{2}-1\right)=\frac{\pi}{2}j+x,$$

т. е.

$$\lambda=\frac{\pi j+2x}{N-2}, \tag{2,15}$$

где j — целое число, а x — малая поправка ($|x|\ll 1$). Причём

j нечётное, если s чётное,
j чётное, если s нечётное.

Подставляя (2,15) в (2,14а) или (2,14б), получим:

$$x=(-1)^j\frac{\beta}{\alpha'-\alpha}\sin\lambda\approx(-1)^j\frac{\beta}{\alpha'-\alpha}\sin\frac{\pi j}{N-2}.$$

Итак, искомые корни суть:

$$\boxed{\lambda_j=\frac{\pi j}{N-2}+(-1)^j\frac{2\beta}{\alpha'-\alpha}\cdot\frac{1}{N-2}\sin\frac{\pi j}{N-2}.} \tag{2,16}$$

Нечётные значения j дают нам корни уравнения (2,14а), а чётные значения j — корни уравнения (2,14б). При $\frac{\beta}{\alpha'-\alpha}\to 0$ получим $\lambda_j\to\frac{\pi j}{N-2}$.

Подсчитаем теперь общее число состояний, содержащихся в зоне. Иначе говоря, посмотрим, сколько существует разных (линейно независимых) функций ψ_j, соответствующих вещественным λ_j. Число j может быть любым целым; однако при изменении j состояния повторяются. Отметим следующих три свойства функции ψ_j, которые могут быть легко установлены на основании (2,16), (2,11) и (2,2):

$$\left.\begin{aligned} &1)\ \psi_{j+2(N-2)}=(-1)\psi_j,\\ &2)\ \psi_{-j}=(-1)^{j+1}\psi_j,\\ &3)\ \psi_0=\psi_{N-2}=0. \end{aligned}\right\} \tag{2,17}$$

Из первого свойства следует, что функции ψ_j и $\psi_{j+2(N-2)}$ представляют собой одно и то же состояние. Поэтому среди бесчисленного множества функций ψ_j следует ограничиться теми, для которых значок j лежит в интервале

$$j_0 \leqslant j \leqslant j_0 + 2(N-2),$$

где j_0 — любое. Например, можно ограничиться для j следующим набором значений:

$$j = 0, \pm 1, \pm 2, \ldots, \pm(N-3), \pm(N-2).$$

Из второго свойства следует, что функции ψ_j и ψ_{-j} также представляют собой одно и то же состояние. Поэтому следует отбросить, например, все состояния с отрицательными j. Итак:

$$j = 0,\ 1,\ 2,\ldots,\ N-3,\ N-2.$$

Из третьего свойства следует, что состояния с $j=0$ и $j=N-2$ также должны быть отброшены, ибо соответствующие им волновые функции равны нулю. Отсюда, окончательно:

$$j = 1,\ 2,\ldots,\ N-3.$$

Мы получаем всего $N-3$ состояния, т. е. на два состояния меньше, чем число атомов. Эти два недостающих состояния, как мы увидим, появляются в качестве локальных состояний.

Случай б. Перепишем (2,13а) и (2,13б) следующим образом:

$$\begin{cases} \cos\lambda\dfrac{N}{2} = \dfrac{\alpha'-\alpha}{\beta}\cos\lambda\left(\dfrac{N}{2}-1\right) \ldots s \text{ чётное}, & (2,18а) \\ \sin\lambda\dfrac{N}{2} = \dfrac{\alpha'-\alpha}{\beta}\sin\lambda\left(\dfrac{N}{2}-1\right) \ldots s \text{ нечётное}. & (2,18б) \end{cases}$$

Так как

$$\left|\cos\lambda\left(\frac{N}{2}-1\right)\right| \leqslant 1 \text{ и } \left|\sin\lambda\left(\frac{N}{2}-1\right)\right| \leqslant 1,$$

то

$$\left|\cos\lambda\frac{N}{2}\right| \leqslant 1 \text{ и } \left|\sin\lambda\frac{N}{2}\right| \leqslant 1.$$

Поэтому можно положить:

$$\lambda\frac{N}{2} = \frac{\pi}{2}j + x, \qquad (2,19)$$

т. е.

$$\lambda = \frac{\pi j + 2x}{N},$$

где j — целое число, а x — малая поправка ($|x| \ll 1$). Причём

j нечётное, если s чётное,
j чётное, если s нечётное.

Подставляя (2,19) в (2,18а) или (2,18б):

$$x = (-1)^j \frac{\alpha' - \alpha}{\beta} \sin\lambda \approx (-1)^j \frac{\alpha' - \alpha}{\beta} \sin\frac{\pi j}{N}.$$

Итак, искомые корни суть:

$$\boxed{\lambda_j = \frac{\pi j}{N} + (-1)^j 2 \frac{\alpha' - \alpha}{\beta} \cdot \frac{1}{N} \sin\frac{\pi j}{N}}. \qquad (2,20)$$

Нечётные значения j дают нам корни уравнения (2,18а), а чётные значения j — корни уравнения (2,18б). При $\frac{\alpha' - \alpha}{\beta} \to 0$ получим $\lambda_j \to \frac{\pi j}{N}$.

Обратимся теперь к подсчёту состояний. Мы можем отметить следующих три свойства функции ψ_j, аналогичные 2,17):

$$\left.\begin{aligned} &1)\ \psi_{j+2N} = (-1)^N \psi_j, \\ &2)\ \psi_{-j} = (-1)^{j+1} \psi_j, \\ &3)\ \psi_0 = \psi_N = 0. \end{aligned}\right\} \qquad (2,21)$$

Из этих свойств, на основании рассуждений, аналогичных тем, которые были проведены в предыдущем случае, вытекает следующий набор возможных значений для j:

$$j = 1, 2, 3, \ldots, N-1.$$

Мы получаем всего $N-1$ состояний, т. е. число состояний, равное числу атомов. В этом случае, как мы увидим ниже, локальные уровни не возникают.

На рис. 45 изображён коэффициент $a_{k,n}(\lambda_j)$ как функция значка n для двух предельных случаев:

а) $\dfrac{\beta}{\alpha'-\alpha} \to 0$ ... светлые кривые,

б) $\dfrac{\alpha'-\alpha}{\beta} \to 0$ ... жирные кривые.

В первом случае узловыми точками являются точки $n=1$ и $n=N-1$. Во втором случае — точки $n=0$ и $n=N$. При чётных j между узловыми точками укладывается чётное число полуволн; при нечётных j — нечётное число полуволн.

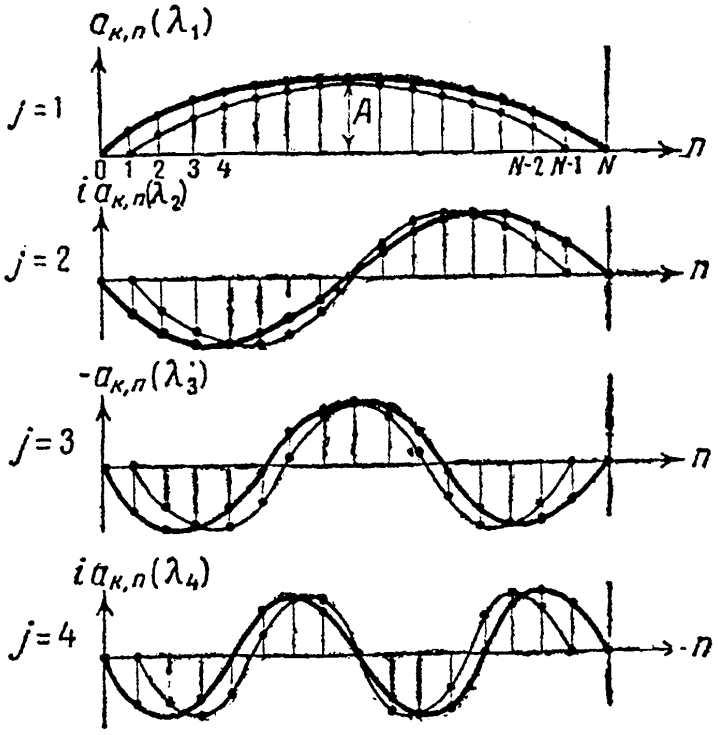


Рис. 45. Коэффициент $a_{k,n}(\lambda_j)$ как функция значка n (для разных j при $\mu=0$).

На рис. 46 изображена зависимость W_j от j опять-таки для тех же предельных случаев:

а) если $\dfrac{\beta}{\alpha'-\alpha} \to 0$, то $W_j = W_0 + \alpha + 2\beta\cos\dfrac{\pi j}{N-2}$,

б) если $\dfrac{\alpha'-\alpha}{\beta} \to 0$, то $W_j = W_0 + \alpha + 2\beta\cos\dfrac{\pi j}{N}$.

Случаю а) соответствует светлая кривая; случаю б) — жирная кривая. Число j можно трактовать как номер уровня внутри зоны. В нашей задаче $\beta < 0$, поэтому нумерация

уровней идёт снизу вверх (рис. 46,*а*). Если бы мы имели $\beta > 0$, то нумерация шла бы сверху вниз (рис. 46,*б*).

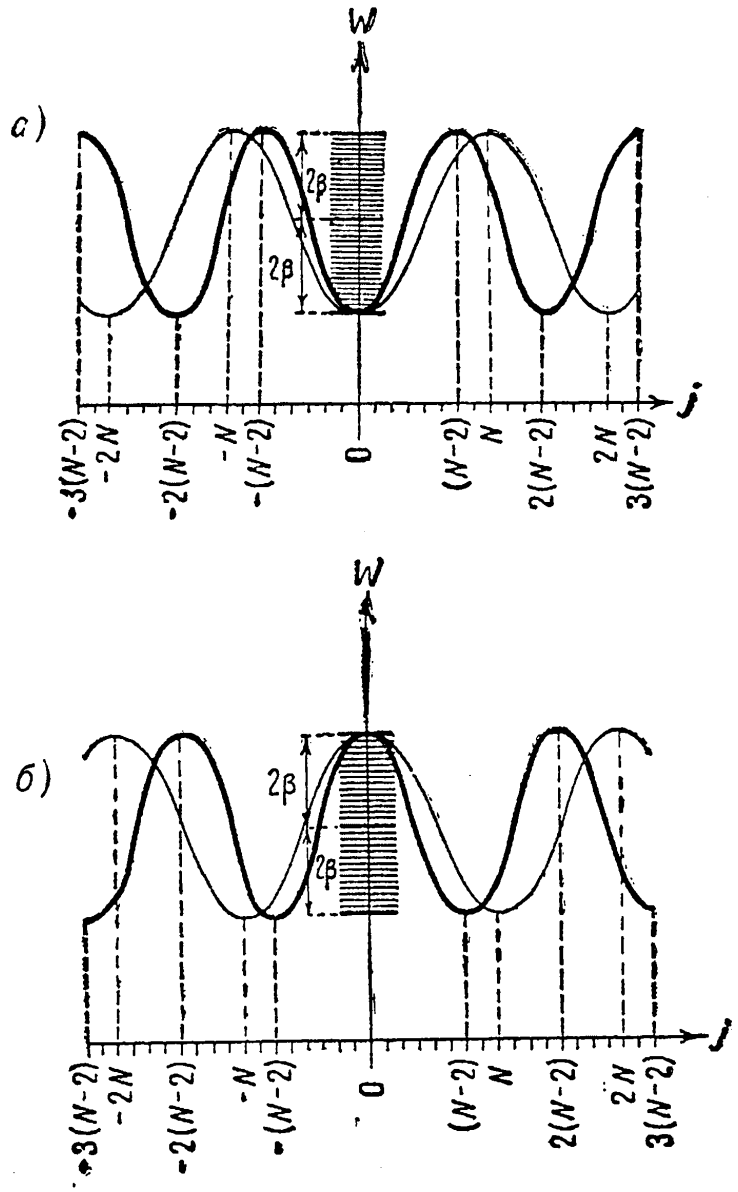


Рис. 46. Зависимость W от j (структура зоны):
а) случай $\beta < 0$, б) случай $\beta > 0$.

D. Локальные уровни. Обратимся теперь к отысканию мнимых корней уравнений (2,13а) и (2,13б). Положим:

$$\lambda = \pi r + i\lambda''. \qquad (2,22)$$

Опять-таки ограничимся рассмотрением двух случаев:

$$\text{а)}\ \frac{\alpha' - \alpha}{\beta} \gg 1, \quad \text{б)}\ \frac{\alpha' - \alpha}{\beta} \ll 1.$$

Случай а. Используя (2,22), перепишем (2,13а) и (2,13б)

следующим образом:

$$(-1)^r \frac{\operatorname{ch} \lambda'' \frac{N}{2}}{\operatorname{ch} \lambda'' \left(\frac{N}{2}-1\right)} = \frac{\alpha'-\alpha}{\beta} \quad \ldots \; s \text{ чётное}, \tag{2,23а}$$

$$(-1)^r \frac{\operatorname{sh} \lambda'' \frac{N}{2}}{\operatorname{sh} \lambda'' \left(\frac{N}{2}-1\right)} = \frac{\alpha'-\alpha}{\beta} \quad \ldots \; s \text{ нечётное}. \tag{2,23б}$$

Заметим, каково бы ни было λ'' (как при $\lambda'' > 0$, так и при $\lambda'' < 0$), мы имеем всегда

$$\frac{\operatorname{ch} \lambda'' \frac{N}{2}}{\operatorname{ch} \lambda \left(\frac{N}{2}-1\right)} > 1, \quad \frac{\operatorname{sh} \lambda'' \frac{N}{2}}{\operatorname{sh} \lambda'' \left(\frac{N}{2}-1\right)} > 1, \tag{2,24}$$

откуда следует:

r должно быть чётным, если $\frac{\alpha'-\alpha}{\beta} > 0$,

r должно быть нечётным, если $\frac{\alpha'-\alpha}{\beta} < 0$.

Если это не имеет места, то уравнения (2,13а) и (2,13б) вообще не имеют корней вида (2,22). В нашей задаче $\frac{\alpha'-\alpha}{\beta} > 0$ [см. (2,6а)], т. е. r чётное; это значит, что локальные уровни располагаются *под* зоной.

В рассматриваемом нами случае, когда $\frac{\alpha'-\alpha}{\beta} \gg 1$, мы можем переписать (2,23а) так:

$$e^{|\lambda''|} = \frac{\alpha'-\alpha}{\beta}. \tag{2,25}$$

Совершенно в таком же виде может быть переписано и уравнение (2,23б). Итак, уравнения (2,23а) и (2,23б) имеют общий корень. Положение соответствующего энергетического уровня может быть определено, если подставить (2,25) в (2,12)

$$\boxed{W = W_0 + \alpha'.} \tag{2,26}$$

Этот уровень двукратно вырожден: ему соответствуют две волновые функции [согласно (2,11)]:

$$a'_n = C(-1)^{r\left(\frac{N}{2}-n\right)}\operatorname{ch}\lambda''\left(\frac{N}{2}-n\right) = C'\operatorname{ch}\lambda''\left(\frac{N}{2}-n\right);$$

$$a''_n = -C(-1)^{r\left(\frac{N}{2}-n\right)}\operatorname{sh}\lambda''\left(\frac{N}{2}-n\right) =$$

$$= -C'\operatorname{sh}\lambda''\left(\frac{N}{2}-n\right),$$

где $C' = (-1)^N C$, ибо, так как r чётное:

$$(-1)^{r\frac{N}{2}} = (-1)^N \text{ и } (-1)^{rn} = 1.$$

Удобнее пользоваться линейными комбинациями из этих функций:

$$\begin{cases} a_n^1 = a'_n - a''_n = C'\left(\dfrac{a'-a}{\beta}\right)^{\frac{N}{2}-n}, & (2,27а) \\ a_n^2 = a'_n + a''_n = C'\left(\dfrac{a'-a}{\beta}\right)^{n-\frac{N}{2}}. & (2,27б) \end{cases}$$

Обе функции быстро спадают по мере удаления от места повреждения: одна — влево, другая — вправо. Электрон практически локализован на атомах, непосредственно соседних к пустому узлу.

На рис. 47 изображён ход волновых функций (2,27а) (светлая кривая) и (2,27б) (жирняя кривая). На рис. 48 изображено положение локального уровня (2,26) в энергетическом спектре.

Наши локальные уровни представляют собой не что иное, как поверхностные уровни Тамма, рассмотренные в предыдущем параграфе. Этого и следовало ожидать. Действительно, мы учитывали здесь взаимодействие лишь непосредственно соседних атомов (вследствие слабого перекрывания волновых функций); поэтому наличие пустых узлов привело к тому, что наша цепочка атомов оказалась разорванной на отдельные, не связанные между собой куски. Это является специфической особенностью одномерной модели. При переходе к трёх-

мерному случаю эквивалентность между нашими локальными уровнями, вызванными наличием пустых узлов, и таммовскими поверхностными уровнями, исчезает.

Случай б. Рассмотрим теперь случай, когда

$$\frac{\alpha' - \alpha}{\beta} \ll 1. \tag{2,28}$$

В этом случае уравнения (2,13а) и (2,13б) вовсе не имеют отличных от нуля мнимых корней. В этом легко убедиться.

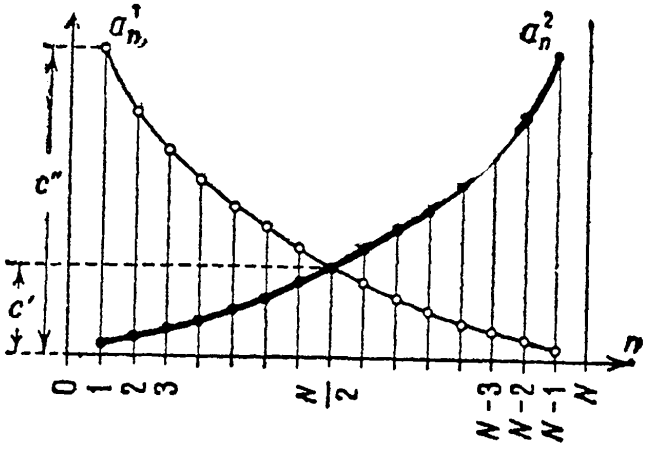


Рис. 47. Коэффициенты a_n^1 и a_n^2 как функции значка n (при $\mu = 0$[1]).

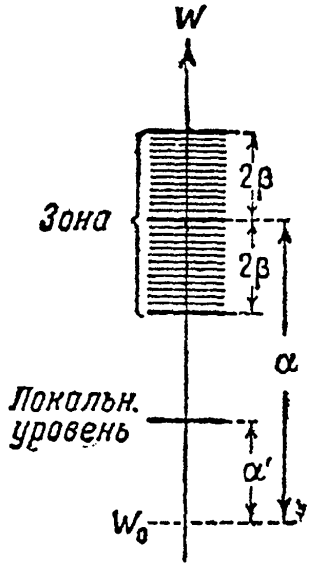


Рис. 48. Положение локального уровня в энергетическом спектре (случай $\alpha > \alpha' > 0 > \beta$).

Перепишем вновь (2,23а) и (2,23б), полагая r чётным (на основании тех же соображений, что и в предыдущем случае):

$$\begin{cases} \dfrac{\operatorname{ch}\lambda''\dfrac{N}{2}}{\operatorname{ch}\lambda''\left(\dfrac{N}{2}-1\right)} = \dfrac{\alpha'-\alpha}{\beta}, \\ \dfrac{\operatorname{sh}\lambda''\dfrac{N}{2}}{\operatorname{sh}\lambda''\left(\dfrac{N}{2}-1\right)} = \dfrac{\alpha'-\alpha}{\beta}. \end{cases} \tag{2,29}$$

[1]) Принято обозначение: $C'' = C'\left(\dfrac{\alpha'-\alpha}{\beta}\right)^{\frac{N}{2}-1}$.

При всех λ'' (как при $\lambda'' > 0$, так и при $\lambda'' < 0$) левая часть этих уравнений больше единицы [см. (2,24)], правая же часть значительно меньше единицы [см. (2,28)]. Поэтому равенства (2,29) могут быть удовлетворены лишь при $\lambda''=0$.

Таким образом, в рассматриваемом случае локальные уровни отсутствуют.

Возвращаясь, в заключение, к вопросу об общем числе состояний, мы можем резюмировать результаты в следующей табличке:

	Случай $\frac{\alpha'-\alpha}{\beta} \gg 1$	Случай $\frac{\alpha'-\alpha}{\beta} \ll 1$
Число состояний внутри зоны	$N-3$	$N-1$
Число локальных состояний	2	0
Всего состояний . .	$N-1$	$N-1$

§ 3. Гетерополярная решётка с пустыми узлами

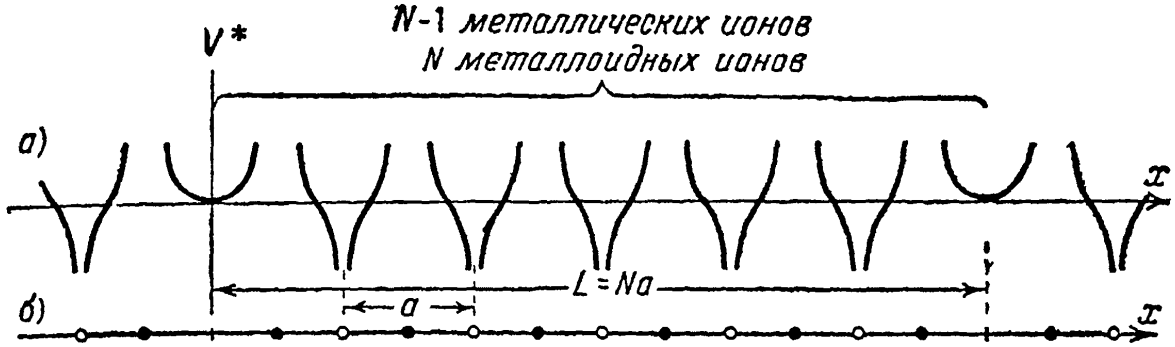
А. Пустые металлические узлы. Задача, рассмотренная в предыдущем параграфе, может быть распространена на случай гетерополярной решётки с пустыми металлическими узлами. Пусть мы имеем цепочку ионов, в которой чередуются положительные (металлические) и отрицательные (металлоидные) ионы. Пусть a — расстояние между соседними одноимёнными ионами (a равно удвоенной постоянной решётки). Представим себе теперь, что каждый N-й металлический ион отсутствует, оставляя на своём месте пустой металлический узел. Таким образом, между двумя пустыми узлами заключено $N-1$ металлических ионов и N металлоидных ионов (см. рис. 49,*б*).

Снабдим каждый металлический узел значком g или парой значков k, n, где, как и в предыдущем параграфе,

$$g = kN + n.$$

Металлоидные узлы остаются ненумерованными. Пусть пустые узлы имеют номера $g = kN$, т. е. для них $n = 0$ (см. рис. 49,*б*).

Наша задача отличается от задачи предыдущего параграфа лишь одним обстоятельством: видом функции $V^*(xyz)$. Апроксимированный ход потенциала V^* изображён на рис. 49,*а*. Теперь рис. 49,*а* заменяет нам рисунок 44,*б*. Апроксимация, как и прежде, сводится, во-первых, к предположению, что



$g=-1$	$g=0$	$g=1$	$g=2$			$g=N-1$	$g=N$	
$n=N-1$	$n=0$	$n=1$	$n=2$			$n=N-1$	$n=0$	
$\kappa=-1$	$\kappa=0$						$\kappa=+1$	

○ *металлические (положительные) ионы*
● *металлоидные (отрицательные) ионы*

Рис. 49. Ход потенциала в гетерополярной решётке с пустыми металлическими узлами.

наличие пустых узлов никак не влияет на положение ионов (все ионы остаются в своих узлах, не испытывая сдвига), и, во-вторых, к предположению, что возмущающий потенциал V' практически отличен от нуля лишь в области, непосредственно соседней к повреждению. В нашем случае потенциал V^* складывается из потенциалов всех металлических и металлоидных ионов:

$$V^*(x, y, z) = \sum_g{}' U_g^M(xyz) + \sum_g U_g^R(xyz),$$

а возмущающий потенциал V' — из потенциалов всех отсутствующих металлических ионов:

$$V'(xyz) = \sum_k U_{kN}^M(xyz).$$

Штрих при знаке суммы означает, что суммирование прово-

дится по всем g, за исключением $g=kN$. Здесь:

$$\left.\begin{aligned} U_g^{\mathrm{M}}(x, y, z) &= U^{\mathrm{M}}(x-ga, y, z), \\ U_g^{\mathrm{R}}(x, y, z) &= U^{\mathrm{R}}\left(x-ga-\frac{1}{2}a, y, z\right), \end{aligned}\right\} \qquad (3,1)$$

причём

$$U^{\mathrm{M}}(x, y, z) = -U^{\mathrm{R}}(x, y, z) = -\frac{e^2}{\sqrt{x^2+y^2+z^2}}.$$

Строго говоря, последнее выражение справедливо лишь при

$$x^2+y^2+z^2 \gg 0.$$

Очевидно,

$$V^*(xyz) = V(xyz) - V'(xyz),$$

где V — потенциал идеальной решётки.

Волновую функцию ищем в виде линейной комбинации из индивидуальных волновых функций отдельных металлических атомов:

$$\psi(x, y, z) = \sum_g{}' a_g \varphi_g(x, y, z). \qquad (3,2)$$

Функция $\varphi_g(x, y, z)$ удовлетворяет, очевидно, уравнению

$$H_g\varphi_g = W_0\varphi_g,$$

где

$$H_g = -\frac{h^2}{8\pi^2 m}\Delta + U_g^M.$$

Система уравнений, служащая для определения коэффициентов a_g, имеет вид (2,4) или (2,7). Повторяя рассуждения предыдущего параграфа, мы можем свести эту систему к уравнениям (2,12) и (2,13). Параметры α, α' и β имеют тот же смысл, что и в предыдущем параграфе [см. (2,3) и (2,5)]. Теперь, однако, вместо (2,5) мы имеем:

$$\alpha' > \alpha > 0, \quad \beta > 0. \qquad (3,3)$$

Заметим следующее. Если ионы считать точечными и если электрон, принадлежащий g-му металлическому атому, считать

сосредоточенным в g-м узле, т. е. если принять

$$|\varphi_g(xyz)|^2 = \begin{cases} =1 & \text{в точке } x=ga,\ y=z=0, \\ =0 & \text{вне этой точки,} \end{cases}$$

то квантово-механические параметры α, α', β превращаются в свои классические аналоги:

$$\left.\begin{aligned} \alpha &\longrightarrow \frac{e^2}{a_0} M, \\ \alpha' &\longrightarrow \frac{e^2}{a_0}\left(M+\frac{1}{2}\right), \\ \beta &\longrightarrow 0, \end{aligned}\right\} \tag{3,4}$$

где a_0 — постоянная решётки; M — коэффициент Маделунга [1]).

В нашей задаче, как и в предыдущем параграфе, мы имеем

$$\frac{\alpha'-\alpha}{\beta} > 0$$

и можем рассмотреть два случая:

$$\text{а)}\ \frac{\alpha'-\alpha}{\beta} \gg 1, \quad \text{б)}\ \frac{\alpha'-\alpha}{\beta} \ll 1.$$

В первом случае (случай а) мы получаем в энергетическом спектре зону, содержащую $N-3$ состояния, и, кроме того, локальный уровень, двукратно вырожденный. В отличие от предыдущего параграфа локальный уровень располагается теперь не под зоной, а над зоной. Действительно, мы имеем теперь $\beta > 0$ (в предыдущем параграфе мы имели $\beta < 0$) и r чётное (как и в предыдущем параграфе, см. формулу (2,22) и следующие). Положение локального уровня определяется формулой (2,26):

$$W = W_0 + \alpha'. \tag{3,5}$$

Рис. 48 должен быть теперь заменён рис. 50. Расстояние от центра зоны до локального уровня равно (по по-

---

[1]) В нашей одномерной задаче коэффициент Маделунга M имеет иное численное значение, чем то, которое обычно ему приписывают, когда имеют дело с трёхмерной решёткой.

рядку величины):

$$a' - a \approx \frac{e^2}{a}. \qquad (3,6)$$

Тот факт, что локальный уровень помещается над зоной проводимости, означает, что электрон, введённый в нашу повреждённую решётку, не улавливается повреждениями, а, наоборот, отталкивается от них. Этого и следовало ожидать. В вопросах электропроводности наличие такого рода локальных уровней не играет сколько-нибудь существенной роли.

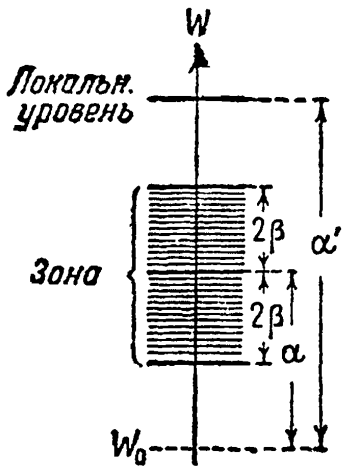


Рис. 50. Положение локального уровня в энергетическом спектре (случай $\alpha' > \alpha > 0$; $\beta > 0$).

Во втором случае (случай б) мы получаем зону, содержащую $N-1$ состояний. Локальные уровни отсутствуют.

В. Пустые металлоидные узлы. Рассмотрим попрежнему цепочку, составленную из ионов двух сортов (металлические и металлоидные), которые чередуются между собой. Пусть попрежнему a — кратчайшее расстояние между двумя одноимёнными ионами. Предположим теперь, что каждый N-й металлоидный ион отсутствует, оставляя на своём месте пустой металлоидный узел. Таким образом, между двумя пустыми узлами содержится N металлических ионов и $N-1$ металлоидных.

Перенумеруем ячейки нашей решётки так, как это показано на рис. 51, *б*. Пусть g — номер ячейки. Иногда нам будет удобнее значок g заменять парой значков: k, n, где k — номер области, а n — номер ячейки внутри области. Каждая ячейка содержит два противоположно заряженных иона (металлоидный и металлический). Оба эти иона, следовательно, имеют один и тот же номер g. Начало координат совместимо с одним из металлических ионов. Если x_g^M — координата ядра g-го металлического иона, а x_g^R — координата ядра g-го металлоидного иона, то

$$\begin{cases} x_g^M = ga, \\ x_g^R = \left(g + \frac{1}{2}\right)a. \end{cases}$$

Пусть пустые металлоидные узлы помещаются между металлическими ионами с номерами $kN-1$ и kN, как это изображено на рис. 51, *б*.

Апроксимированный ход потенциала $V^*(xyz)$ изображён на рис. 51, *а*. Апроксимация та же, что и в предыдущих

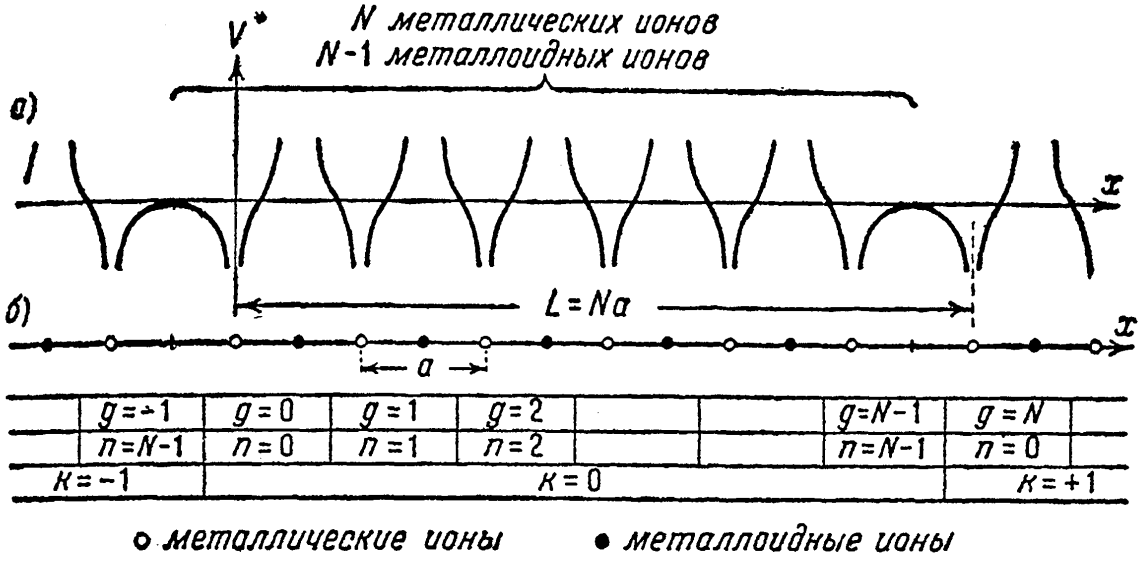


Рис. 51. Ход потенциала в гетерополярной решётке с пустыми металлоидными узлами.

параграфах. Потенциал V^* складывается из потенциалов всех металлических и металлоидных ионов, фактически присутствующих в решётке:

$$V^*(xyz) = \sum_g{}' U_g^{\mathrm{R}}(xyz) + \sum_g U_g^{\mathrm{M}}(xyz),$$

а возмущающий потенциал V' складывается из потенциалов отсутствующих металлоидных ионов:

$$V'(xyz) = \sum_k U_{kN-1}^{\mathrm{R}}(xyz).$$

Штрих при знаке суммы означает, что суммирование охватывает все значения g, за исключением значений $g = kN-1$, причём имеют место соотношения (3,1).

Волновую функцию $\psi(x, y, z)$, как и прежде, строим из индивидуальных волновых функций отдельных металлических атомов:

$$\psi(xyz) = \sum_g a_g \varphi_g(xyz). \qquad (3,7)$$

Система уравнений, служащая для определения коэффициентов a_g, имеет вид:

$$\sum_{g'} a_{g'} \left\{ (W_0 - W) A_{gg'} + B_{gg'} \right\} = 0. \tag{3,8}$$

Здесь, в отличие от (2,4), суммирование охватывает все без исключения значения g. Здесь:

$$A_{gg'} = \int \bar{\varphi}_g \varphi_{g'} \, d\tau;$$
$$B_{gg'} = \int \bar{\varphi}_g (V - U^{M}_{g'}) \varphi_{g'} \, d\tau. \tag{3,9}$$

Учитывая сферическую симметрию функций φ_g (по предположению это суть s-функции) и принимая во внимание нашу апроксимацию относительно потенциала V^*, мы можем положить:

$$A_{gg'} = \begin{cases} = 1 \ldots, \text{ если } g' = g \\ = 0 \ldots, \text{ если } g' \neq g \end{cases}$$

$$B_{gg'} = \begin{cases} = \alpha \ldots, \text{ если } g' = g \neq kN, \; kN - 1 \\ = \alpha' \ldots, \text{ если } g' = g = kN, \; kN - 1 \\ = \beta \ldots, \text{ если } \begin{cases} g' = g + 1 \neq kN \\ g' = g - 1 \neq kN - 1 \end{cases} \\ = \beta' \ldots, \text{ если } \begin{cases} g' = g + 1 = kN \\ g' = g - 1 = kN - 1 \end{cases} \\ = 0 \ldots \text{ во всех остальных случаях.} \end{cases} \tag{3,10}$$

Заметим, что

$$\left.\begin{aligned} \alpha > \alpha' > 0, \\ \beta > \beta' > 0. \end{aligned}\right\} \tag{3,11}$$

Если ионы считать точечными и если электрон, принадлежащий g-ому металлическому атому, считать сосредоточенным в g-м металлическом узле, то квантовомеханические параметры

α, α', β, β' превращаются в свои классические аналоги:

$$\left.\begin{aligned} &\alpha \longrightarrow \frac{2e^2}{a_0} M, \\ &\alpha' \longrightarrow \frac{2e^2}{a_0}(M-1), \\ &\beta \longrightarrow 0, \\ &\beta' \longrightarrow 0. \end{aligned}\right\} \qquad (3,12)$$

На основании (3,10) система уравнений (3,8) может быть переписана в таком виде:

$$\left.\begin{aligned} &\text{для } n = 1, 2, 3 \ldots, N-2;\ k \text{ — любое} \\ &\quad \text{а) } (W_0 - W + \alpha) a_{k,n} + \beta (a_{k,n+1} + a_{k,n-1}) = 0, \\ &\text{для } n = 0;\ k \text{ — любое} \\ &\quad \text{б) } (W_0 - W + \alpha') a_{k,0} + \beta a_{k,1} + \beta' a_{k-1,N-1.} = 0, \\ &\text{для } n = N-1;\ k \text{ — любое} \\ &\quad \text{в) } (W_0 - W + \alpha) a_{k,N-1} + \beta a_{k,N-2} + \beta' a_{k+1,0} = 0. \end{aligned}\right\} \qquad (3,13)$$

Эта система отличается от аналогичной системы (2,7) лишь одной чертой: наличием членов, содержащих β'.

Налагая на функцию $\psi(x, y, z)$ условие Кармана-Борна, получим аналогично тому, как это сделано в предыдущем параграфе:

$$a_n = \begin{cases} = C \cos \lambda \left(\frac{N-1}{2} - n\right) \ldots, & \text{если } s \text{ чётное} \qquad (3,14а) \\ = -iC \sin \lambda \left(\frac{N-1}{2} - n\right) \ldots, & \text{если } s \text{ нечётное} \qquad (3,14б) \end{cases}$$

Теперь уравнения (3,13б) и (3,13в) сливаются в одно уравнение.

Подставляя (3,14а) или (3,14б) в (3,13а), будем иметь:

$$W = W_0 + \alpha + 2\beta \cos \lambda. \qquad (3,15)$$

Подставляя (3,14а) или (3,14б) и (3,15) в уравнение (3,13б), будем иметь для s чётного и нечётного соответственно:

$$(\alpha' - \alpha + \beta') \cos \lambda \left(\frac{N-1}{2}\right) - \beta \cos \lambda \left(\frac{N-1}{2} + 1\right) = 0. \qquad (3,16а)$$

$$(\alpha' - \alpha - \beta') \sin \lambda \left(\frac{N-1}{2}\right) - \beta \sin \lambda \left(\frac{N-1}{2} + 1\right) = 0. \qquad (3,16б)$$

Итак, система (3,13) сводится к двум уравнениям:

(3,15) и (3,16а), если s чётное;
(3,15) и (3,16б), если s нечётное.

Уравнения (3,16) похожи на уравнения (2,13). Поэтому, отыскивая корни уравнений (3,16а) и (3,16б), мы можем сразу же воспользоваться результатами § 2. Во всех формулах § 2 нам надо заменить

$$\begin{cases} N \text{ через } N+1, \\ \alpha'-\alpha \text{ через } \alpha'-\alpha+(-1)^s\beta'. \end{cases}$$

Как и в § 2, мы могли бы рассмотреть два случая:

$$\text{а)}\ \left|\frac{\alpha'-\alpha}{\beta}\right| \gg 1,$$
$$\text{б)}\ \left|\frac{\alpha'-\alpha}{\beta}\right| \ll 1.$$

Однако, случай б отпадает, ибо у нас заведомо

$$\left|\frac{\alpha'-\alpha}{\beta}\right| > 1,$$

в чём можно легко убедиться. Если

$$\left|\frac{\alpha'-\alpha}{\beta}\right| \gg 1, \quad \text{то} \quad \left|\frac{\alpha'-\alpha\pm\beta'}{\beta}\right| \gg 1, \quad \text{ибо} \quad \left|\frac{\beta'}{\beta}\right| < 1. \qquad (3,17)$$

Поэтому, согласно § 2, вещественные корни уравнения (3,16а) имеют вид:

$$\lambda_j = \frac{\pi j}{N-1} - \frac{2\beta}{\alpha'-\alpha+\beta'} \cdot \frac{1}{N-1} \sin\frac{\pi j}{N-1}, \qquad (3,18а)$$

где j нечётное, а вещественные корни уравнения (3,16б) имеют вид:

$$\lambda_j = \frac{\pi j}{N-1} + \frac{2\beta}{\alpha'-\alpha-\beta'} \cdot \frac{1}{N-1} \sin\frac{\pi j}{N-1}, \qquad (3,18б)$$

где j чётное. Формулы (3,18а) и (3,18б) можно записать так:

$$\boxed{\lambda_j = \frac{\pi j}{N-1} + (-1)^j \frac{2\beta}{\alpha'-\alpha-(-1)^j\beta'} \cdot \frac{1}{N-1} \sin\frac{\pi j}{N-1}.} \qquad (3,19)$$

Этим вещественным корням соответствуют $N-2$ отличных от нуля, линейно независимых волновых функций $\psi_j(xyz)$

$$j=1, 2, 3, \ldots, N-2.$$

Мы получаем зону, содержащую $N-2$ состояния (на два меньше, чем число металлических ионов).

Кроме вещественных корней, каждое из уравнений (3,16а) и (3,16б) обладает одним мнимым корнем вида

$$\lambda=\pi r+i\lambda''.$$

Здесь r нечётное, ибо, согласно (3,11) и (3,17):

$$\frac{\alpha'-\alpha\pm\beta'}{\beta}<0$$

(см. § 2). Корень уравнения (3,16а) имеет вид

$$e^{|\lambda''|}=-\frac{\alpha'-\alpha+\beta'}{\beta}. \qquad (3,20а)$$

Ему соответствует волновая функция

$$\boxed{\begin{aligned}a'_n&=C(-1)^{r\left(\frac{N-1}{2}-n\right)}\operatorname{ch}\lambda''\left(\frac{N-1}{2}-n\right)=\\&=(-1)^n C'\operatorname{ch}\lambda''\left(\frac{N-1}{2}-n\right),\end{aligned}} \qquad (3,21а)$$

где

$$C'=C(-1)^{r\left(\frac{N-1}{2}\right)}=C(-1)^{\frac{N-1}{2}}.$$

Корень уравнения (3,16б) имеет вид:

$$e^{|\lambda''|}=-\frac{\alpha'-\alpha-\beta'}{\beta}. \qquad (3,20б)$$

Ему соответствует волновая функция

$$\boxed{\begin{aligned}a''_n&=-C(-1)^{r\left(\frac{N-1}{2}-n\right)}\operatorname{sh}\lambda''\left(\frac{N-1}{2}-1\right)=\\&=-(-1)^n C'\operatorname{sh}\lambda''\left(\frac{N-1}{2}-1\right).\end{aligned}} \qquad (3,21б)$$

Формулы (3,20а) и (3,20б) могут быть соединены и записаны так:

$$e^{|\lambda''|} = \frac{\alpha' - \alpha + (-1)^s \beta'}{\beta}. \qquad (3,22)$$

Итак, мы получаем два локальных уровня:

$$\boxed{W = W_0 + \alpha' \pm \beta'.} \qquad (3,23)$$

В § 2 мы имели один локальный уровень, двукратно вырожденный. Теперь этот уровень расщепляется на два, и вырождение пропадает. Это происходит благодаря тому, что теперь мы учитываем перекрытие волновых функций φ_{kN} и φ_{kN-1}, принадлежащих двум атомам, непосредственно соседним к повреждению (левому и правому соседу). В § 2 расстояние между этими атомами было больше, и мы имели право не учитывать перекрытия соответствующих волновых функций φ_{kN+1} и φ_{kN-1}.

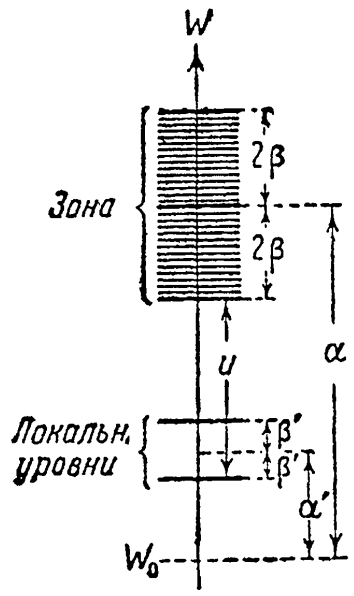


Рис. 52. Положение локальных уровней в энергетическом спектре.

В нашей задаче мы имеем $\beta > 0$ и r нечётное, следовательно, локальные уровни располагаются *под* зоной. Это означает, что повреждения рассматриваемого типа обладают способностью улавливать электрон, странствующий по зоне проводимости: электрон притягивается к пустому металлоидному узлу, что и следовало ожидать.

Положение локальных уровней (3,23) в энергетическом спектре изображено на рис. 52. Соответствующие им волновые функции (3,21а) (верхняя кривая) и (3,21б) (нижняя кривая) изображены на рис. 53. Верхняя кривая соответствует верхнему уровню на рис. 52; нижняя кривая — нижнему уровню.

Заметим, что удаление из решётки металлоидных ионов должно сопровождаться введением в неё равного количества электронов, для того чтобы сохранить нейтральность решётки. Эти электроны рассаживаются по локальным уровням, ибо локальные уровни представляют собой энергетически наиболее

выгодные для них места. Таким образом, около каждого пустого металлоидного узла локализован один электрон. Эта локализация тем сильнее, чем глубже лежит локальный уровень, т. е. чем больше (по абсолютной величине) отношение

$$\frac{\alpha' - \alpha - \beta'}{\beta}.$$

Действительно, чем больше (по абсолютной величине) это отношение, тем быстрее спадает соответствующая волновая функция по мере удаления от повреждения. Практически можно

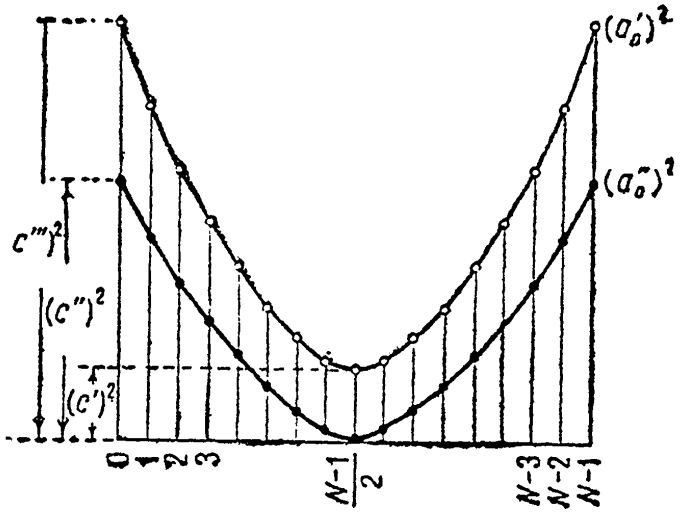


Рис. 53. Коэффициенты $(a_n')^2$ и $(a_n'')^2$, как функции значка n.

(Обозначения: $C'' = C' \left(\frac{\alpha' - \alpha - \beta'}{\beta}\right)^{\frac{N-1}{2}}$;

$C''' = C' \left(\frac{\alpha' - \alpha + \beta'}{\beta}\right)^{\frac{N-1}{2}}$.)

считать, что электрон сидит на одном из шести (в трёхмерной решётке) металлических ионов, окружающих пустой металлоидный узел. Иначе говоря, один из этих шести ионов нейтрализован. С этих нейтральных атомов металла, привязанных к пустым металлоидным узлам, и черпаются электроны проводимости. Энергия ионизации, т. е. энергия, необходимая для того, чтобы перевести электрон с локального уровня в зону проводимости, или, иначе говоря, из связанного состояния в коллективизированное состояние, равна (см. рис. 52):

$$u = \alpha - 2\beta - \alpha' - \beta'. \qquad (3,24)$$

При классическом подходе к задаче мы имеем, согласно (3,12),

$$u = a - a' = \frac{e^2}{a_0}.$$

В заключение вернёмся к вопросу об общем числе состояний. В нашей задаче мы имеем:

число состояний внутри зоны . . .	$N-2$
число локальных состояний	2
всего состояний	N

Общее число состояний равно числу металлических ионов, как и должно быть.

§ 4. Решётка с междуузельными атомами

А. Постановка задачи. Рассмотрим бесконечную цепочку, составленную из одинаковых атомов, расположенных на расстоянии a друг от друга. Представим себе, что в нашу цепочку вкраплены лишние атомы одного и того же сорта,

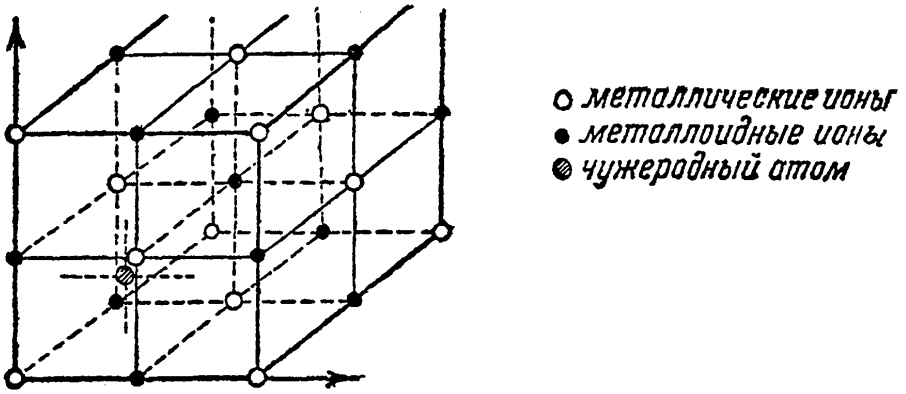


Рис. 54. Гетерополярная решётка с междуузельным атомом.

размещённые в междуузельях. Пусть через каждые N регулярных атомов встречается иррегулярный атом. Эти избыточные иррегулярные атомы могут быть иной или той же химической природы, что и регулярные атомы решётки.

Такая цепочка представляет собой одномерную модель гомеополярной решётки с примесными междуузельными атомами. Она может служить также одномерной моделью гетерополярной решётки с избыточными (междуузельными) атомами металла. Заметим, однако, что наша одномерная модель не вполне точно отражает положение дел, существующее в ре-

альной трёхмерной (кубической) решётке. Действительно, в трёхмерном случае междуузельные атомы и основные атомы решётки не попадают на одну и ту же ось (см. рис. 54).

Перенумеруем все атомы в нашей цепочке (включая и иррегулярные), снабдив каждый атом значком g. Будем характеризовать регулярные атомы целыми значениями g, а иррегулярные атомы — полуцелыми значениями g, как это показано на рис. 55, *в*; именно:

$$\left.\begin{array}{l}\text{для регулярных атомов: } g=kN+n,\\ \text{для иррегулярных атомов: } g=kN-\frac{1}{2},\\ \text{где } \begin{cases} k=0,\ \pm1,\ \pm2,\ \pm3,\ldots \\ n=0,\ 1,\ 2,\ldots,\ N-1. \end{cases}\end{array}\right\} \qquad (4{,}1)$$

Рассмотрим g-й атом, считая его изолированным. Если x_g есть координата ядра g-го атома, то при нашем способе нумерации мы будем иметь для любого атома:

$$x_g=ga. \qquad (4{,}2)$$

Уравнение Шредингера для электрона, принадлежащего g-му атому, имеет вид:

$$H_g\varphi_g=W_g\varphi_g,$$

где

$$H_g=-\frac{h^2}{8\pi^2 m}\Delta+U_g.$$

В общем случае, когда регулярные и иррегулярные атомы обладают различной химической природой, имеем, на основании (4,2):

$$\left.\begin{array}{l} U_g(xyz)=\begin{cases} =U(x-ga,y,z)\ldots\text{для } g=kN+n \\ =U'(x-ga,y,z)\ldots\text{для } g=kN-\frac{1}{2}, \end{cases}\\ \varphi_g(xyz)=\begin{cases} =\varphi(x-ga,y,z)\ \ldots\text{для } g=kN+n \\ =\varphi'(x-ga,y,z)\ \ldots\text{для } g=kN-\frac{1}{2}, \end{cases}\\ W_g=\begin{cases} =W_0\ .\ .\ .\ .\ .\ .\ .\text{для } g=kN+n \\ =W'_0\ .\ .\ .\ .\ .\ .\text{для } g=kN-\frac{1}{2}. \end{cases}\end{array}\right\} \qquad (4{,}3)$$

Если иррегулярные атомы тождественны по своей природе с основными атомами решётки, то штрихи в последних формулировках следует опустить.

Обратимся теперь к решётке в целом. Уравнение Шредингера имеет вид:

$$H\psi = W\psi,$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V^*.$$

Здесь

$$V^*(xyz) = V(xyz) + V'(xyz),$$

где

V — потенциал идеальной решётки: $V(xyz) = V(x+a, y, z)$,

V' — возмущающий потенциал: $V'(x, y, z) = V(x+L, y, z)$.

Очевидно, потенциал V складывается из индивидуальных потенциалов всех регулярных атомов, а потенциал V' — из индивидуальных потенциалов всех иррегулярных атомов:

$$V(xyz) = \sum_g U(x - ga, y, z)\ldots \text{ (сумма по всем } g = kN + n),$$

$$V'(xyz) = \sum_g U'(x - ga, y, z)\ldots \left(\text{сумма по всем } g = kN - \frac{1}{2}\right).$$

Ход потенциала V^* изображён на рис. 55, *а*. Мы предполагаем, что при внесении лишних атомов регулярные атомы решётки продолжают оставаться в своих узлах, не испытывая никакого сдвига. Кроме того, мы будем предполагать, что возмущающий потенциал практически отличен от нуля лишь в непосредственной близости от иррегулярного атома, так что наличие иррегулярного атома чувствуют на себе лишь его непосредственные соседи. Иначе говоря, мы заменяем истинный ход потенциала V^*, изображённый на рис. 55, *а*, апроксимированным ходом, изображённым на рис. 55, *б*.

Наша задача отличается от задач, рассмотренных в предыдущих параграфах. Теперь в энергетическом спектре, как мы увидим ниже, появляется (наряду с зоной и локальными уровнями, обусловленными самим фактом нарушения периодичности) ещё собственный уровень чужеродного атома. Положение этого уровня в спектре зависит от химической природы иррегулярного атома. Этот уровень может попасть внутрь зоны или оказаться расположенным над зоной или под зоной. Здесь следует различать два случая.

1. Случай, когда собственный уровень примесного атома расположен достаточно далеко вне зоны (над ней или под ней).

2. Случай, когда собственный уровень примесного атома находится вблизи зоны или внутри неё. Количественный критерий этих двух случаев будет дан ниже. Соответственно

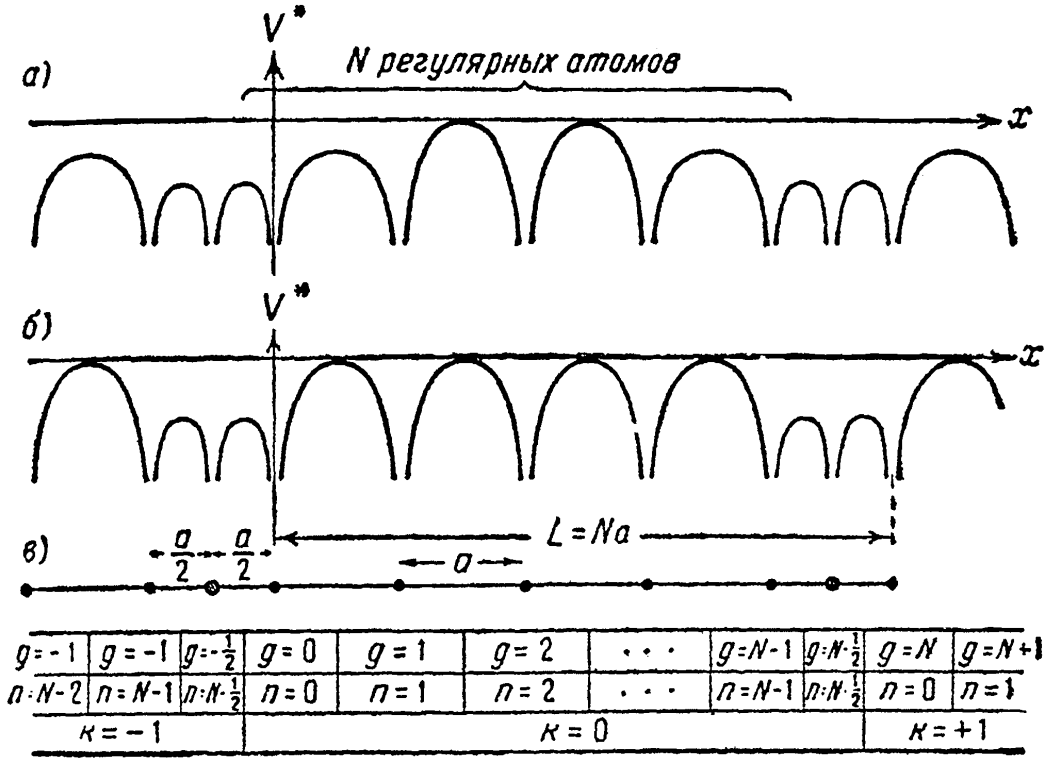


Рис. 55. Ход потенциала в решётке с междуузельными атомами.

этим двум случаям возможны два существенно различных подхода к задаче.

В первом случае волновую функцию ψ можно строить из волновых функций только одних регулярных атомов, или, наоборот, из волновых функций только иррегулярных атомов. Иначе говоря, волновую функцию можно искать в следующем виде:

$$\psi(xyz) = \sum_g a_g \varphi(x - ga, y, z), \qquad (4,4)$$

$$\psi(xyz) = \sum_g a_g \varphi'(x - ga, y, z), \qquad (4,5)$$

где в (4,4) суммирование проводится по всем $g = kN + n$, а в (4,5) — по всем $g = kN - \frac{1}{2}$.

Во втором случае уже недопустимо рассматривать регулярные и иррегулярные атомы раздельно. Волновую функцию следует искать в виде:

$$\psi(xyz) = \sum_{(g=kN+n)} a_g \varphi(x-ga, y, z) + \sum_{\left(g=kN-\frac{1}{2}\right)} a_g \varphi'(x-ga, y, z), \qquad (4,6)$$

т. е. в виде линейной комбинации из функций (4,4) и (4,5).

В. Случай, когда уровень примесного атома расположен достаточно далеко вне зоны. Если искать волновую функцию в виде (4,4), то для определения коэффициентов $a_g (g = kN+n)$ мы получаем систему уравнений:

$$\sum_{g'} a_{g'} \{(W_0 - W) A_{gg'} + B_{gg'}\} = 0, \qquad (4,7)$$

где

$$A_{gg'} = \int \bar{\varphi}_g \varphi_{g'} d\tau, \qquad B_{gg'} = \int \bar{\varphi}_g (V^* - U_{g'}) \varphi_{g'} d\tau, \qquad (4,8)$$

причём здесь

$$\left.\begin{aligned} g &= kN+n, \\ g' &= k'N+n'. \end{aligned}\right\}$$

Полагая

$$\left.\begin{aligned} A_{gg'} &= \begin{cases} =1 \ldots, & \text{если } g'=g \\ =0 \ldots, & \text{если } g' \neq g, \end{cases} \\ B_{gg'} &= \begin{cases} =\alpha \ldots, & \text{если } g'=g \neq kN,\ kN-1 \\ =\alpha' \ldots, & \text{если } g'=g=kN,\ kN-1 \\ =\beta \ldots, & \text{если } \begin{cases} g'=g+1 \neq kN \\ g'=g-1,\ g \neq kN-1 \end{cases} \\ =\beta' \ldots, & \text{если } \begin{cases} g'=g+1=kN \\ g'=g-1=kN-1 \end{cases} \\ =0 \ldots, & \text{во всех остальных случаях,} \end{cases} \end{aligned}\right\} \qquad (4,9)$$

мы можем переписать нашу систему (4,7) следующим образом:

$$\left.\begin{aligned}
&\text{для } n=1, 2, \ldots, N-2;\ k \text{ — любое}\\
&\quad (W_0 - W + \alpha)\, a_{k,n} + \beta\,(a_{k,n+1} + a_{k,n-1}) = 0,\\
&\text{для } n=0;\ k \text{ — любое}\\
&\quad (W_0 - W + \alpha')\, a_{k,0} + \beta a_{k,1} + \beta' a_{k-1,N-1} = 0,\\
&\text{для } n=N-1;\ k \text{ — любое}\\
&\quad (W_0 - W + \alpha') a_{k,N-1} + \beta a_{k,N-2} + \beta' a_{k+1,0} = 0.
\end{aligned}\right\} \quad (4,10)$$

Эта система представляет собой повторение системы (3,13), с которой мы имели дело в предыдущем параграфе. Поэтому при её решении можно сразу же воспользоваться результатами предыдущего параграфа.

Если искать волновую функцию в виде (4,3), то для определения коэффициентов a_g $\left(\text{где } g = kN - \frac{1}{2}\right)$ мы получаем систему уравнений:

$$\sum_{g'} a_{g'} \{(W_0' - W) A_{gg'} + B_{gg'}\} = 0, \quad (4,11)$$

где параметры $A_{gg'}$ и $B_{gg'}$ попрежнему имеют вид (4,8). Теперь, однако,

$$\left.\begin{aligned} g &= kN - \tfrac{1}{2},\\ g' &= k'N - \tfrac{1}{2}. \end{aligned}\right\}$$

Вместо (4,9) теперь следует положить:

$$\left.\begin{aligned}
A_{gg'} &= \begin{cases} = 1\ldots, \text{ если } g' = g = kN - \frac{1}{2} \\ = 0\ldots \text{во всех остальных случаях,} \end{cases}\\
B_{gg'} &= \begin{cases} = \alpha''\ldots, \text{ если } g' = g = kN - \frac{1}{2} \\ = 0\ldots \text{во всех остальных случаях,} \end{cases}
\end{aligned}\right\} \quad (4,12)$$

и систему (4,11) можно переписать так:

$$a_{kN-\frac{1}{2}} (W_0' - W + \alpha'') = 0 \text{ для всех } k. \quad (4,13)$$

Обратимся к исследованию энергетического спектра. Так как системы (4,10) и (4,13) не связаны между собой, мы

получаем в энергетическом спектре уровни двух происхождений: уровни, обеспечивающие условие совместности уравнений (4,10), а также уровни, обеспечивающие совместность уравнений (4,13). Расположение уровней зависит от соотношений между параметрами

$$\alpha'', \ \alpha', \ \alpha, \ \beta', \ \beta.$$

(Различные случаи изображены на рис. 56.)

Так, если

$$\left|\frac{\alpha' - \alpha \pm \beta'}{\beta}\right| \ll 1,$$

то в энергетическом спектре мы получаем зону, содержа-

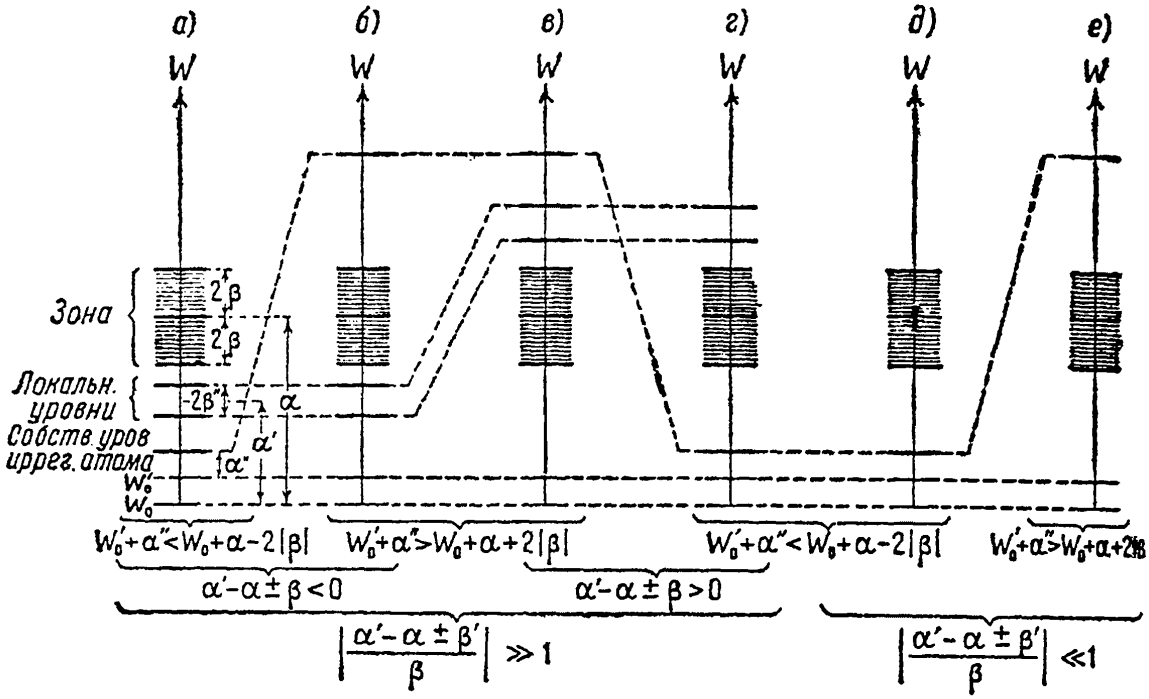


Рис. 56. Энергетический спектр решётки с междуузельными атомами.

щую N состояний (рис. 56, *д* и 56, *г*). Структура зоны определяется формулой

$$\boxed{W_j = W_0 + \alpha + 2\beta\cos\lambda_j.} \qquad (4,14)$$

Соответствующие волновые функции имеют вид (4,4), где

$$\boxed{a_{k,n} = \begin{cases} = C\cos\lambda_j\left(\frac{N-1}{2} - n\right), & \text{если } j \text{ нечётное,} \\ = -iC\sin\lambda_j\left(\frac{N-1}{2} - n\right), & \text{если } j \text{ чётное.} \end{cases}} \qquad (4,15)$$

Здесь

$$\boxed{\lambda_j = \frac{\pi j}{N+1} + (-1)^j \frac{2\beta}{\alpha' - \alpha - (-1)^j \beta'} \cdot \frac{1}{N+1} \sin \frac{\pi j}{N+1}, \quad \text{где } j = 1, 2, \ldots, N.} \tag{4,16}$$

Если же

$$\left| \frac{\alpha' - \alpha \pm \beta'}{\beta} \right| \gg 1,$$

то зона содержит $N-2$ состояний и, кроме того, в спектре появляются два локальных уровня (рис. 56, *а*, *б*, *в*, *г*). Структура зоны попрежнему определяется формулой (4,14), а соответствующие волновые функции (4,4) — формулами (4,15). Теперь, однако, вместо (4,16) мы имеем:

$$\boxed{\lambda_j = \frac{\pi j}{N-1} + (-1)^j \frac{2\beta}{\alpha' - \alpha - (-1)^j \beta'} \cdot \frac{1}{N-1} \sin \frac{\pi j}{N-1}, \quad \text{где } j = 1, 2, \ldots, N-2.} \tag{4,17}$$

Положение локальных уровней даётся формулой:

$$\boxed{W_j = W_0 + \alpha' - (-1)^j \beta'.} \tag{4,18}$$

Этим уровням соответствуют волновые функции (4,4), где коэффициенты $a_{k,n}$ попрежнему имеют вид (4,15), но теперь

$$\lambda_j = \pi r + i\lambda_j'',$$

причём

$$e^{|\lambda_j''|} = \left| \frac{\alpha' - \alpha - (-1)^j \beta'}{\beta} \right|.$$

Если при этом

$$\begin{cases} \dfrac{\alpha' - \alpha \pm \beta'}{\beta} > 0, & \text{то } r \text{ чётное,} \\ \dfrac{\alpha' - \alpha \pm \beta'}{\beta} < 0, & \text{то } r \text{ нечётное.} \end{cases}$$

Локальные уровни (4,18) располагаются

над зоной, если $\alpha' - \alpha \pm \beta' > 0$ (см. рис. 56, *в*, *г*)
под зоной, если $\alpha' - \alpha \pm \beta' < 0$ (см. рис. 56, *а*, *б*).

На всю эту энергетическую картину накладывается еще собственный уровень чужеродного атома:

$$\boxed{W = W_0' + \alpha''.} \tag{4,19}$$

Соответствующая волновая функция имеет вид (4,5). Этот уровень помещается

над зоной, если $W_0' + \alpha'' > W_0 + \alpha + 2|\beta|$ (см. рис. 56, *б*, *в*, *е*)
под зоной, если $W_0' + \alpha'' < W_0 + \alpha - 2|\beta|$ (см. рис. 56, *а*, *г*, *д*).

Электрон, принадлежащий зоне (4,14), является коллективизированным электроном. Электрон, принадлежащий уровням (4,18), практически локализован на каком-либо из регулярных атомов, непосредственно соседних к повреждению. Электрон, принадлежащий уровню (4,19), локализован на каком-либо из иррегулярных атомов.

Если уровень (4,19) помещается над зоной (рис. 56, *б*, *в*, *е*), то это значит, что чужеродный атом, внесённый в решётку, теряет свой валентный электрон. Этот электрон переходит к регулярным атомам решётки. Если же уровень (4,19) оказывается расположенным под зоной (рис. 56, *а*, *г*, *д*), то это значит, что иррегулярный атом сохраняет на себе свой электрон. В первом случае перевод электрона с регулярного атома на иррегулярный требует некоторой затраты энергии. Во втором случае, наоборот, затратой энергии сопровождается перевод электрона с иррегулярного атома на регулярный.

С. Гетерополярная решётка с междуузельным металлическим атомом. Остановимся особо на случае, когда мы имеем дело с гетерополярной решёткой и когда междуузельные атомы тождественны по своей природе с металлическими атомами самой решётки. Такая решётка представляет собой решётку со стехиометрическим избытком металла (например, решётка NaCl с междуузельными атомами Na). В этом случае мы имеем

$$\left.\begin{aligned} &W_0' = W_0, \\ &\alpha > \alpha' > \alpha'' > 0, \\ &\beta > \beta' > 0. \end{aligned}\right\}$$

Если считать ионы точечными, а электрон, принадлежащий g-му атому, считать сосредоточенным в точке $x = ga$, то

параметры α, α', α'', β, β' превращаются в свои классические аналоги. В трёхмерном случае имеем:

$$\left.\begin{aligned} &\alpha \longrightarrow \frac{e^2}{a_0} M, \\ &\alpha' \longrightarrow \frac{e^2}{a_0}\left(M - \frac{1}{\sqrt{3}}\right), \\ &\alpha'' \longrightarrow 0, \\ &\beta \longrightarrow 0, \\ &\beta' \longrightarrow 0. \end{aligned}\right\}$$

Энергетический спектр изображён на рис. 56, *а*. Собственные уровни междуузельных атомов являются донорами (они заняты электронами). Действительно, введение в решётку лишнего атома металла можно рассматривать как введение положительного металлического иона и компенсирующего электрона. Решётка, обладающая избыточными металлическими ионами, содержит также и избыточные электроны. Эти электроны, однако, остаются привязанными к своим атомам, ибо собственные уровни междуузельных атомов являются энергетически наивыгоднейшими для электронов.

Стехиометрический избыток металла может быть осуществлён, как мы знаем, или за счёт введения в решётку лишних атомов металла (междуузельные металлические атомы) или за счёт удаления атомов металлоида (пустые металлоидные узлы). Первый случай рассматривается в этом параграфе, в то время как второй случай был рассмотрен в предыдущем параграфе. Сравнивая результаты этого параграфа с результатами предыдущего, мы замечаем, что в обоих случаях характер энергетического спектра получается в известном смысле одинаковым: в обоих случаях мы получаем систему доноров, расположенных под зоной проводимости. Однако уровни междуузельных металлических атомов располагаются глубже, чем уровни пустых металлоидных узлов. Энергия ионизации u в первом случае больше, чем во втором. При классическом подходе к задаче:

для междуузельных металлических атомов: $u = \frac{e^2}{a_0} M$,

для пустых металлоидных узлов: $u = \frac{e^2}{a_0}$.

Это значит, что зоне проводимости легче черпать электроны с пустых металлоидных узлов, чем с иррегулярных металлических атомов.

D. Случай, когда собственный уровень примесного атома расположен вблизи зоны. Напомним в заключение, что всё сказанное здесь остаётся справедливым лишь до тех пор, пока собственный уровень междуузельного атома лежит достаточно далеко за пределами зоны. Для гетерополярной решётки, обладающей стехиометрическим избытком металла, это условие выглядит так:

$$\alpha' \gg \alpha''.$$

Если это условие не соблюдено, то собственный уровень междуузельного атома вступает во взаимодействие с основными уровнями самой решётки, и вся картина искажается. При этом уже нельзя пользоваться волновой функцией (4,4) и (4,5), как мы это делали здесь, а следует искать её в виде (4,6). Вместо двух систем (4,10) и (4,13), не связанных между собой, из которых первая служит для определения коэффициентов a_{kN+n}, а вторая — для определения коэффициентов $a_{kN-\frac{1}{2}}$, мы получим теперь одну общую систему уравнений. Решение её связано с некоторыми математическими трудностями.

Литература к главе III

Цитированные оригинальные статьи:

[1] Frenkel, *Sow. Phys.* **9**, 158, 1936.
[2] Landau, *Sow. Phys.* **3**, 664, 1933.
[3] Tamm, *Sow. Phys.* **1**, 733, 1932; *Zs. f. Ph.* **76**, 849, 1932.
[4] Maue, *Zs. f. Phys.* **94**, 717, 1935.
[5] Goodwin, *Proc. of Cambr. Phil. Soc.* **35**, 205.
[6] Shockey, *Phys. Rev.* **56**, 317, 1939.
[7] Sokolow und Machalowa, *Zs. f. Ph.* **99**, 503, 1936.

ГЛАВА IV

ВЛИЯНИЕ ТЕМПЕРАТУРЫ И ПРИМЕСЕЙ НА ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

§ 1. Обзор экспериментальных и теоретических работ

А. Сводка экспериментальных данных. Электрические свойства полупроводников стали предметом экспериментального исследования сравнительно недавно. Однако к настоящему времени накоплен уже очень богатый экспериментальный материал.

В отношении электропроводности полупроводники занимают промежуточное положение между металлами и диэлектриками. Электропроводность металлов имеет порядок величины $\sigma \approx 10^5$ *ом*$^{-1}$ *см*$^{-1}$. Электропроводность диэлектриков при обычных условиях порядка $\sigma \approx 10^{-15}$ *ом*$^{-1}$ *см*$^{-1}$. Полупроводники охватывают широкую промежуточную область. Электропроводность полупроводников колеблется в пределах от $\sigma \approx 10^2$ до $\sigma \approx 10^{-10}$ *ом*$^{-1}$ *см*$^{-1}$. Таким образом, к группе полупроводников относятся тела с весьма разнообразными значениями σ. Полупроводники отличаются от металлов, с одной стороны, и от диэлектриков, с другой стороны, не только количественно, но и качественно. Они имеют ряд черт, общих с металлами, и в то же время ряд черт, объединяющих их с диэлектриками. Мы отметим три таких черты.

1. Электропроводность металлов имеет электронную природу. Диэлектрические кристаллы, наоборот, обладают ионной проводимостью. В этом отношении полупроводники схожи с металлами: как и в металлах, электропроводность полупроводников имеет электронное происхождение.

2. Электропроводность металлов уменьшается при введении примесей. Электропроводность диэлектриков, наоборот, при введении примесей возрастает. В этом отношении полу-

проводники похожи на диэлектрики: включение примесей приводит к резкому увеличению электропроводности полупроводников.

3. Электропроводность металлов при нагревании медленно падает. Электропроводность диэлектриков, наоборот, резко возрастает. В этом отношении полупроводники можно разделить на два класса. Известны полупроводники, которые обладают металлическим ходом электропроводности. В то же время существует обширная группа полупроводников, у которых зависимость электропроводности от температуры имеет тот же характер, что и у диэлектриков.

Электропроводность полупроводников существенно зависит от внешних условий. Воздействуя на полупроводник теми или иными физическими факторами, можно изменять его проводимость часто в миллионы раз. К таким факторам относятся: а) концентрация примесей, б) температура, в) сильное электрическое поле, г) освещение.

На эти факторы электропроводность полупроводников, как правило, реагирует чрезвычайно резко.

Влияние сильного электрического поля на электропроводность будет рассмотрено в следующей главе. Влиянию освещения будет посвящена глава VI. Здесь же мы рассмотрим зависимость электропроводности от примесей и температуры.

а) Слово «примесь» не следует понимать слишком буквально как наличие чужеродных атомов, введённых в решётку извне. Этот термин имеет более широкий смысл, обозначая дефекты любого сорта, присутствующие в решётке и нарушающие строгую периодичность её структуры. В частности, в полупроводниках, представляющих собой соединения типа M_mR_r, под наличием «примеси» обычно понимают всякое отклонение от стехиометрического соотношения. В этом случае «примесным» полупроводником называют всякий полупроводник, в котором стехиометрическое соотношение в той или иной степени нарушено, в отличие от «чистого» полупроводника, в котором стехиометрическое соотношение соблюдено.

Среди полупроводников типа M_mR_r нам известны два сорта примесных полупроводников: полупроводники со стехиометрическим избытком металла и полупроводники со стехиометрическим избытком металлоида. В полупроводниках с избытком металла проводимость имеет электронную природу. В полупроводниках с избытком металлоида — дырочную природу.

Это является достаточно твёрдо установленным экспериментальным фактом. Заметим, что в некоторых случаях мы умеем делать полупроводник электронным или дырочным по нашему желанию, вызывая в нём стехиометрические отклонения в ту или иную сторону.

Для всех полупроводников характерна одна и та же закономерность: малейшие стехиометрические изменения вызывают резкое изменение в величине электропроводности. Электропроводность полупроводника тем больше, чем больше содержание «примеси», т. е. чем сильнее нарушено стехиометрическое соотношение. Это также является достаточно прочно установленным экспериментальным фактом. Принято считать, что именно наличию «примесей» полупроводник обязан своею электропроводностью. «Чистые» полупроводники, с которыми, впрочем, мы никогда не имеем дела в действительности, с этой точки зрения являются по существу диэлектриками. Эта мысль впервые была высказана Гуденом [1].

Резкая зависимость электропроводности от концентрации примеси может служить, повидимому, объяснением того, почему измерения различных авторов обычно столь резко расходятся между собой. Различные авторы при исследовании одного и того же материала, естественно, имеют дело с различными образцами, различным образом приготовленными и, следовательно, различающимися содержанием примеси. Действительно, данные, полученные на различных образцах, как правило, настолько различны, что это обстоятельство вряд ли может считаться случайным и вряд ли может быть отнесено за счёт несовершенства экспериментов, а имеет, повидимому, более глубокое происхождение. В связи с этим возникает сомнение в том, может ли вообще электропроводность полупроводника считаться константой материала, подобно тому как мы это считаем в отношении металлов или диэлектриков.

б) Зависимость электропроводности от температуры для различных полупроводников различна. Различные типы этой зависимости схематически изображены на рис. 57. На этом же рисунке выписаны некоторые характерные представители каждого типа. Таким образом, в не слишком широком температурном интервале электропроводность полупроводника всегда может быть выражена экспоненциальным законом:

$$\sigma = Ce^{-\frac{B}{T}}. \qquad (1,1)$$

В частности, мы можем иметь здесь $B = 0$, что соответствует металлическому типу проводимости. Этот тип проводимости (рис. 57, *а*) встречается у полупроводников с достаточно большим σ.

Коэффициент C в (1,1) или вовсе не зависит от концентрации примеси, или же возрастает вместе с возрастанием концентрации. В последнем случае это соответствует перемещению прямых на рис. 57 вверх, параллельно самим себе, как это показано на рис. 58, *а*.

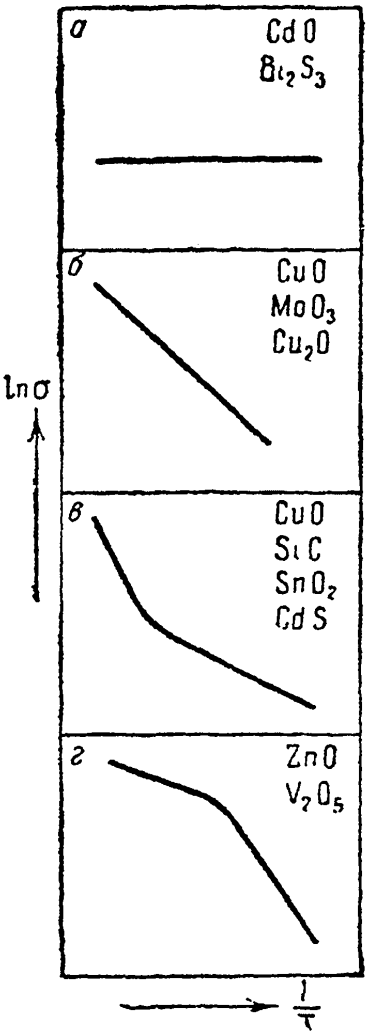


Рис. 57. Различные типы зависимости электропроводности от температуры.

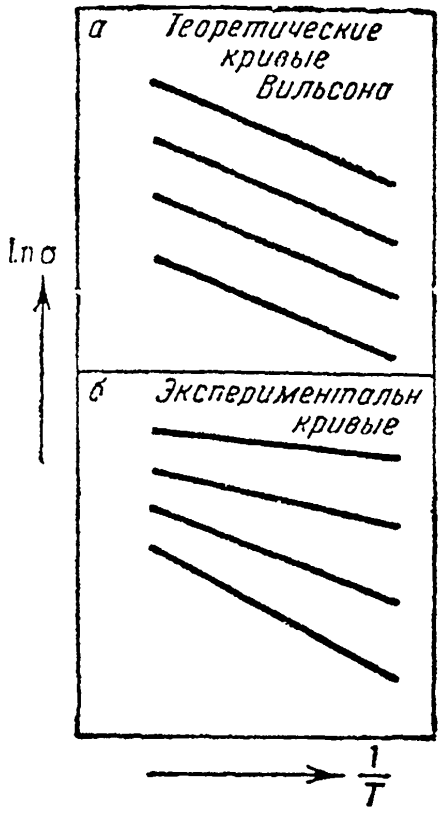


Рис. 58. Температурный ход электропроводности для образцов с различным содержанием примеси.

Коэффициент B также или вовсе не зависит от концентрации примеси, или же убывает при возрастании концентрации. В последнем случае это соответствует уменьшению угла наклона прямых на рис. 57. По мере увеличения концентрации прямые становятся более пологими, как это изображено на рис. 58, *б*.

При переходе от одного температурного интервала к другому значения коэффициентов C и B меняются, что соответствует излому прямых на рис. 57. При переходе от низких температур в область высоких температур коэффициенты C и B у одних полупроводников увеличиваются (рис. 57, *в*), в то время как у других, наоборот, уменьшаются (рис. 57, *г*).

Вместо обзора экспериментальных работ мы ограничимся здесь более или менее полной сводкой экспериментальных кривых. Эта сводка приводится здесь в виде целой серии рисунков, на которых воспроизведены зависимости $\sigma = \sigma(T)$, полученные разными авторами для разных материалов (см. рис. 59). Над каждой кривой приведено название материала и фамилия автора; номер в прямых скобках обозначает ссылку на библиографию, приведённую в конце главы. В тех случаях, когда на одном и том же рисунке изображены сразу несколько кривых, эти кривые относятся к различным образцам одного и того же материала (т. е. к образцам, по-разному обработанным). За подробностями мы отсылаем читателя к соответствующим оригинальным работам. Обзоры экспериментального материала (хотя и несколько устаревшие) даны в статьях Бронштейна и Арсеньевой [13], Гуддена [14] и в книжке Иоффе [15].

Рассматривая рис. 59, мы видим, что различные типы температурной зависимости не связаны непосредственно с химической природой полупроводника. Различные образцы одного и того же полупроводника могут давать зависимости различного типа. Таким образом, для полупроводников не только величина электропроводности, но и температурный ход электропроводности не являются характеристикой материала, как это имеет место для металлов и диэлектриков.

Теории электропроводности полупроводников, предложенные до сих пор, не охватывают всей совокупности экспериментальных фактов, перечисленных выше. Они объясняют лишь отдельные частные случаи. Мы ограничимся здесь рассмотрением работы Вильсона, явившейся первой попыткой построения теории полупроводников и дающей теоретическую интерпретацию формуле (1,1), а также работы Нейбура, объясняющей изломы температурной кривой для ZnO (см. рис. 59).

В. Теория Вильсона. Согласно Гуддену, проводимость полупроводников обусловливается дефектами кристаллической решётки. Теория Вильсона [16] представляет собой развитие

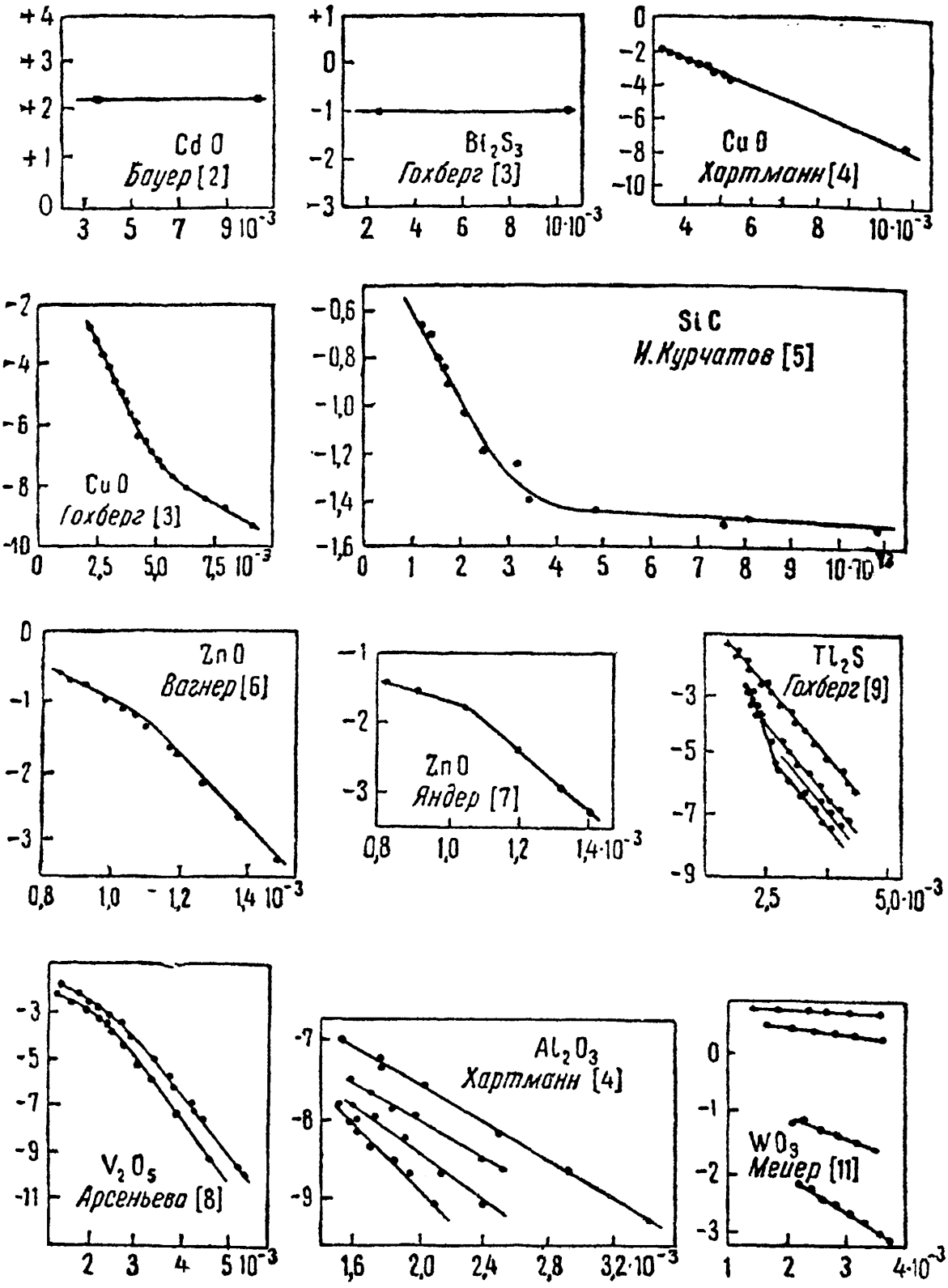


Рис. 59. Зависимость электропроводности от температуры $\ln \sigma = f\left(\frac{1}{T}\right)$ (сводка экспериментальных кривых).

(По оси ординат отложено $\ln \sigma$ [в случае Cu_2O — $\lg (\sigma \cdot 10^4)$]; по оси абсцисс — $\frac{1}{T}$).

этой идеи. Теория относится к случаю, когда решётка содержит в себе дефекты лишь одного какого-либо сорта. Это значит,

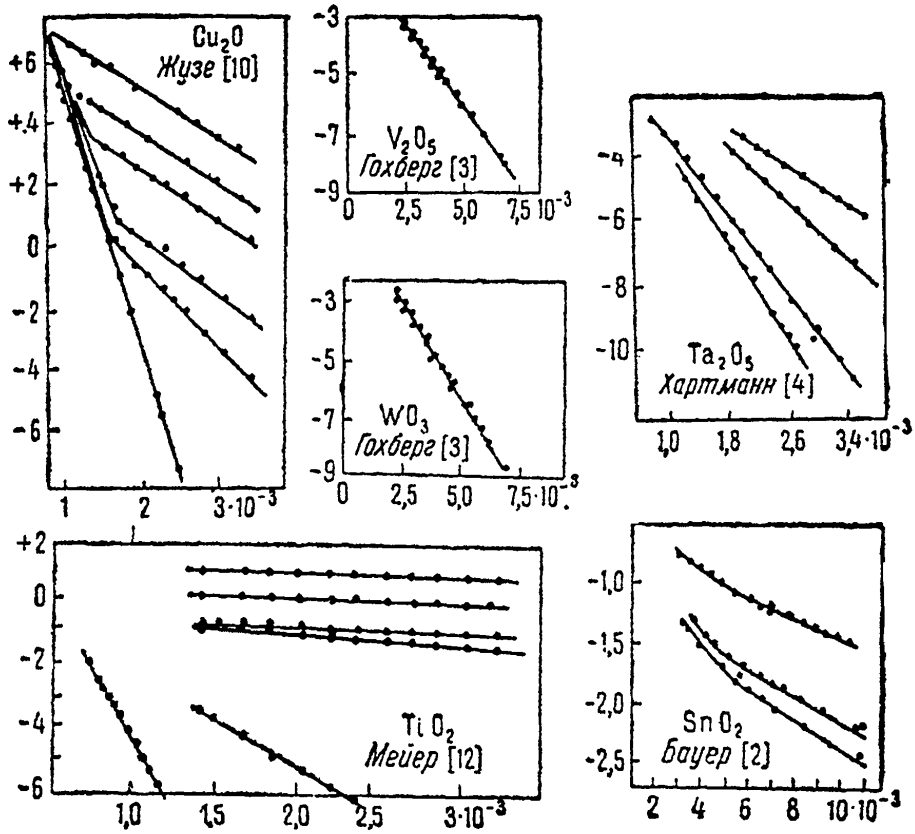


Рис. 59 (продолжение).

что всякая гетерополярная решётка при соблюдённом стехиометрическом соотношении должна считаться идеальной; дефекты в такой гетерополярной решётке всегда свидетельствуют о нарушении стехиометрии.

Энергетический спектр решётки изображён на рис. 60: в запрещённом участке между нормальной зоной и зоной проводимости мы имеем систему локальных уровней. Эти уровни порождены дефектами решётки. Все уровни расположены на одной горизонтали; это свидетельствует о том, что решётка содержит дефекты только одного определённого сорта. В нормальном состоянии (т. е. при абсолютном нуле темпе-

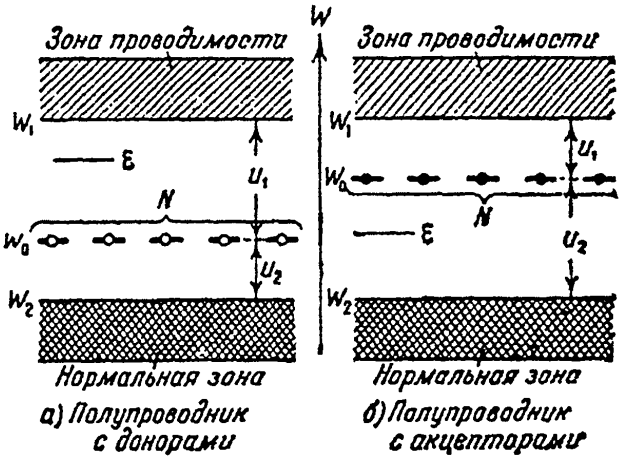


Рис. 60. Схема уровней для полупроводника с дефектами одного сорта (схема Вильсона).

ратуры) нижняя зона заполнена электронами до отказа; верхняя зона свободна. Что же касается локальных уровней, то в зависимости от их природы возможны два случая: в нормальном состоянии локальные уровни могут быть заняты электронами или же могут оказаться свободными. Эти два случая следует различать между собой.

Первый случай изображён на рис 60, *а*. В этом случае локальные уровни служат поставщиками электронов: они могут отдавать свои электроны зоне проводимости. Такие локальные уровни мы будем называть *донорами*. Второй случай изображён на рис. 60, *б*. В этом случае локальные уровни играют роль ловушек: они могут принимать на себя электроны из нормальной зоны. Такие локальные уровни мы будем называть *акцепторами*.

Рисунок 60, *а* относится к полупроводникам со стехиометрическим избытком металла. В таких полупроводниках электропроводность имеет электронную природу. Рисунок 60, *б* относится к полупроводникам со стехиометрическим избытком металлоида. Такие полупроводники обладают дырочной электропроводностью.

Размещение электронов по уровням, изображённое на рис. 60, *а* и 60 *б*, соответствует абсолютному нулю температуры. При повышении температуры электроны переводятся с более низких на более высокие уровни: в нормальной зоне возникают дырки; в зоне проводимости появляются электроны; доноры частично освобождаются от электронов; акцепторы частично заполняются электронами.

Рассмотрим условия теплового равновесия в полупроводнике. Ограничимся полупроводником с донорами (рис. 60. *а*). Результаты могут быть перенесены на случай акцепторного полупроводника (рис. 60, *б*).

Обозначим через $n_1(W, T)$ число электронов на некотором уровне W при температуре T, а через $n_2(W, T)$ — число дырок на том же уровне при той же температуре (отнесённые к единице объёма кристалла). Очевидно (см. главу II, § 1, Е):

$$\left.\begin{aligned} n_1(W, T) + n_2(W, T) &= Z(W), \\ n_1(W, T) &= \frac{Z(W)}{e^{\frac{W-\varepsilon}{kT}} + 1}, \\ n_2(W, T) &= \frac{Z(W)}{e^{\frac{\varepsilon-W}{kT}} + 1}, \end{aligned}\right\} \qquad (1,2)$$

где ε — уровень химического потенциала, $Z(W)$ — число состояний, соответствующих энергии W.

Функция $Z(W)$ имеет вид (см. главу II, § 2, C и § 1, E)
а) для зоны проводимости вблизи её нижней границы ($W \geqslant W_1$):

$$Z(W) = G_1 \frac{2\pi}{h^3} (2m_1)^{3/2} \sqrt{W - W_1}, \qquad (1,3а)$$

б) для нормальной зоны вблизи её верхней границы ($W \leqslant W_2$):

$$Z(W) = G_2 \frac{2\pi}{h^3} (2m_2)^{3/2} \sqrt{W_2 - W}, \qquad (1,3б)$$

в) для локальных уровней ($W = W_0$):

$$Z(W) = N. \qquad (1,3в)$$

Здесь

W_1 — нижняя граница зоны проводимости,
W_2 — верхняя граница нормальной зоны,
N — концентрация дефектов в решётке (число локальных уровней на единицу объёма),
m_i — эффективная масса электрона в зоне проводимости ($i = 1$) и в нормальной зоне ($i = 2$),
G_i — статистический вес зоны проводимости ($i = 1$) и нормальной зоны ($i = 1$) [1]).

Положим далее, что при температуре T мы имеем (на 1 $см^3$ кристалла):

$n_0(T)$ — число электронов на локальных уровнях,
$n_1(T)$ — число электронов в зоне проводимости,
$n_2(T)$ — число дырок в нормальной зоне.

Очевидно:

$$n_0(0) = N, \quad n_1(0) = n_2(0) = 0.$$

Кроме того, очевидно, что между числами $n_0(T)$, $n_1(T)$, $n_2(T)$ при любом T соблюдается следующее соотношение:

$$n_0(T) + n_1(T) - n_2(T) = N. \qquad (1,4)$$

Это соотношение выражает собой закон сохранения заряда.

[1]) Для s-зоны мы имеем $G = 2$. Для p-зоны — $G = 6$.

Согласно (1,2) и (1,3), мы имеем:

$$\left\{\begin{array}{l} n_0(T)=\dfrac{N}{e^{\frac{W_0-\varepsilon}{kT}}+1}, \qquad (1,5)\\[2ex] n_1(T)=\int n_1(W,T)\,dW=\dfrac{2}{\sqrt{\pi}}A_1\int\limits_{e^{\frac{W-\varepsilon}{kT}}+1}^{}\dfrac{\sqrt{W-W_1}\,dW}{}, \qquad (1,6)\\[2ex] n_2(T)=\int n_2(W,T)\,dW=\dfrac{2}{\sqrt{\pi}}A_2\int\limits_{e^{\frac{\varepsilon-W}{kT}}+1}^{}\dfrac{\sqrt{W_2-W}\,dW}{}, \qquad (1,7)\end{array}\right.$$

где

$$A_i=\frac{G_i}{h^3}(2\pi m_i kT)^{3/2}.$$

В выражении (1,6) интегрирование производится от нижней границы зоны проводимости (т. е. от $W=W_1$) до её верхней границы (практически до $W=+\infty$); в выражении (1,7) — от нижней границы нормальной зоны (практически от $W=-\infty$) до её верхней границы (т. е. до $W=W_2$). Интегралы (1,6) и (1,7) могут быть вычислены приближённо в предположении, что электронный и дырочный газы слабо вырождены, т. е.

$$\left.\begin{array}{l} e^{\frac{W_1-\varepsilon}{kT}}\gg 1,\\ e^{\frac{\varepsilon-W_2}{kT}}\gg 1.\end{array}\right\} \qquad (1,8)$$

Условия (1,8) заключают в себе предположение, что

$$W_1<\varepsilon<W_2,$$

т. е., что уровень химического потенциала лежит между зонами. В этом случае выражения (1,6) и (1,7) принимают такой вид:

$$\left.\begin{array}{l} n_1(T)=A_1 e^{-\frac{W_1-\varepsilon}{kT}},\\ n_2(T)=A_2 e^{-\frac{\varepsilon-W_2}{kT}}.\end{array}\right\} \qquad (1,9)$$

Эти выражения содержат неизвестный параметр ε. Он может быть определён из уравнения (1,4). Действительно, подставляя (1,9)

и (1,5) в (1,4), мы получим

$$A_1 e^{-\frac{W_1-\varepsilon}{kT}} - A_2 e^{-\frac{\varepsilon-W_2}{kT}} = \frac{N}{e^{\frac{\varepsilon-W_0}{kT}}+1}. \qquad (1,10)$$

Это уравнение решается относительно ε особенно просто, если предположить заранее, что уровень ε располагается над локальными уровнями, и притом так, что

$$e^{\frac{\varepsilon-W_0}{kT}} \gg 1. \qquad (1,11)$$

В этом случае, принимая обозначения

$$\begin{cases} u_1 = W_1 - W_0, \\ u_2 = W_0 - W_2, \end{cases}$$

мы получаем из (1,10)

$$\varepsilon = W_1 - \frac{kT}{2} \ln \frac{A_1 e^{\frac{u_1}{kT}}}{N + A_2 e^{-\frac{u_2}{kT}}}. \qquad (1,12)$$

Таким образом, положение уровня ε в спектре определяется температурой T и концентрацией дефектов N.

Подставляя (1,12) в (1,9), будем иметь:

$$\left.\begin{aligned} n_1(T) &= \sqrt{A_1 e^{-\frac{u_2}{kT}} \left(N + A_2 e^{-\frac{u_2}{kT}}\right)}, \\ n_2(T) &= A_2 e^{-\frac{u_2}{kT}} \sqrt{\frac{A_1 e^{-\frac{u_1}{kT}}}{N + A_2 e^{-\frac{u_2}{kT}}}}. \end{aligned}\right\} \qquad (1,13)$$

Проводимость σ полупроводника складывается из двух составляющих:

$$\sigma = e v_1 n_1(T) + e v_2 n_2(T), \qquad (1,14)$$

где

$\left\{\begin{aligned} &v_1 \text{ — средняя подвижность электронов в зоне проводимости,} \\ &v_2 \text{ — средняя подвижность дырок в нормальной зоне,} \\ &e \text{ — абсолютная величина заряда электрона.} \end{aligned}\right.$

Выражение для v_i может быть заимствовано из теории металлов. Мы имеем:

$$v_i = \frac{4}{3}\frac{l_i}{\sqrt{2\pi m_i kT}},$$

где l_i — длина свободного пробега для электрона ($i=1$) или для дырки ($i=2$).

Первое слагаемое в (1,14) соответствует электронной составляющей, а второе слагаемое — дырочной составляющей проводимости. Соотношение между этими составляющими меняется с температурой.

Подставляя (1,13) в (1,14) и полагая

$$\begin{cases} v_1 \approx v_2 \approx v, \\ A_1 \approx A_2 \approx A, \end{cases}$$

мы получим σ как функцию T и N:

$$\boxed{\sigma = ev\sqrt{Ae^{-\frac{u_1}{kT}}}\left\{\frac{N+2Ae^{-\frac{u_2}{kT}}}{\sqrt{N+Ae^{-\frac{u_2}{kT}}}}\right\}.} \qquad (1,15)$$

В этой формуле параметры v и A можно считать практически не зависящими от температуры (по сравнению с экспоненциальными членами).

Формулы (1,12) и (1,15) существенно упрощаются в двух крайних случаях:

а) в случае достаточно низких температур, когда

$$N \gg Ae^{-\frac{u_2}{kT}},$$

б) в случае достаточно высоких температур, когда

$$N \ll Ae^{-\frac{u_2}{kT}}.$$

В первом случае мы имеем:

$$\boxed{\begin{aligned} &\varepsilon = W_1 - \frac{1}{2}u_1 - \frac{1}{2}kT\ln\frac{A}{N}, \\ &\sigma = ev\sqrt{AN}e^{-\frac{u_1}{2kT}}. \end{aligned}} \qquad (1,16)$$

Во втором случае:

$$\boxed{\begin{aligned}\varepsilon &= W_1 - \frac{1}{2}(u_1 + u_2),\\ \sigma &= ev2Ae^{-\frac{u_1+u_2}{2kT}}.\end{aligned}} \qquad (1,17)$$

Эти формулы совпадают с эмпирической формулой (1,1), причём параметры B и C в (1,1) приобретают теперь следующий смысл:

$$\left.\begin{aligned} B &= \frac{u_1}{2k}, \quad C = ev\sqrt{AN},\\ \text{или}\qquad & \\ B &= \frac{u_1+u_2}{2k}, \quad C = ev2A.\end{aligned}\right\} \qquad (1,18)$$

Зависимости $\varepsilon = \varepsilon(T)$ и $\sigma = \sigma(T)$ для разных N [согласно (1,12) и (1,15)] изображены соответственно на рис. 61 и 62. На этих рисунках различные кривые соответствуют различным значениям N; кривые перенумерованы в порядке возрастания N. Область, заштрихованная на рис. 61, представляет собой ту область, внутри которой теория перестаёт быть применимой,

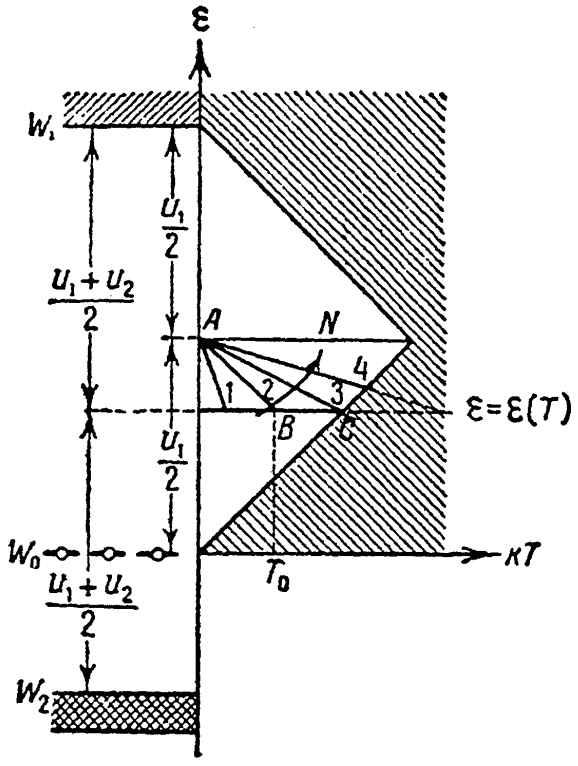


Рис. 61. Положение в энергетическом спектре уровня химического потенциала (ε как функция T при разных N).

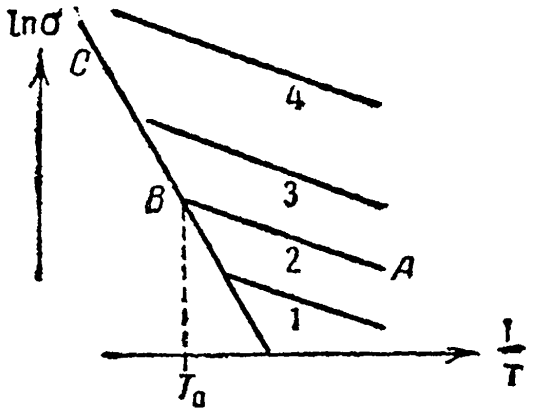


Рис. 62. Зависимость электропроводности от температуры при разных концентрациях дефектов (схема теоретических кривых).

т. е. в которой нарушаются исходные условия (1,8) и (1,11). Из рис. 61 видно, как с изменением T перемещается в энер-

гетическом спектре уровень ε: при абсолютном нуле этот уровень располагается посредине между зоной проводимости и локальными уровнями; при достаточно высоких температурах он опускается на середину запрещённого участка между зоной проводимости и нормальной зоной.

Точка перелома на кривых рис. 61 и рис. 62 (точка B) соответствует температуре $T = T_0$, которая определяется из условия

$$N = Ae^{-\frac{u_2}{kT_0}}.$$

При $T \ll T_0$ мы имеем *область низких температур* (участок AB). В этой области справедливы формулы (1,16). При $T \gg T_0$ мы имеем *область высоких температур* (участок BC). В этой области справедливы формулы (1,17).

Переход из области низких температур в область высоких температур (от участка AB к участку BC) свидетельствует о перемене механизма проводимости. На участке AB электроны проводимости черпаются с локальных уровней; нормальная зона при этом практически не участвует в игре. На этом участке электропроводность имеет чисто электронную природу. На участке BC электроны поставляются в зону проводимости в основном из нормальной зоны, а роль локальных уровней, как поставщиков электронов, практически сведена к нулю. На этом участке электропроводность имеет смешанную природу. Участок AB соответствует *примесной* проводимости, т. е. проводимости, целиком обусловленной дефектами решётки; участок BC соответствует *собственной* проводимости, т. е. той проводимости, которой должна была бы обладать идеальная решётка, вовсе лишённая дефектов.

Заметим в заключение, что все полученные результаты могут быть перенесены на случай, когда локальные уровни в полупроводнике являются не донорами, а акцепторами. (рис. 60, *б*). В этом случае мы приходим к тем же формулам, с той лишь разницей, что во всех этих формулах значки 1 и 2 следует теперь поменять местами.

Сравнивая теоретические результаты с экспериментальными данными, мы видим, что теория Вильсона

1) даёт правильный экспоненциальный закон (1,1);

2) объясняет зависимость константы C в (1,1) от концентрации дефектов;

3) позволяет теоретически интерпретировать семейства кривых, полученные Гохбергом и Жузе для Tl_2S и Cu_2O (ср. соответствующие кривые на рис. 59 с теоретическими кривыми рис. 62).

В то же время мы замечаем, что теория Вильсона обладает следующими недостатками:

1. Теория Вильсона не объясняет влияния концентрации дефектов на значение коэффициента B в (1,1). В теории Вильсона коэффициент B целиком определяется природой решётки и природой дефектов (ср. рис. 58, *а* с рис. 58, *б*).

2. Характер зависимости коэффициента C в (1,1) от концентрации дефектов, к которому приводит теория Вильсона [см. формулы (1,18)], оправдывается на опыте далеко не всегда.

3. Не все типы изломов на температурной кривой могут быть объяснены теорией Вильсона.

4. Полупроводники с металлическим ходом электропроводности (см. рис. 57, *а*) остаются за рамками вильсоновской теории.

Таким образом, теория Вильсона даёт правдоподобные результаты для отдельных частных случаев, далеко не охватывая всей совокупности экспериментальных фактов.

С. Теория Нейбура. Рассмотрим полупроводник, энергетический спектр которого содержит одновременно как уровни-доноры, так и уровни-акцепторы, расположенные при этом на одной и той же горизонтали, как это изображено на рис. 63. У полупроводников такого типа, как это показал Нейбур [7], кривая электропроводности имеет вид, изображённый на рис. 57, *г*. Таким температурным ходом электропроводности обладает ZnO (по измерениям Вагнера [6] и Яндера [7], а также V_2O_5 (по измерениям Арсеньевой [8]).

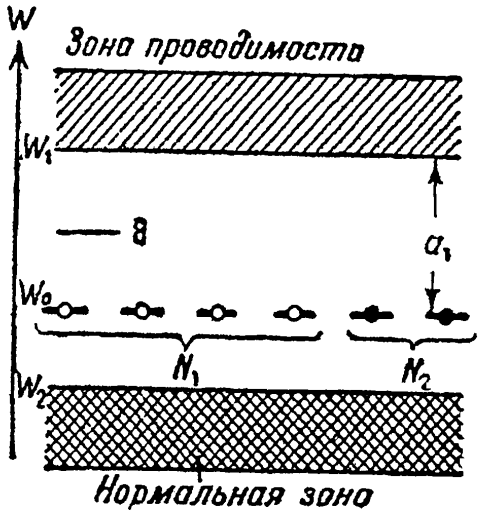


Рис. 63. Полупроводник с акцепторами и донорами (схема Нейбура).

Рассмотрим условия теплового равновесия в полупроводнике, характеризующемся энергетической схемой, изображённой на рис. 63. При этом будем пренебрегать тепловыми перебросами

электронов из нормальной зоны. Будем считать, что в тепловом обмене электронами принимают участие лишь зона проводимости и локальные уровни. Таким образом, нормальную зону мы вовсе исключаем из нашего рассмотрения. Это можно делать до тех пор, пока уровень химического потенциала, расположенный в запрещённом участке, лежит значительно ближе к верхней зоне, чем к нижней.

Пусть мы имеем (на единицу объёма):
N_1 — число доноров,
N_2 — число акцепторов,
$n_1(T)$ — число дырок в зоне проводимости при температуре T,
$n_2(T)$ — число дырок на локальных уровнях при температуре T.

Очевидно

$$n_1(0)=0, \qquad n_2(0)=N_2.$$

Считая, что уровень химического потенциала ε помещается между локальными уровнями и зоной проводимости и притом так, что

$$\begin{cases} e^{\frac{W_1-\varepsilon}{kT}} \gg 1, \\ e^{\frac{\varepsilon-W_0}{kT}} \gg 1, \end{cases} \tag{1,19}$$

мы будем иметь:

$$\begin{cases} n_1(T)=A_1 e^{-\frac{W_1-\varepsilon}{kT}}, \\ n_2(T)=(N_1+N_2)\, e^{-\frac{\varepsilon-W_0}{kT}}. \end{cases} \tag{1,20}$$

Условие сохранения заряда может быть записано так:

$$n_1(T)+N_2=n_2(T). \tag{1,21}$$

Подставляя сюда (1,20), мы можем определить из этого уравнения неизвестный параметр ε. Уравнение (1,21) решается относительно ε особенно просто в двух крайних случаях: в случае достаточно высоких и в случае достаточно низких температур. Действительно, первый член в левой части равенства (1,21) возрастает по мере возрастания температуры, поэтому в области достаточно высоких температур мы имеем:

$$n_1(T) \gg N_2, \tag{1,22}$$

и в области достаточно низких температур:

$$n_1(T) \ll N_2. \tag{1,23}$$

Рассмотрим оба эти случая.

а) В первом случае уравнение (1,21) принимает такой вид:

$$n_1(T) = n_2(T),$$

или

$$A_1 e^{-\frac{W_1-\varepsilon}{kT}} = (N_1 + N_2)\, e^{-\frac{\varepsilon - W_0}{kT}}. \tag{1,24}$$

Согласно (1,22) и (1,19), мы имеем:

$$A_1 e^{-\frac{W_1-\varepsilon}{kT}} \gg N_2 \gg N_2\, e^{-\frac{\varepsilon-W_0}{kT}},$$

и, следовательно, выражение (1,24) можно переписать так:

$$A_1 e^{-\frac{W_1-\varepsilon}{kT}} = N_1 e^{-\frac{\varepsilon-W_0}{kT}}.$$

Отсюда

$$\boxed{\begin{gathered}\varepsilon = W_1 - \frac{1}{2}u_1 - \frac{1}{2}kT\ln\frac{A_1}{N_1}, \\ \sigma = ev_1 n_1(T) = ev_1\sqrt{A_1 N_1}\, e^{-\frac{u_1}{2kT}}.\end{gathered}} \tag{1,25}$$

Мы получаем формулы Вильсоновской теории [ср. (1,25) с (1,16)].

Во втором случае уравнение (1,21) принимает вид:

$$N_2 = n_2(T),$$

или

$$N_2 = (N_1 + N_2)\, e^{-\frac{\varepsilon-W_0}{kT}}.$$

Согласно (1,19), это выражение можно переписать так:

$$N_2 = N_1 e^{-\frac{\varepsilon-W_0}{kT}}.$$

Отсюда

$$\boxed{\begin{gathered}\varepsilon = W_0 + kT\ln\frac{N_1}{N_2}, \\ \sigma = ev_1 n_1(T) = ev_1 A_1 \frac{N_1}{N_2}\, e^{-\frac{u_1}{kT}}.\end{gathered}} \tag{1,26}$$

Исходные условия (1,19) могут быть расшифрованы, если в первое из них подставить (1,25), а во второе (1,26). Мы получим:

$$A_1 e^{-\frac{u_1}{kT}} \gg N_1 \gg N_2.$$

Этими условиями определяются пределы применимости теории.

Зависимость ε от T изображена на рис. 64, а зависимость σ от T — на рис. 65 (для разных N_1 при заданном N_2; кривые перенумерованы в порядке возрастания N_1). В области, заштрихованной на рис. 64, исходные условия (1,19) нарушаются, т. е. теория перестаёт быть применимой.

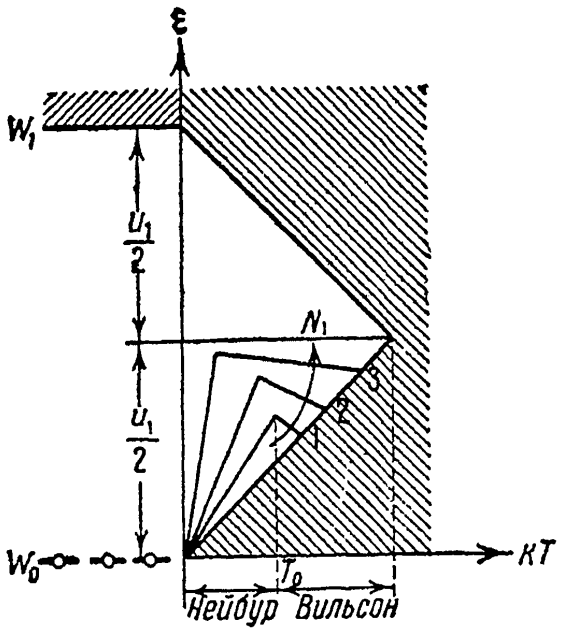


Рис. 64. Положение в энергетическом спектре уровня химического потенциала ε как функция T (при разных N_1 и заданном N_2).

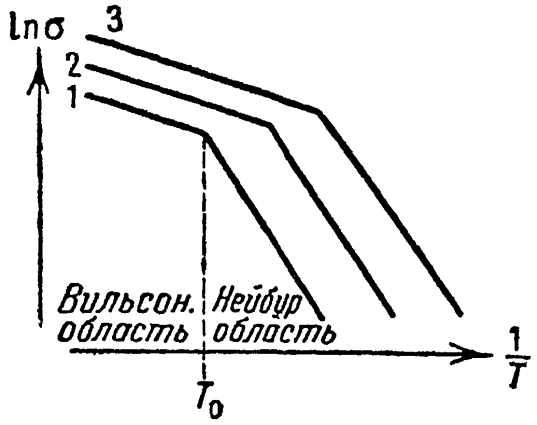


Рис. 65. Зависимость электропроводности от температуры при разных концентрациях доноров (и заданной концентрации акцепторов).

Температура T_0, соответствующая точке излома на кривых рис. 64 и рис. 65, определяется из уравнения:

$$n_1(T_0) = N_2.$$

При $T \gg T_0$ полупроводник ведёт себя как вильсоновский полупроводник. Эту область температур мы можем назвать *вильсоновской областью*. Область $T \ll T_0$ мы можем назвать *нейбуровской областью*. При переходе из вильсоновской области в нейбуровскую наклон кривых на рис. 65 увеличивается вдвое. Такой излом на температурной кривой электропроводности

с изменением наклона в два раза был обнаружен Яндером в кристаллах ZnO.

В главе VI мы вернёмся вновь к полупроводникам рассмотренного здесь типа. Мы увидим, что такие полупроводники обладают специфическими фотоэлектрическими свойствами.

Заметим в заключение, что, наряду с вильсоновской и нейбуровской схемами (рис. 60 и 63), можно было бы рассмотреть и более сложные энергетические схемы, характеризующиеся тем или иным размещением доноров и акцепторов в энергетическом спектре. Рассмотрение таких схем не встречает принципиальных затруднений. Однако такое рассмотрение только тогда представляет интерес, когда выбор той или иной схемы физически достаточно мотивирован.

D. Полупроводники с металлическим ходом электропроводности. Среди различных типов полупроводников особое место занимают полупроводники с металлическим ходом электропроводности (см. рис. 57, *а*). Такие полупроводники иногда называют полуметаллами. Сюда относятся CdO, Bi_2S_3, SiC (при достаточно низких температурах) и др. Обычно почти все полупроводники превращаются в полуметаллы при достаточном содержании примеси (иначе говоря: при достаточно сильно нарушенной стехиометрии). Иллюстрацией этого правила могут служить кривые электропроводности для WO_3 и TiO_2, полученные Мейером (см. рис. 59).

Схемы Вильсона и Нейбура непригодны для объяснения температурных кривых в полупроводниках этого типа. Все попытки теоретической интерпретации этих кривых, делавшиеся до сих пор, приходится признать неудачными.

Одна из таких попыток принадлежит И. Курчатову, Костиной и Русинову [5]. С точки зрения этих авторов, примесные атомы, при достаточной их концентрации, порождают в энергетическом спектре кристалла свою собственную «примесную» зону. Эта «примесная» зона помещается в запрещённом участке под зоной проводимости, на месте обычных локальных уровней вильсоновской теории. Если, например, примесные атомы одновалентны, то такая «примесная» зона заполнена электронами наполовину. Электропроводность кристалла (при не слишком высоких температурах) осуществляется этими электронами в рамках «примесной» зоны. Число электронов, равное числу чужеродных атомов, не зависит от температуры. Эти электроны

ведут себя внутри примесной зоны как обычные металлические электроны в зоне проводимости.

Эти представления, однако, явно неудовлетворительны. Действительно, даже при максимальной возможной концентрации примесных атомов расстояние между ними порядка нескольких атомных расстояний. Таким образом, индивидуальные волновые функции отдельных примесных атомов практически не перекрываются. Это значит, что ширина «примесной» зоны сводится к нулю, т. е. зона вырождается в дискретный уровень. Электрон в такой бесконечно узкой зоне обладает бесконечно большой эффективной массой и, следовательно, бесконечно малой подвижностью. Тем самым электропроводность сводится к нулю. Таким образом, различие между «примесной» зоной Курчатова и локальными уровнями Вильсона скорее терминологическое, нежели физическое.

Рассмотрим другую, широко распространённую точку зрения, объясняющую электропроводность полуметаллов в духе вильсоновской теории (см., например, статью Давыдова [18]). Представим себе, что уровни-доноры на рис. 60, *а* расположены не в запрещённом участке, а внутри зоны проводимости или над ней. Или же представим себе, что уровни-акцепторы на рис. 60, *б* расположены внутри нормальной зоны или под этой зоной. В этом случае зона проводимости будет содержать в себе определённое число электронов, равное числу доноров и, следовательно, не зависящее от температуры. Или же нормальная зона будет содержать в себе число дырок, равное числу акцепторов и также, следовательно, не зависящее от температуры.

Эту точку зрения также нельзя признать удовлетворительной. В действительности уровни-доноры обязаны лежать *под* зоной проводимости, а уровни-акцепторы обязаны лежать *над* нормальной зоной. Такое заключение вытекает из совершенно общих соображений. Действительно, дефекты, порождающие уровни-доноры, являются электроположительными дефектами, т. е. дефектами, притягивающими к себе электроны и образующими, в совокупности с сидящими на них электронами, электрически нейтральные образования Уровни-акцепторы, наоборот, соответствуют электроотрицательным дефектам. Очевидно, отрыв электрона от электроположительного дефекта и перевод его в коллективизированное состояние должен сопровождаться всегда затратой энергии и не может сопровождаться выделе-

нием энергии. Это и значит, в переводе на язык энергетической схемы, что уровни-доноры должны помещаться в энергетическом спектре ниже, чем зона проводимости. Из аналогичных соображений вытекает, что уровни-акцепторы должны располагаться выше, чем нормальная зона.

Таким образом, приходится констатировать, что металлический ход электропроводности, наблюдающийся в некоторых полупроводниках, не получил до сих пор удовлетворительной теоретической интерпретации.

§ 2. Общая теория электропроводности реальных гетерополярных решёток

А. Энергетическая схема. Реальная гетерополярная решётка, в отличие от идеальной решётки, содержит в себе дефекты, вообще говоря, двух типов. Эти дефекты, вследствие той роли, которую они играют в решётке, могут быть названы *электроположительными* и *электроотрицательными* дефектами. По своей физической природе электроположительные дефекты представляют собой пустые металлоидные узлы или междуузельные металлические ионы. К электроотрицательным дефектам относятся пустые металлические узлы или междуузельные металлоидные ионы. Эта классификация дефектов иллюстрируется схемой, изображённой на рис. 66.

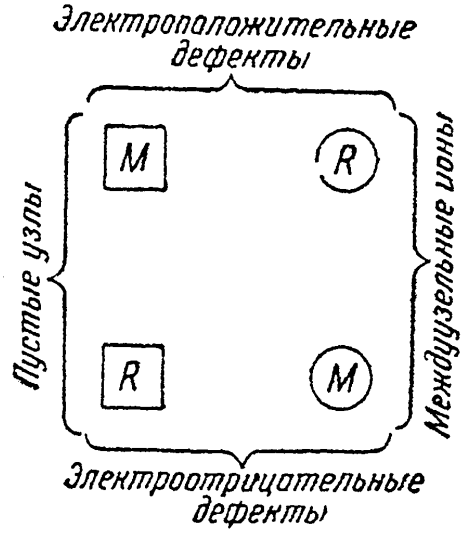


Рис. 66. Классификация дефектов по их роли в решётке и по их физической природе.

Мы будем считать, что решётка, в зависимости от её природы и в зависимости от той обработки, которой она была подвергнута, содержит в качестве дефектов или только *пустые узлы* (обоих сортов), или только *междуузельные ионы* (обоих сортов). Кроме того, мы будем считать, что число дефектов, присутствующих в решётке, определяется технологией образца и практически не меняется с температурой. Мы видели (глава I, § 1, С), что во всех практически интересных для нас случаях такое предположение допустимо.

Сумма электроположительных и электроотрицательных дефектов определяет собой степень отклонения нашей реальной решётки от идеальной. *Разность* между числом электроположительных и числом электроотрицательных дефектов характеризует степень отклонения от стехиометрического соотношения. Таким образом, всякая реальная решётка для своего описания требует введения двух независимых между собой параметров.

Электроположительные дефекты в решётке играют роль ловушек для электронов, а электроотрицательные дефекты, являясь центрами отталкивания для электронов, тем самым служат ловушками для дырок.

В решётке, содержащей в себе электроположительные и электроотрицательные дефекты, электрическая нейтральность оказывается, вообще говоря, нарушенной. Для восстановления нейтральности в такую решётку должны быть введены дополнительно электроны и дырки в определённой пропорции, так, чтобы их суммарный заряд компенсировал заряд, созданный дефектами. Введение в решётку пары электрон + дырка может быть для решётки энергетически А) выгодным, В) невыгодным. В соответствии с этим, следует различать два типа решёток, которые мы назовём условно решётками «типа А» и решётками «типа В» (см. главу I, § 5, В).

А. Для решёток «типа А» дополнительные электроны и дырки выгодно ввести в таком количестве, чтобы, разместив их соответственно вокруг электроположительных и электроотрицательных дефектов, можно было бы нейтрализовать каждый из этих дефектов в отдельности. В таких решётках каждый дефект, в совокупности с локализованными около него электронами или дырками, представляет собой электрически нейтральное образование. Энергетические схемы для такой решётки изображены на рис. 67 (см. также рис. 16). Мы имеем две системы локальных уровней, что свидетельствует о наличии дефектов двух сортов. Уровни, соответствующие электроположительным дефектам, являются донорами. Действительно, каждый такой электроположительный дефект является обладателем сконцентрированных вокруг него и нейтрализующих его электронов. Уровни, соответствующие электроотрицательным дефектам, являются акцепторами, поскольку электроотрицательные дефекты могут рассматриваться как обладатели дырок.

В. Для решёток «типа В» одновременное введение электронов и дырок является невыгодным, т. е. электроны и дырки в таких решётках рекомбинируют друг с другом. В таких решётках, при соблюдённом стехиометрическом соотношении, электроположительные и электроотрицательные дефекты вовсе лишены нейтрализующих их электронных и дырочных оболочек.

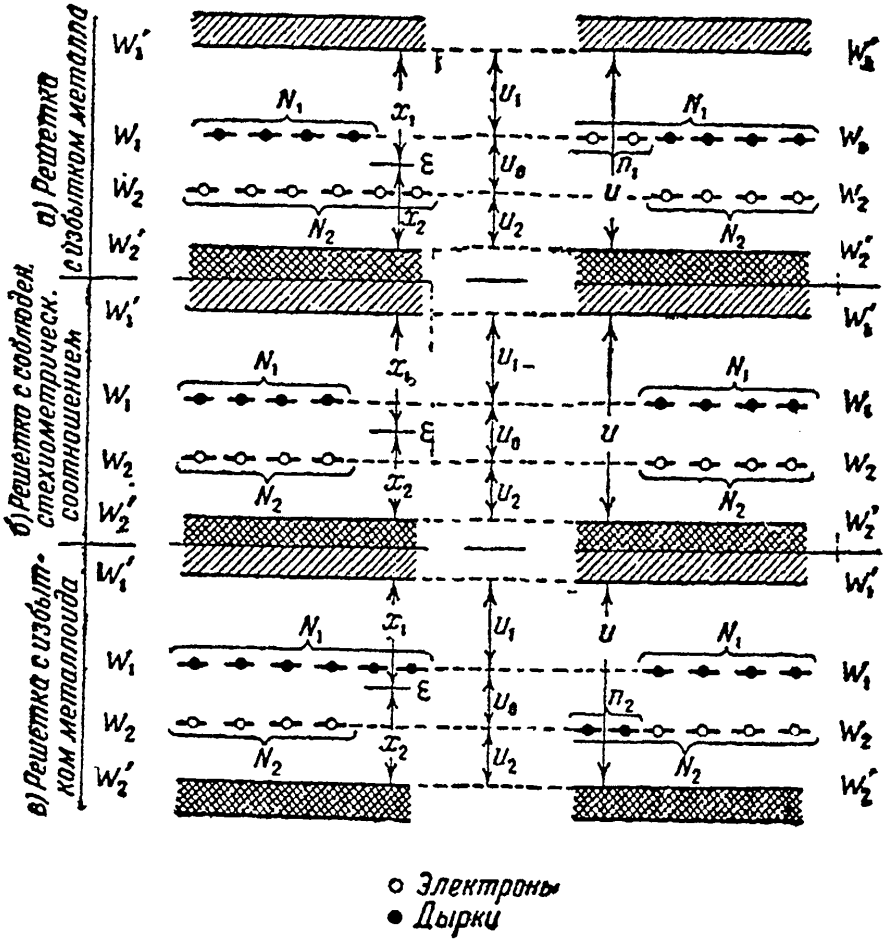


Рис. 67. Схема уровней для решёток типа *A*.

Рис. 68. Схема уровней для решёток типа *B*.

Энергетическая схема, соответствующая этому случаю, изображена на рис. 68, *б*. На этот раз электроположительные дефекты фигурируют в энергетическом спектре в качестве акцепторов, поскольку каждый такой электроположительный дефект является ловушкой для электрона, а электроотрицательные дефекты служат донорами, поскольку они могут рассматриваться как ловушки для дырок. В случае стехиометрического избытка металла в решётку необходимо ввести некоторое количество компенсирующих электронов, которые рас-

саживаются по электроположительным дефектам (по одному на каждый дефект), превращая соответствующие уровни из акцепторов в доноры, как это изображено на рис. 68, *а*. Стехиометрический избыток металлоида связан, наоборот, с появлением компенсирующих дырок и превращением некоторого числа электроотрицательных уровней из доноров в акцепторы. К этому последнему случаю относится рис. 68, *б*.

На рисунках 67 и 68 уровни-акцепторы изображены тёмными, а уровни-доноры — светлыми кружками. На рис. 67 электроположительные уровни расположены над электроотрицательными, а на рис. 68, наоборот, под электроотрицательными. К этому и сводится, с точки зрения энергетического спектра, различие между решётками «типа А» и «типа В». Если на рис. 67 электроположительные и электроотрицательные уровни поменять местами, то электроны с верхних уровней рекомбинируют с дырками нижних уровней, и рис. 67 превратится в рис. 68. Заметим, что схема, изображённая на рис. 67, превращается в вильсоновскую схему в том частном случае, когда $N_1 = 0$ (см. рис. 60, *а*) или $N_2 = 0$ (см. рис. 60, *б*), или в нейбуровскую схему в том случае, когда $W_1 = W_2$ (см. рис. 63).

Схемы, представленные на рис. 67 и 68, однозначно вытекают из наших физических представлений о реальных гетерополярных решётках. Они являются отражением принятой нами модели [1]). Этими схемами, в свою очередь, предопределяется температурный ход электропроводности. Мы увидим, что, варьируя соотношение между числом электроположительных и электроотрицательных дефектов, мы можем передвигать теоретическую кривую по шкале температур, в результате чего в поле зрения различных экспериментаторов будут попадать различные участки этой теоретической кривой, и мы получим в итоге всё разнообразие экспериментальных кривых, изображённых на рис. 59.

В. Вывод общей формулы для электропроводности. Рассмотрим условия теплового равновесия в полупроводниках, характеризующихся энергетическими схемами, изображёнными на рис. 67 и 68. Обозначения на этих рисунках выбраны так, чтобы все дальнейшие формулы могли быть отнесены к любому из случаев, представленных на рис. 67 и 68.

---

[1]) О пределах применимости этих схем см. главу I, § 5,А.

Пусть мы имеем на единицу объёма кристалла при температуре T:

$$\left\{\begin{array}{ll} n_1'(T) & \text{— число электронов в зоне проводимости,} \\ n_1(T) & \text{— число электронов на локальных уровнях } W_1, \\ n_2(T) & \text{— число дырок на локальных уровнях } W_2, \\ n_2'(T) & \text{— число дырок в нормальной зоне,} \end{array}\right.$$

причём

$$\left\{\begin{array}{l} n_1'(0)=0, \\ n_1(0)=n_1, \\ n_2(0)=n_2, \\ n_2'(0)=0. \end{array}\right.$$

Относительно чисел n_1 и n_2 мы условимся считать:

$$\left\{\begin{array}{ll} \text{для решёток «типа A» (рис. 67):} & n_1=n_2=0 \\ \text{для решёток «типа B»:} & \\ \quad \text{при стехиом. избытке металла:} & n_2=0 \\ \quad \text{при соблюдённом стехиом. соотн.:} & n_1=n_2=0 \\ \quad \text{при стехиом. избытке металлоида:} & n_1=0 \end{array}\right\} \quad (2,1)$$

Предположим, что уровень химического потенциала располагается между электроположительными и электроотрицательными уровнями, т. е.:

$$W_2 < \varepsilon < W_1,$$

причём соблюдаются условия:

$$\left.\begin{array}{l} e^{\frac{W_1-\varepsilon}{kT}}=e^{\frac{x_1-u_1}{kT}}\gg 1, \\ e^{\frac{\varepsilon-W_2}{kT}}=e^{\frac{x_2-u_2}{kT}}\gg 1, \end{array}\right\} \quad (2,2)$$

из которых, в свою очередь, вытекают условия:

$$\left.\begin{array}{l} e^{\frac{W_1'-\varepsilon}{kT}}=e^{\frac{x_1}{kT}}\gg 1, \\ e^{\frac{\varepsilon-W_2'}{kT}}=e^{\frac{x_2}{kT}}\gg 1; \end{array}\right\} \quad (2,3)$$

здесь (см. рис. 67 и 68):

$$\left.\begin{array}{l} x_1=W_1'-\varepsilon, \\ x_2=\varepsilon-W_2', \end{array}\right\}$$

так что

$$u = x_1 + x_2.$$

Условиями (2,3) ограничивается область применения теории. Физический смысл этих условий будет раскрыт в дальнейшем.

Принимая условия (2,2) и (2,3), мы будем иметь:

$$\left.\begin{aligned} n_1'(T) &= A_1 e^{-\frac{W_1 - \varepsilon}{kT}} = A_1 e^{-\frac{x_1}{kT}}, \\ n_1(T) &= N_1 e^{-\frac{W_1 - \varepsilon}{kT}} = N_1 e^{-\frac{x_1 - u_1}{kT}}, \\ n_2(T) &= N_2 e^{-\frac{\varepsilon - W_2}{kT}} = N_2 e^{-\frac{x_2 - u_2}{kT}}, \\ n_2'(T) &= A_2 e^{-\frac{\varepsilon - W_2'}{kT}} = A_2 e^{-\frac{x_2}{kT}}. \end{aligned}\right\} \qquad (2,4)$$

Неизвестный параметр x_1 (или x_2) может быть определён из закона сохранения заряда:

$$n_1'(T) + n_1(T) - n_1 = n_2'(T) + n_2(T) - n_2. \qquad (2,5)$$

Подставляя сюда (2,4) и решая это уравнение относительно x_1 (или x_2), будем иметь:

$$\left.\begin{aligned} e^{-\frac{x_1}{kT}} &= \frac{(n_1 - n_2)}{2\left(A_1 + N_1 e^{\frac{u_1}{kT}}\right)} + \\ &+ \frac{\sqrt{(n_1 - n_2)^2 + 4\left(A_1 + N_1 e^{\frac{u_1}{kT}}\right)\left(A_2 + N_2 e^{\frac{u_2}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(A_1 + N_1 e^{\frac{u_1}{kT}}\right)}, \\ e^{-\frac{x_2}{kT}} &= \frac{(n_2 - n_2)}{2\left(A_2 + N_2 e^{\frac{u_2}{kT}}\right)} + \\ &+ \frac{\sqrt{(n_2 - n_1)^2 + 4\left(A_2 + n_2 e^{\frac{u_2}{kT}}\right)\left(A_1 + N_1 e^{\frac{u_1}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(A_2 + N_2 e^{\frac{u_2}{kT}}\right)}. \end{aligned}\right\} \qquad (2,6)$$

Этими выражениями определяется положение уровня ε в спектре в зависимости от температуры T и параметров n_1, n_2, N_1, N_2, характеризующих решётку.

Обратимся теперь к электропроводности. Электропроводность полупроводника σ складывается, вообще говоря, из двух составляющих (электронной и дырочной):

$$\sigma(T) = ev_1 n_1'(T) + ev_2 n_2'(T) = ev_1 A_1 e^{-\frac{x_1}{\kappa T}} + ev_2 A_2 e^{-\frac{x_2}{\kappa T}}. \quad (2,7)$$

Подставляя (2,6) в (2,7) и полагая

$$A_1 \approx A_2 \approx A,$$
$$v_1 \approx v_2 \approx v,$$

будем иметь:

$$\frac{\sigma}{evA} = \quad (2,8)$$

$$= \frac{n_1 - n_2 + \sqrt{(n_1 - n_2)^2 + 4\left(A + N_1 e^{\frac{u_1}{\kappa T}}\right)\left(A + N_2 e^{\frac{u_2}{\kappa T}}\right)e^{-\frac{u}{\kappa T}}}}{2\left(A_1 + N_1 e^{\frac{u_1}{\kappa T}}\right)} +$$

$$+ \frac{n_2 - n_1 + \sqrt{(n_2 - n_1)^2 + 4\left(A + N_2 e^{\frac{u_2}{\kappa T}}\right)\left(A + N_1 e^{\frac{u_1}{\kappa T}}\right)e^{-\frac{u}{\kappa T}}}}{2\left(A + N_2 e^{\frac{u_2}{\kappa T}}\right)}.$$

Формула (2,8) является общей формулой для решёток всех рассматриваемых типов. Полагая в ней, согласно (2,1), $n_1 = 0$, или $n_2 = 0$, или $n_1 = n_2 = 0$, мы получаем все случаи, изображённые на рис. 67 и 68.

Формулы (2,6) и (2,8) можно представить в ином виде, заменив параметры n_1, n_2, N_1, N_2 параметрами μ и ν, характеризующими, соответственно, степень отклонения нашей реальной решётки от идеального состояния и степень нарушения стехиометрического соотношения (см. главу I, § 1 В). Ограничиваясь случаем, когда

$$|\nu| \ll \mu,$$

мы имеем, согласно формулам (5,1) и (1,16) главы I:

$$\left.\begin{aligned} N_1 &= \alpha_1 \mu, \\ N_2 &= \alpha_2 \mu, \\ n_1 - n_2 &= \beta\nu, \end{aligned}\right\} \quad (2,9)$$

где

$$\alpha_{1,2} = \frac{1}{2} mZ \quad \text{или} \quad \alpha_{1,2} = \frac{1}{2} rZ; \quad \beta = pmZ = qrZ.$$

Полагая, кроме того:

$$\left.\begin{aligned} \frac{A_1}{\alpha_1} &\approx \frac{A_2}{\alpha_2} \approx a, \\ v_1 &\approx v_2 \approx v, \\ \frac{\beta}{\sqrt{\alpha_1\alpha_2}} &\approx 1, \\ \frac{\alpha_1}{\alpha_2} &\approx 1, \\ ev\sqrt{\alpha_1\alpha_2} = \frac{1}{2}evZ\sqrt{mr} &= b, \end{aligned}\right\} \qquad (2,10)$$

мы получим, подставляя (2,9) в (2,6) и в (2,8) и учитывая (2,10):

$$\boxed{\begin{aligned} e^{-\frac{x_1}{kT}} &= \frac{\nu + \sqrt{\nu^2 + 4\left(a+\mu e^{\frac{u_1}{kT}}\right)\left(a+\mu e^{\frac{u_2}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(a+\mu e^{\frac{u_1}{kT}}\right)}, \\ e^{-\frac{x_2}{kT}} &= \frac{-\nu + \sqrt{\nu^2 + 4\left(a+\mu e^{\frac{u_2}{kT}}\right)\left(a+\mu e^{\frac{u_1}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(a+\mu e^{\frac{u_2}{kT}}\right)} \end{aligned}} \qquad (2,11)$$

и

$$\boxed{\sigma(T) = ab\left\{\frac{\nu + \sqrt{\nu^2 + 4\left(a+\mu e^{\frac{u_1}{kT}}\right)\left(a+\mu e^{\frac{u_2}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(a+\mu e^{\frac{u_1}{kT}}\right)} + \right.} \qquad (2,12)$$

$$\boxed{\left. + \frac{-\nu + \sqrt{\nu^2 + 4\left(a+\mu e^{\frac{u_2}{kT}}\right)\left(a+\mu e^{\frac{u_1}{kT}}\right)e^{-\frac{u}{kT}}}}{2\left(a+\mu e^{\frac{u_2}{kT}}\right)}\right\}.}$$

Общая формула (2,12), как мы в этом убедимся ниже, охватывает всё разнообразие экспериментальных кривых. Все типы кривых (см. рис. 59) могут быть извлечены из этой формулы путём варьирования параметров μ и ν. Формулы (2,11) и (2,12) относятся к полупроводникам «типа B». В случае полупроводников «типа A» в этих формулах, согласно

(2,1) и (2,9), следует положить $\nu = 0$. Таким образом, для полупроводников этого типа электропроводность практически зависит только от μ, но не зависит от ν до тех пор, пока сохраняется условие $|\nu| \ll \mu$.

Заметим, что параметры a и b в формулах (2,11) и (2,12) будем считать в дальнейшем практически не зависящими от температуры. Действительно, как это видно из (1,7а) и (1,14,а), параметры a и b меняются с температурой очень медленно по сравнению с экспоненциальными членами, присутствующими в этих формулах. Поэтому, полагая в этих формулах $a = \text{const}$ и $b = \text{const}$, мы совершаем незначительную ошибку. По порядку величины имеем

$$\begin{cases} a \approx 10^{-1} - 10^{-3}, \\ b \approx 10^{3} - 10^{6}\ ом^{-1} \cdot см^{-1}, \end{cases}$$

так что можно считать:

$$|\nu| \ll \mu \ll a. \qquad (2,13)$$

С. Применимость предыдущей формулы при учёте тепловой диссоциации. Формула (2,8) была получена в предположении, что решётка содержит два сорта дефектов (металлические и металлоидные пустые узлы *или* металлические и металлоидные междуузельные ионы), причём концентрация дефектов не меняется с температурой. Такое предположение правомерно лишь до тех пор, пока «беспорядок», вносимый в решётку температурой, мал по сравнению с «беспорядком», данным ей, так сказать, от рождения, т. е. лишь до тех пор, пока можно пренебрегать диссоциацией ионов из узлов в междуузелья. Мы видели (глава I, § 1 С), что для полупроводников, во всей практически интересной для нас области температур, число тепловых повреждений действительно ничтожно мало по сравнению с числом повреждений, возникших в процессе приготовления образца и присутствующих в решётке, так сказать, «от рождения». В щёлочно-галоидных кристаллах, которые обладают более идеальными решётками, мы встречаемся, повидимому, с противоположным случаем. Здесь число «врождённых» дефектов обычно сравнимо с числом дефектов, обусловленных температурой. В этом случае учёт диссоциации ионов является необходимым.

Формула (2,8) может быть распространена на те случаи, когда концентрация дефектов является функцией температуры,

если только предполагать, что мы попрежнему имеем дело с двумя (и только с двумя) сортами дефектов. При этом безразлично, какие предположения мы делаем о физической природе этих дефектов. От тех или иных предположений изменяется лишь смысл параметров N_1 и N_2, фигурирующих в формуле (2,8).

Рассмотрим гетерополярную решётку M_mR_r, которая содержит в качестве дефектов пустые узлы одного определённого сорта и соответствующие междуузельные ионы. Для определённости пусть это будут металлические пустые узлы и металлические междуузельные ионы. Первые играют роль электроотрицательных, а вторые — электроположительных дефектов. Будем учитывать тепловую диссоциацию металлических ионов из узлов в междуузелья. Предположим, что при температуре T мы имеем на единицу объёма кристалла:

$$\begin{cases} X_{\text{м}}(T) — \text{число пустых металлических узлов,} \\ Y_{\text{м}}(T) — \text{число междуузельных металлических ионов.} \end{cases}$$

Заметим, что числа $X_{\text{м}}$ и $Y_{\text{м}}$, вообще говоря, не обязаны быть одинаковыми, если считать, что стехиометрическое соотношение в нашей решётке может быть нарушено. Мы имеем [см. формулы (1,10) главы I]:

$$\left.\begin{array}{l} \text{а) при стехиометрическом избытке металла:} \\ X_{\text{м}}(T) = Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}, \quad Y_{\text{м}}(T) = Y^0_{\text{м}} + Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}; \\ \text{б) при соблюдённом стехиометрическом соотношении:} \\ X_{\text{м}}(T) = Y_{\text{м}}(T) = Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}; \\ \text{в) при стехиометрическом недостатке металла:} \\ X_{\text{м}}(T) = X^0_{\text{м}} + Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}, \quad Y_{\text{м}}(T) = Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}, \end{array}\right\} \quad (2,14)$$

где

$$\begin{cases} U_{\text{м}} — \text{энергия диссоциации металлического иона,} \\ X^0_{\text{м}} — \text{число «врождённых» пустых узлов,} \\ Y^0_{\text{м}} — \text{число «врождённых» междуузельных ионов,} \\ Z_{\text{м}} — \text{число металлических узлов в решётке.} \end{cases}$$

Числа $X_{\text{м}}^0$ и $Y_{\text{м}}^0$ обусловливаются технологией образца и не зависят от температуры.

Наша решётка может принадлежать к «типу А» или к «типу В». Принадлежность решётки к «типу А» означает в нашем случае, что междуузельные ионы фигурируют в решётке не в виде ионов, а в виде нейтральных атомов, т. е. нейтрализованы компенсирующими электронами (в этом случае пустые узлы также будут нейтрализованы «прилипшими» к ним дырками).

Схемы, изображённые на рис. 67 и 68, вполне применимы в нашем случае. Однако теперь следует считать, что число локальных уровней является функцией температуры. Вместе с тем сохраняется формула (2,8). Если наша решётка принадлежит к «типу А», то в формуле (2,8) следует теперь положить:

$$\begin{cases} N_1 = X_{\text{м}}(T), \\ N_2 = Y_{\text{м}}(T), \\ n_1 - n_2 = 0. \end{cases}$$

Если же наша решётка принадлежит к «типу В», то, сохраняя соответствие с обозначениями, принятыми на рис. 68, следует положить в (2,8):

$$\begin{cases} N_1 = Y_{\text{м}}(T), \\ N_2 = X_{\text{м}}(T), \\ n_1 - n_2 = pY_{\text{м}}(T) - pX_{\text{м}}(T), \end{cases}$$

где $X_{\text{м}}(T)$ и $Y_{\text{м}}(T)$ имеют вид (2,14); p — заряд металлического иона решётки (в единицах абсолютного заряда электрона).

При достаточно высоких температурах и достаточно слабых отклонениях от стехиометрии, когда число тепловых дефектов велико по сравнению с числом «врождённых», мы можем положить в (2,14):

$$X_{\text{м}}^0 = Y_{\text{м}}^0 = 0,$$

а в формуле (2,8) следует положить:

$$\begin{cases} N_1 = N_2 = Z_{\text{м}} e^{-\frac{U_{\text{м}}}{kT}}, \\ n_1 = n_2 = 0. \end{cases}$$

Рис. 67 в этом случае совпадает с рис. 68.

D. Области апроксимации. Обратимся к анализу формулы (2,12).

Вся область температур, для которой справедлива формула (2,12), может быть разбита на отдельные температурные участки, внутри которых формула (2,12) вырождается в простые экспоненциальные зависимости типа

$$\sigma = Ce^{-\frac{B}{T}}. \tag{2,15}$$

Коэффициенты C и B остаются постоянными внутри каждого данного участка, но меняются при переходе от одного участка к другому. Границы между участками фиксируются параметрами μ и ν, фигурирующими в формуле (2,22). Иначе говоря, от параметров μ и ν зависит положение на температурной шкале тех точек, при переходе через которые меняются значения коэффициентов C и B в (2,15).

Чтобы убедиться в этом, прежде всего введём в рассмотрение три точки на температурной шкале, которые мы обозначим через T_0, T_1, T_2 и определим из следующих уравнений:

$$\left.\begin{aligned} \left(a+\mu e^{\frac{u_1}{kT_0}}\right)\left(a+\mu e^{\frac{u_2}{kT_0}}\right)e^{-\frac{u}{kT_0}} &= \nu^2, \\ \mu e^{\frac{u_i}{kT_1}} &= a, \\ \mu e^{\frac{u_k}{kT_2}} &= a, \end{aligned}\right\} \tag{2,16}$$

откуда

$$\left.\begin{aligned} T_0 &= f_0(\mu, \nu), \\ T_1 &= f_1(\mu) = \frac{u_i}{k \ln \frac{a}{\mu}}, \\ T_2 &= f_2(\mu) = \frac{u_k}{k \ln \frac{a}{\mu}}. \end{aligned}\right\} \tag{2,17}$$

Здесь значения индексов i и k мы выберем так, чтобы мы имели заведомо $T_1 \geqslant T_2$, т. е. $u_i \geqslant u_k$:

$$\left.\begin{aligned} i=1, \quad k=2 \ldots\ldots, \quad &\text{если } u_1 \geqslant u_2, \\ i=2, \quad k=1 \ldots\ldots, \quad &\text{если } u_2 \geqslant u_1. \end{aligned}\right\} \tag{2,18}$$

Точки T_1 и T_2 делят всю шкалу температур на три области

$T > T_1$	$T_1 > T > T_2$	$T_2 > T$
в этой области: $\mu e^{\frac{u_i}{kT}} < a$, $\mu e^{\frac{u_k}{kT}} < a$.	в этой области: $\mu e^{\frac{u_i}{kT}} > a$, $\mu e^{\frac{u_k}{kT}} < a$.	в этой области: $\mu e^{\frac{u_i}{kT}} > a$, $\mu e^{\frac{u_k}{kT}} > a$.

(2,19)

С другой стороны, точка T_0 делит всю шкалу температур на две области:

$T > T_0$	$T_0 > T$
в этой области: $\left(a+\mu e^{\frac{u_1}{kT}}\right)\left(a+\mu e^{\frac{u_2}{kT}}\right) e^{-\frac{u}{kT}} > \nu^2$.	в этой области: $\left(a+\mu e^{\frac{u_1}{kT}}\right)\left(a+\mu e^{\frac{u_2}{kT}}\right) e^{-\frac{u}{kT}} < \nu^2$.

(2,20)

Положение точек T_1 и T_2 фиксируется числом μ. Уменьшая μ, мы сдвигаем точки T_1 и T_2 в сторону низких температур, так что при $\mu \to 0$ мы имеем $T_1 \to 0$ и $T_2 \to 0$.

Положение точки T_0 при заданном μ фиксируется значением числа ν. При уменьшении ν точка T_0 сдвигается в сторону низких температур, так что при $\nu \to 0$ мы имеем $T_0 \to 0$. На рис. 69 изображена T_0 как функция от ν (жирная ломаная). По осям отложены $\frac{1}{T}$ и $\ln|\nu|$. Положение T_0 относительно T_1 и T_2, как мы видим из рисунка, определяется значением $|\nu|$. Мы имеем:

I) $T_0 \gg T_1$ . . . . . при $|\nu| \gg \nu_1$,
II) $T_1 \gg T_0 \gg T_2$ . . . . . при $\nu_1 \gg |\nu| \gg \nu_2$,
III) $T_2 \gg T_0$ . . . . . при $\nu_2 \gg |\nu|$.

Соответственно, мы будем различать полупроводники:

I) со «значительными» стехиометрическими нарушениями, для которых $|\nu| \gg \nu_1$;

II) с «умеренными» стехиометрическими нарушениями, для которых $\nu_1 \gg |\nu| \gg \nu_2$;

III) со «слабыми» стехиометрическими нарушениями, для которых $\nu_2 \gg |\nu|$.

Для этих трёх случаев первое из уравнений (2,16) может быть легко апроксимировано следующим образом:

$$\nu^2\left(a+\mu e^{\frac{u_1}{kT_0}}\right)\left(a+\mu e^{\frac{u_2}{kT_0}}\right)e^{-\frac{u}{kT_0}}=\begin{cases}=a^2e^{-\frac{u}{kT_0}}=\nu^2\ldots, \\ \qquad \text{если} \quad |\nu|\gg\nu_1, \\ =a\mu e^{-\frac{u-u_i}{kT_0}}=\nu^2\ldots, \\ \qquad \text{если} \quad \nu_1\gg|\nu|\gg\nu_2, \\ =\mu^2e^{-\frac{u_0}{kT_0}}=\nu^2\ldots, \\ \qquad \text{если} \quad \nu_2\gg|\nu|, \end{cases}$$

откуда

$$T_0=f_0(\mu,\nu)=\left\{\begin{array}{ll}=\dfrac{u}{k\ln\frac{a^2}{\nu^2}}\ldots & \text{при } T_0\gg T_1 \\ & \text{т. е. } |\nu|\gg\nu_1, \\ =\dfrac{u-u_l}{k\ln\frac{a\mu}{\nu^2}}\ldots & \text{при } T_1\gg T_0\gg T_2, \\ & \text{т. е. } \nu_1\gg|\nu|\gg\nu_2, \\ =\dfrac{u_0}{k\ln\frac{\mu^2}{\nu^2}}\ldots & \text{при } T_2\gg T_0, \\ & \text{т. е. } \nu_2\gg|\nu|. \end{array}\right\} \quad (2,21)$$

Точки ν_1 и ν_2 определяются, соответственно, из условий:

$$\left.\begin{array}{l} T_0=T_1\ldots\ldots, \text{ откуда } \nu_1=\mu\left(\dfrac{\mu}{a}\right)^{\frac{u-2u_l}{2u_l}}, \\ T_0=T_2\ldots\ldots, \text{ откуда } \nu_2=\mu\left(\dfrac{\mu}{a}\right)^{\frac{u-u_k-u_l}{2u_k}}. \end{array}\right\} \quad (2,22)$$

Наряду с точками T_0, T_1, T_2 введём ещё в рассмотрение точку T_3, которую определим из уравнения:

$$\mu^2e^{-\frac{u-2u_j}{kT_3}}=\nu^2, \quad (2,23)$$

где $j=i$ или $j=k$. Отсюда

$$T_3=f_3(\mu,\nu)=\frac{u-2u_j}{k\ln\frac{\mu^2}{\nu^2}}.$$

Положение точки T_3 при заданном μ фиксируется значением ν. При уменьшении ν точка T_3 сдвигается в сторону низких температур. Функция $T_3=f_3(\nu)$ также изображена на рис. 69.

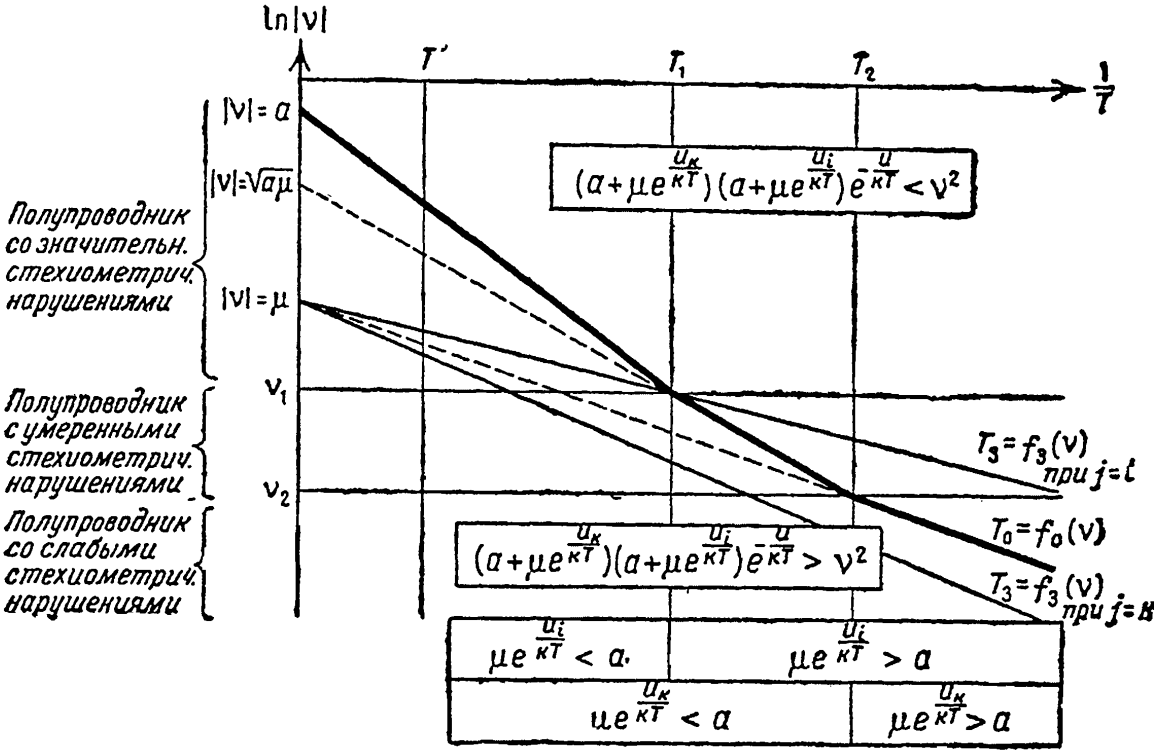


Рис. 69. T_0 и T_3 как функции ν.

Положение точки T_3 относительно T_0, как это видно из рисунка, определяется следующим образом:

$$\text{если } j=i\ldots\begin{cases}T_3>T_0 \ \ldots\ldots\ \text{при}\ |\nu|>\nu_1,\\ T_3=T_0 \ \ldots\ldots\ \text{при}\ |\nu|=\nu_1,\\ T_3<T_0 \ \ldots\ldots\ \text{при}\ |\nu|<\nu_1,\end{cases}$$

$$\text{если } j=k\ldots\quad T_3>T_0 \ \ldots\ldots\ \text{при всех}\ |\nu|.$$

Прямая $T_3=f(\nu)$ вовсе не пересекается с кривой $T_0=f_0(\nu)$, если $j=k$. Если же $j=i$, то прямая $T_3=f_3(\nu)$ пересекается с кривой $T_0=f_0(\nu)$ в точке $T=T_1$, $\nu=\nu_1$.

Теперь вся плоскость на рис. 69 может быть разделена на отдельные области, как это показано на рис. 70, *а* и 70, *б*.

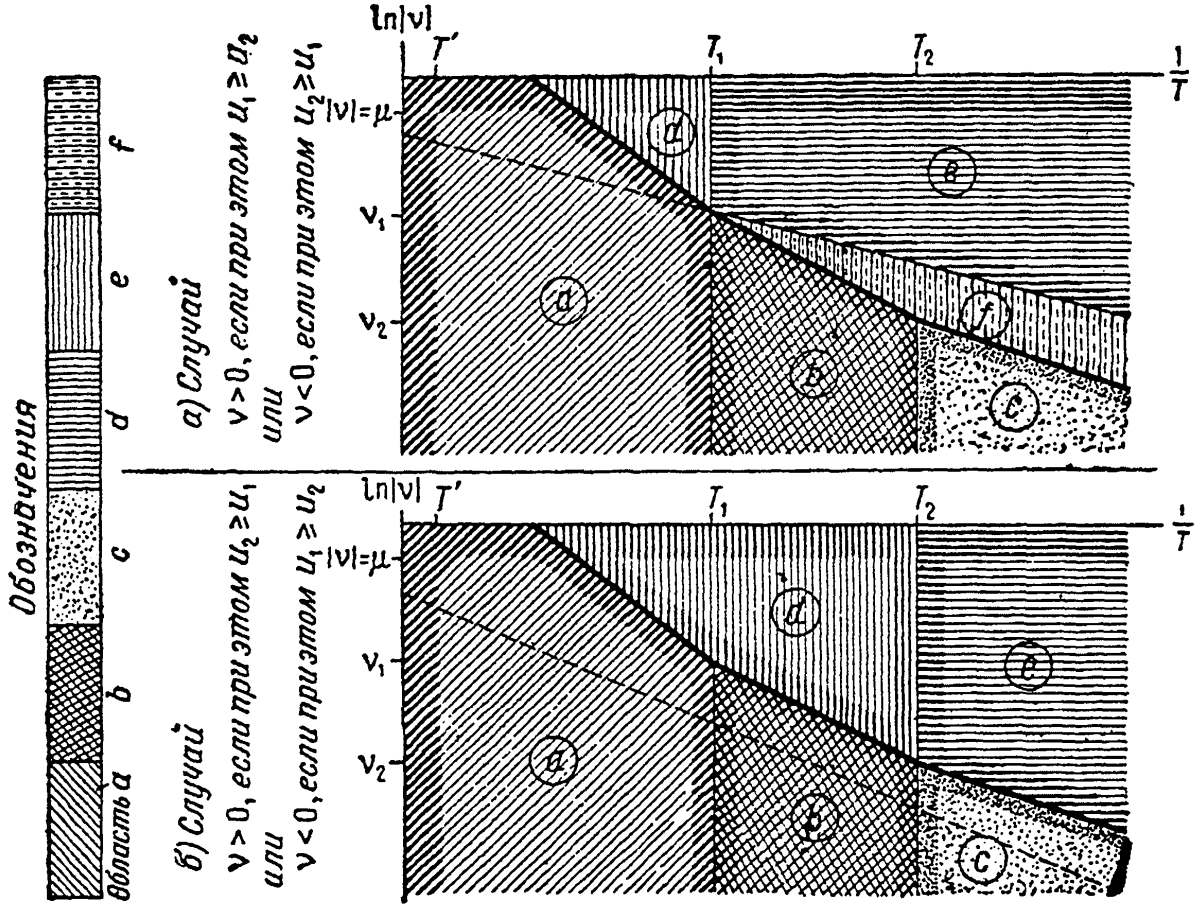


Рис. 70. Области апроксимации функции $\sigma(T)$ [см. формулу (3.21)]

Внутри каждой такой области функция (2,12) апроксимируется простой экспоненциальной зависимостью типа (2,15).

Действительно, при $T \gg T_0$, согласно (2,20), имеем вместо (2,11):

$$\left.\begin{aligned} e^{-\frac{x_1}{kT}} &= e^{-\frac{u}{2kT}} \sqrt{\frac{a+\mu e^{\frac{u_2}{kT}}}{a+\mu e^{\frac{u_1}{kT}}}}, \\ e^{-\frac{x_2}{kT}} &= e^{-\frac{u}{2kT}} \sqrt{\frac{a+\mu e^{\frac{u_1}{kT}}}{a+\mu e^{\frac{u_2}{kT}}}}. \end{aligned}\right\} \qquad (2,23a)$$

Из этих формул получаем для областей a, b, c (см. рис. 70 и 69), согласно (2,19):

a)	$e^{-\frac{x_k}{kT}} = e^{-\frac{u}{2kT}}$	$\sigma_a = 2abe^{-\frac{u}{2kT}}$,
b)	$e^{-\frac{x_k}{kT}} = \sqrt{\frac{\mu}{a}}\, e^{-\frac{u-u_i}{2kT}} = \sqrt{\frac{\mu}{a}}\, e^{-\frac{u_0+u_k}{2kT}}$	$\sigma_b \approx b\sqrt{a\mu}\, e^{-\frac{u_0+u_k}{2kT}}$,
c)	$e^{-\frac{x_k}{kT}} = e^{-\frac{u+u_k-u_i}{2kT}} = e^{-\frac{u_0+2u_k}{2kT}}$	$\sigma_c \approx abe^{-\frac{u_0+2u_k}{2kT}}$,

(2,24)

где $k=1$ или $k=2$ в соответствии с (2,18). В области *a* проводимость имеет смешанную природу: обе компоненты (электронная и дырочная) входят в неё в одинаковой пропорции. В областях *b* и *c* мы имеем электронную проводимость, если $u_2 > u_1$ (т. е. $k=1$), и дырочную проводимость, если $u_1 > u_2$ (т. е. $k=2$). Если же $u_1 = u_2$, то точки T_1 и T_2 сливаются, как это видно из (2,19); в этом случае $\sigma_a = \sigma_b = \sigma_c$.

При $T_0 \gg T$ на основании (2,20) имеем вместо (2,11):

$$\left.\begin{aligned} e^{-\frac{x_1}{kT}} &= \frac{\nu + |\nu| \left\{1 + \frac{2}{\nu^2}\left(a+\mu e^{\frac{u_1}{kT}}\right)\left(a+\mu e^{\frac{u_2}{kT}}\right)e^{-\frac{u}{kT}}\right\}}{2\left(a+\mu e^{\frac{u_1}{kT}}\right)}, \\ e^{-\frac{x_2}{kT}} &= \frac{-\nu + |\nu| \left\{1 + \frac{2}{\nu^2}\left(a+\mu e^{\frac{u_2}{kT}}\right)\left(a+\mu e^{\frac{u_1}{kT}}\right)e^{-\frac{u}{kT}}\right\}}{2\left(a+\mu e^{\frac{u_2}{kT}}\right)}, \end{aligned}\right\}$$

откуда:

$$\left.\begin{aligned} &\text{в случае полупроводников с избытком металла } (\nu > 0) \left\{\begin{aligned} e^{-\frac{x_1}{kT}} &= \frac{|\nu|}{a+\mu e^{\frac{u_1}{kT}}}, \\ e^{-\frac{x_2}{kT}} &= \frac{a+\mu e^{\frac{u_1}{kT}}}{|\nu|} e^{-\frac{u}{kT}}, \end{aligned}\right. \\ &\text{в случае полупроводников с избытком металлоида } (\nu < 0) \left\{\begin{aligned} e^{-\frac{x_1}{kT}} &= \frac{a+\mu e^{\frac{u_2}{kT}}}{|\nu|} e^{-\frac{u}{kT}}, \\ e^{-\frac{x_2}{kT}} &= \frac{|\nu|}{a+\mu e^{\frac{u_2}{kT}}}. \end{aligned}\right. \end{aligned}\right\} \quad (2,25)$$

Из формул (2,25) мы получаем для областей d, e, f (см. рис. 69 и 70), согласно (2,19) и (2,23) при $\nu > 0$:

$d)$	$e^{-\frac{x_1}{kT}} = \frac{\lvert\nu\rvert}{a}$	$\sigma_d = ab\left(\frac{\lvert\nu\rvert}{a} + \frac{a}{\lvert\nu\rvert} e^{-\frac{u}{kT}}\right) \approx b\lvert\nu\rvert,$
$e)$	$e^{-\frac{x_1}{kT}} = \frac{\lvert\nu\rvert}{\mu} e^{-\frac{u_1}{kT}}$	$\sigma_e \approx ab \frac{\lvert\nu\rvert}{\mu} e^{-\frac{u_1}{kT}}$
$f)$		$\sigma_f \approx ab \frac{\mu}{\lvert\nu\rvert} e^{-\frac{u-u_1}{kT}}.$

(2,26)

В областях d и e — проводимость электронная; в области f — проводимость дырочная. Формулы (2,26) относятся к случаю $\nu > 0$; при $\nu < 0$ мы получим те же формулы (2,26), в которых, однако, значок 1 надо заменить значком 2, причём электронная и дырочная проводимости меняются местами. При $\nu = 0$ области d, e, f вообще исчезают, как это видно из (2,20).

Мы видим, что в различных областях апроксимации электропроводность, вообще говоря, различна по своей природе. В каждой данной области природа проводимости определяется, с одной стороны, характером решётки:

$$\begin{cases} 1) \text{ случай } u_1 > u_2, \\ 2) \text{ случай } u_2 > u_1, \end{cases}$$

а с другой стороны, характером отклонения от стехиометрического соотношения:

$$\begin{cases} 1) \text{ случай } \nu > 0, \\ 2) \text{ случай } \nu < 0. \end{cases}$$

Все четыре случая сведены на рис. 71. Рисунки 71, *а*, *б*, *в*, *г* представляют собой повторение рис. 70 с той лишь разницей, что теперь одной и той же штриховкой покрыты области, имеющие одинаковую (по природе) проводимость.

Пределы применимости формул (2,24) и (2,26) ясны из рис. 70. Они устанавливаются границами между областями. Вблизи от этих границ приближённые формулы (2,24) и (2,26) перестают быть справедливыми. Выпишем теперь формулы для промежуточных областей, остающиеся справедливыми при переходе через границы из одной температурной области в другую. Эти формулы могут быть записаны так:

$$
\begin{aligned}
&ab) \quad \sigma_{ab} = \frac{\sigma_a^2 + 2\sigma_b^2}{\sqrt{\sigma_a^2 + 4\sigma_b^2}},\\
&ad) \quad \sigma_{ad} = \sqrt{\sigma_a^2 + \sigma_d^2},\\
&bc) \quad \sigma_{bc} = \frac{\sigma_b \sigma_c}{\sqrt{\sigma_b^2 + \sigma_c^2}},\\
&bd) \quad \sigma_{bd} = \frac{1}{2}\sigma_d + \sqrt{\frac{1}{4}\sigma_d^2 + \sigma_b^2},\\
&bf) \quad \sigma_{bf} = \frac{1}{2}\frac{\sigma_b}{\sigma_f}\left\{-\sigma_b + \sqrt{\sigma_b^2 + 4\sigma_f^2}\right\},\\
&ce) \quad \sigma_{ce} = \frac{1}{2}\sigma_e + \sqrt{\frac{1}{4}\sigma_e^2 + \sigma_c^2},\\
&cf) \quad \sigma_{cf} = \frac{1}{2}\frac{\sigma_c}{\sigma_f}\left\{-\sigma_c + \sqrt{\sigma_c^2 + 4\sigma_f^2}\right\},\\
&de) \quad \sigma_{de} = \frac{\sigma_d \sigma_e}{\sigma_d + \sigma_e},\\
&ef) \quad \sigma_{ef} = \sigma_e + \sigma_f.
\end{aligned}
\tag{2,27}
$$

В пределе из формул (2,27) получаются формулы (2,24) и (2,26). Так, например, на основании (2,27) мы имеем:

$$
\sigma_{ab} \approx \begin{cases} \approx \sigma_a \ldots \text{ при } \sigma_a \gg \sigma_b, \text{ т. е. при } T \gg T_1, \\ \approx \sigma_b \ldots \text{ при } \sigma_b \gg \sigma_a, \text{ т. е. при } T_1 \gg T, \end{cases}
$$

и т. д. Таким образом, с помощью формулы (2,27) формулы (2,24) или (2,26) «сшиваются» между собой, и тем самым обеспечивается плавный (без разрывов) переход от одной температурной области в другую, соответствующий непрерывному изменению T при фиксированном значении ν.

Обратимся теперь к вопросу о пределах применимости вообще всей нашей теории. Эти пределы устанавливаются условиями (2,2), которые были приняты нами в качестве исходных предположений. Раскроем физическое содержание этих

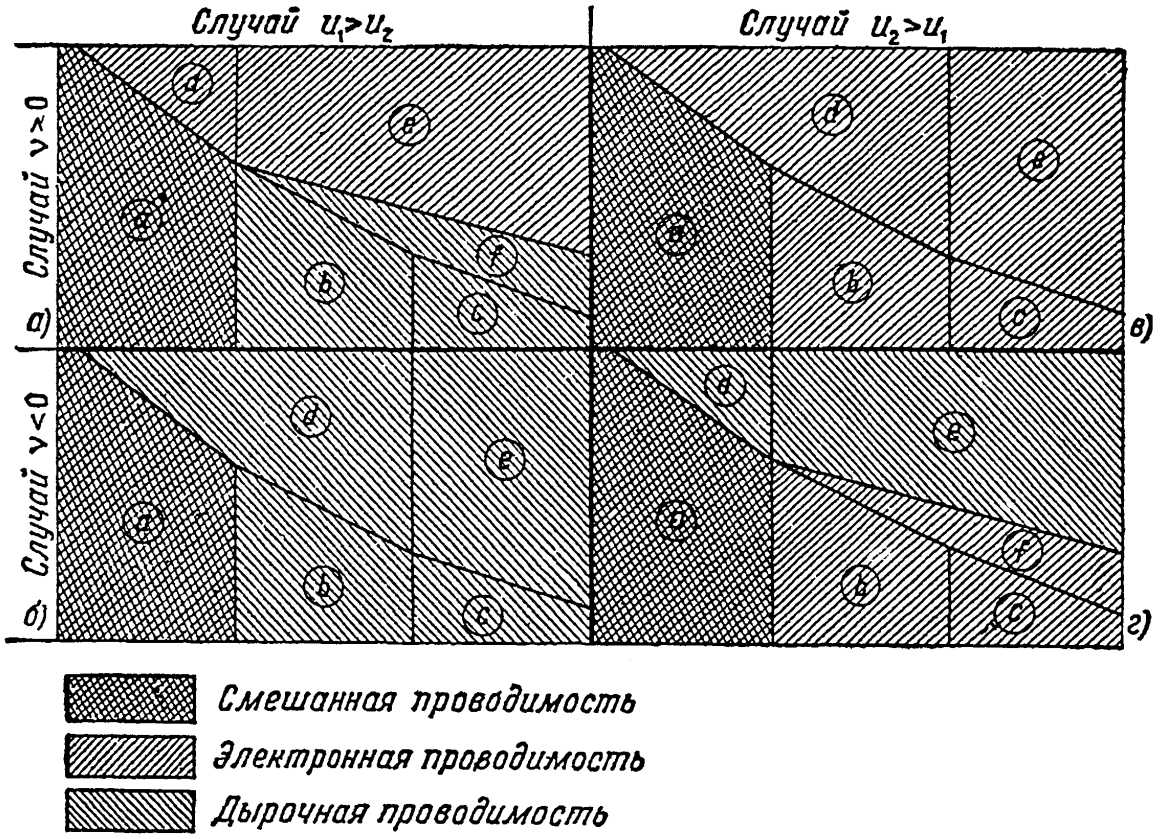


Рис. 71. Природа проводимости в различных областях апроксимации (см. рис. 70).

условий. Можно показать [подставляя в (2,2) выражения (2,23а) и (2,25)], что эти условия могут быть сведены к следующим:

$$\begin{cases} u_i < \frac{1}{2} u, \\ T \ll T', \\ |\nu| \ll \mu, \end{cases} \tag{2,28}$$

где

$$T' = \frac{1}{k}\left(\frac{1}{2} u - u_i\right). \tag{2,29}$$

Легко видеть, что мы имеем всегда

$$\mu > \nu_1 \quad \text{и} \quad T' > T_1,$$

как это изображено на рис. 70. Действительно, на основании (2,22), (2,17), (2,29) и (2,13):

$$\frac{\nu}{\mu} = e^{\frac{T'}{T_1}} = \left(\frac{\mu}{a}\right) e^{\frac{\frac{1}{2}u - u_i}{u_i}} < 1.$$

В заключение заметим, что параметр u_0, фигурирующий в наших формулах, в случае решёток «типа B» является функцией μ (см. главу I, § 5 B). Вместе с тем, функциями μ оказываются также параметры u_1 и u_2, ибо

$$u_0(\mu) + u_1(\mu) + u_2(\mu) = u.$$

Согласно формуле (5,11) главы I, мы имеем:

$$u_0(\mu) = u_0(0) + \frac{2e^2\gamma M}{a}\left(\frac{\mu}{2}\right)^{1/3}, \qquad (2,30)$$

где

a — наикратчайшее расстояние между ионами;
M — коэффициент Маделунга для нашей решётки;
γ — наименьшее из чисел $\frac{q}{r}$ или $\frac{p}{m}$;
p, q — заряды, соответственно, металлического и металлоидного ионов решётки;
m, r — индексы в формуле решётки $M_m R_r$.

Мы имеем по порядку величины:

$$\left.\begin{aligned} u_0 &\approx 0{,}1 - 1{,}0\,\text{eV}, \\ \frac{2e^2\gamma M}{a} &\approx 5 - 10\,\text{eV}, \end{aligned}\right\}$$

и, следовательно, можно считать, что зависимость u_0 от μ практически начинает сказываться лишь при

$$\mu \gtrsim 10^{-9},$$

т. е. когда расстояние между дефектами в среднем делается меньше 1000 атомных расстояний.

E. Обсуждение результатов. По формулам (2,24), (2,26) и (2,27) может быть построена температурная кривая электропроводности $\sigma(T)$ на всём интервале температур. Для случая $\nu > 0$, $u_2 > u_1$ или $\nu < 0$, $u_1 > u_2$ это сделано на рис. 72. На этом рисунке воспроизведена система кривых,

соответствующих различным значениям ν при $\mu = \text{const}$. Кривые перенумерованы в порядке возрастания ν. Каждая кривая распадается на ряд прямолинейных участков, которые обозначены через a, b, c, d, e ... в соответствии с обозначениями областей на рис. 70, б. Вертикалью $T = T'$ и горизонталью $|\nu| = \mu$ наша система кривых ограничивается *слева* и *сверху*.

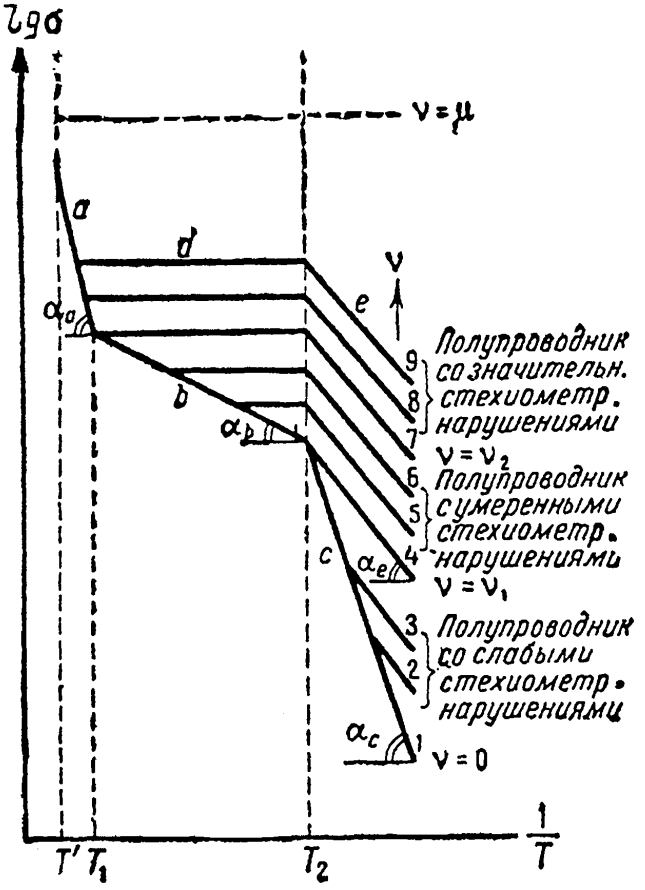


Рис. 72. Зависимость электропроводности σ от температуры T при различных ν и заданном μ. Рисунок относится к случаю $\nu > 0$, $u_2 > u_1 < u_0$; или к случаю $\nu < 0$, $u_1 > u_2 > u_0$.

Наклон кривых:

$$\begin{cases} k \operatorname{tg} \alpha_a = \dfrac{u}{2}, \\ k \operatorname{tg} \alpha_b = \dfrac{u_0}{2} + \dfrac{u_k}{2} = \dfrac{u - u_i}{2}, \\ k \operatorname{tg} \alpha_c = \dfrac{u_0}{2} + u_k, \\ k \operatorname{tg} \alpha_d = 0, \\ k \operatorname{tg} \alpha_e = u_k, \end{cases}$$

где $i = 1$, $k = 2$, если $u_1 > u_2$, или $i = 2$, $k = 1$, если $u_2 > u_1$.

По формулам (2,24) и (2,26) может быть также построена кривая $\varepsilon(T)$, определяющая положение в энергетическом спектре уровня химического потенциала ε как функцию от температуры T. Система таких кривых, соответствующих различным ν при заданном μ, изображена на рис. 73, который относится к случаю $\nu > 0$, $u_2 > u_1$. По этим кривым можно судить о том, как перемещается по спектру уровень ε в зависимости от T. Кривые на рис. 73 перенумерованы в порядке возрастания ν и соответствуют кривым на рис. 72, так что к каждой паре значений μ, ν может быть отнесена определённая кривая $\sigma(T)$ на рис. 72 и определённая кривая $\varepsilon(T)$ на рис. 73, обозначенная тем же номером. На рис. 73 светлой штриховкой покрыты те области, для которых

$\frac{x_1-u_1}{kT}>1$ и $\frac{x_2-u_2}{kT}>1$. Это суть те области, внутри которых наша теория перестаёт быть применимой.

Смена одного прямолинейного участка на кривой $\sigma(T)$ (см. рис. 72) другим прямолинейным участком при переходе из одного температурного интервала в другой свидетельствует об изменении коэффициентов B и C в формуле (2,15):

$$\sigma = Ce^{-\frac{B}{T}}.$$

Это отнюдь, однако, не значит, что «работа ионизации» B является функцией температуры, т. е. что работа высвобождения связанного электрона (или дырки) меняется с температурой. Это означает лишь, что подача электронов в зону проводимости (или дырок в нормальную зону) в разных областях температур происходит из разных резервуаров. Таким образом, изломы на кривой $\sigma(T)$ свидетельствуют о смене механизмов, обусловливающих проводимость.

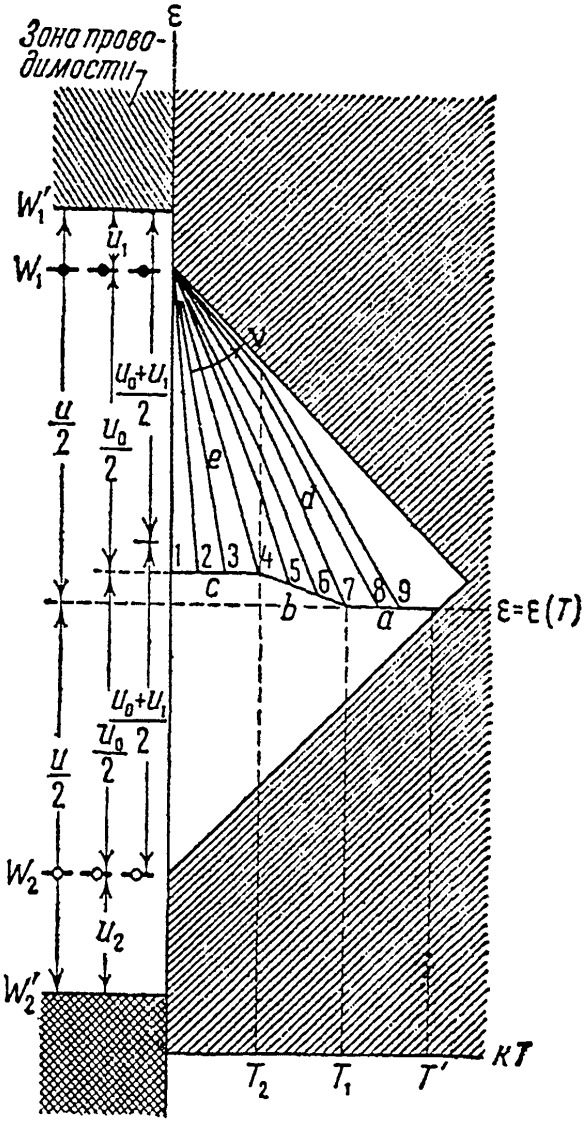


Рис. 73. Положение в спектре уровня химического потенциала ε как функция температуры T при разных ν и заданном μ. Рисунок относится к случаю $\nu>0$, $u_0>u_2>u_1$.

В этом можно легко убедиться, если вернуться к исходному уравнению (2,5), которое служило нам для определения ε. Приближённые формулы (2,24) и (2,26) могут быть получены непосредственно из этого уравнения, если в нём пренебрегать одними членами по сравнению с другими. Действительно, уравнение (2,5) на основании (2,9) можно переписать так:

$$n_1'(T)+n_2(T)=n_2'(T)+n_2(T)+\beta\nu. \qquad (2,31)$$

Отсюда:

	Если в ур-нии (2,31) положить:	то ур-ние (2,31) принимает вид:	Откуда вытекает на основании (2,4) (2,9), (2,10):
a	$\begin{cases} n_1'(T) \gg n_1(T) \\ n_2'(T) \gg n_2(T) + \beta\nu \end{cases}$	$n_1'(T) = n_2'(T)$	$e^{-\frac{x_k}{kT}} = e^{-\frac{u}{2kT}}$
b	$\begin{cases} n_i'(T) \gg n_i(T) \\ n_k(T) \gg n_k'(T) + \beta\nu \end{cases}$	$n_i'(T) = n_k(T)$	$e^{-\frac{x_k}{kT}} = \sqrt{\frac{\mu}{a}}\, e^{-\frac{u_0+u_k}{2kT}}$
c	$\begin{cases} n_1(T) \gg n_1'(T) \\ n_2(T) \gg n_2'(T) + \beta\nu \end{cases}$	$n_1(T) = n_2(T)$	$e^{-\frac{x_k}{kT}} = e^{-\frac{u_0+2u_k}{2kT}}$
d	$\begin{cases} n_i'(T) \gg n_i(T) \\ \beta\nu \gg n'_k(T) + n_k(T) \end{cases}$	$n_j'(T) = \beta\nu$	$e^{-\frac{x_j}{kT}} = \frac{\lvert\nu\rvert}{a} e^{-\frac{u_j}{kT}}$
e f	$\begin{cases} n_i(T) \gg n_i'(T) \\ \beta\nu \gg n_k'(T) + n_k(T) \end{cases}$	$n_j(T) = \beta\nu$	$e^{-\frac{x_j}{kT}} = \frac{\lvert\nu\rvert}{\mu} e^{-\frac{u_j}{kT}}$,

(2,32)

где попрежнему

$$u_i > u_k \text{ и}$$
$$j = 1 \text{ при } \nu > 0,$$
$$j = 2 \text{ при } \nu < 0.$$

Последний столбец этой таблички совпадает с соответствующим столбцом таблички (2,24) и (2,26).

Точное уравнение (2,31) является уравнением баланса, выражающим сохранение общего числа электронов и дырок, запасённых в нашей решётке. Приближённые уравнения, выписанные во втором столбце таблички (2,32), получаются из этого точного уравнения при различных предположениях о том, какие именно уровни в энергетическом спектре решётки при данных условиях являются *основными* носителями электронов и дырок. Пояснением к табличке (2,32) служит рис. 74,

на котором воспроизведён энергетический спектр решётки (рисунок построен для случая $\nu > 0$, $u_2 > u_1$) и стрелками соединены попарно уровни, отдающие электроны, с уровнями, принимающими электроны. Строго говоря, в температурных перебросах электронов участвуют одновременно все уровни, т. е. все четыре процесса, изображённые стрелками на рис. 74, сосуществуют. Однако относительная роль этих процессов меняется с температурой так, что в данном более или менее широком интервале температур один из этих процессов существенно превалирует над всеми остальными. Таким образом, перечисленные процессы при постепенном изменении температуры практически включаются в действие один за другим, причём последовательность, в которой эти механизмы вступают в игру, зависит от природы полупроводника и от того, насколько велики в нём стехиометрические нарушения. Имеем (для случая $\nu > 0$, $u_2 > u_1$, изображённого на рис. 74):

Полупроводники со значительными стехиометрическими нарушениями ($\lvert\nu\rvert > \nu_1$):	$a \leftarrow\!\!\!

(процесс *e* на рис. 74; участок *e* на рис. 72; область *e* на рис. 70, *а*).

При повышении температуры может наступить насыщение, когда все электроны с электроотрицательных уровней W_1 уже исчерпаны, а электроны с электроположительных уровней W_2 ещё не вступили в игру (процесс *d* на рис. 74; горизонтальный участок *d* на рис. 72; область *d* на рис. 70, *а*).

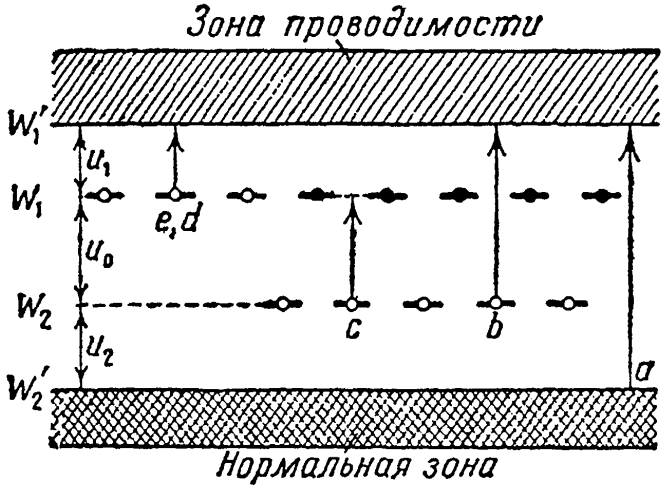


Рис. 74.

Может, однако, случиться, что ещё до наступления насыщения запас электронов на уровнях W_1 начнёт восполняться электронами с уровней W_2 (процесс *c* на рис. 74; участок *c* на рис. 72; область *c* на рис. 70, *а*). В этом случае насыщение так и не будет достигнуто.

При дальнейшем повышении температуры зона проводимости начинает обогащаться непосредственно за счёт электронов с электроположительных уровней W_2, число которых определяется параметром μ, т. е. степенью отклонения нашей реальной решётки от идеальной (процесс *b* на рис. 74; участок *b* на рис. 72; область *b* на рис. 70, *а*).

Наконец, при достаточно высокой температуре, в игру вступает нижняя зона в качестве основного поставщика электронов, и все остальные поставщики отходят на второй план (процесс *a* на рис. 74; участок *a* на рис. 72; область *a* на рис. 70, *а*).

§ 3. Сравнение теории с экспериментом

Прежде чем обратиться к сравнению теории с экспериментом, сделаем ряд замечаний о параметрах

$$u_1,\ u_2,\ u,\ a,\ b,\ \mu,\ \nu,$$

фигурирующих в теоретических формулах предыдущего параграфа. Все эти параметры можно разделить на две группы.

а) Параметры, являющиеся характеристикой материала. Сюда относятся:

$$u_1,\ u_2,\ u,\ a,\ b.$$

Значения этих параметров остаются одними и теми же для всех образцов [1]).

б) Параметры, зависящие от биографии образца, т. е. от того, какой обработке был подвергнут образец. Сюда относятся:

$$\mu,\ \nu.$$

Значения этих параметров меняются от образца к образцу.

Таким образом, поскольку в выражение для σ входят параметры μ и ν [см. общую формулу (2,12)], проводимость σ ни в коем случае нельзя рассматривать как константу вещества.

А. Параметры u_1, u_2, u. В физическом отношении эти параметры представляют собой работу ионизации, т. е. работу образования свободного электрона (дырки). Этими параметрами определяется наклон температурной кривой $\sigma(T)$ на рис. 72; через эти параметры выражается коэффициент B в формуле

$$\sigma = Ce^{-\frac{B}{T}}.$$

Коэффициент B, как это видно из рис. 72, изменяется, если двигаться по рис. 72 справа налево, т. е. в сторону возрастающих температур при $\nu = \text{const}$, или снизу вверх, т. е. в сторону возрастающих ν при $T = \text{const}$. Это, однако, не значит, что работа ионизации является непосредственной функцией от T и ν. Это означает лишь, что, перемещаясь по рис. 73 вдоль горизонтали или вдоль вертикали, мы переходим постепенно от одного механизма проводимости к другому.

Путь к непосредственному экспериментальному определению параметров u_1, u_2, u лежит в изучении оптического спектра поглощения нашей кристаллической решётки. Этот спектр обычно содержит в себе, наряду с яркой *собственной*

---

[1]) Имеются в виду значения u_1 и u_2, соответствующие $\mu = 0$ [см. (2,30)].

полосой поглощения, интенсивность и положение которой в спектре целиком определяются природой решётки, ещё две значительно более слабые, так называемые «*примесные*» полосы, лежащие с красной стороны от собственной полосы, интенсивность которых определяется не только природой решётки, но находится в явной зависимости от всей её биографии и может быть искусственно увеличена или уменьшена при соответствующей обработке образца. Собственную полосу поглощения естественно интерпретировать как переход электронов из нормальной зоны в зону проводимости [1]), а примесные полосы — как переброс электронов из нормальной зоны на уровни W_1 и с уровней W_2 в зону проводимости (см. рис. 67 и 68). При такой интерпретации спектра поглощения параметры u, u_1, u_2 могут быть определены из красных границ собственной полосы и примесных полос. Если частоты этих границ обозначить соответственно через ν, ν_1, ν_2, то

$$\left.\begin{aligned} h\nu &= u, \\ h\nu_1 &= u - u_1, \\ h\nu_2 &= u - u_2. \end{aligned}\right\}$$

Заметим, однако, что оптическая работа ионизации $h\nu$ не всегда равна термической работе ионизации u, а иногда может быть больше неё, в зависимости от тех правил отбора, которые руководят перебросами электронов из нормальной зоны в зону проводимости. В том случае, когда $h\nu > u$, измерение красных границ является недостаточным для определения параметров u, u_1, u_2.

Для полупроводников u, u_1, u_2 имеют порядок величины десятых долей электрон-вольта, а для щёлочно-галоидных кристаллов — порядок нескольких электрон-вольтов. К этой и только к этой чисто количественной разнице и сводится различие между полупроводниками и щёлочно-галоидными кристаллами, электрические и оптические свойства которых на первый взгляд кажутся качественно столь различными. В оптическом отношении полупроводник можно рассматривать, как щёлочно-га-

[1]) Впрочем, такая интерпретация приводит к ряду затруднений, например, к невозможности объяснить без дополнительных гипотез отсутствие фотопроводимости в области собственной полосы (см. гл. VI).

лоидный кристалл, все оптические особенности которого как бы сдвинуты в инфракрасную область спектра, мало удобную для экспериментатора. С другой стороны, в электрическом отношении щёлочно-галоидный кристалл можно рассматривать как полупроводник, все электрические особенности которого как бы сдвинуты в сторону высоких температур, выходящих за рамки, доступные экспериментатору.

Заметим, что параметры u_1 и u_2 можно считать константами материала лишь до тех пор; пока решётка не слишком сильно отличается от идеальной (μ мало). Иначе говоря, параметры u_1 и u_2 служат характеристикой материала лишь до тех пор, пока повреждения, присутствующие в решётке, можно считать невзаимодействующими. Учёт этого взаимодействия, которым при значительных μ уже нельзя пренебрегать, приводит к ослаблению связей, удерживающих электроны и дырки около мест повреждений, и, следовательно, к уменьшению u_1 и u_2.

В. Параметры μ и ν. Обратимся теперь к параметрам μ и ν, которые представляют собой две независимые характеристики всякой реальной решётки и играют важную роль в теории. Параметр μ характеризует относительное число дефектов, содержащихся в решётке. Этим параметром определяется степень отклонения нашей реальной решётки от идеальной. Параметр ν характеризует преобладание дефектов одного сорта над дефектами другого сорта. Этим параметром определяется степень отклонения от стехиометрического соотношения. Существенно, что в теории фигурируют именно два параметра. Физически это означает, что свойства решётки определяются не общим числом дефектов, а числом дефектов каждого сорта в отдельности (см. главу I, § 1).

Подвергая образец той или иной обработке, мы тем самым варьируем параметры μ и ν. Заметим, что эти два теоретически независимых параметра практически до известной степени оказываются связанными между собой, хотя эта связь носит коррелятивный, но не строго функциональный характер. Переходя от образца к образцу, мы не можем изменять параметр ν, оставляя при этом неизменным параметр μ (вопреки тому, что изображено на теоретическом рис. 72). Изменяя ν, мы практически не можем обеспечить постоянство μ, ибо те способы, которыми пользуется экспериментатор для изменения ν (прокаливание образца в вакууме или в той или иной атмосфере

при том или ином давлении), являются в то же время способами, вызывающими изменение μ. Возрастание ν практически неизбежно сопровождается возрастанием μ, причём эта не поддающаяся количественному учёту связь может нами регулироваться лишь до некоторой степени за счёт изменений условий обработки. Поэтому можно заведомо утверждать, что два образца с различными ν будут, как правило, отличаться также и значениями μ.

Варьируя параметры μ и ν, мы получаем право перемещать прямолинейные участки b, d, e на нашей теоретической кривой $\sigma(T)$, изображённой на рис. 72, вверх и вниз по чертежу, оставляя их при этом параллельными самим себе, в то время, как положение прямых a и c, фиксированное самой природой решётки, должно оставаться неизменным. Таким образом, нашу теоретическую кривую можно представлять себе составленной из отдельных прямолинейных кусков, которые мы можем до известной степени произвольно прикладывать друг к другу.

Заметим, что не следует искать непосредственной количественной связи между начальными ординатами этих прямых и степенью нарушения стехиометрического соотношения ν, как это делают многие авторы и как это следует из элементарной теории Вильсона. Начальная ордината может зависеть не только от ν, но и от μ, или только от μ, или же может вовсе не зависеть ни от ν, ни от μ [как это видно из формул (2,24) и (2,26)], в зависимости от того, какой именно из прямолинейных участков нашей теоретической кривой попадает в поле зрения экспериментатора, иначе говоря, в зависимости от того, каков механизм проводимости, действующий в нашем образце в исследуемом интервале температур.

Для непосредственного экспериментального определения параметров μ и ν можно указать следующие пути:

а) оптический метод (из интенсивности примесных полос поглощения);

б) электрический метод (из постоянной Холла);

в) химический метод (непосредственным анализом).

Эти методы позволяют определить значения μ и ν лишь при условии введения тех или иных дополнительных гипотез.

а) Оптический метод основан на том факте, что интенсивность примесной полосы должна быть пропорциональна

числу доноров или числу акцепторов, содержащихся в решётке. Для того чтобы отсюда сделать какие-либо заключения о значениях параметров μ и ν, необходимо заранее принять ту или иную гипотезу о механизме возникновения данной примесной полосы, т. е. необходимо заранее знать, какие именно электронные переходы обусловливают возникновение данной полосы. Так, если наша полоса связана с перебросом электронов из нормальной зоны на уровни W_1, или с уровней W_2 в зону проводимости, и если при этом считать $|\nu| \ll \mu$, то в этом случае интенсивность примесной полосы позволяет судить о величине μ, но не даёт нам никаких сведений о величине ν. Если же наша примесная полоса вызвана переходами электронов, например, с уровней W_1 в зону проводимости (случай $\nu > 0$) или из нормальной зоны на уровни W_2 (случай $\nu < 0$), то в этом случае, наоборот, из интенсивности примесной полосы может быть определён параметр ν, но неопределённым остаётся параметр μ.

б) Из эффекта Холла может быть вычислена концентрация электронов в зоне проводимости или дырок в нормальной зоне, если исходить из предположения, что проводимость обусловлена носителями только одного какого-либо знака. Из концентрации свободных электронов (или дырок) в некоторых случаях можно определить ν или μ, но для этого надо заранее знать, с каким именно температурным участком (участки a, b, c, d, e на рис. 72) мы имеем дело, ибо на разных участках, как это видно, например, из формул (2,24) и (2,26), концентрация связана с параметрами ν и μ различным образом. Знание концентрации само по себе, без введения дополнительных гипотез, ещё не гарантирует однозначного определения параметров ν и μ.

в) Непосредственный химический анализ позволяет однозначно определить количество содержащейся в образце примеси, т. е. степень отклонения от стехиометрического соотношения. Однако из химических данных можно вычислить лишь *среднюю* концентрацию примеси (ибо химический метод определяет общее содержание примеси во всём объёме), в то время как в теоретических формулах фигурирует *истинная* концентрация ν, господствующая внутри образца, которая заведомо меньше, чем средняя концентрация. Действительно, примесь распределяется по объёму неравномерно: в поверхностном слое концентрация, конечно, гораздо больше, чем внутри

образца. Поэтому значения ν, полученные из химического анализа, не могут быть с достаточной уверенностью использованы в теоретических формулах.

С. Параметры a и b. Обратимся, наконец, к параметрам a и b, играющим в теории второстепенную роль. Мы имеем:

$$a = \frac{g}{h^3 a}(2\pi m_0 kT)^{3/2}, \quad b = \frac{1}{2} evZ\sqrt{mr},$$

где

$$a \approx \frac{1}{2} mZ \approx \frac{1}{2} rZ.$$

Здесь:

g — статистический вес рассматриваемой зоны,
m_0 — эффективная масса электрона (или дырки),
v — подвижность свободного электрона (или дырки),
Z — общее число молекул $M_m R_r$ в единице объёма,
m, r — индексы в химической формуле.

Подвижность v зависит от эффективной массы m_0, которая, в свою очередь, определяется шириной ΔW рассматриваемой зоны:

$$v = \frac{4}{3}\frac{el}{(2\pi m_0 kT)^{1/2}}, \quad m_0 = \frac{3}{2}\frac{h^2}{\pi^2 a^2 \Delta W}.$$

(Вторая из этих формул относится к случаю, когда рассматриваемая зона есть s-зона, т. е. $g = 2$, и решётка кубическая; a — постоянная решётки.) Здесь l — длина свободного пробега (l убывает при росте T).

Мы видим, что параметры a и b суть функции температуры: a медленно возрастает, а b медленно убывает при повышении температуры. Однако мы можем считать в наших формулах $a = \text{const}$ и $b = \text{const}$. Это допустимо не только потому, что обе эти величины меняются с температурой медленно (медленно по сравнению с экспоненциальным множителем), но и потому, что они меняются в противоположные стороны и поэтому, входя в наши формулы, как правило, в виде произведения, если не целиком, то во всяком случае в значительной степени, погашают друг друга.

Параметр a можно считать константой вещества. Его значение в конечном счёте определяется шириной зоны ΔW. Значение ΔW может быть оценено, хотя и очень грубо, из

оптических данных: можно считать, что ΔW равно ширине примесной полосы поглощения. Обычно ΔW по порядку величины равно $\sim 1\,\text{eV}$.

Параметр b, строго говоря, нельзя считать константой вещества, поскольку подвижность v зависит не только от ΔW, но и от l. Действительно, длина свободного пробега l косвенным образом может зависеть от количества дефектов в решётке, ибо такие дефекты, помимо всего прочего, играют роль дополнительных центров рассеяния для электронных волн.

Подвижность v в некоторых случаях может быть определена, если одновременно измерены электропроводность σ и постоянная Холла R. Действительно:

$$\left.\begin{aligned} \sigma &= e\,(v_1 n_1' + v_2 n_2'), \\ R &= \frac{3\pi}{8e}\cdot\frac{v_2^2 n_2' - v_1^2 n_1'}{(v_2 n_2' + v_1 n_1')^2}, \end{aligned}\right\}$$

где попрежнему n_1' — число свободных электронов, n_2' — число свободных дырок, v_1 и v_2 — подвижности, соответственно, электронов и дырок (в § 2 мы считали $v_1 \approx v_2 \approx v$). Если в проводимости участвуют носители только одного какого-нибудь знака, то

$R\sigma = \frac{3\pi}{8} v_2$ ..., если $n_1' = 0$ (дырочная проводимость),

$R\sigma = -\frac{3\pi}{8} v_1$ ..., если $n_2' = 0$ (электронная проводимость).

В этом случае подвижность однозначно определяется произведением $R\sigma$. Обычно мы имеем:

$$v_1 \approx v_2 \approx 1 - 100 \text{ вольт}^{-1} \cdot \text{сек}^{-1} \cdot \text{см}^{-2}.$$

Если же в проводимости участвуют носители обоих знаков, то знания R и σ ещё недостаточно, чтобы судить о подвижности.

Полагая

$$\left\{\begin{aligned} &\Delta W \approx 0{,}5 - 5\,\text{eV}, \\ &v \approx 1 - 100 \text{ вольт}^{-1} \cdot \text{сек}^{-1} \cdot \text{см}^2, \\ &Z \approx 10^{22} - 10^{23}\ \text{см}^{-3}, \end{aligned}\right.$$

мы получаем

$$\frac{m_0}{m} \approx 0{,}1 - 1,$$

где m — истинная масса электрона, и, следовательно:

$$\begin{cases} a \approx 10^{-1} - 10^{-3}, \\ b \approx 10^{3} - 10^{6}\ \text{ом}^{-1} \cdot \text{см}^{-1}. \end{cases}$$

Итак, из анализа параметров, входящих в теоретические формулы, видно, что для того, чтобы сравнение теории с экспериментом могло бы быть проведено во всех деталях, необходимо, чтобы экспериментатор дал в руки теоретика одновременно следующие сведения:

а) кривые поглощения в достаточно широком интервале частот;

б) σ как функцию T в достаточно широком интервале температур;

в) R как функцию T в том же интервале температур.

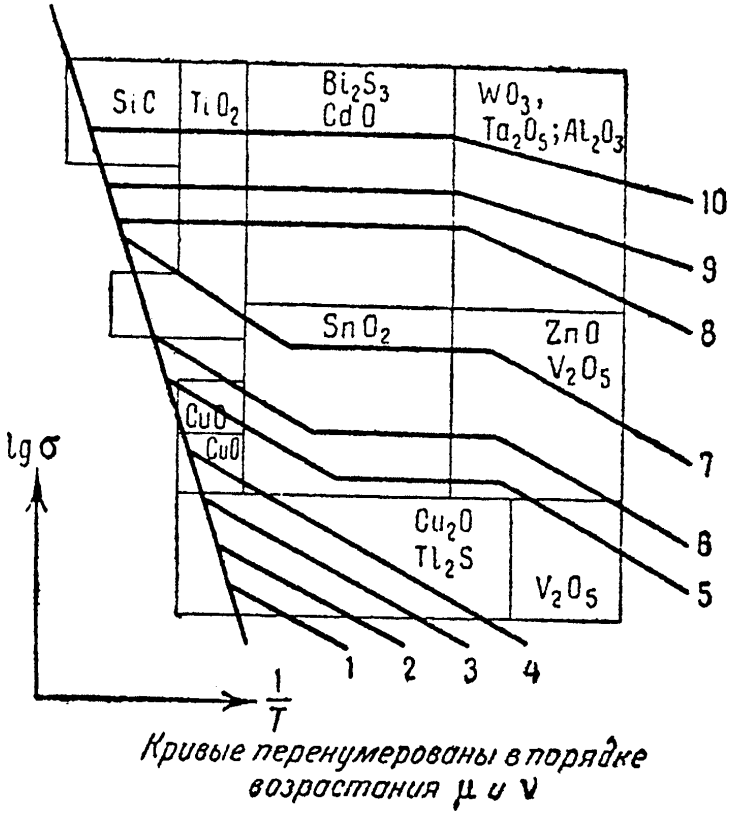
При этом существенно, чтобы все эти сведения относились обязательно *к одному и тому же образцу*. Нельзя, например, вычислять подвижность v, измерив σ на одном образце, а R — на другом. К сожалению, существующий в настоящее время экспериментальный материал не удовлетворяет этому требованию. Поэтому, сравнивая теорию с экспериментом, теоретик стоит перед необходимостью до некоторой степени произвольно выбирать значения для констант, входящих в теоретические формулы, заботясь лишь о том, чтобы по порядку величины они оставались правдоподобными.

D. Сравнение теоретических кривых с экспериментальными. Обратимся к качественному сопоставлению теории, развитой в предыдущем параграфе, с экспериментальными данными.

На рис. 75 изображено семейство теоретических кривых. Здесь различные кривые соответствуют различным значениям параметров μ и ν, т. е. образцам, отличающимся друг от друга общим количеством дефектов и степенью отклонения от стехиометрического соотношения. Таким образом, на рис. 75, в отличие от рис. 72, параметр μ не фиксирован, а варьируется одновременно с параметром ν. На рис. 75 кривые перенумерованы в порядке возрастания μ и ν.

Сопоставим рис. 75 с рис. 59. Весь экспериментальный материал, собранный на рис. 59, может быть размещён на

рис. 75. Пунктирные линии делят рис. 75 на отдельные участки; каждому такому участку соответствует своя группа кривых на рис. 59. Заметим, что сопоставление экспериментальной кривой с тем или иным участком теоретической кривой не всегда является однозначным. Заметим ещё, что рис. 75 безмасштабный. В действительности резкие углы между прямолинейными участками теоретической кривой могут быть сглажены в той или иной степени в зависимости от масштаба, принятого на чертеже.



Кривые перенумерованы в порядке возрастания μ и ν

Рис. 75. Схема теоретических кривых и сопоставление их с экспериментальными (ср. с рис. 59).

Таким образом, разнообразие экспериментальных кривых проистекает не от разнообразия моделей, а является лишь результатом того, что в поле зрения различных экспериментаторов попадают различные участки теоретической кривой, которая имеет универсальный характер. Все полупроводники могут быть описаны с помощью одной и той же энергетической модели (см. рис. 67 и 68), причём качественные различия между ними имеют чисто количественное происхождение.

Из сопоставления рисунков 59 и 75 мы видим, что теория находится в удовлетворительном качественном согласии с экспериментом.

Обратимся теперь к количественному сравнению некоторых экспериментальных кривых с теоретическими.

Здесь в первую очередь нас должны интересовать экспериментальные кривые с явно выраженной кривизной, которые, с точки зрения теоретической интерпретации, соответ-

ствуют переходу от одного прямолинейного участка к другому. Мы рассмотрим следующие материалы (см. рис. 59):

1) SiC — по измерениям Курчатова,
2) SnO_2 — по измерениям Бауера (три образца),
3) CuO — по измерениям Хартмана и Гохберга (два образца),
4) ZnO — по измерениям Вагнера и Яндера (два образца),
5) V_2O_5 — по измерениям Арсеньевой, Курчатова и Гохберга (три образца).

Экспериментальные точки и соответствующие теоретические кривые приводятся на рис. 76—80.

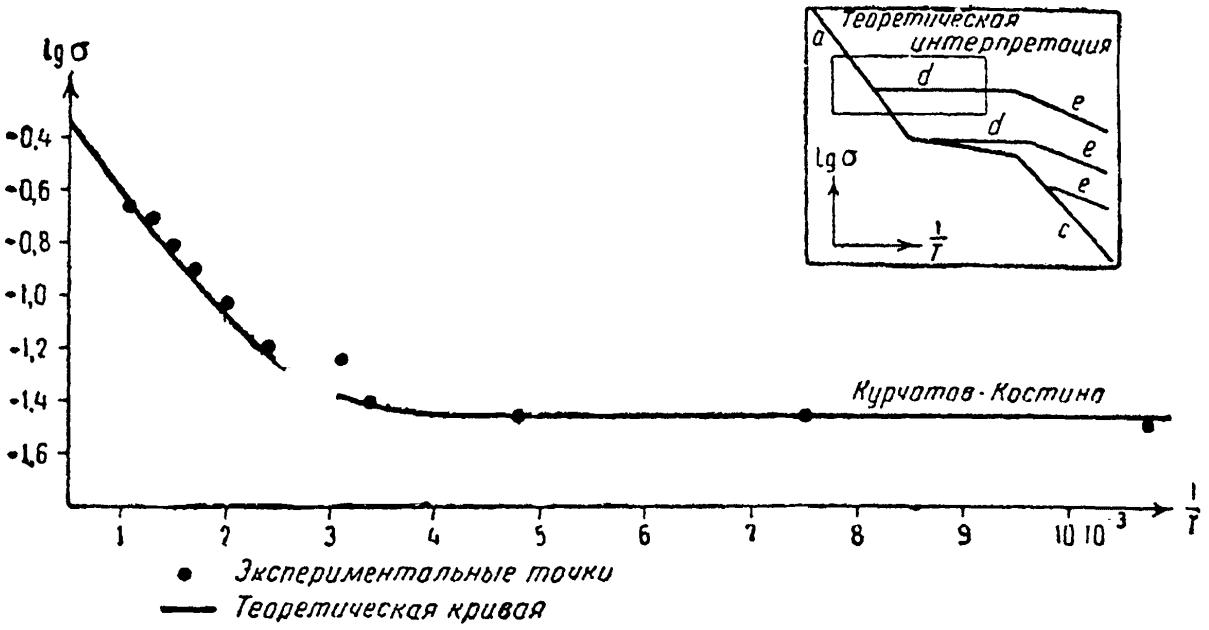


Рис. 76. Зависимость электропроводности от температуры для SiC

Теоретическая формула:

$$\sigma_{ad} = \sqrt{\sigma_d^2 + \sigma_a^2} = b\sqrt{\nu^2 + 4a^2 e^{-\frac{a}{kT}}},$$

где

$$a = 2\frac{A}{Z}, \qquad b = \frac{1}{2} evZ.$$

Приняты следующие значения параметров:

$$\begin{cases} A = 0{,}25 \cdot 10^{19}\ см^{-3}, \\ Z = 10^{22}\ см^{-3}, \\ v = 1\ \text{вольт}^{-1} \cdot сек^{-1} \cdot см^2, \\ a = 0{,}2\ \text{eV}, \\ \nu = 0{,}4 \cdot 10^{-4}. \end{cases}$$

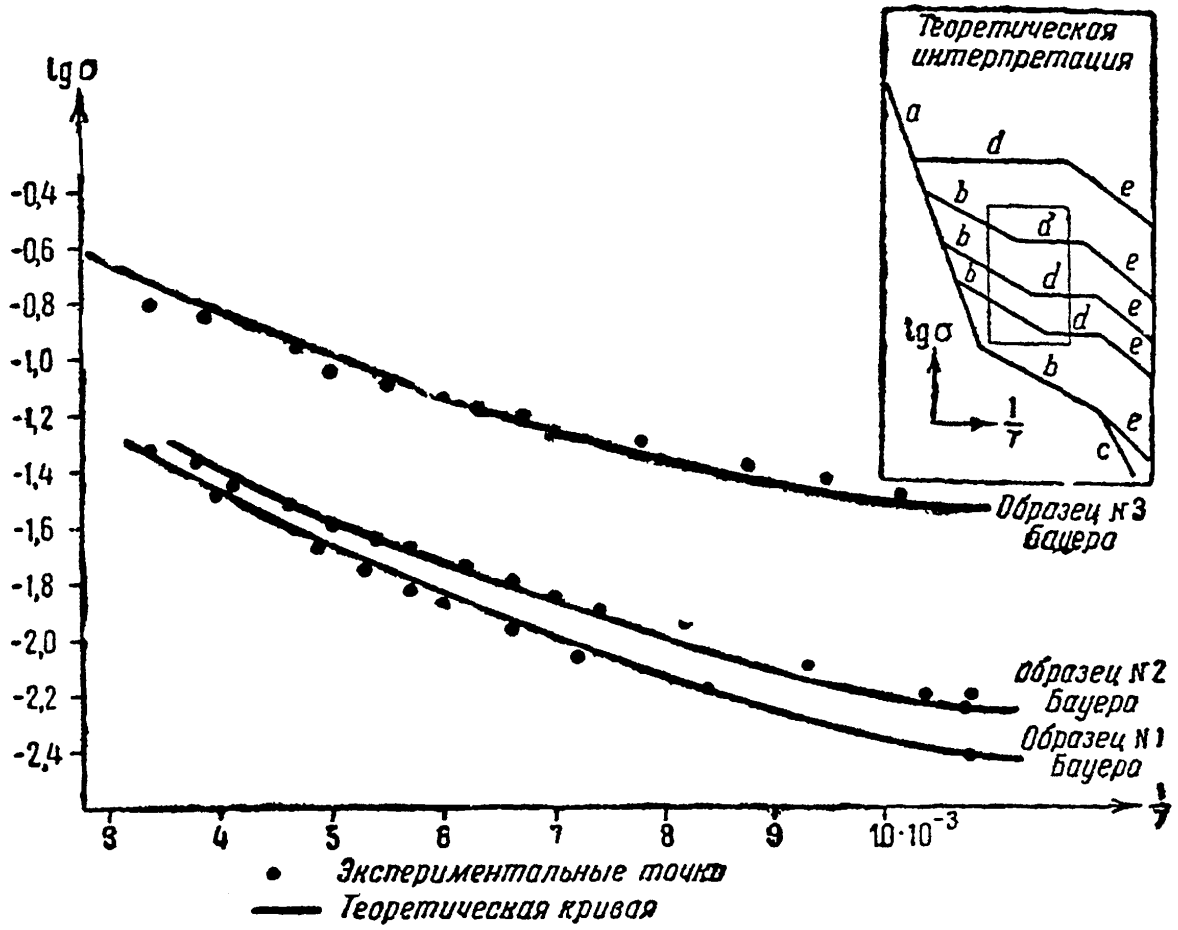


Рис. 77. Зависимость электропроводности от температуры для SnO_2. Теоретическая формула:

$$\sigma_{bd}=\frac{1}{2}\sigma_d+\sqrt{\sigma_b^2+\frac{1}{4}\sigma_d^2}=\frac{1}{2}b\left\{\nu+\sqrt{\nu^2+4a\mu e^{\frac{u-u_i}{kT}}}\right\},$$

где

$$a=\frac{A}{Z},\quad b=evZ.$$

Приняты следующие значения параметров:

$$\begin{cases} A=10^{19}\ см^{-3} \\ Z=10^{22}\ см^{-3} \\ v=1\ \text{вольт}^{-1}сек^{-1}см^2 \\ u-u_i=0{,}08\ \text{eV} \end{cases}$$

$$\begin{cases} \mu=5\cdot10^{-5} \\ \nu=2{,}5\cdot10^{-6} \end{cases}\text{образец № 1}$$

$$\begin{cases} \mu=8\cdot10^{-5} \\ \nu=3{,}0\cdot10^{-6} \end{cases}\text{образец № 2}$$

$$\begin{cases} \mu=8\cdot10^{-4} \\ \nu=2{,}5\cdot10^{-5} \end{cases}\text{образец № 3}$$

Значение для v выбрано в согласии с результатами измерений холл-эффекта на тех же образцах.

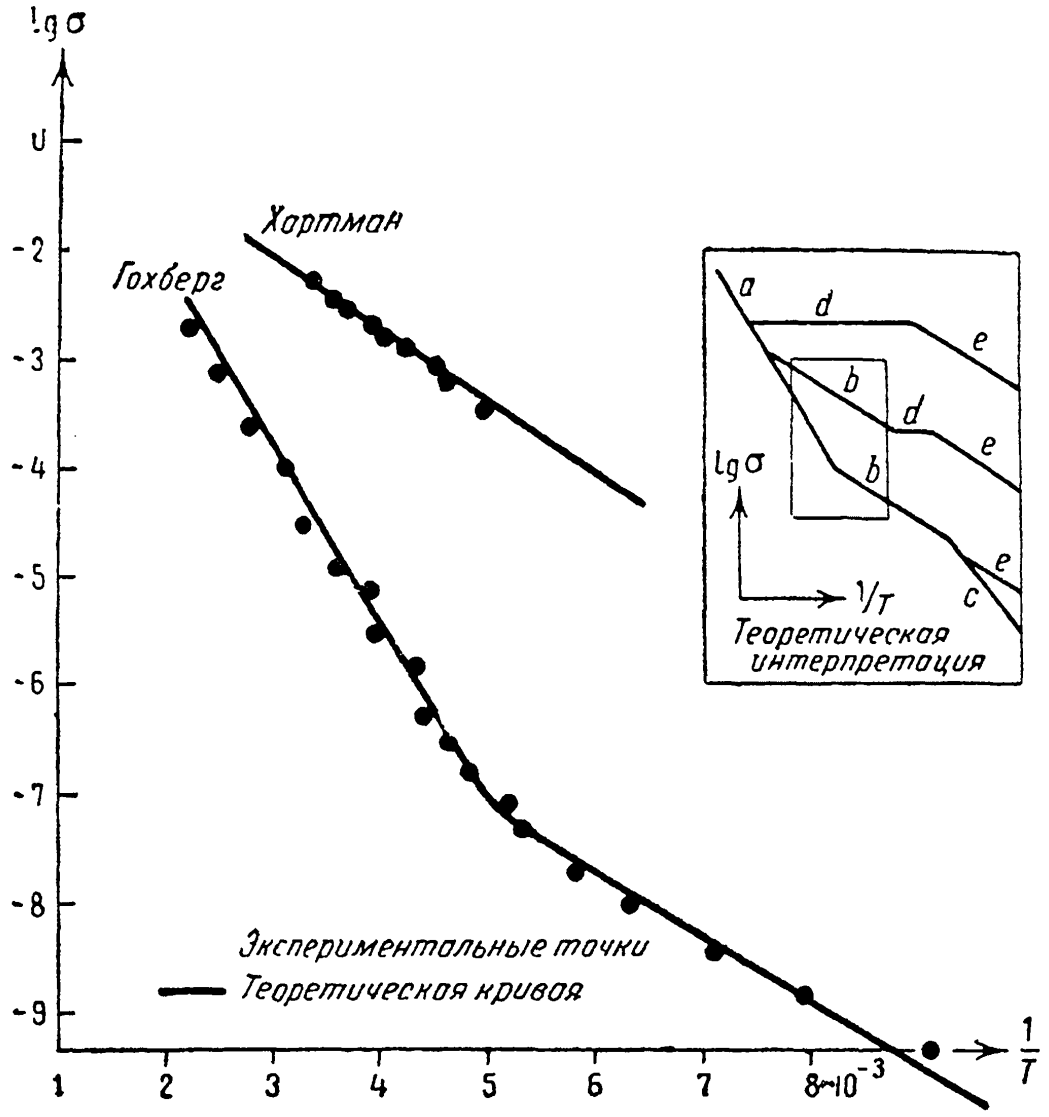


Рис. 78. Зависимость электропроводности от температуры для CuO.

Теоретическая формула:

$$\sigma_{ab}=\frac{\sigma_a^2+2\sigma_b^2}{\sqrt{\sigma_a^2+4\sigma_b^2}}=\frac{\sqrt{a}\cdot be^{-\frac{u}{2kT}}\left(2a+\mu e^{\frac{u_1}{kT}}\right)}{\sqrt{a+\mu e^{\frac{u_1}{kT}}}},$$

где, согласно (3,19):

$$a=2\frac{A}{Z},\quad b=\frac{1}{2}evZ.$$

Приняты следующие значения параметров:

$A=10^{21}$ *см*$^{-3}$
$Z=10^{22}$ *см*$^{-3}$
$v=20$ вольт$^{-1}$*сек*$^{-1}\cdot$*см*^{2}
$u=0{,}64$ eV

$u-u_1=0{,}24$ eV
$\mu=0{,}8\cdot10^{-4}$...образец Хартмана
$\mu=0{,}8\cdot10^{-15}$...образец Гохберга

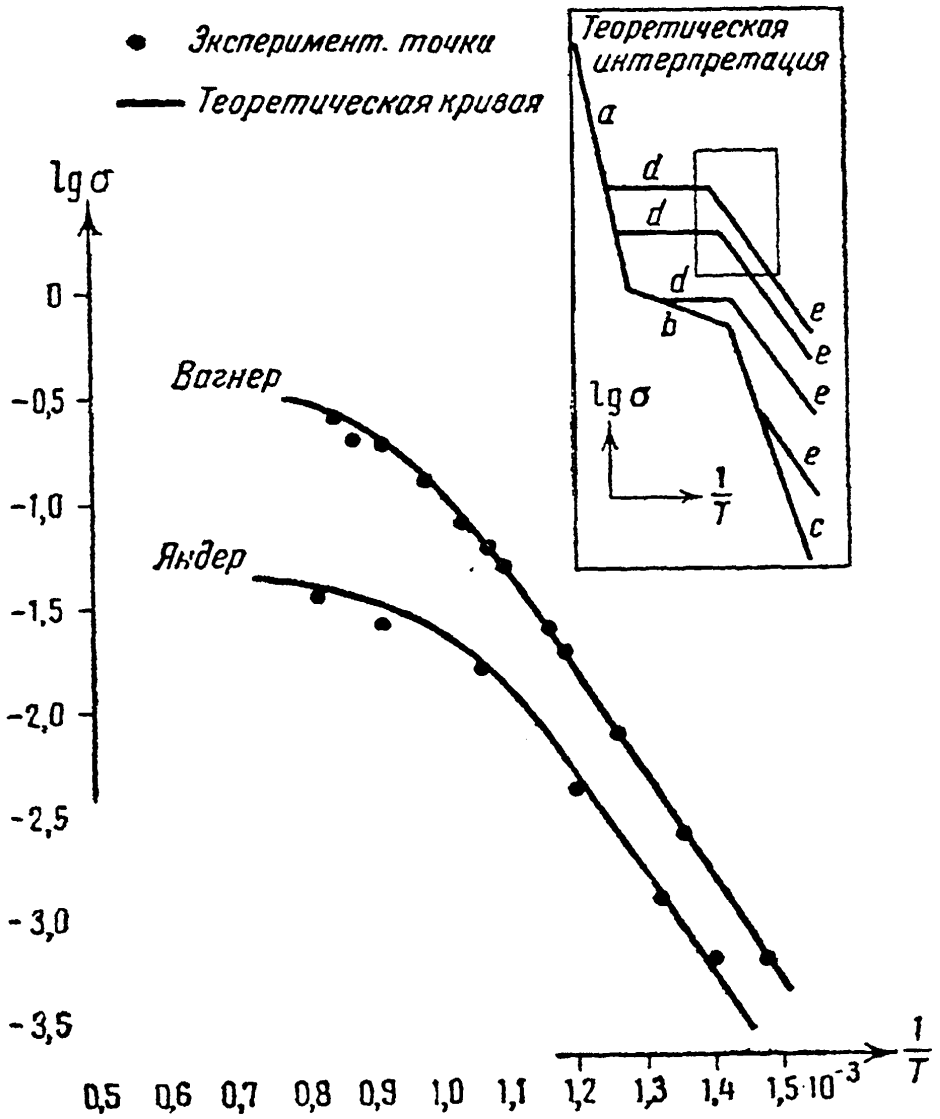


Рис. 79. Зависимость электропроводности от температуры для ZnO.

Теоретическая формула:

$$\sigma_{de} = \frac{\sigma_d \sigma_e}{\sigma_d + \sigma_e} = \frac{ab\nu}{a + \mu e^{\frac{u}{kT}}},$$

где, согласно (3,19):

$$a = 2\frac{A}{Z}, \quad b = \frac{1}{2} ev Z.$$

Приняты следующие значения параметров:

$A = 10^{21}$ $см^{-3}$
$Z = 10^{22}$ $см^{-3}$
$v = 500$ вольт$^{-1}$·сек$^{-1}$·см^2
$u^1 = 1$ eV

$\mu = 1{,}4 \cdot 10^{-6}$, $\nu = 1{,}1 \cdot 10^{-8}$ — образец Яндера
$\mu = 5 \cdot 10^{-6}$, $\nu = 1{,}0 \cdot 10^{-6}$ — образец Вагнера

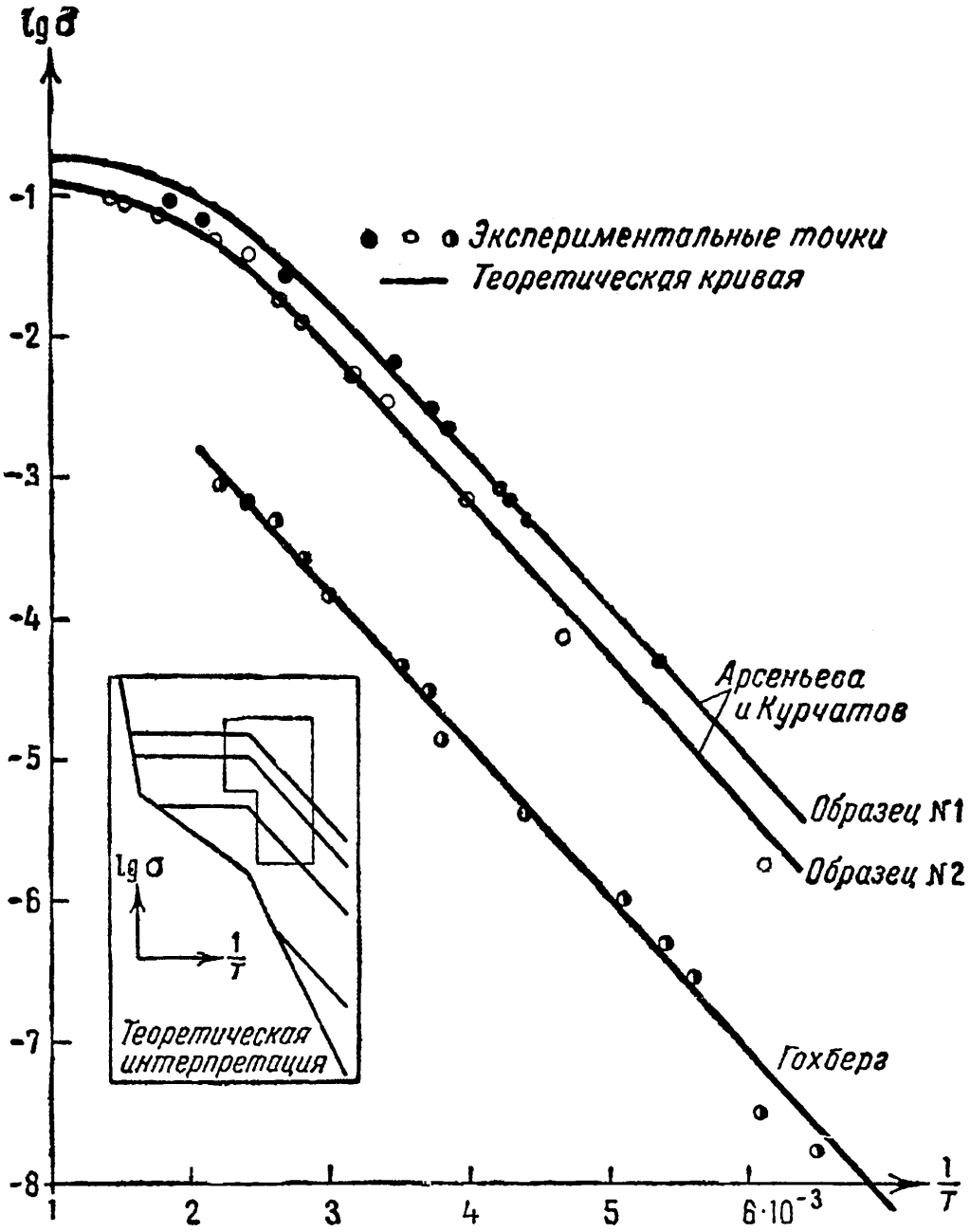


Рис. 80. Зависимость электропроводности от температуры для V_2O_5.

Теоретическая формула: $\sigma_{de} = \frac{\sigma_d \sigma_e}{\sigma_d + \sigma_e} = \frac{ab\nu}{a + \mu e^{\frac{u_1}{kT}}}$,

где, согласно (3,19): $a = \frac{A}{\frac{1}{2}mZ} = \frac{A}{Z}$, $b = \frac{1}{2} evZ\sqrt{mr} = \frac{\sqrt{10}}{2} evZ$.

Приняты следующие значения параметров: $A = 10^{21} см^{-1}$; $Z = 10^{22} см^{-3}$; $v = 6{,}4$ вольт$^{-1}\cdot$сек$^{-1}\cdot см^2$; $u_1 = 0{,}22$ eV; $\mu = 1{,}0\cdot10^{-3}\nu = 1{,}2\cdot10^{-5}$...образец № 1 Арсеньевой; $\nu = 0{,}8\cdot10^{-5}$... образец № 2 Арсеньевой; $\nu = 0{,}64\cdot10^{-7}$...образец Гохберга. Значение $u_1 = 0{,}22$ eV находится в согласии с оптическими измерениями Арсеньевой, которая для красной границы примесной полосы получила $u_1 = h\nu \approx 0{,}25$ eV. Значение для ν ($\nu = 1{,}2\cdot10^{-5}$ и $\nu = 0{,}8\cdot10^{-5}$) заимствованы у Арсеньевой, которая получила эти значения, измеряя интенсивность примесной полосы поглощения.

Наличие кривизны в температурной кривой указывает нам на наличие двух разных механизмов проводимости, которые действуют одновременно, но относительная роль которых изменяется с изменением температуры. Количественное сравнение теории с экспериментом имеет смысл именно на этих промежуточных криволинейных участках, которые выражаются формулами (2,27). Мы видим из (2,27), что прямолинейными участками

$$\sigma_a,\ \sigma_b,\ \sigma_c,\ \sigma_d,\ \sigma_e,\ \sigma_f \tag{3,1}$$

вполне предопределяется ход кривой в промежуточных областях (2,27).

В наших теоретических формулах фигурируют параметры

$$\left.\begin{array}{l} a,\ b,\ \mu,\ \nu, \\ u_0,\ u_1,\ u_2. \end{array}\right\} \tag{3,2}$$

Из прямолинейных участков (3,1), согласно (2,24) и (2,26), мы можем определить параметры (3,2) и, наоборот, задавая эти параметры, мы тем самым фиксируем прямолинейные участки (3,1). Заметим, что при выборе значений для этих параметров мы должны учитывать совокупность всех экспериментальных фактов, относящихся к данному материалу, сопоставляя данные по электропроводности, холлэффекту и спектрам поглощения, полученные часто различными авторами. Однако в большинстве случаев экспериментальные данные не отличаются полнотой, и поэтому при сравнении экспериментальных кривых с теоретическими неизбежен известный произвол при выборе значений для параметров (3,2). Однако наличие свободных параметров отнюдь ещё не обеспечивает априори попадания экспериментальных точек на теоретическую кривую, ибо этими параметрами фиксируются лишь прямые (3,1), являющиеся асимптотами интересующих нас кривых. На каждом из рисунков 76—80 выписана соответствующая теоретическая формула и принятые значения для параметров, причём специально оговорены те случаи, когда эти значения возможно было заимствовать из эксперимента. Рассматривая эти рисунки, видим, что всё разнообразие экспериментальных кривых действительно может быть уложено в общую теоретическую формулу (2,12), из которой при различных апроксимациях получается серия формул (2,24), (2,26), (2,27).

Литература к главе IV

[1] Gudden, *Erlangen Ber.*, **62**, 289, 1930.
[2] Bauer, *Ann. d. Phys.*, **30**, 433, 1937.
[3] Гохберг, *ЖЭТФ*, **7**, 1090, 1937.
[4] Hartmann, *Zs. f. Phys.*, **102**, 709, 1936.
[5] J. Kurtschatow, Kostina, Rusinow, *Sov. Phys.*, **7**, 129, 1935.
[6] Wagner und Baumbach, *Zs. f. Phys. Chem.*, **B**, 22, 199, 1935.
[7] Jander, *Zs. f. anorg. und allgemein. Chem.*, **192**, 236, 295, 1930.
[8] Арсеньева и Курчатов, *ЖЭТФ*, **4**, 576, 1934.
[9] Гохберг и Саминский, *ЖЭТФ*, **7**, 1099, 1937.
[10] Jusé und B. Kurtschatow, *Sov. Phys.*, **2**, 454, 1932.
[11] Meyer, *Zs. f. Phys.*, **85**, 278, 1933.
[12] Meyer, *Zs. f. techn. Phys.*, **16**, 355, 1935.
[13] Арсеньева и Бронштейн, *ЖЭТФ*,
[14] Гудден, *УФН*, **15**, 703, 1935.
[15] Иоффе, Электронные полупроводники, ГТТИ, 1933.
[16] Wilson, *Proc. of. R. Soc.*, **A**, **133**, 458, 1931; **134**, 277, 1932.
[17] Nijboer, *Proc. of. R. Soc.*, **A**, **51**, 575, 1939.
[18] Давыдов и Шмушкевич, *УФН*, **24**, 21, 1940.

ГЛАВА V

ВЛИЯНИЕ СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ ПОЛЕЙ НА ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

§ 1. Сводка экспериментальных данных

В предыдущей главе мы рассматривали зависимость электропроводности σ полупроводников от температуры T и параметров μ и ν, характеризующих концентрацию повреждений в решётке. В этой главе мы рассмотрим зависимость электропроводности от другого, также весьма важного фактора: от напряжённости электрического поля E.

Мы будем различать область *слабых* полей, для которой $E < E_0$, и область *сильных* полей, для которой $E > E_0$, где E_0 — некоторая критическая точка.

В слабых полях электропроводность не зависит от поля, т. е. удовлетворяет закону Ома:

$$\sigma = \sigma_0 = \text{const} \quad \text{при } E < E_0. \tag{1,1}$$

В сильных полях электропроводность резко возрастает с полем, подчиняясь закону Пуля:

$$\sigma = \sigma_0 e^{a(E-E_0)} \quad \text{при } E > E_0. \tag{1,2}$$

Зависимость электропроводности от поля в полулогарифмических координатах схематически изображена на рис. 81.

Рассмотрим параметры σ_0, a, E_0, фигурирующие в (1,2). Значения этих параметров зависят не только от природы полупроводника, но и от температуры и от количества содержащейся в решётке примеси:

1) Множитель σ_0 в ограниченном интервале температур может быть представлен в виде:

$$\sigma_0 = Ce^{-\frac{B}{T}}. \tag{1,3}$$

Анализу этого множителя была посвящена вся предыдущая глава. Там же было выяснено происхождение этого множителя.

2) Коэффициент a характеризует наклон кривой электропроводности в области сильных полей (см. рис. 81). Этот коэффициент также существенно зависит от температуры. Чем выше температура, тем меньше a. Таким образом, при повышении температуры вся кривая на рис. 81 сдвигается вверх и делается более пологой. Зависимость a от T для образцов Cu_2O и V_2O_5 была подробно изучена А. В. и А. Ф. Иоффе [1]. Анализируя данные Иоффе, мы получаем следующий вид зависимости a от T:

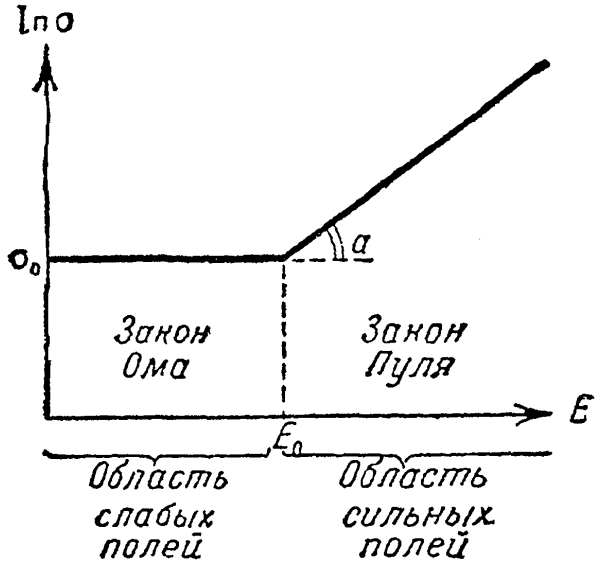


Рис. 81. Зависимость электропроводности σ от напряжённости поля E.

$$a = \frac{\alpha}{T} - \beta, \qquad (1,4)$$

где α и β имеют следующий порядок величины:

$$\begin{cases} \alpha \approx 1 - 10 \dfrac{см\ абс.\ град}{kV}, \\ \beta \approx 0{,}01 - 0{,}1\ \dfrac{см}{kV}. \end{cases} \qquad (1,5)$$

Соответствующие кривые $a\left(\frac{1}{T}\right)$ для Cu_2O и V_2O_5, построенные по данным Иоффе, изображены на рис. 82.

3) Параметр E_0 представляет собой критическую точку, в которой закон Ома сменяется законом Пуля. Положение этой точки также зависит от природы полупроводника, от температуры и концентрации примеси. Чем больше работа ионизации B в формуле (1,3), тем правее на рис. 81 лежит точка E_0. Также E_0 сдвигается вправо по мере увеличения количества примесей, содержащихся в решётке и, повидимому, по мере возрастания температуры. Однако вид зависимости E_0 от всех этих факторов пока ещё нельзя считать прочно уста-

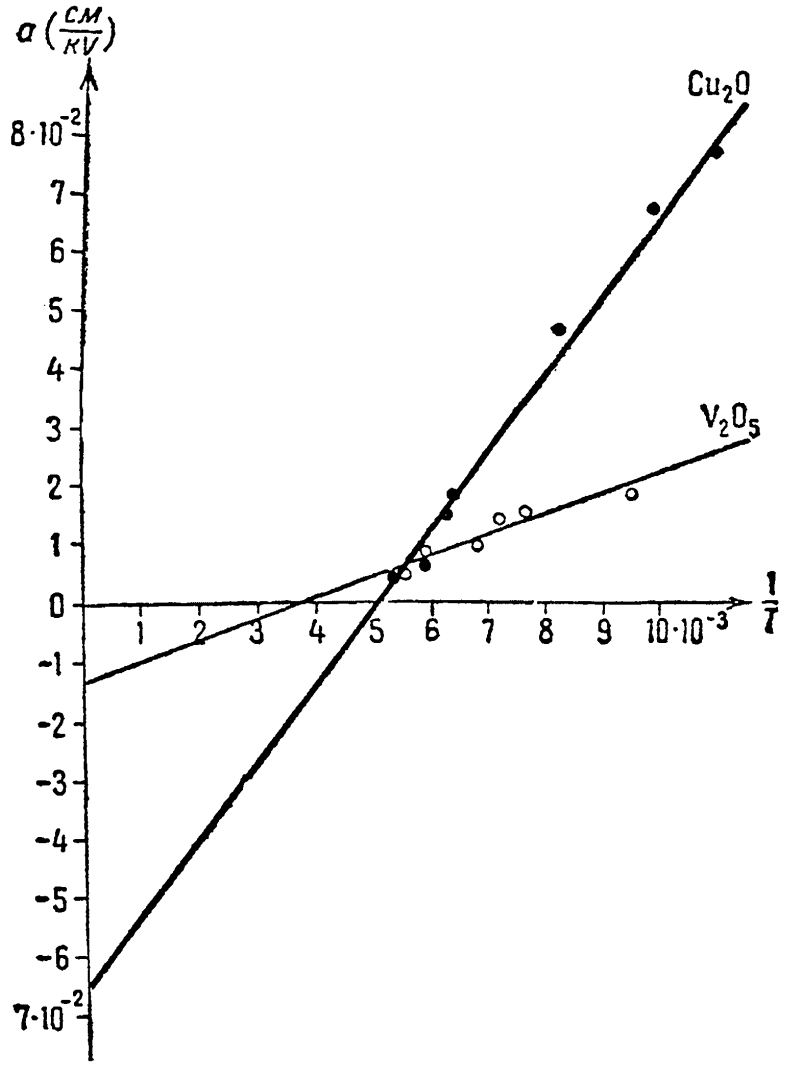


Рис. 82. Зависимость коэффициента α в законе Пуля $\sigma = \sigma_0 e^{\alpha(E-E_0)}$ от температуры (по данным А. В. и А. Ф. Иоффе).

$$\alpha(T) = \frac{\alpha}{T} - \beta$$

	β $см/kV$	α $\frac{см\cdot абс.\ град.}{kV}$
Cu_2O	$6{,}5\cdot10^{-2}$	13,0
V_2O_5	$1{,}3\cdot10^{-2}$	3,5

новленным. По порядку величины имеем

$$E_0 \approx 1 - 10 \frac{\text{kV}}{см}.$$

Заметим, что в некоторых случаях в области сильных полей возрастание электропроводности с полем происходит несколько медленнее, чем это требуется законом Пуля. Это видно из рис. 83, на котором приведены кривые Иоффе для различных полупроводников (см., например, Sb_2S_3, MoS_2). С другой стороны, в области слабых полей в некоторых случаях имеют место отступления от закона Ома (см., например, Cu_2O). Эти отступления, однако, являются результатом вторичных эффектов (непроводящие слои на границе между полупроводником и электродом) и могут быть игнорированы при построении общей теории.

Заметим также, что возрастание проводимости в сильных полях не связано непосредственно с явлением пробоя. Пробой вызывается, повидимому, иным механизмом, вступающим в игру внезапно при значительно более сильных полях. Действительно, экстраполируя закон Пуля к пробивным напряжениям, мы не получаем токов, достаточных для того, чтобы привести к пробою.

Заметим, наконец, что зависимость проводимости от поля, схематически изображённая на рис. 81, имеет общий характер и относится как к полупроводникам, так и к диэлектрикам. Для полупроводников (Cu_2O, V_2O_5, Tl_2S, WO_3, MoS_2, Sb_2S_3) эта зависимость была изучена Иоффе [1]. Для диэлектриков (слюда, NaCl и др.) — Пружининой-Грановской, а также Вальтером и Инге и др. Диэлектрики отличаются от полупроводников тем, что у диэлектриков в области слабых полей (участок Ома) проводимость ионная, в то время как у полупроводников — электронная. В области же сильных полей (участок Пуля) проводимость в обоих случаях имеет электронную природу. Можно считать, что как в полупроводниках, так и в диэлектриках закон Пуля имеет одно и то же происхождение, т. е. вызывается одним и тем же механизмом. Такое заключение вытекает из следующих соображений.

Во всякой решётке проводимость σ, строго говоря, складывается из двух составляющих: электронной $\sigma_{эл}$ и ионной $\sigma_{ион}$, т. е.

$$\sigma = \sigma_{эл} + \sigma_{ион}. \qquad (1,6)$$

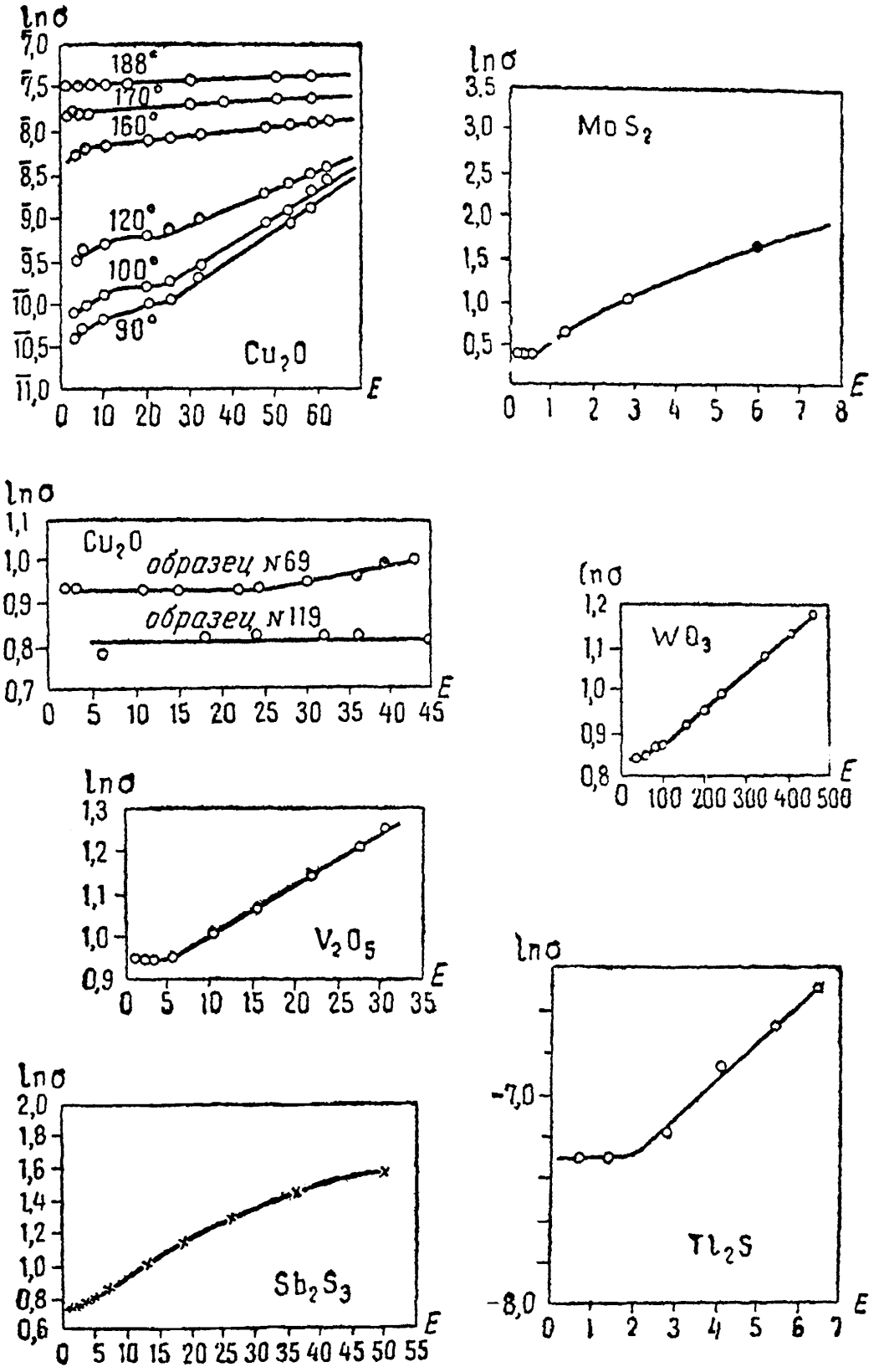


Рис. 83. Зависимость электропроводности от поля для различных полупроводников (по данным Иоффе).
По оси абсцисс отложено E в kV/*см*.

Для электронной составляющей мы имеем, согласно (1,1), (1,2) и (1,3):

$$\sigma_{эл} = \begin{cases} = Ce^{-\frac{B}{T}} \ldots\ldots \text{в слабых полях,} \\ = Ce^{-\frac{B}{T} + a(E - E_0)} \ldots \text{в сильных полях,} \end{cases} \quad (1,7)$$

где, согласно (1,4):

$$a = \frac{\alpha}{T} - \beta.$$

Для ионной составляющей, согласно теории Шоттки (см. главу I) и вычислениям Браунбека [2], в полной мере применимым к этой теории, мы можем написать:

$$\sigma_{ион} = C'e^{-\frac{B'}{T}} \frac{\operatorname{sh} a'E}{E} \ldots \text{для любых полей,} \quad (1,8)$$

где

$$a' = \frac{\alpha'}{T}.$$

Так как по порядку величины $\alpha' \approx 0{,}1 \frac{см\ абс.\ град}{\text{kV}}$, то мы можем разложить гиперболический синус в ряд по степеням E и ограничиться первым членом разложения, вплоть до полей порядка $10^3 - 10^4 \frac{\text{kV}}{см}$. Это даёт нам закон Ома. Таким образом, в области сильных полей, по мере возрастания поля, соотношение между ионной и электронной составляющими быстро изменяется в пользу последней. Если в слабых полях мы можем иметь

$$\text{или} \quad \begin{cases} \sigma_{эл} \gg \sigma_{ион} \ldots\ldots \text{для полупроводников,} \\ \sigma_{ион} \gg \sigma_{эл} \ldots\ldots \text{для диэлектриков,} \end{cases}$$

то при достаточно сильных полях мы имеем всегда

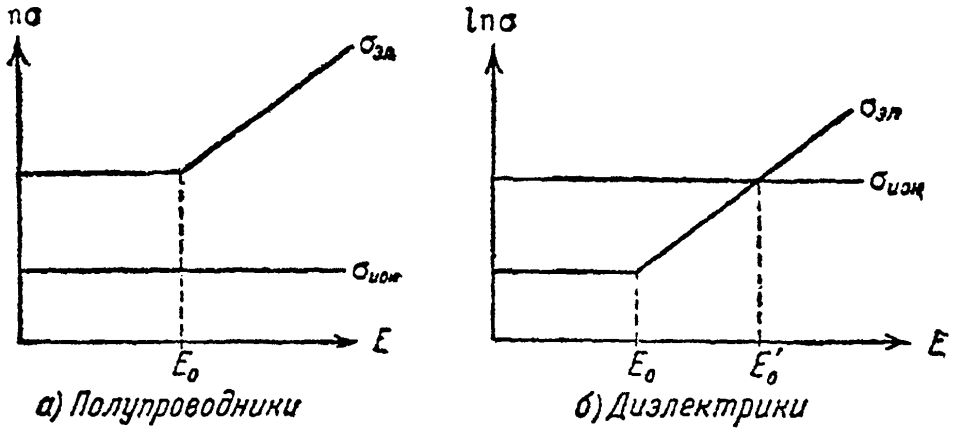
$$\sigma_{эл} \gg \sigma_{ион}.$$

Это поясняется на рис. 84. Естественно, что переход от закона Ома к закону Пуля у диэлектриков происходит при значительно более сильных полях, нежели у полупроводников (точка E_0' на рис. 84, *б*).

Настоящую главу мы посвятим выяснению тех причин, которые вызывают повышение проводимости в сильных полях, иначе говоря, анализу механизмов, обусловливающих появление экспоненциального множителя в формуле (1,2). Поскольку проводимость выражается как произведение двух множителей:

$$\sigma = evn,$$

где v — подвижность, а n — концентрация свободных электронов (дырок), то повышение проводимости может быть



а) Полупроводники б) Диэлектрики

Рис. 84. Электронная и ионная составляющие проводимости как функции поля.

отнесено или за счёт увеличения v или же за счёт увеличения n. Опыты Иоффе совершенно исключают предположение, что рост проводимости обязан повышению подвижности. Также и с теоретической точки зрения нет оснований ожидать, чтобы подвижность сколько-нибудь существенно увеличивалась с полем. Поэтому здесь мы ограничимся рассмотрением того, почему и каким образом под действием поля может увеличиваться число носителей тока n. Теоретически это увеличение может быть вызвано несколькими физически различными механизмами, которые, строго говоря, развиваются параллельно, но преобладание того или иного из них зависит от конкретных условий. Все эти механизмы, рано или поздно вступающие в игру при достаточно сильном поле, обусловлены теми деформациями, которые поле вызывает в структуре энергетического спектра, и тем влиянием, которое оно оказывает на поведение электрона в решётке.

§ 2. Влияние электрического поля на поведение электрона в решётке

А. Движение электрона в фазовом пространстве. Когда внешнее поле отсутствует, электрон в решётке может быть охарактеризован тройкой квантовых чисел λ_1, λ_2, λ_3, которые можно трактовать как компоненты некоторого вектора $\vec{\Lambda}$. Этот вектор определяет волновую функцию электрона ψ, его энергию W и его среднюю скорость $\vec{v}$ (групповую скорость пакета). Если известна W как функция λ_1, λ_2, λ_3, то $\vec{v}$ может быть вычислена по формуле (2,7) главы II.

$$\boxed{\vec{v} = \frac{2\pi a}{h} \operatorname{grad}_{\vec{\Lambda}} W,} \qquad (2,1)$$

где a — постоянная решётки (решётка предполагается кубической). Значок $\vec{\Lambda}$ при градиенте обозначает, что здесь имеется в виду градиент в пространстве чисел λ_1, λ_2, λ_3, которое мы в дальнейшем будем называть фазовым пространством, так что:

$$v_1 = \frac{2\pi a}{h} \frac{\partial W}{\partial \lambda_1}, \quad v_2 = \frac{2\pi a}{h} \frac{\partial W}{\partial \lambda_2}, \quad v_3 = \frac{2\pi a}{h} \frac{\partial W}{\partial \lambda_3}.$$

Предположим для простоты, что дело происходит в s-зоне. Тогда, согласно (2,39) главы II:

$$W = A + B(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3),$$

где

$$A = W^s + a^s, \quad B = 2\beta^s,$$

и, следовательно:

$$v_1 = -\frac{2\pi a B}{h} \sin\lambda_1, \quad v_2 = -\frac{2\pi a B}{h} \sin\lambda_2, \quad v_3 = -\frac{2\pi a B}{h} \sin\lambda_3.$$

Таким образом, при отсутствии поля электрон остаётся неподвижным в фазовом пространстве, а в координатном пространстве движется с постоянной скоростью $\vec{v}$. Под влиянием поля поведение электрона искажается.

Когда на решётку наложено внешнее электрическое поле $\vec{E}$, электрон уже не может быть охарактеризован каким-либо

определённым значением вектора $\vec{\Lambda}$. Если в начальный момент времени (до наложения поля) электрон находился в состоянии $\vec{\Lambda}^0$, то в последующие моменты ему могут быть приписаны определённые значения $\vec{\Lambda}$, однако различные для различных моментов времени и изменяющиеся со временем по закону:

$$\boxed{\vec{\Lambda} = \vec{\Lambda}^0 - \frac{2\pi a e}{h}\vec{E}t.} \qquad (2,2)$$

(Здесь и в дальнейшем e — абсолютная величина заряда электрона.) Это значит, что электрон может быть описан невозмущённой волновой функцией, в которой, однако, $\vec{\Lambda}$ следует теперь считать зависящим от времени, а не постоянным. Убедимся в этом.

Уравнение Шредингера при наличии поля имеет вид:

$$H\Psi = i\frac{h}{2\pi}\frac{\partial}{\partial t}\Psi,$$

где

$$\boxed{H = -\frac{h^2}{8\pi^2 m}\Delta + V(xyz) + e\vec{E}\,\vec{R}.}$$

Здесь $\vec{R}$ — радиус-вектор (компоненты: x, y, z), а $V(x, y, z)$ — периодический потенциал решётки. Решение этого уравнения может быть развёрнуто по собственным функциям уравнения без поля:

$$\Psi(x, y, z, t) = \sum_i \int C^i(\lambda_1, \lambda_2, \lambda_3, t)\,\psi^i(\lambda_1, \lambda_2, \lambda_3, x, y, z)\,d\lambda_1\,d\lambda_2\,d\lambda_3, \qquad (2,3)$$

где функция ψ^i удовлетворяет, следовательно, уравнению

$$H^0\psi^i = W^i\psi^i,$$

где

$$H^0 = -\frac{h^2}{8\pi^2 m}\Delta + V(xyz).$$

Так как функция ψ^i, как показал Блох (см. главу II, § 2, С), есть периодическая функция от параметров $\lambda_1, \lambda_2, \lambda_3$ с периодом 2π, то в (2,3) область интегрирования следует ограничить величиной периода, принимая в качестве пределов, например,

$\lambda_s = -\pi$ и $\lambda_s = +\pi$. По этой же причине, заметим, коэффициенты C^i в (2,3) следует считать периодическими функциями от $\lambda_1, \lambda_2, \lambda_3$ с тем же периодом 2π. Если первоначально на функцию ψ^i было наложено уравнение Кармана-Борна [см. (1,15) главы II], то в (2,3) интеграл должен быть заменён суммой.

Джонс и Зинер [3] показали, что коэффициенты C^i в (2,3) удовлетворяют следующему дифференциальному уравнению:

$$\left\{\frac{h}{2\pi}\frac{\partial}{\partial t} - e\vec{E}\,a\,\mathrm{grad}_{\vec{\Lambda}}\right\}\sum_i |C^i(\lambda_1, \lambda_2, \lambda_3, t)|^2 = 0,$$

откуда вытекает

$$\sum_i |C^i(\lambda_1, \lambda_2, \lambda_3, t)|^2 =$$

$$= f\left(\lambda_1 + \frac{2\pi e a}{h}E_1 t,\ \lambda_2 + \frac{2\pi e a}{h}E_2 t,\ \lambda_3 + \frac{2\pi e a}{h}E_3 t\right), \quad (2,4)$$

где f — произвольная функция своих аргументов. Вид этой функции определяется выбором начальных условий. Действительно, полагая $t = 0$:

$$f(\lambda_1, \lambda_2, \lambda_3) = \sum_i |C^i(\lambda_1, \lambda_2, \lambda_3, 0)|^2,$$

и, следовательно, (2,4) можно переписать так:

$$\sum_i |C^i(\lambda_1, \lambda_2, \lambda_3, t)|^2 =$$

$$= \sum_i \left| C^i\left(\lambda_1 + \frac{2\pi e a}{h}E_1 t, \lambda_2 + \frac{2\pi e a}{h}E_2 t, \lambda_3 + \frac{2\pi e a}{h}E_3 t\right)\right|^2. \quad (2,5)$$

Предположим, что в начальный момент электрон находился в некоторой зоне i', т. е. что $C^i(\lambda_1, \lambda_2, \lambda_3, 0) = 0$ при всех i, кроме $i = i'$, и будем пренебрегать переходами электрона при $t \neq 0$ из данной зоны в другие зоны.

Тогда в (2,3) можно вовсе опустить суммирование по значку i:

$$\Psi(x, y, z, t) =$$

$$= \int C(\lambda_1, \lambda_2, \lambda_3, t)\,\psi(\lambda_1, \lambda_2, \lambda_3, x, y, z)\,d\lambda_1\,d\lambda_2\,d\lambda_3. \quad (2,6)$$

Формула (2,5) примет при этом такой вид:

$$|C(\lambda_1, \lambda_2, \lambda_3, t)|^2 =$$

$$= \left|C\left(\lambda_1 + \frac{2\pi e a}{h}E_1 t, \lambda_2 + \frac{2\pi e a}{h}E_2 t, \lambda_3 + \frac{2\pi e a}{h}E_3 t\right)\right|^2. \quad (2,7)$$

Если предположить, кроме того, что начальное состояние было «чистым», т. е. в момент $t=0$ электрон находился на уровне $\vec{\Lambda}^0$, так что

$$|C(\lambda_1, \lambda_2, \lambda_3, 0)|^2 = \begin{cases} =1 \text{ при } \lambda_1=\lambda_1^0,\ \lambda_2=\lambda_2^0,\ \lambda_3=\lambda_3^0 \\ =0 \text{ в иных случаях,} \end{cases}$$

то тогда, согласно (2,7):

$$|C(\lambda_1, \lambda_2, \lambda_3, t)|^2 = \begin{cases} =1 \text{ при } \begin{cases} \lambda_1+\frac{2\pi ea}{h}E_1 t=\lambda_1^0, \\ \lambda_2+\frac{2\pi ea}{h}E_2 t=\lambda_2^0\lambda_3, \\ \lambda_3+\frac{2\pi ea}{h}Et=\lambda_3^0, \end{cases} \\ =0 \text{ в иных случаях.} \end{cases}$$

При этом выражение (2,6) сведётся к

$$\Psi(x, y, z, t) = C(\lambda_1, \lambda_2, \lambda_3,)\,\psi(\lambda_1, \lambda_2, \lambda_3, x, y, z),$$

где λ_1, λ_2, λ_3 являются функциями времени:

$$\left.\begin{aligned} \lambda_1 &= \lambda_1^0 - \frac{2\pi ea}{h}E_1 t, \\ \lambda_2 &= \lambda_2^0 - \frac{2\pi ea}{h}E_2 t, \\ \lambda_3 &= \lambda_3^0 - \frac{2\pi ea}{h}E_3 t. \end{aligned}\right\}$$

Мы приходим, таким образом, к формуле (2,2). Возмущающее действие поля сводится, следовательно, к тому, что электрон начинает путешествовать по невозмущённым состояниям.

До наложения поля электрон покоился в фазовом пространстве (в точке $\vec{\Lambda}^0$). Теперь же он начинает двигаться с постоянной скоростью

$$\dot{\vec{\Lambda}} = -\frac{2\pi ea}{h}\vec{E}.$$

Заметим, что скорость $\dot{\vec{\Lambda}}$ не зависит от начального состояния электрона, т. е. она одна и та же для всех электронов. Таким образом, вся совокупность электронов, заполняющих некоторый объём фазового пространства, начинает

двигаться как одно целое. Так как в начальный момент времени $t=0$ все электроны обладали различными $\vec{\Lambda}^0$ (если оставить без внимания спин), то они будут обладать различными $\vec{\Lambda}$ и во все последующие моменты времени $t>0$, что видно из (2,2). Следовательно, можно быть уверенным, что возможность столкновений между электронами в фазовом пространстве заведомо исключена. В этом смысле принцип Паули, имевший место до начала движения, сохраняется и в процессе движения.

Заметим, кроме того, что две точки фазового пространства $\vec{\Lambda}$ и $\vec{\Lambda}'$, для которых

$$\vec{\Lambda}'=\vec{\Lambda}+2\pi\vec{G}$$

(где $\vec{G}$ — вектор с целочисленными компонентами g_1, g_2, g_3) соответствуют одному и тому же физическому состоянию [1]), ибо волновая функция электрона, как мы уже отмечали, является периодической функцией от параметров $\lambda_1, \lambda_2, \lambda_3$ с периодом 2π. Таким образом, электрон, двигаясь в фазовом пространстве равномерно и прямолинейно, периодически возвращается к своему исходному состоянию. Если, например, предположить, что поле направлено антипараллельно оси OX так, что

$$E_1=-E, \quad E_2=E_3=0,$$

то

$$\lambda_1=\lambda_1^0+\frac{2\pi ae}{h}Et, \quad \lambda_2=\lambda_2^0, \quad \lambda_3=\lambda_3^0,$$

где $\lambda_1^0, \lambda_2^0, \lambda_3^0$ — координаты электрона в фазовом пространстве в начальный момент времени (т. е. до наложения поля). В этом случае состояния электрона периодически повторяются через одинаковые промежутки времени

$$\tau=\frac{h}{aeE}.$$

Это иллюстрируется рис. 85. На этом рисунке квадраты $ABB'A'$ и $BCC'B'$ физически эквивалентны друг другу. Стрелки обозначают путь, проходимый электроном за время τ. За это время каждый электрон проходит через все возможные со-

[1]) Если, конечно, ограничиваться одной только зоной.

стояния, отличающиеся друг от друга значением λ_1, и вновь возвращается к исходному состоянию.

Таково движение электрона в фазовом пространстве.

В. Движение электрона в координатном пространстве. Рассмотрим теперь движение электрона в координатном пространстве x, y, z. Оно делается весьма своеобразным, коль скоро мы накладываем на решётку внешнее электрическое

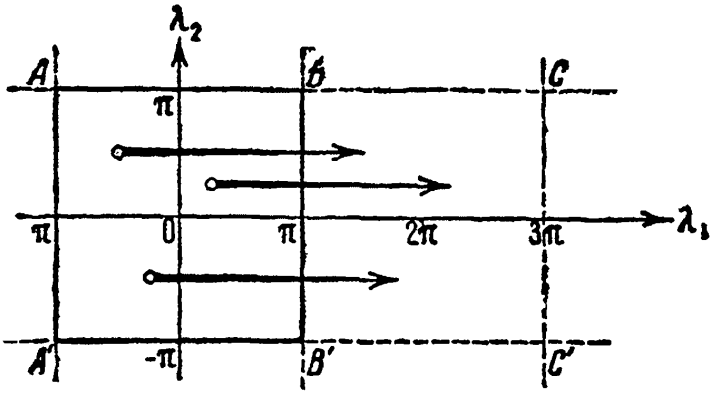


Рис. 85. Движение электрона в фазовом пространстве (поле предполагается направленным антипараллельно оси OX).

поле. Если при отсутствии поля электрон двигался сквозь кристалл с постоянной скоростью $\vec{v}$, то при наличии поля он начинает двигаться с некоторым переменным во времени ускорением. Джонс и Зинер [3] показали, что

$$\boxed{\dot{\vec{v}} = -\frac{4\pi^2 a^2 e}{h^2} \operatorname{grad}_{\vec{\lambda}} \left(\vec{E} \operatorname{grad}_{\vec{\lambda}} W\right).} \qquad (2,8)$$

Эта формула легко может быть получена из (2,1) и (2,2). Действительно, на основании (2,1), мы имеем:

$$\left\{\begin{aligned}
\dot{v}_1 &= \frac{2\pi a}{h}\frac{d}{dt}\left(\frac{\partial W}{\partial \lambda_1}\right) = \frac{2\pi a}{h}\left(\frac{\partial^2 W}{\partial \lambda_1^2}\dot{\lambda}_1 + \frac{\partial^2 W}{\partial \lambda_1 \partial \lambda_2}\dot{\lambda}_2 + \frac{\partial^2 W}{\partial \lambda_1 \partial \lambda_3}\dot{\lambda}_3\right),\\
\dot{v}_2 &= \frac{2\pi a}{h}\frac{d}{dt}\left(\frac{\partial W}{\partial \lambda_2}\right) = \frac{2\pi a}{h}\left(\frac{\partial^2 W}{\partial \lambda_2 \partial \lambda_1}\dot{\lambda}_1 + \frac{\partial^2 W}{\partial \lambda_2^2}\dot{\lambda}_2 + \frac{\partial^2 W}{\partial \lambda_2 \partial \lambda_3}\dot{\lambda}_3\right),\\
\dot{v}_3 &= \frac{2\pi a}{h}\frac{d}{dt}\left(\frac{\partial W}{\partial \lambda_3}\right) = \frac{2\pi a}{h}\left(\frac{\partial^2 W}{\partial \lambda_3 \partial \lambda_1}\dot{\lambda}_1 + \frac{\partial^2 W}{\partial \lambda_2 \partial \lambda_3}\dot{\lambda}_2 + \frac{\partial^2 W}{\partial \lambda_3^2}\dot{\lambda}_3\right),
\end{aligned}\right.$$

откуда на основании (2,2):

$$\left\{\begin{aligned}
-\dot{v}_1 &= \frac{4\pi^2a^2e}{h^2}\left(\frac{\partial^2W}{\partial\lambda_1^2}E_1+\frac{\partial^2W}{\partial\lambda_1\partial\lambda_2}E_2+\frac{\partial^2W}{\partial\lambda_1\partial\lambda_3}E_3\right)= \\
&= \frac{eE_1}{m_{11}}+\frac{eE_2}{m_{12}}+\frac{eE_3}{m_{13}}, \\
-\dot{v}_2 &= \frac{4\pi^2a^2e}{h^2}\left(\frac{\partial^2W}{\partial\lambda_2\partial\lambda_1}E_1+\frac{\partial^2W}{\partial\lambda_2^2}E_2+\frac{\partial^2W}{\partial\lambda_2\partial\lambda_3}E_3\right)= \\
&= \frac{eE_1}{m_{21}}+\frac{eE_2}{m_{22}}+\frac{eE_3}{m_{23}}, \\
-\dot{v}_3 &= \frac{4\pi^2a^2e}{h^2}\left(\frac{\partial^2W}{\partial\lambda_3\partial\lambda_2}E_1+\frac{\partial^2W}{\partial\lambda_3\partial\lambda_2}E_2+\frac{\partial^2W}{\partial\lambda_3^2}\right)= \\
&= \frac{eE_1}{m_{31}}+\frac{eE_2}{m_{32}}+\frac{eE_3}{m_{33}}.
\end{aligned}\right. \tag{2,9}$$

Отсюда вытекает формула (2,8). Здесь

$$\boxed{\frac{1}{m_{sk}} = \frac{4\pi^2a^2}{h^2}\frac{\partial^2W}{\partial\lambda_s\partial\lambda_k} \quad \ldots\ldots \quad s, k = 1, 2, 3} \tag{2,10}$$

— так называемый *тензор обратной массы.*

Если считать поле направленным антипараллельно оси OX так, что

$$E_1 = -E, \quad E_2 = E_3 = 0,$$

и если предположить, что функция $W(\lambda_1, \lambda_2, \lambda_3)$ распадается на сумму трёх членов: $W(\lambda_1, \lambda_2, \lambda_3) = W_1(\lambda_1) + W_2(\lambda_2) + W_3(\lambda_3)$, как это имеет место, например, для s- или p-зоны или для случая свободных электронов, то

$$\frac{1}{m_{sk}} = \begin{cases} = \dfrac{4\pi^2a^2}{h^2}\dfrac{\partial^2W_s}{\partial\lambda_s^2} & \ldots \quad \text{при } s = k, \\ = 0 \ldots\ldots\ldots\ldots & \text{при } s \neq k, \end{cases} \tag{2,11}$$

и, следовательно, формулы (2,9) принимают вид:

$$\begin{cases} \dot{v}_1 = \dfrac{eE}{m_{11}}, \\ \dot{v}_2 = 0, \\ \dot{v}_3 = 0. \end{cases}$$

Таким образом, движение электрона может быть описано классическими законами, если считать, что электрон обладает

эффективной массой $m^* = m_{11}$, зависящей от λ_1, т. е. меняющейся в процессе движения. Эффективная масса m^* сохраняется постоянной и совпадает с истинной массой электрона m только в том случае, когда мы пренебрегаем периодическим полем решётки, ибо в этом случае

$$W(\lambda_1, \lambda_2, \lambda_3) = \frac{h^2}{8\pi^2 m a^2}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2),$$

и, следовательно,

$$m_{11} = m_{22} = m_{33} = m.$$

Ограничимся для определённости рассмотрением s-зоны. Для s-зоны мы имеем:

$$W = A + B(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3), \tag{2,12}$$

и, следовательно, на основании (2,11):

$$\frac{1}{m^*} = \frac{1}{m_{11}} = -\frac{4\pi^2 a^2 B}{h^2}\cos\lambda_1,$$

где

$$\begin{cases} \lambda_1 = \lambda_1^0 + \dfrac{2\pi ea}{h}Et, \\ \lambda_2 = \lambda_2^0, \\ \lambda_3 = \lambda_3^0. \end{cases} \tag{2,13}$$

Таким образом, эффективная масса периодически изменяется со временем. Она положительна, когда электрон находится на дне зоны ($\cos\lambda_1 \approx -1$), и отрицательна под потолком зоны ($\cos\lambda_1 \approx +1$). В верхней половине зоны осуществляется совершенно аномальное, с классической точки зрения, поведение электрона: электрон ускоряется в направлении, прямо противоположном действующей на него силе.

Из (2,1) для s-зоны получаем:

$$\begin{cases} v_1 = -\dfrac{2\pi aB}{h}\sin\left(\lambda_1^0 + \dfrac{2\pi ea}{h}Et\right), \\ v_2 = -\dfrac{2\pi aB}{h}\sin\lambda_2^0, \\ v_3 = -\dfrac{2\pi aB}{h}\sin\lambda_3^0. \end{cases} \tag{2,14}$$

Мы видим, что при наложении поля те две составляющие скорости, которые перпендикулярны к полю, остаются постоянными (такими же, какими они были при отсутствии поля), а

третья составляющая, параллельная полю, делается периодической функцией времени, изменяясь в пределах

$$\text{от } v_1 = \frac{2\pi aB}{h} \text{ до } v_1 = -\frac{2\pi aB}{h}.$$

Отсюда следует, что электрон оказывается запертым в более или менее узкой области кристалла: его траектория лежит между двумя стенками, перпендикулярными к полю. Таким образом, как это ни странно на первый взгляд, наложение поля делает электрон в известном смысле более локализованным, чем это было при отсутствии поля.

Продолжая наше полуклассическое рассмотрение для s-зоны, получаем из (2,14):

$$\begin{cases} x = x_0 + \dfrac{B}{eE}\left\{\cos\left(\lambda_1^0 + \dfrac{2\pi ae}{h}Et\right) - \cos\lambda_1^0\right\}, \\ y = y_0 + v_2 t, \\ z = z_0 + v_3 t, \end{cases} \tag{2,15}$$

где x_0, y_0, z_0 — координаты электрона в начальный момент времени. Движение электрона вполне определено, коль скоро известно E и начальные условия, т. е. x_0, y_0, z_0 и λ_1^0, λ_2^0, λ_3^0. Прямолинейная траектория, имевшая место при отсутствии поля, превращается теперь в косинусоиду, изображённую на рис. 86. Плоскость траектории проходит через точку x_0, y_0, z_0; она всегда перпендикулярна к плоскости YOZ и составляет с положительным направлением оси OY угол α:

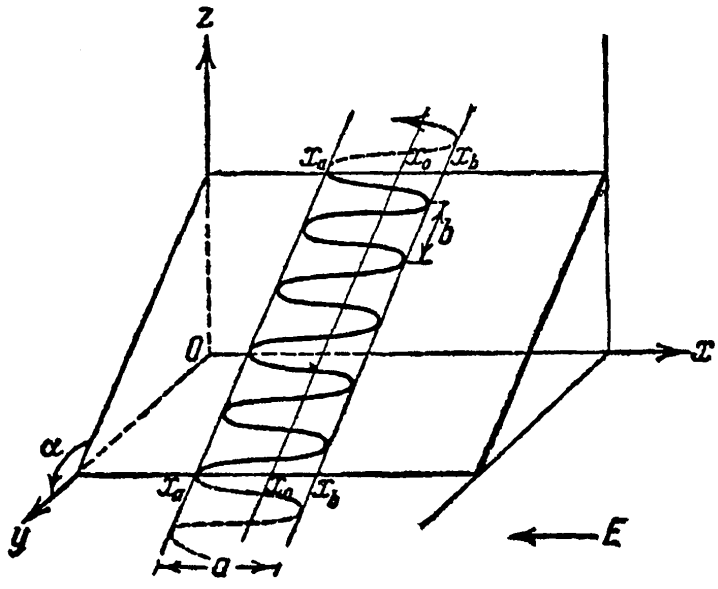


Рис. 86. Движение электрона в координатном пространстве.

$$\operatorname{tg}\alpha = \frac{v_3}{v_2} = \frac{\sin\lambda_3}{\sin\lambda_2}.$$

Тем самым плоскость траектории вполне определена.

Траектория зажата между двумя стенками, параллельными плоскости YOZ (см. рис. 86):

$$x_a = x_0 - \frac{B}{eE}\cos\lambda_1^0 - \frac{B}{eE}, \quad x_b = x_0 - \frac{B}{eE}\cos\lambda_1^0 + \frac{B}{eE}; \quad (2,16)$$

расстояние a между стенками зависит только от E:

$$a = x_b - x_a = \frac{2B}{eE}. \quad (2,17)$$

Величина a определяет ширину области, внутри которой заперт электрон. Как видно, эта область тем уже, чем больше E. Положение этой области относительно x_0 зависит, как это видно из (2,16), от того, каково абсолютное значение λ_1^0: по мере возрастания $|\lambda_1^0|$ вся область передвигается слева направо. Если $\lambda_1^0 = 0$, то точка x_0, y_0, z_0 лежит на правой стенке (т. е. $x_0 = x_b$); если $\lambda_1^0 = \pm\frac{\pi}{2}$, то она расположена как раз посредине между стенками $\left(\text{т. е. } x_0 = \frac{x_b - x_a}{2}\right)$; если же $\lambda_1^0 = \pm\pi$ — она попадает на левую стенку (т. е. $x_0 = x_b$).

Двигаясь по косинусоиде, электрон испытывает отражения от правой и левой стенок через одинаковые промежутки времени τ:

$$\tau = \frac{h}{aeE}.$$

Расстояние b между точками, соответствующими последовательным отражениям (см. рис. 86), равно

$$b = \frac{2\pi B}{eE}\sqrt{\sin^2\lambda_2^0 + \sin^2\lambda_3^0}.$$

Таким образом, косинусоида тем более растянута, чем больше по абсолютной величине λ_2^0 и λ_3^0. Если $\lambda_2^0 = \lambda_3^0 = 0$, то косинусоида вырождается в отрезок прямой.

Если переменить знак одновременно у λ_1^0, λ_2^0, λ_3^0, то электрон будет двигаться по той же траектории, но в обратном направлении. Если переменить знак только у λ_2^0 или только у λ_3^0, мы будем иметь ту же траекторию, но расположенную в плоскости, перпендикулярной к первоначальной. Если переменить знак только у λ_1^0 — это будет та же траектория, ле-

жащая в той же плоскости и между теми же стенками x_a и x_b, но несколько сдвинутая параллельно самой себе.

Итак, коль скоро на решётку накладывается внешнее поле, электрон попадает в ловушку. Двигаясь против поля, он затормаживается полем. Таково происхождение левой стенки на рис. 86. Двигаясь по направлению поля, он затормаживается решёткой. Это затормаживание в конечном счёте вызвано наличием периодического потенциала решётки, обусловливающего отрицательные значения эффективной массы. Действительно, если кристалл рассматривать как потенциальную яму с гладким дном (случай «свободных» электронов), этот эффект не имеет места. Таково происхождение правой стенки. Электрон, запертый в этой ловушке, начинает совершать гармонические колебания, параллельные полю, и в то же время сносится в сторону, в направлении, перпендикулярном к полю.

Таково поведение электрона в координатном пространстве, описанное в терминах классической картины. Это поведение представляется парадоксальным. Такое впечатление создаётся от того, что здесь мы имеем дело в сущности с квантово-механическими результатами, всего лишь грубо переведёнными на классический язык.

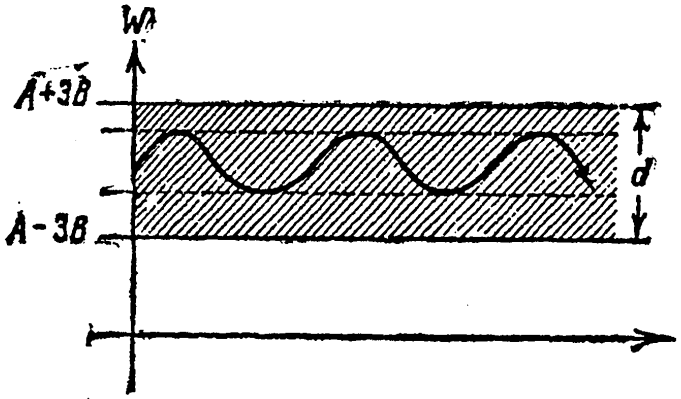


Рис. 87. Движение электрона по уровням энергетической зоны.

С. Движение электрона по уровням зоны. Рассмотрим теперь, как изображается поведение электрона на энергетической диаграмме. Электрон, связанный, при отсутствии поля, с определённым уровнем энергетической зоны, под влиянием поля начинает странствовать по зоне, как это вытекает непосредственно из (2,12) и (2,13). Движение электрона по уровням зоны изображено на рис. 87. Чтобы вернуться в рамки классической картины, достаточно в выражении для энергии (2,12) ввести дополнительный член, выражающий потенциальную энергию электрона во внешнем поле. Тогда вместо (2,12) мы должны будем написать

$$W = A + B(\cos\lambda_1 + \cos\lambda_2 + \cos\lambda_3) - eEx, \quad (2,18)$$

где λ_1, λ_2, λ_3 имеют вид (2,13). Заметим, что при замене выражения (2,12) выражением (2,18) основные формулы (2,1) и (2,8) сохраняют силу. Заметим также, что выражение (2,18), в отличие от (2,12), обеспечивает постоянство энергии. Действительно, на основании (2,1) и (2,2):

$$\frac{dW}{dt} = \dot{\vec{\Lambda}} \operatorname{grad}_{\vec{\Lambda}} W + e\vec{E}\vec{v} = 0.$$

Выражение (2,18) приводит нас к представлению о *наклонной* зоне. Как и при отсутствии поля, мы будем называть зоной совокупность таких значений W, которые соответствуют вещественным значениям λ_1, λ_2, λ_3. При отсутствии поля мы имели границы зоны:

$$\begin{cases} W_{\min} = A - 3B, \\ W_{\max} = A + 3B, \end{cases}$$

ширина зоны:

$$d = W_{\max} - W_{\min} = 6B.$$

При ниличии поля условие вещественности параметров λ_1, λ_2, λ_3 ещё не накладывает ограничения на W, поскольку остаётся

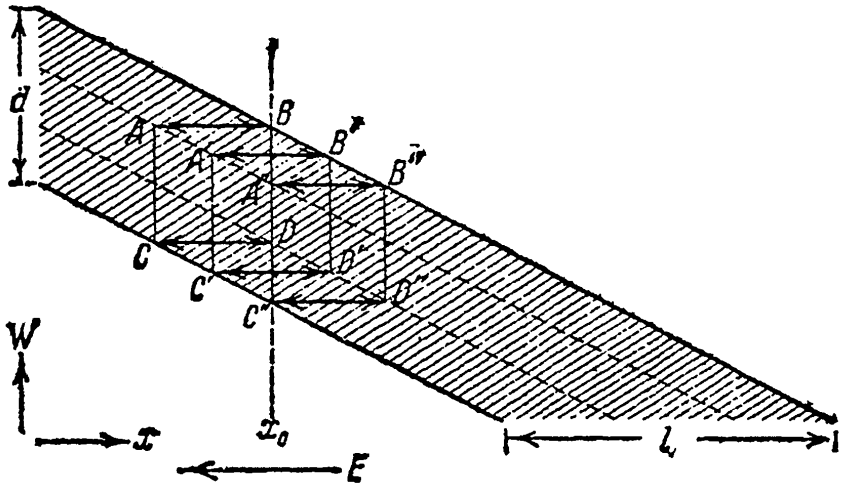


Рис. 88. Наклонная зона. Отрезки AB, $A'B'$ изображают различные траектории электрона (при заданном x).

свободным x, как это видно из (2,18). Таким образом, понятие зоны, в первоначальном смысле, сохраняется лишь в применении к определённому x. На рис. 88 изображена наклонная зона; она представляет собой такую область значений W

и x, для которых $\lambda_1, \lambda_2, \lambda_3$ вещественны (эта область заштрихована на рисунке). Мы будем иметь теперь:

	В вертикальном направлении, т. е. при $x = \text{const}$:	В горизонтальном направлении, т. е. при $W = \text{const}$:
Границы зоны.	$\begin{cases} W_{\max} = A + 3B - eEx \\ W_{\min} = A - 3B - eEx \end{cases}$	$x_{\max} = \frac{1}{eE}(A + 3B - W)$ $x_{\min} = \frac{1}{eE}(A - 3B - W)$
Ширина зоны.	$d = W_{\max} - W_{\min} = 6B$	$l = x_{\max} - x_{\min} = \frac{6B}{eE}$

Итак, при отсутствии поля электроны в решётке обладали лишь определёнными значениями энергии, ограниченными интервалом (2,19), но зато они могли двигаться по всему кристаллу. При наличии поля, наоборот, электроны могут обладать какими угодно значениями энергии, но зато каждый из них ограничен в своём движении более или менее узкой областью кристалла.

Каждый электрон колеблется внутри зоны между границами x_a и x_b, которые определяются из (2,16). На рис. 88 его движение изображается прямолинейным горизонтальным отрезком, который всеми своими точками лежит внутри зоны (это есть проекция траектории на ось OX). Длина этого отрезка равна

$$a = x_b - x_a = \frac{1}{3} l;$$

его положение внутри зоны зависит от абсолютных значений $\lambda_1^0, \lambda_2^0, \lambda_3^0$ и от x_0. При заданном x_0 границы x_a и x_b зависят только от $|\lambda_1^0|$, причём всегда $x_a \leqslant x_0 \leqslant x_b$. Если же заданы x_0 и λ_1^0, т. е. границы x_a и x_b зафиксированы, то отрезок, изображающий движение электрона, будет лежать выше или ниже, в зависимости от значений $|\lambda_2^0|$ и $|\lambda_3^0|$. Так, например, если $\lambda_1^0 = 0$, то он лежит где-то внутри прямоугольника $ABDC$ и тем ниже, чем больше $|\lambda_2^0|$ и $|\lambda_3^0|$.

Отрезки AB, $A'B$, ... изображают различные траектории электрона (при заданном x_0).

Мы имеем в качестве траектории:

	$\lambda_1^0 = 0$	$\lambda_1^0 = \pm \frac{\pi}{2}$	$\lambda_1^0 = \pm \pi$
$\lambda_2^0 = \lambda_3^0 = 0$ $\lambda_2^0 = \lambda_3^0 = \pm \pi$	AB CD	$A'B'$ $C'D'$	$A''B''$ $C''D''$

Таким образом, электрон, колеблясь внутри зоны, охватывает не всю зону, а лишь треть её. Не всякий электрон достигает правой (верхней) границы зоны: это делают только те из них, для которых $\lambda_2^0 = \lambda_3^0 = 0$ (при любом λ_1^0); зато они не достигают её левой (нижней) границы. Левой границы зоны достигают только те электроны, для которых $\lambda_2^0 = \lambda_3^0 = \pm \pi$ (при любом λ_1^0); зато они не достигают её правой границы. Все остальные электроны, колеблясь внутри зоны, отражаются, не доходя до её границ. Это обстоятельство отличает трёхмерную задачу от одномерной. В одномерной решётке, исследованной Зинером [4], электрон проходит всю зону насквозь; в этом случае амплитуда его колебаний как раз равна ширине зоны ($a = l$).

К представлению о наклонных зонах мы вновь вернёмся в следующем параграфе, в котором мы получим те же результаты, но на основании иных, более строгих соображений.

§ 3. Деформация энергетических зон под влиянием электрического поля

А. Постановка задачи. На рис. 89, *а* изображена кривая потенциальной энергии для электрона в решётке при отсутствии внешнего поля. При наложении поля кривая приобретает вид, изображённый на рис. 89, *б*. Вы видим, что поле вызывает два эффекта:

а) оно приподнимает потенциальный кратер каждого атома (так, $\vec{G}$-й атом оказывается поднятым на величину $ea\vec{E}\vec{G}$);

б) кроме того, оно возмущает каждый атом, т. е. искажает, форму этого кратера.

Первый эффект приводит к *наклонению* зон. Предварительное представление об этом эффекте было дано в предыдущем параграфе.

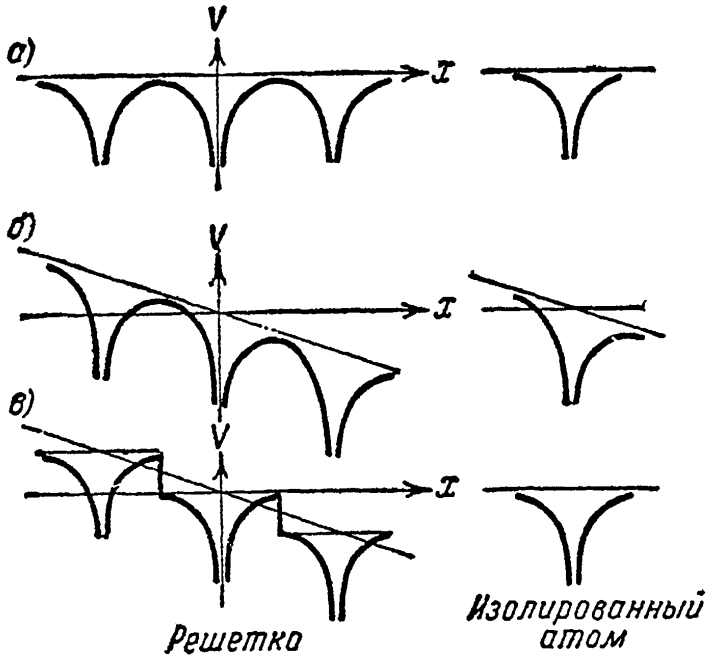


Рис. 89. Кривая потенциальной энергии для электрона в решётке:
а) при отсутствии поля;
б) точная кривая при наличии поля;
в) апроксимированная кривая при наличии поля.

Второй эффект приводит к *расширению* зон, обусловленному штарковским расщеплением термов изолированного атома.

Мы рассмотрим здесь оба эффекта.

Представление о *наклонных* зонах содержится уже в самом виде волновой функции, описывающем поведение электрона в поле, совершенно так же, как представление о *горизонтальных* зонах уже содержится в волновой функции, описывающей поведение электрона при отсутствии поля.

Пусть мы имеем неограниченную кубическую решётку. Пусть a — постоянная решётки. При отсутствии внешнего поля уравнение Шредингера для электрона имеет вид:

$$H^0\psi = W\psi, \tag{3,1}$$

где

$$H^0 = -\frac{h^2}{8\pi^2 m}\Delta - V(x, y, z).$$

Здесь $V(xyz)$ — периодический потенциал. Уравнение (3,1) имеет решения вида:

$$\psi^{l}_{\vec{\Lambda}}(x, y, z) = e^{\frac{l}{a}\vec{\Lambda}\vec{R}}\, u^{l}_{\vec{\Lambda}}(x, y, z), \tag{3,2}$$

где $\vec{R}$ — вектор с компонентами x, y, z; $\vec{\Lambda} = \vec{\Lambda}(W)$ — вектор с компонентами $\lambda_1 = \lambda_1(W)$, $\lambda_2 = \lambda_2(W)$, $\lambda_3 = \lambda_3(W)$, где $\lambda_1, \lambda_2, \lambda_3$ — прозвольные, вообще говоря, комплексные параметры

$u^{l}_{\vec{\Lambda}}(x, y, z)$ — периодическая функция от координат с периодом решётки.

Значения W, соответствующие вещественному $\vec{\Lambda}$, мы называем «разрешёнными», а соответствующие комплексному $\vec{\Lambda}$ — «запрещёнными». Первые образуют энергетические зоны. Из (3,2) видно, что для «разрешённых» W модуль функции ψ периодичен с периодом решётки. Для «запрещённых» же W модуль функции ψ, оставаясь модулированным с периодом решётки, обращается в бесконечность на бесконечности, и таким образом волновые функции, соответствующие «запрещённым» значениям W, не имеют физического смысла.

Заметим, что если (как это обычно делают) на функцию ψ наложить условия Кармана-Борна, то тем самым исключаются из рассмотрения все решения (3,2) с комплексными $\vec{\Lambda}$.

Представим себе теперь, что на нашу решётку наложено внешнее электрическое поле $\vec{E}$. Уравнение Шредингера будет иметь вид:

$$H\Psi = W\Psi, \qquad (3,3)$$

где

$$H = -\frac{h^2}{8\pi^2 m}\Delta + V(x,y,z) + e\vec{E}\vec{R}.$$

Если eEa мало (т. е. если потенциал внешнего поля мало изменяется на протяжении одного периода функции V), то, как показал Хилл [5], уравнение (3,3) имеет решения вида:

$$\Psi^{l}_{\vec{\Lambda}}(x, y, z) = e^{a\int^{\vec{\Lambda}d\vec{R}}} u^{l}_{\vec{\Lambda}}(x, y, z), \qquad (3,4)$$

где теперь $\vec{\Lambda}$ является такой же функцией от $(W - e\vec{E}\vec{R})$, какой раньше (при $E = 0$) оно было только от W. Функция $u^{l}_{\vec{\Lambda}}(x, y, z)$ есть периодическая функция от координат, как и в (3,2), с той лишь разницей, что теперь сам индекс $\vec{\Lambda}$ зависит от $\vec{R}$. При $E = 0$ выражение (3,4) превращается в выражение (3,2).

Представление о «наклонных» зонах получается уже из рассмотрения решения (3,4), так же как представление о «го-

ризонтальных» зонах при отсутствии поля получалось уже из рассмотрения решения (3,2). Мы попрежнему будем называть i-й зоной совокупность тех значений W, которые соответствуют вещественным $\vec{\Lambda}$ в (3,4). Однако теперь $\vec{\Lambda}$ является уже функцией двух независимых аргументов: W и $\vec{R}$. Поэтому условие вещественности $\vec{\Lambda}$ уже не накладывает ограничений на W, поскольку остаётся свободным второе переменное $\vec{R}$. Понятие зон в первоначальном смысле сохраняется здесь лишь для заданного $\vec{R}$.

Функция Ψ (3,4) для заданного произвольного W имеет ту особенность, что для неё существуют такие области пространства, внутри которых $\vec{\Lambda}$ комплексно, и такие, внутри которых $\vec{\Lambda}$ вещественно. В областях с вещественными $\vec{\Lambda}$ модуль функции Ψ периодичен с периодом решётки; в областях с комплексным $\vec{\Lambda}$ модуль функции Ψ быстро затухает, так что в этих областях его можно считать практически равным нулю. Таким образом, электрон с заданной энергией теперь можно считать запертым в более или менее узкой области пространства.

Мы приходим к схеме, изображённой на рис. 88. Эта схема отнюдь не выражает собой зависимости энергии от координат. Она должна восприниматься лишь как характеристика функции Ψ.

Решение уравнения (3,3), написанное в форме (3,4), не позволяет обнаружить эффект расширения зон, происходящий под влиянием поля. Действительно, приближение, заключённое в решении Хилла (eEa мало), равносильно предположению, что потенциал внешнего поля мало меняется в пределах одного атома, так что в области каждого атома его можно считать постоянным. Это означает, что истинная кривая потенциальной энергии (рис. 89, *б*) заменяется ступенчатой апроксимированной кривой (рис. 89, *в*). Такая апроксимация, очевидно, есть пренебрежение эффектом Штарка.

Мы учтём эффект расширения зон, подойдя к решению уравнения (3,3) с иной стороны. Будем отыскивать это решение в виде линейной комбинации атомных функций, соответ-

ствующих изолированным атомам, возмущённым внешним полем [6]:

$$\boxed{\Psi(W, xyz) = \sum_i \sum_{\vec{G}} a^i_{\vec{G}}(W)\, \vartheta^i_{\vec{G}}(xyz).} \qquad (3,5)$$

Здесь $\vartheta^i_{\vec{G}}$ — i-я собственная функция $\vec{G}$-го изолированного атома, удовлетворяющая, следовательно, уравнению:

$$H_{\vec{G}}\, \vartheta^i_{\vec{G}} = W^i_{\vec{G}}\, \vartheta^i_{\vec{G}}, \qquad (3,6)$$

где

$$\begin{cases} H_{\vec{G}} = -\dfrac{h^2}{8\pi^2 m}\Delta - U_{\vec{G}}(xyz) + e\vec{E}\vec{R}, \\ W^i_{\vec{G}} = W^i + \Delta W^i + ea\vec{E}\vec{G}. \end{cases} \qquad (3,7)$$

Здесь

W^i — соответствующий терм изолированного атома при отсутствии поля,

ΔW^i — штарковский сдвиг i-го терма,

$U_{\vec{G}}$ — потенциал $\vec{G}$-го атома.

Если внешнее поле отсутствует ($\vec{E} = 0$), то в (3,5), (3,6) и (3,7) мы будем иметь (в соответствии с обозначениями, принятыми в главе II, § 2, С):

$$\begin{cases} \vartheta^i_{\vec{G}} = \varphi^i_{\vec{G}}, \\ \Delta W^i = 0, \\ W^i_{\vec{G}} = W^i. \end{cases}$$

При этом функция (3,5) превращается в блоховскую функцию [см. формулу (2,32) главы II, § 2, С].

Если же поле отлично от нуля, но достаточно слабо (так, что внутри каждого атома внешний потенциал можно считать постоянным), то в этом случае в (3,5), (3,6) и (3,7) можно положить:

$$\begin{cases} \vartheta^i_{\vec{G}} \approx \varphi^i_{\vec{G}}, \\ e\vec{E}\vec{R} \approx ea\vec{E}\vec{G}, \\ \Delta W^i = 0; \end{cases} \qquad (3,8)$$

при этом, однако, коэффициенты $a^{i}_{\vec{G}}$ отнюдь ещё не обращаются в блоховские.

Заметим, что функция (3,5) предполагается построенной из функций $\vartheta^{i}_{\vec{G}}$, обладающих *близкими* собственными значениями. Таким образом, суммирование в (3,5) проводится по тем значениям $\vec{G}$ и i, для которых

$$\left| W^{i}_{\vec{G}} - W^{i'}_{\vec{G}'} \right| \leqslant \left| W - W^{i'}_{\vec{G}'} \right|.$$

Так как собственные значения $W^{i}_{\vec{G}}$ зависят не только от значка i, но и от значка $\vec{G}$, то в (3,5) суммирование по $\vec{G}$ должно охватывать не всю решётку, как это было бы при отсутствии поля, а лишь ограниченную область пространства. Будем считать, что суммирование по $\vec{G}$ в (3,5) распространено вокруг некоторой точки $\vec{G} = \vec{N}$, т. е. вокруг $\vec{N}$-го атома. Заметим далее, что если термы $W^{i}_{\vec{G}}$, $W^{i'}_{\vec{G}}$, $W^{i''}_{\vec{G}}$, ... достаточно удалены друг от друга, то суммирование по значку i в (3,5) можно вовсе опустить.

Задавая решения уравнения (3,3) в виде (3,5) и считая при этом атомные функции ϑ^{i} известными, мы сводим задачу к отысканию коэффициентов $a^{i}_{\vec{G}}$. Мы будем искать коэффициенты $a^{i}_{\vec{G}}$ в следующем виде:

$$a^{i}_{\vec{G}}(W) = c^{i}(W)\, b_{\vec{G}}(W).$$

Коэффициенты $b_{\vec{G}}$ могут быть определены из общих свойств уравнения (3,3) (см. об этом ниже). Что же касается коэффициентов c^{i}, то мы их найдём из вариационного принципа. Для этого будем искать минимум интеграла

$$I = \int \overline{\Psi} (H - W) \Psi \, d\tau,$$

распространённого по всему тому объёму, внутри которого наша функция Ψ считается нормированной. Условие минимума

даёт нам систему уравнений

$$\frac{\partial I}{\partial \bar{c}^{l'}} = 0,$$

из которой и определяются все интересующие нас коэффициенты c^l. Эта система в развёрнутом виде может быть записана так:

$$\frac{\partial I}{\partial \bar{c}^{l'}} = \sum_l c^l \sum_{\vec{G}} \sum_{\vec{G}'} \bar{b}_{\vec{G}'} b_{\vec{G}} \left\{ (W^l_{\vec{G}} - W) A^{l' l}_{\vec{G}'\vec{G}} + B^{l' l}_{\vec{G}'\vec{G}} \right\} = 0, \quad (3,9)$$

где

$$\begin{cases} A^{l' l}_{\vec{G}'\vec{G}} = \int \bar{\vartheta}^{l'}_{\vec{G}} \vartheta^l_{\vec{G}} d\tau \approx \int \bar{\varphi}^{l'}_{\vec{G}} \varphi^l_{\vec{G}} d\tau, \\ B^{l' l}_{\vec{G}'\vec{G}} = \int \bar{\vartheta}^{l'}_{\vec{G}'} (V - U_{\vec{G}}) \vartheta^l_{\vec{G}} d\tau \approx \int \bar{\varphi}^{l'}_{\vec{G}'} (V - U_{\vec{G}}) \varphi^l_{\vec{G}} d\tau. \end{cases}$$

Условия совместности уравнений (3,9) дадут нам систему интересующих нас собственных значений W уравнения (3,3). При этом обнаруживаются оба эффекта, которые производит внешнее электрическое поле в энергетическом спектре электрона: наклонение и расширение зон.

Мы решим задачу для одномерного случая, а затем обобщим результаты на случай трёх измерений.

В. Решение задачи для одномерного случая. Рассмотрим одномерный случай, когда наша решётка превращается просто в цепочку атомов. Будем считать, что атомы расположены по оси x и внешнее поле E направлено по положительному направлению этой оси. Формулы (3,3), (3,5) и (3,9) в нашем случае могут быть переписаны так:

$$H\Psi = W\Psi, \quad (3,10)$$

где теперь

$$H = -\frac{h^2}{8\pi^2 m} \Delta + V(x, y, z) + eEx,$$

$$\Psi(W, xyz) = \sum_l c^l(W) \sum_g b_g(W) \vartheta^l_g(xyz); \quad (3,11)$$

$$\sum_l c^l \sum_g \sum_{g'} \bar{b}_{g'} b_g \left\{ (W^l_g - W) A^{l'l}_{g'g} + B^{l'l}_{g'g} \right\} = 0. \quad (3,12)$$

Займёмся отысканием коэффициентов b_g и c^l.

Прежде всего покажем, что два соседних коэффициента b_g в (3,11) связаны между собой следующим образом:

$$\boxed{\begin{array}{l} b_{g+1}(W) = b_g(W) e^{i\xi_1 (W - eaE_g)}, \\ b_{g-1}(W) = b_g(W) e^{i\xi_2 (W - eaE_g)}, \end{array}} \quad \begin{array}{l} (3,13) \\ (3,14) \end{array}$$

или

где ξ_1 и ξ_2 следует считать, вообще говоря, комплексными:

$$\begin{cases} \xi_1 = \xi_1' + i\xi_1'', \\ \xi_2 = \xi_2' + i\xi_2''. \end{cases} \tag{3,15}$$

Соотношения (3,13) и (3,14) могут быть получены из общих свойств функции Ψ. Действительно, если функция (3,11) удовлетворяет уравнению (3,10), то этому же уравнению при том же W будет также удовлетворять и такая функция:

$$\Psi(W - eaEh, x - ha, y, z) =$$
$$= \sum_i c^i(W - eaEh) \sum_g b_g(W - eaEh)\, \vartheta_g^i(x - ha, y, z), \tag{3,16}$$

где h — любое. Поэтому, если функция Ψ не вырождена, то можно написать:

$$\Psi(W, x, y, z) = C\Psi(W - eaEh, x - ha, y, z). \tag{3,17}$$

Если же Ψ вырождена, можно всегда исходные функции выбрать так, чтобы соотношение (3,17) имело место для каждой из них. Подставляя в левую часть равенства (3,17) выражение (3,11), а в правую часть — выражение (3,16) и принимая во внимание, что

$$\vartheta_g^i(x - ha, y, z) = \vartheta_{g+h}^i(x, y, z),$$

мы получим, сравнивая коэффициенты при ϑ_g^i в левой и в правой частях:

$$c^i(W)\, b_g(W) = Cc^i(W - eaEh)\, b_{g-h}(W - eaEh).$$

Это имеет место для любого g. Отсюда:

$$\frac{b_{g'}(W)}{b_g(W)} = \frac{b_{g'-h}(W - eaEh)}{b_{g-h}(W - eaEh)}. \tag{3,18}$$

Полагая здесь $g' = g + 1$ и $h = g$, мы получим:

$$\frac{b_{g+1}(W)}{b_g(W)} = \frac{b_{+1}(W - eaEg)}{b_0(W - eaEg)}.$$

Полагая в (3,18) $g'=g-1$ и $h=g$:

$$\frac{b_{g-1}(W)}{b_g(W)}=\frac{b_{-1}(W-eaEg)}{b_0(W-eaEg)}.$$

Принимая обозначения:

$$\frac{b_{+1}(W)}{b_0(W)}=e^{i\xi_1(W)} \quad \text{и} \quad \frac{b_{-1}(W)}{b_0(W)}=e^{i\xi_2(W)},$$

мы приходим к соотношениям (3,13) и (3,14).

Рекуррентные формулы (3,13) и (3,14) позволяют выражать любой коэффициент b_g через ξ_1 или ξ_2. Из (3,13) мы имеем:

$$b_g=b_g^1(\xi_1)=\begin{cases}=b_0\prod\limits_{n=0}^{n=g-1}e^{i\xi_1(W-eaEn)}\ldots\text{для } g>0,\\ =b_0\prod\limits_{n=-1}^{n=0}e^{-i\xi_1(W-eaEn)}\ldots\text{для } g<0.\end{cases}\tag{3,19}$$

Из (3,14) мы имеем:

$$b_g=b_g^2(\xi_2)=\begin{cases}=b_0\prod\limits_{n=1}^{n=g}e^{-i\xi_2(W-eaEn)}\ldots\text{для } g>0,\\ =b_0\prod\limits_{n=0}^{n=g+1}e^{i\xi_2(W-eaEn)}\ldots\text{для } g<0.\end{cases}\tag{3,20}$$

В предельном случае при $E=0$ эти формулы превращаются в такие:

$$\left.\begin{aligned}b_g^1(\xi_1)&=b_0e^{i\xi_1 g},\\ b_g^2(\xi_2)&=b_0e^{-i\xi_2 g}.\end{aligned}\right\}\ldots\text{для } g\gtrless 0.$$

Соответственно (3,19) и (3,20) мы получаем две собственные функции:

$$\Psi_1(\xi_1, xyz)=\sum_i c^i\sum_g b_g^1(\xi_1)\,\vartheta_g^i(xyz),\tag{3,21}$$

$$\Psi_2(\xi_2, xyz)=\sum_i c^i\sum_g b_g^2(\xi_2)\,\vartheta_g^i(xyz).\tag{3,22}$$

Обратимся теперь к коэффициентам c^i. Ограничимся рассмотрением изолированного s-состояния. В этом случае суммирование по значку i в (3,11) и в (3,12) можно вовсе опустить (если при этом суммирование по значку g проводить лишь в ограниченной области решётки). Система уравнений

(3,12) сводится в этом случае к одному уравнению. Принимая для коэффициентов $A^s_{g'g}$ и $B^s_{g'g}$ обозначения Блоха [см. (2.36) главы II], мы можем переписать (3,12) так:

$$c^s\{\sum_g b_g\bar{b}_g(W^s_g - W + \alpha^s) + \\ + \beta^s(\sum_g b_g\overline{b_{g+1}} + \sum_g b_g\overline{b_{g-1}})\} = 0. \quad (3,23)$$

Так как заведомо $c^s \neq 0$, то следует приравнять нулю выражение в фигурных скобках. Заметим, что, согласно (3,13), (3,14) и (3,15):

$$\sum_g b_g\bar{b}_{g+1} + \sum_g b_g\bar{b}_{g-1} = \left\{\begin{array}{l}\sum\limits_g \bar{b}_g b_{g+1} + \sum\limits_g \bar{b}_g b_{g+1} \\ \sum\limits_g b_g\bar{b}_{g-1} + \sum\limits_g \bar{b}_g b_{g-1}\end{array}\right\} = \\ = \left\{\begin{array}{l}= 2\sum\limits_g b_g\bar{b}_g e^{-\xi_1(W-eaEg)}\cos\xi'_1(W - eaEg), \\ = 2\sum\limits_g b_g\bar{b}_g e^{-\xi_2(W-eaEg)}\cos\xi'_2(W - eaEg).\end{array}\right\} \quad (3,24)$$

Заметим, кроме того, что

$$W^s_g = W^s + \Delta W^s + eaEg. \quad (3,25)$$

Подставляя (3,24) и (3,25) в (3,23) и приравнивая нулю коэффициенты при $b_g\bar{b}_g$, мы будем иметь:

$$\left\{\begin{array}{l} W^s + \Delta W + eaEg - W + \alpha^s + 2\beta^s e^{-\xi''_1}\cos\xi'_1 = 0, \\ W^s + \Delta W + eaEg - W + \alpha^s + 2\beta^s e^{-\xi''_2}\cos\xi'_2 = 0. \end{array}\right.$$

Отсюда следует:

$$\xi_1 = \xi_2.$$

Принимая обозначение $\xi = \xi_1 = \xi_2$, мы можем написать:

$$\boxed{W - eaEg = W^s + \Delta W^s + \alpha^s + 2\beta^s e^{-\xi''}\cos\xi'.} \quad (3,26)$$

Таким образом, одной и той же энергии W соответствуют, как это видно из (3,21) и (3,22), две собственные функции: $\Psi_1(\xi, xyz)$ и $\Psi_2(\xi, xyz)$ (двукратное вырождение). В частном случае, при $E = 0$, эти две вырожденные функции принимают

такой вид:

$$\begin{cases} \Psi_1(\xi, xyz) = C\sum\limits_g e^{i\xi g}\varphi_g(xyz), \\ \Psi_2(\xi, xyz) = C\sum\limits_g e^{-i\xi g}\varphi_g(xyz). \end{cases}$$

Из (3,26) мы видим, что если параметр ξ вещественен, т. е. если $\xi'' = 0$, то

$$C_{\min} \leqslant W - eaEg \leqslant C_{\max}, \qquad (3,27)$$

где

$$\begin{cases} C_{\min} = W^s + \Delta W^s + \alpha^s - 2\beta^s, \\ C_{\max} = W^s + \Delta W^s + \alpha^s + 2\beta^s. \end{cases}$$

Область значений W и g, удовлетворяющих условию (3,27), мы будем называть *зоной* (см. рис. 90, *а*). Вне зоны, т. е. при

$$\begin{cases} W - eaEg < C_{\min}, \\ W - eaEg > C_{\max}, \end{cases}$$

параметр ξ заведомо комплексен, причём $\xi'' < 0$.

Заметим, что, согласно (3,13) и (3,14), мы имеем:

$$\left|\frac{b^1_{g+1}(\xi)}{b^1_g(\xi)}\right|^2 = \left|\frac{b^2_{g1-}(\xi)}{b^2_g(\xi)}\right|^2 = e^{-\xi''}.$$

Отсюда можно заключить, что внутри зоны (т. е. при $\xi'' = 0$) модуль функции $\Psi_1(\xi, xyz)$, так же как и модуль функции $\Psi_2(\xi, xyz)$, периодичен с периодом решётки. Вне зоны (т. е. при $\xi'' < 0$) модуль функции $\Psi_1(\xi, xyz)$ быстро спадает в сторону отрицательных x-ов, а модуль функции $\Psi_2(\xi, xyz)$ — в сторону положительных x-ов (оставаясь модулированными по периоду решётки) (см. рис. 90, *б*).

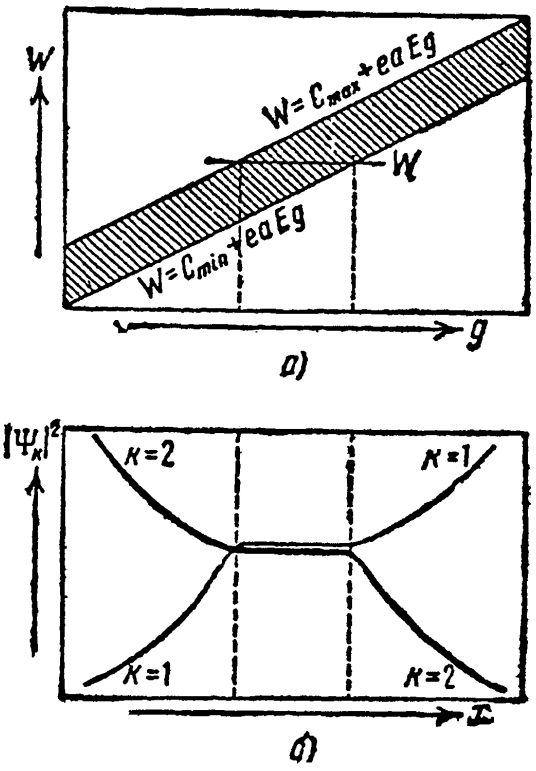


Рис. 90. *а*) Наклонная зона (поле направлено по оси g). *б*) Поведение модуля волновой функции внутри зоны и вне её.

С. Решение задачи для трёхмерного случая. В трёхмерном случае рекуррентные формулы для коэффициентов $b^{\vec{\sigma}}$

имеют вид:

$$\boxed{\begin{aligned} b_{g_1+1,\,g_2,\,g_3}(W) &= b_{\vec{G}}(W)\, e^{i\xi_1(W-ea\vec{E}\vec{G})}, \\ b_{g_1,\,g_2+1,\,g_3}(W) &= b_{\vec{G}}(W)\, e^{i\eta_1(W-ea\vec{E}\vec{G})}, \\ b_{g_1,\,g_2,\,g_3+1}(W) &= b_{\vec{G}}(W)\, e^{i\zeta_1(W-ea\vec{E}\vec{G})}, \end{aligned}} \tag{3,28}$$

или

$$\boxed{\begin{aligned} b_{g_1-1,\,g_2,\,g_3}(W) &= b_{\vec{G}}(W)\, e^{i\xi_2(W-ea\vec{E}\vec{G})}, \\ b_{g_1,\,g_2-1,\,g_3}(W) &= b_{\vec{G}}(W)\, e^{i\eta_2(W-ea\vec{E}\vec{G})}, \\ b_{g_1,\,g_2,\,g_3-1}(W) &= b_{\vec{G}}(W)\, e^{i\zeta_2(W-ea\vec{E}\vec{G})}, \end{aligned}} \tag{3,29}$$

где ξ_k, η_k, ζ_k, вообще говоря, комплексны:

$$\left.\begin{aligned} \xi_k &= \xi_k' + i\xi_k'', \\ \eta_k &= \eta_k' + i\eta_k'', \\ \zeta_k &= \zeta_k' + i\zeta_k''. \end{aligned}\right\} k = 1,\ 2.$$

Это можно показать из общих свойств функции Ψ так же, как это было сделано для одномерного случая.

Формулы (3,28) и (3,29) позволяют выразить любой коэффициент $b_{\vec{G}}$ через ξ_k, η_l, ζ_m

$$b_{\vec{G}} = b_{\vec{G}}^{klm}(\xi_k,\ \eta_l,\ \zeta_m).$$

Здесь каждый из значков k, l, m может иметь любое из двух возможных значений:

$$k,\ l,\ m = 1,\ 2. \tag{3,30}$$

Мы получаем восемь различно построенных собственных функций:

$$\Psi_{klm}(\xi_k,\ \eta_l,\ \zeta_m;\ x,\ y,\ z) = \sum_i c^i \sum_{\vec{G}} b_{\vec{G}}^{klm}(\xi_k,\ \eta_l,\ \zeta_m)\ \vartheta_{\vec{G}}^{i}(x,\ y,\ z)$$

соответственно восьми возможным комбинациям (3,30).

Рассмотрим s- и p-состояния.

1) В случае изолированного s-состояния, принимая обозначения Блоха [см. (2,36) главы II], мы получаем из (3,9)

совершенно тем же путём, как и в одномерной задаче:

$$W^s + \Delta W^s + ea\vec{E}\vec{G} - W + \alpha^s + \\ + 2\beta^s (e^{-\xi''_k} \cos \xi'_k + e^{-\eta''_l} \cos \eta'_l + e^{-\xi''_m} \cos \zeta'_m) = 0.$$

Мы имеем восемь таких уравнений, сравнивая которые, мы вынуждены заключить, что

$$\xi_1 = \xi_2, \quad \eta_1 = \eta_2, \quad \zeta_1 = \zeta_2.$$

Принимая обозначения

$$\xi = \xi_1 = \xi_2, \quad \eta = \eta_1 = \eta_2, \quad \zeta = \zeta_1 = \zeta_2,$$

мы можем написать:

$$\boxed{\begin{array}{c} W - ea\vec{E}\vec{G} = W^s + \Delta W^s + \alpha^s + \\ + 2\beta^s (e^{-\xi''} \cos \xi' + e^{-\eta''} \cos \eta' + e^{-\zeta''} \cos \zeta'). \end{array}} \quad (3{,}31)$$

Мы видим, что одному и тому же W соответствует восемь различных собственных функций $\Psi_{klm}(\xi, \eta, \zeta; x, y, z)$. При отсутствии поля эти функции имеют вид:

$$\Psi(\pm\xi, \pm\eta, \pm\zeta; x, y, z) = \\ = b_0 \sum_i c^i \sum_{\vec{G}} e^{i(\pm\xi g_1 \pm \eta g_2 \pm \zeta g_3)} \varphi^i_{\vec{G}}(x, y, z),$$

где может быть выбрана любая комбинация знаков + и — (таких комбинаций — восемь). Заметим, что вырождение здесь, однако, более, чем восьмикратное, ибо, кроме различных комбинаций значений k, l, m, возможны перестановки чисел ξ, η, ζ одного на место другого, что будет давать нам различные собственные функции, относящиеся к одному и тому же собственному значению W, как это видно из (3,31).

Вещественные значения параметров ξ, η, ζ приводят к представлению о зоне, границы которой устанавливаются условием

$$C_{\min} \leqslant W - ea\vec{E}\vec{G} \leqslant C_{\max},$$

где теперь

$$\begin{cases} C_{\min} = W^s + \Delta W^s + \alpha^s - 6\beta^s, \\ C_{\max} = W^s + \Delta W^s + \alpha^s + 6\beta^s. \end{cases}$$

Внутри зоны модуль функции Ψ_{klm} периодичен с периодом решётки; вне зоны, по мере удаления от её границ, этот модуль убывает или возрастает (оставаясь модулированным по периоду решётки) в зависимости от того, каковы выбранные нами значения индексов k, l, m.

2) Рассмотрим теперь изолированный p-терм. При отсутствии поля этот терм трёхкратно вырожден. При наложении поля он расщепляется на два, из которых один (с магнитным квантовым числом $m=0$) не вырожден, а другой имеет двукратное вырождение ($m=+1$ и $m=-1$). Оба эти терма достаточно близки друг к другу. Значок i в формулах (3,5) и (3,9) имеет теперь три значения ($i=1, 2, 3$). Функции ϑ^1, ϑ^2, ϑ^3 мы выберем так, чтобы при $E=0$ они переходили, соответственно, в

$$\begin{cases} \vartheta^1(xyz) \longrightarrow \varphi^1(xyz) = xf(R), \\ \vartheta^2(xyz) \longrightarrow \varphi^2(xyz) = yf(R), \\ \vartheta^3(xyz) \longrightarrow \varphi^3(xyz) = zf(R), \end{cases}$$

где квантовому числу

$$\begin{cases} m=0 \text{ соответствует функция } \varphi^3, \\ m=+1 \text{ соответствует функция } \frac{1}{2}(\varphi^1 + i\varphi^2), \\ m=-1 \text{ соответствует функция } \frac{1}{2}(\varphi^1 - i\varphi^2). \end{cases}$$

Таким образом:

$$\begin{cases} W^1 = W^2 = W^3, \\ \Delta W^1 = \Delta W^2 \neq \Delta W^3. \end{cases}$$

Положим для определённости $\Delta W^1 = \Delta W^2 > \Delta W^3$ и примем обозначения:

$$\begin{cases} \Delta W^p_{\max} = \Delta W^1 = \Delta W^2, \\ \Delta W^p_{\min} = \Delta W^3, \\ W^p = W^1 = W^2 = W^3. \end{cases}$$

Система (3,9) теперь сводится к трём уравнениям. Совершенно так же, как и в случае s-терма, мы можем убедиться, что

$$\xi_1 = \xi_2 \equiv \xi, \quad \eta_1 = \eta_2 \equiv \eta, \quad \zeta_1 = \zeta_2 \equiv \zeta.$$

Принимая обозначения Вильсона (2,41) (глава II), мы будем иметь:

$$\left.\begin{aligned}
W-ea\vec{E}\vec{G} &= W^p+\Delta W^p_{\max}+\alpha^p+2\beta^p e^{-\xi''}\cos\xi'+\\
&\quad+2\gamma^p(e^{-\eta''}\cos\eta'+e^{-\zeta''}\cos\zeta'),\\
W-ea\vec{E}\vec{G} &= W^p+\Delta W^p_{\max}+\alpha^p+2\beta^p e^{-\eta''}\cos\eta'+\\
&\quad+2\gamma^p(e^{-\xi''}\cos\xi'+e^{-\zeta''}\cos\zeta'),\\
W-ea\vec{E}\vec{G} &= W^p+\Delta W^p_{\min}+\alpha^p+2\beta^p e^{-\zeta''}\cos\zeta'+\\
&\quad+2\gamma^p(e^{-\xi''}\cos\xi'+e^{-\eta''}\cos\eta').
\end{aligned}\right\} \quad (3,32)$$

Вещественные ξ, η, ζ образуют зоны. Мы видим, что p-терм порождает три зоны, соответственно трём уравнениям (3,32). Границы этих зон определяются условиями:

$$\begin{cases}
C^1_{\min}\leqslant W-ea\vec{E}\vec{G}\leqslant C^1_{\max},\\
C^2_{\min}\leqslant W-ea\vec{E}\vec{G}\leqslant C^2_{\max},\\
C^3_{\min}\leqslant W-ea\vec{E}\vec{G}\leqslant C^3_{\max},
\end{cases}$$

где

$$\left.\begin{cases}
C^1_{\min}=C^2_{\min}=W^p+\Delta W^p_{\max}+\alpha^p-2\beta^p-4\gamma^p,\\
C^1_{\max}=C^2_{\max}=W^p+\Delta W^p_{\max}+\alpha^p+2\beta^p+4\gamma^p,\\
C^3_{\min}=W^p+\Delta W^p_{\min}+\alpha^p-2\beta^p-4\gamma^p,\\
C^3_{\max}=W^p+\Delta W^p_{\max}+\alpha^p+2\beta^p+4\beta^p,
\end{cases}\right\} \quad (3,33)$$

так что

$$\begin{cases}
C^1_{\min}=C^2_{\min}=C^3_{\min}+(\Delta W^p_{\max}-\Delta W^p_{\min}),\\
C^1_{\max}=C^2_{\max}=C^3_{\max}+(\Delta W^p_{\max}-\Delta W^p_{\min}).
\end{cases}$$

Мы видим, что границы трёх зон не совпадают: две из них ($i=1$ и $i=2$) наложены одна на другую, но третья ($i=3$) несколько сдвинута. Этот эффект есть результат штарковского расщепления p-терма. Действительно, при $E=0$ мы имеем $\Delta W^p_{\max}=\Delta W^p_{\min}=0$, откуда следует, на основании (3,34), что при отсутствии поля все три зоны имеют общие нижние и верхние границы.

Ширина каждой из трёх зон при наложении поля не меняется. Действительно, из (3,33) мы имеем:

$$C^i_{\max} - C^i_{\min} = 4\beta^p + 8\gamma^p,$$

где $i = 1, 2, 3$. Однако вся p-зона в целом под действием поля оказывается расширенной на величину

$$\Delta W^p_{\max} - \Delta W^p_{\min},$$

равную расстоянию между двумя крайними компонентами p-терма в изолированном *возмущённом* атоме. Этот эффект

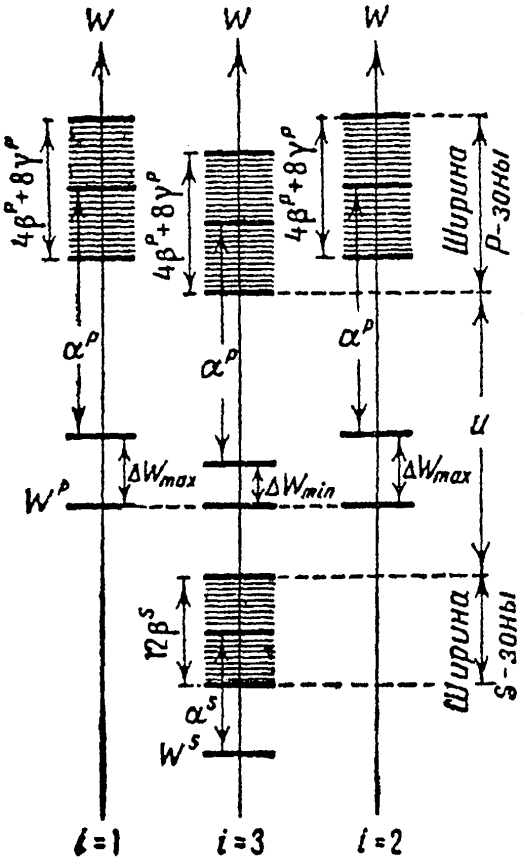


Рис. 91. Расширение зон в электрическом поле.

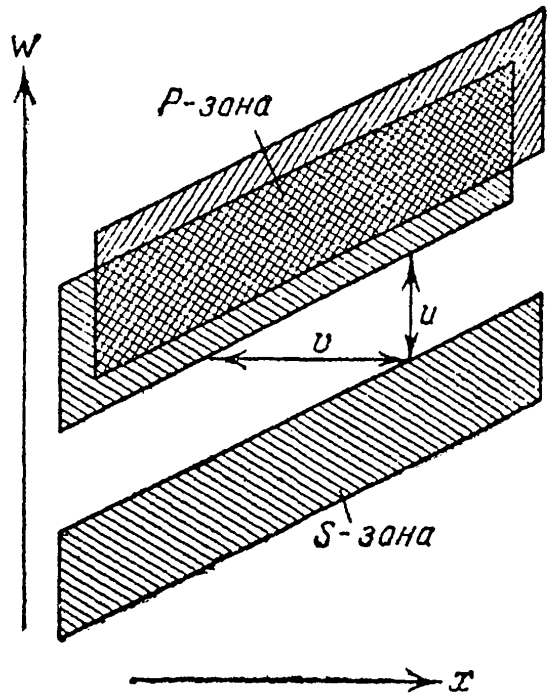


Рис. 92. Наклонные зоны при учёте штаркэффекта.

поясняется схематическим рисунком 91 (ср. с рис. 33), на котором изображены две соседних s- и p-зоны. (На этом рисунке положено $\Delta W^s = 0$, что обычно имеет место).

Очевидно, эффект расширения должен наблюдаться во всякой зоне, соответствующей *вырожденному* терму. Действительно, такую зону следует рассматривать как результат наложения нескольких зон, число которых равно кратности

вырождения. При наложении поля совпадавшие термы раздвигаются, причём сдвиг у каждого из них, вообще говоря, различен, поэтому в результате мы получаем *расщепление* терма. Расщепление термов вызывает раздвижение зон, в результате чего зона в целом оказывается *расширенной*.

Вследствие этого эффекта ширина u запрещённого участка между двумя соседними зонами оказывается функцией поля:

$$u(E) = u(0) + \Delta u(E),$$

где

$\begin{cases} u(0) \text{ — расстояние между зонами при отсутствии поля;} \\ \Delta u(E) \text{ — добавка, обусловленная расширением зон.} \end{cases}$

Обычно мы имеем $\Delta u < 0$, т. е. учёт штаркэффекта приводит к сближению зон, возрастающему по мере увеличения E. Если нижняя зона представляет собой s-зону, а верхняя зона — p-зону, как это изображено на рис. 90, то

$$\begin{cases} u(0) = (W^p - W^s) + (\alpha^p - \alpha^s) - (2\beta^p + 4\gamma^p + 2\beta^s), \\ \Delta u(E) = \Delta W^p_{\min} - \Delta W^s. \end{cases}$$

На рис. 92 изображены наклонные зоны при учёте штаркэффекта. На этом рисунке:

$\begin{cases} u(E) \text{ — расстояние между зонами в «вертикальном» направлении;} \\ v(E) \text{ — расстояние между зонами в «горизонтальном» направлении.} \end{cases}$

Очевидно:

$$v(E) = \frac{u(E)}{eE} = \frac{u(0) + \Delta u(E)}{eE}.$$

Наклонение зоны обусловлено членом $ea\vec{E}\vec{G}$ в формулах (3,33), в то время как *расширение* зоны обусловлено членом $\Delta W^p_{\max} \neq \Delta W^p_{\min}$. При отсутствии поля эти члены исчезают и формулы (3,33) (при $\xi'' = \eta'' = \zeta'' = 0$) превращаются в соответствующие формулы Вильсона (2,44) главы II.

§ 4. Ионизирующее действие поля

А. Механизмы ионизации. Во всех вопросах, касающихся электронной проводимости кристаллической решётки, руководящая роль, как мы уже отмечали, принадлежит её энергетическому спектру. Спектр решётки представляет собой систему уровней, среди которых следует различать уровни двух сортов:

а) собственные уровни решётки, образующие практически сплошные полосы (зоны);

б) локальные уровни, образующие дискретный спектр, наложенный на систему энергетических полос.

Происхождение этих двух сортов уровней, как мы видели, различно. Сплошные полосы появляются как результат периодичности в структуре решётки. Дискретные локальные уровни, наоборот, появляются в результате нарушений этой периодичности.

Поведение электронов, принадлежащих этим двум сортам уровней, также различно. Электроны, принадлежащие зоне, способны с той или иной степенью лёгкости странствовать в пределах зоны. Действительно, для таких электронов, как мы видели, модуль волновой функции периодичен с периодом решётки. Электроны, сидящие на локальных уровнях, наоборот, ограничены в своём движении весьма узкими областями кристалла. Действительно, для них модуль волновой функции обладает резко выраженными максимумами лишь в определённых точках кристалла, расположенных достаточно далеко друг от друга, и быстро спадает при удалении от этих точек.

Пока речь идёт об электропроводности, достаточно рассматривать две соседние зоны, из которых нижняя практически целиком заполнена электронами (нормальная зона), а верхняя — практически совершенно пуста (зона проводимости). Запрещённый участок между этими зонами содержит ряд локальных уровней: доноров и акцепторов.

Таким образом, решётка обладает достаточным запасом электронов. Однако при температуре абсолютного нуля эти электроны совершенно бесполезны с точки зрения проводимости. Электроны, сидящие на локальных уровнях, бесполезны вследствие своей неподвижности. Электроны, принадлежащие нормальной зоне, бесполезны не потому, что они не способны передвигаться (они обладают этой способностью),

а потому, что внешнее электрическое поле не в силах вызвать асимметрию в их распределении по скоростям, что необходимо для того, чтобы можно было говорить о токе.

Для того чтобы заставить диэлектрик проводить, необходимо:

1) обогатить электронами зону проводимости,

2) обеднить нормальную зону.

Необходимо выполнить хотя бы одно из этих требований или оба одновременно. Соответственно этому электронный ток может быть:

а) чисто электронным, если выполняется первое требование, но не выполняется второе;

б) чисто дырочным, если выполняется второе, но не выполняется первое;

в) смешанным, если сразу выполняются оба требования.

Чисто электронный ток имеет место в том случае, когда зона проводимости черпает свои электроны с локальных уровней, расположенных в запрещённом участке. Электрон и дырка рождаются одновременно; однако электрон свободно уходит к аноду по зоне проводимости, а дырка остаётся закреплённой на локальном уровне. Таким образом, для осуществления чисто электронного тока необходимо, чтобы локальные уровни содержали электроны.

Чисто дырочный ток имеет место в том случае, когда связанные электроны переходят из нормальной зоны на пустые локальные уровни в запрещённом участке. В этом случае также электрон рождается вместе с дыркой. Теперь, однако, закреплённым оказывается электрон, а дырка свободно уходит к катоду по нормальной зоне. Необходимое условие для осуществления чисто дырочного тока состоит, следовательно, в том, чтобы локальные уровни не были заняты электронами.

Смешанный ток появляется в том случае, когда электрон попадает в зону проводимости непосредственно из нормальной зоны. И в этом случае электрон рождается одновременно с дыркой, но теперь оба партнёра оказываются свободными. Возникнув, они расходятся в разные стороны: электрон — к аноду по зоне проводимости, дырка — к катоду по нормальной зоне.

Итак, чтобы вызвать в решётке проводимость, необходимо воздействовать на неё такими агентами, которые снабжали бы

электронами зону проводимости или опустошали бы нормальную зону (ионизирующие агенты). Такими агентами служат: 1) температура, 2) освещение, 3) сильное электрическое поле.

Что касается температуры, то она играет существенную роль лишь в тех случаях, когда уровни, заполненные электронами (те, с которых происходит переход), и уровни пустые (те, на которые происходит переход) разделены достаточно узким запрещённым участком, т. е. в случае полупроводников. В этих случаях, как известно, решётка действительно обладает, даже в сравнительно слабых полях, значительной электронной проводимостью, быстро возрастающей с температурой. В случае широких запрещённых участков, т. е. в случае диэлектриков, температура сама по себе оказывается недостаточно мощным фактором для того, чтобы создать сколько-нибудь заметное количество носителей тока. Диэлектрик необходимо было бы нагреть до температур, значительно превышающих температуру плавления, для того чтобы создать в нём такой же электронный ток, какой наблюдается в полупроводниках при комнатных температурах. Таким образом, для диэлектриков температура как ионизирующий агент практически отпадает. Что же касается электрического поля, то, будучи достаточно сильным, оно также может играть роль ионизирующего агента, причём его ионизирующая способность зависит от ширины запрещённого участка: ионизация полем наступает тем раньше, чем уже этот запрещённый участок.

Ионизирующая способность поля делается возможной благодаря тем деформациям, которые оно вызывает в энергетическом спектре решётки. Эта деформация, как мы видели, сводится к 1) наклонению зон, 2) расширению зон.

Благодаря наклонению зон электрон приобретает способность переходить с одного уровня на другой (через разделяющий их запрещённый участок) двумя путями, которые отмечены на рис. 93 вертикальной и горизонтальной стрелками 4 и 3. Вертикальный переход связан с затратой энергии; этот переход аналогичен возбуждению. Переходы этого рода обусловлены механизмом *ударной ионизации* (ионизация столкновениями). Горизонтальный переход происходит без затраты энергии; он подобен просачиванию через потенциальный барьер. Такие переходы мы будем называть *электростатической ионизацией*. Поле, вызывая наклонение зон, делает возможными переходы как того, так и другого рода. В обоих

случаях ионизирующая способность поля тем больше, чем уже запрещённый участок, который должен быть преодолён электроном.

Как горизонтальные, так и вертикальные переходы могут совершаться:

1) с локальных уровней на зону проводимости (чисто электронный ток);

2) с нормальной зоны на локальные уровни (чисто дырочный ток);

3) с нормальной зоны на зону проводимости (смешанный ток).

Горизонтальные переходы, кроме того, могут совершаться:

4) с катода на зону проводимости (чисто электронный ток);

5) с нормальной зоны на анод (чисто дырочный ток).

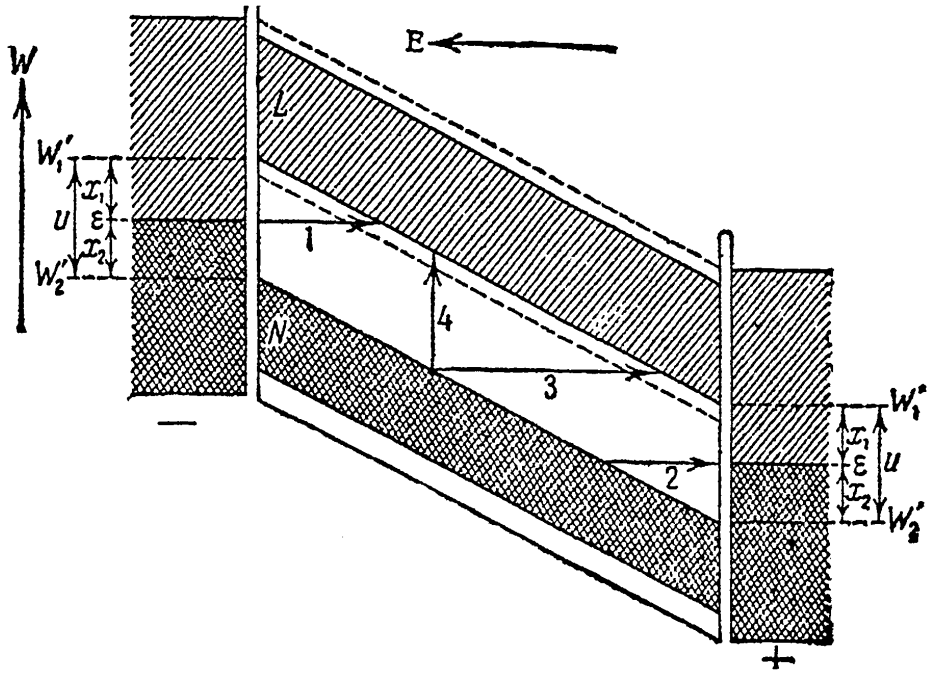


Рис. 93. Энергетическая схема полупроводника в сильном поле. N — нормальная зона; L — зона проводимости (пунктиром показаны границы расширенной зоны). Локальные уровни не изображены на схеме. Стрелки иллюстрируют механизмы ионизации.

Таким образом, ионизация может развиваться или во всём объёме диэлектрика (полупроводника) (случаи 1, 2, 3) или только в его приэлектродных слоях (случаи 4, 5).

Благодаря расширению зон запрещённый участок между зонами сужается (как в горизонтальном, так и в вертикальном направлении). Таким образом, штарковское расширение служит фактором, облегчающим ионизацию при любом из механизмов.

Учитывая штаркэффект, мы тем самым заставляем оба ионизационных механизма (ударная ионизация, электростатическая ионизация) вступать в действие при полях несколько более слабых, чем это имеет место при пренебрежении эффектом Штарка. Здесь сужение запрещённого участка оказывает такое же действие, как дополнительное увеличение поля. Кроме того, штаркэффект усиливает роль температурной ионизации (что может быть существенным в случае полупроводников). Действительно, сужение запрещённого участка увеличивает концентрацию тепловых электронов в зоне проводимости или дырок в нормальной зоне. Здесь сужение запрещённого участка оказывает такое же действие, как дополнительное повышение температуры.

В. Ударная ионизация. Вертикальные переходы происходят в результате соударений между электронами. Здесь энергия одного электрона увеличивается за счёт энергии другого. При этих процессах зона проводимости обогащается электронами или нормальная зона обогащается дырками, или же одновременно происходит и то и другое.

В первом случае свободный электрон (ионизирующий электрон), двигаясь в зоне проводимости, сталкивается с электроном, сидящим на локальном уровне (ионизируемый электрон), и подымает его в зону проводимости, теряя при этом как раз ту энергию, которую приобретает электрон ионизируемый. При этом ионизировавший электрон продолжает оставаться в зоне проводимости, падая, однако, с более высоких её уровней на более низкие. Очевидно, такого рода соударения свободный электрон может совершать лишь тогда, когда, двигаясь по зоне проводимости, он уходит достаточно далеко от её нижней границы. Расстояние его от нижней границы должно быть не меньше $u_1 + u_0$, если считать по вертикали, или не меньше $(u_1 + u_0)/eE$, если считать по горизонтали (см. рис. 94, *а*), где $u_1 + u_0$ — ширина запрещённого участка между нижней границей зоны проводимости и тем уровнем, на котором сидит ионизируемый электрон. Заходит ли вообще свободный электрон в ту область зоны проводимости, где он делается способным ионизировать? Это имеет место лишь при достаточно сильных полях. Действительно, свободный электрон, двигаясь к аноду, спускается по ступенькам лестницы, лежащей внутри зоны проводимости (рис. 94 *а*); каждый переход с одной ступеньки на следующую нижнюю

свидетельствует об акте соударения с решёткой. Длина ступеньки λ есть средняя длина свободного пробега, а высота ступеньки ε есть та средняя порция энергии, которую теряет электрон при каждом таком соударении. Будет ли заходить этот ступенчатый путь электрона в ту область зоны проводимости, где делается возможной ударная ионизация, или будет целиком лежать ниже этой области — это зависит (при заданных λ и ε) от того, насколько сильно наклонена зона, т. е. от величины внешнего поля E. Условие, вообще необходимое для того, чтобы была возможна такого рода ионизация, состоит в том, чтобы зона проводимости была шире запрещённого участка $u_1 + u_0$. Кроме того, чтобы этот механизм ионизации мог развиваться, необходимо, чтобы в зоне проводимости уже заранее был приготовлен хотя бы один случайный электрон. В рассматриваемом случае, когда ионизация происходит при помощи свободного электрона, двигающегося по зоне проводимости, энергия ионизирующего электрона до соударения, конечно, больше энергии ионизируемого. После соударения энергия ионизировавшего электрона может оказаться как больше, так и меньше, чем энергия электрона ионизированного.

Во втором случае ионизация происходит при помощи дырки, двигающейся в нормальной зоне. В действительности же здесь сталкиваются два электрона, оба принадлежащие нормальной зоне, в результате чего один из них (ионизируемый) выбрасывается вверх за пределы нормальной зоны на пустой локальный уровень, а другой (ионизирующий), оставаясь в нормальной зоне, падает на её более низкий уровень, который предполагается свободным. Очевидно, что такое соударение может происходить лишь при условии, что дырка находится достаточно далеко от верхней границы нормальной зоны. Её расстояние до верхней границы должно быть не меньше $u_2 + u_0$, если считать по вертикали, или не меньше $(u_2 + u_0)/eE$, если считать по горизонтали, где $u_2 + u_0$ есть ширина запрещённого участка между верхней границей нормальной зоны и тем локальным уровнем, на который перебрасывается электрон. Способна ли вообще дырка удаляться от верхней границы нормальной зоны настолько, чтобы сделать возможными такого рода соударения? Это имеет место лишь при достаточно сильных полях. Действительно, двигаясь к катоду, дырка подымается по ступенькам лестницы, лежащей внутри

нормальной зоны (рис. 94, *б*). Будет ли эта лестница заходить в требуемую область — это зависит от того, насколько сильно наклонена зона. Условие, вообще необходимое для того, чтобы была возможна такая ионизация, состоит в том, чтобы нормальная зона была шире запрещённого участка, и, кроме того, необходимо, чтобы нормальная зона с самого начала содержала, по крайней мере, одну дырку. В рассматриваемом

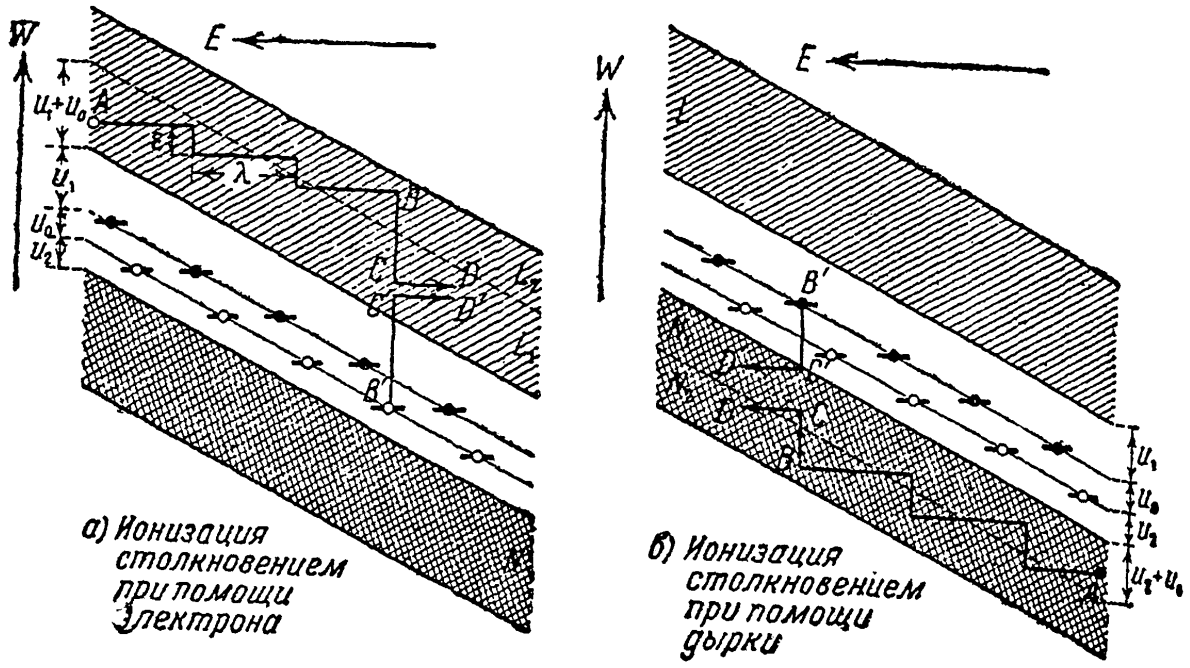


Рис. 94. Механизм ударной ионизации.
N — нормальная зона; L — зона проводимости; N_1, L_1 — та часть зоны, в которой ударная ионизация невозможна; N_2, L_2 — та часть зоны, в которой происходит ударная ионизация; $ABCD$ — путь ионизирующего электрона (дырки). Скачок BC соответствует моменту ионизации; $B'C'D'$ — путь ионизуемого электрона (дырки). Очевидно: $BC = B'C'$.

случае до соударения между электронами энергия ионизирующего электрона не обязана быть больше энергии ионизируемого. После же соударения энергия ионизировавшего электрона (т. е. оставшегося в нормальной зоне), конечно, заведомо меньше энергии ионизированного (т. е. выброшенного из зоны).

Наконец, в третьем случае мы имеем столкновения между свободным электроном зоны проводимости и связанным электроном нормальной зоны. При таком столкновении связанный электрон делается свободным, и, таким образом, зона проводимости, обогащаясь электронами, тем самым обогащает дырками нормальную зону.

Наряду с процессом ударной ионизации, при котором электрон так или иначе совершает вертикальный переход снизу вверх, имеет место обратный процесс, при котором электрон совершает такой же вертикальный переход, но сверху вниз, отдавая свою энергию решётке. Этот процесс приводит к связыванию свободного электрона, если электрон падает из зоны проводимости на пустой локальный уровень, или к связыванию дырки, если электрон падает с локализованного уровня в нормальную зону, или, наконец, к связыванию одновременно и электрона и дырки, если электрон падает с зоны проводимости в нормальную зону. Если бы этого процесса не существовало, то ток в диэлектрике возрастал бы с расстоянием (подобно лавине). Стационарный ток, постоянный по всему объёму кристалла, устанавливается лишь тогда, когда оба эти процесса (ионизация столкновением и связывание) приходят в равновесие. Процесс связывания может происходить: а) по линейному закону, б) по квадратичному закону.

Линейный закон имеет место в том случае, когда при связывании электроны садятся не на те уровни, с которых они были сорваны. (Так происходит, например, если электроны подняты из нормальной зоны, а прилипают к пустым локальным уровням.) Действительно, в этом случае число электронов, связываемых за 1 *сек*, пропорционально наличному числу ионизованных электронов. Если же электроны при связывании падают на ту же самую систему уровней, с которой они подымаются в результате ионизации, то мы имеем квадратичный закон. Здесь число связываемых электронов пропорционально квадрату от числа наличных электронов. (Коэффициент пропорциональности, т. е. вероятность связывания, конечно, не обязан быть одинаковым для обоих этих случаев.)

Учёт процесса связывания представляет собой как раз то, чем современные теории ударной ионизации отличаются от старых теорий [7]. В старых теориях при рассмотрении процесса высвобождения электронов обратный процесс связывания вовсе не учитывался. Поэтому классические теории давали лавинообразное возрастание тока с расстоянием. В современных теориях учитываются оба эти противоположных процесса, которые, находясь в равновесии, обеспечивают некоторую постоянную концентрацию электронного (дырочного) газа, тем

большую, чем больше поле. Поэтому новые теории дают ток, постоянный по всему объёму диэлектрика.

Квантовомеханической теории ударной ионизации посвящены работы Фрелиха [8] и скончавшегося в 1942 г. в Ленинграде Н. Л. Писаренко [9]. Оба автора рассматривают ударную ионизацию как механизм, приводящий к пробою. Оба автора получают для пробивных напряжений значения, по порядку величины совпадающие с экспериментальными. Это обстоятельство также является характерной чертой, отличающей квантовомеханические теории от классических: все классические теории ударной ионизации приводили к пробивным напряжениям, более высоким, чем это имеет место в действительности.

Работа Писаренко относится к случаю примесных полупроводников. Её недостатком является то, что она даёт неправильную температурную зависимость для электропроводности. Зависимость $\ln\sigma$ от E у Писаренко имеет, примерно, такой вид, как на рис. 81. Это находится в согласии с экспериментом. Однако при повышении температуры кривая Писаренко, сдвигаясь вверх, становится более крутой, в то время как в действительности, как мы видели, она должна становиться более пологой.

Работа Фрелиха относится к щёлочно-галоидным кристаллам. Её недостатком является то, что она даёт неправильную зависимость пробивного напряжения от ширины запрещённого участка. У Фрелиха пробивное напряжение возрастает при сужении запрещённого участка, в то время как в действительности оно должно убывать.

Против механизма ударной ионизации можно высказать следующее общее возражение экспериментального характера. Если из фототока, измеренного в кристалле, вычесть темновой ток, то мы получим ток, обусловленный исключительно световыми электронами. Этот ток строго следует закону Ома даже при таких высоких полях, при которых темновой ток уже существенно отклоняется от закона Ома [1]. Из этого экспериментального факта можно сделать следующие два важных вывода:

1) можно заключить, что подвижность электронов не зависит от поля;

2) можно заключить, что отклонения темнового тока от закона Ома вызваны не ударной ионизацией, а каким-либо

другим механизмом, ибо в противном случае было бы непонятно, почему световые электроны не могут участвовать в ударной ионизации наравне с темновыми.

Кроме того, по поводу существующих теорий ударной ионизации можно высказать следующее соображение теоретического характера. В теоретических работах ионизирующие электроны (или дырки) рассматриваются как свободные, с той лишь разницей, что их действительная масса заменяется некоторой эффективной массой, которая считается постоянной. Это справедливо до тех пор, пока электронный газ располагается на дне зоны проводимости или дырочный газ — под потолком нормальной зоны, как это и имеет место в полупроводниках при слабых полях. Однако, для того чтобы электрон или дырка приобрели способность ионизировать, они должны быть оттянуты в глубину зоны. Можно попрежнему рассматривать их как свободные, но их эффективная масса при этом возрастает, как мы видели в § 2, В; около середины зоны это возрастание происходит особенно быстро. Ионизирующий электрон (или дырку) можно характеризовать некоторой эффективной массой, средней для промежутка времени от одного акта ионизации до следующего. Однако такая средняя эффективная масса будет зависеть, во-первых, от того, насколько глубоко внутрь зоны должен забраться электрон (дырка), для того чтобы произвести ионизацию, а, во-вторых, от того, сколько соударений с решёткой он испытает, прежде чем совершит ионизацию, т. е. от того, сколько ступенек (рис. 94, *а*, *б*) лежит на его пути между точками *A* и *B*. Таким образом, средняя эффективная масса оказывается функцией от ширины запрещённого участка и от длины свободного пробега (иначе говоря, от температуры). Учёт этого обстоятельства несомненно должен отразиться на результатах расчёта.

Пока мы имеем дело с полупроводниками, механизм ударной ионизации (во всяком случае, в рамках теории Писаренко), повидимому, может быть исключён из рассмотрения. Действительно, если бы этот механизм осуществлялся, то, как это следует из теории Писаренко, он вступал бы в действие уже при полях 1—10 $\frac{\text{kV}}{см}$, т. е. на участке Пуля. При этом он приводил бы к закономерностям, которые, как мы видели, находятся в противоречии с экспериментом.

С. Вероятность электростатической ионизации. Горизонтальные переходы, т. е. переходы без изменения энергии, становятся возможными в результате наклонения зон. Действительно, волновая функция для электрона с заданной энергией быстро затухает в запрещённом участке, однако её хвост простирается достаточно далеко. Это и обусловливает то обстоятельство, что электрон всегда имеет некоторый шанс очутиться по ту сторону запрещённого участка.

Вероятность такого просачивания чрезвычайно чувствительна к длине запрещённого участка, преодолеваемого электроном. Этот участок укорачивается по мере наклонения зон, т. е. по мере возрастания E. Можно показать (см. ниже), что вероятность горизонтального перехода, которую мы обозначим через P, будучи практически равна нулю при не слишком сильных полях, почти внезапно становится весьма значительной, коль скоро зоны приобретают достаточный наклон. Таким образом, при достаточно сильном поле зона проводимости всегда может ожидать, что в ней появится электрон, просочившийся в неё с локального уровня или непосредственно с катода, или, наконец, из нормальной зоны. Нормальная зона, в свою очередь, всегда можеть ожидать, что в ней возникнет дырка за счёт электрона, ушедшего сквозь тоннель на пустой локализованный уровень или непосредственно на анод, или, наконец, в зону проводимости.

Обозначим через P_1 вероятность перехода, изображённого на рис. 93 стрелкой 1. Переходы этого типа вызывают в диэлектрике (полупроводнике) чисто электронный ток. Через P_2 обозначим вероятность перехода, соответствующего стрелке 2 на рис. 93. Эти переходы приводят к возникновению чисто дырочного тока. Наконец, через P_3 обозначим вероятность перехода, изображённого стрелкой 3. Эти переходы вызывают смешанный ток. Вычислим вероятности P_1, P_2, P_3.

Вероятность P_3 может быть вычислена по формуле:

$$P_3 = \left| \frac{\Psi^i(\lambda_1, \lambda_2, \lambda_3; x_1, y_1, z_1)}{\Psi^i(\lambda_1, \lambda_2, \lambda_3; x_2, y_2, z_2)} \right|^2, \qquad (4,1)$$

где Ψ^i — волновая функция для электрона с энергией W, запертого в i-й зоне, а точки x_2, y_2, z_2 и x_1, y_1, z_1 соответствуют началу и концу тоннеля. Как мы видели [см. (3, 4)],

функция Ψ^l имеет вид:

$$\Psi^l(\lambda_1, \lambda_2, \lambda_3; x, y, z) = e^{\frac{i}{a}\int^{\vec{\Lambda}}\vec{dR}} u^l(\lambda_1, \lambda_2, \lambda_3; x, y, z), \quad (4,2)$$

где λ_1, λ_2, λ_3 — некие функции, зависящие от $W - e\vec{E}\vec{R}$, а u^l — функция обладающая свойством:

$$u^l(x, y, z) = u^l(x + an_1, y + an_2, z + zn_3),$$

где n_1, n_2, n_3 — любые целые числа и a — постоянная решётки (решётка предполагается кубической). Если положить $E = 0$, то волновая функция (4,2) превращается в

$$\psi^l(\lambda_1, \lambda_2, \lambda_3; x, y, z) = e^{\frac{i}{a}\vec{\Lambda}\vec{R}} u^l(\lambda_1, \lambda_2, \lambda_3; x, y, z), \quad (4,3)$$

где теперь λ_1, λ_2, λ_3 играют роль параметров, а функция u^l становится периодической функцией от координат с периодом a. Таким образом, $\vec{\Lambda}$ в (4,2) зависит от $W - e\vec{E}\vec{R}$ так же, как $\vec{\Lambda}$ в (4,3) зависит от W.

Заметим, что если пренебречь периодичностью потенциала в решётке, как это, например, делается в зоммерфельдовской теории металлов, то функция u^l делается постоянной и функция (4,3) превращается в

$$\psi(\lambda_1, \lambda_2, \lambda_3; x, y, z) = Ce^{\frac{i}{a}\vec{\Lambda}\vec{R}}.$$

Связь между Λ и W при этом оказывается чрезвычайно простой:

$$\Lambda^2 = \frac{4\pi^2 a^2}{h^2} 2mW. \quad (4,4)$$

Вектор $\vec{\Lambda}$ в (4,2) или в (4,3) может быть вещественным или комплексным. Он вещественен внутри зоны и комплексен вне её. Полагая в области запрещённого участка

$$\vec{\Lambda} = \vec{\Lambda}' + i\vec{\Lambda}''$$

и подставляя (4,2) в (4,1), получаем:

$$P_s = e^{-\frac{2}{a}\int^{\vec{\Lambda}''}\vec{dR}}.$$

Интеграл, стоящий в показателе этого выражения, был вычислен Зинером [4] для тоннеля длины u/eE, что соответствует просачиванию из зоны в зону. Зинер получил:

$$\frac{2}{a}\int \vec{\Lambda}''d\vec{R} = \frac{\pi^2 ma}{h^2}\frac{u^2}{eE}. \qquad (4,5)$$

Если измерять u в eV и E в kV/*см* и положить $a = 3\cdot 10^{-8}$ *см*, мы получаем отсюда

$$P_3 = 10^{-5\cdot 10^3 \frac{u^2}{E}}. \qquad (4,6)$$

Вероятность просачивания P_3, конечно, совершенно одинакова как в направлении, указанном на рис. 93 стрелкой 3, так и в обратном направлении. Отсюда следует:

$$\begin{cases} P_1 = 10^{-5\cdot 10^3 \frac{x_1^2}{E}}, \\ P_2 = 10^{-5\cdot 10^3 \frac{x_2^2}{E}}, \end{cases} \qquad (4,7)$$

где, в соответствии с обозначениями, принятыми в § 2 главы IV:

$\begin{cases} x_1 \text{ — расстояние от нижней границы зоны проводимости до уровня химического потенциала } \varepsilon, \\ x_2 \text{ — расстояние от верхней границы нормальной зоны до уровня } \varepsilon, \end{cases}$

так что

$$x_1 + x_2 = u.$$

Таким образом, P_1 соответствует переходу с уровня Ферми в катоде на зону проводимости, а P_2 — переходу с нормальной зоны на уровень Ферми в аноде.

Заметим, что в том случае, когда связь между Λ и W выражается формулой (4,4), т. е. при пренебрежении периодическим потенциалом решётки, интеграл (4,5) будет иметь вид:

$$\frac{2}{a}\int \vec{\Lambda}''d\vec{R} = \frac{4\pi}{h}\int \sqrt{2m(e\vec{E}\vec{R} - W)}\,dR = \frac{8\pi\sqrt{2mu^3}}{3heE}, \qquad (4,8)$$

откуда, если измерять u в *e*V, а E — в kV/*см*:

$$P_3 = 10^{-3\cdot 10^4 \frac{u^{3/2}}{E}}. \qquad (4,9)$$

В этом случае задача о просачивании электрона из одной зоны в соседнюю через разделяющий их запрещённый участок эквивалентна задаче о просачивании сквозь потенциальный барьер треугольной формы, высота которого u, а ширина u/eE (см. рис. 95). Аналогично получаем:

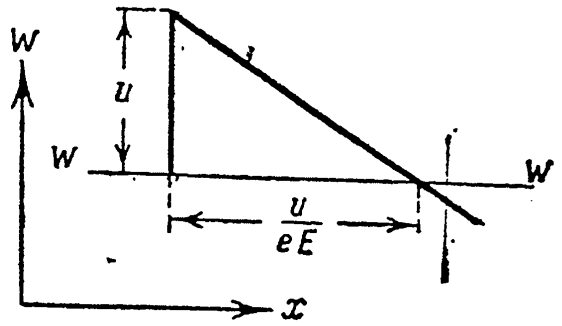


Рис. 95. Треугольный потенциальный барьер.

$$\begin{cases} P_1 = 10^{-3\cdot 10^4 \frac{x_1^{3/2}}{E}}, \\ P_2 = 10^{-3\cdot 10^4 \frac{x_2^{3/2}}{E}}. \end{cases} \quad (4,10)$$

Принимая обозначение $x_3 = u$, мы можем переписать формулы (4,9) и (4,10), с одной стороны, и формулы (4,7) и (4,6), с другой стороны, так:

если пренебрегать периодическим полем в полупроводнике:

$$P_i = e^{-\frac{8\pi\sqrt{2m}}{3he}\cdot\frac{x_i^{3/2}}{E}} = 10^{-3\cdot 10^4 \frac{x_i^{3/2}}{E}} \ldots, \quad (4,11)$$

если учитывать периодическое поле в полупроводнике:

$$P_i = e^{-\frac{\pi^2 ma}{h^2 e}\cdot\frac{x_i^2}{E}} = 10^{-5\cdot 10^3 \frac{x_i^2}{E}} \ldots, \quad (4,12)$$

где $i = 1, 2, 3$. Мы видим, что показатель в формуле (4,12) всегда меньше, чем в формуле (4,11):

$$\frac{\frac{\pi^2 ma}{h^2 e}\cdot\frac{x_i^2}{E}}{\frac{8\pi\sqrt{2m}}{3he}\frac{x_i^{3/2}}{E}} = \frac{3\pi a}{8h}\sqrt{\frac{mx_i}{2}} = 0{,}17\sqrt{x_i} < 1. \quad (4,13)$$

(Это неравенство имеет место до тех пор, пока $x_i < 40$ eV.) Таким образом, учёт периодического поля облегчает просачивание.

Пренебрежение периодическим полем допустимо лишь непосредственно вблизи границы зоны (как со стороны разрешённой, так и со стороны запрещённой области), так как

только вблизи этой границы электроны ведут себя подобно свободным электронам зоммерфельдовского металла. При этом они все отличаются от свободных электронов своей массой, так что в формулу (4,4), а следовательно, и в (4,8), вместо истинной массы электрона m должна быть подставлена некоторая эффективная масса m^*, зависящая от ширины зоны. Заметим, что при оценке вероятности просачивания пользование формулой (4,11) вместо правильной формулы (4,12) в некоторых случаях может привести к довольно существенной ошибке. Действительно, например, при $u = 0{,}5$ eV (случай полупроводника) одинаковое значение для P_3 достигается при полях, в десять раз более слабых в том случае, когда учитывается периодичность потенциала [формула (4,12)], чем в том случае, когда этой периодичностью пренебрегают [формула (4,11)] [10].

D. Электростатическая компонента проводимости. Возвращаясь к рис. 93, мы видим, что свободные заряды, появляющиеся в полупроводнике вследствие тоннельного эффекта, по своему происхождению могут быть двух сортов:

а) Это могут быть электроны и дырки, принадлежащие самому полупроводнику и переведённые при помощи тоннельного механизма из связанного состояния в свободное. Таким образом, полупроводник может обогащаться свободными зарядами за счёт своих собственных запасов.

б) Это могут быть электроны или дырки, не принадлежавшие ранее полупроводнику, но попавшие в него при помощи тоннельного эффекта из металлических электродов. Таким образом, полупроводник может обогащаться электронами за счёт катода и дырками — за счёт анода («холодное вырывание»).

Роль, которую играют в полупроводнике его собственные электроны и дырки, рассмотрена в работе Зинера [4]. Роль, которую могут играть чужие электроны и дырки, сорванные с поверхности электродов, рассматривается в работе Фаулера [11].

Если учитывать электростатическую ионизацию, то полная проводимость σ полупроводника будет:

$$\sigma = \sigma_{\text{терм}} + \sigma_{\text{элст}}, \tag{4,14}$$

где $\sigma_{\text{терм}}$ — обычная термическая проводимость, которую мы здесь будем считать не зависящей от поля, т. е. подчиняю-

щейся закону Ома. Это есть та доля проводимости, которая обусловлена электронами и дырками, проходящими из электродов в полупроводник *поверх* потенциального барьера, т. е. теми электронами катода, энергия W которых лежит выше уровня W_1', т. е. для которых $W \geqslant W_1'$, и теми дырками на аноде, энергия W которых лежит ниже уровня W_2', т. е. для которых $W \leqslant W_2'$ (см. рис. 93);

$\sigma_{элст}$ — электростатическая часть проводимости, существенно зависящая от напряжённости поля E. Это есть та часть проводимости, которая обусловлена электронами и дырками, проникающими в полупроводник с электродов сквозь потенциальный барьер, т. е. теми электронами катода, для которых $W < W_1'$, и теми дырками анода, для которых $W < W_2'$.

Мы имеем в (4,14):

$$\begin{cases} \text{в «слабых» полях} \ldots\ldots \sigma_{терм} \gg \sigma_{элст}, \\ \text{в «сильных» полях} \ldots\ldots \sigma_{терм} \ll \sigma_{элст}. \end{cases}$$

Обозначим через E_0' то значение поля, при котором $\sigma_{терм} = \sigma_{элст}$. Это есть то значение поля, при котором механизм электростатической ионизации практически вступает в игру.

Вычислим электростатическую часть проводимости $\sigma_{элст}$. Пренебрегая толщиной промежуточного слоя между металлом и полупроводником, теми искажениями в энергетическом спектре решётки, которые вызваны близостью поверхности, и теми изменениями в свойствах приконтактного слоя, которые происходят под влиянием тока, иначе говоря, до известной степени схематизируя процесс, который в действительности является более сложным, мы будем иметь:

$$\sigma_{элст} = \frac{e^2E}{8\pi h}\left(\frac{P_1}{x_1} + \frac{P_2}{x_2}\right) = \begin{cases} \approx \dfrac{e^2E}{8\pi h x_1} P_1 \ldots\ldots, \text{ если } x_1 < x_2, \\ = \dfrac{2e^2E}{8\pi h x_1} P_1 = \dfrac{2e^2E}{8\pi h x_2} P_2 .., \text{ если } x_1 = x_2, \\ \approx \dfrac{e^2E}{8\pi h x_2} P_2 \ldots\ldots, \text{ если } x_2 < x_1. \end{cases} \tag{4,15}$$

В этих формулах коэффициент при P_i заимствован из теории холодного вырывания (см., например, Зоммерфельд-Бете [12]),

что является допустимым, пока мы остаёмся в рамках принятых нами приближений. Из (4,15) мы видим, что электростатический ток по своей природе может быть:

1) чисто электронным, если $x_1 < x_2$. Это имеет место, как мы видели в главе IV, в полупроводниках, обладающих стехиометрическим избытком металла (при не слишком высоких температурах);

2) чисто дырочным, если $x_2 < x_1$. Это имеет место в полупроводниках с избытком металлоида (при не слишком высоких температурах);

3) смешанным, если $x_1 = x_2$. Это имеет место в случае идеальной решётки, а также в случае реальной решётки при достаточно высоких температурах.

Если в формулы (4,15) подставить P_i (где $i = 1,2$) из (4,11), как это делает Фаулер, то это будет означать замену полупроводника вакуумом. В этом случае формулы (4,15) совпадают с формулой Нордгейма для обычного холодного выравнивания (без учёта сил отражения), причём электростатическая проводимость в результате этого оказывается существенно сниженной. Если же учитывать периодический потенциал решётки, т. е. в формулы (4,15) подставить P_i из (4,12), то, в отличие от Фаулера, мы будем иметь:

$$\sigma_{\text{элст}} = \frac{e^2 E}{8\pi h x_i} e^{-\frac{\pi^2 m a x_i^2}{h^2 e E}}, \qquad (4,16)$$

где, согласно (4,12):

$$\begin{cases} i = 1 \text{ при } x_1 < x_2, \\ i = 2 \text{ при } x_2 < x_1, \end{cases}$$

или, измеряя x_i в eV, E — в kV/*см* и полагая $a = 3 \cdot 10^{-8}$ *см*:

$$\sigma_{\text{элст}} = 1{,}55 \cdot 10^{-3} \frac{E}{x_i} \cdot 10^{-5 \cdot 10^3 \frac{x_i^2}{E}} \text{ ом}^{-1}\ \text{см}^{-1}. \qquad (4,17)$$

Формула (4,16) [или (4,17)], строго говоря, верна лишь для абсолютного нуля температуры, т. е. для того случая, когда в металле нет ни одного электрона с энергией, лежащей выше уровня Ферми ε, и ни одной дырки, находящейся ниже этого уровня. Однако практически этой формулой можно пользоваться и при более высоких температурах, ибо число

электронов с энергией $W > \varepsilon$ и число дырок с энергией $W < \varepsilon$ уменьшается по мере удаления от границы ε слишком быстро для того, чтобы это уменьшение могло быть скомпенсировано увеличением коэффициента просачивания. Отсюда, однако, не следует, что $\sigma_{элст}$ практически не зависит от температуры, как это имеет место при обычном холодном вырывании в вакуум. В отличие от обычного холодного вырывания здесь зависимость от T скрыта в x_i. Действительно, как мы видели в предыдущей главе, положение уровня химического потенциала ε (или, что то же самое: границы Ферми) на нашей энергетической схеме зависит от T, μ, ν. При возрастании температуры от нуля до бесконечности уровень ε сдвигается так, что x_i увеличивается (см., например, рис. 73):

$$\begin{cases} \text{от } x_i = u_i \text{ при } T = 0, \\ \text{до } x_i = \dfrac{u}{2} \text{ при } T = \infty. \end{cases}$$

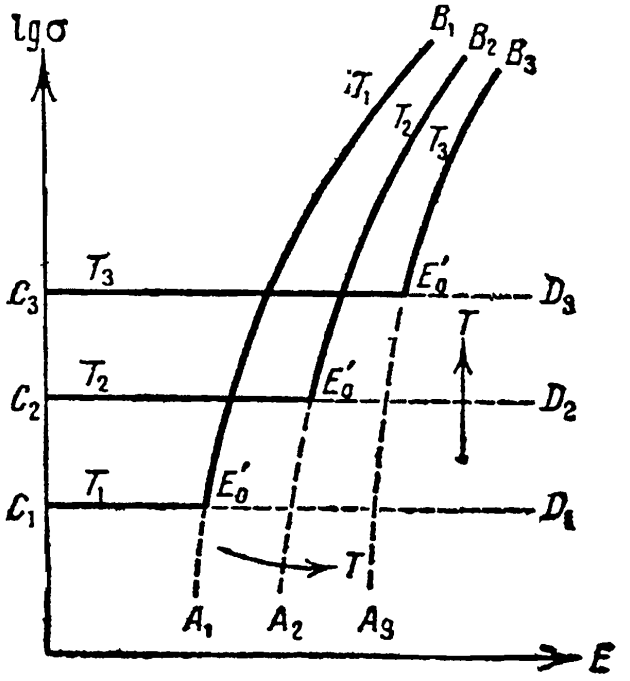


Рис. 96. Соотношение между термической и электростатической компонентами проводимости при различных температурах. Термическая проводимость $\sigma_{терм}$ — прямые CD. Электростатическая проводимость $\sigma_{элст.}$ — кривые AB (на рисунке $T_1 < T_2 < T_3$).

Таким образом, с возрастанием температуры тонельный эффект ослабевает, и электростатическая компонента проводимости $\sigma_{элст}$ уменьшается, в то время как термическая компонента $\sigma_{терм}$, как мы знаем, при возрастании температуры, наоборот, резко увеличивается, так что критическая точка E_0' при повышении температуры перемещается в сторону высоких полей (см. рис. 96).

Анализируя формулу (4,16), мы приходим сразу к двум заключениям:

1) можно быть уверенным, что электростатическая ионизация ни в коей мере не является тем механизмом, с помощью которого можно было бы объяснить пулевское возрастание проводимости в сильных полях;

2) в то же время представляется довольно вероятным, что электростатическая ионизация является как раз тем механизмом, который обусловливает собой пробой полупроводника и предшествующее ему катастрофическое увеличение проводимости, наблюдающееся в области предпробивных напряжений [10].

Эти выводы вытекают из следующих соображений:

а) Критическая точка E_0', при которой практически вступает в действие электростатическая ионизация, лежит значительно выше, чем точка E_0, в которой закон Ома сменяется законом Пуля. Действительно, определяя E_0' из условия

$$\sigma_{\text{терм}} = \sigma_{\text{элст}}(E_0'),$$

имеем, согласно (4,17):

$$\left(\lg\frac{E_0'}{x_i} + 3{,}2 - \lg\sigma_{\text{терм}}\right) E_0' = 5\cdot 10^8 x_i^2.$$

Полагая здесь $x_i = 0{,}5$ eV и $\sigma_{\text{терм}} \geqslant 10^{-7}$ ом$^{-1}$ см$^{-1}$, получаем:

$$E_0' \geqslant 100\,\frac{\text{kV}}{см}, \tag{4,18}$$

в то время как для E_0 имеем [см. (1,6)]:

$$E_0 \approx 1 - 10\,\frac{\text{kV}}{см}.$$

Таким образом, в области закона Пуля, т. е. в области напряжений $E = 1 - 50\,\frac{\text{kV}}{см}$, электростатическая компонента проводимости практически отсутствует. Действительно, в этой области напряжений, согласно (4,17):

$$\sigma_{\text{элст}} < 10^{-25}\ \text{ом}^{-1}см^{-1}.$$

б) Механизм электростатической ионизации, коль скоро он вступает в игру, вызывает гораздо более быстрое возрастание проводимости, чем это наблюдается в области закона Пуля. Действительно, разлагая показатель в (4,17) в ряд по степеням $(E - E_0')$, получим

$$-5\cdot 10^8\,\frac{x_i^2}{E} = b' + a'(E - E_0') + \ldots, \tag{4,19}$$

где

$$\begin{cases} b' = -\dfrac{5\cdot 10^3 x_i^2}{E_0'}, \\ a' = +\dfrac{5\cdot 10^3\cdot x_i^2}{E_0'^2}. \end{cases} \tag{4,20}$$

Полагая $x_i = 0{,}5$ eV и, в соответствии с (4,18), $E_0' = 125\,\frac{\text{kV}}{см}$, находим:

$$a' = 10\,\frac{см}{\text{kV}},$$

в то время как для соответствующего коэффициента в законе Пуля (1,2) имели:

$$a \approx 0{,}01\,\frac{см}{\text{kV}}.$$

в) Коэффициент a' в разложении (4,19) возрастает с температурой, как это видно из (4,20) (ибо с возрастанием температуры растёт x_i), в то время как коэффициент a в законе Пуля с возрастанием температуры убывает [см. (1,4)].

г) Ток в полупроводнике в результате электростатической ионизации возрастает и быстро достигает таких значений, которые заведомо должны привести к расплавлению решётки, т. е. к пробою. Действительно, на основании (4,17) получаем:

$$i = 1{,}55\cdot\frac{E^2}{x_i}\cdot 10^{-5\cdot 10^3\cdot\frac{x_i^2}{E}}\ \text{amp}/см^2.$$

Полагая здесь $x_i = 0{,}5$ eV, будем иметь:

$$\begin{cases} \text{при } E = 100\,\frac{\text{kV}}{см}\ldots i = 2{,}4\cdot 10^{-9}\ \text{amp}/см^2, \\ \text{при } E = 500\,\frac{\text{kV}}{см}\ldots i = 0{,}6\cdot 10^3\ \text{amp}/см^2. \end{cases}$$

Мы видим, что пробивное напряжение $E_{\text{пр}}$ заведомо попадает внутрь этого интервала, т. е.

$$100\,\frac{\text{kV}}{см} < E_{\text{пр}} < 500\,\frac{\text{kV}}{см},$$

что находится в согласии с экспериментальными данными. За-

метим, что сравнительно медленное возрастание проводимости на участке Пуля само по себе, т. е. без помощи электростатической ионизации или какого-либо другого механизма, не может привести к пробою. Действительно, как мы уже отмечали, экстраполируя закон Пуля в область пробивных напряжений, мы получаем токи, далеко не достаточные для того, чтобы вызвать пробой.

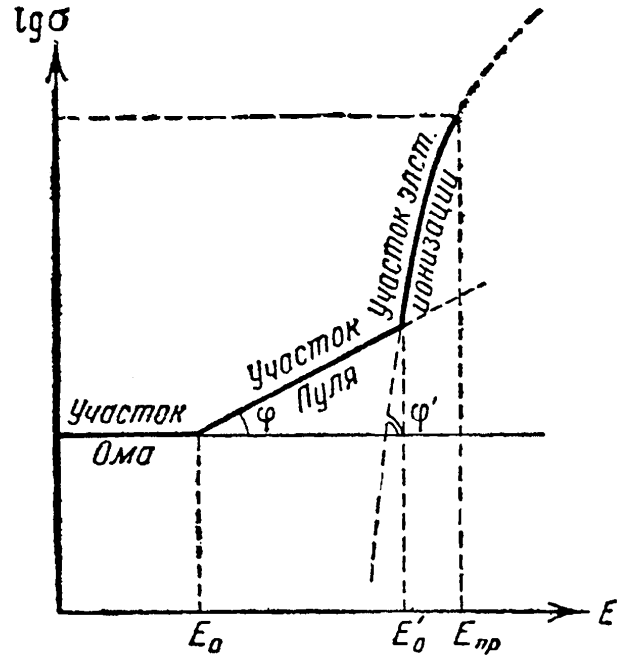


Рис. 97. Роль электростатической ионизации в проводимости полупроводника.

Все эти факты, вытекающие из рассмотрения формулы (4,17), приводят к заключениям, сформулированным выше. Это иллюстрируется рисунком 97. Мы должны различать три области проводимости, показанные на рис. 97:

1) область закона Ома,
2) область закона Пуля,
3) предпробивная область.

Возрастание проводимости на участке Пуля не имеет непосредственного отношения к пробою. Пробой подготовляется и осуществляется иным механизмом. С другой стороны, электростатическая ионизация не имеет никакого отношения к участку Пуля.

$$\begin{cases} E_0 \approx 1 - 10 \dfrac{\text{kV}}{см}, \\ E_0 \geqslant 100 \dfrac{\text{kV}}{см}, \\ E_{пр} \leqslant 500 \dfrac{\text{kV}}{см}. \end{cases}$$

$$\begin{cases} \operatorname{tg}\varphi = a \approx 0{,}01 \dfrac{см}{\text{kV}}, \\ \operatorname{tg}\varphi' = a' \approx 10 \dfrac{см}{\text{kV}}. \end{cases}$$

Электростатический ток практически равен нулю вплоть до полей, близких к пробивным. Однако коль скоро поле достигает предпробивных значений, электростатический ток с электродов внезапно делается существенным.

Если при этом $x_1 = x_2$, то оба электрода вступают в игру одновременно. Роли катода и анода в этом случае совершенно симметричны. Электронный ток, текущий с катода по зоне проводимости, равен дырочному току, текущему из анода по

нормальной зоне. Здесь обе зоны принимают одинаковое участие в проводимости.

Если же $x_1 \neq x_2$, то симметрия между катодом и анодом нарушается. При $x_1 < x_2$ переход электронов из катода в зону проводимости полупроводника оказывается облегчённым, а их переход из нормальной зоны на анод — затруднённым. В этом случае руководящая роль принадлежит катоду. Пробой назревает в зоне проводимости и обусловливается бурным выделением электронов из катода в полупроводник. Нормальная зона при этом остаётся пассивной. При $x_2 < x_1$ переход электронов из катода в зону проводимости оказывается затруднённым, а их переход из нормальной зоны на анод — облегчённым. В этом случае руководящая роль передаётся аноду. Пробой развивается в «подземной» области за счёт катастрофического опустошения нормальной зоны. Зона проводимости при этом остаётся безучастной.

§ 5. Влияние поля на термическую ионизацию

А. Механизм штаркэффекта. В предыдущем параграфе мы рассмотрели два механизма ионизации полем: механизм *ударной ионизации* (ионизация столкновениями) и механизм *электростатической ионизации* (тоннельный эффект). В обоих этих механизмах поле выступает в качестве самостоятельного ионизирующего агента. Оба эти механизма делаются возможными благодаря наклонению зон.

Здесь мы рассмотрим ещё один механизм ионизации, в котором ионизирующим агентом является температура, а поле выступает в качестве помощника. Здесь вспомогательная роль поля делается возможной благодаря штарковскому сдвигу уровней.

Штарковское расщепление термов вызывает следующую деформацию в структуре энергетического спектра:

а) Энергетические зоны испытывают сдвиг, вообще говоря, различный для различных зон. Зоны, соответствующие вырожденным термам, кроме того, испытывают расширение. Это расширение надо понимать как раздвижение отдельных зон вырожденного терма, наложенных друг на друга при отсутствии поля.

б) Локальные уровни испытывают обычное штарковское расщепление. Это значит, что связанный электрон получает

возможность выбирать между несколькими энергетически различными состояниями. Наинизшее из этих состояний является нормальным состоянием для связанного электрона. Таким образом, среди всех компонент штарковского расщепления практический интерес представляет крайняя (наинизшая) из компонент и её сдвиг относительно исходного уровня.

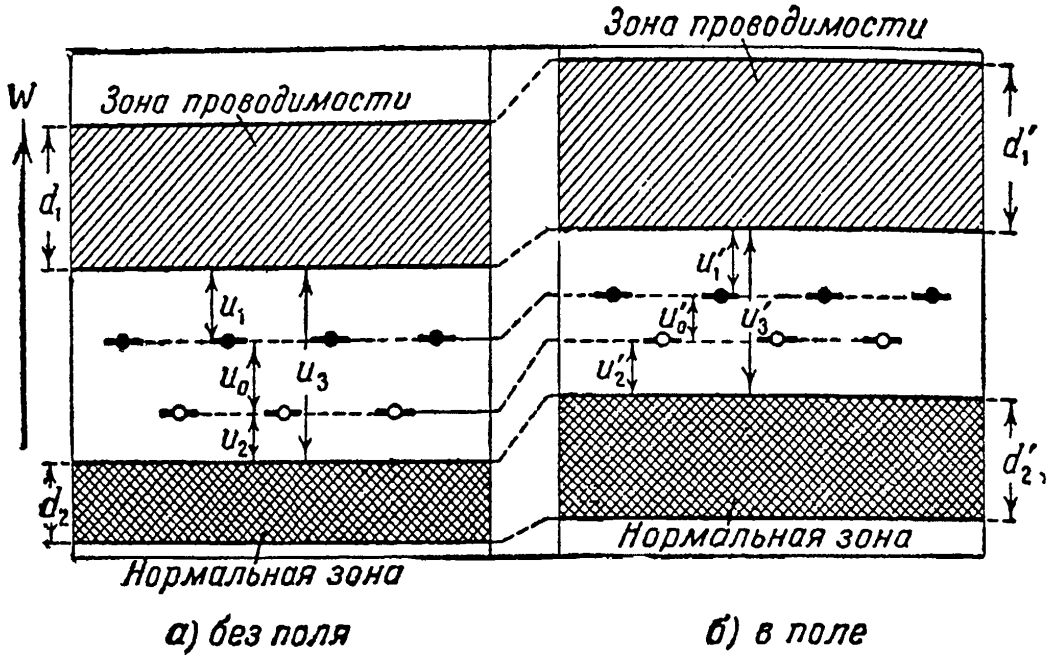


Рис. 98. Штарковский сдвиг уровней в энергетическом спектре решётки. Здесь $u_i' = u_i - \Delta u_i$, $d_i' = d_i + \Delta d_i$.

Вследствие деформации энергетического спектра изменяются расстояния между уровнями. Примем следующие обозначения:

расстояния между уровнями при отсутствии поля:

$$u_1,\ u_2,\ u_3;$$

расстояния между теми же уровнями при наличии поля:

$$u_1' = u_1 - \Delta u_1,\quad u_2' = u_2 - \Delta u_2,\quad u_3' = u_3 - \Delta u_3.$$

Смысл этих обозначений поясняется рис. 98. Среди чисел

$$\Delta u_1,\ \Delta u_2,\ \Delta u_3$$

могут быть положительные, отрицательные или равные нулю в зависимости от природы соответствующих уровней.

В случае $\Delta u_i > 0$ мы имеем сближение уровней. В случае $\Delta u_i < 0$ мы имеем раздвижение уровней. Наконец, слу-

чай $\Delta u_i = 0$ означает, что расстояние между уровнями при включении поля остаётся неизменным. Этот последний случай имеет место, например, тогда, когда оба уровня, расстояние между которыми измеряется отрезком u_i, происходят от одного и того же терма одного и того же атома и, следовательно,

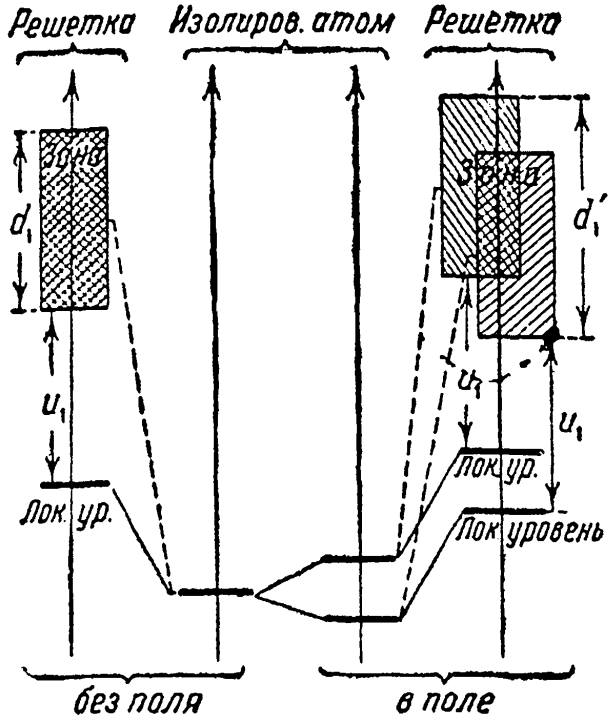


Рис. 99. Штарковский сдвиг в случае, когда зона проводимости и локальный уровень происходят от одного и того же терма изолированного атома.

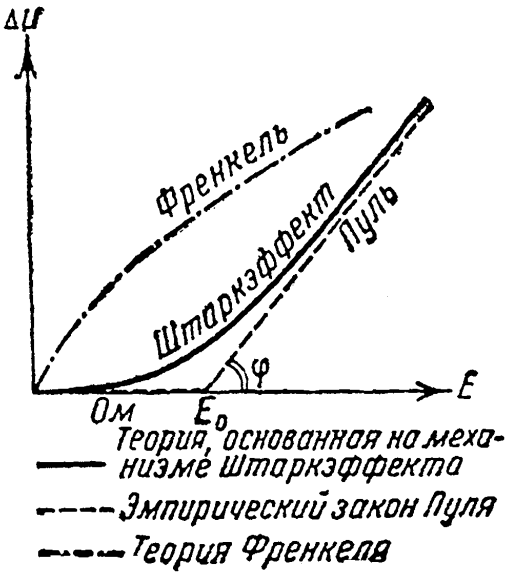


Рис. 100. Уменьшение работы ионизации под влиянием поля.

претерпевают одинаковое штарковское расщепление. Такой случай изображён на рис. 99.

Величина Δu_i (в случае $\Delta u_i \neq 0$) монотонно возрастает с полем. Зависимость Δu_i от E может быть выражена эмпирической формулой [13]:

$$\Delta u_i = a_i E^2 + b_i E. \qquad (5;1)$$

Здесь параметры a_i и b_i сами медленно меняются с полем, причём при возрастании E отношение a_i/b_i убывает, так что

$$\begin{cases} \text{в «слабых» полях} \ldots\ldots\ldots a_i \gg b_i, \\ \text{в «сильных» полях} \ldots\ldots\ldots a_i \ll b_i. \end{cases}$$

При $E \ll 1$ kV/см можно пренебречь вторым членом в (5,1) по сравнению с первым и коэффициент a_i считать постоян-

ным. В этой области мы имеем чисто квадратичный эффект. При $E \gg 1\ \mathrm{kV/см}$, наоборот, можно пренебречь первым членом по сравнению со вторым и считать коэффициент b_i постоянным. В этой области мы получаем чисто линейный эффект. При полях $E \approx 1 - 100\ \mathrm{kV/см}$ величина Δu_i достигает по порядку величины значений $\Delta u_i \approx 10^{-2} - 10^{-3}\ \mathrm{eV}$.

Зависимость (5,1) изображена на рис. 100 жирной кривой. Эта кривая с большей или меньшей точностью может быть апроксимирована следующим образом:

$$\Delta u_i = a_i E^2 + b_i E = \begin{cases} \approx c_i (E - E_0) \ldots \text{при } E \geqslant E_0, \\ \approx 0 \ldots\ldots\ldots \text{при } E \leqslant E_0, \end{cases} \tag{5,2}$$

где, очевидно:

$$\left\{ \begin{array}{l} E_0 \approx 1\ \mathrm{kV/см}, \\ c_i \approx 10^{-3} - 10^{-4} \dfrac{\mathrm{eV} \cdot \text{см}}{\mathrm{kV}}. \end{array} \right\} \tag{5,3}$$

Апроксимированная кривая изображена на том же рис. 100 пунктиром.

Посмотрим теперь, как сказывается учёт штаркэффекта на термической и электростатической компонентах проводимости. Вернёмся к выражению (4,14):

$$\sigma = \sigma_{\text{терм}} + \sigma_{\text{элст}}. \tag{5,4}$$

На основании (2,7) главы IV и (4,16) мы имеем:

$$\left\{ \begin{array}{l} \sigma_{\text{терм}} = A_1 e^{-\frac{x_1}{kT}} + A_2 e^{-\frac{x_2}{kT}}, \\ \sigma_{\text{элст}} = B_1 e^{-C\frac{x_1^2}{E}} + B_2 e^{-C\frac{x_2^2}{E}}, \end{array} \right. \tag{5,5}$$

где, согласно (1,7а), (1,14 а) главы IV и (4,16):

$$\left\{ \begin{array}{l} A_i = \dfrac{4}{3} \dfrac{e g_i l_i (T)}{h^3} (2\pi m_i kT), \\ B_i = \dfrac{e^2 E}{8\pi h x_i}, \\ C = \dfrac{\pi^2 ma}{h^2 e}, \end{array} \right. \tag{5,6}$$

где $i = 1,2$. При $x_1 > x_2$ вторым членом в (5,5) можно пренебречь по сравнению с первым; при $x_2 > x_i$, наоборот, пер-

вым членом можно пренебречь по сравнению со вторым. Отсюда следует, что термическая и электростатическая компоненты проводимости имеют всегда одинаковую природу, т. е. проводимость остаётся электронной или дырочной, или смешанной на всём участке от $E=0$ до $E=E_{\text{пр}}$ (см. рис. 97).

Точное выражение для x_i (в рамках тех предположений, которые были сделаны в § 2 главы IV) даётся формулой (2,11) гл. IV. Мы имеем:

$$x_i = x_i(\nu, \mu, T, u_1, u_2, u_3). \tag{5,7}$$

При включении поля x_i делается функцией от E. Мы должны теперь заменить x_i через x_i', где

$$x_i' = x_i(\nu, \mu, T, u_1', u_2', u_3'). \tag{5,8}$$

В отдельных температурных областях, которые в главе IV мы обозначали как области

$$a, b, c, d, e, f$$

и границы которых определяются значениями параметров μ и ν, точное выражение (2,11) главы IV для x_i может быть, как мы видели, апроксимировано формулами (2,24) и (2,26) главы IV. Внутри этих температурных областей зависимость x_i от E выражается особенно просто; вместо (5,8) мы получаем:

$$x_i' = x_i - \Delta u, \tag{5,9}$$

где, согласно (2,24) и (2,26) главы IV:

$$\Delta u = \begin{cases} = \frac{1}{2}\Delta u_3 \ldots\ldots \text{в области } a \\ = \frac{1}{2}(\Delta u_0 + \Delta u_i) \ldots \text{в области } b \\ = \frac{1}{2}(\Delta u_0 + 2\Delta u_i) \ldots \text{в области } c \end{cases} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \begin{array}{l} \text{где} \\ i=1, \text{если } u_1 < u_2 \\ i=2, \text{если } u_2 < u_1 \end{array}$$

$$\Delta u = \begin{cases} = 0 \ldots\ldots\ldots \text{в области } d \\ = \Delta u_i \ldots\ldots\ldots \text{в области } e \\ = \Delta u_i \ldots\ldots\ldots \text{в области } f \end{cases} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \begin{array}{l} \text{где} \\ i=1, \text{если } \nu \geqslant 0 \\ i=2, \text{если } \nu \leqslant 0. \end{array} \tag{5,10}$$

Таким образом, внутри этих областей штаркэффект может быть учтён путём введения в (5,4) коэффициентов η и ζ:

$$\sigma = \eta\sigma_{\text{терм}} + \zeta\sigma_{\text{элст}}, \tag{5,11}$$

где, согласно (5,6) и (5,9):

$$\begin{cases} \eta = e^{\frac{\Delta u}{kT}}, \\ \zeta = e^{C\frac{(2x_i - \Delta u)\,\Delta u}{E}}. \end{cases} \tag{5,12}$$

Полагая здесь

$$\begin{cases} x_i \approx 0{,}5\ \text{eV}, \\ T \geqslant 100^\circ\ \text{K}, \\ E \leqslant 100\,\frac{\text{kV}}{см}, \\ \Delta u \leqslant 0{,}01\ \text{eV}, \end{cases}$$

мы будем иметь

$$\begin{cases} 1 \leqslant \eta \leqslant 3, \\ 1 \leqslant \zeta \leqslant 3. \end{cases}$$

Отсюда мы заключаем, что в области термической ионизации учёт штаркэффекта может играть существенную роль, и притом бóльшую, чем больше температура, приводя к росту термической проводимости с полем. В области электростатической ионизации учёт штаркэффекта играет малосущественную роль, проводя к незначительному снижению точки E_0', в которой электростатическая ионизация вступает в действие, и к несколько более значительному снижению точки $E_{\text{пр}}$, т. е. сужая область электростатической ионизации в основном за счёт понижения пробивного напряжения. При пренебрежении эффектом Штарка мы имеем $\eta = \zeta = 1$.

Для области термической ионизации, пренебрегая в (5,11) вторым членом по сравнению с первым, мы можем написать, учитывая (5,1):

$$\sigma(E) = \begin{cases} = \sigma(0) \ldots\ldots\ldots \text{при } E < E_0, \\ = \sigma(0)\, e^{a(E - E_0)} \ldots\ldots \text{при } E > E_0, \end{cases} \tag{5,13}$$

где $\sigma(0)$ — электропроводность при отсутствии поля. Формула (5,13) совпадает с эмпирическим законом Пуля [см. (1,1) и (1,2)], причём параметры a и E_0 в (5,13) имеют правильный порядок величины. Действительно, как мы видели [см. (5,3)], $E \approx 1\ \text{kV}/см$, что находится в соответствии с (1,6).

С другой стороны, на основании (5,10), (5,11) и (5,2):

$$a \approx \frac{c}{kT} = \frac{a}{T}, \tag{5,14}$$

где, согласно (5,2):

$$a = \frac{c}{k} \approx 1 - 10 \frac{см \text{ абс. град.}}{\text{kV}},$$

что находится в соответствии с экспериментальным значением для a [см. (1,5)]. Различие между приближённой теоретической формулой (5,13), вытекающей из учёта штаркэффекта, и эмпирическим законом Пуля (1,2) заключается в том, что в теоретическом выражении (5,14) для a как функции T отсутствует свободный член β, фигурирующий в эмпирической формуле (1,4). Однако было бы преждевременно в настоящее время расценивать это различие как аргумент против теории, ибо эмпирическая формула (1,4) базируется на недостаточно ещё богатом экспериментальном материале; она проверена лишь для двух образцов, исследованных Иоффе (Cu_2O и V_2O_5, см. рис. 82).

Итак, учёт штаркэффекта состоит в том, что величины u_i (где $i = 1, 2, 3$) считаются, вообще говоря, не постоянными, а рассматриваются как некие функции поля:

$$u_i(E) = u_i(0) - \Delta u_i(E),$$

где Δu_i, в случае $\Delta u_i > 0$, представляет собой сужение соответствующего запрещённого участка, возрастающее по мере возрастания E. Поскольку любой механизм ионизации содержит в своих формулах величины u_i, то, полагая в этих формулах u_i зависящим от E, мы тем самым учитываем роль штаркэффекта для этих механизмов. В некоторых случаях, как мы видели, это может дать существенный эффект.

В. Деформация полос поглощения вследствие штаркэффекта. Деформация в структуре энергетического спектра, вызываемая штаркэффектом, должна отражаться, конечно, на структуре оптического спектра, приводя к деформации примесных и собственных полос поглощения. На первый взгляд может показаться, что при сужении запрещённого участка между нормальной зоной и зоной проводимости (случай $\Delta u_3 > 0$) красная граница собственной полосы должна сдвигаться в крас-

ную сторону. Однако сужение запрещённого участка отнюдь не всегда обязано сопровождаться сдвигом красной границы в красную сторону. При $\Delta u_s > 0$ собственная полоса может испытывать деформацию различного характера, как это изображено на рис. 102, в зависимости от того, какова природа нормальной зоны и зоны проводимости. В этом легко убедиться, учитывая те правила отбора, которыми регулируются оптические перебросы электронов из одной зоны в другую. Согласно этим правилам, с каждого данного уровня данной зоны возможен переход лишь на один вполне определённый уровень другой зоны, так что если все уровни внутри каждой зоны

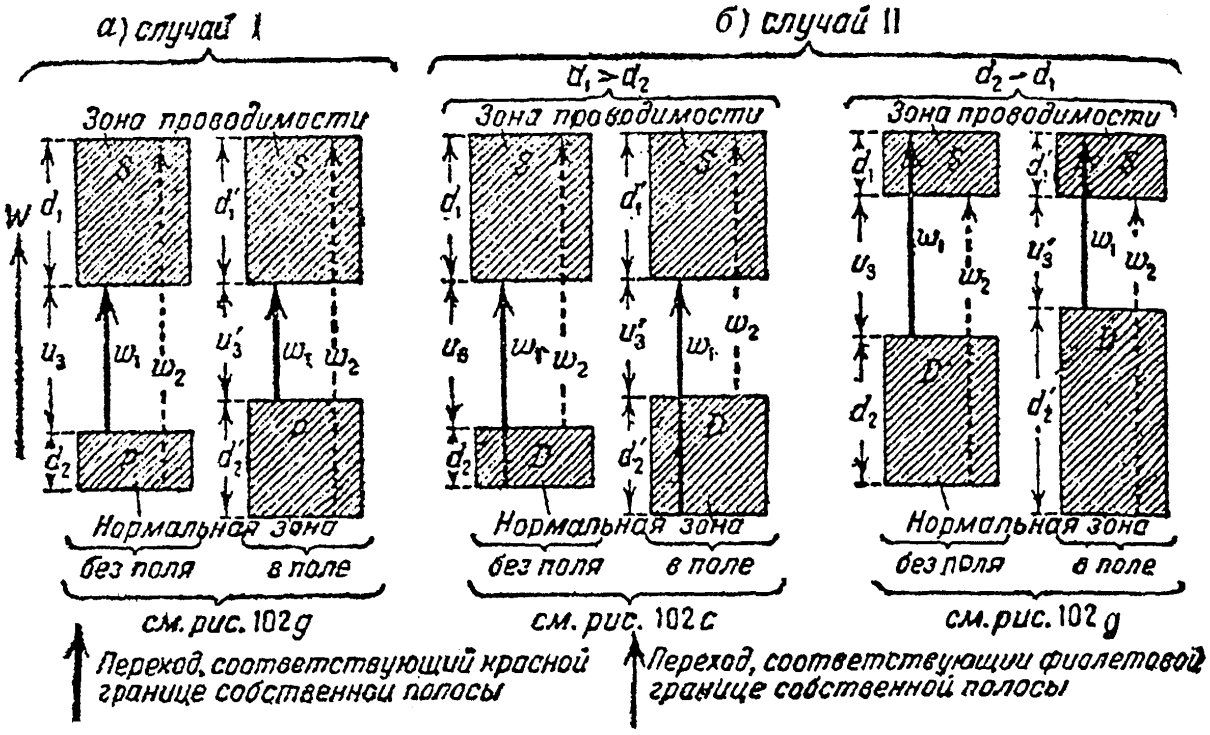


Рис. 101. Происхождение собственной полосы поглощения. На рисунке зона проводимости является s-зоной; нормальная зона — p-зоной или d-зоной; локальные уровни на рисунке не изображены.

считать перенумерованными, то оптический переход может осуществляться лишь между двумя уровнями с одинаковыми номерами. Существенно при этом, что нумерация уровней должна проводиться для одних зон снизу вверх, а для других — сверху вниз, в зависимости от природы зоны. Соответственно этому следует различать два случая.

I. Случай, когда в нижней и в верхней зонах нумерация проведена в противоположных направлениях. В этом случае с верхней границы нормальной зоны электрон переходит на

нижнюю границу зоны проводимости, а с нижней границы нормальной зоны на верхнюю границу зоны проводимости (рис. 101, *а*).

II. Случай, когда нумерация уровней в обеих зонах проведена в одном и том же направлении. В этом случае с верхней границы нормальной зоны электрон переходит на верхнюю границу зоны проводимости, а с нижней границы нормальной зоны на нижнюю границу зоны проводимости (рис. 101, *б*).

В зависимости от природы обеих зон мы будем иметь дело с тем или иным из этих случаев, как это показано в следующей табличке:

Нормальная зона	Зона проводимости *s*	*p*	*d*	*f*
s	II	I	II	I
p	I	II	I	II
d	II	I	II	I
f	I	II	I	II

Сформулированные здесь правила отбора справедливы, как известно, при отсутствии внешнего поля. Однако нет оснований думать, что эти правила отбора снимаются внешним полем. Напротив того, следует ожидать, что они будут сохраняться при включении внешнего поля, ибо внешние поля, с которыми мы практически имеем дело, значительно слабее тех внутренних электрических полей, которые, осуществляясь в решётке, тем не менее обеспечивают выполнение этих правил.

Обозначим ширину зоны проводимости и нормальной зоны, соответственно, через

$$\begin{cases} \text{при отсутствии поля. . . .} & d_1,\ d_2, \\ \text{при наличии поля . . . . .} & d_1',\ d_2', \end{cases}$$

где $d_i' = d_i + \Delta d_i$, причём $\Delta d_i \geqslant 0$. Обозначим красную и фиолетовую границы собственной полосы поглощения, обусловленной перебросом электронов из нормальной зоны в зону проводимости, соответственно:

$$\begin{cases} \omega_1, \ \omega_2, \ \ldots \ . \text{ при отсутствии поля} \\ \omega_1', \ \omega_2', \ \ldots \ . \text{ при наличии поля} \end{cases}$$

где $\omega_i' = \omega_i + \Delta\omega_i$, причём границы полосы сдвигаются

$$\begin{cases} \text{в красную сторону, если} \quad \Delta\omega_i > 0, \\ \text{в фиолетовую сторону, если} \quad \Delta\omega_i < 0. \end{cases}$$

Очевидно, мы будем иметь:

	В случае I (рис. 101, *а*)		В случае II (рис. 101, *б*)	
	красная граница:	фиолетовая граница:	красная граница:	фиолетовая граница:
При отсутствии поля	$h\omega_1 = u_3$	$h\omega_2 = u_3 + d_k + d_i$	$h\omega_1 = u_3 + d_k$	$h\omega_2 = u_3 + d_i$
При наличии поля	$h\omega_1' = u_3'$	$h\omega_2' = u_3' + d_k' + d_i'$	$h\omega_1' = u_3' + d_k'$	$h\omega_2' = u_3' + d_i'$
Сдвиг границы в поле	$h\Delta\omega_1 = -\Delta u_3$	$h\Delta\omega_2 = \Delta d_k + \Delta d_i - \Delta u_3$	$h\Delta\omega_1 = \Delta d_k - \Delta u_k$	$h\Delta\omega_2 = \Delta d_i - \Delta u_3$

где $d_i \geqslant d_k$, т. е.

$$\begin{cases} i=1, \ k=2 \ \ldots, \text{ если } d_1 \geqslant d_2, \\ i=2, \ k=1 \ \ldots, \text{ если } d_2 \geqslant d_1. \end{cases}$$

Отсюда вытекает рис. 102, на котором представлены различные случаи деформации собственной полосы поглощения, вызванные сужением запрещённого участка между зонами, и выписаны условия, при которых осуществляется каждый из этих случаев. Мы видим, что сдвиг красной границы в сторону длинных волн не является во всех случаях обязательным результатом штаркэффекта. В частности, нет основания ожидать этого эффекта, например, в кристаллах Cu_2O, в которых нормальная зона является *d*-зоной, а зона проводимости *s*-зоной.

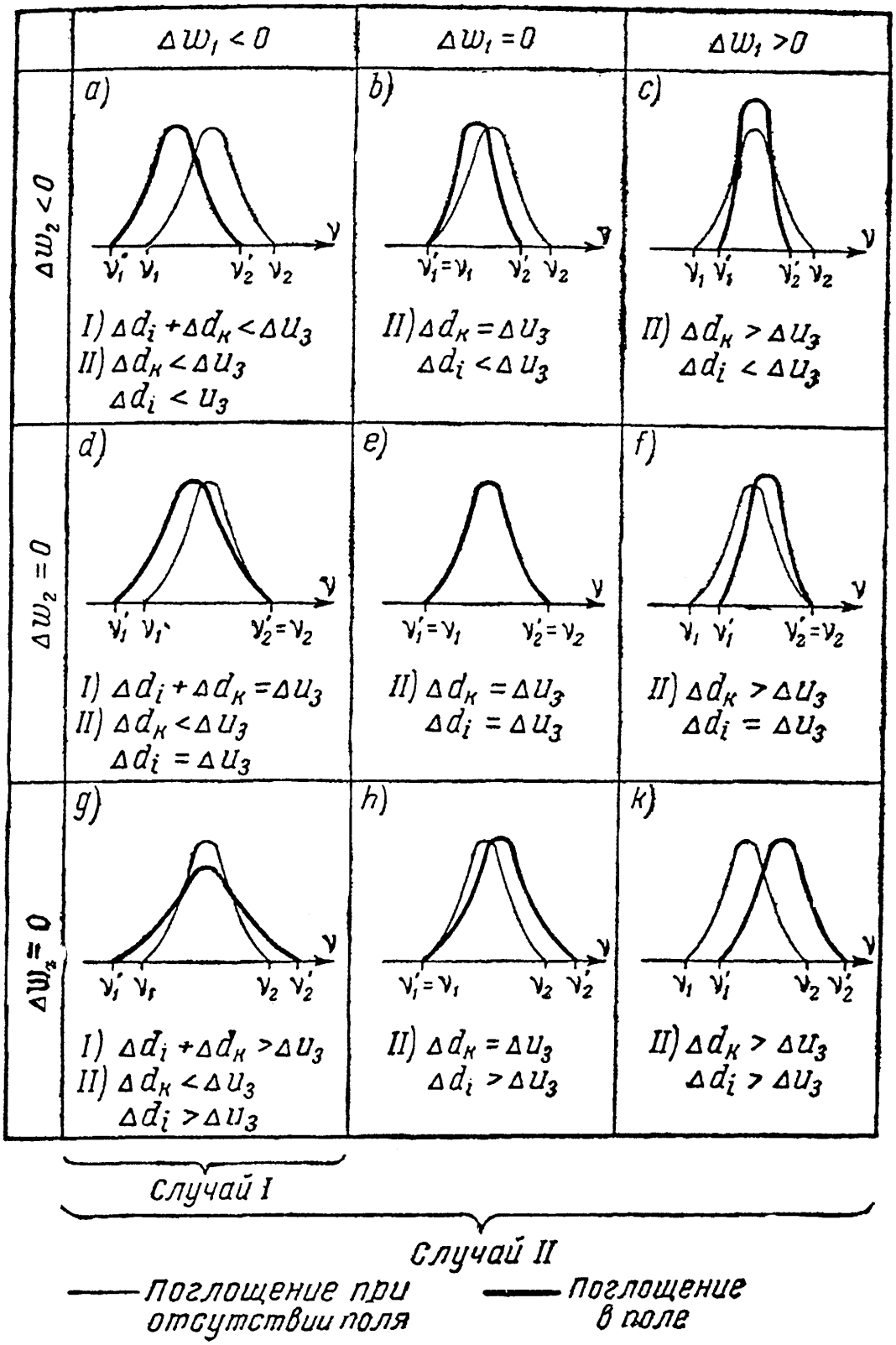


Рис. 102. Различные случаи деформации собственной полосы поглощения (в результате штарковского сужения запрещённого участка между зонами).

Рисунок 102 показывает, как должно отражаться сужение запрещённого участка на структуре собственной полосы. В действительности же влияние поля на оптический спектр кристалла имеет более сложный характер, чем это изображено на рис. 102, и механизм штаркэффекта вряд ли можно считать доминирующим среди тех механизмов, которыми обусловливается это влияние.

С. Механизм Френкеля. В заключение заметим, что облегчение термической ионизации под влиянием поля может происходить не только в результате штаркэффекта, но и за счёт совершенно иного механизма, рассмотренного Я. И. Френкелем [14]. Механизм Френкеля подобен механизму Шоттки, вызывающему понижение работы выхода при термической эмиссии электронов из металлов. Теория Френкеля не выходит за рамки тех модельных представлений, которые были развиты нами в главе I. С точки зрения этих представлений, работа u, необходимая для того, чтобы пересадить электрон с одного иона решётки на другой и развести образовавшиеся таким образом аномальные ионы на бесконечность, в присутствии внешнего электрического поля E, должна уменьшаться на величину

$$\Delta u = 2\sqrt{\frac{e^3 E}{\varepsilon}}, \tag{5,15}$$

где ε — диэлектрическая постоянная кристалла. Таким образом, теория Френкеля приводит к выражению

$$\sigma(E) = \sigma(0)\, e^{a\sqrt{E}} \tag{5,16}$$

вместо выражения (5,13), основанного на механизме штаркэффекта. Здесь

$$a = \frac{e}{kT}\sqrt{\frac{e}{\varepsilon}} = \frac{a}{T}. \tag{5,17}$$

Полагая $\varepsilon \approx 4$, мы будем иметь:

$$a = \frac{e}{k}\sqrt{\frac{e}{\varepsilon}} \approx 60\,\frac{\text{абс. град.}\ см^{1/2}}{\text{kV}^{1/2}}.$$

По поводу теории Френкеля, сравнивая её с экспериментальными данными, и, кроме того, с теорией, основанной на механизме штаркэффекта, можно сделать следующие замечания:

1. Формула Френкеля (5,16) отличается от эмпирического закона Пуля (1,2), во-первых, заменой E на $\sqrt{E}$. Это обстоятельство не следует расценивать как возражение против теории, ибо нельзя сказать, чтобы имеющиеся экспериментальные данные свидетельствовали бы против такой замены.

2. Во-вторых, формула (5,16) отличается от формулы (1,2) отсутствием свободного члена E_0. Это значит, что теория Френкеля не оставляет места для участка Ома, как это видно из рис. 100, на котором зависимость Δu от E, даваемая формулой (5,15), изображена коричневой кривой. Таким образом, критическое напряжение поля E_0, при котором перестаёт действовать закон Ома, не получает никакой интерпретации в теории Френкеля.

3. Коэффициент a в формуле Френкеля (5,16) имеет правильный порядок величины, но в теоретической формуле (5,17), так же, впрочем, как и в формуле (5,14), отсутствует свободный член β, фигурирующий в эмпирической формуле (1,4).

4. Индивидуальность решётки входит в формулу (5,16) только через диэлектрическую постоянную ε, в то время как в теории, основанной на штаркэффекте, индивидуальность атомов, составляющих решётку, содержится уже в величине c, как это видно из (5,2), и через эту величину входит в коэффициент a в окончательной формуле (5,13).

5. Теория Френкеля требует возрастания проводимости с полем ото всех решёток и во всех случаях, в то время как с точки зрения теории, основанной на штаркэффекте, возможно существование таких температурных областей, определяемых концентрацией примесей, содержащихся в образце, внутри которых пулевское возрастание проводимости вообще должно отсутствовать. В действительности среди данных Иоффе мы имеем случаи, когда отклонение от закона Ома в сильных полях вовсе не было обнаружено (см., например, кривую для образца № 119 Cu_2O на рис. 85).

6. Френкель рассматривает свой механизм как механизм, приводящий при достаточно сильных полях к пробою. В действительности же, если экстраполировать закон Френкеля (5,16) в область пробивных напряжений, то мы не получаем токов, достаточных для пробоя, так же как мы не получаем пробоя, экстраполируя в область пробивных напряжений эмпирический закон Пуля.

Литература к главе V

[1] А. В. и А. Ф. Иоффе. *ЖЭТФ*, **9**, 1428, 1939; **9**, 1451, 1939.
[2] Braunbeck, *Zs f. Phys.* **48**, 684, 1927.
[3] Jones and Zener, *Proc. of R. Soc.*, A, **144**, 101, 1934.
[4] Zener, *Proc. of R. Soc.*, A, **145**, 523, 1934.
[5] Hill, *Acta Mat.*, **8**, 27, 1886.
[6] Ф. Ф. Волькенштейн, *Труды ФИАН СССР*, I, вып. 2, стр. 123, 1935.
[7] Ф. Ф. Волькенштейн, *ЖЭТФ*, вып. 5, 583, 1935.
[8] Fröhlich, *Proc. of R. Soc.*, A, **160**, 230, 1937.
[9] Писаренко, *Изв. АН, СССР*, 631, 1938.
[10] Ф. Ф. Волькенштейн, *ЖТФ*, **9**, 171, 1939.
[11] Fowler, *Proc. of R. Soc.* A, **141**, 56, 1933.
[12] Зоммерфельд-Бете, Электронная теория металлов, ОНТИ, 1938, стр. 109 и дальше.
[13] Stark, *Handb. d. Exp. Phys.*, B. 21, стр. 463—502.
[14] Я. И. Френкель, *ЖЭТФ*, **8**, 1292, 1938.

ГЛАВА VI

ВЛИЯНИЕ ОСВЕЩЕНИЯ НА ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

§ 1. Нормальный фотоэлектрический эффект

А. Экспериментальные данные. Как мы видели, проводимость полупроводников, как правило, очень чувствительна к температуре и к концентрации примесей, содержащихся в решётке. Освещение также является фактором, на который проводимость полупроводников реагирует обычно чрезвычайно резко. Влияние освещения на проводимость неметаллических кристаллов в настоящее время изучено весьма подробно в отношении щёлочно-галоидных кристаллов. Работы школы Поля [1] за границей и покойного П. С. Тартаковского [2] и его учеников в СССР предоставляют в руки теоретика очень богатый и совершенно отчётливый экспериментальный материал. Иначе обстоит дело в отношении полупроводников. Экспериментальный материал здесь ещё очень беден, во многом противоречив и совершенно недостаточен для построения теории. Тем не менее, можно указать на некоторые фотоэлектрические особенности полупроводников, которые уже сейчас (во всяком случае с качественной стороны) могут быть предметом теоретического исследования.

В первую очередь внимание теоретика должно быть привлечено к самому тому факту, что один и тот же полупроводник в различных условиях или различные полупроводники в одних и тех же условиях могут реагировать на освещение существенно различным образом: как правило, поглощение света вызывает значительное увеличение проводимости, причём зависимость проводимости от различных факторов оказывается при освещении, вообще говоря, иной, чем в темноте; в некоторых же случаях даже весьма сильное поглощение света

не вызывает сколько-нибудь заметного изменения в величине проводимости; наконец, известны случаи, когда под влиянием освещения проводимость полупроводника уменьшается. Таким образом, следует различать следующие случаи:

$$\begin{cases} 1)\ \Delta\sigma > 0 \ldots \text{положительный (нормальный) фотоэффект,} \\ 2)\ \Delta\sigma \approx 0 \ldots \text{фотоэлектрически неактивное поглощение,} \\ 3)\ \Delta\sigma < 0 \ldots \text{отрицательный (аномальный) фотоэффект,} \end{cases}$$

где

$$\Delta\sigma = \sigma - \sigma_0,$$

$$\begin{cases} \sigma \text{ — проводимость при наличии освещения,} \\ \sigma_0 \text{ — темновая проводимость при тех же условиях.} \end{cases}$$

Отметим некоторые характерные черты для каждого из этих трёх принципиально различных случаев.

1. При положительном фотоэффекте световая проводимость σ возрастает вместе с количеством поглощённых квантов Q, причём зависимость σ от Q имеет вид:

а) $\sigma = aQ$. Этот закон имеет место, например, в Cu_2O по наблюдениям А. В. и А. Ф. Иоффе [3] и в ряде других случаев;

б) $\sigma = \beta\sqrt{Q}$. Этот закон имеет место, например, в M_0S_2 [4], в Sb_2S_3 [5] и в некоторых других материалах.

Световая проводимость σ зависит от напряжённости электрического поля E; характер зависимости такой же, как и для темновой проводимости σ_0; однако σ возрастает с полем гораздо медленнее, чем σ_0 при тех же условиях. Соответствующие кривые, заимствованные из работы Иоффе [3], приведены на рис. 103.

Световая проводимость σ зависит также от температуры. Форма зависимости весьма сложна и различна для различных случаев, так что нельзя установить сколько-нибудь общих закономерностей (см., например, [6]).

Световая добавка $\Delta\sigma$, как правило, не зависит ни от E ни от T. В отдельных случаях, однако, в особенности в полупроводниках, обладающих значительными стехиометрическими нарушениями, $\Delta\sigma$ медленно растёт с E, подчиняясь закону Пуля, как это видно, например, на рисунке 103, *г*

Сложность всех этих зависимостей, являющаяся отличительной чертой полупроводников по сравнению со щёлочно-

галоидными кристаллами, есть результат наложения различных вторичных явлений, которые часто совершенно маскируют основной фотоэлектрический эффект.

2. Для фотоэлектрически неактивного поглощения света характерно то, что это явление обусловливается не природой полупроводника, а природой поглощаемого света. Так что в одном и том же образце и при одних и тех же условиях одни полосы поглощения обладают значительным фото-

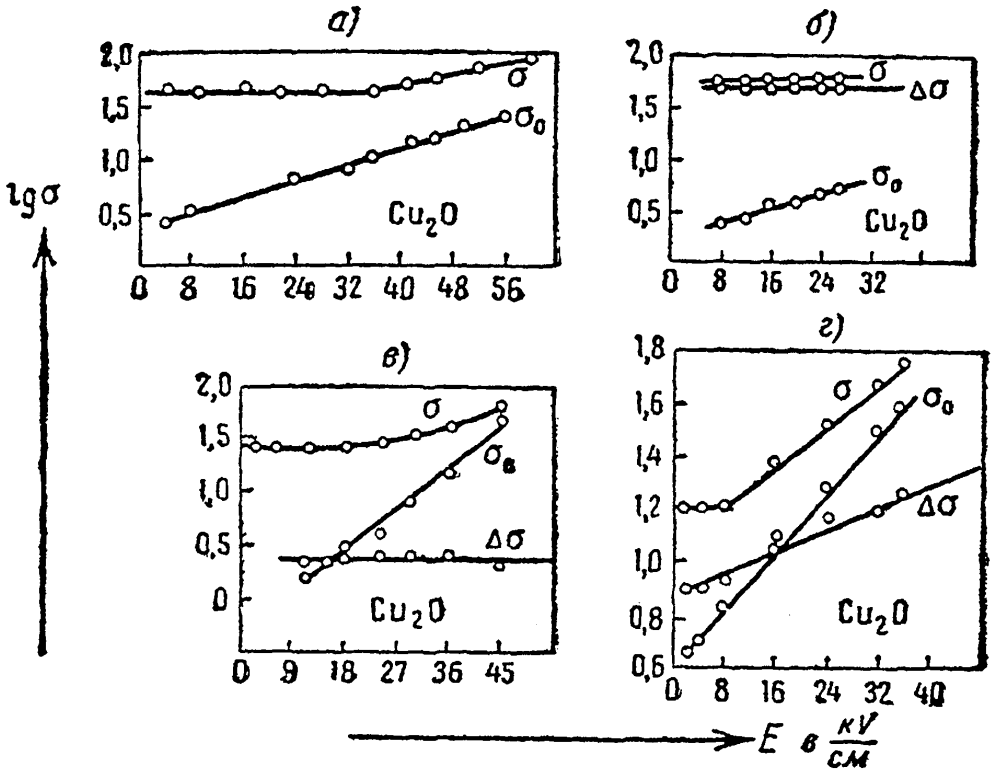


Рис. 103. Влияние электрического поля на световую и темновую проводимости для различных образцов закиси меди (по данным Иоффе).

эффектом, в то время как в других полосах фотоэффект практически совсем отсутствует. Такое фотоэлектрически неактивное поглощение наблюдается как в полупроводниках (например, V_2O_5), так и в щёлочно-галоидных кристаллах, в которых при сильном фотоэффекте в F-полосе, в U-полосе и в собственной полосе поглощение света не вызывает сколько-нибудь заметной проводимости. Фотоэлектрически неактивное поглощение не укладывается в элементарную вильсоновскую схему полупроводника. Этот факт послужил поводом для развития Я. И. Френкелем представлений об экситоне.

3. Отрицательный фотоэффект наблюдался в ряде полупроводников (селен — Рис [7] и Браун [8]; стибнит — Эллиот [5]; молибденит — Кобленц [9], закись меди — Иоффе [3]). Результаты исследований весьма разноречивы. Общим, однако, является тот факт, что отрицательный фотоэффект, если вообще он имеет место, наблюдается лишь при не слишком низких температурах и в достаточно сильных электрических полях. При понижении температуры или при понижении поля (при неизменных прочих условиях) отрицательный эффект исчезает и обращается в нормальный эффект. Критическая температура, при которой фотоэффект меняет знак, лежит тем ниже, чем выше E. Отрицательный фотоэффект теоретически ещё не был интерпретирован. Принято считать, что это явление непосредственно не связано с механизмом поглощения света, а является результатом тех или иных вторичных эффектов.

С теоретической точки зрения положительный фотоэффект является в достаточной мере тривиальным, если не пытаться объяснить отдельные частные закономерности, что в настоящее время следует считать преждевременным. Что же касается фотоэлектрически неактивного поглощения и отрицательного фотоэффекта, то сам факт существования подобных явлений представляет несомненный теоретический интерес, если только допустить, что эти явления обусловлены некиим самостоятельным механизмом, а не являются результатом вторичных процессов, накладывающихся на нормальный фотоэлектрический эффект.

В. Интерпретация полос поглощения. Если свет поглощается кристаллической решёткой, то это свидетельствует о переводе решётки в возбуждённое состояние. Среди таких возбуждённых состояний, как мы видели (см. главу I), следует различать состояния проводящие и непроводящие. Таким образом, поглощение света отнюдь ещё не означает создания в решётке свободных носителей тока. Действительно, каждый акт поглощения кванта сопровождается перебросом электрона с одного иона на другой. В идеальной гетерополярной решётке это приводит к возникновению двух аномальных ионов с противоположными по знаку дополнительными зарядами. До тех пор, пока эти ионы остаются связанными между собой, они не принимают участия в проводимости. Такие пары связанных аномальных ионов, представляющие собой электри-

чески нейтральные образования, являются в гетерополярной решётке аналогами экситонов, представление о которых применительно к гомеополярным решёткам было развито Я. И. Френкелем. Реальные гетерополярные решётки отличаются от идеальных, как мы видели, тем, что уже с самого начала они могут обладать некоторым запасом аномальных ионов, что является необходимым для обеспечения электрической нейтральности решётки. В нормальном состоянии эти аномальные ионы локализованы около мест повреждений. В таких реальных решётках поглощение света может и не сопровождаться возникновением новых пар аномальных ионов, а может быть связано с расширением области локализации для уже существующих аномальных ионов. Однако и в этом случае аномальный ион в совокупности с повреждением, около которого он локализован, составляет электрически нейтральное образование, не принимающее участия в проводимости. Электронная (дырочная) проводимость обусловливается лишь свободными аномальными ионами, т. е. такими, взаимодействием которых с локальными повреждениями или с другими аномальными ионами можно пренебречь. С другой стороны, локальные повреждения (пустые узлы или междуузельные ионы), лишённые нейтрализующих их аномальных ионов, делаются источниками ионной проводимости.

Фотоэлектрически неактивное поглощение света естественно связывать именно с такими непроводящими возбуждёнными состояниями решётки. Однако это не является необходимым для объяснения фотоэлектрически неактивного поглощения. Можно показать (см. § 2), что свет может переводить электроны на уровни проводимости и, тем не менее, при определённых условиях, этот процесс может и не вызывать сколько-нибудь заметного увеличения проводимости, а в некоторых случаях (см. § 3) может приводить даже к снижению проводимости. Перевод электронов в проводящее состояние, не сопровождающийся увеличением проводимости, может быть обусловлен, как мы увидим ниже, изменениями в условиях рекомбинации, происходящими под действием света.

Если игнорировать непроводящие возбуждённые состояния, то мы можем вернуться к энергетической схеме, с которой мы оперировали в предыдущих главах и которая ещё раз изображена на рис. 104. Рис. 104, *а* относится к решётке со

стехиометрическим избытком металла; рис. 104, *б* — к решётке с избытком металлоида. На энергетической схеме рис. 104 изображены лишь нормальное состояние и наинизшее из проводящих состояний.

С точки зрения этой схемы, спектр поглощения решётки, схематически изображённый на рис. 105, может быть легко

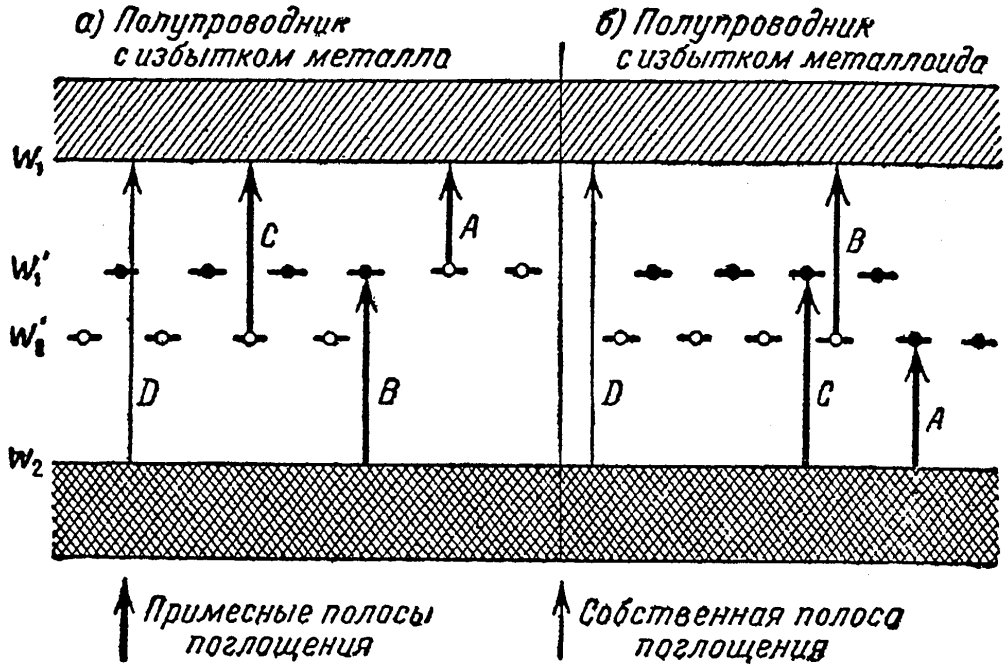


Рис. 104. Происхождение различных полос поглощения в реальном кристалле.

интерпретирован. В спектре поглощения реального кристалла мы имеем всегда две системы полос:

а) *собственные* полосы поглощения в коротковолновой части спектра;

б) *примесные* полосы поглощения в длинноволновой части спектра.

Собственные полосы довольно устойчивы: они лишь в очень слабой степени реагируют на внешние воздействия. Примесные же полосы представляют собой весьма хрупкие образования: их структура чрезвычайно чувствительна к внешним воздействиям. В некоторых образцах эти полосы могут вовсе отсутствовать. Они могут быть созданы искусственно и искусственно уничтожены. Под действием некоторых факторов они могут превращаться друг в друга, т. е. поглощение кристалла может возрастать в одной примесной полосе за счёт ослабления

поглощения в другой примесной полосе. Коэффициент поглощения в примесных полосах в среднем значительно меньше, чем в собственных полосах (на схематическом рисунке 105 масштаб не выдержан). С точки зрения схемы, изображённой на рис. 104, примесные полосы соответствуют переходам электронов с локальных уровней в зону проводимости или из нормальной зоны на локальные уровни (эти полосы на рис. 104 и 105 условно названы полосами A, B, C), а собственные полосы соответствуют переходам из зоны в зону (полоса D на рис. 104 и 105).

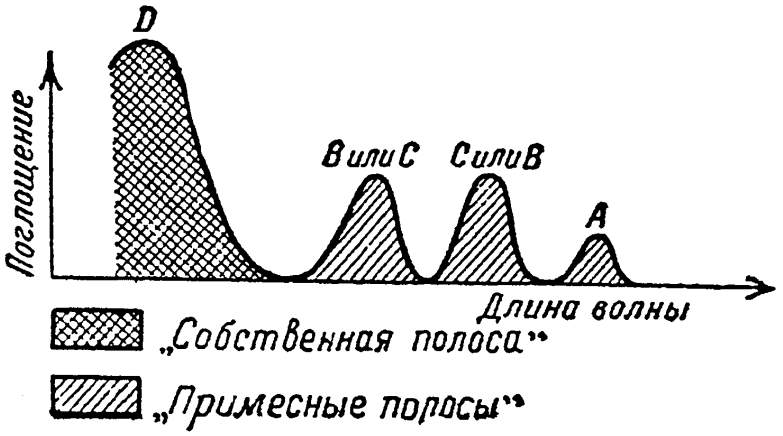


Рис. 105. Спектр поглощения реального кристалла.

Такая интерпретация спектра поглощения приводит к следующим следствиям:

1. Интенсивность примесных полос должна определяться не только природой полупроводника, но и степенью отклонения нашей реальной решётки от идеальной, а также степенью отклонения от стехиометрического соотношения, т. е. должна зависеть от параметров μ и ν. Иначе говоря, интенсивность этих полос должна находиться в непосредственной зависимости от обработки образца и, таким образом, может быть различной у различных образцов. Интенсивность собственной полосы, напротив того, определяется исключительно природой полупроводника, чем и обусловливается её устойчивость. В полупроводниках с соблюдённым стехиометрическим соотношением полоса A (см. рис. 104 и 105) должна вовсе отсутствовать. При превращении реальной решётки в идеальную должны также исчезать полосы B и C. Что же касается полосы D, то она должна сохраняться во всех случаях. Эти свойства полос поглощения, непосредственно вытекающие из рис. 104, находятся в соответствии с действительностью.

2. Из рис. 104 также видно, что примесные полосы поглощения должны лежать в более длинноволновой части спектра, нежели собственная полоса. При этом примесные полосы

должны быть значительно менее интенсивными, чем собственная полоса, поскольку интенсивность примесных полос обусловлена числом повреждений в решётке, а интенсивность собственной полосы — числом регулярных атомов. Эти свойства полос также соответствуют действительности.

3. Если освещать кристалл частотами, принадлежащими какой-либо примесной полосе, то, в результате поглощения этих частот, интенсивность других полос поглощения может, вообще говоря, измениться. Иначе говоря, освещение кристалла может вызвать изменения в его окраске. Действительно,

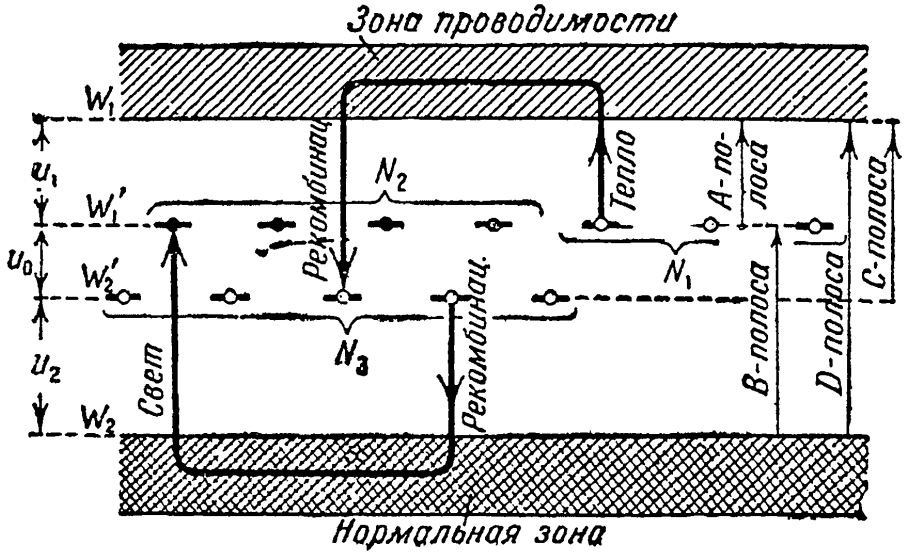


Рис. 106. Механизм изменения окраски кристалла под действием освещения.

если, например, освещать кристалл частотами, принадлежащими полосе B (см. рис. 104, a), предполагая при этом, что температура настолько низка, что можно пренебречь тепловыми перебросами электронов, то поглощение будет сопровождаться переводом электронов из нормальной зоны на уровни W_1', причём дырки, образующиеся в нормальной зоне, будут заполняться электронами с уровней W_2'. Так как непосредственная рекомбинация электронов, принадлежащих уровням W_1', с дырками, возникающими на уровнях W_2', не возможна, то процесс поглощения света сведётся к постепенному перекачиванию электронов с уровней W_2' на уровни W_1' через нормальную зону, как это показано на рис. 106 нижней жирной стрелкой. В результате все уровни W_2' окажутся истощёнными, а все уровни W_1' — насыщенными электронами. Кристалл

окажется переведённым в своего рода метастабильное состояние, а поглощательная способность в полосах B и C будет сведена к нулю. Однако исчезновение полос B и C приведёт к соответственному усилению полосы A. Исчезнувшие полосы будут с течением времени постепенно восстанавливаться, причём этот процесс восстановления должен протекать тем быстрее, чем выше температура. Действительно, тепловые перебросы электронов с уровней W_1' в зону проводимости и последующая их рекомбинация с дырками, сидящими на уровнях W_2', приведут к перекачиванию электронов с уровней W_1' на уровни W_2' (через зону проводимости, как это изображено на рис. 106 верхней жирной стрелкой), так что постепенно будет восстанавливаться первоначальное состояние. Такого рода процессы, существование которых вытекает непосредственно из анализа рис. 104, как известно, наблюдаются в действительности.

4. Из рис. 104 видно, что поглощение света в любой полосе всегда приводит к образованию электронов в зоне проводимости или дырок в нормальной зоне, или же к тому и другому одновременно. Таким образом, если оставаться в рамках рисунка 104, то на первый взгляд кажется, что все полосы поглощения должны быть фотоэлектрически активными, т. е. поглощение света во всех случаях должно сопровождаться положительным фотоэффектом. Этот вывод находится в противоречии с экспериментальными фактами. В дальнейшем (см. § 2 и 3) мы увидим, что это кажущееся противоречие может быть разрешено без введения каких-либо новых гипотез, и, таким образом, интерпретация спектра поглощения, данная на рис. 104, может быть сохранена.

В заключение заметим, что всё, что мы говорим о полупроводниках, может быть в полной мере перенесено на щёлочно-галоидные кристаллы (2,10), которые в оптическом и фотоэлектрическом отношении исследованы более тщательно. Переход от полупроводников к щёлочно-галоидным кристаллам означает лишь изменение масштаба на рис. 104.

С. Некоторые закономерности нормального фотоэффекта. Будем исходить из схемы уровней, изображённой на рис. 104 или 107. Эта схема относится к случаю полупроводника, обладающего стехиометрическим избытком металла. Впрочем, все дальнейшие рассуждения могут быть в полной

мере перенесены на полупроводники со стехиометрическим избытком металлоида (рис. 104, *б*).

Пусть мы имеем (на единицу объёма):

$$\begin{cases} N_1 + N_2 \text{ — общее число электроположительных уровней; среди них: } N_1 \text{ доноров и } N_2 \text{ акцепторов;} \\ N_3 \text{ — общее число электроотрицательных уровней; все эти уровни являются донорами.} \end{cases}$$

Числа N_1, N_2, N_3 связаны с параметрами μ и ν, характеризующими, соответственно, степень отклонения нашей реальной решётки от идеальной и степень отклонения от стехиометрического соотношения (см. главу I, § 1). В реальной решётке при соблюдённом стехиометрическом соотношении мы имеем: $N_1 = 0$; в идеальной решётке мы имеем: $N_1 = N_2 = N_3 = 0$.

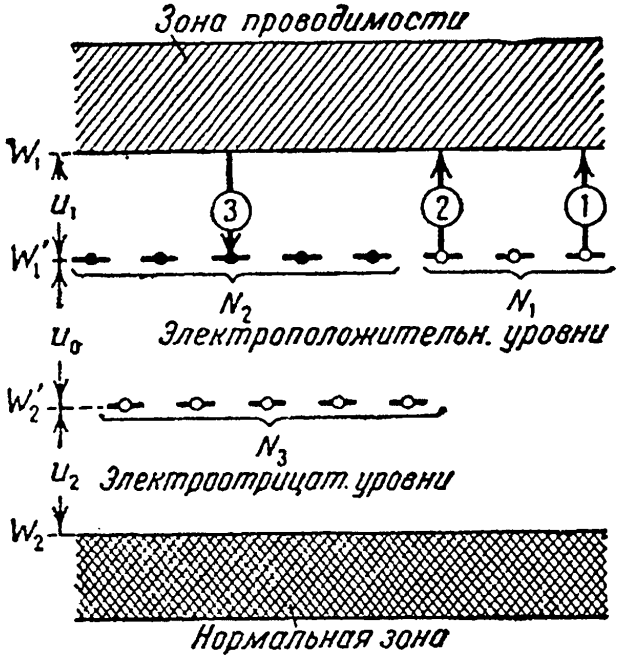


Рис. 107. Схема уровней.

Соотношение между числами N_1 и $N_2 \approx N_3$ характеризует степень нарушения стехиометрического соотношения. Будем различать:

1. Полупроводники со «слабыми» стехиометрическими нарушениями —

$$\text{если } N_2 \gg N_1, \quad \text{т. е. } \nu \ll \mu. \tag{1,1}$$

2. Полупроводники со «значительными» стехиометрическими нарушениями —

$$\text{если } N_2 \ll N_1, \text{ т. е. } \nu \approx \mu. \tag{1,2}$$

Заметим, что в главе IV, исследуя темновую проводимость полупроводников, мы ограничились рассмотрением лишь первого из этих двух случаев.

Предположим, что полупроводник находится при температуре T и облучается светом, так что за единицу времени в единице объёма поглощается Q квантов. Примем следую-

щие обозначения. Пусть

$$\begin{cases} n_1(T,Q) — \text{число электронов в зоне проводимости,} \\ n_1'(TQ) — \text{число электронов на электроположительных уровнях,} \\ n_2'(TQ) — \text{число дырок на электроотрицательных уровнях,} \\ n_2(TQ) — \text{число дырок в нормальной зоне.} \end{cases}$$

В этом параграфе и в следующем мы будем предполагать, что освещение производится частотами, принадлежащими примесной полосе, соответствующей переходам электронов с электроположительных уровней W_1' в зону проводимости (полоса A на рис. 104, a). В § 3 мы рассмотрим случай, когда кристалл освещается частотами, перебрасывающими электроны из нормальной зоны на электроположительные уровни W_1' (полоса B на рис. 104, a). Таким образом, в случае, рассматриваемом здесь, локальные уро ни W_1' играют роль резервуара, из которого черпаются световые электроны, в то время как в случае, который будет рассмотрен в § 3, те же уровни играют роль резервуара, принимающего в себя световые электроны.

Будем предполагать, что (см. рис. 107)

$$u_1 < u_0 \text{ и } u_1 < u_2, \tag{1,3}$$

и ограничим себя не слишком высокими температурами, когда участием электроотрицательных уровней W_2' и нормальной зоны в тепловых перебросах электронов можно вовсе пренебречь. Мы будем иметь:

$$\begin{cases} n_1(0,0) = 0, \\ n_1'(0,0) = N_1, \\ n_2'(0,0) = n_2'(T,Q) = 0, \\ n_2(0,0) = n_2(T,Q) = 0, \end{cases}$$

причём, очевидно:

$$n_1(T,Q) + n_1'(T,Q) = n_1(0,0) + n_1'(0,0) = N_1. \tag{1,4}$$

Примем обозначения:

$$\begin{cases} n = n_1(T,Q), \\ n_0 = n_1(T,0), \\ n^0 = n_1(0,Q). \end{cases}$$

Считая, что подвижность электронов в зоне проводимости не меняется при включении и выключении освещения, мы будем иметь:

$$\begin{cases} \sigma = evn, \\ \sigma_0 = evn_0, \\ \sigma^0 = evn^0, \end{cases}$$

где σ — проводимость при комбинированном действии тепла и света ($T \neq 0$, $Q \neq 0$); σ_0 — чисто темновая проводимость ($T \neq 0$, $Q = 0$), σ^0 — чисто световая проводимость ($T = 0$, $Q \neq 0$).

Между локальными уровнями W_1' и зоной проводимости существует постоянный обмен электронами: электроны выпадают из зоны проводимости вследствие прилипания (рекомбинаций) и появляются в ней вновь под действием тепла и под действием света, как это схематически изображено стрелками на рис. 107. Уровни W_2' и нормальная зона не принимают участия в игре и могут быть выброшены из рассмотрения.

Обозначим через Q_1 число электронов, которые за 1 *сек* в 1 *см*^{3} кристалла перебрасываются в зону проводимости при помощи тепла. Очевидно,

$$Q_1 = \beta (N_1 - n). \tag{1,5}$$

Обозначим через Q_2 число электронов, которые за 1 *сек* в 1 *см*^{3} кристалла перебрасываются в зону проводимости при помощи света. Очевидно,

$$Q_2 = Q = \varkappa Q_0, \tag{1,6}$$

где Q_0 — число падающих квантов (на 1 *см*^{2} поверхности кристалла за 1 *сек*), $\varkappa = \varkappa(\omega)$ — коэффициент поглощения, причём $\varkappa$ зависит от числа поглощающих центров, т. е. от числа электронов на локальных уровнях:

$$\varkappa = \gamma (N_1 - n). \tag{1,7}$$

Обозначим через Q_3 число электронов, выпадающих из зоны проводимости (за 1 *сек* на 1 *см*^{3} кристалла) вследствие рекомбинаций, иначе говоря, вследствие прилипания к локальным уровням. Очевидно,

$$Q_3 = \alpha (N_2 + n) n, \tag{1,8}$$

где коэффициент пропорциональности α мы будем называть *коэффициентом прилипания*, или *коэффициентом реком-*

бинации. Этот коэффициент следующим образом связан с вероятностью прилипания w и продолжительностью жизни в свободном состоянии τ:

$$\alpha(N_2+n)=w=\frac{1}{\tau}.$$

В дальнейшем мы будем считать, что ни температура, ни освещение не вызывают сколько-нибудь заметного истощения поглощающих центров, т. е.

$$n\ll N_1, \tag{1,9}$$

так что

$$Q_1=\beta N_1 \text{ и } \varkappa=\gamma N_1,$$

что справедливо при не слишком больших T и не слишком больших Q. Кроме того, будем считать, что коэффициенты α и β не меняются при включении и выключении освещения и в первом приближении не зависят от напряжённости внешнего электрического поля E.

Напишем условия равновесия

$$Q_1+Q_2=Q_3 \tag{1,10}$$

или, на основании (1,5), (1,6) и (1,8):

$$\boxed{Q+\beta N_1=\alpha(N_2+n)\,n.} \tag{1,11}$$

В этом уравнении, как мы увидим ниже, наряду со случаем фотоэлектрически активного поглощения, содержится также случай неактивного поглощения. Заметим, что этот второй случай оказывается возможным (см. § 2) благодаря наличию среди локальных уровней акцепторов, наряду с донорами. Если положить в наших формулах $N_2=0$, что означает возвращение в рамки элементарной вильсоновской теории, то случай фотоэлектрически неактивного поглощения выпадает из рассмотрения.

Обращаясь к анализу уравнения (1,11), рассмотрим следующие случаи:

1. Тепловое возбуждение при отсутствии светового, т. е. $Q_2=0$ при $Q_1\neq 0$.
2. Световое возбуждение при отсутствии теплового, т. е. $Q_1=0$ при $Q_2\neq 0$.
3. Комбинированное действие тепла и света, т. е. $Q_1\neq 0$ и $Q_2\neq 0$.

1. Предположим, что освещение отсутствует ($Q=0$). Это — случай тепловой проводимости в полупроводнике (см. § 1, С гл. IV). В этом случае, заменяя в (1, 11) n через n_0, мы получаем из (1,11):

$$\frac{\beta}{\alpha}=\frac{n_0(N_2+n_0)}{N_1}. \tag{1,12}$$

С другой стороны, мы имеем

$$\frac{\beta}{\alpha}=A(T)\,e^{-\frac{u_1}{kT}}, \tag{1,13}$$

где

$$A(T)=2\left(\frac{2\pi m^* kT}{h}\right)^{3/2}.$$

Здесь u_1 — ширина запрещённого участка между локальными уровнями и зоной проводимости (см. рис. 107); m^* — эффективная масса электрона в зоне проводимости.

Рассмотрим два предельных случая:

а) Предположим, что

$$N_2 \ll n_0. \tag{1,14а}$$

Это может иметь место лишь в полупроводниках со значительными стехиометрическими нарушениями ($N_1 \gg N_2$), ибо, согласно нашему предположению, $n_0 \ll N_1$. При этом мы возвращаемся к результатам элементарной вильсоновской теории. Поэтому полупроводники, для которых имеет место (1,14а), мы будем для краткости называть *вильсоновскими* полупроводниками.

Формулы (1,12) и (1,13) в этом случае дают нам:

$$n_0=\sqrt{\frac{\beta}{\alpha}N_1}=\sqrt{AN_1}e^{-\frac{u_1}{2kT}}. \tag{1,15а}$$

б) Предположим, что

$$N_2 \gg n_0. \tag{1,14б}$$

Этот случай может осуществляться как в полупроводниках со значительными стехиометрическими нарушениями, т. е. при $N_1 \gg N_2$, так и в полупроводниках со слабыми стехиометрическими нарушениями, т. е. при $N_2 \gg N_1$ (см. рис. 108). Этот случай рассмотрен в работе Нейбура (§ 1, С гл. IV).

Поэтому полупроводники, для которых имеет место (1,14б), мы будем для краткости называть *нейбуровскими* полупроводниками.

Формулы (1,12) и (1,13) дают в этом случае:

$$n_0 = \frac{\beta}{\alpha} \cdot \frac{N_1}{N_2} = \frac{N_1}{N_2} A e^{-\frac{u_1}{kT}}. \qquad (1,15б)$$

Заметим, что полупроводник с сильно нарушенным стехиометрическим составом может быть вильсоновским полупроводником при высоких температурах и нейбуровским — при низких. Зависимость электропроводности σ_0 от температуры для такого полупроводника [согласно формулам (1,12) — (1,13)] изображена на рис. 109 (см. также рис. 65). Точка перегиба соответствует значению $n_0 = N_2$, как это легко сообразить на основании (1,15а) и (1,15б).

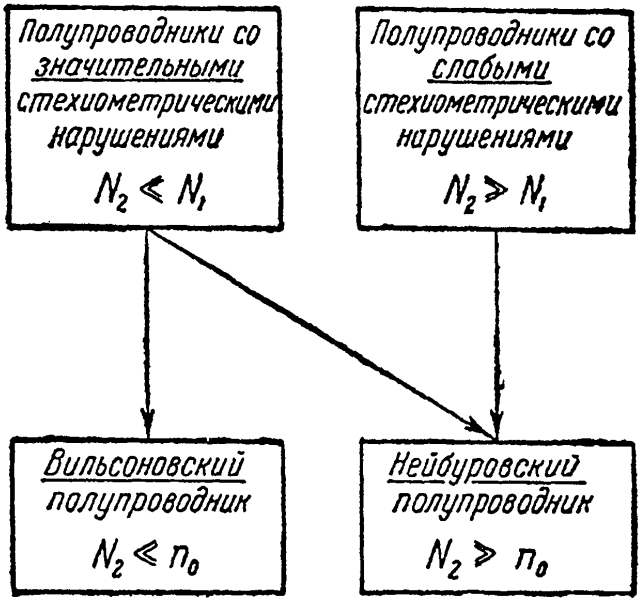


Рис. 108. Различные типы полупроводников (при $n_0 \ll N_1$).

2. Предположим теперь, что кристалл освещается, но тепловое возбуждение при этом отсутствует ($Q_2 = 0$). Это — случай фотопроводимости в диэлектрике или в полупроводнике при достаточно низкой температуре. В этом случае, заменяя в (1,11) n через n^0, мы получаем из (1,11):

$$n^0 = \frac{1}{2} N_2 \left\{ \sqrt{1 + 4\frac{Q}{\alpha N_2^2}} - 1 \right\}. \qquad (1,16)$$

Рассмотрим два предельных случая:

а) Предположим, что $N_2 \ll n$ или, что то же самое, $4\frac{Q}{\alpha N_2^2} \gg 1$.

Тогда (1,16) превращается в

$$n^0 = \sqrt{\frac{Q}{\alpha}}. \qquad (1,17а)$$

В этом случае фотопроводимость пропорциональна корню квадратному из поглощённой энергии. Это имеет место при заданном N_2 до тех пор, пока Q не слишком мало, т. е. пока

$$Q \gg \frac{1}{4} aN_2^2.$$

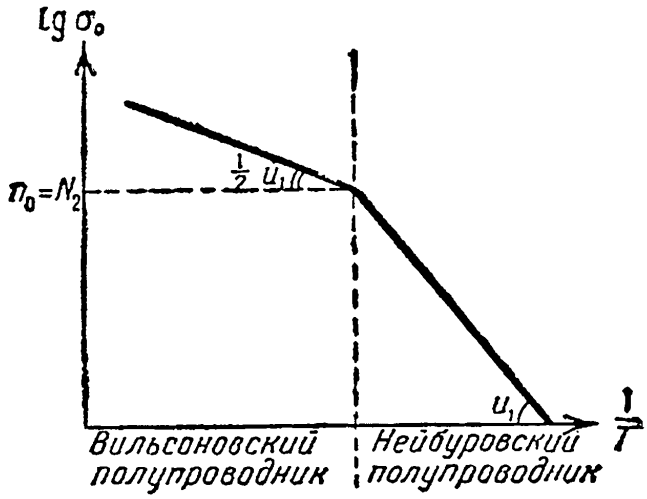


Рис. 109. Зависимость темновой проводимости σ_0 от температуры для полупроводника со значительными стехиометрическими нарушениями (в области низких температур).

б) Предположим, что $N_2 \gg n^0$ или, что то же самое, $4\frac{Q}{aN_2^2} \ll 1$. В этом случае (1,16) превращается в

$$n^0 = \frac{Q}{aN_2}, \quad (1,17б)$$

т. е. фотопроводимость пропорциональна поглощённой энергии. При заданном N_2 это имеет место до тех пор, пока Q не слишком велико, т. е. пока

$$Q \ll \frac{1}{4} aN_2^2.$$

Зависимость фотопроводимости от количества поглощённых квантов [согласно формуле (1,16)] изображена на рис. 110. Различные кривые соответствуют различному числу акцепторов N_2. Чем больше акцепторов, тем меньше проводимость (при одном и том же Q), тем длиннее прямолинейный участок и тем меньше наклон прямой $\left(\operatorname{tg}\varphi = \frac{ev}{aN_2}\right)$. Значение Q, при котором происходит переход от линейной зависимости (1,17б) к параболической 1,17а), определяется условием:

$$Q = aN_2^2.$$

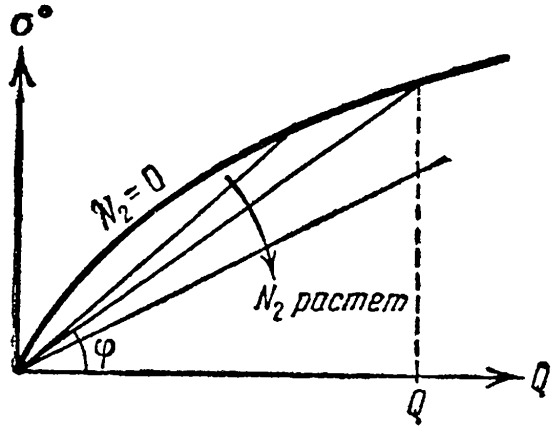


Рис. 110. Зависимость фотопроводимости от количества поглощённых квантов для полупроводников с различным содержанием акцепторов:

$$\operatorname{tg}\varphi = \frac{ev}{aN_2}, \quad Q' = aN_2^2.$$

3. Обратимся теперь к общему случаю (одновременное действие теплового и светового возбуждения). Это — случай фотопроводимости в полупроводнике при таких температурах,

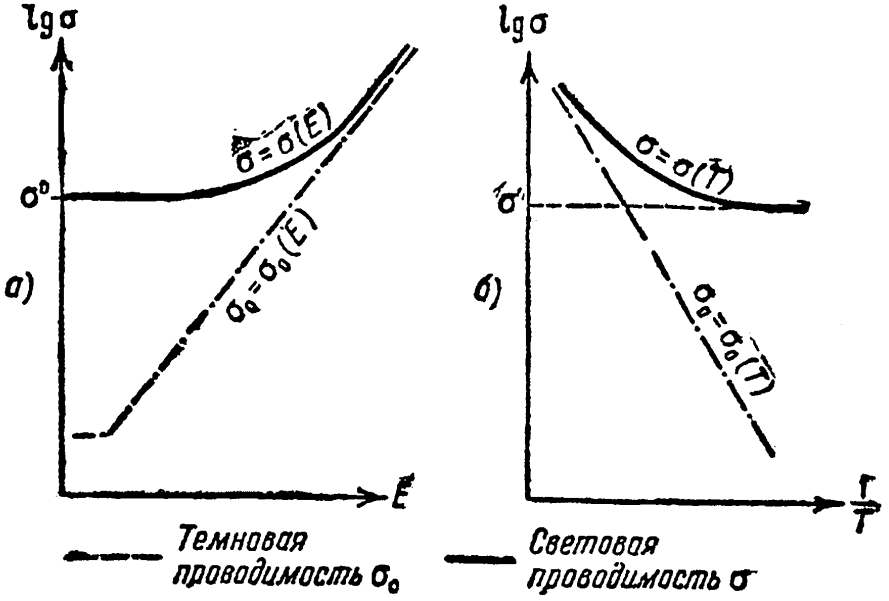


Рис. 111. Зависимость темновой и световой проводимостей от E и от T.

когда полупроводник обладает уже заметной темновой проводимостью.

Из (1,11) мы имеем

$$n=\frac{1}{2}\left\{\sqrt{N_2^2+4\frac{\beta}{\alpha}N_1+4\frac{Q}{\alpha}}-N_2\right\} \qquad (1,18)$$

или , на основании (1,12) и (1,16):

$$\boxed{n=\frac{1}{2}\left\{\sqrt{(N_2+2n^0)^2+4n_0(N_2+n_0)}-N_2\right\}.} \qquad (1,19)$$

Из (1,19) мы имеем:

$$n=\begin{cases}=n_0\ldots, & \text{если } n^0=0,\\ =n^0\ldots, & \text{если } n_0=0,\end{cases}$$

как и следовало ожидать.

Очевидно, n зависит

$$\begin{cases}\text{от } T-\\ \text{от } E-\end{cases}\Big\} \text{ через } n_0, \text{ ибо } n_0=n_0(T,E),$$

$$\text{от } Q- \quad \text{через } n^0, \text{ ибо } n^0=n^0(Q).$$

Соответствующие кривые изображены на рис. 111 (ср. с рис. 103). Эти результаты следует считать в достаточной мере тривиальными. Более интересный случай, скрытый в выражении (1,19), будет рассмотрен в следующем параграфе.

§ 2. Фотоэлектрически неактивное поглощение света

А. Условия исчезновения фотоэффекта. Пусть σ_0 — проводимость полупроводника при отсутствии освещения, а σ — при наличии освещения (при прочих равных условиях). Очевидно,

$$\Delta\sigma = \sigma - \sigma_0$$

представляет собой добавку к проводимости, создаваемую светом. Фотоэлектрическую активность следует характеризовать не абсолютной величиной добавки $\Delta\sigma$, а её относительной величиной по отношению к темновой проводимости:

$$\eta = \frac{\Delta\sigma}{\sigma_0}. \qquad (2,1)$$

Величина η показывает, во сколько раз добавочная проводимость, создаваемая светом, больше основной темновой проводимости. Эта величина характеризует, насколько проводимость полупроводника чувствительна к освещению. Величину η мы и будем называть *фотоэлектрической активностью* полупроводника.

Обычно

$$\eta \gg 1.$$

В этом случае мы говорим, что в полупроводнике имеет место внутренний фотоэффект или что полупроводник фотоэлектрически активен. Иногда же

$$\eta \ll 1.$$

В этом случае мы говорим, что внутренний фотоэффект отсутствует или что полупроводник фотоэлектрически не активен. Это значит, что свет хотя и поглощается, но не создаёт сколько-нибудь заметной дополнительной проводимости.

Заметим что, сравнивая фотоэлектрическую активность двух полупроводников, мы должны относить фотоэлектрическую активность к одним и тем же условиям освещения. Под одинаковыми условиями освещения следует понимать не

одну и ту же частоту света и не одну и ту же интенсивность света, а одно и то же число поглощённых квантов.

Задача заключается в том, чтобы объяснить, почему два полупроводника с одинаковой темновой проводимостью, при одном и том же числе поглощённых квантов, могут обладать существенно различной фотоэлектрической активностью, так что, например, в одном полупроводнике может наблюдаться значительный фотоэффект, а в другом — фотоэффект может вовсе отсутствовать [11].

Возвратимся к случаю, рассмотренному в предыдущем параграфе. На основании (1,18), (1,12) и (2,1), мы будем иметь:

$$\eta = \frac{n - n_0}{n_0} = \frac{1}{2n_0}\left\{ \sqrt{N_2^2 + 4n_0(N_2 + n_0) + 4\frac{Q}{\alpha}} - (N_2 + 2n_0) \right\}. \quad (2,2)$$

Рассмотрим два предельных случая:

а) Предположим, что $N_2 \ll n_0$ (или $N_1 \gg n_0 \gg N_2 \geqslant 0$). Это — случай вильсоновского полупроводника. В этом случае (2,2) даёт:

$$\eta = \sqrt{1 + \frac{Q}{\alpha n_0^2}} - 1. \quad (2,3а)$$

Если

$$\frac{Q}{\alpha n_0^2} \gg 1, \text{ т. е. } Q \gg \alpha n_0^2, \quad (2,4а)$$

то

$$\eta = \sqrt{\frac{Q}{\alpha n_0^2}} \gg 1, \quad (2,5а)$$

т. е. фотоэлектрический эффект существует. При этом

$$\sigma = ev\left(n_0 + \sqrt{\frac{Q}{\alpha}}\right) \approx ev\sqrt{\frac{Q}{\alpha}}. \quad (2,6а)$$

Если же

$$\frac{Q}{\alpha n_0^2} \ll 1, \text{ т. е. } Q \ll \alpha n_0^2, \quad (2,7а)$$

то

$$\eta = \frac{Q}{2\alpha n_0^2} \ll 1, \quad (2,8а)$$

т. е. фотоэлектрический эффект отсутствует. В этом случае

$$\sigma = ev\left(n_0 + \frac{Q}{2\alpha n_0^2}\right) \approx \sigma_0.$$

б) Предположим теперь, что $N_2 \gg n_0$ (или N_1, $N_2 \gg n_0 \gg 0$). Это — случай нейбуровского полупроводника. В этом случае (2,2) даёт:

$$\eta = \frac{N_2}{2n_0}\left\{\sqrt{1 + 4\frac{Q}{\alpha N_2^2}} - 1\right\}. \qquad (2,3б)$$

Если

$$4\frac{Q}{\alpha N_2^2} \gg 1, \text{ т. е. } Q \gg \frac{1}{4}\alpha N_2^2, \qquad (2,4б)$$

то

$$\eta = \sqrt{\frac{Q}{\alpha n_0^2}} \gg 1, \qquad (2,5б)$$

т. е. фотоэлектрический эффект существует. При этом

$$\sigma = ev\left(n_0 + \sqrt{\frac{Q}{\alpha}}\right) \approx ev\sqrt{\frac{Q}{\alpha}}. \qquad (2,6б)$$

Если же

$$4\frac{Q}{\alpha N_2^2} \ll 1, \text{ т. е. } Q \ll \frac{1}{4}\alpha N_2^2, \qquad (2,7б)$$

то

$$\eta = \frac{Q}{\alpha n_0 N_2}. \qquad (2,8б)$$

При этом фотоэлектрический эффект может отсутствовать, но может и существовать. Действительно, из условия (2,7б) отнюдь ещё не следует, что $\eta \ll 1$, ибо $n_0 \ll N_2$. Фотоэлектрический эффект существует, если

$$\frac{Q}{\alpha n_0 N_2} \gg 1, \text{ т. е. } Q \gg \alpha n_0 N_2; \qquad (2,9)$$

при этом

$$\sigma = ev\left(n_0 + \frac{Q}{\alpha N_2}\right) \approx ev\frac{Q}{\alpha N_2}. \qquad (2,10)$$

Фотоэлектрический эффект отсутствует, если

$$\frac{Q}{\alpha n_0 N_2} \ll 1, \text{ т. е. } Q \ll \alpha n_0 N_2. \qquad (2,11)$$

При этом

$$\sigma = ev\left(n_0 + \frac{Q}{\alpha N_2}\right) \approx \sigma_0. \qquad (2,12)$$

Все эти различные случаи сведены в схему, изображённую на рис. 112.

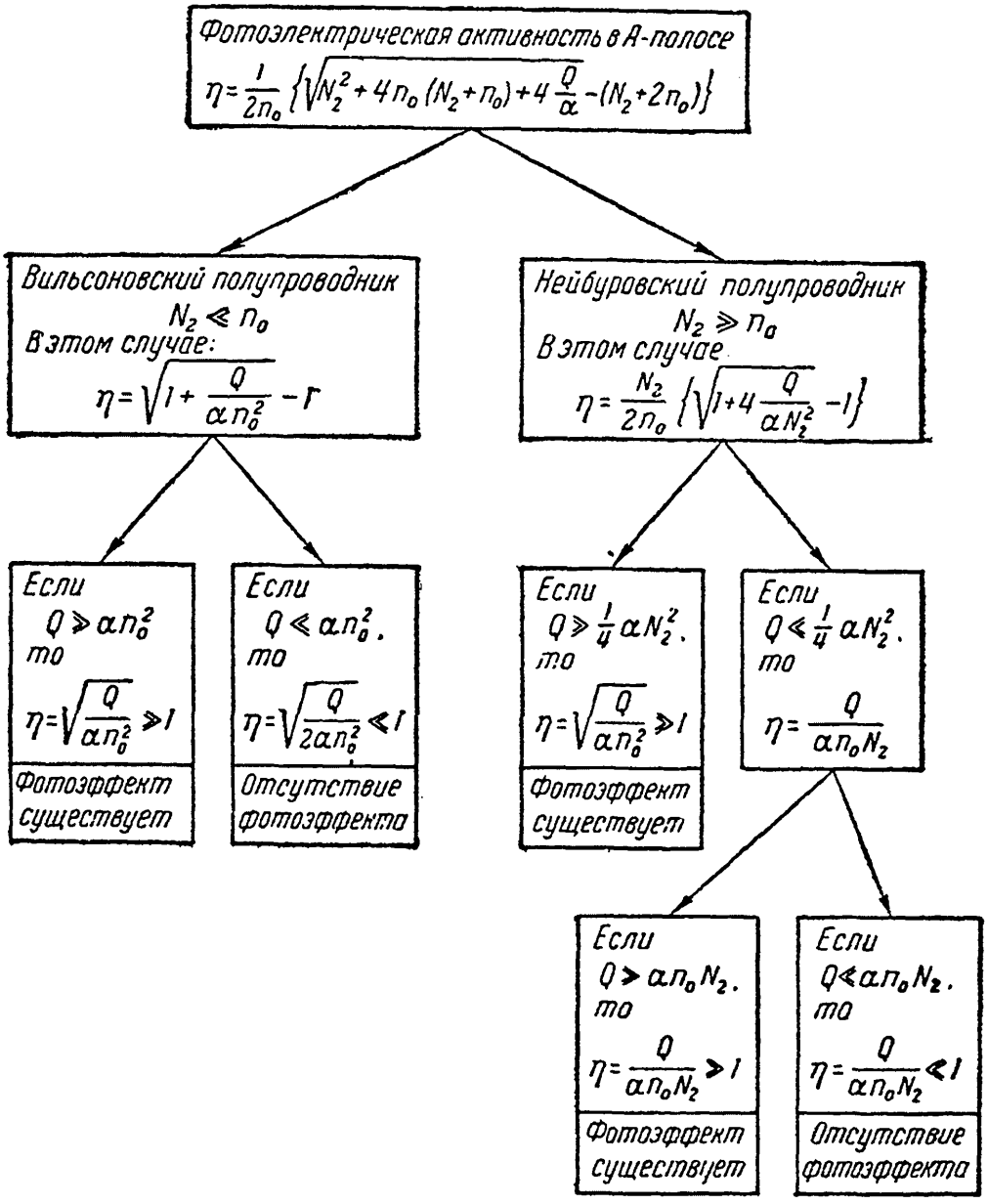


Рис. 112. Условия существования и отсутствия внутреннего фотоэффекта.

В. Некоторые выводы. Мы видим, что условия существования или отсутствия внутреннего фотоэффекта выглядят по-разному, в зависимости от того, имеем ли мы дело с полупроводником, принадлежащим к типу вильсоновских полу-

проводников, или — к типу нейбуровских. Действительно, примем для сокращённости такие обозначения:

$$\begin{cases} a = \dfrac{Q}{\alpha n_0^2}, \\ b = \dfrac{Q}{\alpha n_0 N_2}. \end{cases} \tag{2,13}$$

Тогда:

	Полупроводник типа Вильсона ($n_0 \gg N_2$)	Полупроводник типа Нейбура ($n_0 \ll N_2$)
Условие *существования* внутреннего фотоэффекта . . .	$a \gg 1$	$b \gg 1$
Условие *отсутствия* внутреннего фотоэффекта	$a \ll 1$	$b \ll 1$

Расшифровывая эти условия, мы приходим к следующим трём выводам:

1) Вспомним, что

$$\alpha (N_2 + n_0) n_0 = \beta N_1.$$

Очевидно, для вильсоновского полупроводника ($n_0 \gg N_2$):

$$\alpha n_0^2 = \beta N_1 = Q_2;$$

для нейбуровского ($n_0 \ll N_2$):

$$\alpha n_0 N_2 = \beta N_1 = Q_2.$$

Следовательно,

$$\begin{cases} a = \dfrac{Q}{Q_2}, \\ b = \dfrac{Q}{Q_2}. \end{cases} \tag{2,14}$$

Таким образом, независимо от того, принадлежит ли наш полупроводник к вильсоновскому или к нейбуровскому типу, для того чтобы в нём существовал внутренний фотоэффект, необходимо, чтобы $Q \gg Q_2$, т. е. необходимо, чтобы число электронов, перебрасываемых светом, было бы значительно

больше, чем число электронов, перебрасываемых теплом (при отсутствии освещения). Наоборот, для того чтобы внутренний фотоэффект отсутствовал, необходимо, чтобы $Q_2 \gg Q$, т. е. тепло должно перебрасывать электронов больше, чем свет. Этот вывод вполне естественен, если не сказать тривиален.

2) Выражения (2,13) можно ещё иначе переписать так:

$$\begin{cases} a = \dfrac{e^2 v^2}{\alpha} \cdot \dfrac{Q}{\sigma_0^2}, \\ b = \dfrac{ev}{\alpha} \cdot \dfrac{Q}{\sigma_0 N_2}. \end{cases} \tag{2,15}$$

Отсюда видно, что в вильсоновском полупроводнике фотоэлектрическая активность (отнесённая к одному и тому же числу поглощённых квантов Q) может быть уменьшена только за счёт увеличения темновой проводимости σ_0. В нейбуровском полупроводнике, в отличие от вильсоновского, фотоэлектрическая активность может быть уменьшена не только за счёт повышения σ_0, но также при неизменном σ_0, за счёт увеличения числа акцепторов N_2.

Темновую проводимость σ_0 можно увеличить следующими двумя способами (это относится как к вильсоновскому, так и к нейбуровскому полупроводникам):

Во-первых, увеличивая температуру и оставляя неизменным число доноров и число акцепторов ($N_1 = \text{const}$, $N_2 = \text{const}$). Фотоэлектрическая активность полупроводника при этом будет падать, как это видно из (2,15), ибо σ_0 растёт с температурой гораздо быстрее, чем уменьшается v. (Коэффициент прилипания α можно считать не зависящим от температуры; не следует путать коэффициент прилипания α с вероятностью прилипания $w = \alpha(N_2 + n)$, которая быстро растёт с температурой.)

Во-вторых, σ_0 можно увеличить, оставляя постоянной температуру и увеличивая содержание доноров N_1 при неизменном числе акцепторов ($N_2 = \text{const}$). При этом опять-таки фотоэлектрическая активность, отнесённая к одному и тому же числу поглощённых квантов Q, будет падать. Заметим, что при возрастании N_1 растёт также и коэффициент поглощения $\varkappa = \gamma N_1$. Поэтому, чтобы сохранить $Q = \text{const}$, надо, по мере увеличения N_1, уменьшать интенсивность освещающего света. Иначе говоря, чем больше содержание доноров, тем при меньшей интенсивности света поглощается данное число квантов.

Таким образом, если фотоэлектрическую активность относить не к одному и тому же числу поглощённых квантов (как мы условились), а к одному и тому же числу падающих квантов, то в этом случае фотоэлектрическую активность надо считать вовсе не зависящей ни от числа доноров N_1, ни от числа акцепторов N_2. Действительно, если в выражение для a [формула (2,13)] подставить (1,15а), а в выражение для b [формула (3,13)] подставить (1,15б) и, кроме того, вспомнить, что $Q=\gamma N_1 Q_0$, то легко обнаружить, что параметры N_1 и N_2 совсем выпадают как из a, так и из b.

Посмотрим теперь, каким образом в нейбуровском полупроводнике можно уменьшить фотоэлектрическую активность, не изменяя темновой проводимости σ_0. Для этого, оставляя постоянной температуру, надо вводить в нейбуровский полупроводник новые акцепторы и одновременно вводить новые доноры так, чтобы при этом отношении N_1/N_2 всё время оставалось постоянным. Тогда темновая проводимость не будет изменяться, как это видно из (1,15 б), но фотоэлектрическая активность тем не менее будет падать. На этот раз она будет падать уже не за счёт увеличения темновой проводимости σ_0, а за счёт уменьшения фотопроводимости. (Здесь опять-таки имеется в виду фотоэлектрическая активность, отнесённая к одному и тому же числу поглощённых, а не падающих квантов.)

3) Поскольку подвижность v и коэффициент прилипания a можно считать более или менее одинаковыми (одинаковыми по порядку величины) для всех полупроводников, то, рассматривая (2,15), мы можем сделать ещё следующий вывод.

Два вильсоновских полупроводника с одинаковой темновой проводимостью обязаны обладать также одинаковой (или более или менее одинаковой) фотоэлектрической активностью. Иначе говоря, если в каком-нибудь вильсоновском полупроводнике при некоторой заданной температуре существует значительный фотоэффект, то невозможно себе представить, чтобы в каком-нибудь другом вильсоновском полупроводнике, обладающем такой же темновой проводимостью, при той же температуре и при том же количестве поглощённых квантов, фотоэффект вовсе отсутствовал бы.

Иначе обстоит дело у нейбуровских полупроводников. Из (2,15) видно, что два нейбуровских полупроводника, обладая одинаковой темновой проводимостью, могут давать

существенно различный фотоэлектрический эффект, если N_2 у них существенно различны.

Таким образом, имея дело с нейбуровскими полупроводниками, мы можем себе представить два таких полупроводника, у которых темновая проводимость одинаковая, но при этом один полупроводник обладает значительным фотоэффектом, а у другого полупроводника при том же количестве поглощённых квантов фотоэффект отсутствует совсем. Это будет иметь место в том случае, если второй полупроводник содержит значительно больше уровней прилипания, чем первый. Это обстоятельство является следствием того факта, что в полупроводнике с бо́льшим содержанием акцепторов тепловой переброс электронов в зону проводимости должен быть бо́льшим, для того чтобы концентрация свободных электронов оставалась такой же. Иначе говоря, если σ_0 при некоторой заданной температуре у обоих полупроводников одинаково, то это значит, что в полупроводнике с бо́льшим количеством акцепторов Q_2 должно быть больше, чем в полупроводнике с меньшим количеством акцепторов.

Заметим, что для того, чтобы фотоэлектрическая активность у двух нейбуровских полупроводников (с одинаковой темновой проводимостью) была бы существенно различной, вовсе не необходимо, чтобы эти два полупроводника были приготовлены из различных материалов. Это может иметь место также и для двух разных образцов одного и того же полупроводника. Для этого достаточно, чтобы число акцепторов в одном из образцов было бы гораздо больше, чем в другом, но и число доноров было бы во столько же раз бо́льшим. Таким образом, наличие или отсутствие фотоэффекта, как и многие другие свойства полупроводников, определяется не столько природой материала, сколько историей каждого данного образца, т. е. тем, какой обработке был подвергнут образец. Фотоэлектрическая активность нейбуровского полупроводника может резко меняться от образца к образцу.

Наконец, сравнивая два полупроводника, у которых темновая проводимость при данной температуре одинакова, но один из которых является вильсоновским полупроводником, а другой — нейбуровским, мы замечаем, что в нейбуровском полупроводнике фотоактивность должна быть меньше, чем в вильсоновском. Это видно из (2,13): действительно, α и n_0

для обоих полупроводников более или менее одинаковы, но зато $N_2 \gg n_0$ и поэтому $a \gg b$.

С. Два численных примера. В заключение приведём два численных примера.

Пример первый. Рассмотрим два нейбуровских полупроводника. Пусть это будут два по-разному приготовленных образца одного и того же полупроводника. Примем для них следующие численные значения параметров:

	N_1 $см^3$	N_2 $см^3$	A (при $T=292°$) $см^3$	u_1 eV	v $см^2/V \cdot сек$
Образец № 1 .	10^{14}	10^{16}	10^{19}	0,4	100
Образец № 2 .	10^{18}	10^{20}			

Такие два образца вплоть до достаточно высоких температур обладают одинаковой темновой проводимостью. В частности при комнатной температуре ($T=292°$ K; $1/T=3{,}42\cdot10^{-3}$) темновая проводимость для обоих образцов [согласно (2,15б)]:

$$\sigma_0 = 2{,}17\cdot10^{-7} ом^{-1} см^{-1}.$$

Предположим, что первый образец обладает значительной фотоэлектрической чувствительностью. Пусть для него при комнатной температуре:

$$\eta = \frac{ev}{a}\frac{Q}{\sigma_0\cdot10^{16}} = 10^3,$$

т. е. под действием освещения проводимость его увеличивается в 1000 раз. Тогда для второго образца при той же температуре и при том же количестве поглощённых квантов фотоэлектрическая активность будет:

$$\eta = \frac{ev}{a}\frac{Q}{\sigma_0\cdot10^{20}} = \frac{ev}{a}\cdot\frac{Q}{\sigma_0\cdot10^{16}}\cdot\frac{10^{16}}{10^{20}} = 10^3\cdot\frac{10^{16}}{10^{20}} = 0{,}1.$$

Таким образом, у первого образца существует значительный фотоэффект, а у второго образца он практически отсутствует. Первый образец являет собой пример фотоэлектрически активного поглощения; второй образец — пример неактивного поглощения.

Заметим, однако, что здесь фотоэлектрическая активность рассчитана для одного и того же количества поглощённых квантов. Если же её относить к одному и тому же количеству падающих квантов, то у второго образца она будет такой же, как и у первого. Это происходит потому, что коэффициент поглощения у второго образца в 10^4 раз больше, чем у первого.

Пример второй. Рассмотрим два полупроводника, один из которых пусть будет вильсоновским полупроводником, а другой — нейбуровским. Примем для параметров следующие численные значения:

	N_1 $см^{-3}$	N_2 $см^{-3}$	A (при $T=300°$) $см^{-3}$	u_1 eV	v $см^2/V\cdot сек$
Полупроводник № 1 (вильсоновский)	10^{19}	0	10^{19}	1,2	100
Полупроводник № 2 (нейбуровский)	10^{19}	10^{19}	10^{19}	0,6	100

Темновая проводимость у обоих полупроводников одинакова во всём интервале температур, доступном для теории. В этом можно убедиться, подставляя значения параметров в формулы (1,15а) и (1,15б). В частности, при $T=300°K$:

$$\sigma = 1{,}59\cdot 10^{-8} ом^{-1} см^{-1}.$$

Заметим, что в спектре поглощения примесная полоса у второго (нейбуровского) полупроводника сдвинута по сравнению с первым (вильсоновским) полупроводником на 0,6 eV в сторону длинных волн. Поэтому будем освещать первый полупроводник светом частоты ω, принадлежащей его примесной полосе, а второй полупроводник — светом частоты $\omega' = \omega - \frac{0{,}6}{\hbar}$. При этом коэффициент поглощения у обоих полупроводников можно считать одинаковым (или почти одинаковым), поскольку оба полупроводника обладают одинаковым числом поглощающих центров N_1 (в предположении, что в обоих случаях зона проводимости имеет одно и то же происхождение, например, является s-зоной). Это значит, что при одинаковом

числе падающих квантов оба полупроводника поглощают также одинаковое число квантов.

Предположим, что первый полупроводник обладает значительным фотоэффектом. Пусть для него при $T=300°\,\mathrm{K}$

$$\eta=\frac{ev}{\sigma_0}\sqrt{\frac{Q}{\alpha}}=10^4.$$

Тогда для второго полупроводника при той же температуре и том же количестве поглощённых квантов будем иметь:

$$\eta=\frac{ev}{\sigma_0}\cdot\frac{Q}{\alpha\cdot 10^{19}}=\left(\frac{ev}{\sigma_0}\sqrt{\frac{Q}{\alpha}}\right)^2\cdot\frac{\sigma_0}{ev\cdot 10^{19}}=10^8\cdot\frac{\sigma_0}{ev\cdot 10^{19}}=10^{-2},$$

т. е. фотоэлектрическая активность у второго полупроводника оказывается в миллион раз меньшей, чем у первого. Здесь фотоэлектрическая активность рассчитана по отношению к одному и тому же числу поглощённых квантов, или, что в нашем примере то же самое, к одному и тому же числу падающих квантов.

§ 3. Аномальный фотоэлектрический эффект

А. Условия возникновения отрицательного фотоэффекта. В предыдущих параграфах мы рассматривали тот случай, когда световые и тепловые электроны имеют одно и то же происхождение, т. е. когда и свет и тепло срывают электроны с одних и тех же уровней. Этот случай привёл нас к нормальному фотоэффекту (увеличение проводимости под действием света). Кроме того, этот случай, при определённом соотношении между числом доноров и акцепторов, привёл нас к объяснению фотоэлектрически неактивного поглощения света. Однако рассмотренный случай не оставил места для отрицательного фотоэффекта. Из схемы, с которой мы оперировали в предыдущих параграфах, ни при каких условиях нельзя получить понижения проводимости при освещении. Действительно, для Δn мы имеем выражение (2,2):

$$\Delta n=\frac{1}{2}\left\{\sqrt{N_2^2+4n_0(N_2+n_0)+4\frac{Q}{\alpha}}-(N_2+2n_0)\right\}.$$

Легко видеть, что при любых значениях параметров, входящих в это выражение, мы имеем всегда

$$\Delta n>0.$$

В этом параграфе мы рассмотрим случай, когда световые и тепловые электроны различны по своему происхождению, т. е. когда свет черпает электроны с иных уровней, чем тепло. Этот случай существенно отличается от предыдущего тем, что дополнительные уровни прилипания, возникающие под действием света, имеют теперь иную природу. Свободные электроны получают теперь возможность рекомбинировать с дырками двух сортов: с дырками, созданными теплом, и с дырками, созданными светом, причём соответствующие коэффициенты рекомбинации теперь, вообще говоря, не обязаны быть одинаковыми. Это обстоятельство имеет своим следствием то, что, наряду с положительным фотоэффектом и фотоэлектрически неактивным поглощением, может существовать также и отрицательный фотоэффект. Условия, при которых осуществляется тот или иной из этих случаев, будут выведены ниже.

Заметим, что во всех случаях (как в случае положительного, так и в случае отрицательного фотоэффекта) число электронов, высвобождаемых ежесекундно при комбинированном действии тепла и света, конечно, больше, чем при действии одного лишь тепла. Однако освещение, изменяя условия рекомбинации, может привести к её значительному облегчению, в результате чего, при определённых условиях, суммарная электропроводность при наличии освещения может оказаться меньше, чем в темноте. Таково, как мы убедимся в этом ниже, возможное происхождение отрицательного фотоэффекта.

Мы будем исходить из энергетической схемы, изображённой на рис. 113. Эта схема относится к полупроводнику, обладающему стехиометрическим избытком металла. Однако наши результаты в полной мере могут быть перенесены на полупроводники со стехиометрическим избытком металлоида.

Будем предполагать, что кристалл освещается частотами той примесной полосы, которая соответствует перебросу электронов из нормальной зоны на электроположительные уровни W_1' (полоса B на рис. 104, a), а тепло осуществляет перебросы электронов с электроположительных уровней W_1' в зону проводимости. Таким образом, при чисто тепловом возбуждении, т. е. при отсутствии освещения, электроотрицательные уровни W_2' и нормальная зона вовсе не участвуют в игре. Это значит,

что мы ограничиваемся областью достаточно низких температур. При включении освещения все уровни включаются в игру, в том числе и электроотрицательные уровни W_2', электроны которых получают возможность рекомбинировать с дырками, освобождающимися в нормальной зоне под действием освещения. Выпадение электронов с уровней W_2' приводит к тому, что эти уровни приобретают способность принимать на себя

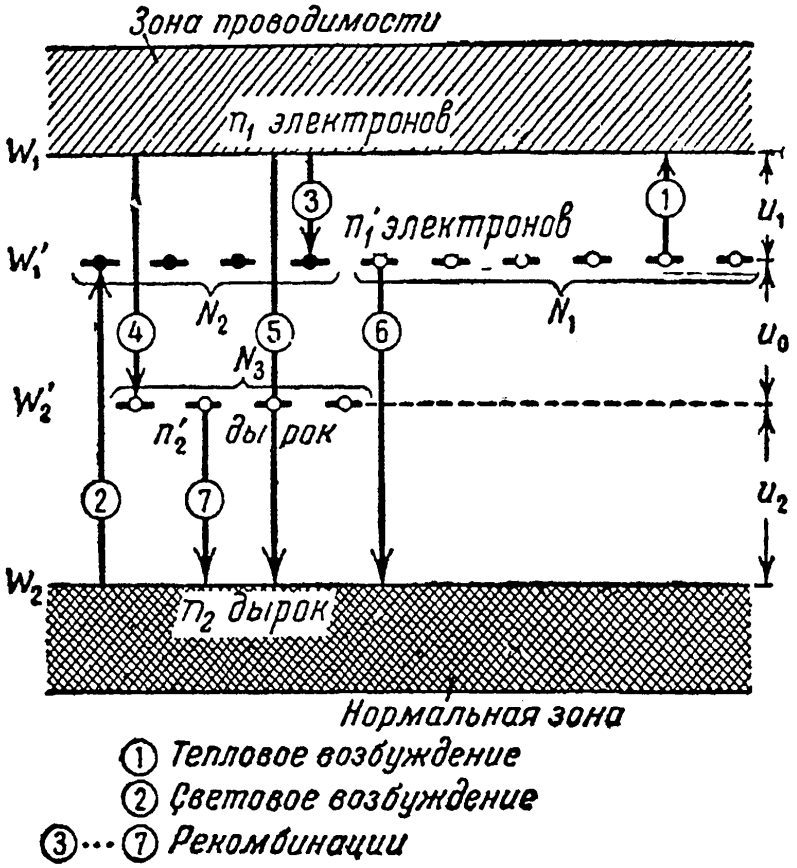


Рис. 113. Схема уровней.

электроны из зоны проводимости. Что касается тепловых перебросов электронов из нормальной зоны на уровни W_2', которые, вообще говоря, делаются возможными при включении освещения, то мы будем считать, полагая $u_1 \ll u_2$, что эти перебросы в рассматриваемой области температур практически отсутствуют. Таким образом, уровни W_2', не участвуя непосредственно ни в тепловом, ни в световом возбуждении, тем не менее принимают участие в игре, выполняя роль промежуточной станции в кругообращении электронов.

Примем, как и в § 1, следующие обозначения:

$$\begin{cases} n_1(T, Q) — \text{число электронов в зоне проводимости,} \\ n_1'(T, Q) — \text{число электронов на электроположительных уровнях,} \\ n_2'(T, Q) — \text{число дырок на электроотрицательных уровнях,} \\ n_2(T, Q) — \text{число дырок в нормальной зоне,} \end{cases}$$

где Q — количество квантов, поглощаемых единицей объёма в единицу времени; T — абсолютная температура. Очевидно, мы имеем:

$$\begin{cases} n_1(0, 0) = 0, \\ n_1'(0, 0) = N_1, \\ n_2'(0, 0) = n_2'(T, 0) = 0, \\ n_2(0, 0) = n_2(T, 0) = 0, \end{cases}$$

причём, конечно:

$$n_1(T, Q) + n_1'(T, Q) - n_2'(T, Q) - n_2(T, Q) = \\ = (n_1(0, 0) + n_1'(0, 0) - n_2'(0, 0) - n_2(0, 0) = N_1. \quad (3,1)$$

Пусть

$$\begin{cases} \sigma = \sigma(T, Q) — \text{проводимость при комбинированном действии тепла и света } (T \neq 0,\ Q \neq 0), \\ \sigma_0 = \sigma(T, Q) — \text{чисто тепловая проводимость в темноте } (Q = 0), \\ \Delta\sigma = (\sigma - \sigma_0) — \text{добавка в проводимости, вызванная освещением.} \end{cases}$$

Считая подвижность электронов в зоне проводимости и дырок в нормальной зоне примерно одинаковыми ($v_1 \approx v_2 \approx v$), мы будем иметь:

$$\begin{cases} \sigma = evn \ \ldots, \text{ где } n = n_1(T, Q) + n_2(T, Q), \\ \sigma_0 = evn_0 \ \ldots, \text{ где } n_0 = n_1(T, 0), \\ \Delta\sigma = ev\Delta n \ \ldots, \text{ где } \Delta n = n - n_0. \end{cases}$$

Таким образом, при отсутствии освещения мы имеем чисто электронную проводимость. При комбинированном действии тепла и света проводимость делается смешанной.

Положим, что ежесекундно в каждом кубическом сантиметре кристалла с заданных уровней на другие заданные

уровни переходит Q_i электронов, где значок i принимает значения

$$i=1,2,3,4,5,6,7,$$

соответственно тем переходам, которые изображены соответствующе перенумерованными стрелками на рис. 113. Переходы снизу вверх, вызываемые теплом ($i=1$) и светом ($i=2$), свидетельствуют об актах возбуждения; переходы сверху вниз ($i=3,4,5,6,7$), происходящие самопроизвольно, свидетельствуют об актах рекомбинации.

Условия равновесия имеют вид:

$$\left\{\begin{array}{ll} \text{зона проводимости} — & Q_1=Q_3+Q_4+Q_5,\\ \text{электроположительные уровни} — & Q_2+Q_3=Q_1+Q_6,\\ \text{электроотрицательные уровни} — & Q_4=Q_7,\\ \text{нормальная зона} — & Q_5+Q_6+Q_7=Q_2. \end{array}\right. \tag{3,2}$$

Здесь, в правых частях равенства, записаны электроны, уходящие с данных уровней, а в левых частях — электроны, приходящие на эти уровни. Мы имеем:

$$\left\{\begin{array}{l} Q_1=\alpha_1 n_1',\\ Q_2=Q=\varkappa Q_0=\alpha_2(N_1+N_2-n_1'),\\ Q_3=\alpha_3(N_1+N_2-n_1')\ n_1,\\ Q_4=\alpha_4 n_1 n_2',\\ Q_5=\alpha_5 n_1 n_2,\\ Q_6=\alpha_6 n_1' n_2,\\ Q_7=\alpha_7(N_2-n_2')\,n_2, \end{array}\right. \tag{3,3}$$

где $Q=Q_2$ — число поглощённых квантов (в 1 *сек* в 1 *см*^{3}), Q_0 — число падающих квантов (в 1 *сек* на 1 *см*^{2}), $\varkappa$ — коэффициент поглощения, α_i — коэффициенты пропорциональности, среди которых α_3, α_4, α_5, α_6, α_7 представляют собой коэффициенты рекомбинации различных сортов электронов с различными сортами дырок. Мы будем считать

$$\left.\begin{array}{c} \alpha_3\approx\alpha_4\approx\alpha_6\approx\alpha_7\approx\alpha,\\ \alpha_5=\beta. \end{array}\right\} \tag{3,4}$$

Иначе говоря, мы будем полагать, что условия рекомбинации примерно одинаковы для всех тех случаев, когда один из партнёров свободный, а другой связанный, но отличаются от усло-

вий рекомбинации для того случая, когда оба партнёра свободны.

На основании (3,3) и (3,4) уравнения (3,2) можно переписать так:

$$\left.\begin{aligned}
&\text{а)}\ \frac{\alpha_1}{\alpha} n_1' - (N_1 + N_2 - n_1')\, n_1 - n_1 n_2' - \frac{\beta}{\alpha} n_1 n_2 = 0,\\
&\text{б)}\ \frac{Q}{\alpha} + (N_1 + N_2 - n_1')\, n_1 - \frac{\alpha_1}{\alpha} n_1' - n_1' n_2 = 0,\\
&\text{в)}\ n_1 n_2' - (N_2 - n_2')\, n_2 = 0,\\
&\text{г)}\ \frac{\beta}{\alpha} n_1 n_2 + n_1' n_2 + (N_2 - n_2')\, n_2 - \frac{Q}{\alpha} = 0.
\end{aligned}\right\} \quad (3,5)$$

Среди этих четырёх уравнений, содержащих четыре неизвестных (n_1, n_1', n_2', n_2), одно является следствием трёх остальных. К системе (3,5) следует прибавить ещё уравнение (3,1), выражающее собой закон сохранения заряда:

$$n_1 + n_1' - n_2' - n_2 = N_1. \quad (3,6)$$

Система (3,5) — (3,6) может быть преобразована к следующему виду:

$$\left.\begin{aligned}
n_1' &= N_1 - (n_1 - n_2) + N_2 \frac{n_2}{n_1 + n_2},\\
n_2' &= N_2 \frac{n_2}{n_1 + n_2},\\
\frac{\alpha_1}{\alpha} &\left\{ \left\{ N_1 - (n_1 - n_2) + N_2 \frac{n_2}{n_1 + n_2} \right\} - \right.\\
&\left. - \{ N_2 + (n_1 - n_2) \}\, n_1 - \frac{\beta}{\alpha} n_1 n_2 = 0,\right.\\
\frac{Q}{\alpha} &- \{ N_1 + N_2 - (n_1 - n_2) \}\, n_2 - \frac{\beta}{\alpha} n_1 n_2 = 0.
\end{aligned}\right\} \quad (3,7)$$

Положим здесь

$$N_1 \gg N_2 \gg |n_1 - n_2|. \quad (3,8)$$

Это означает, во-первых, что мы ограничиваемся рассмотрением полупроводников, обладающих значительными стехиометрическими нарушениями [см. (1,3)]. Это означает, во-вторых, что ни температура, ни поглощение света не вызывают сколько нибудь заметного истощения поглощающих центров (что справедливо при не слишком больших T и не слишком больших Q). Принимая (3,8), мы можем переписать два последних урав-

нения (3,7) так:

$$\left.\begin{aligned} N_2 n_1 + \frac{\beta}{\alpha} n_1 n_2 &= \frac{\alpha_1}{\alpha} N_1, \\ N_1 n_2 + \frac{\beta}{\alpha} n_1 n_2 &= \frac{Q}{\alpha}, \end{aligned}\right\}$$

откуда

$$\left\{\begin{aligned} n_1 &= \frac{1}{2\alpha N_2}\left\{-\frac{\alpha}{\beta}\alpha N_1 N_2 - (Q-\alpha_1 N_1) + \right. \\ &\quad \left. + \sqrt{\left[\frac{\alpha}{\beta}\alpha N_1 N_2 + (Q-\alpha_1 N_1)\right]^2 + 4\frac{\alpha}{\beta}\alpha N_1 N_2 \alpha_1 N_1}\right\}, \\ n_2 &= \frac{1}{2\alpha N_1}\left\{-\frac{\alpha}{\beta}\alpha N_1 N_2 + (Q-\alpha_1 N_1) + \right. \\ &\quad \left. + \sqrt{\left[\frac{\alpha}{\beta}\alpha N_1 N_2 + (Q-\alpha_1 N_1)\right]^2 + 4\frac{\alpha}{\beta}\alpha N_1 N_2 \alpha_1 N_1}\right\}. \end{aligned}\right.$$

Полагая здесь, согласно (1,15б):

$$\frac{\alpha_1}{\alpha} = n_0 \frac{N_2}{N_1},$$

мы получаем:

$$\Delta n = n_1 + n_2 - n_0 = \frac{1}{2}\left(\frac{1}{N_2} + \frac{1}{N_1}\right)\left\{\frac{Q}{\alpha} - N_2 n_0 - \frac{\alpha}{\beta} N_1 N_2 + \right.$$
$$\left. + \sqrt{\left(\frac{Q}{\alpha} - N_2 n_0 + \frac{\alpha}{\beta} N_1 N_2\right)^2 + 4\frac{\alpha}{\beta} N_1 N_2 \cdot N_2 n_0}\right\} - \frac{Q}{\alpha N_2}. \quad (3,9)$$

Это выражение может быть положительным или отрицательным в зависимости от того, каковы соотношения между параметрами, входящими в (3,9). Отрицательные Δn соответствуют аномальному фотоэффекту. Полагая

$$\Delta n < 0$$

и подставляя сюда (3,9), мы получаем после элементарных алгебраических преобразований:

$$n_0 > \frac{Q}{\alpha N_1} + \frac{\alpha}{\beta} N_2. \quad (3,10)$$

Этим условием обеспечивается осуществление отрицательного эффекта.

Для того чтобы (3,10) имело место, мы должны считать

$$n_0 > \frac{\alpha}{\beta} N_2,$$

что возможно лишь в том случае, если

$$\frac{\beta}{\alpha} \gg 1,$$

ибо, согласно исходному предположению, $N_2 \gg n_0$. Таким образом, условием, необходимым (хотя и недостаточным) для того, чтобы отрицательный фотоэффект вообще был возможен, является требование, чтобы рекомбинация свободных электронов со свободными дырками была существенно облегчена по сравнению с рекомбинацией свободных электронов со связанными дырками или свободных дырок со связанными электронами.

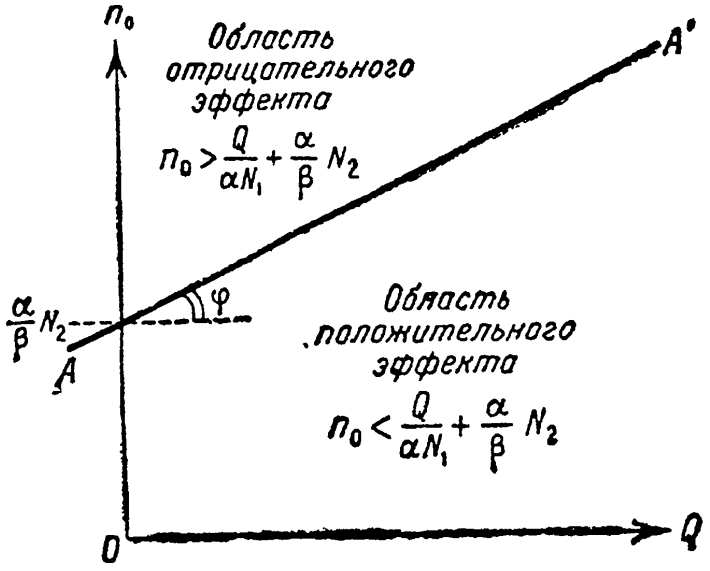


Рис. 114. Области положительного и отрицательного фотоэффектов. $\operatorname{tg}\varphi = 1/\alpha N_1$.

В. Области положительного и отрицательного фотоэффекта. Вся область значений параметров (n_0, Q) может быть разделена на области положительного и отрицательного фотоэффектов, как это показано на рис. 114. Точки, лежащие над прямой AA', соответствуют отрицательному фотоэффекту; точки, лежащие под этой прямой — положительному фотоэффекту. В одном и том же полупроводнике (если только для него соблюдается условие $\beta \gg \alpha$) может наблюдаться как положительный, так и отрицательный фотоэффект в зависимости от условий. Критическая точка, в которой фотоэффект меняет знак, определяется из уравнения:

$$n_0 = \frac{Q}{\alpha N_1} + \frac{\alpha}{\beta} N_2 \,. \qquad (3,11)$$

Положение этой точки зависит от природы полупроводника, от температуры, от концентрации повреждений, содержащихся в решётке, и от величины внешнего электрического поля, приложенного к полупроводнику. Роль этих факторов может быть

вскрыта, если в (3,11) положить, согласно (1,15б):

$$n_0 = \frac{N_1}{N_2} A e^{-\frac{u_1}{kT}},$$

где

$$A(T) = 2\left(\frac{2\pi mkT}{h}\right)^{3/2} \approx \text{const}.$$

Уравнение (3,11) при этом может быть переписано так:

$$kT_0 = \frac{u_1}{\lg\left\{\frac{AN_1}{\left(\frac{\beta}{\alpha}Q + \alpha N_1 N_2\right)} \cdot \frac{N_1}{N_2}\right\}}. \qquad (3,12)$$

Мы имеем (пренебрегая зависимостью A от T):

$$\begin{cases} \text{положительный фотоэффект при } T < T_0, \\ \text{отрицательный фотоэффект при } T > T_0. \end{cases}$$

Таким образом, при понижении температуры отрицательный фотоэффект должен исчезать, как это и наблюдается в действительности.

Из (3,19) видим, что критическая температура T_0 зависит от напряжённости внешнего электрического поля E. Действительно, при повышении E работа ионизации u_1 уменьшается (вследствие механизма штаркэффекта или механизма Френкеля), и, следовательно, точка T_0 сдвигается в сторону низких температур (если предполагать, что коэффициенты рекомбинации α и β не зависят или мало зависят от E). Таким образом, если при некоторой заданной температуре в отсутствии внешнего поля полупроводник обладает нормальным фотоэффектом, то при достаточно сильном поле в том же полупроводнике, при той же температуре и при тех же условиях освещения может наблюдаться аномальный эффект. Это также находится в качественном согласии с экспериментальными данными.

Из (3,12) видно также, что критическая температура T_0 может быть понижена при соответствующей обработке образца. По мере увеличения N_1 (при $N_2 = \text{const}$) или уменьшения N_2 (при $N_1 = \text{const}$), т. е. по мере удаления от стехиометрического соотношения, область отрицательного фотоэффекта рас-

пространяется в сторону низких температур. С другой стороны, по мере приближения реальной решётки к идеальной отрицательный фотоэффект сдвигается к высоким температурам и в идеальной решётке делается вообще невозможным.

В настоящее время отрицательный фотоэффект ещё недостаточно изучен экспериментально для того, чтобы можно было судить о том, насколько соответствует рассмотренный здесь механизм явлению, наблюдающемуся в действительности. Как бы то ни было, представляется интересным сам тот факт, что понижение проводимости при освещении полупроводника является теоретически возможным при определённых условиях.

С. Фотоэлектрическая активность в различных полосах поглощения. Возвратимся к выражению (3,9). Рассмотрим область положительного фотоэффекта. Если какой-либо один из трёх членов

$$N_2 n_0, \quad \frac{Q}{\alpha}, \quad \frac{\alpha}{\beta} N_1 N_2,$$

фигурирующих в (3,9), значительно больше двух остальных, так что двумя из них можно пренебречь по сравнению с третьим (наибольшим), то в этом случае, как легко проверить, выражение (3,9) апроксимируется следующим образом:

$$\Delta n \approx \frac{Q}{\alpha N_1}.$$

При этих условиях для фотоэлектрической активности полупроводника мы получаем:

$$\eta = \frac{Q}{\alpha n_0 N_1}. \tag{3,13}$$

Мы имеем

$$\begin{cases} \text{фотоэлектрически активное поглощение, если } \eta \gg 1; \\ \text{фотоэлектрически неактивное поглощение, если } \eta \ll 1. \end{cases}$$

Мы видим, сравнивая (3,13) с (2,8 б), что условия существования фотоэффекта выглядят по-разному в зависимости от того, производим ли мы освещение частотами, принадлежащими *B*-полосе (случай, рассматриваемый в настоящем параграфе),

или частотами, принадлежащими A-полосе (случай, рассмотренный в предыдущем параграфе). Действительно, обозначая через η_A и η_B фотоэлектрическую активность для соответствующих полос, мы можем переписать (2,8 б) и (3,13) так:

$$\begin{cases} \eta_A = \dfrac{Q}{\alpha N_2 n_0}, \\ \eta_B = \dfrac{Q}{\alpha N_1 n_0}. \end{cases}$$

Так как, согласно исходному предположению, $N_1 \gg N_2$, то

при $\eta_A \gg 1$ мы можем иметь $\begin{cases} \eta_B \gg 1, \\ \eta_B \ll 1. \end{cases}$

Это значит, что в одном и том же кристалле фотоэлектрическая активность различных полос поглощения может быть существенно различной, так что при значительном фотоэффекте

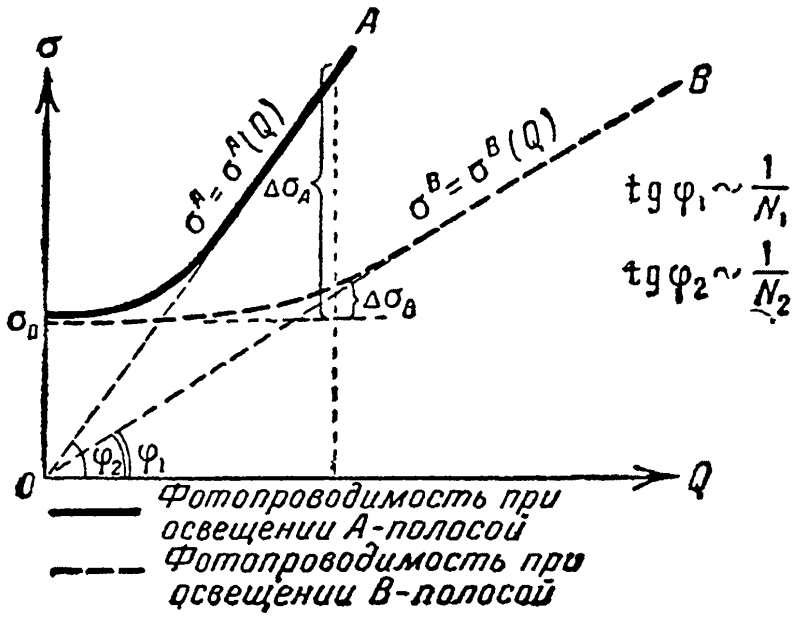


Рис. 115. Световая проводимость σ как функция количества поглощённых квантов (для разных полос поглощения).

в одной из полос в другой полосе (при тех же условиях освещения) фотоэффект может практически совсем отсутствовать. Это поясняется рис. 115, на котором изображена световая проводимость σ как функция Q в случае освещения A-полосой (жирная сплошная кривая) и в случае освещения B-полосой (жирная пунктирная кривая). Из рисунка видно, что при

освещении частотами, принадлежащими к различным полосам поглощения, одно и то же число поглощённых квантов может вызвать существенно различную по величине световую добавку $\Delta\sigma$.

Заметим, что этот факт обусловлен не различием в коэффициентах поглощения для обеих полос, а связан в конечном счёте с различием в условиях рекомбинации электронов. Действительно, если относить фотоэлектрическую активность не к одному и тому же числу поглощённых квантов Q, а к одному и тому же числу падающих квантов Q_0, то вместо (3,13) мы будем иметь, согласно (1,6) и (3,3):

$$\begin{cases} \eta_A = \dfrac{\varkappa_A Q_0}{\alpha N_2 n_0}, \\ \eta_B = \dfrac{\varkappa_B Q_0}{\alpha N_1 n_0}, \end{cases}$$

где $\varkappa_A$ и $\varkappa_B$ — коэффициенты поглощения для соответствующих частот в A-полосе и в B-полосе соответственно, причём

$$\begin{cases} \varkappa_A \approx \gamma N_1, \\ \varkappa_B \approx \gamma N_2 \end{cases}$$

и, следовательно,

$$\frac{\eta_A}{\eta_B} = \frac{\varkappa_A}{\varkappa_B} \cdot \frac{N_1}{N_2} = \frac{N_1^2}{N_2^2}.$$

Мы видим, что при переходе от полосы A к полосе B фотоэлектрическая активность ослабляется в N_1/N_2 раз быстрее, чем коэффициент поглощения.

Различие в фотоэлектрической активности для разных полос поглощения одного и того же кристалла — хорошо известный экспериментальный факт. Он имеет место, например, в щёлочно-галоидных кристаллах, у которых F-полоса даёт значительный фотоэффект, в то время как в U-полосе фотоэффект обычно вовсе не наблюдается. С точки зрения представлений об экситонах этот факт вряд ли может быть объяснён, поскольку фотоэлектрически неактивная полоса (U-полоса) лежит в более коротковолновой части спектра, чем активная полоса (F-полоса). Действительно, в рамках этих представлений это означало бы, что для ионизации требуется меньшая энергия, чем для возбуждения.

Литература к главе VI

[1] Gudden, Lichtelektrische Erscheinungen, Berlin, 1928.
[2] П. С. Тартаковский, Внутренний фотоэффект в диэлектриках, ГТТИ, 1940.
[3] А. В. и А. Ф. Иоффе, *ЖЭТФ*, **9**, 1428, 1939.
[4] Coblentz and Kahler, *U. S. Bur. Stand. Bull.*, **15**, 121, 1919.
[5] Elliott, *Phys. Rev.*, **5**, 50, 1915.
[6] Юз и Дюбридж, Фотоэлектрические явления, ОНТИ, 1936.
[7] Ries. *Ph. Zs.*, **9**, 569, 1908.
[8] Brown, *Phys. Rev.*, **33**, 1, 1911.
[9] Coblentz, *U. S. Bur. Stand. Bull.* **16**. 597, 1920.
[10] Ф. Ф. Волькенштейн, *Уч. Зап. ЛГПИ им. Герцена*, 33, 1941.
[11] Ф. Ф. Волькенштейн, *ЖЭТФ*, **10**, 1229, 1940.

Опечатки.

Стр.	Строка	Напечатано	Должно быть	По чьей вине
172	4 сн.	$g' = g - 1,\ g \neq kN - 1$	$g' = g - 1 \neq kN - 1$	Ред.
194	11 св.	дырок	электронов	»
222	ф-ла (2,32)	$e^{-\frac{x_j}{kF}} = \frac{\lvert\nu\rvert}{a} e^{-\frac{u_j}{kT}}$	$e^{-\frac{x_j}{kT}} = \frac{\lvert\nu\rvert}{a}$	Авт.