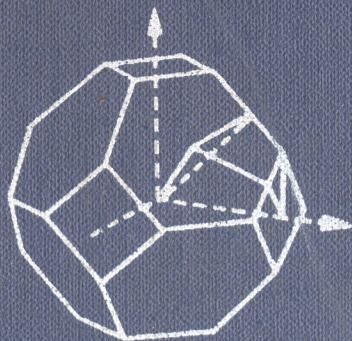
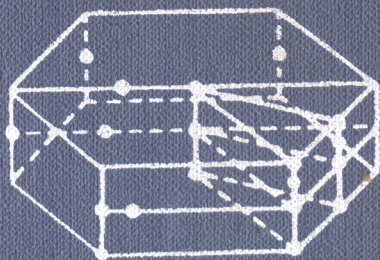


П. С. КИРЕЕВ • ФИЗИКА ПОЛУПРОВОДНИКОВ •

П. С. КИРЕЕВ

ФИЗИКА ПОЛУПРОВОДНИКОВ



531.9

К 43

УДК 537.311(0.75)

Рецензент — докт. физ.-мат. наук, проф. Фистуль В. И.

Киреев П. С.

К 43 Физика полупроводников. Учеб. пособие
для втузов. М., «Высш. школа», 1975.

584 с. с ил.

В книге рассматриваются элементы электронной теории проводимости и проводимости полупроводников; зонная теория на основе теории возмущения; статистика электронов и дырок; кинетические явления в полупроводниках; теория рассеяния, контактные и неравновесные явления на основе уравнения непрерывности; теории оптических и фотоэлектрических явлений в полупроводниках.

К $\frac{30407-125}{001(01)-75}$ 172-75

531.9

© Издательство «Высшая школа», 1975 г.

Предисловие ко второму изданию

Первое издание «Физики полупроводников», выпущенное в 1969 г., быстро разошлось. Широкое использование книги в учебном процессе и специалистами в практической работе показало целесообразность второго издания пособия. Поскольку за годы, прошедшие после выхода в свет первого издания, кардинальных изменений наших представлений о физических явлениях в полупроводниках не произошло, материал не подвергался существенной переработке. Основные изменения связаны с добавлениями, в которых либо раскрываются большие возможности математических соотношений для анализа физических явлений, либо освещается физическое содержание полученных выводов. Заметно увеличено число примеров экспериментальных зависимостей. Добавлены два новых параграфа, в которых рассматриваются эффект Фарадея и спин-орбитальное расщепление уровней и зон энергии.

Приношу искреннюю благодарность всем товарищам за доброжелательную критику и конструктивные предложения, способствующие улучшению книги.

Автор

Предисловие к первому изданию

В последние годы нашими издательствами выпущен целый ряд превосходных монографий, в которых нашли отражение сущность различных физических явлений в полупроводниках и методы их теоретического и экспериментального изучения. К числу таких пособий относятся монографии А. И. Ансельма «Введение в теорию полупроводников», С. М. Рывкина «Фотоэлектрические явления в полупроводниках», В. С. Вавилова «Действие излучений на полупроводники», Я. Тауца «Фото- и термоэлектрические явления в полупроводниках», Р. Бьюба «Фотопроводимость твердых тел» и многие другие. В большом числе пособий читатель может найти практически все, что нам известно в области физики полупроводников и их применения. Однако будучи крайне полезной для научных работников и лиц, изучающих определенный вопрос, монографическая литература оказывается мало пригодной для студентов, приступающих к изучению физики полупроводников. Для этого необходимы учебник или учебное пособие, в котором нашел бы отражение наиболее важный материал, считающийся твердо установленным. Создание учебника в настоящее время необходимо и возможно, однако в качестве первого шага реальнее учебное пособие для студентов таких специальностей, как «Полупроводники и диэлектрики», «Полупроводниковые приборы» и некоторых других.

В основу данного учебного пособия была положена программа по физике полупроводников, чем определились отбор и распределение материала книги.

Книга написана как учебное пособие для студентов, это наложило определенный отпечаток на метод изложения материала. В ней практически отсутствует экспериментальный материал, связанный с описанием свойств конкретных полупроводниковых веществ. Это обусловлено тем, что объем информации в настоящее время крайне велик и к тому же он непрерывно возрастает, поэтому сведения, полученные сегодня, завтра могут оказаться устаревшими. Учебник или учебное пособие в отличие от обзоров и монографий должен содержать общие для всех веществ вопросы, одинаково пригодные для объяснения физических явлений в любом веществе. И таким общим вопросом является метод теоретического анализа явлений. Это придает учебному пособию ярко выраженный теоретический характер, который подчеркивается широким использованием метода теоретической физики. Для освоения студентами не только результатов, но и методов теоретического анализа физических явлений

в полупроводниках, в книге дается подробный вывод основных соотношений с приведением всех промежуточных преобразований. Это необходимо по двум соображениям. Во-первых, литература по полупроводникам грешит излишне частым употреблением непригодных для учебного пособия выражений типа «очевидно», «как известно», «легко получить» и т. п. Записав такое «очевидное», соотношение, студент не только не знает, как оно получено (это еще полбеды!), но и не представляет вследствие этого границ применимости «очевидного» соотношения, пытаясь впоследствии применить его в случаях, для которых оно непригодно. Недопустимость этого действительно очевидна.

Во-вторых, издание учебного пособия, в котором содержится подробный вывод основных соотношений, позволит лектору опустить их, оставить для самостоятельной проработки студентами, посвятив больше времени изложению последних экспериментальных результатов.

Данное пособие написано на основе лекций по физике полупроводников, которые читались студентам факультета полупроводниковых материалов и приборов Московского института стали и сплавов. Как показал опыт работы, предлагаемый уровень изложения материала успешно усваивается студентами.

Общая подготовка, получаемая студентами в вузах, вполне позволяет им изучать теоретические работы «средней» сложности и проводить теоретический анализ экспериментального материала на более высоком уровне, чем это имеет место в настоящее время.

Более того, в течение нескольких ближайших лет уровень обучения должен быть поднят еще выше. Необходимо будет ознакомить студентов, например, с такими методами теоретического исследования, как метод функции Грина, матрица плотности, метод вторичного квантования, теория групп, которые все шире используют в физике полупроводников и позволяют получать результаты, недоступные при других методах исследования.

Как уже указывалось выше, результаты экспериментальных работ в данном пособии использованы в основном в иллюстративных целях. Это с неизбежностью приводит к тому, что теряется возможность показать расхождение между теорией и экспериментом. Несомненно, в дальнейшем необходимо провести сопоставление теории и эксперимента. Однако в настоящее время важнее научить студентов пользоваться, например, кинетическим уравнением или теорией возмущений, показать им сильные стороны существующих методов теоретического анализа, после чего студенты легко и правильно поймут и их слабые стороны.

Автор исходит при этом из того, что физику полупроводников можно и нужно излагать как стройную физическую науку, а не как собрание противоречивых экспериментальных фактов.

При выборе материала автор учитывал тот факт, что физика полупроводников изучается после прохождения студентами теоретической физики, физической кристаллографии и некоторых других

предметов, поэтому в данном учебном пособии отсутствуют все вопросы, связанные со структурой твердых тел, характером взаимодействия атомов, которые обычно излагаются в пособиях по физике полупроводников. С другой стороны, студенты изучают физику полупроводниковых приборов, поэтому эти вопросы также не вошли в данное пособие. Физика полупроводников, базируясь на некоторых дисциплинах и будучи основой для изучения других предметов, является в то же время самостоятельным учебным курсом со своими задачами и методами исследования.

Автор выражает глубокую благодарность Е. В. Загорянской за всестороннюю помощь и поддержку.

Автор приносит благодарность сотрудникам и аспирантам кафедры физики полупроводников МИСиС за помощь в работе над созданием курса лекций и данного пособия, сотрудникам и аспирантам кафедры диэлектриков и полупроводников Ленинградского электротехнического института им. В. И. Ульянова (Ленина) за ценные советы при обсуждении рукописи, доценту А. Ф. Городецкому за большое число ценных пожеланий и замечаний.

Автор благодарит профессора К. В. Шалимову, без помощи которой создание курса было бы невозможно.

Автор будет благодарен за все замечания, которые могут способствовать улучшению учебного пособия, за советы и пожелания по отбору материала и методике его изложения.

Автор

ЭЛЕКТРОННАЯ ТЕОРИЯ ПРОВОДИМОСТИ

§ 1. ЭЛЕКТРОННАЯ ТЕОРИЯ ПРОВОДИМОСТИ.
ЗАКОН ОМА

Многие понятия физики полупроводников основаны на электронной теории металлов. Прежде всего это относится к таким понятиям, как электрическая проводимость и подвижность носителей заряда.

Электронная теория проводимости металлов, развитая в прошлом веке, рассматривает электронный газ как газ, находящийся в тепловом равновесии с решеткой кристалла. Предполагается, что электронный газ подобен идеальному газу молекулярной физики: *он не имеет собственного объема и электроны не взаимодействуют друг с другом*. Состояние движения каждой частицы определяется шестью величинами: тремя координатами x , y , z и тремя компонентами скорости v_x , v_y , v_z (или импульса p_x , p_y , p_z) или двумя векторными величинами \mathbf{r} и \mathbf{v} (или \mathbf{p}). Предположение о малости собственного объема кажется оправданным, так как по классической теории радиус электрона $r_0 \approx 10^{-15}$ м, объем $V_0 = 10^{-45}$ м³. Если число электронов в единице объема $n \approx 10^{28}$ м⁻³, то собственный объем электронов b составляет $b = nV_0 = 10^{-17}$ от объема тела. Однако предположение о невзаимодействии электронов кажется совершенно необоснованным. Действительно, заряд электронов равен $e = 1,6 \cdot 10^{-19}$ Кл, сила, с которой они взаимодействуют, находясь на расстоянии 10^{-10} м, равна $2 \cdot 10^{-8}$ Н. Ускорение, которое должен получить электрон при действии этой силы, имеет величину $2 \cdot 10^{22}$ м/с², а энергия кулоновского взаимодействия двух электронов при $r = 10^{-10}$ м составляет около 14 эВ.

Полная энергия кулоновского взаимодействия (отталкивания) электронов должна была бы достигать колоссальной положительной величины. Однако, как показывает опыт, энергия электронов в металлах отрицательна (по отношению к энергии удаленного на бесконечность электрона). Это связано с тем, что помимо кулоновского отталкивания электронов существует кулоновское притяжение между электронами и ядрами. Сила и энергия взаимодействия электронов с ядрами имеют тот же порядок величины, что и при взаимодействии электронов друг с другом. Двигаясь в поле всех электронов и ядер, каждый электрон испытывает как притяжение, так и отталкивание. В результате этих двух видов взаимодействия и создается «кажу-

щаяся независимость» движения отдельных электронов. Как будет показано во второй главе, законы квантовой механики действительно позволяют рассматривать электроны как невзаимодействующие частицы.

Электроны двигаются в кристалле хаотически. При этом они «сталкиваются» с ионами решетки, что приводит к изменению скорости электронов как по *модулю*, так и по *направлению*. Но изменение модуля скорости электрона связано с изменением его кинетической энергии. В условиях термодинамического равновесия температура электронного газа должна быть равной температуре ионов решетки. Это означает, что в среднем не происходит передачи энергии ни от электронов к решетке, ни от решетки к электронам.

Но если изменить температуру электронного газа, то вследствие обмена энергией между электронами и ионами должна измениться и температура решетки. Этот факт играет важную роль при объяснении проводимости металлов и полупроводников, который мы используем в дальнейшем.

В результате случайного характера рассеяния электронов при столкновении их с решеткой средняя скорость *одного* электрона за длительный промежуток времени и его среднее перемещение, рассматриваемые как векторные величины, должны быть равны нулю. Все электроны находятся в одинаковых условиях, следовательно, это утверждение справедливо для *любого* электрона. Поскольку среднее перемещение электронов при хаотическом (тепловом) движении равно нулю, то хаотическое движение не может вызвать электрического тока, который характеризует перенос заряда через некоторое сечение. Для создания тока необходимо направленное движение электронов, которое может быть вызвано различными факторами: электрическим полем, градиентом температуры, неоднородным освещением и некоторыми другими причинами.

Если создать в металле электрическое поле напряженности E , то электроны будут ускоряться полем. Ускорение, сообщаемое полем электрону, равно

$$a = \frac{e}{m} E. \quad (1.1)$$

За время t электрон приобретает скорость

$$v = at = \frac{et}{m} E, \quad (1.2)$$

направленную против поля. Если начальная скорость электрона v_T , то скорость в момент времени t будет равна

$$at + v_T = \frac{et}{m} E + v_T. \quad (1.3)$$

Мы видим, что составляющая скорости электронов по полю уменьшается, а против поля — увеличивается, и совокупность электронов получает некоторую скорость направленного движения. Элек-

троны, двигаясь хаотически, совершают в то же время движение против поля. *Направленное движение совокупности электронов в электрическом поле называется дрейфом, а скорость направленного движения называется дрейфовой скоростью и обозначается v_d .* За время t под действием поля E электрон переместится на расстояние l :

$$l(t) = \frac{et^2}{2m} E. \quad (1.4)$$

В классической электронной теории предполагается, что изменение скорости происходит в результате кратковременного акта взаимодействия электрона с решеткой (с атомами или ионами решетки). Другими словами, предполагается, что взаимодействие электрона с решеткой подобно явлению удара в механике. Между двумя соударениями электрон движется как свободная частица, не испытывая влияния поля решетки и остальных электронов. Для характеристики движения электрона вводятся понятия времени τ и длины l свободного пробега; τ имеет смысл среднего времени между двумя соударениями, а l — среднего пути.

Средняя длина свободного пробега l связана со средним временем свободного пробега τ соотношением

$$l = v_T \tau, \quad (1.5)$$

где v_T — среднеарифметическая скорость теплового движения электронов, т. е. среднее значение модуля скорости.

Найдем среднюю скорость дрейфа электронов в электрическом поле.

При $t = 0$ скорость направленного движения электрона равна нулю, а при $t = \tau$ равна соответственно

$$a\tau = \frac{e\tau}{m} E. \quad (1.6)$$

Скорость дрейфа будет равна средней скорости направленного движения, т. е. полусумме начальной и конечной скоростей:

$$v_d = \frac{0 + a\tau}{2} = \frac{e\tau}{2m} E. \quad (1.7)$$

Из (1.7) видно, что средняя скорость направленного движения пропорциональна напряженности электрического поля E . *Величина, связывающая дрейфовую скорость с напряженностью поля, называется подвижностью электронов и обозначается обычно буквой μ :*

$$\mu = \frac{e\tau}{2m}; \quad v_d = \mu E. \quad (1.8)$$

Подвижность электронов численно равна скорости дрейфа в электрическом поле единичной напряженности.

Если концентрация электронов равна n , то в единицу времени через единичное сечение пройдет заряд, заключенный в объеме параллелепипеда с единичным сечением и длиной $1 \cdot v_d$. Поскольку вели-

чина, равная заряду, проходящему в единицу времени через единичное сечение, называется плотностью тока, для плотности тока j можем записать

$$j = env_d = en\mu E = \sigma E. \quad (1.9)$$

Соотношение (1.9) представляет собой закон Ома в дифференциальной форме. Для удельной электрической проводимости из (1.9) и (1.8) имеем

$$\sigma = en\mu \quad (1.10)$$

и

$$\sigma = e^2 n \tau / 2m. \quad (1.11)$$

Выражение (1.11) было получено Друде. Если выразить τ из (1.5), то выражение (1.11) можно записать в таком виде:

$$\sigma = e^2 n l / 2m v_T. \quad (1.12)$$

Закон Ома справедлив, пока электрическое поле не меняет концентрацию n и подвижность μ электронов. Однако с ростом поля E концентрация электронов и их подвижность могут измениться. Покажем это на примере подвижности. При выводе закона Ома мы предположили, что скорость направленного движения и связанная с ним энергия полностью передаются решетке. В слабых электрических полях дрейфовая скорость много меньше тепловой скорости, поэтому τ от поля E не зависит. Но с ростом поля дрейфовая скорость может быть сравнима с тепловой. В этом случае время свободного пробега будет уменьшаться, поскольку теперь

$$\tau = \frac{l}{v_T + v_d}, \quad (1.13)$$

что приведет к уменьшению подвижности и проводимости металла. Необходимая для этого критическая величина поля $E_{кр}$ будет тем меньше, чем меньше v_T , т. е. температура тела, и чем больше подвижность электронов в слабых полях E .

Время свободного пробега можно интерпретировать несколько иначе. Если выключить в некоторый момент времени электрическое поле, то совокупность электронов будет продолжать направленное движение до тех пор, пока в результате столкновений она не передаст решетке всю приобретенную в поле скорость. Это направленное движение прекратится в среднем через время τ . После этого электроны вернуться в состояние хаотического, теплового движения. *Столкновения приводят совокупность электронов в равновесное состояние, в то время как электрическое поле приводит к нарушению равновесного состояния. Переход некоторой системы из неравновесного состояния в равновесное называют релаксационным процессом, или релаксацией, а время восстановления нарушенного равновесия называют временем релаксации.* Таким образом, можно сказать, что время свободного пробега представляет собой время релаксации.

В системе единиц СИ электропроводность измеряется в сименсах (Сим), а удельная электрическая проводимость — в Сим/м. Размерность подвижности в СИ можно найти из (1.8):

$$[\mu] = (A \cdot c^2) \text{ кг}^{-1}.$$

Подвижность можно измерять также величиной скорости в единичном поле, т. е. величиной с наименованием м/с: $V/m = m^2/(V \cdot c)$. В лабораторной практике чаще используют внесистемные единицы (см, В, с) и измеряют подвижность в $\text{см}^2/(V \cdot c)$, а удельную электрическую проводимость в $\text{Ом}^{-1} \text{ см}^{-1}$. Очевидно, что

$$1 \text{ м}^2/(V \cdot c) = 10^4 \text{ см}^2/(V \cdot c);$$

$$1 \text{ Сим/м} = 10^{-2} 1/(\text{Ом} \cdot \text{см}).$$

Резюме § 1

1. Основные положения классической электронной теории проводимости сводятся к следующему:

а) электроны образуют идеальный (электронный) газ, находящийся в хаотическом тепловом движении, которое характеризуется средней длиной свободного пробега l и средним временем свободного пробега τ ;

б) электроны обмениваются энергией и импульсом с ионами решетки, благодаря чему электронный газ находится в термодинамическом равновесии с решеткой;

в) электрическое поле сообщает электронам направленное движение, тем самым вызывает электрический ток.

2. Плотность тока пропорциональна напряженности электрического поля:

$$j = \sigma E. \quad (1.1p)$$

3. Удельная электрическая проводимость связана с концентрацией n и подвижностью μ электронов соотношением

$$\sigma = en\mu. \quad (1.2p)$$

4. Подвижность определяется характеристиками электрона и хаотического движения (в слабых полях):

$$\mu = e\tau/2m = el/2m v_T. \quad (1.3p)$$

5. Подвижность численно равна скорости дрейфа в электрическом поле единичной напряженности:

$$\mu = v_d/E. \quad (1.4p)$$

6. В сильных электрических полях закон Ома может нарушаться.

7. Время свободного пробега можно рассматривать как время релаксации.

§ 2. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ВРЕМЕН И ДЛИН СВОБОДНОГО ПРОБЕГА

Выражения для подвижности и проводимости, полученные в предположении одинаковости времени свободного пробега для всех электронов, необходимо преобразовать, учитывая, что в действительности будут наблюдаться самые различные времена свободного пробега, меняющиеся от 0 до ∞ . Для этого необходимо знать вероятности определенных значений времен свободного пробега, которые рассматриваются как случайные величины. Найдем функцию распределения времен свободного пробега.

Для этого предположим, что:

1) *вероятность испытать столкновение (рассеяние) в течение интервала времени dt для электрона пропорциональна величине интервала dt ;*

2) *вероятность столкновения в единицу времени — постоянная во времени величина.*

Этих двух предположений достаточно для получения функции распределения времен свободного пробега. Обозначим вероятность того, что частица двигалась без столкновения в течение времени $t, t + dt$, через

$$dw = dw(dt). \quad (2.1)$$

Величина $w(t)$ есть вероятность свободного движения в интервале времени $(t, t + 1)$, а $w(t + dt)$ есть вероятность свободного движения в интервале $t + dt, t + dt + 1$. Но величину $w(t + dt)$ можно выразить двояко. С одной стороны,

$$w(t + dt) = w(t) + \frac{dw}{dt} dt. \quad (2.2)$$

С другой стороны, факт свободного движения в течение времени $t + dt$ — событие C — можно рассматривать как произведение двух событий: A — свободного движения в течение времени t и B — свободного движения в течение времени dt ; так что

$$C = AB. \quad (2.3)$$

Вероятность произведения двух событий есть произведение вероятности одного на условную вероятность другого:

$$w(C) = w(A) w(B/A) = w(B) w(A/B). \quad (2.4)$$

Но так как событие A не зависит от события B , то $w(A/B) = w(A)$. Вместе с тем событие B оказывается не зависящим от A^* , поэтому

$$w(t + dt) = w(t) \cdot dw(dt). \quad (2.5)$$

* Утверждение о независимости события B от события A (в нашем случае), как правило, вызывает у изучающих резкое возражение, в то время независимость события A от B принимается как очевидный факт. Основным аргументом обычно выдвигается утверждение, что если частицы рассеиваются за время, мень-

Вероятность свободного движения за время dt можно выразить через вероятность рассеяния за то же время. Обозначим вероятность рассеяния (соударения) за единицу времени через a , тогда вероятность рассеяния за время dt будет равна adt , а вероятность свободного движения будет равна $1 - adt$, т. е.

$$dw(dt) = 1 - adt. \quad (2.6)$$

Учитывая выражения (2.2), (2.5) и (2.6), можно записать

$$w(t + dt) = w(t) + \frac{dw}{dt} dt = w(t) [1 - adt] = w(t) - w(t) adt, \quad (2.7)$$

что приводит к дифференциальному уравнению для функции $w(t)$

$$dw/dt = -wa. \quad (2.8)$$

Решая уравнение (2.8), получим

$$w(t) = ce^{-at}. \quad (2.9)$$

Константа интегрирования c определяется из условия нормировки

$$\int_0^{\infty} w(t) dt = 1 = c \int_0^{\infty} e^{-at} dt = \frac{c}{a}, \quad (2.10)$$

откуда

$$c = a. \quad (2.11)$$

Таким образом, *нормированная функция распределения времен свободного пробега имеет вид*

$$w(t) = ae^{-at}. \quad (2.12)$$

Найдем среднее время свободного пробега $\langle t \rangle$:

$$\langle t \rangle = \int_0^{\infty} tw(t) dt = \int_0^{\infty} tae^{-at} dt = \frac{1}{a}. \quad (2.13)$$

Если обозначить среднее время свободного пробега через τ : $\langle t \rangle = \tau$, то согласно (2.13) *вероятность столкновения за единицу времени равна обратному среднему времени свободного пробега:*

$$a = \frac{1}{\tau} = \tau^{-1}. \quad (2.14)$$

Нормированная к единице функция распределения может быть записана в виде

$$w(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}. \quad (2.15)$$

шее t , то ни о каком движении за время dt после t говорить нельзя. Но в этом случае нет не только события B , но и C и даже A , в то время как теорема формулируется для события $AB = C$, а произведение двух событий есть осуществление обоих событий — сомножителей.

Аналогичным образом можно найти функцию распределения длин свободного пробега x :

$$\omega(x) = \frac{1}{l} e^{-\frac{x}{l}}, \quad (2.16)$$

где l — средняя длина свободного пробега.

Полученные функции распределения справедливы для самых общих случаев. Применим их для случая движения электронов в электрическом поле. Нас будет интересовать при этом только направленное движение электронов против поля. Пусть ось x совпадает по направлению с полем. Электрон, двигаясь в поле, за время свободного движения t приобретает скорость

$$v(t) = \frac{eE}{m} t \quad (2.17)$$

и проходит путь длиной x :

$$x = \frac{eE}{2m} t^2; \quad x = \frac{eE}{2m} t^2. \quad (2.18)$$

Средняя скорость дрейфового движения

$$v_d = \int_0^{\infty} v(t) \omega(t) dt = \frac{eE}{m} \int_0^{\infty} t e^{-\frac{t}{\tau}} \frac{dt}{\tau} = \frac{e\tau}{m} E, \quad (2.19)$$

а среднее перемещение против поля

$$l = \int_0^{\infty} x(t) \omega(t) dt = \frac{eE}{2m} \int_0^{\infty} t^2 e^{-\frac{t}{\tau}} \frac{dt}{\tau} = \frac{e\tau^2}{m} E. \quad (2.20)$$

Мы видим, что скорость дрейфа равна

$$v_d = \frac{e\tau}{m} E = \mu E, \quad (2.21)$$

где

$$\mu = \frac{e\tau}{m} \quad (2.22)$$

есть подвижность.

Выражение (2.22) для подвижности отличается от ранее найденного значения (1.8) множителем 2, обусловленного тем, что в (2.2) учтен бóльший вклад движений с бóльшими временами свободного пробега. Это особенно отчетливо видно из выражения для среднего смещения:

$$\langle x \rangle = \frac{e\tau^2}{m} E \quad \left(\text{но не } \frac{e\tau^2}{2m} E \right), \quad (2)$$

так как всегда

$$\langle t^2 \rangle \geq \langle t \rangle^2;$$

в данном случае

$$\langle t^2 \rangle = 2 \langle t \rangle^2 = 2\tau^2.$$

В заключение этого параграфа необходимо сделать одно простое, но очень существенное замечание: τ есть среднее время свободного пробега, т. е. время между двумя соударениями. Оно определяется длиной свободного пробега и полной скоростью частицы, но не скоростью дрейфового движения. Полная скорость зависит от энергии частицы. Тем самым *время свободного пробега является функцией энергии частицы*. Эта зависимость усложняется в том случае, когда длина свободного пробега также зависит от энергии. Поэтому для нахождения дрейфовой скорости, как и целого ряда других характеристик вещества, необходимо усреднять времена свободного пробега и другие величины с учетом функции распределения электронов по энергии, что будет сделано в IV главе.

Резюме § 2

1. Для получения функции $\omega(t)$ достаточно предположить, что:
 а) вероятность рассеяния в течение времени dt пропорциональна dt ;
 б) существует независимая от времени вероятность рассеяния в единицу времени, равная a .

2. Времена свободного пробега вследствие случайного характера столкновения могут принимать различные значения. Вероятность того, что время свободного пробега лежит в пределах $t, t + 1$, равна

$$\omega(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}. \quad (2.1p)$$

3. Среднее время свободного пробега τ и вероятность соударения за единицу времени a связаны между собой соотношением

$$a\tau = 1. \quad (2.2p)$$

4. Вероятность того, что длина свободного пробега лежит в пределах $x, x + 1$, равна

$$\omega(x) = \frac{1}{l} e^{-\frac{x}{l}}. \quad (2.3p)$$

5. Средняя длина свободного пробега l связана с вероятностью рассеяния на единице длины b соотношением

$$bl = 1.$$

6. Если учесть статистический разброс времен свободного пробега и учесть, что большие времена свободного пробега в силу квадратичной зависимости пути от времени дают больший вклад в проводимость, то для подвижности получим выражение

$$\mu = e\tau/m. \quad (2.4p)$$

**§ 3. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ЭЛЕКТРОНОВ ПО СОСТОЯНИЯМ.
СРЕДНИЕ ЗНАЧЕНИЯ ФИЗИЧЕСКИХ ВЕЛИЧИН**

Пространство координат x, y, z и импульсов p_x, p_y, p_z называют фазовым пространством. Рассмотрим в фазовом пространстве элемент объема

$$d\Gamma = dx dy dz dp_x dp_y dp_z = d\tau_r d\tau_p, \quad (3.1)$$

построенного вокруг точки с координатами $(x, y, z; p_x, p_y, p_z)$. Число электронов dN , находящихся в этом элементе объема, должно быть пропорционально величине объема $d\Gamma$, если $d\Gamma$ достаточно мало. Кроме того, это число должно зависеть от точки (\mathbf{r}, \mathbf{p}) фазового пространства, оно определяется некоторой функцией $F(\mathbf{r}, \mathbf{p}, t)$, так что

$$dN = F(\mathbf{r}, \mathbf{p}, t) d\Gamma. \quad (3.2)$$

Интегрируя по всему фазовому пространству, получим полное число электронов N :

$$N = \int_{(V_\Gamma)} F(\mathbf{r}, \mathbf{p}, t) d\Gamma. \quad (3.3)$$

Функция $F(\mathbf{r}, \mathbf{p}, t)$ определяет число электронов в единице объема фазового пространства. Вместо нее можно ввести функцию $f(\mathbf{r}, \mathbf{p}, t)$ в N раз меньшую:

$$f(\mathbf{r}, \mathbf{p}, t) = \frac{1}{N} F(\mathbf{r}, \mathbf{p}, t). \quad (3.4)$$

Функция $f(\mathbf{r}, \mathbf{p}, t)$ определяет вероятность нахождения одного электрона в единичном объеме фазового пространства, а равенство

$$\int_{(V_\Gamma)} f(\mathbf{r}, \mathbf{p}, t) d\Gamma = 1 \quad (3.5)$$

называют условием нормировки.

Интегрирование проводится по объему кристалла V и по всем возможным значениям импульса. Условие нормировки не зависит от времени, функция же распределения в общем случае зависит от времени. С помощью функции распределения можно вычислить средние значения произвольной физической величины, зависящей от координат и импульса. Напомним, как производится это вычисление. Пусть α — некоторая величина, которая может зависеть от состояния движения электрона: $\alpha = \alpha(\mathbf{r}, \mathbf{p}, t)$. Возьмем элемент фазового объема $d\Gamma$. В нем находится

$$dN = N f(\mathbf{r}, \mathbf{p}, t) d\Gamma \quad (3.6)$$

электронов, каждый из которых обладает величиной $\alpha(\mathbf{r}, \mathbf{p}, t)$.

Суммарное значение физической величины для dN электронов равно

$$\alpha dN. \quad (3.7)$$

Если проинтегрировать (3.7) по всему фазовому пространству, то получим суммарное для всех электронов значение величины α ; разделив эту величину на N , получим среднее значение $\langle \alpha \rangle$:

$$\langle \alpha \rangle = \frac{1}{N} \int_{(V_r)} \alpha dN = \int_{(V_r)} \alpha(r, p, t) f(r, p, t) d\Gamma. \quad (3.8)$$

Если величину (3.7) проинтегрировать только по импульсу и разделить на $N d\tau_r$, то получим среднее значение величины $\alpha(r)^p$, которую имеют электроны в окрестности точки r .

Отметим один практически важный частный случай. Пусть функция распределения не зависит от координат. В этом случае интегрирование по координатам соотношения (3.5) дает объем тела V , вследствие чего интегрирование по импульсу функции распределения дает обратный объем. Если α не зависит от координат r , то усреднение только по импульсу дает величину $\bar{\alpha}^p = \frac{\langle \alpha \rangle^p}{V}$, где V — объем тела.

Плотность тока можно выразить через функцию распределения следующим образом:

$$j = en \langle v \rangle = enV \int_{-\infty}^{\infty} v f(r, p, t) d\tau_p = env_d. \quad (3.9)$$

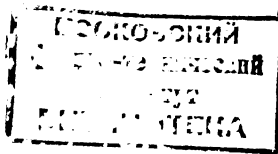
Если $f(r, p, t) = f(r, -p, t)$, что означает равновероятность движений со скоростями v и $-v$, то средняя скорость $\langle v \rangle$ в силу нечетности функции $\frac{p}{m} f(r, p, t)$ при интегрировании в симметричных пределах равна нулю.

Для металла, находящегося в состоянии *термодинамического равновесия*, движение электронов описывается четной (симметричной) функцией распределения. Отсюда следует хорошо известный факт, что *тепловое движение не приводит к возникновению тока*. Для его появления необходимо создать условия, при которых нарушается симметрия функции распределения, благодаря чему одни направления скоростей становятся более вероятными, чем другие. Это приводит к тому, что вся совокупность электронов должна перемещаться в пространстве; на хаотическое движение должно быть наложено направленное движение электронов, которое может быть вызвано различными причинами: электрическим полем, градиентом температуры, неоднородным освещением и т. п.

Функция распределения электронов по состояниям в общем случае должна быть найдена из кинетического уравнения Больцмана, о котором будем говорить при описании кинетических явлений в полупроводниках.

Для равновесного состояния системы квантовая статистика приводит к следующему выражению ($f = f_0$):

$$f_0(E, T) = \frac{1}{e^{\frac{E-F}{kT}} + 1}. \quad (3.10)$$



В этом выражении E есть полная энергия, характеризующая состояние электрона, T — абсолютная температура, k — постоянная Больцмана, F — энергия (уровень) Ферми.

Величина $f_0(E, T)$ представляет собой вероятность того, что состояние с энергией E занято электроном. Выражение $f_0(E, T)$ называют функцией Ферми — Дирака. Вероятность нахождения электрона в состоянии с энергией E зависит от величины энергии E и температуры T . Кроме того, вероятность зависит от энергии Ферми, физический смысл которой можно понять из рассмотрения некоторых простейших случаев.

Пусть $T \rightarrow 0$, тогда

$$\lim_{T \rightarrow 0} f_0 = \begin{cases} 1 & \text{при } E < F, \\ 0 & \text{при } E > F. \end{cases} \quad (3.11)$$

При $E = F$ функция не определена, кроме того, она терпит разрыв. На рис. 1 приведен график $f_0(E, T)$. Из рисунка видно, что при $T = 0$ все состояния с $E < F$ заняты, в то время как состояния с $E > F$ совершенно свободны, не заняты электроном. Состояние $E = F$ можно считать занятым с вероятностью 0,5, как будет показано ниже.

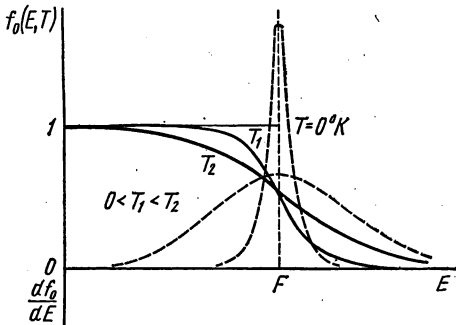


Рис. 1. Функция Ферми — Дирака и ее производная по энергии при различных температурах

В данном случае физический смысл энергии Ферми очень нагляден — это максимальная энергия, которую имеют электроны в металле при температуре абсолютного нуля. Вместе с тем можно сказать, что уровень Ферми отделяет занятые энергетические состояния от свободных.

В общем случае энергия Ферми представляет собой термодинамический потенциал Гиббса, отнесенный к одной частице. Величину F называют также химическим потенциалом. Энергия Ферми численно равна работе, которую необходимо затратить, чтобы увеличить число частиц в системе на единицу.

Рассмотрим теперь случай $T \neq 0$. При любом значении $T \neq 0$ вероятность равна 0,5 для $E = F$, что и позволило доопределить значение $f_0(E = F; 0) = 0,5$. При $E \ll F$ $f_0(E, T) \approx 1$, а при $E \gg F$ функция Ферми — Дирака может быть записана в виде

$$f_0(E, T) \approx e^{\frac{F}{kT}} e^{-\frac{E}{kT}}. \quad (3.12)$$

Вероятность $f_0(E, T)$ испытывает резкое изменение в окрестности $E = F$ в интервале энергии в несколько kT . Действительно, пусть $E = F + \xi kT$, где ξ — переменная величина, представляющая собой энергию в единицах kT , отсчитываемую от уровня Ферми:

$$\xi = \frac{E - F}{kT}. \quad (3.13)$$

В таком случае

$$f_0 = \frac{1}{e^{\xi} + 1}. \quad (3.14)$$

В табл. 1 приведена величина $f_0(\xi)$ для нескольких значений ξ .

Таблица 1

ξ	-5	-4	-3	-2	-1	0	1	2	3	4	5
f_0	0,994	0,980	0,954	0,882	0,730	0,500	0,270	0,118	0,048	0,018	0,007

На рис. 2 приведен график $f_0(\xi)$ - из которого видно, что функция Ферми - Дирака испытывает сильное изменение: от 0,88 до 0,12 - при изменении ξ в интервале от -2 до 2. Другими словами, *существенное изменение функция распределения $f_0(E, T)$ испытывает при изменении энергии от $F - 2kT$ до $F + 2kT$, и $f_0 \approx 1$ при $\xi < -2$, а при $\xi > 2$ она ведет себя подобно экспоненте.*

Интервал энергии, определяющий резкое изменение $f_0(E, T)$, зависит от температуры: при $T \rightarrow 0$ он стремится к нулю. Для характеристики изменения функции распределения удобно использовать ее производную по энергии

$$f'_{0E} = \frac{\partial f_0}{\partial E} = \frac{\partial f_0}{\partial \xi} \cdot \frac{\partial \xi}{\partial E} = -\frac{1}{2kT} \cdot \frac{1}{1 + \text{ch } \xi}, \quad (3.15)$$

где $\text{ch } \xi = \frac{e^{\xi} + e^{-\xi}}{2}$.

При $T \rightarrow 0$ всюду, кроме $E = F$ ($\xi = 0$), $\text{ch } \xi \rightarrow \infty$ и поэтому при $T \neq 0$ f'_{0E} будет отлична от нуля только в малой окрестности точки $\xi = 0$. В пределе $T \rightarrow 0$ эта окрестность стягивается в одну точку $\xi = 0$, так что

$$-\frac{\partial f_0(E, 0)}{\partial E} = \begin{cases} \infty & \text{при } E = F, \\ 0 & \text{при } E \neq F, \end{cases} \quad (3.16)$$

при этом

$$-\int_{-\infty}^{\infty} f'_{0E}(E, 0) dE = 1. \quad (3.17)$$

Выражение (3.17) справедливо при любой температуре $T \geq 0$. Мы видим, что $-f'_{0E}(E, 0)$ представляет собой δ -функцию Дирака.

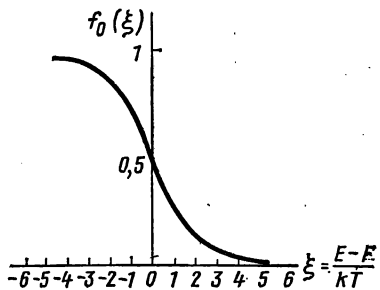


Рис. 2. Функция Ферми - Дирака. Энергия выражена в kT и отсчитывается от уровня Ферми ($\xi = \frac{E-F}{kT}$)

При $T > 0$ функция $f'_{0E}(E, T)$ будет тем ближе к δ -функции, чем меньше температура. На рис. 1 (пунктирная линия) дан график $f'_{0E}(E, T)$ при разных температурах.

Квантовые системы, описываемые функцией Ферми—Дирака, называют вырожденными.

Рассматривая большие значения ξ , видим, что f_0 может быть представлена экспоненциальной функцией

$$f_0(\xi) = \frac{1}{e^{\xi} + 1} \approx e^{-\xi}, \quad (3.18)$$

или

$$f_0(E, T) \approx e^{\frac{F}{kT}} \cdot e^{-\frac{E}{kT}}. \quad (3.19)$$

Функцию $f_0(E, T) = e^{\frac{F}{kT}} \cdot e^{-\frac{E}{kT}}$ называют классической функцией распределения, или функцией Больцмана. *Системы, описываемые функцией Больцмана, т. е. подчиняющиеся классической статистике, называют невырожденными.* Отличие классической функции распределения от квантовой состоит не только в различии функционального выражения, но имеет более глубокий смысл, связанный, в частности, с возможностью в классической физике раздельного определения кинетической и потенциальной энергии, что невозможно в квантовой физике, в которой измеряется лишь полная энергия.

Для вычисления средних значений физических величин с использованием функций распределения $f_0(E, T)$ необходимо знать связь между энергией и скоростью (или импульсом p) частиц. Предположим, что для свободного электронного газа

$$E = \frac{mv^2}{2} = \frac{p^2}{2m} = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2). \quad (3.20)$$

Так как энергия не зависит от направления импульса, то элемент объема в пространстве импульсов $d\tau_p = dp_x dp_y dp_z$ удобнее выразить в сферической системе координат:

$$d\tau_p = p^2 dp d\Omega = p^2 dp \sin \theta d\theta d\varphi. \quad (3.21)$$

Интегрирование по углам θ, φ дает множитель 4π .

Рассмотрим случай статистики Больцмана и вычислим среднее значение энергии электрона:

$$\langle E \rangle = \frac{4\pi V \int_0^{\infty} \frac{p^2}{2m} e^{-\frac{p^2}{2mkT}} p^2 dp}{4\pi V \int_0^{\infty} e^{-\frac{p^2}{2mkT}} p^2 dp}. \quad (3.22)$$

Наличие знаменателя в (3.22) связано с тем, что функция распределения Максвелла, которая использована в (3.22), не нормирована.

Интегрирование по координатам, от которых в данном случае функция распределения не зависит, дает объем тела V . Удобно перейти к новым переменным, положив

$$\frac{p^2}{2mkT} = x; \quad p^2 = (2mkT)x; \quad dp = \frac{1}{2}(2mkT)^{1/2} \frac{dx}{x^{1/2}}. \quad (3.23)$$

В таком случае

$$\langle E \rangle = \frac{(kT) \int_0^{\infty} x^{3/2} e^{-x} dx}{\int_0^{\infty} x^{1/2} e^{-x} dx} = \frac{3}{2} kT. \quad (3.24)$$

При вычислении мы учли, что

$$\int_0^{\infty} x^{3/2} e^{-x} dx = \frac{3}{2} \int_0^{\infty} x^{1/2} e^{-x} dx. \quad (3.25)$$

Это легко доказывается интегрированием по частям, последний же интеграл (3.25) равен $\frac{V\pi}{2}$ (интеграл Пуассона). Кроме того, мы нашли нормировочный коэффициент для функции Больцмана: нормированная к единице при интегрировании по фазовому пространству функция Больцмана имеет вид

$$f_0(E, T) = \frac{1}{V} \left(\frac{1}{2\pi mkT} \right)^{3/2} e^{-\frac{p^2}{2mkT}}. \quad (3.26)$$

Найдем теперь среднюю энергию электронов вырожденной системы при тех же предположениях:

$$f_0(E, T) = \frac{1}{e^{\frac{p^2}{2m} - F} + 1} \quad (3.27)$$

и

$$\langle E \rangle = \frac{4\pi V \int_0^{\infty} \frac{p^2}{2m} \cdot \frac{1}{e^{\left\{ \frac{p^2}{2m} - F \right\} / kT} + 1} p^2 dp}{4\pi V \int_0^{\infty} \frac{1}{e^{\left\{ \frac{p^2}{2m} - F \right\} / kT} + 1} p^2 dp}. \quad (3.28)$$

Переходя к переменным (3.23) и сокращая на одинаковые величины числитель и знаменатель, можем записать

$$\langle E \rangle = kT \frac{\int_0^{\infty} \frac{x^3 dx}{e^{(x - \frac{F}{kT})} + 1}}{\int_0^{\infty} \frac{x^2 dx}{e^{(x - \frac{F}{kT})} + 1}}. \quad (3.29)$$

При большой температуре, когда $kT \gg F$ и $e^{F/kT} \approx 1$, можно пренебречь единицей по сравнению с e^x в подынтегральном выражении уже при малых x . В этом случае квантовая статистика переходит в классическую и

$$\langle E \rangle = \frac{3}{2} kT. \quad (3.30)$$

При очень низких температурах функция Ферми — Дирака имеет вид прямоугольной единичной ступеньки, интеграл можно вычислить от 0 до $x = F/kT$, поэтому

$$\langle E \rangle = \frac{kT \int_0^{F/kT} x^{3/2} dx}{\int_0^{F/kT} x^{1/2} dx} = \frac{3}{5} F. \quad (3.31)$$

Как видно из (3.31), энергия вырожденного электронного газа не зависит от температуры, поэтому он не вносит вклада в теплоемкость тела, что полностью объясняет закон Дюлонга — Пти для твердых тел.

Если провести более строгий расчет, то для $\langle E \rangle$ получим выражение

$$\langle E \rangle = \frac{3}{5} F_0 \left[1 + \frac{5\pi^2}{12} \left(\frac{kT}{F_0} \right)^2 \right]. \quad (3.32)$$

Объяснение независимости теплоемкости твердого тела при обычных температурах от движения электронов явилось крупным успехом квантовой теории, поскольку по классической теории электроны должны давать вклад в удельную теплоемкость металла величиной $\frac{3}{2} kn$.

Однако при низких температурах распределение электронов по энергетическим состояниям определяется не тепловым движением, а волновыми свойствами электронов и вытекающим из них принципом Паули, благодаря чему $\langle E \rangle = \frac{3}{5} F$, что составляет величину

в несколько электронвольт, поэтому тепловое движение слабо влияет на величину энергии электронного газа. Однако при больших температурах, когда $kT > F_0$, энергия электронного газа будет определяться его температурой.

Переход от квантовой статистики к классической можно определить условием

$$kT_b = F_0, \quad (3.33)$$

или

$$T_b = F_0/k. \quad (3.34)$$

Температуру T_b называют температурой вырождения, поскольку при $T > T_b$ электронный газ является классическим, а при $T < T_b$ — квантовым (вырожденным).

При описании свойств полупроводников будут использованы как функция Ферми — Дирака (вырожденные полупроводники), так и функция Больцмана (невырожденные полупроводники).

В дальнейшем будет необходимо вычислять интегралы вида

$$I = \int_0^{\infty} \varphi(E) f_0(E, T) dE. \quad (3.35)$$

Если через $\psi(E)$ обозначим первообразную для функции $\varphi(E)$, то, интегрируя по частям, получим:

$$I = \psi(E) f_0(E, T) \Big|_0^{\infty} - \int_0^{\infty} \psi(E) \frac{\partial f_0(E, T)}{\partial E} dE. \quad (3.36)$$

Первый член в (3.36) обычно обращается в нуль, для этого достаточно, чтобы $\psi(0) = 0$, а $\psi(E)$ возрастала с ростом энергии медленнее, чем экспонента. Считая, что эти условия выполнены, опустим первый член в (3.36). Разложим $\psi(E)$ в ряд Тейлора в окрестности точки $E = F$:

$$I = - \int_0^{\infty} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{d^n \psi(E)}{dE^n} \Big|_F (E - F)^n \frac{\partial f_0(E, T)}{\partial E} dE. \quad (3.37)$$

Переходя к безразмерной переменной $\xi = \frac{E - F}{kT}$ и учитывая соотношение (3.15), запишем:

$$I = - \sum_{n=0}^{\infty} \frac{d^n \psi(E)}{dE^n} \Big|_F (kT)^n \frac{1}{n!} \int_{-F/kT}^{\infty} \xi^n \frac{df_0}{d\xi} d\xi. \quad (3.38)$$

Для сильного вырождения нижний предел можно заменить на $-\infty$. В этом случае члены нечетной степени дают нуль, и (3.38) можно представить в виде:

$$I = \sum_{r=0}^{\infty} \frac{d^{2r} \psi(E)}{dE^{2r}} \Big|_F (kT)^{2r} C_{2r}, \quad (3.39)$$

где через C_{2r} обозначены определенные интегралы вида

$$C_{2r} = \frac{-1}{(2r)!} \int_{-\infty}^{\infty} \xi^{2r} \frac{df_0}{d\xi} d\xi = \frac{1}{(2r)!} \int_{-\infty}^{\infty} \frac{\xi^{2r} e^{-\xi}}{(1 + e^{-\xi})^2} d\xi. \quad (3.40)$$

Вычисление коэффициентов C_{2r} проводится с использованием некоторых специальных функций, например дзета-функции Римана

$$\zeta(2r) = \sum_{l=1}^{\infty} l^{-2r}:$$

$$C_{2r} = 2(1 - 2^{1-2r}) \zeta(2r). \quad (3.41)$$

Используя значения $\zeta(2r)$, можно найти C_{2r} :

$$\zeta(2) = \frac{\pi^2}{6}; \quad \zeta(4) = \frac{\pi^4}{90}; \quad \dots$$

$$C_0 = 1; \quad C_2 = \frac{\pi^2}{6}; \quad C_4 = \frac{7\pi^4}{360}; \quad C_6 = \frac{31\pi^6}{15120}; \quad \dots$$

или $C_0 = 1$; $C_2 \approx 1,64$; $C_4 \approx 1,89$; $C_6 \approx 1,96$ и т. д.

Ограничиваясь несколькими членами ряда, запишем:

$$\int_0^{\infty} \varphi(E) f_0(E, T) dE = \psi(F) + \frac{\pi^2 (kT)^2}{6} \psi''(F) + \frac{7\pi^4 (kT)^4}{360} \psi''''(F) + \dots \quad (3.42)$$

Во многих случаях можно ограничиться двумя членами

$$\int_0^{\infty} \varphi(E) f_0(E, T) dE \approx \psi(F) + \frac{\pi^2}{6} (kT)^2 \psi''(F). \quad (3.43)$$

Выражение (3.43) можно использовать, например, для вычисления (3.29).

Положив $\psi = \frac{2}{5} E^{5/2}$ и $\psi = \frac{2}{3} E^{3/2}$ для верхнего и нижнего интегралов, найдем:

$$\begin{aligned} \langle E \rangle &= \frac{\frac{2}{5} F^{5/2} + \frac{2}{5} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{\pi^2}{6} (kT)^2 F^{1/2} + \dots}{\frac{2}{3} F^{3/2} + \frac{2}{3} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \frac{\pi^2}{6} (kT)^2 F^{-1/2} + \dots} \approx \\ &\approx \frac{3}{5} F \frac{1 + \frac{15}{4} \cdot \frac{\pi^2}{6} \left(\frac{kT}{F}\right)^2}{1 + \frac{3}{4} \cdot \frac{\pi^2}{6} \left(\frac{kT}{F}\right)^2} \approx \frac{3}{5} F \left[1 + \frac{\pi^2}{2} \left(\frac{kT}{F}\right)^2 \right]. \end{aligned} \quad (3.44)$$

Видим, что при $T \rightarrow 0$ (3.44) переходит в (3.31). Аналогично можно найти

$$F = F_0 \left[1 - \frac{\pi^2}{12} \left(\frac{kT}{F_0}\right)^2 \right] \quad (3.45)$$

при условии постоянства концентрации электронов. Подставив (3.45) в (3.44), получим (3.32).

Резюме § 3

1. Для описания электронного газа может быть введена функция распределения $f(\mathbf{r}, \mathbf{p}, t)$, представляющая собой вероятность нахождения электрона в единичном фазовом объеме.

2. Для любой физической величины α можно найти среднее значение $\langle \alpha \rangle$ на основании соотношения (3.8).

3. Квантовая функция распределения электронов по состояниям (3.10) зависит только от энергии и температуры. Она близка к единице при $E < F - 2kT$, близка к нулю при $E > F + 2kT$, испытывает резкое изменение в интервале энергии порядка $\pm 2kT$ относительно энергии Ферми F . Системы, подчиняющиеся квантовой статистике, называют вырожденными. С ростом температуры, когда F становится меньше kT , функция Ферми — Дирака переходит в функцию Больцмана, или классическую функцию распределения (3.12).

4. Средняя энергия вырожденного электронного газа определяется выражением (3.32), невырожденного газа — выражением (3.24):

$$\langle E \rangle = \frac{3}{5} F_0 \left[1 + \frac{5\pi^2}{12} \left(\frac{kT}{F_0} \right)^2 \right]; \quad \langle E \rangle = \frac{3}{2} kT. \quad (3.1p)$$

5. Температура вырождения T_v электронного газа определяется энергией Ферми F_0 :

$$T_v = F_0/kT. \quad (3.2p)$$

При $T > T_v$ газ является классическим (невырожденным), при $T < T_v$ — квантовым (вырожденным).

6. Для вырожденных полупроводников и металлов вычисления с функцией Ферми — Дирака дают:

$$\int_0^{\infty} \varphi(E) f_0(E, T) dE = \varphi(E) f_0(E, T) \Big|_0^{\infty} + \varphi(E) + \frac{\pi^2}{6} (kT)^2 \varphi'(F) + \frac{7\pi^4}{360} (kT)^4 \varphi'''(F) + \dots, \quad (3.3p)$$

где $\varphi(E) = \frac{d\psi(E)}{dE}$.

§ 4. ПОЛУПРОВОДНИКИ, КЛАССИФИКАЦИЯ ВЕЩЕСТВ ПО ВЕЛИЧИНЕ ПРОВОДИМОСТИ

Реальные вещества характеризуются различной величиной удельной электрической проводимости σ . Примеры проводимости некоторых веществ даны в табл. 2:

Из табл. 2 видно, что для таких веществ, как золото, серебро, медь, проводимость имеет величину в несколько единиц на 10^7 Сим·м⁻¹, а эбонит и янтарь имеют проводимость, характеризующуюся величиной порядка 10^{14} Сим·м⁻¹. Вещества с проводимостью $\sigma \approx$

$\approx (10^7 \div 10^6)$ Сим·м⁻¹ принято называть проводниками, или металлами. К изоляторам, или диэлектрикам, относят вещества с проводимостью $\sigma \approx (10^{-8} \div 10^{-16})$ Сим·м⁻¹.

Таблица 2

Вещество	σ , Сим·м ⁻¹	Вещество	σ , Сим·м ⁻¹
Алюминий	$3,12 \cdot 10^7$	Алмаз	10^{-10}
Золото	$4,13 \cdot 10^7$	Эбонит	$5 \cdot 10^{-14}$
Медь тянутая	$5,62 \cdot 10^7$	Пирекс	$1 \cdot 10^{-12}$
Медь отожженная	$6,30 \cdot 10^7$	Слюда	$1,1 \cdot 10^{-11}$
Серебро	$6,03 \cdot 10^7$	Парафинированный воск	$3,3 \cdot 10^{-17}$
Нихром	$9 \cdot 10^5$	Кварц	$5 \cdot 10^{-13}$

Вещества, имеющие промежуточную между металлами и диэлектриками проводимость, были названы полупроводниками. *Полупроводники — это вещества, удельная электрическая проводимость которых лежит в интервале от 10^{-8} до 10^6 Сим·м⁻¹, т. е. может принимать значения, отличающиеся на 14 порядков. Однако это определение полупроводников совершенно не передает специфических особенностей их проводимости. Действительно, рассмотрим, например, температурную зависимость проводимости металлов и полупроводниковых веществ. Для металлов с ростом температуры сопротивление увеличивается*

$$R(t) = R_0(1 + \alpha t), \quad (4.1)$$

где R_0 — сопротивление при $t = 0^\circ\text{C}$, $R(t)$ — сопротивление при $t^\circ\text{C}$, α — термический коэффициент сопротивления, равный примерно $1/273$. Для металлов

$$\alpha = dR/dt = dR/dT > 0.$$

Для полупроводников сопротивление с ростом температуры быстро уменьшается. Эмпирическая формула, дающая связь между сопротивлением и абсолютной температурой T , справедливая для некоторого интервала температур, имеет вид

$$R(T) = R_0 e^{B/T}, \quad (4.2)$$

где R_0 , B — некоторые постоянные для данного интервала температур величины, характерные для каждого полупроводникового вещества. Для удельной проводимости эту формулу можно переписать в виде

$$\sigma = \sigma_0 e^{-B/T}. \quad (4.3)$$

Если умножить числитель и знаменатель показателя экспоненты на постоянную Больцмана k и обозначить $kB = E_a$, то

$$\sigma = \sigma_0 e^{-E_a/kT}. \quad (4.4)$$

Характерная для данного полупроводника величина E_a носит название энергии активации. По своему физическому смыслу она

различна для разных интервалов температур. На рис. 3 приведены графики зависимости сопротивления металла и полупроводника от температуры T . На рис. 4 приведены графики зависимости $\ln \sigma$ от обратной температуры.

Наличие энергии активации E_a означает, что для увеличения проводимости к полупроводниковому веществу необходимо подвести энергию. Опыт показывает, что проводимость полупроводников увеличивается не только при нагревании (т. е. при подведении к полупроводнику тепловой энергии), но и при освещении, при облучении ядерными частицами; она меняется при наложении электрических и магнитных полей, при изменении внешнего давления и т. п. Это означает, что полупроводники — это вещества, проводимость которых сильно зависит от внешних условий: температуры, давления, внешних полей, освещения, облучения ядерными частицами.

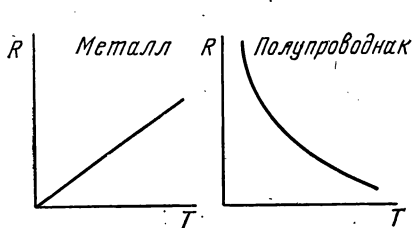


Рис. 3. Зависимость сопротивления металлов и полупроводников от температуры

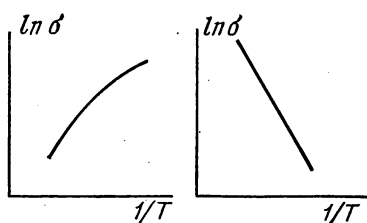


Рис. 4. Зависимость $\ln \sigma$ от обратной температуры для металлов и полупроводников

Так как при $T \rightarrow 0$ и при отсутствии подвода энергии извне проводимость (невырожденных) полупроводников стремится к нулю, то мы можем сказать, что полупроводники — это вещества, обладающие проводимостью только в возбужденном состоянии. При таком определении в принципе нет различия между полупроводниками и диэлектриками, в то время как отличие полупроводников от металлов отражено достаточно четко.

Влияние условий на проводимость полупроводников проявляется по-разному в зависимости от структуры и свойств вещества. При неизменных внешних условиях проводимость одного и того же вещества в виде чистого и совершенного монокристалла, монокристалла с дефектами и примесями и поликристалла различна.

Учитывая все сказанное, можем определить полупроводники следующим образом: полупроводники — это вещества, имеющие при комнатной температуре удельную электрическую проводимость в интервале от 10^{-8} до 10^6 Сим \cdot м $^{-1}$, которая зависит в сильной степени от вида и количества примеси и структуры вещества и от внешних условий: температуры, освещения, электрических и магнитных полей, и т. п.

Это определение позволяет отличать полупроводники от металлов — проводимость металлов значительно слабее зависит от внеш-

них условий, поскольку для создания проводимости в полупроводниках их необходимо перевести в возбужденное (активированное) состояние, в то время как состояние проводимости металлов является состоянием невозбужденным. Указание на интервал возможных значений удельной проводимости позволяет отличать полупроводники от диэлектриков. Отсюда следует, что различие между полупроводниками и диэлектриками чисто количественное и в значительной мере условное.

Существует два типа полупроводниковых веществ: ионные и электронные. В ионных полупроводниках ток переносится ионами вещества, вследствие чего состав и структура ионного полупроводника меняется при прохождении через него электрического тока. Для приборов, преобразующих энергию, такие вещества не при-

годны, так как они будут разрушаться при прохождении через них тока. Такие вещества в данном курсе не рассматриваются.

В электронных полупроводниках ток переносится электронами, вследствие чего переноса вещества при прохождении тока не происходит, и приборы, изготовленные из этого материала, могут работать длительное время.

К полупроводникам (электронным) относится огромное количество самых различных

Рис. 5. Положения простых полупроводниковых веществ в таблице Менделеева

Группы \ Периоды	II	III	IV	V	VI	VII	
II	Be	B	C	N	O		
III		Al	Si	P	S	Cl	
IV		Ca	Ge	As	Se	Br	
V		In	Sn	Sb	Te	I	Xe
VI			Pb	Bi	Po	At	

веществ. Но практическое значение в настоящее время имеют лишь некоторые из них. Однако по мере развития химии и технологии получения чистых веществ число полупроводников, находящихся широкого практического применения, быстро возрастает.

К числу полупроводников относятся 12 простых веществ: бор B, углерод C, кремний Si, фосфор P, сера S, германий Ge, мышьяк As, селен Se, серое олово Sn, сурьма Sb, теллур Te, иод I. На рис. 5 указано расположение полупроводников среди остальных элементов периодической таблицы Менделеева. Наибольшее значение из атомарных полупроводников в настоящее время имеют германий и кремний.

Полупроводниковыми свойствами обладают многие бинарные соединения типа $A^x B^{8-x}$, где A—элемент группы X и B—элемент группы (8—X).

К числу соединений типа $A^I B^{VI}$ относятся: AgCl, CuBr, KBr, LiF и др. Как полупроводниковые вещества они еще не нашли широкого применения.

К соединениям типа $A^{II} B^{VI}$ относятся сульфиды, теллуриды, селениды, окислы металлов второй группы. Наиболее известными

среди них являются соединения CdS, CdSe, CdTe, ZnS, ZnO, ZnSe, HgTe, HgSe и некоторые другие. В ближайшем будущем многие из них найдут широкое практическое применение.

Одними из важнейших в настоящее время полупроводниковых соединений являются вещества типа $A^{III}B^V$. К ним принадлежат антимоныды, арсениды, фосфиды, нитриды элементов второй подгруппы третьей группы: алюминия, галлия, индия, бора.

В число «соединений» типа $A^{IV}B^{IV}$ входят SiC, SiGe.

Помимо соединений вида $A^x B^{8-x}$, к полупроводниковым веществам относятся такие соединения, как $A^{IV}B^{VI}$ (PbS, PbSe, PbTe), $A^{IV}B^{VI}$ (CuS, CuO, Cu₂O) и др.

Большой интерес представляют более сложные соединения и твердые растворы, как вещества вида $A^x B_1^{8-x} B_2^{8-x}$, $A_x A_y B^{8-x}$, $A_x A_y B_1^{8-x} B_2^{8-x}$. Примерами твердых растворов могут служить GaAsP, InGaSb или ZnCdSeTe.

Комбинируя различные элементы, можно будет получать соединения с наиболее подходящими для данных практических целей свойствами. Помимо этого, полупроводниками являются различные более сложные соединения.

Кроме неорганических веществ, к полупроводникам причисляются и некоторые органические вещества, такие, как антрацен метиленовый голубой, фталоцианин, коронен и др.

Разнообразие полупроводниковых веществ с большим набором свойств обусловило их широкое практическое применение для изготовления различных приборов.

Полупроводниковые диоды позволяют выпрямить ток от миллиампер до тысяч ампер, от низких частот до сверхвысоких частот, от напряжения в доли вольта до сотен вольт. Триоды, или транзисторы, используют для усиления и генерации колебаний в широком диапазоне частот. Регистрация светового и корпускулярного излучения, преобразование энергии излучения и тепловой в электрическую достигается полупроводниковыми приемниками и преобразователями с большим коэффициентом полезного действия.

В измерительной технике широко используют всевозможные «датчики», позволяющие преобразовывать различные воздействия в электрический сигнал для измерения давлений, магнитных полей, температур, энергии излучения.

Полупроводники находят широкое применение для получения когерентного излучения с большим к.п.д. преобразования электрической энергии в световую. Трудно перечислить все те области, где полупроводниковые приборы уже нашли применение.

В основе каждого прибора лежат определенные физические процессы и явления, без знания которых невозможно правильно использовать существующие и разрабатывать новые приборы.

§ 5. МОДЕЛЬНЫЕ ПРЕДСТАВЛЕНИЯ О ПРОВОДИМОСТИ ПОЛУПРОВОДНИКОВ. ПОНЯТИЕ ДЫРКИ

Кристаллическая решетка какого-либо вещества возникает в результате взаимодействия атомов. Характер взаимодействия определяется строением электронных оболочек атомов, образующих кристалл. Основную роль при этом играет так называемый обменный эффект, в результате которого какие-либо два атома могут передавать электроны друг другу, что приводит к возникновению сил притяжения между атомами. Эта связь носит название гомополярной, поскольку она чаще всего наблюдается между одинаковыми атомами. Так как в ее образовании участвуют в первую очередь валентные электроны, то она называется и ковалентной. Наибольшей силы связь возникает в том случае, когда атомы обмениваются парой электронов с противоположно направленными спинами. Отсюда очевидно, что ковалентная связь должна обладать свойством насыщения — добавление третьего электрона не может увеличить энергию связи, так как при этом всегда два электрона будут иметь параллельные спины. Это свойство ковалентной, или парноэлектронной, связи вытекает из принципа Паули, который в данном случае гласит, что *два электрона с одинаковыми проекциями спина не могут находиться в одной и той же области пространства между атомами*. Примерами веществ с ковалентной связью являются алмаз, кремний, германий.

Если во взаимодействие вступают два различных атома, то максимальная плотность электронного облака (плотность вероятности нахождения электронов) может сместиться ближе к одному из взаимодействующих атомов, и именно к тому атому, который имеет большее число электронов в валентной оболочке. В предельном случае максимальная плотность электронного облака находится у одного из взаимодействующих атомов, который тем самым превращается в отрицательно заряженный ион, в то время как другой атом превращается в положительный ион. В этом случае связь носит название ионной, поскольку ее можно рассматривать как результат кулоновского притяжения разноименных ионов. Наиболее отчетливо ионная связь проявляется в щелочно-галогидных кристаллах (NaCl , LiF и др.). Ионная связь есть предельный случай ковалентной связи. О таких веществах, как окислы или сульфиды элементов второй группы, говорят, что связь на столько-то процентов ионная. Другим предельным видом обменного взаимодействия является обмен электронами любой пары атомов кристалла. Электроны принадлежат решетке в целом, они коллективизируются. Эта связь носит название металлической, так как она наиболее характерна для металлов. Однако нужно помнить, что резкой грани между тремя видами связи провести нельзя. В органических кристаллах наблюдается сравнительно слабая молекулярная связь, обусловленная так называемым взаимодействием Ван-дер-Ваальса, в основе которого лежит взаимодействие между индуцированными дипольными моментами молекул.

Рассмотрим наглядную картину проводимости полупроводниковых веществ на примере кремния.

Атом кремния имеет 14 электронов, которые распределены по оболочкам следующим образом: $(1s^2) (2s^2) (2p^6) (3s^2) 3p^2$.

Незаполненная внешняя оболочка содержит четыре электрона. Электронное облако атома, вступившего во взаимодействие при образовании кристалла, имеет тетраэдрическую структуру. Подобная структура электронного облака обуславливает так называемый алмазный тип кристаллической решетки кремния. Решетка алмазного типа (рис. 6) является кубической. Каждый атом связан с четырьмя ближайшими атомами кремния парноэлектронной связью, в осуществлении которой принимают участие четыре его валентных электрона. В идеальной решетке все электроны связаны, свободных носителей заряда нет, и поэтому при наложении электрического поля электрический ток возникнуть не может. Для его возникновения необходимо часть электронов сделать свободными, не связанными. Но для отрыва электрона необходимо затратить энергию. Ее можно подвести к решетке в виде энергии фотонов или корпускулярного излучения, или в виде энергии тепловых колебаний решетки. Для отрыва электрона в кремнии при комнатной температуре необходимо затратить энергию 1,08 эВ. Обратим внимание на то, что освобождение одного электрона порождает одну незавершенную связь.

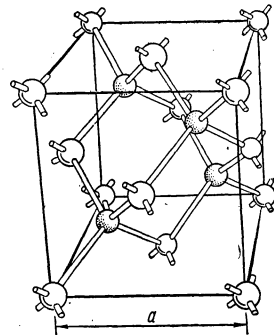


Рис. 6. Решетка типа алмаза

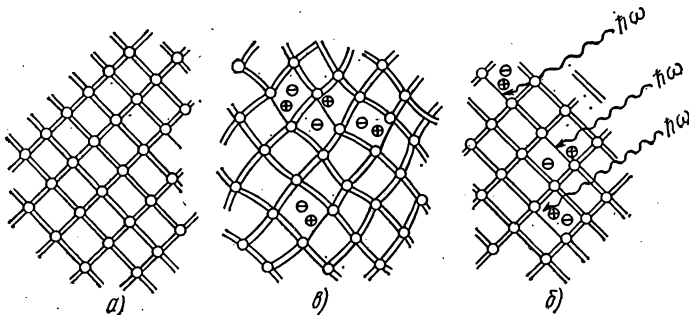


Рис. 7. Возникновение проводимости в результате тепловых колебаний решетки и облучения полупроводника

На рис.7 представлены диаграммы идеального кристалла (а), возникновения свободного электрона за счет энергии тепловых колебаний решетки (б) и в результате поглощения фотона подходящей энергии (в). Освобожденный электрон будет хаотически перемещаться по кристаллу. Если электрон подойдет вновь к тому атому, от которого он был оторван, то он может снова соединиться с ато-

мом, отдав свою энергию решетке или испустив квант света. Процесс превращения свободного электрона в связанный электрон носит название рекомбинации. Он является обратным процессу освобождения связанного электрона (генерации). Отметим, что рекомбинация приводит к исчезновению не только свободного электрона, но и одной незавершенной связи между атомами. Число свободных электронов и вакантных, незавершенных связей одинаково. При этом в области незавершенной связи имеется избыточный положительный заряд, не компенсируемый теперь зарядом электрона. Однако для достаточно большого объема вещества полный заряд остается по-прежнему равным нулю, так что появление в равных количествах свободных электронов и вакантных связей не нарушает электронейтральности кристалла в целом.

При наложении на кристалл электрического поля E свободные электроны, участвуя в хаотическом тепловом движении, будут испытывать действие силы $e_n E$ и придут в дрейфовое движение против поля. Если обозначить концентрацию электронов через n , их подвижность через μ_n , то плотность электронного тока будет равна

$$j_n = e_n \mu_n n E = \sigma_n E, \quad (5.1)$$

где через e_n обозначен заряд электрона.

В полупроводниках проводимость действительно зависит от внешних условий, поскольку, меняя интенсивность освещения, облучение или температуру, можно менять концентрацию носителей заряда в широких пределах, в то время как в металлах число электронов остается неизменным при изменении внешних условий и температуры. Однако это не единственное различие между металлами и полупроводниками. В последних существует два механизма проводимости.

Действительно, незавершенная связь вследствие движения электронов может перемещаться от атома к атому, т. е. может совершать хаотическое движение по кристаллу. При наложении внешнего электрического поля E на связанные электроны будет действовать сила $e_n E$, поэтому они, перемещаясь против поля, будут занимать вакантную связь. В идеальном кристалле, когда все связи заполнены, движение связанных электронов в силу принципа Паули было бы невозможным. Наличие вакансий в связях позволяет валентным электронам перемещаться против поля. Тем самым совокупность валентных электронов также участвует в образовании проводимости полупроводников. При этом подвижность связанных электронов должна зависеть от числа вакансий. Если обозначить число связанных электронов через N , их подвижность через μ_N , то

$$j_N = e_n \mu_N N E. \quad (5.2)$$

Таким образом, в полупроводниках существуют два вида носителей заряда — свободные электроны и связанные электроны, поэтому

$$j = j_n + j_N = (e_n \mu_n n + e_n \mu_N N) E. \quad (5.3)$$

Однако удобнее рассматривать не движение совокупности валентных (т. е. связанных) электронов, а движение вакантных связей. Движение электрона по связям против поля приводит к перемещению вакансии по полю (рис.8). Но перемещение вакантной связи по полю равносильно перемещению по полю положительного заряда e^+ . Однако нужно помнить, что положительный заряд перемещается по полю не в результате действия на него силы e^+E , а в результате перемещения электронов против поля. Непосредственное перемещение положительного заряда e^+ по полю означало бы перемещение ионизированного атома кремния. Однако подвижность ионов на много порядков меньше подвижности электронов, поэтому ионный ток в кремнии практически не возникает. Именно поэтому перемещение нескомпенсированного заряда ядер по полю есть только результат движения валентных электронов. Обозначив число вакантных связей через p , а их подвижности через μ_p , можно выразить ток совокупности связанных электронов следующим образом:

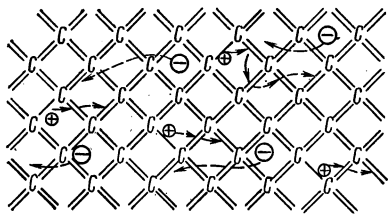


Рис. 8. Движение электронов и вакансий в электрическом поле

$$\mathbf{j}_N = e_N \mu_N N E = e_p \mu_p p E = \mathbf{j}_p. \quad (5.4)$$

При этом μ_p не должно зависеть от p , что будет показано в дальнейшем с помощью метода зон Бриллюэна.

Вакантная связь получила название дырки в связях, или просто дырки, и механизм проводимости посредством связанных электронов получил название дырочной проводимости. Величину p называют при этом концентрацией дырок, μ_p — подвижностью дырок, а $e_p = e^+$ — зарядом дырки. При этом рассматривают дырки как некие квазичастицы, движение которых вполне адекватно движению валентных электронов. Для того чтобы прийти к непротиворечивым результатам при описании физических явлений с помощью понятия дырок, необходимо правильно определить все свойства дырок. Только в этом случае можно говорить о движении дырок без рассмотрения движения совокупности валентных электронов (см. § 25).

§ 6. СОБСТВЕННАЯ И ПРИМЕСНАЯ ПРОВОДИМОСТЬ

Полупроводник, в котором число электронов равно числу дырок $n = p$, называется собственным полупроводником, для него

$$\mathbf{j} = \sigma E = (e_n \mu_n n + e_p \mu_p p) E = (\sigma_n + \sigma_p) E. \quad (6.1)$$

Если обозначить через b отношение модулей подвижностей (считая подвижность скалярной величиной):

$$b = |\mu_n| / |\mu_p|, \quad (6.2)$$

то для проводимости собственного полупроводника можно записать

$$\sigma = e_n \mu_n n + e_p \mu_p p = e_p \mu_p (1 + b). \quad (6.3)$$

Однако в большинстве случаев число дырок и электронов в полупроводниках различно, именно на этом основана работа большинства полупроводниковых приборов. Различие в концентрациях дырок и электронов достигается введением примесей. *Проводимость, созданная введением примеси, называется примесной.* Для понимания существа примесной проводимости рассмотрим кремний с примесью элементов пятой и третьей групп.

Пусть в кристалле кремния атом мышьяка замещает атом кремния. Наружная оболочка атома мышьяка содержит пять электронов. Из них четыре будут участвовать в образовании ковалентной связи с четырьмя ближайшими атомами кремния. Пятый электрон не может принять участие в образовании связи, поскольку все связи заполнены. В то же время он будет испытывать воздействие со стороны окружающих атомов кремния, в силу чего энергия его связи с атомом мышьяка уменьшается, как показывает расчет, примерно в ϵ^2 раз (ϵ — относительная диэлектрическая проницаемость, для кремния $\epsilon = 12$). Это значит, что пятый электрон атома мышьяка может стать свободным при затрате энергии в десятки раз меньшей, чем энергия, необходимая для отрыва электрона от атома основного вещества, в данном случае кремния. Поэтому примесь может быть легко ионизирована, и в кристалле тем самым появится большое число свободных электронов, намного превышающее число свободных электронов в кремнии, лишенном примеси. Видно, что элементы пятой группы, введенные в кремний, приводят к появлению в нем значительно большего числа свободных электронов, чем в чистом кремнии. *Примесь, которая отдает электроны, носит название донорной.* Атомы примеси, отдав электрон, превращаются в положительно заряженные ионы (As^+). Но положительно заряженный ион примеси в создании тока принимать участие не может, поэтому донорная примесь является поставщиком только свободных электронов.

Наряду с ионизацией атомов примеси будут наблюдаться процессы ионизации и атомов основного вещества. Однако количество электронов, поставляемых основным веществом, будет намного меньше числа электронов, поставляемых примесью. Благодаря этому число дырок будет намного меньше общего числа электронов. *При протекании тока заряд в таком кристалле будет переноситься в основном электронами, которые в силу этого называются основными носителями заряда, а дырки — неосновными носителями заряда. Такой полупроводник носит название электронного, или n-типа.* Проводимость электронного полупроводника может быть записана в виде

$$\sigma = \sigma_n + \sigma_p \approx \sigma_n = e_n \mu_n n, \quad (6.4)$$

так как $p \ll n$ и $\sigma_p \ll \sigma_n$.

Характер проводимости при введении примеси можно в ряде случаев определить на основании простого правила валентности: если валентность примеси больше валентности основного вещества, то примесь является донорной. Так, например, элементы V группы в кремнии и германии являются донорами. В соединениях $A^{III}B^V$ донорами являются элементы VI группы. Однако это правило применимо не всегда.

Рассмотрим теперь другой случай. Пусть в кремний введен атом индия или любого другого элемента III группы. Атом индия имеет три валентных электрона, поэтому одна связь его с атомами кремния будет незаполнена. Для заполнения этой связи необходимо перевести к индию один из электронов какого-либо атома кремния. Энергия, необходимая для перевода электрона от атома основного вещества к атому примеси, который превращается в отрицательно заряженный ион, намного меньше энергии, необходимой для освобождения электрона, поэтому процессы перехода электронов от атомов основного вещества к атомам примеси будут наблюдаться значительно чаще процессов ионизации атомов основного вещества с образованием свободных электронов. Но отрицательный заряд локализуется на атоме примеси и поэтому участия в создании тока принимать не будет. Незавершенная же связь (дырка в связях), находящаяся теперь между двумя атомами основного вещества, является носителем положительного заряда — дыркой.

Примесь, принимающая электрон, называется акцепторной. Число дырок в этом случае может намного превосходить число свободных электронов, поэтому проводимость полупроводника будет в основном дырочной:

$$\sigma = \sigma_n + \sigma_p \approx \sigma_p = e_p \mu_p p, \quad (6.5)$$

так как $n \ll p$ и $\sigma_n \ll \sigma_p$.

Дырки называются основными носителями заряда, а электроны — неосновными. Полупроводник с акцепторной примесью носит название дырочного, или p-типа. Акцепторами часто являются вещества, валентность которых меньше валентности основного вещества; например, элементы III группы в Si, Ge; элементы II группы в соединениях $A^{III}B^V$.

Так как концентрация носителей заряда в примесных полупроводниках превосходит концентрацию носителей заряда в собственном полупроводнике, то сопротивление полупроводника с примесью одного типа меньше сопротивления чистого вещества. Другими словами, введение примеси (легирование) в полупроводниковое вещество понижает его сопротивление.

Если в веществе содержится примесь двух типов — и акцепторы, и доноры, — то происходит взаимная компенсация примеси. При равенстве концентраций доноров и акцепторов легированный полупроводник подобен собственному. Такие полупроводники будут обладать высоким сопротивлением. Они называются скомпенсированными.

Полупроводники, тип проводимости которых зависит от вида примеси, называются амфотерными.

Резюме § 4—6

1. К полупроводникам относят вещества, проводимость которых в сильной степени зависит от состава, структуры кристалла и внешних условий. Проводимость полупроводников, как правило, возрастает при сообщении им энергии нагревом, освещением, облучением ядерными частицами, она зависит от давления, внешних электрических и магнитных полей.

2. В полупроводниках существует два механизма проводимости: носителями заряда являются свободные электроны и свободные дырки. Дырочная проводимость есть проводимость, создаваемая движением связанных электронов по связям.

3. В чистом полупроводнике число дырок равно числу электронов, такой полупроводник называют собственным. Примесь, поставляющую свободные электроны, называют донорной; примесь, поставляющую свободные дырки, называют акцепторной. Носители заряда, имеющиеся в большем количестве, называют основными; носители заряда, имеющиеся в меньшем количестве, называют неосновными.

Полупроводниковое вещество, в котором концентрации акцепторной и донорной примеси равны, называют скомпенсированным полупроводником.

ОСНОВЫ ЗОННОЙ ТЕОРИИ ПОЛУПРОВОДНИКОВ

§ 7. УРАВНЕНИЕ ШРЕДИНГЕРА ДЛЯ КРИСТАЛЛА

Кристалл представляет собой единую систему легких (электроны) и тяжелых (ядра) частиц. Обозначим координаты электронов через $\mathbf{r}_1, \mathbf{r}_2, \dots$, а ядер — через $\mathbf{R}_1, \mathbf{R}_2, \dots$. Стационарное состояние частиц описывается уравнением Шредингера

$$\hat{H}\Psi = E\Psi, \quad (7.1)$$

где \hat{H} — гамильтониан кристалла, Ψ — его собственная волновая функция, E — его собственное значение, или энергия кристалла. Волновая функция кристалла зависит от координат всех частиц:

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots; \mathbf{R}_1, \mathbf{R}_2, \dots) = \Psi(\mathbf{r}_i; R_\alpha), \quad (7.2)$$

где для краткости через \mathbf{r}_i обозначены координаты электронов, а через \mathbf{R}_α — координаты ядер.

Оператор Гамильтона включает в себя все виды энергии, а именно:

1) кинетическую энергию электронов \hat{T}_e :

$$\hat{T}_e = \sum_i \hat{T}_i = \sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right), \quad (7.3)$$

где m — масса электрона, \hbar — постоянная Планка h , деленная на 2π : $\hbar = \frac{h}{2\pi}$, $\Delta_i = \nabla_i^2$ — оператор Лапласа для i -го электрона:

$$\Delta_i = \nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}; \quad (7.4)$$

2) кинетическую энергию ядер \hat{T}_Z :

$$\hat{T}_Z = \sum_\alpha \hat{T}_\alpha = \sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \Delta_\alpha \right), \quad (7.5)$$

где M_α — масса ядра;

$$\Delta_\alpha = \frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}; \quad (7.6)$$

3) энергию попарного взаимодействия электронов \hat{U}_e :

$$\hat{U}_e = \frac{1}{2} \sum_{i,j \neq i} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} = \frac{1}{2} \sum_{i \neq j} \hat{U}_{ij}; \quad (7.7)$$

4) энергию попарного взаимодействия ядер \hat{U}_Z :

$$\hat{U}_Z = \frac{1}{2} \sum_{\alpha \neq \beta} \frac{Z_\alpha Z_\beta e^2}{4\pi\epsilon_0 |R_\alpha - R_\beta|} = \frac{1}{2} \sum_{\alpha \neq \beta} \hat{U}_{\alpha\beta}, \quad (7.8)$$

где $Z_\alpha e$, $Z_\beta e$ — заряды ядер α и β ;

5) энергию взаимодействия электронов с ядрами \hat{U}_{eZ} :

$$\hat{U}_{eZ} = - \sum_{i, \alpha} \frac{Z_\alpha e^2}{4\pi\epsilon_0 |r_i - R_\alpha|} = \sum_{i, \alpha} \hat{U}_{i\alpha}; \quad (7.9)$$

6) энергию всех частиц во внешнем поле \hat{V} :

$$\hat{V} = \hat{V}(r_1, r_2, \dots; R_1, R_2, \dots). \quad (7.10)$$

В таком случае гамильтониан кристалла во внешнем поле может быть записан в следующем виде:

$$\hat{H} = \hat{T}_e + \hat{T}_Z + \hat{U}_e + \hat{U}_Z + \hat{U}_{eZ} + \hat{V}; \quad (7.11)$$

$$\hat{H}\Psi = E\Psi.$$

Уравнение Шредингера (7.11) содержит $3(Z+1)N$ переменных, если через N обозначить число атомов в кристалле. Так как в 1 см^3 содержится примерно $5 \cdot 10^{22}$ атомов, то при $Z=14$ число переменных составляет $2 \cdot 10^{24} \text{ см}^{-3}$. Ясно, что такое уравнение не может быть решено в общем виде. Это связано не только с техническими вычислительными трудностями, но и с трудностями принципиальными, поскольку современная квантовая механика по существу не имеет аппарата при решении задач для системы с большим числом частиц. Для того чтобы решить уравнение Шредингера системы взаимодействующих частиц, необходимо свести такую систему к системе невзаимодействующих частиц. В последнем случае уравнение для системы частиц распадается на систему уравнений, каждое из которых описывает движение одной частицы. Действительно, если гамильтониан системы может быть представлен в виде суммы гамильтонианов:

$$\hat{H} = \sum_k \hat{H}_k, \quad (7.12)$$

где \hat{H}_k зависит только от координат k -ой частицы:

$$\hat{H}_k = - \frac{\hbar^2}{2m_k} \Delta_k + \hat{U}_k(r_k), \quad (7.13)$$

т. е. частицы не взаимодействуют, то уравнение Шредингера для системы решается следующим образом. Волновая функция системы представляется в виде произведения волновых функций отдельных частиц, а энергия системы равна сумме энергий частиц:

$$\Psi(r_1, r_2, \dots) = \psi_1(r_1) \psi_2(r_2) \dots \quad (7.14)$$

и

$$E = \sum_k E_k, \quad (7.15)$$

где E_k и ψ_k связаны между собой соотношением

$$\hat{H}_k \psi_k(\mathbf{r}_k) = E_k \psi_k(\mathbf{r}_k), \quad (7.16)$$

что легко доказывается простой подстановкой (7.14) в уравнение Шредингера для системы частиц, откуда с учетом (7.16) следует (7.15).

Переход от системы (7.11) взаимодействующих частиц к системе невзаимодействующих частиц возможен только в результате приближенного решения уравнения при более или менее очевидных упрощениях.

Ниже будем предполагать, что внешние поля отсутствуют:

$$\hat{V}(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots) = 0. \quad (7.17)$$

Прежде чем переходить к упрощениям уравнения Шредингера, запишем выражение для энергии кристалла

$$E = \int \Psi^*(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots) \hat{H} \Psi(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots) d\tau, \quad (7.18)$$

где интегрирование ведется по координатам всех частиц:

$$d\tau = dx_1 dy_1 dz_1 \dots dX_1 dY_1 dZ_1 \dots = d\tau_e d\tau_z. \quad (7.19)$$

Волновая функция $\Psi(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots)$ позволяет найти движение любой частицы кристалла. С ее помощью, рассматривая состояния с минимальной энергией, теоретически можно было бы найти кристаллическую структуру вещества и ее возможные модификации.

§ 8. АДИАБАТИЧЕСКОЕ ПРИБЛИЖЕНИЕ

Адиабатическое приближение, или приближение Борна — Оппенгеймера, учитывает различный характер движения легких (электронов) и тяжелых (ядер) частиц. Для быстро движущихся электронов важно мгновенное положение ядер, в то время как на движение ядер должно оказывать влияние не мгновенное положение электронов, а только их усредненное движение. Это утверждение легко понять применительно к многоэлектронным атомам. Ясно, движение ядра в силу его инерционности не следует за движением каждого электрона, а движется в усредненном поле всех электронов. В то же время сравнительно медленное движение ядра увлекает за собой электроны, вследствие чего целостность атома сохраняется. Подобное же положение должно иметь место и в кристалле.

Наиболее грубое предположение должно состоять в том, что ядра покоятся: $\mathbf{R}_\alpha = \mathbf{R}_\alpha^0$. В этом случае уравнение Шредингера существенно упрощается. Действительно, кинетическая энергия ядер обращается в нуль, энергия взаимодействия ядер U_Z представляет собой константу, которую выбором начала отсчета энергии можно обратить в нуль. *Учитывая, что $\hat{T}_Z = 0$ и $\hat{U}_Z = 0$, запишем выраже-*

ние для гамильтониана, который обозначим теперь через \hat{H}_e и назовем его гамильтонианом электронов:

$$\hat{H}_e = \hat{T}_e + \hat{U}_e + \hat{U}_{eZ}. \quad (8.1)$$

Волновую функцию электронов обозначим через Ψ_e . Она должна зависеть от координат электронов r_i и координат покоящихся ядер R_α^0 , причем $\Psi_e(r_i; R_\alpha^0)$ должна быть нормирована к единице при любых значениях координат ядер при интегрировании по координатам электронов:

$$\int \Psi_e^*(r_1, \dots; R_1^0, \dots) \Psi_e(r_1, \dots; R_1^0, \dots) d\tau_e. \quad (8.2)$$

Уравнение Шредингера может быть записано в следующем виде:

$$\begin{aligned} \hat{H}_e \Psi_e = E_e \Psi_e; \\ \left[\sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) + \frac{1}{8\pi\epsilon_0} \sum_{i \neq j} \frac{e^2}{|r_i - r_j|} - \frac{1}{4\pi\epsilon_0} \sum_{i,\alpha} \frac{Ze^2}{|r_i - R_\alpha^0|} \right] \Psi_e = E_e \Psi_e. \end{aligned} \quad (8.3)$$

В это уравнение R_α^0 уже входит не в качестве переменной дифференциального уравнения, а в качестве параметра, выбор которого влияет в конечном счете на волновую функцию и значение энергии кристалла E_e :

$$E_e = \int \Psi_e^* \hat{H}_e \Psi_e d\tau_e = E_e(R_1^0, R_2^0, \dots). \quad (8.4)$$

E_e представляет собой энергию электронов, движущихся в поле покоящихся ядер.

Однако предположение о покоящихся ядрах является излишне грубым. Можно считать, что ядра движутся, и учесть это движение введением волновой функции ядер $\Phi_Z(R_1, \dots)$. С этой целью представим гамильтониан кристалла в следующем виде:

обозначим через \hat{H}_Z оператор

$$\hat{H}_Z = \sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \Delta_\alpha \right) + \hat{U}_Z + \hat{E}_e(\dots, R_\alpha, \dots), \quad (8.5)$$

который будем называть ядерной частью гамильтониана кристалла. Очевидно, что гамильтониан кристалла можно выразить через \hat{H}_e и \hat{H}_Z . Учитывая (7.11), (7.17), (8.1) и (8.5), запишем

$$\hat{H} = \hat{H}_e + \hat{H}_Z - \hat{E}_e. \quad (8.6)$$

Волновую функцию кристалла Ψ представим в виде произведения

$$\Psi(\dots, r_i, \dots; \dots, R_\alpha, \dots) = \Psi_e(\dots, r_i, \dots; \dots, R_\alpha, \dots) \Phi_Z(\dots, R_\alpha, \dots), \quad (8.7)$$

подставим его в уравнение Шредингера для кристалла и найдем уравнение для $\Phi_Z(\mathbf{R}_1, \dots)$:

$$\hat{H}\Psi = (\hat{H}_e + \hat{H}_Z - \hat{E}_e) \Psi_e \Phi_Z = \Phi_Z \hat{H}_e \Psi_e + \mathbf{H}_Z \Psi_e \Phi_Z - \hat{E}_e \Psi_e \Phi_Z = E \Psi_e \Phi_Z, \quad (8.8)$$

или, учитывая (8.3), запишем

$$\hat{H}\Psi = \hat{H}_Z \Psi_e \Phi_Z = E \Psi_e \Phi_Z = E \Psi. \quad (8.9)$$

Так как Ψ_e зависит от координат ядер, то \hat{H}_Z будет действовать на Ψ_e ; найдем, например, $\Delta_\alpha \Psi_e \Phi_Z$:

$$\begin{aligned} \Delta_\alpha \Phi_Z \Psi_e &= \nabla_\alpha (\Psi_e \nabla_\alpha \Phi_Z + \Phi_Z \nabla_\alpha \Psi_e) = \\ &= \Psi_e \Delta_\alpha \Phi_Z + \Phi_Z \Delta_\alpha \Psi_e + 2(\nabla_\alpha \Phi_Z \cdot \nabla_\alpha \Psi_e). \end{aligned}$$

Учитывая эту формулу, перепишем (8.9):

$$\begin{aligned} \hat{H}_Z \Psi_e \Phi_Z &= \sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \right) [\Psi_e \Delta_\alpha \Phi_Z + \Phi_Z \Delta_\alpha \Psi_e + 2(\nabla_\alpha \Psi_e \cdot \nabla_\alpha \Phi_Z)] + \\ &+ U_Z \Psi_e \Phi_Z + E_e \Psi_e \Phi_Z = E \Psi_e \Phi_Z. \end{aligned} \quad (8.10)$$

Умножим уравнение (8.10) слева на Ψ^* и проинтегрируем по координатам электронов; учитывая, что

$$\int \Psi_e^* \Psi_e d\tau_e = 1, \quad (8.11)$$

получим

$$\begin{aligned} \sum_\alpha \left(-\frac{\hbar^2}{2M_\alpha} \right) \left[\Delta_\alpha \Phi_Z + \Phi_Z \int \Psi_e^* \Delta_\alpha \Psi_e d\tau_e + 2 \left(\nabla_\alpha \Phi_Z \int \Psi_e^* \nabla_\alpha \Psi_e d\tau_e \right) \right] + \\ + U_Z \Phi_Z + E_e \Phi_Z = E \Phi_Z. \end{aligned} \quad (8.12)$$

Если учесть обозначение (8.5) для \hat{H}_Z , то (8.12) можно переписать следующим образом:

$$\hat{H}_Z \Phi_Z = E \Phi_Z + \sum_\alpha \frac{\hbar^2}{2M_\alpha} \left[\Phi_Z \int \Psi_e^* \Delta_\alpha \Psi_e d\tau_e + 2 \left(\nabla_\alpha \Phi_Z \int \Psi_e^* \nabla_\alpha \Psi_e d\tau_e \right) \right]. \quad (8.13)$$

Умножив уравнение (8.13) на Φ_Z^* слева и интегрируя по координатам ядер, получим выражение для энергии кристалла:

$$\int \Phi_Z^* \hat{H}_Z \Phi_Z d\tau_Z = E + \delta E. \quad (8.14)$$

Чтобы оценить величину δE :

$$\begin{aligned} \delta E &= \sum_\alpha \frac{\hbar^2}{2M_\alpha} \int \Psi_e^* \Delta_\alpha \Psi_e d\tau_e + \\ &+ \sum_\alpha \frac{\hbar^2}{2M_\alpha} \left(2 \int \Phi_Z^* \nabla_\alpha \Phi_Z d\tau_Z \int \Psi_e^* \nabla_\alpha \Psi_e d\tau_e \right), \end{aligned} \quad (8.15)$$

необходимо сделать некоторые предположения о характере зависимости Ψ_e от координат \mathbf{r}_i и \mathbf{R}_α . Если пренебречь в \hat{H}_e взаимодействием электронов, то \hat{H}_e превращается в гамильтониан системы не взаимодействующих частиц, и волновая функция такой системы будет представлять собой комбинацию атомных волновых функций, которые зависят от разности $|\mathbf{r}_i - \mathbf{R}_\alpha|$. Но в таком случае

$$\nabla_\alpha^n \Psi_e = (-1)^n \nabla_i^n \Psi_e \quad (8.16)$$

и первый член в (8.15) можно переписать в следующем виде:

$$\begin{aligned} \sum_\alpha \frac{\hbar^2}{2M_\alpha} \int \Psi_e^* \Delta_\alpha \Psi_e d\tau_e &= \sum_{i, \alpha} \frac{\hbar^2}{2M_\alpha} \int \Psi_e^* \Delta_i \Psi_e d\tau_e = \\ &= - \sum_{\alpha, i} \frac{m}{M_\alpha} \int \Psi_e^* \left(-\frac{\hbar^2}{2m} \Delta_i \right) \Psi_e d\tau_e = - \sum_{i, \alpha} \frac{m}{M_\alpha} \langle T_i \rangle, \end{aligned} \quad (8.17)$$

где $\langle T_i \rangle$ — средняя кинетическая энергия одного электрона. Таким образом,

$$\sum_\alpha \frac{\hbar^2}{2M_\alpha} \int \Psi_e^* \Delta_\alpha \Psi_e d\tau_e = -ZN \cdot \frac{m}{M} \langle T_i \rangle. \quad (8.18)$$

Но этим членом по сравнению с E можно пренебречь, ошибка, вносимая при этом в выражение для энергии кристалла, крайне мала; она составляет величину порядка отношения массы электрона к массе ядра, которая для Ge, например, составляет величину около 10^{-4} .

Второй член в (8.15) оценивается аналогично:

$$\frac{\hbar^2}{M_\alpha} \left(\int \Phi_Z^* \nabla_\alpha \Phi_Z d\tau_Z \int \Psi_e^* \nabla_\alpha \Psi_e d\tau_e \right) = -\frac{1}{M_\alpha} \langle \mathbf{p}_\alpha \rangle \cdot \langle \mathbf{p}_i \rangle. \quad (8.19)$$

Так как в состоянии термодинамического равновесия для случая классической статистики

$$\left\langle \frac{\mathbf{p}_i^2}{2m} \right\rangle = \left\langle \frac{\mathbf{p}_\alpha^2}{2M_\alpha} \right\rangle \quad (8.20)$$

и

$$\langle p_e \rangle^2 = \frac{8}{3\pi} \langle p_e^2 \rangle, \quad (8.21)$$

то

$$\langle p_e \rangle \approx \sqrt{\frac{m}{M}} \langle p_Z \rangle, \quad (8.22)$$

поэтому второй член в (8.15) составляет величину порядка $\sqrt{m/M}$ от полной энергии кристалла.

Мы выбрали при этом самый «невыгодный» случай, когда волновая функция электронов представляется комбинацией атомных волновых функций (так называемое приближение сильно связанных электронов). Если бы волновая функция электронов не зависела от координат ядер (приближение свободных электронов), то обе поправки в (8.15) были бы равны нулю.

Таким образом приходим к выводу, что, отбрасывая обе поправки, мы делаем ошибку в величине энергии E не более $\sqrt{m/M}$ от E (для германия $\sqrt{m/M}$ составляет примерно 0,3%), поэтому для определения энергии кристалла с достаточной точностью можно решать уравнение

$$\hat{H}_Z \Phi_Z = E_Z \Phi_Z = E \Phi_Z; \quad (8.23)$$

другими словами, полная энергия кристалла с большой точностью совпадает с собственным значением ядерной части гамильтониана.

Из сказанного видно, что адиабатическое приближение для кристалла с гамильтонианом (7.11) приводит к достаточно точному значению энергии, если предположить, что волновая функция кристалла может быть записана в виде

$$\Psi(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots) = \Phi_Z(\mathbf{R}_1, \dots) \Psi_e(\mathbf{r}_1, \dots; \mathbf{R}_1, \dots), \quad (8.24)$$

причем Ψ_e и Φ_Z находятся из уравнений

$$\hat{H}_e \Psi_e = E_e \Psi_e; \quad (8.25)$$

$$\hat{H}_Z \Phi_Z = E_Z \Phi_Z = E \Phi_Z. \quad (8.26)$$

В адиабатическом приближении волновая функция электронов определяется мгновенным положением ядер (член U_{eZ} в \hat{H}_e), в то время как волновая функция ядер определяется усредненным полем электронов (член E_e в \hat{H}_Z).

Резюме § 7—8

1. Гамильтониан кристалла учитывает все виды энергии электронов и ядер: кинетическую энергию электронов и ядер, энергию взаимодействия частиц и энергию всех частиц во внешнем поле. Уравнение Шредингера для кристалла имеет вид (7.11). Оно содержит $3(Z+1)N$ переменных и в общем виде не решается.

2. Адиабатическое приближение позволяет разделить координаты электронов и ядер, что упрощает уравнение (7.11).

Гамильтониан кристалла разбивается на электронную \hat{H}_e — (8.1) — и ядерную \hat{H}_Z — (8.6) — части, а волновая функция кристалла Ψ представляется в виде произведения волновой функции ядер Φ_Z и электронов Ψ_e :

$$\Psi = \Phi_Z \Psi_e, \quad (8.1p)$$

при этом

$$\hat{H}_e \Psi_e = E_e \Psi_e, \quad (8.2p)$$

$$\hat{H}_Z \Phi_Z = E \Phi_Z \quad (8.3p)$$

и члены

$$\sum_{\alpha} \frac{\hbar}{2M_{\alpha}} \left[\int \Psi_e^* \Delta_{\alpha} \Psi_e d\tau_e + 2 \left(\int \Phi_Z^* \nabla_{\alpha} \Phi_Z d\tau_Z \int \Psi_e^* \nabla_{\alpha} \Psi_e d\tau_e \right) \right] \quad (8.4p)$$

отбрасываются вследствие их малости. По энергии при этом достигается точность порядка $\sqrt{\frac{m}{M}}$, но из теоретического рассмотрения выпадают некоторые процессы, связанные с тепловым колебанием решетки.

Волновая функция ядер Φ_Z определяется усредненным движением электронов — членом E_e в \hat{H}_Z , а Ψ_e зависит от мгновенного положения ядер (член U_{eZ} в \mathbf{H}_e). Пренебрежение поправкой δE равносильно тому, что действие Δ_α на Ψ сводится только к действию на Φ_Z , что отражает факт медленного движения ядер по сравнению с электронами.

§ 9. ОДНОЭЛЕКТРОННОЕ ПРИБЛИЖЕНИЕ

В результате адиабатического приближения волновая функция электронов должна удовлетворять уравнению

$$\hat{H}_e \Psi_e = E_e \Psi_e, \quad (9.1)$$

или

$$\left[\sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) + \frac{1}{2} \sum_{i \neq j} U_{ij} + \sum_{i, \alpha} U_{i\alpha} \right] \Psi_e = E_e \Psi_e. \quad (9.2)$$

Это уравнение также не может быть решено, оно должно быть сведено к уравнению для одной частицы. Из уравнения (9.2) непосредственно видно, что для системы электронов оно распадается на систему уравнений, если предположить, что электроны не взаимодействуют: $U_{ij} = 0$. Поэтому возникает задача учесть взаимодействие электронов таким образом, чтобы в конечном итоге рассматривать вместо системы взаимодействующих электронов систему невзаимодействующих электронов. Это достигается посредством введения так называемого самосогласованного поля. Возьмем некоторый i -й электрон. Он находится в поле всех ядер и остальных электронов. Допустим, с помощью внешнего источника удалось создать в каждый момент времени такое же поле в точке нахождения i -го электрона, какое создается остальными электронами. Обозначим потенциальную энергию i -го электрона в этом поле через Ω_i . Очевидно, что она зависит от координат только i -го электрона: $\Omega_i = \Omega_i(\mathbf{r}_i)$.

Если бы удалось создать такие поля для каждого электрона, то энергию попарного взаимодействия электронов можно было бы заменить суммой членов $\Omega_i(\mathbf{r}_i)$:

$$\frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \rightarrow \sum_i \Omega_i(\mathbf{r}_i). \quad (9.3)$$

Потенциальная энергия $\Omega_i(\mathbf{r}_i)$ -го электрона зависит не только от движения всех остальных электронов, но косвенно зависит от движения и самого i -го электрона, поскольку движение i -го электрона

оказывает воздействие на движение остальных электронов. Поэтому поле $\Omega_i(\mathbf{r}_i)$ не только определяет движение i -го электрона, но и само зависит от его движения, в силу чего оно было названо самосогласованным. Поле $\Omega_i(\mathbf{r}_i)$ в принципе может быть найдено методом последовательных приближений. Идею метода нахождения этого поля, развитого первоначально для атома, и рассмотрим дальше, но вначале остановимся на некотором следствии введения самосогласованного поля. Предполагая, что такое поле найдено, можно гамильтониан \hat{H}_e записать в виде

$$\begin{aligned} \hat{H}_e &= \sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) + \frac{1}{2} \sum_{i \neq j} U_{ij} + \sum_{i, \alpha} U_{i\alpha} = \\ &= \sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) + \sum_i \Omega_i(\mathbf{r}_i) + \sum_i \left(\sum_{\alpha} U_{i\alpha} \right) = \sum_i \hat{H}_i, \end{aligned} \quad (9.4)$$

где гамильтониан i -го электрона

$$\hat{H}_i = -\frac{\hbar^2}{2m} \Delta_i + \Omega_i(\mathbf{r}_i) + U_i(\mathbf{r}_i); \quad (9.5)$$

$\Omega_i(\mathbf{r}_i)$ есть потенциальная энергия i -го электрона в поле всех остальных электронов, а $U_i(\mathbf{r}_i)$ — в поле всех ядер. Так как теперь гамильтониан не содержит энергии взаимодействия электронов, то волновая функция системы электронов имеет вид произведения волновых функций отдельных электронов, а энергия системы равна сумме энергий отдельных электронов:

$$\Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots) = \prod_i \psi_i(\mathbf{r}_i), \quad (9.6)$$

$$E_e = \sum_i E_i, \quad (9.7)$$

причем

$$\hat{H}_i \psi_i = E_i \psi_i, \quad (9.8)$$

т. е. введение самосогласованного поля сводит многоэлектронную задачу к одноэлектронной.

Чтобы найти вид $\Omega_i(\mathbf{r}_i)$, запишем уравнение Шредингера для электронной части гамильтониана кристалла в двух формах:

$$\hat{H}_e \Psi_e = \left[\sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) \Psi_e + \frac{1}{2} \sum_{i \neq j} U_{ij} \Psi_e + \sum_i U_i(\mathbf{r}_i) \Psi_e \right] = E_e \Psi_e \quad (9.9)$$

$$\hat{H}_e \Psi_e = \left[\sum_i \left(-\frac{\hbar^2}{2m} \Delta_i \right) \Psi_e + \sum_i \Omega_i(\mathbf{r}_i) \Psi_e + \sum_i U_i(\mathbf{r}_i) \Psi_e \right] = E_e \Psi_e. \quad (9.10)$$

Эти две формы одного и того же уравнения определяют оператор $\Omega_i(\mathbf{r}_i)$, однако его нельзя определить равенством

$$\Omega_i(\mathbf{r}_i) = \frac{1}{2} \sum_{i (i \neq j)} U_{ij}, \quad (9.11)$$

поскольку $\Omega_i(\mathbf{r}_i)$ должно зависеть только от координат i -го электрона, в то время как $\sum_{j (j \neq i)} U_{ij}$ зависит от координат всех электронов. Очевидно, уравнения (9.11) и (9.3) лишены смысла, поскольку необходимо приравнивать не отдельные сомножители членов дифференциальных уравнений (9.9) и (9.10), а сами уравнения.

Для нахождения $\Omega_i(\mathbf{r}_i)$ умножим оба уравнения (9.9) и (9.10) слева на Ψ_e^* и проинтегрируем по координатам всех электронов, после чего вычтем уравнение (9.10) из (9.9). Справа получим нуль. Первые и третьи члены дадут нуль, и мы получим

$$\frac{1}{2} \int \Psi_e^* \sum_{i, j \neq i} U_{ij} \Psi_e d\tau_e - \int \Psi_e^* \sum_i \Omega_i(\mathbf{r}_i) \Psi_e d\tau_e = 0 \quad (9.12)$$

или

$$\sum_i \int \Psi_e^* \Omega_i(\mathbf{r}_i) \Psi_e d\tau_e = \sum_i \int \Psi_e^* \frac{1}{2} \sum_{j (j \neq i)} U_{ij} \Psi_e d\tau_e. \quad (9.13)$$

Так как введение $\Omega_i(\mathbf{r}_i)$ сводит задачу для электронов к системе невзаимодействующих частиц, то Ψ_e можно представить в виде произведения волновых функций отдельных частиц:

$$\Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots) = \prod_i \psi_i(\mathbf{r}_i). \quad (9.14)$$

Учитывая, что $d\tau_e = d\tau_1 d\tau_2 \dots$, можем переписать (9.13) таким образом:

$$\begin{aligned} & \sum_i \int \psi_i^*(\mathbf{r}_i) \dots \Omega_i(\mathbf{r}_i) \psi_i(\mathbf{r}_i) \dots d\tau_1 d\tau_2 \dots = \\ & = \sum_i \int \psi_i^*(\mathbf{r}_i) \Omega_i(\mathbf{r}_i) \psi_i(\mathbf{r}_i) d\tau_i = \\ & = \frac{1}{2} \sum_{i, j \neq i} \int \psi_i^*(\mathbf{r}_i) \dots U_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \psi_i(\mathbf{r}_i) \dots d\tau_1 d\tau_2 \dots = \\ & = \sum_i \int \psi_i^*(\mathbf{r}_i) \left[\frac{1}{2} \sum_{j \neq i} \int \psi_j^*(\mathbf{r}_j) U_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) \psi_j(\mathbf{r}_j) d\tau_j \right] \psi_i d\tau_i. \quad (9.15) \end{aligned}$$

Сравнивая второй и последние члены цепи равенств (9.15), можем записать

$$\Omega_i(\mathbf{r}_i) = \frac{1}{2} \sum_{j \neq i} \int |\psi_j(\mathbf{r}_j)|^2 \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} d\tau_j. \quad (9.16)$$

Выражение для $\Omega_i(\mathbf{r}_i)$ имеет следующий смысл: $e|\psi_j(\mathbf{r}_j)|^2$ есть плотность заряда электронного облака j -го электрона в точке \mathbf{r}_j ; $e|\psi_j(\mathbf{r}_j)|^2 d\tau_j$ есть элемент заряда, который определяет потенциал в точке \mathbf{r}_i . Интегрируя по всем значениям координат j -го электрона, получаем энергию взаимодействия i -го электрона с «размазанным» j -электроном.

Если подставить выражение (9.16) для $\Omega_i(\mathbf{r}_i)$ в (9.8), то уравнение для нахождения функций $\psi_i(\mathbf{r}_i)$ будет иметь вид

$$-\frac{\hbar^2}{2m} \Delta \psi_i(\mathbf{r}_i) + \left[\frac{1}{2} \sum_{i \neq j} \int |\psi_j(\mathbf{r}_j)|^2 \frac{e^2 d\tau_j}{4\pi\epsilon_0 r_{ij}} \right] \psi_i(\mathbf{r}_i) + U_i(\mathbf{r}_i, \mathbf{R}_1, \mathbf{R}_2, \dots) \psi_i(\mathbf{r}_i) = E_i \psi_i(\mathbf{r}_i).$$

Чтобы найти $\Omega_i(\mathbf{r}_i)$, надо знать все $\psi_j(\mathbf{r}_j)$, для нахождения которых в свою очередь необходимо знать все $\Omega_i(\mathbf{r}_i)$. Уравнение (9.17) носит название уравнения Хартри, оно является интегродифференциальным уравнением. Его решение можно искать методом последовательных приближений. Взяв в качестве нулевого приближения некоторые функции $\psi_j^{(0)}(\mathbf{r}_j)$, необходимо с их помощью вычислить $\Omega_i^{(0)}(\mathbf{r}_i)$. Подставив $\Omega_i^{(0)}(\mathbf{r}_i)$ в (9.17), найдем некоторые функции $\psi_i^{(1)}(\mathbf{r}_i)$, с их помощью вычисляется $\Omega_i^{(1)}(\mathbf{r}_i)$ и т. д. Эта процедура повторяется до тех пор, пока $(n+1)$ -е приближение не совпадет с n -м приближением в пределах заданной точности. Уравнение Хартри имеет тот недостаток, что не учитывает принцип Паули. Учет принципа Паули приводит к уравнению Хартри — Фока. Принцип Паули требует, чтобы волновая функция электронов была антисимметричной относительно перестановки двух электронов с учетом их координат и проекции спина. В то же время $\prod_i \psi_i(\mathbf{r}_i)$ не удовлетворяет

этому условию. Правильная комбинация волновых функций отдельных электронов задается в виде определителя Слэтера:

$$\Psi_e(\mathbf{q}_1, \mathbf{q}_2, \dots) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{q}_1) \psi_1(\mathbf{q}_2) \dots \\ \psi_2(\mathbf{q}_1) \psi_2(\mathbf{q}_2) \dots \\ \dots \dots \dots \end{vmatrix}, \quad (9.18)$$

где N — число электронов, а \mathbf{q}_i обозначает набор четырех переменных x_i, y_i, z_i и s_{z_i} . Волновая функция Ψ_e удовлетворяет условиям:

$$\Psi_e(\dots \mathbf{q}_i \dots \mathbf{q}_k \dots) = -\Psi_e(\dots \mathbf{q}_k \dots \mathbf{q}_i \dots) \quad (9.19)$$

$$\int \Psi_e^* \Psi_e d\mathbf{q}_e = 1.$$

Используя выражение для Ψ_e в виде определителя Слэтера, найдем выражение для энергии E_i

$$E_i = \int \Psi_e^*(\mathbf{q}_1, \dots) \left[-\frac{\hbar^2}{2m} \Delta_i + U_i(\mathbf{r}_i, \mathbf{R}_1, \mathbf{R}_2, \dots) \right] \Psi_e d\mathbf{q}_e + \frac{1}{8\pi\epsilon_0} \sum_{i \neq j} \int \Psi_e^*(\mathbf{q}_1, \dots) \frac{e^2}{r_{ij}} \Psi_e(\mathbf{q}_1, \dots) d\mathbf{q}_e. \quad (9.20)$$

Через $d\mathbf{q}_e$ обозначен элемент объема в конфигурационном пространстве системы электронов, включающий спиновую переменную, интегрирование по $d\mathbf{q}_e$ означает интегрирование по координатам и суммирование по спиновым переменным всех электронов.

Если первый интеграл в (9.20) не отличается от соответствующего члена в уравнении Хартри, то второй член отличается наличием обменных членов, отсутствующих в уравнении Хартри. Это связано с тем, что при интегрировании по dq_e нужно оставить члены, содержащие координаты i и j электронов, которые могут теперь находиться в любом состоянии ψ_k и ψ_l и ψ'_k, ψ'_l :

$$\begin{aligned} & \frac{1}{8\pi\epsilon_0} \sum_{i \neq j} \Psi_e(q_1, \dots) \frac{e^2}{r_{ij}} \Psi_e(q_1, \dots) dq_e = \\ & = \frac{1}{8\pi\epsilon_0 N!} \sum_i \sum_{k, l} (-1)^{k+l} \int \psi_k^*(q_i) \psi_l^*(q_j) \frac{e^2}{r_{ij}} \psi_k(q_i) \psi_l(q_j) dq_i dq_j. \quad (9.21) \end{aligned}$$

При $k=l$ имеем обычную усредненную кулоновскую энергию взаимодействия электронов, а при $k \neq l$ получаем обменную энергию. Однако решение методом Хартри—Фока уравнения Шредингера для кристалла практически невозможно.

Резюме § 9

1. Движение некоторого электрона зависит от состояния движения всех остальных электронов, но поскольку оно само влияет на их движение, то движение всех электронов является взаимосогласованным. Это учитывается введением величины Ω_i , представляющей собой энергию i электрона в поле всех остальных электронов с учетом его воздействия на их движение. Поле Ω_i носит название самосогласованного.

2. Благодаря введению самосогласованного поля уравнение для системы электронов сводится к системе уравнений для одного электрона. Самосогласованное поле в принципе может быть найдено из решения системы уравнений Хартри (9.17) или Хартри—Фока (9.20).

3. Введение самосогласованного поля позволяет рассматривать электроны как частицы не взаимодействующие. Тем самым квантовая механика подтверждает представление об электронном газе как газе идеальном.

§ 10. ПЕРИОДИЧЕСКОЕ ПОЛЕ РЕШЕТКИ КРИСТАЛЛА. ОПЕРАТОР ТРАНСЛЯЦИИ

Введение самосогласованного поля сводит задачу для системы взаимодействующих частиц к одноэлектронной задаче. Обозначив потенциальную энергию произвольного i -го электрона буквой U без указания значка i :

$$U = U(\mathbf{r}) = \Omega(\mathbf{r}) + U(\mathbf{r}, \mathbf{R}_1, \mathbf{R}_2, \dots), \quad (10.1)$$

можем представить гамильтониан произвольного электрона в виде

$$\hat{H} = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}). \quad (10.2)$$

Энергия E электрона и его волновая функция находятся из уравнения

$$\left[-\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (10.3)$$

Относительно потенциальной энергии $U(\mathbf{r})$ можно сказать, что она должна быть периодической функцией координат. Обозначим периоды идентичности решетки по трем произвольным направлениям через $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ и через \mathbf{n} обозначим вектор

$$\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \quad (10.4)$$

где n_1, n_2, n_3 — произвольные целые числа; вектор \mathbf{n} называется вектором трансляции. Условие периодичности имеет вид

$$U(\mathbf{r} + \mathbf{n}) = U(\mathbf{r}). \quad (10.5)$$

Введем оператор трансляции $\hat{T}(\mathbf{n})$, который перемещает пространство на вектор \mathbf{n} , вследствие чего координаты всех точек $M(\mathbf{r})$ изменятся на \mathbf{n} , поэтому определим $\hat{T}(\mathbf{n})$ из условия

$$\hat{T}(\mathbf{n}) f(\mathbf{r}) = f(\mathbf{r} + \mathbf{n}), \quad (10.6)$$

где $f(\mathbf{r})$ — произвольная функция от координат. Разлагая $f(\mathbf{r} + \mathbf{n})$ в ряд Тейлора в точке \mathbf{r} , получим

$$\begin{aligned} f(\mathbf{r} + \mathbf{n}) &= \sum_{m=0}^{\infty} \frac{1}{m!} \frac{d^m f(\mathbf{r})}{d\mathbf{r}^m} \mathbf{n}^m = \left(\sum_{m=0}^{\infty} \frac{1}{m!} \mathbf{n}^m \nabla^m \right) f(\mathbf{r}) = \\ &= e^{(\mathbf{n}\nabla)} f(\mathbf{r}) = \hat{T}(\mathbf{n}) f(\mathbf{r}) \quad \left(\nabla \equiv \frac{d}{d\mathbf{r}} \right). \end{aligned} \quad (10.7)$$

Следовательно, оператор трансляции имеет вид

$$\hat{T}(\mathbf{n}) = e^{(\mathbf{n}\nabla)}. \quad (10.8)$$

Если подействовать оператором трансляции на периодическую функцию, то

$$\hat{T}(\mathbf{n}) U(\mathbf{r}) = U(\mathbf{r} + \mathbf{n}) = U(\mathbf{r}). \quad (10.9)$$

Таким образом, произвольная периодическая функция является собственной функцией оператора $\hat{T}(\mathbf{n})$, соответствующей единичному собственному значению.

Если рассмотреть действие оператора трансляции на произведение потенциальной энергии (оператор) и произвольной волновой функции, то

$$\begin{aligned} \hat{T}(\mathbf{n}) \hat{U}(\mathbf{r}) \psi(\mathbf{r}) &= \hat{T}(\mathbf{n}) U(\mathbf{r}) \psi(\mathbf{r}) = U(\mathbf{r} + \mathbf{n}) \psi(\mathbf{r} + \mathbf{n}) = \\ &= U(\mathbf{r}) \psi(\mathbf{r} + \mathbf{n}) = \hat{U}(\mathbf{r}) \hat{T}(\mathbf{n}) \psi(\mathbf{r}), \end{aligned} \quad (10.10)$$

$$\hat{T}(\mathbf{n}) \hat{U} - \hat{U} \hat{T}(\mathbf{n}) = 0. \quad (10.11)$$

Другими словами, оператор трансляции коммутирует с оператором потенциальной энергии $\hat{U}(\mathbf{r}) \equiv U(\mathbf{r})$ электрона в кристалле. Легко видеть, что оператор кинетической энергии коммутирует с оператором трансляции:

$$\hat{T}\hat{T}(\mathbf{n}) = -\frac{\hbar^2}{2m}\nabla^2 \sum_{l=0}^{\infty} \frac{1}{l!} (\mathbf{n}^l \nabla^l) = \sum_{l=0}^{\infty} \frac{1}{l!} (\mathbf{n}^l \nabla^l) \left(-\frac{\hbar^2}{2m} \nabla^2 \right) = \hat{T}(\mathbf{n}) \hat{T}. \quad (10.12)$$

Из (10.11) и (10.12) следует, что оператор трансляции коммутирует с оператором Гамильтона для периодического потенциального поля:

$$\hat{H}\hat{T}(\mathbf{n}) - \hat{T}(\mathbf{n})\hat{H} = 0, \quad (10.13)$$

из чего следует, что $\hat{T}(\mathbf{n})$ является сохраняющейся во времени величиной

$$\frac{d\hat{T}(\mathbf{n})}{dt} = \frac{1}{i\hbar} \{ \hat{T}(\mathbf{n}) \hat{H} - \hat{H} \hat{T}(\mathbf{n}) \} \equiv [\hat{H}, \hat{T}(\mathbf{n})] = 0 \quad (10.14)$$

и что операторы $\hat{T}(\mathbf{n})$ и \hat{H} имеют одну и ту же систему собственных функций.

Рассмотрим оператор, обратный к $\hat{T}(\mathbf{n})$, который обозначим как $\hat{T}^{-1}(\mathbf{n})$. Произведение прямого и обратного операторов должно оставлять пространство неизменным, должно дать тождественное преобразование, которое определяется единичным оператором \hat{I} :

$$\hat{T}(\mathbf{n}) \hat{T}^{-1}(\mathbf{n}) \equiv \hat{T}^{-1}(\mathbf{n}) \hat{T}(\mathbf{n}) \equiv \hat{I}, \quad (10.15)$$

$$\hat{I}f(\mathbf{r}) \equiv f(\mathbf{r}). \quad (10.16)$$

Из явного выражения $\hat{T}(\mathbf{n})$ следует, что $\hat{T}^{-1}(\mathbf{n})$ приводит к трансляции пространства на вектор $-\mathbf{n}$:

$$\hat{T}^{-1}(\mathbf{n}) = [e^{(\mathbf{n}\nabla)}]^{-1} = e^{(-\mathbf{n}\nabla)} = \hat{T}(-\mathbf{n}). \quad (10.17)$$

Квадрат оператора трансляции $\hat{T}^2(\mathbf{n})$ есть оператор смещения на вектор $2\mathbf{n}$, и аналогично

$$\hat{T}^s(\mathbf{n}) = \hat{T}(s\mathbf{n}). \quad (10.18)$$

Пусть $\psi(\mathbf{r})$ есть собственная функция оператора трансляции, соответствующая собственному значению $T(\mathbf{n})$:

$$\hat{T}(\mathbf{n})\psi(\mathbf{r}) = T(\mathbf{n})\psi(\mathbf{r}). \quad (10.19)$$

Собственные значения $\hat{T}^2(\mathbf{n})$ связаны с $T(\mathbf{n})$ условием

$$\hat{T}^2(\mathbf{n})\psi(\mathbf{r}) = \hat{T}(\mathbf{n})T(\mathbf{n})\psi(\mathbf{r}) = T^2(\mathbf{n})\psi(\mathbf{r}). \quad (10.20)$$

Из (10.18) и (10.20) следует

$$T'(n) = T(ln), \quad (10.21)$$

что справедливо для экспоненциальной функции.

Запишем $T(n)$ в виде

$$T(n) = e^{i\varphi(n)}. \quad (10.22)$$

Так как трансляция на вектор n есть совокупность независимых трансляций по трем осям, то для того, чтобы выполнялось условие (10.21) и (10.22), необходимо, чтобы $\varphi(n)$ была линейной скалярной функцией:

$$\varphi(n) = (kn) = \kappa_1 n_1 a_1 + \kappa_2 n_2 a_2 + \kappa_3 n_3 a_3. \quad (10.23)$$

Из условия нормировки, которая не зависит от выбора начала координат

$$1 = \int |\psi(r+n)|^2 d\tau = |e^{i\varphi(n)}|^2 \int |\psi(r)|^2 d\tau = |e^{i\varphi(n)}|^2, \quad (10.24)$$

следует, что $\varphi(n)$ является вещественной функцией, т. е. *вектор k должен быть вещественным*. Так как размерность вектора k — обратная длина, то он получил название волнового вектора. Ниже увидим, что это название оправдано не только его размерностью.

Итак,

$$\hat{T}(n)\psi(r) = \psi(r+n) = e^{i(kn)}\psi(r) = T(n)\psi(r). \quad (10.24')$$

Но собственные функции $\hat{T}(n)$ и \hat{H} совпадают, поэтому для собственных функций гамильтониана можно записать

$$\psi(r+n) = e^{i(kn)}\psi(r). \quad (10.25)$$

Условие (10.25) называется *трансляционным свойством волновой функции*. Можем сказать, что собственные функции оператора Гамильтона при движении электрона в периодическом поле удовлетворяют трансляционному условию (10.25).

Вектор k является характерным для данной волновой функции, поэтому его необходимо указывать в виде индекса внизу:

$$\psi(r) = \psi_k(r). \quad (10.26)$$

Но так как

$$\hat{H}\psi_k(r) = E\psi_k(r), \quad (10.27)$$

то можно утверждать, что *энергия E должна быть функцией волнового вектора: $E = E(k)$* . Нахождение этой зависимости представляет собой одну из важнейших задач современной физики твердого тела.

Мы видим, что нахождение собственных функций оператора трансляции связано с решением уравнения Шредингера, поэтому их вид должен зависеть от функционального выражения потенциальной

энергии $U(\mathbf{r})$. Однако общий характер волновой функции можно найти, не решая уравнение Шредингера. Для этого необходимо вспомнить, что собственная функция оператора ∇ имеет вид

$$\psi(\mathbf{r}) = Ae^{i(\mathbf{k}\mathbf{r})} \quad (10.28)$$

и удовлетворяет уравнению

$$\nabla\psi(\mathbf{r}) = i\mathbf{k}\psi(\mathbf{r}). \quad (10.29)$$

Она же является собственной функцией оператора ∇^l и, следовательно, оператора e^{∇} . Легко проверить непосредственной подстановкой, что функция $e^{i(\mathbf{k}\mathbf{r})}$ является собственной функцией оператора трансляции:

$$\begin{aligned} \hat{T}(\mathbf{n}) e^{i(\mathbf{k}\mathbf{r})} &= \left(\sum_{m=0}^{\infty} \frac{1}{m!} \mathbf{n}^m \nabla^m \right) e^{i(\mathbf{k}\mathbf{r})} = \sum_{m=0}^{\infty} \frac{1}{m!} (\mathbf{n} \cdot i\mathbf{k})^m e^{i(\mathbf{k}\mathbf{r})} = e^{i(\mathbf{k}\mathbf{n})} e^{i(\mathbf{k}\mathbf{r})} = \\ &= T(\mathbf{n}) e^{i(\mathbf{k}\mathbf{r})}. \end{aligned} \quad (10.30)$$

Однако это не единственно возможный вид собственной функции оператора трансляции $\hat{T}(\mathbf{n})$. Если умножить $e^{i(\mathbf{k}\mathbf{r})}$ на произвольную периодическую функцию $\varphi(\mathbf{r} + \mathbf{n}) = \varphi(\mathbf{r})$, то это произведение будет также собственной функцией $T(\mathbf{n})$. Выбор функции $\varphi(\mathbf{r})$ должен определяться тем, что $\psi_{\mathbf{k}}$ должна являться также решением и уравнения Шредингера.

Приходим к важному результату, что решение уравнения Шредингера для электрона в периодическом поле $U(\mathbf{r} + \mathbf{n}) = U(\mathbf{r})$ должно иметь вид

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i(\mathbf{k}\mathbf{r})} \varphi(\mathbf{r}) = e^{i(\mathbf{k}\mathbf{r})} \varphi_{\mathbf{k}}(\mathbf{r}), \quad (10.31)$$

где $\varphi_{\mathbf{k}}(\mathbf{r} + \mathbf{n}) = \varphi_{\mathbf{k}}(\mathbf{r})$ является периодической функцией с периодичностью потенциального поля $U(\mathbf{r})$; $e^{i(\mathbf{k}\mathbf{r})}$ представляет собой плоскую волну, идущую в направлении вектора \mathbf{k} . Выражение (10.31) для $\psi_{\mathbf{k}}(\mathbf{r})$ носит название волны Блоха, поскольку наглядно волновую функцию $\psi_{\mathbf{k}}(\mathbf{r})$ можно представить в виде плоской волны $e^{i(\mathbf{k}\mathbf{r})}$ с переменной амплитудой $\varphi_{\mathbf{k}}(\mathbf{r})$, модулированной в такт решетке кристалла. Мы указали \mathbf{k} в качестве индекса у $\varphi(\mathbf{r})$, поскольку при разных \mathbf{k} функции $\varphi(\mathbf{r})$ могут быть различными.

Резюме § 10

1. Оператор трансляции $\hat{T}(\mathbf{n})$ смещает пространство на вектор трансляции \mathbf{n} , в силу чего функции координат точек пространства меняются при действии на них $\hat{T}(\mathbf{n})$ согласно (10.6) $\hat{T}(\mathbf{n})$ коммутирует с \hat{H} периодического поля.

2. Собственные значения оператора трансляции равны

$$T(\mathbf{n}) = e^{i(\mathbf{k}\mathbf{n})}, \quad (10.1)$$

где \mathbf{k} — произвольный вещественный вектор, называемый волновым.

3. Решение уравнения Шредингера для электрона в периодическом поле имеет вид волны, или функции, Блоха:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i(\mathbf{k}\mathbf{r})} \varphi_{\mathbf{k}}(\mathbf{r}), \quad (10.2)$$

при этом

$$\varphi_{\mathbf{k}}(\mathbf{r} + \mathbf{n}) = \varphi_{\mathbf{k}}(\mathbf{r}).$$

4. Энергия электрона является функцией волнового вектора:

$$E = E(\mathbf{k}). \quad (10.3)$$

§ 11. КВАЗИИМПУЛЬС

Многие рассматриваемые в физике величины обладают важнейшим свойством — для них существуют законы сохранения, т. е. при определенных условиях эти величины сохраняются. Действительно, импульс $\mathbf{p} = m\mathbf{v}$ сохраняется при движении частицы в пространстве с постоянной потенциальной энергией. Момент импульса $\mathbf{M} = [\mathbf{r}\mathbf{p}]$ сохраняется в поле с центральной симметрией $U(\mathbf{r}) = U(r)$. Энергия изолированной системы сохраняется, если функция Гамильтона $H(\mathbf{p}, \mathbf{r})$ не зависит явно от времени. Эти же величины являются интегралами движения в квантовой механике и записываются в форме операторов: *не зависящая явно от времени величина L сохраняется, если соответствующий ей оператор \hat{L} коммутирует с оператором Гамильтона*, так как

$$\frac{d\hat{L}}{dt} = [\hat{H}, \hat{L}] = \frac{1}{i\hbar} [\hat{L}\hat{H} - \hat{H}\hat{L}]. \quad (11.1)$$

Физической основой законов сохранения являются определенные свойства симметрии пространства и времени.

Действительно, закон сохранения импульса отражает однородность пространства; изотропность пространства приводит к закону сохранения момента количества движения; однородность времени приводит к закону сохранения энергии; неразличимость левого и правого винтов приводит к закону сохранения четности. Другими словами, *наличие некоторой симметрии пространства и времени приводит к сохранению определенной физической величины*. Если некоторое воздействие нарушает симметрию, то соответствующая данной симметрии величина начинает изменяться под влиянием этого же воздействия. Три из указанных выше закона легко получить, используя канонические уравнения Гамильтона.

Рассматривая движение электрона в периодическом потенциальном поле решетки кристалла, можем высказать следующее утверждение: *трансляционной симметрии потенциального поля решетки должна соответствовать некоторая физическая величина, сохраняющаяся при движении частицы в этом поле. Назовем эту величину квазиимпульсом*. Простейшим основанием для названия «квазиим-

пульс» служит его размерность. Так как трансляционная симметрия отражает неизменность свойств пространства при его смещении на любое целое число периодов идентичности решетки, то квазиимпульс должен иметь ту же размерность, что и импульс, отражающий однородность, неизменность свойств пространства при произвольном смещении. В дальнейшем будет показано, что свойства квазиимпульса во многом подобны свойствам импульса. Квазиимпульсу \hat{P} должен соответствовать оператор \hat{P} , коммутирующий с гамильтонианом решетки:

$$\hat{P}\hat{H} - \hat{H}\hat{P} = 0. \quad (11.2)$$

Таким образом, мы можем утверждать, что при движении электрона в поле решетки собственные функции операторов \hat{H} и \hat{P} должны совпадать, а между их собственными значениями должна быть некоторая функциональная связь:

$$E = E(P). \quad (11.3)$$

Другими словами, энергия электрона должна быть функцией квазиимпульса.

Совершенно очевидно, что из условия коммутации \hat{P} и \hat{H} следует, что оператор квазиимпульса не может иметь вид $-i\hbar\nabla$, т. е. вид обычного оператора импульса, поскольку определенный таким образом оператор квазиимпульса не коммутирует с гамильтонианом, что и приводит к несохранению обычного импульса при движении частицы в поле решетки:

$$\frac{d\hat{p}}{dt} = \frac{1}{i\hbar}(\hat{p}\hat{H} - \hat{H}\hat{p}) = -(\nabla U) = F_i. \quad (11.4)$$

В то же время очевидно, что между \hat{p} и \hat{P} должна быть определенная связь. Действительно, пусть $\nabla U \rightarrow 0$, т. е. потенциальная энергия поля решетки стремится к константе. В этом предельном случае квазиимпульс и импульс должны быть тождественными. Но это означает, во-первых, что оператор квазиимпульса должен содержать в себе величину, зависящую от вида потенциального поля $U(\mathbf{r})$, которая стремится к нулю при $\nabla U \rightarrow 0$. Это позволяет написать для \hat{P} :

$$\hat{P} = -i\hbar\nabla + i\hbar\hat{\gamma}(\mathbf{r}), \quad (11.5)$$

где $\hat{\gamma}(\mathbf{r}) \rightarrow 0$ при $\nabla U \rightarrow 0$. Наличие $\hat{\gamma}(\mathbf{r})$ должно обеспечить коммутацию \hat{P} и \hat{H} .

Будем искать вид оператора \hat{P} из уравнения на собственные функции и собственные значения, учитывая, что ψ_k является собственной функцией оператора квазиимпульса:

$$\hat{P}\psi_k(\mathbf{r}) = P\psi_k(\mathbf{r}). \quad (11.6)$$

Чтобы найти отсюда вид оператора \hat{P} , необходимо представить ψ_k в виде волны Блоха, \hat{P} в виде $-i\hbar\nabla + i\hbar\hat{\gamma}$. Для нахождения $\hat{\gamma}$ мы получим уравнение

$$\hat{P}\psi_k(\mathbf{r}) = -i\hbar k\psi_k(\mathbf{r}) + e^{i(k\mathbf{r})}(-i\hbar\nabla\psi_k(\mathbf{r}) + i\hbar\hat{\gamma}\psi_k(\mathbf{r})) = \\ = \hbar k\psi_k(\mathbf{r}) + i\hbar[\hat{\gamma} - \nabla \ln \varphi_k(\mathbf{r})]\psi_k(\mathbf{r}) = P\psi_k(\mathbf{r}). \quad (11.7)$$

Из (11.7) можно записать

$$P = \hbar k \quad (11.8)$$

и

$$\hat{\gamma} \equiv (\nabla \ln \varphi_k(\mathbf{r})). \quad (11.9)$$

Мы видим, что $\hat{\gamma}$ есть оператор умножения, он зависит от вида потенциального поля $U(\mathbf{r})$ посредством зависимости через периодическую функцию $\varphi_k(\mathbf{r})$. При стремлении $\nabla U(\mathbf{r})$ к нулю $\varphi_k(\mathbf{r})$ будет стремиться к постоянной и $\hat{\gamma}$ будет стремиться к нулю, что обеспечивает тождество квазиимпульса с обычным импульсом \hat{p} в этом предельном случае.

Таким образом, оператор квазиимпульса имеет вид

$$\hat{P}(\mathbf{r}) = -i\hbar\nabla + i\hbar(\nabla \ln \varphi_k(\mathbf{r})). \quad (11.10)$$

Пусть теперь на периодическое поле $U(\mathbf{r})$ наложено некоторое дополнительное поле $V(\mathbf{r})$, не обладающее той же периодичностью. В этом случае

$$\hat{H} = \hat{T} + \hat{U} + \hat{V} = \hat{H}_0 + \hat{V}. \quad (11.11)$$

Так как P коммутирует с \hat{H}_0 — гамильтонианом поля решетки, а $(\nabla \ln \varphi_k)$ коммутирует с \hat{V} , поскольку они оба являются операторами умножения, то производная по времени от квазиимпульса будет равна внешней силе \hat{F}_α :

$$\hat{F}_\alpha = -(\nabla V). \quad (11.12)$$

Действительно,

$$\frac{d\hat{P}}{dt} = [\hat{H}, \hat{P}] = [\hat{V}\hat{P}] = \frac{1}{i\hbar} \{(-i\hbar\nabla)V - V(-i\hbar\nabla)\} = -(\nabla V) = F_\alpha. \quad (11.13)$$

Таким образом, квазиимпульс меняется под действием неперiodической части потенциального поля $-\nabla V(\mathbf{r})$. Это означает, что при любых нарушениях идеальности поля решетки происходит изменение квазиимпульса P и, следовательно, на любых нарушениях идеальной структуры решетки должно происходить рассеяние электронных волн. Такими нарушениями периодичности $U(\mathbf{r})$ являются тепловые колебания и дефекты решетки. Рассеяние на них является физической причиной конечного сопротивления электрическому току. Если на

идеальный кристалл наложить внешнее силовое поле $V(\mathbf{r})$, то квазиимпульс меняется только под действием внешней силы \mathbf{F}_a , в то время как импульс меняется в результате действия внешних \mathbf{F}_a и внутренних $\mathbf{F}_i = -\nabla U(\mathbf{r})$ сил:

$$d\mathbf{p}/dt = \mathbf{F}_i + \mathbf{F}_a. \quad (11.14)$$

Энергия электрона должна зависеть от волнового вектора \mathbf{k} или квазиимпулься $\mathbf{P} = \hbar\mathbf{k}$. Конкретный вид функциональной зависимости $E(\mathbf{k})$ или $E(\mathbf{P})$ может быть найден только при решении уравнения Шредингера

$$\hat{H}\psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r}). \quad (11.15)$$

Найдем уравнение, которому должна удовлетворять функция $\psi_{\mathbf{k}}$. Для этого учтем, что

$$\nabla\psi_{\mathbf{k}}(\mathbf{r}) = i\mathbf{k}\psi_{\mathbf{k}}(\mathbf{r}) + e^{i(\mathbf{k}\mathbf{r})}[\nabla\varphi_{\mathbf{k}}(\mathbf{r})]; \quad (11.16)$$

$$\Delta\psi_{\mathbf{k}}(\mathbf{r}) = i\mathbf{k}[i\mathbf{k}\psi_{\mathbf{k}}(\mathbf{r}) + e^{i(\mathbf{k}\mathbf{r})}(\nabla\varphi_{\mathbf{k}}(\mathbf{r}))] + i\mathbf{k}e^{i(\mathbf{k}\mathbf{r})}[\nabla\varphi_{\mathbf{k}}(\mathbf{r})] + e^{i(\mathbf{k}\mathbf{r})}[\Delta\varphi_{\mathbf{k}}(\mathbf{r})]. \quad (11.17)$$

Следовательно, подставляя (11.17) в уравнение (11.15) и сокращая на $e^{i(\mathbf{k}\mathbf{r})}$, получим уравнение для $\varphi_{\mathbf{k}}(\mathbf{r})$:

$$-\frac{\hbar^2}{2m}[-\mathbf{k}^2\varphi_{\mathbf{k}}(\mathbf{r}) + 2i(\mathbf{k} \cdot \nabla\varphi_{\mathbf{k}}(\mathbf{r})) + \Delta\varphi_{\mathbf{k}}(\mathbf{r})] + U(\mathbf{r})\varphi_{\mathbf{k}} = E(\mathbf{k})\varphi_{\mathbf{k}}(\mathbf{r}), \quad (11.18)$$

или

$$-\frac{\hbar^2}{2m}\Delta\varphi_{\mathbf{k}}(\mathbf{r}) + \left[\frac{\hbar^2\mathbf{k}^2}{2m} - \frac{i\hbar^2}{m}(\mathbf{k}\nabla) + U(\mathbf{r})\right]\varphi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\varphi_{\mathbf{k}}(\mathbf{r}). \quad (11.19)$$

Уравнение (11.19) показывает, что $\varphi_{\mathbf{k}}(\mathbf{r})$ зависит от значения \mathbf{k} , поэтому мы и указываем \mathbf{k} в виде индекса внизу. Так как энергия — вещественная функция, т. е.

$$E^*(\mathbf{k}) = E(\mathbf{k}), \quad (11.20)$$

то уравнение Шредингера для комплексно-сопряженной волновой функции $\psi_{\mathbf{k}}^*(\mathbf{r})$ можно записать в виде

$$\hat{H}^*\psi_{\mathbf{k}}^*(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}^*(\mathbf{r}). \quad (11.21)$$

Учитывая, что

$$\hat{H}^* = -\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}) \quad (11.22)$$

и

$$\psi_{\mathbf{k}}^*(\mathbf{r}) = e^{-i(\mathbf{k}\mathbf{r})}\varphi_{\mathbf{k}}^*(\mathbf{r}), \quad (11.23)$$

получим для $\varphi_{\mathbf{k}}^*(\mathbf{r})$

$$-\frac{\hbar^2}{2m}\Delta\varphi_{\mathbf{k}}^*(\mathbf{r}) + \left[\frac{\hbar^2\mathbf{k}^2}{2m} + \frac{i\hbar^2}{m}(\mathbf{k}\nabla) + U(\mathbf{r})\right]\varphi_{\mathbf{k}}^*(\mathbf{r}) = E(\mathbf{k})\varphi_{\mathbf{k}}^*(\mathbf{r}). \quad (11.24)$$

Запишем теперь уравнение (11.19) для функции с вектором $(-\mathbf{k})$:

$$-\frac{\hbar^2}{2m} \Delta \varphi_{-\mathbf{k}}(\mathbf{r}) + \left[\frac{\hbar^2 \mathbf{k}^2}{2m} + \frac{i\hbar^2}{m} (\mathbf{k}\nabla) + U(\mathbf{r}) \right] \varphi_{-\mathbf{k}}(\mathbf{r}) = E(-\mathbf{k}) \varphi_{-\mathbf{k}}(\mathbf{r}). \quad (11.25)$$

Если $\varphi_{\mathbf{k}}^*(\mathbf{r}) = \varphi_{-\mathbf{k}}(\mathbf{r})$, то уравнения (11.24) и (11.25) для них совпадают, что означает выполнение условия

$$E(\mathbf{k}) = E(-\mathbf{k}), \quad (11.26)$$

т. е. энергия является четной функцией волнового вектора. В окрестности точки $\mathbf{k} = 0$ энергия зависит по крайней мере от \mathbf{k}^2 .

Если рассматривать пространство k_x, k_y, k_z , то уравнение

$$E(\mathbf{k}) = \text{const} \quad (11.27)$$

определяет некоторую поверхность — поверхность постоянной энергии, или изоэнергетическую поверхность. Форма изоэнергетических поверхностей определяет многие свойства полупроводников.

Резюме § 11

1. Трансляционной симметрии поля решетки должна соответствовать сохраняющаяся физическая величина, называемая квазиимпульсом.

2. Оператор квазиимпульса коммутирует с гамильтонианом поля решетки; его собственные функции есть функции Блоха $\psi_{\mathbf{k}}(\mathbf{r})$, собственные значения \mathbf{P} связаны с волновым вектором \mathbf{k} :

$$\mathbf{P} = \hbar \mathbf{k}, \quad (11.1p)$$

а сам оператор $\hat{\mathbf{P}}$ имеет вид

$$\hat{\mathbf{P}} = -i\hbar \nabla + i\hbar [\nabla \ln \varphi_{\mathbf{k}}(\mathbf{r})]. \quad (11.2p)$$

3. Энергия является функцией квазиимпульса и волнового вектора

$$E = E(\mathbf{P}); \quad E = E(\mathbf{k}), \quad (11.3p)$$

а уравнение

$$E(\mathbf{P}) = \text{const} \quad \text{или} \quad E(\mathbf{k}) = \text{const} \quad (11.4p)$$

определяет в пространстве \mathbf{P} или \mathbf{k} поверхность, называемую изоэнергетической.

4. Если на кристалл наложить поле $V(\mathbf{r})$, не обладающее периодичностью решетки, то квазиимпульс меняется в соответствии с уравнением

$$d\mathbf{P}/dt = -\nabla V(\mathbf{r}) = \mathbf{F}_a, \quad (11.5p)$$

В то время как импульс \mathbf{p} меняется в результате действия внешних сил \mathbf{F}_a и сил поля решетки $\mathbf{F}_i = -\nabla U(\mathbf{r})$:

$$d\mathbf{p}/dt = \mathbf{F}_a + \mathbf{F}_i. \quad (11.6p)$$

§ 12. ЭФФЕКТИВНАЯ МАССА ЭЛЕКТРОНА

Пусть точка \mathbf{k}_0 (или \mathbf{P}_0) является точкой экстремума

$$E(\mathbf{k}_0) = E_0 = E_{extr}. \quad (12.1)$$

Помимо точки \mathbf{k}_0 должны быть и другие экстремальные точки, например симметричная ей точка $-\mathbf{k}_0$. Поскольку четность энергии должна проявляться не только относительно \mathbf{k} , но и его проекций:

$$E(k_x, k_y, k_z) = E(-k_x, k_y, k_z) = \dots, \quad (12.2)$$

то можно утверждать, что число экстремальных точек должно делиться элементами симметрии поля решетки, т. е. потенциальной энергии. К этому можно прийти, применяя преобразование симметрии к гамильтониану. В кубической решетке, имеющей 24 элемента симметрии, экстремальных эквивалентных точек может быть в общем случае 24.

Разложим $E(\mathbf{k})$ в ряд Тейлора относительно одной из экстремальных точек \mathbf{k}_0 :

$$E(\mathbf{k}) = \sum_{l=0}^{\infty} \frac{d^l E(\mathbf{k})}{d\mathbf{k}^l} \Big|_{\mathbf{k}_0} \frac{(\mathbf{k} - \mathbf{k}_0)^l}{l!} = E(\mathbf{k}_0) + \frac{dE(\mathbf{k}_0)}{d\mathbf{k}} (\mathbf{k} - \mathbf{k}_0) + \dots \quad (12.3)$$

Дифференцирование по векторному аргументу $-d/d\mathbf{k}$ означает совокупность трех величин, получающихся при дифференцировании соответственно по k_x, k_y, k_z , поэтому записанная кратко в произведении производная представляет собой сложное выражение, называемое (ковариантным) тензором l -ранга.

При $l=0$ получаем скаляр, при $l=1$ имеем тензор первого ранга, представляющий собой вектор; при $l=2, 3$ и т. д. имеем тензор 2-го, 3-го и т. д. ранга. Выпишем только два первых члена $-l=1; l=2$:

$$\frac{dE}{d\mathbf{k}} = \left(\frac{\partial E}{\partial k_x}, \frac{\partial E}{\partial k_y}, \frac{\partial E}{\partial k_z} \right); \quad (12.4)$$

$$\begin{aligned} \frac{d^2 E}{d\mathbf{k}^2} &= \left(\frac{d}{d\mathbf{k}} \frac{\partial E}{\partial k_x}, \frac{d}{d\mathbf{k}} \frac{\partial E}{\partial k_y}, \frac{d}{d\mathbf{k}} \frac{\partial E}{\partial k_z} \right) = \\ &= \begin{bmatrix} \frac{\partial^2 E}{\partial k_x^2}; & \frac{\partial^2 E}{\partial k_y \partial k_x}; & \frac{\partial^2 E}{\partial k_z \partial k_x} \\ \frac{\partial^2 E}{\partial k_x \partial k_y}; & \frac{\partial^2 E}{\partial k_y^2}; & \frac{\partial^2 E}{\partial k_z \partial k_y} \\ \frac{\partial^2 E}{\partial k_x \partial k_z}; & \frac{\partial^2 E}{\partial k_y \partial k_z}; & \frac{\partial^2 E}{\partial k_z^2} \end{bmatrix}, \end{aligned} \quad (12.5)$$

т. е. $d^2 E/(d\mathbf{k})^2$ представляет собой величину из девяти частных производных энергии по волновому вектору второго порядка. Величины $\partial^2 E/(\partial k_i \partial k_j)$ называют компонентами, или элементами, тензора. Величина смешанной производной не зависит от порядка дифференцирования

$$\partial^2 E/(\partial k_i \partial k_j) = \partial^2 E/(\partial k_j \partial k_i); \quad (12.6)$$

такой тензор называют симметричным, члены вида $\partial^2 E / (\partial \kappa_i^2)$ называют диагональными. Производная l -порядка образует тензор l -ранга с числом элементов 3^l .

Рассмотрим малую окрестность точки κ_0 . В этом случае можно ограничиться первыми членами ряда. Так как разложение в ряд Тейлора производилось относительно экстремальной точки, то $\left. \frac{dE}{d\kappa} \right|_{\kappa_0} = 0$ (условие экстремума), поэтому ряд начинается с квадратичных членов:

$$E(\kappa) = E_0 + \frac{1}{2} \frac{d^2 E}{d\kappa^2} (\kappa - \kappa_0)^2 + \dots = E_0 + \frac{1}{2} \sum_{i,j} \frac{\partial^2 E}{\partial \kappa_i \partial \kappa_j} (\kappa_i - \kappa_{0i}) (\kappa_j - \kappa_{0j}) + \dots \quad (12.7)$$

Мы видим, что *изоэнергетическая поверхность вблизи экстремума изображается с достаточной степенью точности поверхностью второго порядка*. Это будет тем точнее, чем ближе величина энергии к экстремальной величине E_0 . Условием этого является малость отброшенных членов по сравнению с первым, например,

$$\frac{1}{6} \frac{d^3 E}{d\kappa^3} (\kappa - \kappa_0)^3 \ll \frac{1}{2} \frac{d^2 E}{d\kappa^2} (\kappa - \kappa_0)^2. \quad (12.8)$$

Тензор второго ранга подходящим выбором осей координат можно привести к диагональному виду, т. е. недиагональные элементы тензора обращаются в этих осях в нули. Предположим, что мы нашли такие оси, что

$$\partial^2 E / (\partial \kappa_i \partial \kappa_j) = 0 \quad (\text{при } i \neq j). \quad (12.9)$$

В этом случае уравнение *изоэнергетической поверхности* имеет вид

$$E(\kappa) = E(\kappa_0) + \frac{1}{2} \sum_{i=1}^3 \frac{\partial^2 E}{\partial \kappa_i^2} (\kappa_i - \kappa_{0i})^2 = \text{const} = E, \quad (12.10)$$

или

$$E(\kappa) - E_0 = \frac{1}{2} \frac{\partial^2 E}{\partial \kappa_x^2} (\kappa_x - \kappa_{0x})^2 + \frac{1}{2} \frac{\partial^2 E}{\partial \kappa_y^2} (\kappa_y - \kappa_{0y})^2 + \frac{1}{2} \frac{\partial^2 E}{\partial \kappa_z^2} (\kappa_z - \kappa_{0z})^2 = \text{const}' \quad (12.11)$$

Так как разложение проводится относительно экстремальной точки, то знак у всех трех производных один и тот же — плюс в минимуме, и минус в максимуме, поэтому *изоэнергетическая поверхность представляет собой эллипсоид*.

Рассмотрим вид *изоэнергетических поверхностей* в пространстве квазиимпульсов. Очевидно, что

$$E(P) = E(P_0) + \frac{1}{2} \frac{d^2 E}{dP^2} (P - P_0)^2 + \frac{1}{6} \frac{d^3 E}{dP^3} (P - P_0)^3 + \dots, \quad (12.12)$$

так как $\mathbf{P}_0 = \hbar \mathbf{k}_0$ есть точка экстремума. Ограничиваясь достаточно малой окрестностью точки \mathbf{P}_0 , можем записать

$$E(\mathbf{P}) = E_0 + \frac{1}{2} \frac{d^2 E}{d\mathbf{P}^2} (\mathbf{P} - \mathbf{P}_0)^2 + \dots \quad (12.13)$$

Введем новое обозначение для $d^2 E / (d\mathbf{P}^2)$; положим

$$d^2 E / d\mathbf{P}^2 = \mathbf{m}^{*-1}. \quad (12.14)$$

Очевидно, что составляющими тензора \mathbf{m}^{*-1} являются

$$m_{ij}^{*-1} = \partial^2 E / \partial P_i \partial P_j. \quad (12.15)$$

Так как размерность квазиимпульса совпадает с размерностью импульса, то размерность компонентов тензора \mathbf{m}^{*-1} есть размерность обратной массы; другими словами, размерность $[m_{ij}^*] = = [\partial^2 E / (\partial P_i \partial P_j)]^{-1} = [M]$ — есть размерность массы.

Величина $m^{*-1} = d^2 E / d\mathbf{P}^2$ называется тензором обратной эффективной массы. Выражение для энергии с использованием тензора обратной эффективной массы может быть записано в виде

$$E(\mathbf{P}) = E_0 + \frac{1}{2m^*} (\mathbf{P} - \mathbf{P}_0)^2, \quad (12.16)$$

что по форме подобно выражению для кинетической энергии свободной частицы с импульсом $\mathbf{p} - \mathbf{p}_0$.

Обратим внимание на знак эффективной массы: эффективная масса положительна в минимуме и отрицательна в максимуме энергии. Выпишем выражение для \mathbf{m}^{*-1} в общем виде:

$$\mathbf{m}^{*-1} = \begin{pmatrix} m_{xx}^{-1} & m_{xy}^{-1} & m_{xz}^{-1} \\ m_{yx}^{-1} & m_{yy}^{-1} & m_{yz}^{-1} \\ m_{zx}^{-1} & m_{zy}^{-1} & m_{zz}^{-1} \end{pmatrix} = \begin{bmatrix} \frac{\partial^2 E}{\partial P_x^2} & \frac{\partial^2 E}{\partial P_x \partial P_y} & \frac{\partial^2 E}{\partial P_x \partial P_z} \\ \frac{\partial^2 E}{\partial P_y \partial P_x} & \frac{\partial^2 E}{\partial P_y^2} & \frac{\partial^2 E}{\partial P_y \partial P_z} \\ \frac{\partial^2 E}{\partial P_z \partial P_x} & \frac{\partial^2 E}{\partial P_z \partial P_y} & \frac{\partial^2 E}{\partial P_z^2} \end{bmatrix}. \quad (12.17)$$

Если тензор приведен к диагональному виду, то

$$\mathbf{m}^{*-1} = \frac{1}{m^*} = \begin{pmatrix} m_{xx}^{-1} & 0 & 0 \\ 0 & m_{yy}^{-1} & 0 \\ 0 & 0 & m_{zz}^{-1} \end{pmatrix} = \begin{pmatrix} m_1^{-1} & 0 & 0 \\ 0 & m_2^{-1} & 0 \\ 0 & 0 & m_3^{-1} \end{pmatrix}, \quad (12.18)$$

где через m_i^{-1} обозначено

$$m_i^{-1} = \partial^2 E / (\partial P_i^2) = m_{ii}^{-1}. \quad (12.19)$$

Введем тензор, обратный к тензору обратной эффективной массы

$$\{\mathbf{m}^{*-1}\}^{-1} = \mathbf{m}^*. \quad (12.20)$$

Исходя из его размерности, назовем его *тензором эффективной массы*. Его компоненты m_{ij} не равны обратным величинам компонентов тензора обратной эффективной массы, т. е.

$$m_{ij} \neq \frac{1}{\frac{\partial^2 E}{\partial P_i \partial P_j}}. \quad (12.21)$$

Они должны находиться из условия

$$\mathbf{m}^* \mathbf{m}^{*-1} = \mathbf{m}^{*-1} \mathbf{m}^* = \mathbf{I}, \quad (12.22)$$

где \mathbf{I} есть единичный тензор. Рассмотрим частный случай, когда тензор \mathbf{m}^{*-1} имеет диагональный вид

$$\{\mathbf{m}^{*-1}\}_{ij} = m_i^{-1} \delta_{ij}. \quad (12.23)$$

Обозначим компоненты тензора \mathbf{m}^* в виде m_{ij} . Так как

$$I_{ij} = \delta_{ij} = \begin{cases} 1 & \text{при } i = j, \\ 0 & \text{при } i \neq j, \end{cases} \quad (12.24)$$

то на основании правил умножения матриц из (12.22) запишем, учитывая (12.24):

$$\sum_i m_{ii} \{\mathbf{m}^{*-1}\}_{ij} = \sum_i m_{ii} m_i^{-1} \delta_{ij} = m_{ij} m_i^{-1} = \delta_{ij}. \quad (12.25)$$

Мы видим, что

$$m_{ij} = \frac{1}{m_i^{-1}} \delta_{ij} = m_j \delta_{ij}, \quad (12.26)$$

т. е. если \mathbf{m}^{*-1} является диагональным тензором, то тензор \mathbf{m} также является диагональным, причем по диагонали его стоят элементы, обратные элементам тензора обратной эффективной массы. Запишем это подробно. Если обозначим

$$m_{ii}^{-1} = \partial^2 E (\partial P_i^2) = m_i^{-1}, \quad (12.27)$$

то тензоры \mathbf{m}^* и \mathbf{m}^{*-1} будут иметь вид

$$\mathbf{m}^{*-1} = \begin{bmatrix} \frac{1}{m_1} & 0 & 0 \\ 0 & \frac{1}{m_2} & 0 \\ 0 & 0 & \frac{1}{m_3} \end{bmatrix}; \quad \mathbf{m}^* = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}. \quad (12.28)$$

В дальнейшем величины m_i будем называть компонентами эффективной массы, а деление на \mathbf{m}^* будем понимать как умножение на \mathbf{m}^{*-1} (и наоборот).

Изоэнергетическая поверхность описывается уравнением

$$E(P) = E_0 + \frac{(P_x - P_{0x})^2}{2m_1} + \frac{(P_y - P_{0y})^2}{2m_2} + \frac{(P_z - P_{0z})^2}{2m_3} = \text{const}. \quad (12.29)$$

Если ввести полуоси эллипсоида a, b, c и записать уравнение эллипсоида в канонической форме

$$\frac{(P_x - P_{0x})^2}{a^2} + \frac{(P_y - P_{0y})^2}{b^2} + \frac{(P_z - P_{0z})^2}{c^2} = 1, \quad (12.30)$$

то легко видеть, что длина полуосей пропорциональна корню квадратному из соответствующих величин m_i :

$$a^2 = 2(E - E_0) m_1; \quad b^2 = 2(E - E_0) m_2; \quad c^2 = 2(E - E_0) m_3. \quad (12.31)$$

Рассмотрим случай, когда все три элемента эффективной массы равны: $m_1 = m_2 = m_3 = m^*$.

В этом случае тензор эффективной массы вырождается в скаляр (тензор нулевого ранга), и изоэнергетические поверхности представляют собой сферы:

$$\epsilon \quad E = E_0 + \frac{(P - P_0)^2}{2m^*} = \text{const}. \quad (12.32)$$

Если равны какие-либо два элемента тензора, например $m_1 = m_2 \neq m_3$, то

$$E = E_0 + \frac{(P_x - P_{0x})^2 + (P_y - P_{0y})^2}{2m_1} + \frac{(P_z - P_{0z})^2}{2m_3} = \text{const} \quad (12.33)$$

и изоэнергетические поверхности представляют собой эллипсоиды вращения вокруг оси P_z . Если $m_1 < m_3$, то эллипсоид вращения будет вытянут вдоль оси вращения, причем он вытянут тем больше, чем больше отношение эффективных масс m_3/m_1 . Если $m_1 > m_3$, то эллипсоид сжат вдоль оси вращения.

В общем случае различных эффективных масс m_i эллипсоид имеет три различных полуоси. При этом необходимо помнить, что чем сильнее различаются величины эффективных масс m_1, m_2, m_3 , тем сильнее «вытянут» эллипсоид.

С ростом разности энергии $E(\mathbf{k}) - E_0$ необходимо учитывать последующие члены ряда Тейлора; эллипсоиды энергии будут деформироваться, превращаясь в более сложные поверхности энергии.

§ 13. СВЯЗЬ СКОРОСТИ С КВАЗИИМПУЛЬСОМ

Оператор скорости \hat{v} определяется с помощью квантовых скобок Пуассона

$$\hat{v} = d\hat{r}/dt = [\hat{H}, \hat{r}] = \frac{1}{i\hbar} (\hat{r}\hat{H} - \hat{H}\hat{r}), \quad (13.1)$$

где \hat{r} — оператор координаты и \hat{H} — оператор Гамильтона. Для вычисления коммутатора \hat{r} и \hat{H} удобно перейти к E - или \mathbf{k} -представлению, когда все операторы задаются в виде определенных действий над переменной \mathbf{k} , от которой могут зависеть функции. Оператор

Гамильтона в κ - или E -представлении есть оператор умножения, или просто энергия $E(\kappa)$:

$$\hat{H}(\kappa) \equiv E(\kappa). \quad (13.2)$$

Чтобы найти вид оператора $\hat{r}(\kappa)$, учтем, что $\psi_{\kappa}(\mathbf{r})$ тесно связана с собственной функцией оператора $\hat{r}(\kappa)$, которую обозначим через $\psi_r(\kappa)$:

$$\hat{r}(\kappa) \psi_r(\kappa) = r \psi_r(\kappa), \quad (13.3)$$

где r есть собственное значение оператора координаты, $\psi_r(\kappa)$ его собственная функция, заданная в κ -представлении.

Как известно из квантовой механики, собственные функции двух операторов во взаимных представлениях связаны простым соотношением: если собственная функция \hat{L} в M -представлении есть $\psi_L(M)$

$$\hat{L}(M) \psi_L(M) = L \psi_L(M), \quad (13.4)$$

а собственная функция \hat{M} в L -представлении есть $\psi_M(L)$

$$\hat{M}(L) \psi_M(L) = M \psi_M(L), \quad (13.5)$$

то

$$\psi_L(M) = \psi_M^*(L). \quad (13.6)$$

На основании этого можно записать собственную функцию оператора $\hat{r}(\kappa)$ в κ -представлении в виде:

$$\psi_r(\kappa) = \psi_{\kappa}^*(\mathbf{r}) = e^{-i(\kappa r)} \varphi_{\kappa}^*(\mathbf{r}), \quad (13.7)$$

после чего легко найти вид оператора $\hat{r}(\kappa)$ на основании уравнения на собственные функции и собственные значения:

$$\hat{r}(\kappa) \psi_r(\kappa) = r \psi_r(\kappa), \quad (13.8)$$

или

$$\hat{r}(\kappa) \psi_{\kappa}^*(\mathbf{r}) = r \psi_{\kappa}^*(\mathbf{r}). \quad (13.9)$$

Смысл этого соотношения состоит в том, что нужно подобрать такую форму оператора $\hat{r}(\kappa)$, чтобы после его воздействия на функцию $\psi_{\kappa}^*(\mathbf{r})$ получить ту же функцию, умноженную на r . При этом необходимо действовать на $\psi_{\kappa}^*(\mathbf{r})$ только по переменной κ .

Рассмотрим действие оператора $d/(d\kappa) = \nabla_{\kappa}$ на функцию $\psi_{\kappa}^*(\mathbf{r})$:

$$\begin{aligned} \nabla_{\kappa} \psi_{\kappa}^*(\mathbf{r}) &= \nabla_{\kappa} [e^{-i(\kappa r)} \varphi_{\kappa}^*(\mathbf{r})] = -i r \psi_{\kappa}^*(\mathbf{r}) + e^{-i(\kappa r)} \nabla_{\kappa} \varphi_{\kappa}^*(\mathbf{r}) = \\ &= -i r \psi_{\kappa}^*(\mathbf{r}) + \lambda_{\kappa}^*(\mathbf{r}) [\nabla_{\kappa} \ln \varphi_{\kappa}^*(\mathbf{r})], \end{aligned} \quad (13.10)$$

или

$$r \psi_{\kappa}^*(\mathbf{r}) \equiv [i \nabla_{\kappa} - (i \nabla_{\kappa} \ln \varphi_{\kappa}^*)] \psi_{\kappa}^*(\mathbf{r}) \equiv \hat{r}(\kappa) \psi_{\kappa}(\mathbf{r}), \quad (13.11)$$

т. е. оператор

$$\hat{r}(\kappa) = i \nabla_{\kappa} - i (\nabla_{\kappa} \ln \varphi_{\kappa}^*) \quad (13.12)$$

представляется в виде суммы операторов дифференцирования по волновому вектору (или квазиимпульсу) и умножения на некоторую

функцию от \mathbf{k} (и \mathbf{r}). Вместо члена $-i(\nabla_{\mathbf{k}} \ln \varphi_{\mathbf{k}}^*)$ можно ввести некоторый оператор $\hat{\Omega}$, разложив функции $\nabla_{\mathbf{k}} \varphi_{\mathbf{k}}^*$ по функциям $\varphi_{\mathbf{k}}$.

При $U(\mathbf{r}) = \text{const}$ второй член обращается в нуль, и $\hat{\mathbf{r}}(\mathbf{k})$ имеет обычный вид оператора координаты в \mathbf{p} -представлении, поскольку в этом случае квазиимпульс тождествен импульсу. Теперь можно записать выражение для оператора скорости в \mathbf{k} -представлении:

$$\hat{\mathbf{v}}(\mathbf{k}) = \frac{1}{i\hbar} \{ \hat{\mathbf{r}}(\mathbf{k}) E(\mathbf{k}) - E(\mathbf{k}) \hat{\mathbf{r}}(\mathbf{k}) \} = \frac{1}{\hbar} \frac{dE(\mathbf{k})}{d\mathbf{k}}, \quad (13.13)$$

который имеет вид оператора умножения на производную энергии по квазиимпульсу:

$$\hat{\mathbf{v}}(\mathbf{k}) = \frac{1}{\hbar} \frac{dE(\mathbf{k})}{d\mathbf{k}} = \frac{dE}{d\mathbf{P}} = \hat{\mathbf{v}}(\mathbf{P}) = \mathbf{v}. \quad (13.14)$$

Это соотношение подобно выражению для групповой скорости волнового пакета

$$\mathbf{v}_{\text{гр}} = dE/d\mathbf{p}. \quad (13.15)$$

Средняя скорость электрона в состоянии с энергией $E(\mathbf{k})$ (с волновой функцией $\psi_{\mathbf{k}}(\mathbf{r})$, но не $\psi_{\mathbf{k}}^*(\mathbf{r})$) имеет вполне определенное значение, зависящее от этого состояния и равное $\langle \mathbf{v} \rangle = \mathbf{v} = \frac{1}{\hbar} \frac{dE}{d\mathbf{k}} = \frac{dE}{d\mathbf{P}}$ (опустим знак $\langle \rangle$). (13.16)

Таким образом, усредненная по состоянию с определенной энергией (точнее: и по бесконечно малому интервалу энергий) скорость электрона определяется в виде производной энергии по квазиимпульсу. В точках экстремума средняя в квантовомеханическом смысле скорость равна нулю. (В дальнейшем слова «средняя в квантовомеханическом смысле» будем опускать.)

Если рассматривать состояния в малой окрестности экстремальных точек, где энергия имеет квадратичную зависимость от квазиимпульса:

$$E - E_0 = \frac{1}{2m^*} (\mathbf{P} - \mathbf{P}_0)^2 = \frac{1}{2} \sum_{i,j=1}^3 \frac{(P_i - P_{0i})(P_j - P_{0j})}{m_{ij}}, \quad (13.17)$$

то

$$v_i = \frac{\partial E}{\partial P_i} = \sum_{j=1}^3 \frac{P_j - P_{0j}}{m_{ij}}, \quad (13.18)$$

или в векторной форме

$$\mathbf{v} = \frac{1}{m^*} (\mathbf{P} - \mathbf{P}_0), \quad (13.19)$$

т. е. в общем случае скорость равна скалярному произведению квазиимпульса на тензор обратной эффективной массы. Если тензор m^{*-1} имеет диагональный вид

$$\{m^{*-1}\}_{ij} = m_i^{-1} \delta_{ij}, \quad (13.20)$$

то выражение (13.18) упрощается:

$$v_i = \frac{P_i - P_{0i}}{m_i} \quad (13.21)$$

Скорость есть градиент энергии в пространстве квазимпульса, поэтому она направлена по нормали к изоэнергетическим поверх-

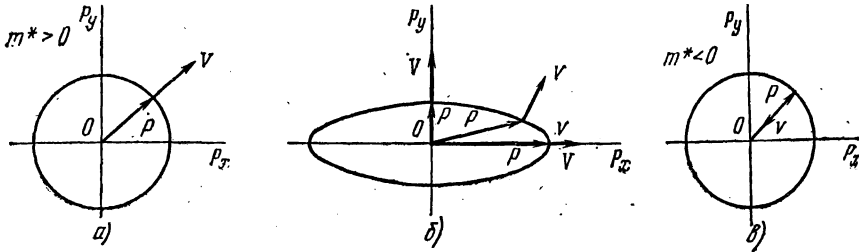


Рис. 9. Направления радиуса-вектора и нормали к поверхности энергии: а — сферические поверхности энергии; б — эллипсоидальные поверхности энергии; в — сферические поверхности энергии, $m^* < 0$

ностям, в то время как $\mathbf{P} - \mathbf{P}_0$ есть радиус-вектор точек поверхности. Для эллипсоидальных поверхностей энергии радиус-вектор и нормаль не коллинеарны, поэтому направления скорости и квазимпульса не совпадают. Коллинеарность будет наблюдаться только для направлений вдоль осей эллипсоидов (рис. 9), при этом

$$P_i = P_{0i} = \sqrt{2m_i(E - E_0)}, \quad (13.22)$$

поэтому

$$v_i = \frac{\sqrt{2(E - E_0)}}{\sqrt{m_i}} \quad (13.23)$$

Таким образом, при одной и той же энергии скорость вдоль осей эллипсоида обратно пропорциональна корню соответствующего компонента эффективной массы. Вводя в рассмотрение оси эллипсоида a_i :

$$a_i = \sqrt{2m_i(E - E_0)}, \quad (13.24)$$

получим

$$v_i = \frac{a_i}{m_i} = \frac{\sqrt{2(E - E_0)}}{\sqrt{m_i}}, \quad (13.25)$$

т. е., чем сильнее вытянут эллипсоид, тем меньше скорость движения в этом направлении. Это становится очевидным, если построить семейство изоэнергетических поверхностей (рис. 10). Чем меньше величина массы в данном направлении, тем гуще расположены изоэнергетические поверхности и, следовательно, тем больше скорость вдоль этой оси.

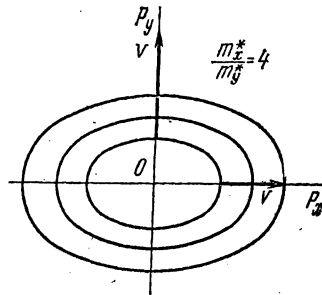


Рис. 10. Связь скорости с величиной эффективной массы и густотой поверхностей энергии

Необходимо остановиться на одном существенном моменте, связанном со знаком эффективной массы. Предположим для простоты, что эффективная масса — скалярная величина. В этом случае векторы \mathbf{v} и $(\mathbf{P} - \mathbf{P}_0)$ коллинеарные, однако их направления зависят от вида экстремума. Для минимума $m^* > 0$ и скорость \mathbf{v} совпадает по направлению с $(\mathbf{P} - \mathbf{P}_0)$. Для максимума энергии $m^* < 0$ и скорость направлена против $(\mathbf{P} - \mathbf{P}_0)$ (см. рис. 9, в).

§ 14. ОПЕРАТОР УСКОРЕНИЯ

Оператор ускорения \hat{a} определяется как производная от оператора скорости по времени:

$$\hat{a} = d\hat{v}/dt = [\hat{H}, \mathbf{v}] = \frac{1}{i\hbar} (\hat{v}\hat{H} - \hat{H}\hat{v}). \quad (14.1)$$

Легко видеть, что $\hat{a} = 0$ при движении электрона в периодическом поле решетки. Действительно, в \mathbf{k} -представлении $\hat{H}(\mathbf{k}) \equiv E(\mathbf{k})$ и $\hat{v}(\mathbf{k}) = dE(\mathbf{k})/d\mathbf{k}$, поэтому

$$\hat{v}\hat{H} - \hat{H}\hat{v} = 0. \quad (14.2)$$

Но так как коммутатор двух операторов не зависит от вида представления, то $\hat{a} \equiv 0$ и

$$\langle \hat{a} \rangle = \langle a \rangle = 0. \quad (14.3)$$

Это значит, что электрон движется в периодическом поле без ускорения. Интуитивное представление о движении электрона как о периодически повторяющемся ускорении и замедлении лишено основания. Причиной этого является наличие (наряду с корпускулярными) волновых свойств у электрона. Это приводит к тому, что его средняя скорость является интегралом движения, подобно тому, как сохраняется скорость свободно движущейся частицы.

Напомним, что одновременно с этим сохраняются во времени энергия и квазиимпульс. Импульс же частицы периодически меняется:

$$dp/dt = -\nabla U(\mathbf{r}) = \mathbf{F}_i, \quad (14.4)$$

так как сила \mathbf{F}_i , действующая на электрон со стороны решетки, имеет период решетки

$$\mathbf{F}_i(\mathbf{r} + \mathbf{n}) = \mathbf{F}_i(\mathbf{r}). \quad (14.5)$$

Пусть теперь на кристалл наложено некоторое «внешнее» поле с потенциальной энергией $V(\mathbf{r})$.

В этом случае, как мы знаем, квазиимпульс начинает изменяться

$$d\mathbf{P}/dt = \mathbf{F}_a = -\nabla V(\mathbf{r}), \quad (14.6)$$

если $V(\mathbf{r} + \mathbf{n}) \neq V(\mathbf{r})$, т. е. поле не является периодическим. Квазиимпульс меняется под действием внешней силы \mathbf{F}_a , обусловленной

любым нарушением периодического поля. Точка в k - (или P)-пространстве, изображающая состояние, начинает перемещаться в соответствии с уравнением

$$dP/dt = F_a. \quad (14.7)$$

Из (14.7) можем записать

$$P(t) = P_0 + \int_0^t F_a(\xi) d\xi. \quad (14.8)$$

Если внешняя сила не зависит от времени, то

$$P(t) = P_0 + F_a t. \quad (14.9)$$

Траекторией частицы в P -пространстве является прямая линия, задаваемая направлением внешней силы F_a . Но если частица начинает двигаться в пространстве квазиимпульса, то она переходит от одной поверхности энергии к другой. Другими словами, *внешняя сила F_a меняет не только квазиимпульс, но и энергию частицы.* К вопросу об изменении энергии можно подойти, рассматривая уравнение Шредингера, что будет проведено ниже.

Вернемся к оператору ускорения. Для этого вычислим скобку Пуассона, учитывая, что на кристалл наложено «внешнее» потенциальное поле $V(\mathbf{r})$. Обозначим гамильтониан поля решетки \hat{H}_0 :

$$\hat{H}_0 = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}), \quad (14.10)$$

для полного гамильтониана имеем выражение

$$\hat{H} = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) + V(\mathbf{r}) = \hat{H}_0 + V(\mathbf{r}). \quad (14.11)$$

Так как оператор скорости \hat{v} коммутирует с \hat{H}_0 , то для оператора ускорения получим

$$\hat{a} = d\hat{v}/dt = (1/i\hbar) \{ \hat{v} (\hat{H}_0 + \hat{V}) - (\hat{H}_0 + \hat{V}) \hat{v} \} = \frac{1}{i\hbar} (\hat{v}\hat{V} - \hat{V}\hat{v}). \quad (14.12)$$

Удобно выразить \hat{a} в k -представлении. Учитывая, что

$$\mathbf{v}(k) = dE/\hbar dk, \quad (14.13)$$

запишем

$$\hat{a}(k) = \frac{1}{i\hbar^2} \left\{ \frac{dE}{dk} \hat{V}(k) - \hat{V}(k) \frac{dE}{dk} \right\}. \quad (14.14)$$

Чтобы вычислить из (14.14) $\hat{a}(k)$, надо задать потенциальную энергию в k -представлении. Для этого в потенциальную энергию, заданную в виде функции от координаты $-V(\mathbf{r})$, необходимо подставить \mathbf{r} в k -представлении, т. е. вместо \mathbf{r} надо подставить оператор

$$\hat{\mathbf{r}}(k) = i \nabla_k - i (\nabla_k \ln \varphi_k^*). \quad (14.15)$$

Рассмотрим наиболее важный практически случай однородного силового поля

$$V(\mathbf{r}) = -(\mathbf{F}_a \mathbf{r}). \quad (14.16)$$

В κ -представлении оно будет иметь вид суммы оператора дифференцирования по κ и оператора умножения:

$$\hat{V}(\kappa) = -i(\mathbf{F}_a \nabla_\kappa) + i(\mathbf{F}_a \nabla_\kappa \ln \varphi_\kappa^*). \quad (14.17)$$

Так как второй член в (14.17) коммутирует с $\frac{dE}{\hbar d\kappa}$, то

$$\hat{\mathbf{a}}(\kappa) = -\frac{i}{\hbar^2} \left\{ \frac{dE}{d\kappa} (-i\mathbf{F}_a \nabla_\kappa) - (i\mathbf{F}_a \nabla_\kappa) \frac{dE}{d\kappa} \right\} = (1/\hbar^2) (\mathbf{F}_a \nabla_\kappa^2 E) = \mathbf{F}_a/m^*. \quad (14.18)$$

Величина

$$\frac{1}{m^*} = \frac{d^2 E}{dP^2} = \frac{1}{\hbar^2} \frac{d^2 E}{d\kappa^2} \quad (14.19)$$

может рассматриваться как обобщенный тензор обратной эффективной массы, который при квадратичной зависимости энергии от квазиимпульса совпадает с введенным ранее тензором \mathbf{m}^{*-1} .

Учитывая (14.19), запишем

$$\hat{\mathbf{a}}(\kappa) = \mathbf{F}_a/m^*(\kappa). \quad (14.20)$$

Так как \mathbf{F}_a не зависит от κ , то при не зависящей от κ эффективной массе $\hat{\mathbf{a}}$ будет иметь один и тот же вид в любом представлении:

$$\hat{\mathbf{a}} = \hat{\mathbf{F}}_a/m^*; \quad \mathbf{a} = \mathbf{F}_a/m^*, \quad (14.21)$$

или

$$\mathbf{m}^* \hat{\mathbf{a}} = \hat{\mathbf{F}}_a; \quad \mathbf{m}^* \mathbf{a} = \mathbf{F}_a. \quad (14.22)$$

Уравнение (14.21) или (14.22) по форме совпадает с уравнением движения Ньютона. Однако уравнение (14.21) или (14.22) имеет ряд особенностей, на которые необходимо обратить внимание.

Во-первых, \mathbf{m}^* является в общем случае тензором второго ранга, поэтому вектор ускорения не совпадает по направлению с вектором силы. Уравнение (14.21) имеет вид, например, для a_x :

$$a_x = \frac{1}{m_{xx}} F_x + \frac{1}{m_{xy}} F_y + \frac{1}{m_{xz}} F_z = \sum_{j=1}^3 \frac{F_j}{m_{xj}}. \quad (14.23)$$

Если \mathbf{m}^{*-1} имеет диагональный вид, то

$$a_j = F_j/m_j. \quad (14.24)$$

Уравнение (14.24) показывает, что ускорение будет коллинеарно силе только в том случае, когда сила направлена по одной из осей эллипсоида энергии.

Во-вторых, ускорение сообщается электрону только внешней силой F_a , внутренние силы F_i никакого ускорения электрону не сообщают.

В-третьих, динамической характеристикой электрона, определяющей его реакцию на действие внешних сил, является эффективная масса m^* , а не обычная масса. Это означает, что несмотря на то, что поле решетки не ускоряет электрон, оно оказывает воздействие на изменение его движения под действием внешних полей. Другими словами, при действии внешних сил поле решетки проявляет себя тем, что динамические свойства электрона определяются не обычной его массой, а массой эффективной.

Для скалярной эффективной массы a и F_a коллинеарны. Но отличие электрона в решетке от свободной частицы заключается не только в отличие величины эффективной массы от массы свободной частицы. В том случае, когда частица находится в окрестности максимума энергии, эффективная масса отрицательна, и ускорение a направлено против внешней силы F_a :

$$a = F_a/m^* = -|F_a|/|m^*|. \quad (14.25)$$

Рассмотрим силу, вызываемую однородным электрическим полем

$$F_a = e_n E, \quad (14.26)$$

где e_n — заряд электрона ($e_n < 0$). В этом случае

$$a = \frac{e_n}{m^*} E = \frac{-|e_n|}{-|m^*|} E = \frac{|e_n|}{|m^*|} E. \quad (14.27)$$

Ускорение, которое испытывает электрон в электрическом поле, направлено по полю, т. е. так, как если бы электрон имел положительный заряд и положительную эффективную массу. Такой аномально ускоряемый электрон будем называть квантовомеханической дыркой.

Если же электрон находится в окрестности минимума энергии, то $m^* > 0$, и ускорение направлено по силе, т. е. против поля E , как и для обычного электрона.

Выражение (14.22) можно получить из (13.19). Действительно, дифференцируя (13.19) по времени, получим

$$\frac{dv}{dt} = \frac{d}{dt} \frac{P - P_0}{m^*} = \frac{1}{m^*} \frac{dP}{dt} = \frac{1}{m^*} F_a = a. \quad (14.28)$$

Выражение (14.28) совпадает по форме с (14.21), однако оно более общее, чем (14.21), где F_a предполагается независимой от координат, в то время как в (14.28) F_a может быть функцией координат. Но с другой стороны m^{*-1} в (14.21) является обобщенным тензором обратной эффективной массы, а в (14.28) m^{*-1} не зависит от k . Другими словами, уравнение (13.19) и, следовательно, (14.28) справедливо до тех пор, пока состояние $P - P_0$ находится в области квадратичной зависимости энергии от квазиимпульса; если же состояние перемещается в область, где энергия является более сложной функцией квазиимпульса, то в (14.28) необходимо добавить производную от m^{*-1} по времени.

В заключение этого параграфа запишем выражение для среднего (в квантовомеханическом смысле) значения импульса в стационарном состоянии, описываемого функцией Блоха $\psi_{\mathbf{k}}(\mathbf{r})$:

$$\langle \mathbf{p} \rangle = \int \psi_{\mathbf{k}}^*(\mathbf{r}) \hat{\mathbf{p}} \psi_{\mathbf{k}}(\mathbf{r}) d\tau = m \int \psi_{\mathbf{k}}^* \frac{\hat{\mathbf{p}}}{m} \psi_{\mathbf{k}} d\tau = m \langle \mathbf{v} \rangle. \quad (14.29)$$

Но, согласно (13.16) и (13.19),

$$\langle \mathbf{v} \rangle = \mathbf{v} = \frac{\mathbf{P} - \mathbf{P}_0}{m^*}, \quad (14.30)$$

поэтому

$$\langle \mathbf{p} \rangle = \frac{m}{m^*} (\mathbf{P} - \mathbf{P}_0); \quad \mathbf{P} - \mathbf{P}_0 \equiv \frac{m^*}{m} \langle \mathbf{p} \rangle, \quad (14.31)$$

или

$$P_i - P_{0i} = \frac{m_i}{m} \langle p_i \rangle. \quad (14.32)$$

Таким образом, понимая под $\langle \mathbf{p} \rangle$ «классический» импульс электрона в решетке, можно выразить квазиимпульс через «классический» импульс и эффективную массу.

Сравним величину кинетической энергии свободной частицы, имеющей импульс $\langle \mathbf{p} \rangle$, с полной энергией электрона в решетке, имеющего тот же средний импульс, что и у свободной частицы:

$$T = \frac{\langle \mathbf{p} \rangle^2}{2m} = \frac{m v^2}{2} = \frac{m}{2} \left[\frac{(\mathbf{P} - \mathbf{P}_0)}{m^*} \right]^2. \quad (14.33)$$

Для скалярной эффективной массы $m_i = m^*$ и

$$T = \frac{m}{m^*} [E(\mathbf{P}) - E(\mathbf{P}_0)]. \quad (14.34)$$

Резюме § 12—14

1. Важнейшим понятием, используемым для описания движения электрона в твердом теле, является понятие эффективной массы m^* . Она определяется как тензор, обратный тензору обратной эффективной массы m^{*-1} . Тензор обратной эффективной массы m^{*-1} есть величина, равная второй производной от энергии по квазиимпульсу:

$$m^{*-1} = \frac{d^2 E}{d\mathbf{P}^2} = \frac{1}{\hbar^2} \frac{d^2 E}{d\mathbf{k}^2}, \quad (14.1p)$$

поэтому

$$m^* = \{d^2 E / d\mathbf{P}^2\}^{-1}. \quad (14.2p)$$

Если m^{*-1} имеет диагональный вид: $\{m^{*-1}\}_{ij} = m_i^{-1} \delta_{ij}$, то m^* также диагонален, причем

$$m_{ij} = m_i \delta_{ij}. \quad (14.3p)$$

Если производная от энергии по \mathbf{P} вычисляется в произвольной точке \mathbf{P} , то $\mathbf{m}^* = \mathbf{m}^*(\mathbf{P})$ и \mathbf{m}^* называется обобщенной эффективной массой.

2. В окрестности экстремума \mathbf{P}_0 энергия является квадратичной функцией от $\mathbf{P} - \mathbf{P}_0$ (или $\mathbf{k} - \mathbf{k}_0$), поэтому \mathbf{m}^{*-1} не зависит от \mathbf{P} и равен второй производной в точке экстремума, при этом

$$E(\mathbf{P}) = E(\mathbf{P}_0) + \frac{1}{2} \frac{(\mathbf{P} - \mathbf{P}_0)^2}{\mathbf{m}^*} = E_0 + \sum_{i=1}^3 \frac{(P_i - P_{0i})^2}{2m_i}. \quad (14.4p)$$

3. В окрестности минимума энергии компоненты $m_i > 0$, в окрестности максимума энергии $m_i < 0$.

4. В окрестности экстремума энергии изоэнергетические поверхности являются эллипсоидами, оси которых пропорциональны компонентам тензора эффективной массы в степени $1/2$. Если все компоненты одинаковы, то эффективная масса является скалярной величиной, а поверхности энергии — сферами. Если два компонента равны, например $m_1 = m_2 \neq m_3$, то поверхность энергии представляет собой эллипсоид вращения, ось вращения при этом соответствует m_3 , которая называется вследствие этого продольной эффективной массой и обозначается m_l , а величина $m_1 = m_2$ называется поперечной эффективной массой m_t .

5. Средняя (в квантовомеханическом смысле) скорость равна

$$\langle \mathbf{v} \rangle = \mathbf{v} = \frac{dE}{d\mathbf{P}} = \frac{1}{\hbar} \frac{dE}{d\mathbf{k}}, \quad (14.5p)$$

она связана с квазиимпульсом соотношением

$$\mathbf{v} = \frac{\mathbf{P} - \mathbf{P}_0}{\mathbf{m}^*}. \quad (14.6p)$$

6. Квазиимпульс связан со средним (в квантовомеханическом смысле) импульсом соотношением

$$\langle \mathbf{p} \rangle = \mathbf{m} \langle \mathbf{v} \rangle = \frac{m}{\mathbf{m}^*} (\mathbf{P} - \mathbf{P}_0). \quad (14.7p)$$

7. Электрон испытывает ускорение только под действием внешней силы \mathbf{F}_a , однако силы \mathbf{F}_l поля решетки проявляют себя тем, что ускорение определяется эффективной массой \mathbf{m}^* :

$$\mathbf{a} = \mathbf{F}_a / \mathbf{m}^* = \mathbf{m}^{*-1} \mathbf{F}_a; \quad \mathbf{F}_a = \mathbf{m}^* \mathbf{a}. \quad (14.8p)$$

8. В случае скалярной эффективной массы электрон движется против электрического поля, если он находится в окрестности минимума энергии. Если же он находится в окрестности максимума энергии, то его ускорение направлено против силы. Он движется в данном электрическом поле как частица с положительной эффективной массой и положительным зарядом.

9. При действии внешней силы F_a квазиимпульс электрона меняется:

$$\frac{dP}{dt} = F_a; \quad P(t) = P(0) + \int_0^t F_a(\xi) d\xi. \quad (14.9p)$$

Если сила F_a не зависит от времени, то траекторией электрона в пространстве квазиимпульса (и волнового вектора) является прямая линия, определяемая направлением силы F_a . Движение электрона в пространстве P означает, что энергия электрона меняется, и он переходит с одной поверхности энергии на другую в результате работы внешних сил.

10. Из (14.6p) и (14.5p) можно получить эффективную массу скорости, или квазиимпульса, часто используемую для описания различных физических явлений (при $P_0 = 0$):

$$\frac{1}{m^*} = \frac{v}{P} = \frac{1}{\hbar k} \frac{dE}{\hbar dk} = \frac{1}{\hbar^2 k} \frac{dE}{dk}. \quad (14.10p)$$

Для квадратичного закона дисперсии различные определения эффективной массы дают одну и ту же величину. Но если закон дисперсии отличается от квадратичного, то эффективные массы, определяемые различными соотношениями, являются разными величинами.

§ 15. ЗОНЫ БРИЛЛЮЭНА

Решетке каждого кристалла соответствует так называемая обратная решетка, которая определяется следующим образом.

Построим три вектора b_1, b_2, b_3 посредством уравнений:

$$b_1 = \frac{[a_2 a_3]}{(a_1 [a_2 a_3])}; \quad b_2 = \frac{[a_3 a_1]}{(a_1 [a_2 a_3])}; \quad b_3 = \frac{[a_1 a_2]}{(a_1 [a_2 a_3])}, \quad (15.1)$$

с помощью которых построим пространственную решетку. Векторы b_1, b_2, b_3 образуют базис этой решетки, называемой обратной к исходной решетке с базисом a_1, a_2, a_3 .

Объем базисной ячейки обратной решетки V_b связан простым соотношением с объемом базисной ячейки прямой решетки V_a :

$$V_a = (a_1 [a_2 a_3]); \quad V_b = (b_1 [b_2 b_3]); \quad V_b = 1/V_a. \quad (15.2)$$

Вектор

$$b = l_1 b_1 + l_2 b_2 + l_3 b_3 \quad (15.3)$$

определяет узлы обратной решетки, если l_1, l_2, l_3 — целые числа.

Из определения базиса обратной решетки следует, что

$$(a_i b_j) = (1/V_a) (a_i [a_s a_t]) = \delta_{ij}, \quad (15.4)$$

поскольку при $i = j$ s и t не могут совпадать с i , в силу чего скалярное произведение должно равняться единице; при $i \neq j$ один из

векторов a_s или a_r совпадает с a_i , поэтому смешанное произведение обращается в нуль, что очевидно из самого определения базиса обратной решетки.

Рассмотрим скалярное произведение вектора \mathbf{b} и произвольного вектора \mathbf{n} . Учитывая (15.4), получим

$$(\mathbf{nb}) = n_1 l_1 + n_2 l_2 + n_3 l_3 = Q, \quad (15.5)$$

где Q — целое число.

Рассмотрим вновь трансляционное условие, накладываемое на волновую функцию электрона, движущегося в поле кристалла:

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{n}) = e^{i(\mathbf{k}\mathbf{n})}\psi_{\mathbf{k}}(\mathbf{r}). \quad (15.6)$$

Если мы вместо вектора \mathbf{k} возьмем вектор $\mathbf{k}' = \mathbf{k} + 2\pi\mathbf{b}$, то трансляционное условие не нарушается:

$$e^{i(\mathbf{k}'\mathbf{n})} = e^{i(\mathbf{k} + 2\pi\mathbf{b}, \mathbf{n})} = e^{i(\mathbf{k}\mathbf{n})}e^{i2\pi(\mathbf{nb})} = e^{i(\mathbf{k}\mathbf{n})}, \quad (15.7)$$

поскольку

$$(\mathbf{nb}) = Q, \quad e^{i2\pi Q} = 1. \quad (15.8)$$

Но это означает, что *состояния, характеризуемые вектором \mathbf{k} и вектором $\mathbf{k} + 2\pi\mathbf{b}$ (или \mathbf{P} и $\mathbf{P} + 2\pi\mathbf{h}\mathbf{b}$ соответственно), физически эквивалентны, и энергия электронов, находящихся в этих двух состояниях, должна быть одной и той же. Другими словами, энергия является периодической функцией волнового вектора (или квазимпульса):*

$$E(\mathbf{k} + 2\pi\mathbf{b}) = E(\mathbf{k}), \quad (15.9)$$

или

$$E(\mathbf{P} + 2\pi\mathbf{h}\mathbf{b}) = E(\mathbf{P}). \quad (15.10)$$

Если в пространстве \mathbf{P} (или \mathbf{k}) построить обратную решетку с базисом $2\pi\mathbf{h}\mathbf{b}_1, 2\pi\mathbf{h}\mathbf{b}_2, 2\pi\mathbf{h}\mathbf{b}_3$ (или $2\pi\mathbf{b}_1, 2\pi\mathbf{b}_2, 2\pi\mathbf{b}_3$), то *все пространство \mathbf{P} (или \mathbf{k}) можно разбить на области, совокупность точек которых представляет собой физически эквивалентные состояния. Такие области называют зонами Бриллюэна. Первой, или основной, зоной называют минимальный по объему многогранник, построенный вокруг начала координат в пространстве \mathbf{P} (или \mathbf{k}), содержащий все возможные различные состояния.*

Предположим, что найден такой многогранник. Если ко всем точкам \mathbf{P} (или \mathbf{k}) выделенной области добавлять различные векторы $2\pi\mathbf{h}\mathbf{b}$ (или $2\pi\mathbf{b}$), то получим все точки пространства \mathbf{P} (или \mathbf{k}). Отсюда следует, что *любую точку пространства \mathbf{P} (или \mathbf{k}) можно перевести в основную зону Бриллюэна с помощью некоторого вектора обратной решетки.* В дальнейшем найдем уравнение, с помощью которого можно разбить пространство \mathbf{k} на зоны Бриллюэна наиболее удобным способом.

Укажем некоторые свойства обратной решетки, которые будут необходимы. Рассмотрим вектор \mathbf{b} с компонентами

$$\mathbf{b} = (l_1 b_1, 0, 0). \quad (15.11)$$

Он ортогонален плоскости, определяемой векторами \mathbf{a}_2 и \mathbf{a}_3 . Модуль вектора \mathbf{b} равен

$$|\mathbf{b}| = l_1 |\mathbf{b}_1| = l_1 \frac{|[\mathbf{a}_2 \mathbf{a}_3]|}{V_a}. \quad (15.12)$$

Для ортогонального базиса прямой решетки

$$|\mathbf{b}| = l_1 |\mathbf{b}_1| = l_1/a_1, \quad (15.13)$$

где a_1 есть расстояние между двумя соседними плоскостями семейства атомных плоскостей, определяемых векторами \mathbf{a}_2 и \mathbf{a}_3 .

Этот результат можно обобщить: вектор \mathbf{b} определяет семейство ортогональных ему атомных плоскостей кристалла, расстояние между которыми d связано с модулем вектора обратной решетки $|\mathbf{b}|$ соотношением

$$|\mathbf{b}| = l/d, \quad (15.14)$$

где l — некоторое целое число.

Для кубической решетки кристалла обратная решетка также кубическая. Если решетка кристалла имеет примитивную элементарную ячейку, то обратная решетка имеет примитивную элементарную ячейку. Для кристалла с простой кубической решеткой зона Бриллюэна в k -пространстве представляет собой куб объемом $\frac{8\pi^3}{a^3}$. Действительно, построим в пространстве k решетку с базисом $2\pi\mathbf{b}_1$, $2\pi\mathbf{b}_2$, $2\pi\mathbf{b}_3$. Куб, построенный на этих векторах, содержит неэквивалентные точки, поскольку они не могут быть получены одна из другой с помощью какого-либо вектора \mathbf{b} . Исключением являются точки, лежащие на гранях этого куба, которые получены друг из друга с помощью векторов $2\pi\mathbf{b}_i$ или $-2\pi\mathbf{b}_i$. Все точки, лежащие вне куба, могут быть получены из точек, лежащих внутри куба. Тем самым можно утверждать, что выделенный объем содержит все физически различные состояния, т. е. является зоной Бриллюэна. Чтобы построить основную зону Бриллюэна, необходимо сместить все точки на вектор $(-\pi\mathbf{b}_1, -\pi\mathbf{b}_2, -\pi\mathbf{b}_3)$, после чего центр куба совместится с началом отсчета $k=0$.

Таким образом, для вектора k можно указать следующий интервал значений:

$$\begin{aligned} -\frac{\pi}{a} &\leq k_x < \frac{\pi}{a}, \\ -\frac{\pi}{a} &\leq k_y < \frac{\pi}{a}, \\ -\frac{\pi}{a} &\leq k_z < \frac{\pi}{a}. \end{aligned} \quad (15.15)$$

Если решетка кристалла не является кубической, то для каждого компонента k_i можно записать

$$-\frac{\pi}{a_i} \leq k_i < \frac{\pi}{a_i}, \quad (15.16)$$

или, в общем виде,

$$-\frac{\pi}{a_i} + 2\pi l_i b_i \leq \kappa_i < \frac{\pi}{a_i} + 2\pi l_i b_i, \quad (15.17)$$

где l_i — произвольное целое число.

Из эквивалентности точек различных зон Бриллюэна следует, что при движении частицы в пространстве \mathbf{k} (или \mathbf{P}) можно рассматривать траекторию частицы только в пределах основной зоны Бриллюэна. Для этого необходимо переносить состояния из некоторой точки, лежащей на границе, в эквивалентную точку противоположной границы зоны Бриллюэна, как это показано на рис. 11.

Покажем теперь, что кубическая решетка с объемноцентрированной (или гранецентрированной) элементарной ячейкой имеет обратную кубическую решетку с гранецентрированной (или объемноцентрированной) элементарной ячейкой.

Объемноцентрированная кубическая решетка (ОЦК-решетка) имеет 8 узлов в вершинах куба и один узел в центре куба. Каждый узел в вершине принадлежит 8 ячейкам, поэтому на одну элементарную ячейку приходится $8 \cdot \frac{1}{8} + 1 = 2$ узла. Объем ячейки равен a^3 , на один узел имеем $a^3/2$. Координаты узлов в вершинах можно записать в виде $a(0, 0, 0)$; $a(1, 0, 0)$; $a(0, 1, 0)$; $a(0, 0, 1)$; $a(1, 1, 1)$; $a(0, 1, 1)$; $a(1, 0, 1)$ и $a(1, 1, 0)$ и в центре $a(1/2, 1/2, 1/2)$. Выберем три вектора вида

$$\begin{aligned} \mathbf{a}_1 &= (a/2)(\mathbf{i} + \mathbf{j} + \mathbf{k}), \\ \mathbf{a}_2 &= (a/2)(-\mathbf{i} + \mathbf{j} + \mathbf{k}), \\ \mathbf{a}_3 &= (a/2)(-\mathbf{i} - \mathbf{j} + \mathbf{k}), \end{aligned} \quad (15.18)$$

т. е. три вектора, идущие в центры трех соседних элементарных ячеек из начала координат, направления осей задаются единичными векторами \mathbf{i} , \mathbf{j} , \mathbf{k} , идущими вдоль ребер элементарной ячейки. Легко видеть, что вектор трансляции $\mathbf{p} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ позволяет получить любой узел решетки. Вычисляя объем базисного параллелепипеда V_a , получим после несложных преобразований $V_a = a^3/2$, т. е. объем, приходящийся на один узел решетки.

Для гранецентрированной кубической решетки с периодом a имеем 8 узлов в вершинах ячейки, 6 узлов в центрах граней, всего на одну элементарную ячейку приходится $(8 \cdot 1/8) + (6 \cdot 1/2) = 4$ узла и объем $a^3/4$. В качестве базисных векторов можно взять три век-

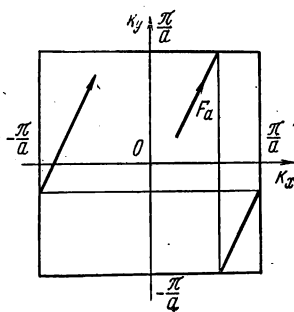


Рис. 11. Движение электрона в пределах основной зоны Бриллюэна

тора, выходящие из начального узла в узлы, лежащие в центрах граней, например:

$$\left. \begin{aligned} \mathbf{a}_1 &= (a/2) (\mathbf{i} + \mathbf{j}) \\ \mathbf{a}_2 &= (a/2) (\mathbf{j} + \mathbf{k}) \\ \mathbf{a}_3 &= (a/2) (\mathbf{k} + \mathbf{i}) \end{aligned} \right\}. \quad (15.19)$$

Объем базового параллелепипеда $V_a = (\mathbf{a}_1 [\mathbf{a}_2 \mathbf{a}_3]) = a^3/4$.

Построим теперь базис обратных решеток. Для ГЦК-решетки согласно (15.1) и (15.19) найдем:

$$\begin{aligned} \mathbf{b}_1 &= \frac{1}{V_a} [\bar{\mathbf{a}}_2 \bar{\mathbf{a}}_3] = \frac{1}{(a^3/4)} \frac{a^2}{4} [\mathbf{j} + \mathbf{k}, \mathbf{k} + \mathbf{i}] = \frac{1}{a} (\mathbf{i} - \mathbf{k} + \mathbf{j}) = \frac{b}{2} (\mathbf{i} + \mathbf{j} - \mathbf{k}); \\ \mathbf{b}_2 &= \frac{b}{2} (\mathbf{j} + \mathbf{k} - \mathbf{i}); \\ \mathbf{b}_3 &= \frac{b}{2} (\mathbf{k} + \mathbf{i} - \mathbf{j}). \end{aligned} \quad (15.19')$$

Таким образом, если положить период обратной решетки $b = \frac{2}{a}$, то обратную решетку получим типа ОЦК. Объем элементарной ячейки обратной решетки равен $8/a^3$.

Для ОЦК-решетки согласно (15.1) и (15.18) найдем:

$$\begin{aligned} \mathbf{b}_1 &= \frac{1}{V_a} [\mathbf{a}_2 \mathbf{a}_3] = \frac{1}{(a^3/2)} \frac{a^2}{4} [-\mathbf{i} + \mathbf{j} + \mathbf{k}, -\mathbf{i} - \mathbf{j} + \mathbf{k}] = \\ &= \frac{1}{a} (\mathbf{i} + \mathbf{k}) = \frac{b}{2} (\mathbf{i} + \mathbf{k}); \\ \mathbf{b}_2 &= \frac{b}{2} (-\mathbf{i} + \mathbf{j}) \\ \mathbf{b}_3 &= \frac{b}{2} (-\mathbf{j} + \mathbf{k}); \quad b = \frac{2}{a}. \end{aligned} \quad (15.20)$$

Как видно из (15.20) и (15.21), обратной для ОЦК-решетки является ГЦК-решетка, а для ГЦК-решетки — ОЦК-решетка, при этом их периоды связаны соотношением $a \cdot b = 2$ при $V_a V_b = 1$.

Для гексагональной решетки с $c/a > 1$ обратной является гексагональная решетка, сжатая вдоль гексагональной оси.

Из сказанного вытекает, что характер симметрии прямой и обратной решеток одинаков, поэтому следует ожидать, что симметрия зависимости $E(\mathbf{k})$ полностью определяется симметрией потенциального поля решетки $U(\mathbf{r})$.

§ 16. НОРМИРОВКА В ЯЩИКЕ И ДИСКРЕТНОСТЬ КВАЗИИМПУЛЬСА

При всех вычислениях, проводимых в квантовой механике, обычно предполагается, что волновая функция нормирована.

Нормировка собственных функций операторов производится по-разному в зависимости от вида спектра его собственных значе-

ний. Для дискретного спектра волновые функции нормируются к единице:

$$\int \psi_n^*(\mathbf{r}) \psi_m(\mathbf{r}) d\tau = \delta_{mn} = \begin{cases} 1 & \text{при } n = m, \\ 0 & \text{при } n \neq m, \end{cases} \quad (16.1)$$

где $\psi_n(\mathbf{r})$, $\psi_m(\mathbf{r})$ — собственные функции (в \mathbf{r} -представлении) некоторого оператора \hat{M} , соответствующие собственным значениям M_m и M_n :

$$\hat{M}\psi_n(\mathbf{r}) = M_n\psi_n(\mathbf{r}); \quad \hat{M}\psi_m(\mathbf{r}) = M_m\psi_m(\mathbf{r}). \quad (16.2)$$

Если оператор \hat{M} имеет непрерывный спектр собственных значений

$$\hat{M}\psi_M(\mathbf{r}) = M\psi_M(\mathbf{r}); \quad \hat{M}\psi_{M'}(\mathbf{r}) = M'\psi_{M'}(\mathbf{r}), \quad (16.3)$$

то собственные функции нормируются на δ -функцию Дирака

$$\int \psi_M^*(\mathbf{r}) \psi_{M'}(\mathbf{r}) d\tau = \delta(M - M'), \quad (16.4)$$

где δ -функция Дирака удовлетворяет известным условиям:

$$\delta(x - a) = \begin{cases} 0 & \text{при } x \neq a, \\ \infty & \text{при } x = a, \end{cases}$$

$$\int_b^c \delta(x - a) dx = \begin{cases} 1 & \text{при } a \in [b, c], \\ 0 & \text{при } a \notin [b, c]. \end{cases} \quad (16.5)$$

Интегрирование при нормировке производится по всему объему, в котором находится частица.

В реальных условиях частицы движутся в ограниченных областях пространства, например в кристалле конечного размера. В этом случае необходимо решить уравнение Шредингера для ограниченного кристалла и при этом надо задать граничные условия, т. е. значения для волновой функции и ее первых производных по координатам на границе кристалла. Это приводит к целому ряду дополнительных трудностей, а именно, нарушается трансляционная симметрия кристаллического поля, требуется вводить дополнительные граничные условия, учесть которые не столь просто, поскольку явления на поверхности кристалла, пожалуй, более сложны, чем явления в объеме. В то же время очевидно, что если размеры кристалла достаточно велики, то поверхностные явления не будут оказывать решающего влияния на процессы внутри объема кристалла. Это позволяет ввести так называемые *циклические граничные условия*. Предположим, что кристалл имеет вид параллелепипеда со сторонами L_1 , L_2 , L_3 с объемом $G = L_1L_2L_3$.

Предположим, что все пространство заполнено подобными кристаллами. В таком случае трансляционное свойство поля кристалла сохраняется. Так как все точки, отличающиеся на целое число L_1 , L_2 , L_3 , эквивалентны, то граничные условия в обычном смысле слова заменяются условием эквивалентности физических свойств кристалла

в точках x и $x + L_1$ и аналогичных точках, поэтому вместо обычных граничных условий задаются циклические граничные условия в виде:

$$\psi(x, y, z) = \psi(x + L_1, y, z) = \psi(x, y + L_2, z) = \psi(x, y, z + L_3). \quad (16.6)$$

В таком случае достаточно рассмотреть движение электрона в основной области G , поэтому и нормировку можно провести по области G :

$$\int_G \psi_M^*(\mathbf{r}) \psi_{M'}(\mathbf{r}) d\tau = 1. \quad (16.7)$$

Это так называемая «нормировка в ящике» приводит к некоторому изменению спектра величины M , которое рассмотрим на примере собственных значений квазиимпульса. Как видно, на квазиимпульс наложено одно условие — он должен быть вещественной величиной. Так как при этом он может быть непрерывным, то его собственные функции, т. е. функции Блоха, должны быть нормированы на δ -функцию:

$$\int \psi_k(\mathbf{r}) \psi_{k'}^*(\mathbf{r}) d\tau = \delta(k - k'). \quad (16.8)$$

При этом интеграл берется по всему пространству. Проводя нормировку в ящике, мы можем записать

$$\int \psi_k^*(\mathbf{r}) \psi_{k'}(\mathbf{r}) d\tau = \delta_{kk'} = \begin{cases} 1 & \text{при } k = k', \\ 0 & \text{при } k \neq k'. \end{cases} \quad (16.9)$$

Другими словами, записываем условие нормировки функции оператора с дискретным спектром. Но дискретность спектра непосредственно следует из циклических граничных условий:

$$\begin{aligned} \psi_k(x + L_1, y, z) &= e^{i[\kappa_x(x + L_1) + \kappa_y y + \kappa_z z]} \varphi_k(x + L_1, y, z) = \\ &= e^{i\kappa_x L_1} e^{i(\kappa \mathbf{r})} \varphi_k(x, y, z) = e^{i(\kappa \mathbf{r})} \varphi_k(\mathbf{r}). \end{aligned} \quad (16.10)$$

Условие $\varphi_k(x + L_1, y, z) = \varphi_k(x, y, z)$ выполняется автоматически, поскольку L_1, L_2, L_3 содержат целое число периодов решетки, а $\varphi_k(\mathbf{r})$ периодична с периодом прямой решетки. Из последнего равенства следует, что

$$e^{i\kappa_x L_1} = 1. \quad (16.11)$$

Но это возможно, если показатель мнимой экспоненты есть целое число, умноженное на 2π , т. е.

$$\kappa_x L_1 = 2\pi n_1, \quad (16.12)$$

где n_1 — произвольное целое число. Отсюда следует дискретность волнового вектора:

$$\kappa_x = \frac{2\pi}{L_1} n_1; \quad n_1 = 0; \pm 1; \pm 2; \dots \quad (16.13)$$

Учитывая условия периодичности по осям y и z , аналогично получим

$$\left. \begin{aligned} \kappa_y &= \frac{2\pi}{L_2} n_2; & n_2 &= 0; \pm 1; \pm 2; \dots \\ \kappa_z &= \frac{2\pi}{L_3} n_3; & n_3 &= 0; \pm 1; \pm 2 \dots \end{aligned} \right\} \quad (16.14)$$

Определим L_i через число узлов N_i решетки, лежащих на ребре кристалла:

$$L_i = N_i a_i; \quad (i = 1, 2, 3). \quad (16.15)$$

Подставляя L_i в выражение для κ_i , получим

$$\kappa_i = \frac{2\pi}{a_i} \frac{n_i}{N_i} = 2\pi b_i \frac{n_i}{N_i}. \quad (16.16)$$

Учитывая, что состояния с \mathbf{k} и $\mathbf{k}' = \mathbf{k} + 2\pi\mathbf{b}$ эквивалентны, мы можем ограничить верхнее значение n_i условием

$$\kappa_i = 2\pi b_i; \quad n_i = N_i, \quad (16.17)$$

а нижнее значение $n_i = 0$.

Можно выбрать и другое условие: $n_i/N_i = \pm 1/2$, что дает симметричный относительно $\mathbf{k} = 0$ интервал значений

$$|\kappa_i| \leq \frac{1}{2} 2\pi b_i = \frac{\pi}{a_i}, \quad (16.18)$$

или

$$-\frac{\pi}{a_i} \leq \kappa_i < \frac{\pi}{a_i}; \quad n_i = 0; \pm 1; \pm 2; \dots; \pm \frac{N_i - 1}{2}; -\frac{N_i}{2}. \quad (16.19)$$

Таким образом, для ограниченного кристалла волновой вектор дискретен. Однако эта дискретность для достаточно большого размера кристалла в большинстве случаев не существенна и будем считать κ_i величиной квазинепрерывной, но в ряде случаев дискретность необходимо учитывать. Так как квазиимпульс дискретен, то и энергия должна быть дискретна: $E(\kappa_i)$. Однако подобной дискретностью энергии в дальнейшем будем пренебрегать, поскольку расстояние между уровнями энергии много меньше kT .

Вернемся к условиям нормировки в «ящике» для волновой функции. Подставляя в (16.9) функцию Блоха

$$\int e^{-i(\mathbf{k}'\cdot\mathbf{r})} \varphi_{\mathbf{k}'}^*(\mathbf{r}) e^{i(\mathbf{k}\cdot\mathbf{r})} \varphi_{\mathbf{k}}(\mathbf{r}) d\tau = \delta_{\mathbf{k}\mathbf{k}'}, \quad (16.20)$$

при $\mathbf{k} = \mathbf{k}'$ получим

$$\int \varphi_{\mathbf{k}}^*(\mathbf{r}) \varphi_{\mathbf{k}}(\mathbf{r}) d\tau = 1. \quad (16.21)$$

Из (16.21) следует, что интеграл от $|\varphi_{\mathbf{k}}(\mathbf{r})|^2$ по области G должен быть равен единице. Этим условием определен выбор мультипликативной константы, входящей в $\varphi_{\mathbf{k}}(\mathbf{r})$.

Однако учитывая, что $\varphi_{\mathbf{k}}(\mathbf{r})$ — функция периодическая, достаточно проинтегрировать $|\varphi_{\mathbf{k}}(\mathbf{r})|^2$ по объему $V_a = (a_1 [a_2 a_3])$ одной элементарной ячейки. Если число таких ячеек в кристалле равно N , то, очевидно,

$$\int_G |\varphi_{\mathbf{k}}(\mathbf{r})|^2 d\tau = N \int_{V_a} |\varphi_{\mathbf{k}}(\mathbf{r})|^2 d\tau = 1, \quad (16.22)$$

и тем самым условие нормировки для $\varphi_{\mathbf{k}}(\mathbf{r})$ может быть записано в виде

$$\int_{V_a} |\varphi_{\mathbf{k}}(\mathbf{r})|^2 d\tau = N^{-1}. \quad (16.23)$$

Однако принято нормировать $\varphi_{\mathbf{k}}(\mathbf{r})$ по объему V_a на единицу:

$$\int_{V_a} |\varphi_{\mathbf{k}}(\mathbf{r})|^2 d\tau = 1. \quad (16.24)$$

В этом случае нормированная на единицу $\psi(\mathbf{r})$ должна иметь вид

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} e^{i(\mathbf{k}\mathbf{r})} \varphi_{\mathbf{k}}(\mathbf{r}). \quad (16.25)$$

Резюме § 15 и 16

1. Каждой решетке с базисом $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ соответствует так называемая обратная решетка, определяемая базисом $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ (15.1).

2. Состояния \mathbf{k} и $\mathbf{k}' = \mathbf{k} + 2\pi\mathbf{b}$ физически эквивалентны, поэтому энергия является периодической функцией волнового вектора (и квазиимпульса):

$$E(\mathbf{k} + 2\pi\mathbf{b}) = E(\mathbf{k}). \quad (16.1p)$$

Пространство \mathbf{k} (или \mathbf{P}) разбивается на эквивалентные области, содержащие все различные состояния. Эти области называют зонами Бриллюэна. Зону, симметричную относительно начала координат, называют основной зоной Бриллюэна. Она определяет набор возможных различных значений волнового вектора:

$$-\pi/a_i \leq k_i < \pi/a_i. \quad (16.2p)$$

3. Чтобы исключить граничные условия уравнения Шредингера для кристалла конечных размеров и сохранить при этом трансляционную симметрию, предполагается, что все пространство занято кристаллами в виде параллелепипеда со сторонами L_1, L_2, L_3 . При этом состояния в точках, отстоящих на расстояния, кратные L_i вдоль каждой оси, считаются эквивалентными, в силу чего волновая функция нормируется на единицу по объему $G = L_1 L_2 L_3$ (по объему «ящика»).

4. Из циклических граничных условий вытекает, что волновой вектор (и квазиимпульс) является дискретной величиной. Разность между соседними значениями $\Delta\kappa_i = \kappa_{i, n+1} - \kappa_{i, n}$ равна

$$\Delta\kappa_i = \pm 2\pi/L_i. \quad (16.3p)$$

5. Дискретность волнового вектора приводит к дискретности поверхностной энергии в зоне Бриллюэна. Они образуют большое число подуровней (или уровней) энергии. Однако, так как $\Delta\kappa_i$ — величина малая (κ_i — квазинепрерывная величина) то и $E(\mathbf{k})$ является величиной квазинепрерывной.

§ 17. ТЕОРИЯ КВАЗИСВОБОДНОГО ЭЛЕКТРОНА

Рассматривая движение электрона в периодическом поле, мы нашли интегралы движения, ввели понятие эффективной массы, установили связь между (средней) скоростью, квазиимпульсом, ускорением и внешней силой. Все эти величины прямо или косвенно связаны с зависимостью энергии от квазиимпульса $E(\mathbf{P})$ или $E(\mathbf{k})$. Как уже упоминалось, нахождение зависимости $E(\mathbf{k})$ в общем виде представляет важную, но пока еще не решенную проблему твердого тела. Однако для понимания многих процессов в твердом теле, и в первую очередь в полупроводниках, большую ценность представляют приближенные методы установления общего характера зависимостей $E(\mathbf{k})$ с использованием в ряде случаев экспериментальных величин, необходимых для практического применения теоретических расчетов.

Воспользуемся методом теории возмущений для нахождения зависимости $E(\mathbf{k})$ в общем виде. Существуют два приближения при решении задачи о движении электрона в периодическом поле, отличающиеся выбором нулевого приближения. Беря в качестве нулевого приближения свободную частицу и рассматривая периодическое поле как возмущение, приходем к так называемой теории квазисвободного электрона. Если за нулевое приближение взять электрон в изолированных атомах, из которых построена решетка кристалла, то приходим к теории квазисвязанного электрона. Рассмотрим первую теорию.

Представим гамильтониан периодического поля

$$\hat{H} = \hat{T} + \hat{U}(\mathbf{r}) \quad (17.1)$$

в виде гамильтониана «невозмущенной» системы $\hat{H}_0 = \hat{T}$ и «возмущения» $\hat{W}(\mathbf{r}) = \hat{U}(\mathbf{r})$:

$$\hat{H} = \hat{H}_0 + \hat{W}(\mathbf{r}) = \hat{T} + \hat{U}(\mathbf{r}); \quad \hat{U}(\mathbf{r}) \equiv U(\mathbf{r}). \quad (17.2)$$

Задачей теории возмущений является нахождение поправок к энергии E^0 и волновой функции ψ^0 невозмущенной системы при наложении возмущения. Найдем спектр энергии свободной частицы и ее волновые функции.

Уравнение Шредингера для свободной частицы, т. е. с гамильтонианом $\hat{H}_0 = \hat{T}$:

$$-\frac{\hbar^2}{2m} \Delta \psi^0(\mathbf{r}) = E^0 \psi^0(\mathbf{r}) \quad (17.3)$$

имеет решение в виде плоских волн де-Бройля

$$\psi_{\mathbf{k}}^0(\mathbf{r}) = A e^{i(\mathbf{k}\mathbf{r})} \quad (17.4)$$

непрерывным спектром энергии

$$E^0(\mathbf{k}) = \hbar^2 \mathbf{k}^2 / (2m) = p^2 / (2m). \quad (17.5)$$

Таким образом, в нулевом приближении энергия электрона в кристалле непрерывна, она имеет квадратичную зависимость от волнового вектора и квазиимпульса, при этом квазиимпульс тождествен с обычным импульсом; эффективная масса есть тензор нулевого ранга (скаляр), тождественная с обычной массой свободного электрона:

$$m^{*-1} = \frac{d^2 E^0}{\hbar^2 d\mathbf{k}^2} = m^{-1}; \quad m^* \equiv m. \quad (17.6)$$

Амплитуда A волны $\psi_{\mathbf{k}}^0(\mathbf{r})$ равна $1/\sqrt{(2\pi)^3}$ при нормировке на δ -функцию при интегрировании по всему пространству, или $1/\sqrt{G}$ при нормировке на единицу при интегрировании по области G , так что нормированная собственная функция имеет вид

$$\psi_{\mathbf{k}}^0(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} e^{i(\mathbf{k}\mathbf{r})}; \quad \int_{\infty} \psi_{\mathbf{k}'}^{0*}(\mathbf{r}) \psi_{\mathbf{k}}^0(\mathbf{r}) d\tau = \delta(\mathbf{k} - \mathbf{k}'); \quad (17.7)$$

$$\psi_{\mathbf{k}}^0(\mathbf{r}) = \frac{1}{G^{1/2}} e^{i(\mathbf{k}\mathbf{r})}; \quad \int_G \psi_{\mathbf{k}'}^{0*}(\mathbf{r}) \psi_{\mathbf{k}}^0(\mathbf{r}) d\tau = \delta_{\mathbf{k}\mathbf{k}'}. \quad (17.8)$$

Для нахождения энергии и собственных функций в первом приближении в соответствии с теорией возмущений понадобятся матричные элементы возмущения:

$$W_{\mathbf{k}'\mathbf{k}} = U_{\mathbf{k}'\mathbf{k}} = \int \psi_{\mathbf{k}'}^{0*}(\mathbf{r}) U(\mathbf{r}) \psi_{\mathbf{k}}^0(\mathbf{r}) d\tau. \quad (17.9)$$

Легко показать, что матричные элементы отличны от нуля только для определенных значений \mathbf{k}' и \mathbf{k} . Действительно, так как $U(\mathbf{r})$ функция периодическая с периодами $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, то можно разложить ее в тройной ряд Фурье:

$$U(\mathbf{r}) = \sum_{l_1, l_2, l_3 = -\infty}^{\infty} c_{l_1 l_2 l_3} e^{i2\pi \left(\frac{l_1 x}{a_1} + \frac{l_2 y}{a_2} + \frac{l_3 z}{a_3} \right)} = \sum_{\mathbf{b}} c_{\mathbf{b}} e^{i2\pi(\mathbf{b}\mathbf{r})}, \quad (17.10)$$

$$\mathbf{b} = l_1 \mathbf{b}_1 + l_2 \mathbf{b}_2 + l_3 \mathbf{b}_3. \quad (17.11)$$

Подставив (17.10) в выражение (17.9), получим

$$\begin{aligned}
 U_{\kappa'\kappa} &= \frac{1}{G} \int_G e^{-i(\kappa'r)} \sum_b c_b e^{i2\pi(b\mathbf{r})} e^{i(\kappa\mathbf{r})} d\tau = \\
 &= \frac{1}{G} \sum_b c_b \int_G e^{i(\kappa+2\pi b-\kappa', \mathbf{r})} d\tau = \sum_b c_b \delta_{\kappa', \kappa+2\pi b} = \\
 &= \begin{cases} 0, & \kappa' \neq \kappa + 2\pi b \\ c_b, & \kappa' = \kappa + 2\pi b. \end{cases} \quad (17.12)
 \end{aligned}$$

Таким образом, матричные элементы $U_{\kappa'\kappa}$ равны или нулю, или коэффициентам ряда Фурье для $U(\mathbf{r})$.

Поправка к энергии $E^0(\kappa)$ в первом приближении $E'(\kappa)$ равна диагональному матричному элементу оператора возмущения:

$$E'(\kappa) = U_{\kappa\kappa} = \frac{1}{G} \int_G U(\mathbf{r}) d\tau = \langle U \rangle, \quad (17.13)$$

т. е. $E'(\kappa)$ равна усредненному по области G значению потенциальной энергии U ; эта поправка не зависит от κ .

Мы видим, что в первом приближении теории возмущений спектр энергии не изменяется при наложении периодического поля, если не считать, что все уровни энергии $E^0(\kappa)$ смещаются на одну и ту же величину $\langle U \rangle$:

$$\begin{aligned}
 E^{(1)}(\kappa) &= E^0(\kappa) + E'(\kappa) = \\
 &= \frac{\hbar^2 \kappa^2}{2m} + \langle U \rangle. \quad (17.14)
 \end{aligned}$$

Величина $\langle U \rangle$ соответствует глубине потенциальной ямы, которая моделирует реальный кристалл в элементарной теории металлов Зоммерфельда. $\langle U \rangle$ — это истинная работа выхода электронов из твердого тела. На рис. 12 представлен спектр электронов в нулевом и первом приближениях. Таким образом, за исключением смещения начала отсчета энергии, спектр энергии не меняется. Выберем начало отсчета энергии так, чтобы $\langle U \rangle = 0$.

Рассмотрим поправку к энергии во втором приближении. Значение энергии во втором приближении может быть записано в виде

$$\begin{aligned}
 E^{(2)}(\kappa) &= E^0(\kappa) + E'(\kappa) + E''(\kappa) = E^0(\kappa) + \langle U \rangle + \\
 &+ \sum_{\kappa'} \frac{|U_{\kappa'\kappa}|^2}{E^0(\kappa) - E^0(\kappa')} = E^0(\kappa) + \sum_b \frac{|c_b|^2}{E^0(\kappa) - E^0(\kappa + 2\pi b)}. \quad (17.15)
 \end{aligned}$$

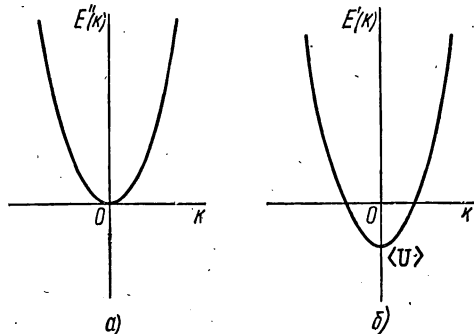


Рис. 12. Спектр энергии электрона в нулевом (а) и первом (б) приближении по возмущению в теории квазисвободного электрона

Мы видим, что поправка к энергии во втором приближении теории возмущений пропорциональна $|c_b|^2$, поэтому при

$$E^0(\kappa) \gtrless E^0(\kappa + 2\pi b) \quad (17.16)$$

поправка к энергии незначительна, и спектр энергии не меняется. Волновая функция имеет при этом вид

$$\psi_{\kappa}^{(1)}(\mathbf{r}) = \psi_{\kappa}^0(\mathbf{r}) + \sum_b \frac{c_b}{E^0(\kappa) - E^0(\kappa + 2\pi b)} \psi_{\kappa + 2\pi b}^0(\mathbf{r}), \quad (17.17)$$

и при тех же условиях можно считать, что возмущение мало меняет волновую функцию свободного движения.

Однако необходимо рассмотреть состояния, когда разность энергий $E^0(\kappa) - E^0(\kappa + 2\pi b)$ сравнима с $|c_b|^2$. В этом случае волновая функция претерпевает резкое изменение, свободное движение электрона сильно возмущено. Это возмущение будет максимальным в том случае, когда знаменатель обращается в нуль для какого-либо состояния. Пусть один из членов суммы, например при $\mathbf{b} = \mathbf{g}$, имеет знаменатель, близкий или равный нулю. Выражения (17.15) и (17.17) для энергии и волновой функции в этом случае теряют смысл, поскольку нарушается условие применимости теории возмущений. При $E^0(\kappa) \rightarrow E^0(\kappa + 2\pi g)$ коэффициент при $\psi_{\kappa + 2\pi g}^0(\mathbf{r})$ стремится к ∞ ; это означает, что доля $\psi_{\kappa + 2\pi g}^0(\mathbf{r})$ в состоянии $\psi_{\kappa}^{(1)}(\mathbf{r})$ не меньше доли $\psi_{\kappa}^0(\mathbf{r})$. При точном равенстве энергии $E^0(\kappa) = E^0(\kappa + 2\pi g)$ состояние $E^0(\kappa)$ является вырожденным, так как одной и той же энергии соответствуют две разные функции $\psi_{\kappa}^0(\mathbf{r})$ и $\psi_{\kappa + 2\pi g}^0(\mathbf{r})$. Но это значит, что уже в нулевом приближении для волновой функции $\psi^0(\mathbf{r})$ необходимо учитывать обе волновые функции, поэтому нужно решать задачу, учитывая вырождение. В соответствии с теорией возмущений для вырожденных состояний волновая функция в нулевом приближении $\psi^0(\mathbf{r})$ должна иметь вид

$$\psi^0(\mathbf{r}) = \alpha \psi_{\kappa}^0(\mathbf{r}) + \beta \psi_{\kappa + 2\pi g}^0(\mathbf{r}), \quad (17.18)$$

где α и β — неизвестные коэффициенты. Для невырожденных состояний имеем $\beta \ll \alpha$, для вырожденных состояний α и β — величины одного порядка. Это будет справедливо не только при точном равенстве энергий, оно будет справедливо до тех пор, пока

$$|E^0(\kappa) - E^0(\kappa + 2\pi g)| \ll |c_g|^2. \quad (17.19)$$

Итак, беря в качестве нулевого приближения функцию (17.18) и подставляя ее в уравнение

$$(\hat{H}^0 + \hat{U})\psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (17.20)$$

получим, учитывая, что $\psi_{\kappa}^0(\mathbf{r})$ и $\psi_{\kappa + 2\pi g}^0(\mathbf{r})$ являются собственными функциями \hat{H}^0 :

$$\alpha E^0(\kappa) \psi_{\kappa}^0(\mathbf{r}) + \beta E^0(\kappa + 2\pi g) \psi_{\kappa + 2\pi g}^0(\mathbf{r}) + U(\mathbf{r}) \psi^0(\mathbf{r}) = E \psi^0(\mathbf{r}). \quad (17.21)$$

Обозначим для простоты

$$\psi_k^0(\mathbf{r}) = \psi_1; \psi_{k+2\pi g}^0(\mathbf{r}) = \psi_2; E^0(k) = E_1; E^0(k+2\pi g) = E_2 \quad (17.22)$$

и перепишем (17.21) в новых обозначениях:

$$\alpha E_1 \psi_1 + \beta E_2 \psi_2 + U(\alpha \psi_1 + \beta \psi_2) = E(\alpha \psi_1 + \beta \psi_2). \quad (17.23)$$

Умножим это уравнение слева на ψ_1^* и ψ_2^* соответственно и, интегрируя по \mathbf{r} , получим

$$\begin{aligned} \alpha E_1 + \alpha U_{11} + \beta U_{12} &= E\alpha, \\ \beta E_2 + \alpha U_{21} + \beta U_{22} &= E\beta, \end{aligned} \quad (17.24)$$

или

$$\begin{aligned} (E_1 + U_{11} - E)\alpha + U_{12}\beta &= 0, \\ U_{21}\alpha + (E_2 + U_{22} - E)\beta &= 0. \end{aligned} \quad (17.25)$$

Два уравнения (17.25) содержат три неизвестные величины E , α , β . Для того чтобы уравнения имели нетривиальное решение, необходимо, чтобы детерминант системы был равен нулю, что дает необходимое третье уравнение, из которого можно определить энергию E :

$$\begin{vmatrix} (E_1 + U_{11} - E) & U_{12} \\ U_{21} & (E_2 + U_{22} - E) \end{vmatrix} = 0. \quad (17.26)$$

Решая это уравнение и учитывая, что $U_{11} = U_{22} = \langle U \rangle = 0$, получим

$$E = \frac{E_1 + E_2}{2} \pm \sqrt{\frac{(E_1 - E_2)^2}{4} + U_{12}U_{21}}. \quad (17.27)$$

Переходя к старым обозначениям, можем записать:

$$\begin{aligned} E^{(1)}(k) &= \frac{E^0(k) + E^0(k+2\pi g)}{2} \pm \\ &\pm \sqrt{\frac{[E^0(k) - E^0(k+2\pi g)]^2}{4} + |U_g|^2}. \end{aligned} \quad (17.28)$$

Мы видим, что при наложении возмущения $W(\mathbf{r}) = U(\mathbf{r})$ энергия $E^0(k)$ терпит разрыв в точке, где $E^0(k) = E^0(k+2\pi g)$. Для этих точек $E(k) = E^0(k) \pm |U_g|$. Величина разрыва равна $2|U_g|$.

Исследуем те состояния, в которых энергия терпит разрыв. Условие разрыва можно переписать иначе:

$$\begin{aligned} E^0(k) - E^0(k+2\pi g) &= \frac{\hbar^2}{2m} [k^2 - (k+2\pi g)^2] = \\ &= \frac{\hbar^2}{2m} [4\pi^2 g^2 + 4\pi(kg)] = 0, \end{aligned} \quad (17.29)$$

или

$$(k + \pi g, g) = 0. \quad (17.30)$$

Это условие имеет наглядный геометрический и физический смысл. Нужно найти все те значения k , которые удовлетворяют условию

(17.30). Оно гласит, что векторы g и $k + \pi g$ ортогональны. Построим вектор $2\pi g$. Поместим начало отсчета волнового вектора k в точку, совпадающую с концом вектора $2\pi g$. Проведем через середину вектора $2\pi g$ перпендикулярную ему плоскость (плоскость BB' на

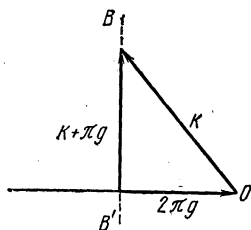


Рис. 13. Построение плоскости, в которой энергия терпит разрыв

рис. 13). Любой вектор, лежащий в этой плоскости, будет перпендикулярен вектору $2\pi g$ или g . Но вектор, лежащий в плоскости BB' , можно получить как сумму векторов πg и k , если k оканчивается на плоскости BB' , т. е. все векторы k , которые оканчиваются на плоскости BB' , удовлетворяют условию (17.30); оно является уравнением плоскости, перпендикулярной вектору g . Чтобы найти все возможные значения k , удовлетворяющие условию (17.19), необходимо задать все векторы g . Для этого надо построить обратную решетку с базисом $2\pi b_1, 2\pi b_2, 2\pi b_3$, наложенную на пространство k , и изо всех узлов решетки провести векторы $2\pi g$ в начало координат $k=0$. Совокупность плоскостей, проходящих через середины векторов $2\pi g$ и перпендикулярных к ним, определит все те векторы k , которые удовлетворяют условиям

$$E^0(k + 2\pi g) = E^0(k), \text{ или } (k + \pi g, g) = 0. \quad (17.31)$$

Таким образом, уравнение (17.30) позволяет построить зоны Бриллюэна для любого кристалла по его обратной решетке. На рис. 14 представлено построение плоскостей, точки которых удовлетворяют уравнению (17.30) для различных векторов g . Различные зоны Бриллюэна заштрихованы по-разному. Построение зон Бриллюэна на основе уравнения (17.30) является наиболее естественным.

Чтобы построить основную зону, необходимо выделить минимальный по объему многогранник, содержащий точку $k=0$. Найдя следующий по объему многогранник и исключив из него первую зону, найдем вторую зону Бриллюэна и т. д.

Для простой кубической решетки обратная решетка является также простой кубической. Если начало отсчета $k=0$ поместить в один из узлов решетки, то ближайших к точке $k=0$ узлов будет шесть; они дадут шесть взаимно ортогональных плоскостей, которые выделяют в пространстве волнового вектора куб со стороной $2\pi/a$ и объемом $8\pi^3/a^3$. При этом начало отсчета $k=0$ совпадает с одним из узлов решетки; обратная решетка строится на основе базиса $2\pi b_1, 2\pi b_2, 2\pi b_3$.

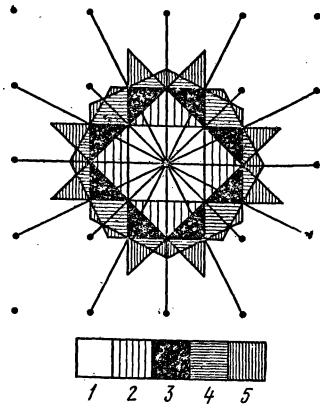


Рис. 14. Зоны Бриллюэна для плоской квадратной решетки

Рассмотрим гранецентрированную кубическую решетку; ее обратная решетка является объемноцентрированной. Если поместить начало координат $k=0$ в одну из вершин куба, то на расстоянии $2\pi/a$ вдоль осей располагается шесть узлов, лежащих в вершинах ячеек. Шесть векторов, идущих из вершин, определяют положение шести взаимно ортогональных плоскостей, т. е. выделяют куб. Вершины этого куба находятся в узлах решетки с координатами $(\pi b, \pi b, \pi b)$, $(-\pi b, \pi b, \pi b)$ и т. д.

Если провести восемь векторов из вершин в начало координат, то они определяют восемь плоскостей, перпендикулярных пространственным диагоналям куба. Эти плоскости отсекают области, прилегающие к вершинам куба. Секущие плоскости проходят через точки диагоналей с координатами $(\pi/a, \pi/a, \pi/a)$, $(-\pi/a, \pi/a, \pi/a)$ и т. д. В результате учета этих плоскостей приходим к зоне Бриллюэна, изображенной на рис. 15. Она является четырнадцатиграничником, шесть граней ее имеют форму квадратов и восемь граней — шестиугольной формы (куб-октаэдр). Аналогично можно построить зону Бриллюэна и для других решеток. Зона Бриллюэна для кристаллов с решеткой типа алмаза совпадает с зоной, изображенной на рис. 15. Для объемноцентрированной решетки первая зона Бриллюэна представляет собой додекаэдр.

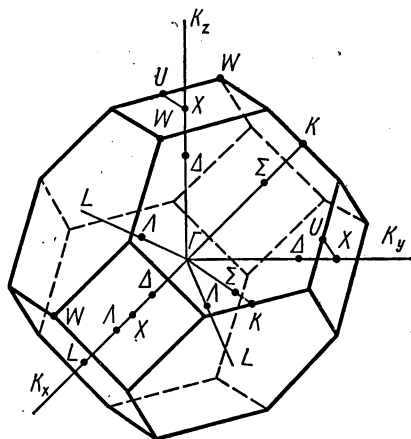


Рис. 15. Зоны Бриллюэна для гранецентрированной кубической решетки. Координаты точек указаны в π/a

Рассмотрев геометрический смысл уравнения (17.30), рассмотрим его физический смысл. Учитывая, что $|k| = 2\pi/\lambda$, где λ — длина волны де-Бройля, $|b| = l/d$, где l — целое число, а d — расстояние между ближайшими плоскостями семейства атомных плоскостей прямой решетки, соответствующих вектору b ; обозначая угол между вектором k и нормалью к плоскости (угол падения) через θ , условие вырождения (17.30) можем записать в виде

$$(kb) + \pi b^2 = 0, \quad (17.32)$$

или

$$-\frac{2\pi}{\lambda} \frac{l}{d} \cos \theta + \pi \frac{l^2}{d^2} = 0. \quad (17.33)$$

Сокращая на l, d и π , получим

$$2d \cos \theta = \lambda \quad (17.34)$$

— известное условие Вульфа — Брэгга для интерференции рентгеновых лучей, если учтем, что угол θ — это угол падения,

а не угол скольжения, который обычно вводится в данное выражение.

Таким образом, вырождение будет наблюдаться для состояний, лежащих на границах зон Бриллюэна, и это те самые состояния, для которых должно наблюдаться вольф-брэгговское отражение.

Двум различным значениям энергии будут соответствовать две различные комбинации волновых функций $\psi_{\mathbf{k}}^0(\mathbf{r})$ и $\psi_{\mathbf{k}+2\pi\mathbf{g}}^0(\mathbf{r})$ в виде их суммы и разности.

Сумма двух волн представляет собой одну волну, идущую в направлении вектора $\pi\mathbf{g} + \mathbf{k}$, т. е. вдоль границы зоны Бриллюэна, которая параллельна соответствующим атомным плоскостям кристалла. Приходим тем самым к простому и наглядному истолкованию причины разрыва энергии. *Вследствие вольф-брэгговского отражения падающая и отраженная волны интерферируют; в направлениях, перпендикулярных атомным плоскостям, возникают стоячие волны; бегущие волны могут быть только вдоль атомных плоскостей.* Поскольку некоторые состояния оказываются запрещенными, то будут запрещены и некоторые значения энергии. Разрывы энергии возникают на всех границах зон Бриллюэна.

Возмущение существенно также и для состояний вблизи границ зон Бриллюэна. Рассмотрим зависимость энергии от квазиимпульса в окрестностях точек разрыва, где энергию можно записать в виде

$$E(\mathbf{k}) = \frac{E^0(\mathbf{k}) + E^0(\mathbf{k} + 2\pi\mathbf{g})}{2} \pm \sqrt{\left[\frac{E^0(\mathbf{k}) - E^0(\mathbf{k} + 2\pi\mathbf{g})}{2}\right]^2 + |U_{\mathbf{g}}|^2}, \quad (17.35)$$

или

$$E(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m} + \frac{\hbar^2 \pi}{m} (\mathbf{g}, \mathbf{k} + \pi\mathbf{g}) \pm \sqrt{\left(\frac{\pi \hbar^2}{m}\right)^2 (\mathbf{g}, \mathbf{k} + \pi\mathbf{g})^2 + |U_{\mathbf{g}}|^2}. \quad (17.36)$$

Обозначим волновой вектор точек на границе зоны Бриллюэна через \mathbf{k}_0 . Так как $(\mathbf{g}, \mathbf{k}_0 + \pi\mathbf{g}) = 0$, то

$$E(\mathbf{k}_0) = \frac{\hbar^2 \mathbf{k}_0^2}{2m} \pm |U_{\mathbf{g}}|. \quad (17.37)$$

В таком случае для произвольного состояния \mathbf{k} можем записать:

$$E(\mathbf{k}) - E(\mathbf{k}_0) = \frac{\hbar^2 \mathbf{k}^2}{2m} - \frac{\hbar^2 \mathbf{k}_0^2}{2m} + \frac{\hbar^2 \pi}{m} (\mathbf{g}, \mathbf{k} + \pi\mathbf{g}) \pm \sqrt{\left(\frac{\pi \hbar^2}{m}\right)^2 (\mathbf{g}, \mathbf{k} + \pi\mathbf{g})^2 + |U_{\mathbf{g}}|^2} \mp |U_{\mathbf{g}}| = f(\mathbf{k}). \quad (17.38)$$

Разложим это выражение в ряд Тейлора относительно точек \mathbf{k}_0 . Очевидно, что $f(\mathbf{k}_0) = 0$. Скорость электронов

$$\mathbf{v} = \frac{dE(\mathbf{k})}{\hbar d\mathbf{k}} = \frac{\hbar \mathbf{k}}{m} + \frac{\pi \hbar \mathbf{g}}{m} \pm \frac{1}{\hbar} \left(\frac{\pi \hbar^2}{m}\right)^2 \frac{(\mathbf{g}, \mathbf{k} + \pi\mathbf{g}) \mathbf{g}}{\sqrt{\left(\frac{\pi \hbar^2}{m}\right)^2 (\mathbf{g}, \mathbf{k} + \pi\mathbf{g})^2 + |U_{\mathbf{g}}|^2}}. \quad (17.39)$$

Для свободного движения (для состояний, не лежащих вблизи границ зон Бриллюэна) скорость равна

$$v = \frac{\hbar k}{m}, \quad (17.40)$$

в то время как в окрестности границ скорость не совпадает по направлению с квазиимпульсом. На самой границе зоны $k = k_0$, поэтому

$$v = \frac{\hbar k_0}{m} + \frac{\pi \hbar g}{m} = \frac{\hbar}{m} (k_0 + \pi g). \quad (17.41)$$

Но вектор $k_0 + \pi g$ лежит в плоскости границы зоны. Для точек разрыва энергии скорость направлена вдоль границы; очевидно, скорость будет иметь преимущественное направление вдоль граничной плоскости и для состояний в небольшой окрестности границы зоны Бриллюэна. Так как $\left. \frac{dE}{dk} \right|_{k_0}$ лежит в плоскости границы, то *изоэнергетические поверхности должны быть ортогональны к границам зон Бриллюэна, на которых они терпят разрыв*. Для состояний $k_0 = -\pi g$ скорость обращается в нуль, что означает невозможность движения в направлении, перпендикулярном атомным плоскостям. Так как для этих точек $\frac{dE}{dk} = 0$, то в них должен находиться экстремум энергии. Найдем вторую производную:

$$\frac{d^2E(k)}{\hbar^2 dk^2} = \frac{1}{m} \pm \frac{\pi^2 \hbar^2}{m^2} \frac{g^2 |U_g|^2}{\left\{ \left(\frac{\pi \hbar^2}{m} \right)^2 (g, k + \pi g)^2 + |U_g|^2 \right\}^{3/2}}. \quad (17.42)$$

Для точек, лежащих на границах зон Бриллюэна,

$$\left. \frac{d^2E(k)}{\hbar^2 dk^2} \right|_{k_0} = \frac{1}{m} \left(1 \pm \frac{\pi^2 \hbar^2 g^2}{m |U_g|} \right). \quad (17.43)$$

Если точка k_0 является экстремальной, то

$$\left. \frac{d^2E}{\hbar^2 dk^2} \right|_{k_0} = \frac{1}{m^*} = \frac{1}{m} \left(1 \pm \frac{\pi^2 \hbar^2 g^2}{m |U_g|} \right), \quad (17.44)$$

или

$$m^* = \frac{m}{1 \pm \frac{\pi^2 \hbar^2 g^2}{m |U_g|}}. \quad (17.45)$$

Знак плюс соответствует энергии $\frac{\hbar^2 k_0}{2m} + |U_g|$, т. е. это значение энергии

Если $|U_g|$ положить равным одному электронвольту, то $m^* = m/(1 \pm 6)$, т. е. для минимума энергии $m^* < 0$, для максимума энергии $m^* > 0$, причем $|m^*|$ меньше m , т. е. эффективная масса меньше массы свободного электрона.

Запишем теперь выражение для энергии в виде ряда Тейлора относительно произвольной точки k_0 на границе зоны Бриллюэна:

$$E(k) = E(k_0) + \frac{\hbar^2}{m} (k_0 + \pi g, k - k_0) + \frac{1}{2} \frac{\hbar^2}{m} \left(1 \pm \frac{\pi^2 \hbar^2 g^2}{m |U_g|} \right) (k - k_0)^2 + \dots, \quad (17.47)$$

где

$$E(k_0) = \frac{\hbar^2 k_0^2}{2m} \pm |U_g|.$$

Разложение начинается с линейного члена. Если вектор $k - k_0$ будет направлен перпендикулярно границе зоны, то энергия является квадратичной функцией $(k - k_0)$.

Теперь рассмотрим основной вопрос — образование зон энергии. Для этого необходимо учесть, что для точек вблизи границы зон Бриллюэна получены две ветви энергии:

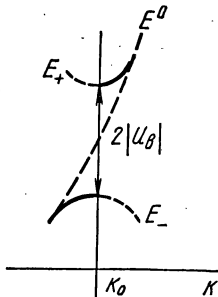


Рис. 16. Две ветви энергии E_+ и E_- — в окрестности границы зоны Бриллюэна

$$E^+(k) = \frac{\hbar^2 k_0^2}{2m} + |U_g| + \frac{\hbar^2}{m} (k_0 + \pi g, k - k_0) + \frac{\hbar^2}{2m} \left(1 + \frac{\pi^2 \hbar^2 g^2}{m |U_g|} \right) (k - k_0)^2,$$

$$E^-(k) = \frac{\hbar^2 k_0^2}{2m} - |U_g| + \frac{\hbar^2}{m} (k_0 + \pi g, k - k_0) + \frac{\hbar^2}{2m} \left(1 - \frac{\pi^2 \hbar^2 g^2}{m |U_g|} \right) (k - k_0)^2. \quad (17.48)$$

Каждое выражение E^+ и E^- является непрерывным на границе зон Бриллюэна. На рис. 16 дана зависимость E^+ и E^- от $(k - k_0)$ для направлений вдоль нормали к границе зоны. Для того чтобы энергия испытывала скачок на границе, будем считать, что каждое выражение E^+ и E^- применимо только с одной ее стороны. Учитывая непрерывность энергии внутри зоны и учитывая, что E^+ и E^- должны переходить в выражение $E^0(k) = \hbar^2 k^2 / (2m)$ вдали от границы, выберем E^- для внутренней стороны границы, а E^+ для внешней стороны (внутренней стороной границы считается та, с которой находится начало отсчета $k = 0$); на рис. 16 они изображены сплошными линиями.

Наличие разрыва энергии в точках k_0 еще не означает разрыва в спектре энергии $E(k)$. Действительно, возьмем две очень близкие к границе точки k_1 и k_2 , соответственно с внутренней и наружной сторон, такие, что при $|k_1| > |k_2|$

$$\frac{\hbar^2 k_1^2}{2m} - |U_g| > \frac{\hbar^2 k_2^2}{2m} + |U_g|, \quad (17.49)$$

т. е. полосы энергии E^- и E^+ перекрываются. Но это означает, что спектр энергии в этом случае непрерывен.

Для того чтобы возник интервал запрещенных значений в спектре энергии, необходимо, чтобы перекрытия полос не происходило. Эта возможность становится очевидной, если учесть существование других граничных поверхностей.

Рассмотрим качественно особенность характера энергетического спектра в окрестности одной из вершин зоны Бриллюэна. Пусть границы зон заданы тремя уравнениями:

$$\left. \begin{aligned} (\mathbf{g}_1, \mathbf{k}_{01} + \pi\mathbf{g}_1) &= 0, \\ (\mathbf{g}_2, \mathbf{k}_{02} + \pi\mathbf{g}_2) &= 0, \\ (\mathbf{g}_3, \mathbf{k}_{03} + \pi\mathbf{g}_3) &= 0, \end{aligned} \right\} \quad (17.50)$$

где \mathbf{g}_1 , \mathbf{g}_2 и \mathbf{g}_3 — некоторые векторы обратной решетки; \mathbf{k}_{01} , \mathbf{k}_{02} , \mathbf{k}_{03} — текущие радиус-векторы соответствующих граничных плоскостей.

Так как для состояний на граничных плоскостях скорости лежат в самих плоскостях, то для состояний, определяемых линией пересечения двух плоскостей, скорость может быть направлена только по этой линии. Если же рассмотрим состояние, соответствующее точке \mathbf{k}_0^0 пересечения трех плоскостей, то скорость $\mathbf{v}(\mathbf{k}_0^0)$ может быть равна только нулю, т. е. в вершине зоны Бриллюэна должен быть экстремум энергии, так как в ней

$$\left. \frac{dE}{d\mathbf{k}} \right|_{\mathbf{k}_0^0} = 0. \quad (17.51)$$

Мы видим, что разложение в ряд Тейлора для экстремумов, лежащих в вершинах зон, не будет содержать линейных членов. Считая векторы \mathbf{g}_1 , \mathbf{g}_2 и \mathbf{g}_3 взаимно ортогональными, можем найти приближенное выражение для энергии в окрестности точки \mathbf{k}_0^0 .

Одновременно должны быть справедливы три уравнения:

$$E^\pm(\mathbf{k}) = E^\pm(\mathbf{k}_0^0) + \frac{1}{2} \frac{\hbar^2}{m} \left(1 \pm \frac{\pi^2 \hbar^2 \mathbf{g}_i^2}{m |U_{\mathbf{g}_i}|} \right) (\mathbf{k} - \mathbf{k}_0^0)^2 \quad (i = 1, 2, 3). \quad (17.52)$$

Для того чтобы они удовлетворялись одновременно, необходимо, чтобы каждое из них было справедливо для проекций $(\mathbf{k} - \mathbf{k}_0^0)$ на соответствующие векторы \mathbf{g}_i , так что

$$\begin{aligned} E^\pm(\mathbf{k}) &= E^\pm(\mathbf{k}_0^0) \pm \sum_{i=1}^3 \frac{\hbar^2}{2m} \left(1 \pm \frac{\pi^2 \hbar^2 \mathbf{g}_i^2}{m |U_{\mathbf{g}_i}|} \right) (\kappa_i - \kappa_{0i}^0)^2 = \\ &= E^\pm(\mathbf{k}_0^0) \pm \sum_{i=1}^3 \frac{\hbar^2}{2m_i^*} (\kappa_i - \kappa_{0i}^0)^2. \end{aligned} \quad (17.53)$$

При этом плюс и минус должны относиться к состояниям внешним и внутренним для каждой поверхности. Величина разрыва

будет разной для различных направлений. Термической шириной запрещенной зоны будет минимальная величина из трех ($2|U_{g_i}|$), поскольку она определяется минимальным энергетическим зазором, существенным для явлений электропроводности. Для явлений, в которых существует закон сохранения квазиимпульса, ширина запрещенной зоны должна изменяться от минимальной до максимальной величины из ($2|U_{g_i}|$).

Наличие разрыва в спектре энергии становится очевидным из самого выражения для энергии — при смещении от точки κ_0^0 энергия или возрастает по сравнению с $E^+(\kappa_0^0)$, или уменьшается по сравнению с $E^-(\kappa_0^0)$, а $E(\kappa_0^0)$ в данном случае не может изменяться, в то время как при разрыве энергии в точках граничной плоскости изменение энергии достигается за счет изменения $E(\kappa_0)$ и линейного члена.

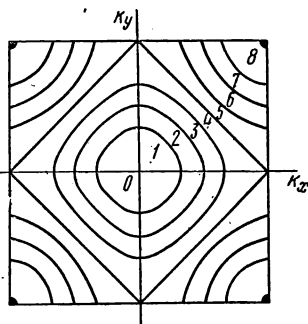


Рис. 17. Изоэнергетические линии для квадратной плоской решетки

Для получения более точного выражения для энергии в окрестности вершин зон Бриллюэна необходимо учесть четырехкратное вырождение энергии, в силу чего волновую функцию нулевого приближения надо брать в виде.

$$\psi^0(\mathbf{r}) = \sum_{i=1}^4 \alpha_i \psi_{k_i}^0(\mathbf{r}), \quad (17.54)$$

где

$$\psi_{k_1}^0(\mathbf{r}) = \frac{1}{\sqrt{G}} e^{i(\kappa \mathbf{r})}; \quad \psi_{k_j}^0(\mathbf{r}) = \frac{1}{\sqrt{G}} e^{i(\kappa + 2\pi \mathbf{g}_j, \mathbf{r})}, \quad (j = 1, 2, 3). \quad (17.55)$$

Для определения $E(\mathbf{k})$ получим уравнение четвертого порядка. В заключение рассмотрим форму изоэнергетических поверхностей в зонах Бриллюэна и зависимость энергии от \mathbf{k} вдоль некоторых направлений.

Вблизи центра первой зоны Бриллюэна при $\mathbf{k} = 0$ поверхности энергии являются сферами (на рис. 17 приведены поверхности энергии, т. е. изоэнергетические линии, для плоской решетки). По мере удаления от центра сферы начинают деформироваться. Характер деформации можно понять, если рассмотреть поверхности энергии в углах зон, где они должны быть частью поверхности эллипсоида.

Рассмотрим теперь зависимость энергии от $|\mathbf{k}|$ для некоторого направления. Поскольку изоэнергетические поверхности имеют сложную форму, графики, выражающие зависимость энергии от $|\mathbf{k}|$ для разных направлений, будут разными. На рис. 18 представлена зависимость $E(\mathbf{k})$ для одного из осевых направлений, перпендикулярных границам зон Бриллюэна. Пунктиром обозначена зависимость $E^0(\mathbf{k})$ для свободного электрона. Расстояние до границы для различных направлений различно, поэтому положение точки, где

энергия терпит разрыв, зависит от направления и характеризуется различной величиной. Зависимость энергии может быть продолжена в соседние зоны, как это показано на рис. 18 пунктиром, тем самым приходим к известному уже результату, что энергия является периодической функцией от k :

$$E(k + 2\pi b) = E(k). \quad (17.56)$$

Однако возможно другое представление. В силу периодической зависимости энергии от волнового вектора участки кривых, указанные на рис. 18 сплошной линией, могут быть перенесены в первую зону Бриллюэна. В этом случае энергия является многозначной

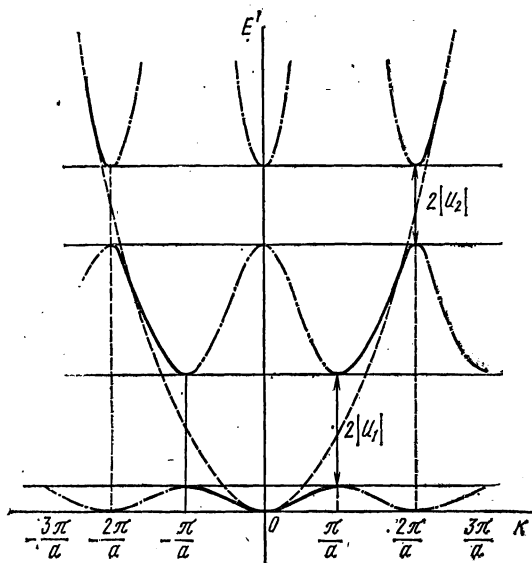


Рис. 18. Зависимость энергии от волнового вектора вдоль некоторого направления

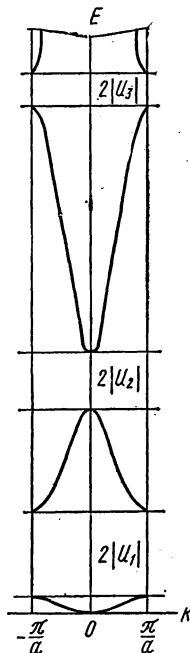


Рис. 19. Спектр энергии в основной зоне Бриллюэна

функцией от волнового вектора, но для каждой энергетической зоны, или полосы, энергия однозначна (в реальных кристаллах даже для одной полосы энергия может быть многозначной функцией). Удобно перестроить графики таким образом, чтобы экстремумы двух соседних зон энергии, необходимых для описания каких-либо процессов, находились в центре первой зоны Бриллюэна (рис. 19). Как увидим ниже, в реальных кристаллах экстремумы энергии разных зон могут находиться в разных точках зоны Бриллюэна.

Резюме § 17

1. В теории квазисвободного электрона в качестве нулевого приближения для электрона в периодическом поле принимается решение уравнения Шредингера для свободного электрона:

$$\hat{H}^0 = \hat{T} = -\frac{\hbar^2}{2m} \Delta; \quad E^0(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m}; \quad (17.1p)$$

$$\Psi_{\mathbf{k}}^0(\mathbf{r}) = \frac{1}{\sqrt{G}} e^{i(\mathbf{k}\mathbf{r})}. \quad (17.2p)$$

2. Потенциальная энергия поля решетки рассматривается в качестве возмущения

$$\hat{W}(\mathbf{r}) \equiv \hat{U}(\mathbf{r}). \quad (17.3p)$$

Матричные элементы оператора \hat{W} , вычисленные с помощью волновых функций $\Psi_{\mathbf{k}}^0(\mathbf{r})$, отличны от нуля, если $\mathbf{k}' = \mathbf{k} + 2\pi\mathbf{b}$; они равны при этом коэффициентам разложения $U(\mathbf{r})$ в ряд Фурье:

$$W_{\mathbf{k}'\mathbf{k}} = U_{\mathbf{b}} = c_{\mathbf{b}}; \quad \mathbf{b} = \frac{\mathbf{k}' - \mathbf{k}}{2\pi}. \quad (17.4p)$$

Для малых значений $|\mathbf{b}|$ величина $U_{\mathbf{b}}$ будет больше, чем для больших $|\mathbf{b}|$, поскольку большие значения $|\mathbf{b}|$ соответствуют «гармоникам» поля $U(\mathbf{r})$ более высокого порядка.

3. В первом приближении теории возмущения спектр энергии смещается вниз на величину $\langle U \rangle$. Во втором приближении для большинства состояний поправка к энергии практически равна нулю. Но существуют состояния, для которых поправки к энергии и волновой функции обращаются в бесконечность. Физически это означает, что такие состояния вырождены.

4. Вырожденные состояния лежат в плоскостях, которые удовлетворяют уравнению

$$(\mathbf{b}, \mathbf{k} + \pi\mathbf{b}) = 0, \quad (17.5p)$$

представляющему собой уравнение Вульф—Брэгга. Эти плоскости удобно выбрать в качестве границ зон Бриллюэна. Решение задачи методами теории возмущений для вырожденных состояний показывает, что энергия терпит разрыв величиной $2|U_{\mathbf{b}}|$ в точках вырождения, т. е. на границах зон Бриллюэна.

5. Разрыв в энергии приводит к разрыву в спектре энергии, возникают запрещенные значения энергии, которые называются запрещенными зонами.

Ширина запрещенной зоны равна $2|U_{\mathbf{b}}|$, поэтому она растет с ростом потенциального поля $U(\mathbf{r})$. С ростом энергии E ширина запрещенных зон уменьшается, а ширина зон энергии увеличивается (см. рис. 19).

6. Физической причиной образования зонной структуры спектра является вульф-брэггова интерференция электронных волн.

7. Эффективная масса имеет различный знак в максимуме и минимуме энергии. Пренебрегая единицей по сравнению с другим членом в знаменателе выражения (17. 45), получим

$$\frac{m^*}{m} = \pm \frac{m}{\pi^2 \hbar^2 g^2} |U_g|. \quad (17.6p)$$

Из (17.6p) следует, что: а) $|m^*|$ пропорциональна ширине запрещенной зоны; б) с ростом энергии $|m^*|$ уменьшается, т. е. в более высоких зонах энергии $|m^*|$ меньше, чем в более низких.

8. Энергия является квадратичной функцией волнового вектора в окрестности экстремумов. Поверхности энергии ортогональны границам зон Бриллюэна.

§ 18. ТЕОРИЯ КВАЗИСВЯЗАННОГО ЭЛЕКТРОНА

В теории квазисвязанного электрона при решении уравнения Шредингера для электрона в периодическом поле

$$\left. \begin{aligned} \hat{H} \psi(\mathbf{r}) &= E \psi(\mathbf{r}), \\ \hat{H} &= -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}); \quad U(\mathbf{r}) = U(\mathbf{r} + \mathbf{n}) \end{aligned} \right\} \quad (18.1)$$

за нулевое приближение принимается состояние электрона в изолированном атоме. Обозначим гамильтониан изолированного атома через \hat{H}_a :

$$\hat{H}_a = -\frac{\hbar^2}{2m} \Delta + V_a(\mathbf{r}) \quad (18.2)$$

и запишем уравнение Шредингера для атома

$$\hat{H}_a \psi_a(\mathbf{r}) = E_a \psi_a(\mathbf{r}), \quad (18.3)$$

где $V_a(\mathbf{r})$ — потенциальная энергия электрона в изолированном атоме; E_a — энергия некоторого уровня, $\psi_a(\mathbf{r})$ — соответствующая уровню E_a волновая функция. На рис. 20 показана схема уровней энергии изолированного атома. Отличительной особенностью атомной волновой функции является ее резкая зависимость от расстояния: начиная с некоторых значений r , волновая функция уменьшается по экспоненте. Решение уравнения для атома считается известным. Для нахождения энергии электрона в кристалле в первом приближении необходимо подобрать волновую функцию в нулевом приближении. С этой целью предположим вначале, что атомы кристалла не взаимо-

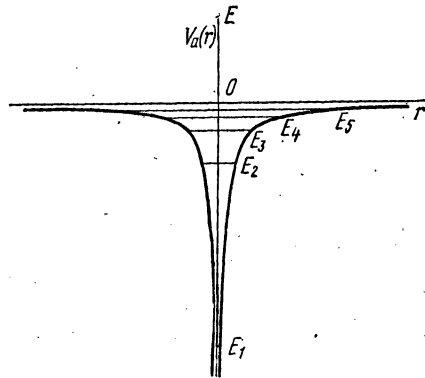


Рис. 20. Потенциальная энергия и уровни изолированного атома

действуют друг с другом. Совместим начало координат с каким-либо ядром, тогда координаты остальных ядер можно представить в виде вектора трансляции:

$$\mathbf{n} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3.$$

Если обозначить текущий радиус-вектор через \mathbf{r} , то расстояние между данной точкой \mathbf{r} и ядром \mathbf{n} равно $|\mathbf{r} - \mathbf{n}|$, волновая функция электрона в \mathbf{n} -ом атоме запишется в виде $\psi_a(\mathbf{r} - \mathbf{n})$.

Волновая функция электрона в кристалле $\psi^0(\mathbf{r})$ должна быть образована из атомных волновых функций $\psi_a(\mathbf{r} - \mathbf{n})$:

$$\psi^0(\mathbf{r}) = \sum_{\mathbf{m}} c_{\mathbf{m}} \psi_a(\mathbf{r} - \mathbf{m}). \quad (18.4)$$

Коэффициенты $c_{\mathbf{m}}$ выбираются таким образом, чтобы $\psi^0(\mathbf{r})$ удовлетворяла трансляционному условию

$$\psi^0(\mathbf{r} + \mathbf{n}) = e^{i(\mathbf{k}\mathbf{n})} \psi^0(\mathbf{r}), \quad (18.5)$$

где \mathbf{n} — вектор трансляции и \mathbf{k} — волновой вектор. Для того чтобы трансляционное условие выполнялось, $c_{\mathbf{m}}$ можно выбрать в виде

$$c_{\mathbf{m}} = e^{i(\mathbf{k}\mathbf{m})}, \quad (18.6)$$

или

$$\psi^0(\mathbf{r}) = \sum_{\mathbf{m}} e^{i(\mathbf{k}\mathbf{m})} \psi_a(\mathbf{r} - \mathbf{m}). \quad (18.7)$$

Проверим, выполняется ли условие (18.5) для функции (18.7):

$$\begin{aligned} \psi^0(\mathbf{r} + \mathbf{n}) &= \sum_{\mathbf{m}} e^{i(\mathbf{k}\mathbf{m})} \psi_a(\mathbf{r} - \mathbf{m} + \mathbf{n}) = \\ &= e^{i(\mathbf{k}\mathbf{n})} \sum_{\mathbf{m}} e^{i(\mathbf{k}, \mathbf{m} - \mathbf{n})} \psi_a[\mathbf{r} - (\mathbf{m} - \mathbf{n})] = \\ &= e^{i(\mathbf{k}\mathbf{n})} \sum_{\mathbf{l}} e^{i(\mathbf{k}\mathbf{l})} \psi_a(\mathbf{r} - \mathbf{l}). \end{aligned} \quad (18.8)$$

Очевидно, что $\mathbf{l} = \mathbf{m} - \mathbf{n}$ пробегает те же значения, что и \mathbf{m} , только в другой последовательности, поэтому

$$\sum_{\mathbf{l}} e^{i(\mathbf{k}\mathbf{l})} \psi_a(\mathbf{r} - \mathbf{l}) = \psi^0(\mathbf{r}), \quad (18.9)$$

и, следовательно, трансляционное условие выполнено.

Так как $\psi_a(\mathbf{r} - \mathbf{l})$ нормированы к единице, то $\psi^0(\mathbf{r})$ не нормирована. Найдем ее нормировочный множитель:

$$\begin{aligned} &\int \psi^{0*}(\mathbf{r}) \psi^0(\mathbf{r}) d\tau = \\ &= \int \left[\sum_{\mathbf{m}} e^{-i(\mathbf{k}\mathbf{m})} \psi_a^*(\mathbf{r} - \mathbf{m}) \right] \left[\sum_{\mathbf{n}} e^{i(\mathbf{k}\mathbf{n})} \psi_a(\mathbf{r} - \mathbf{n}) \right] d\tau = \\ &= \sum_{\mathbf{m}} \sum_{\mathbf{n}} e^{i(\mathbf{k}, \mathbf{n} - \mathbf{m})} \int \psi_a^*(\mathbf{r} - \mathbf{m}) \psi_a(\mathbf{r} - \mathbf{n}) d\tau. \end{aligned} \quad (18.10)$$

Двойная сумма легко сводится к одинарной сумме. Для этого положим

$$\mathbf{r} - \mathbf{m} = \mathbf{r}'; \quad \mathbf{r} = \mathbf{r}' + \mathbf{m}; \quad d\tau = d\tau', \quad (18.11)$$

тогда

$$\begin{aligned} & \int \psi_a^*(\mathbf{r} - \mathbf{m}) \psi_a(\mathbf{r} - \mathbf{n}) d\tau = \\ & = \int \psi_a^*(\mathbf{r}') \psi_a[\mathbf{r}' - (\mathbf{n} - \mathbf{m})] d\tau'. \end{aligned} \quad (18.12)$$

Интеграл от произведения волновых функций $\psi_a^*(\mathbf{r}')$ и $\psi_a[\mathbf{r}' - (\mathbf{n} - \mathbf{m})]$ зависит не от положения ядер \mathbf{n} и \mathbf{m} , а только от расстояния $\mathbf{t} = \mathbf{n} - \mathbf{m}$ между ними. Обозначим

$$\int \psi_a^*(\mathbf{r}') \psi_a(\mathbf{r}' - \mathbf{t}) d\tau' = S_t. \quad (18.13)$$

При $\mathbf{t} = 0$ $S_t = 1$. Если расстояние между ядрами велико, то волновые функции не перекрываются и S_t практически равны нулю.

Величины S_t будут отличны от нуля лишь для малых значений \mathbf{t} . Поскольку $\mathbf{n} - \mathbf{m}$ и \mathbf{t} пробегают одни и те же значения при фиксированном \mathbf{m} , то сумма по \mathbf{n} будет равна сумме по \mathbf{t} :

$$\sum_{\mathbf{n}} e^{i(\mathbf{k}, \mathbf{n} - \mathbf{m})} \int \psi_a^*(\mathbf{r} - \mathbf{m}) \psi_a(\mathbf{r} - \mathbf{n}) d\tau = \sum_{\mathbf{t}} e^{i(\mathbf{k}\mathbf{t})} S_t. \quad (18.14)$$

Под знаком второй суммы стоят одинаковые члены вида $\left[\sum_{\mathbf{t}} e^{i(\mathbf{k}\mathbf{t})} S_t \right]$, их число равно N — числу атомов в кристалле. Следовательно,

$$\int \psi^{*0}(\mathbf{r}) \psi^0(\mathbf{r}) d\tau = N \sum_{\mathbf{t}} e^{i(\mathbf{k}\mathbf{t})} S_t. \quad (18.15)$$

Перейдем к нахождению вида оператора возмущения.

Предположим, что кристалл «растянут» столь сильно, что взаимодействие между атомами отсутствует. В этом случае *периодическое потенциальное поле кристалла есть периодически повторяющееся потенциальное поле изолированного атома* (рис. 21), поэтому

$$U(\mathbf{r}) = \sum_{\mathbf{n}} V_a(\mathbf{r} - \mathbf{n}). \quad (18.16)$$

Отличительной особенностью этой суммы является то, что в каждой точке \mathbf{r} функция $U(\mathbf{r})$ определяется только потенциальной кривой ближайшего атома. При сближении атомов кристалла в каждой точке \mathbf{r} сумма потенциальных кривых изолированных атомов будет по-прежнему давать потенциальную энергию электрона в поле всех

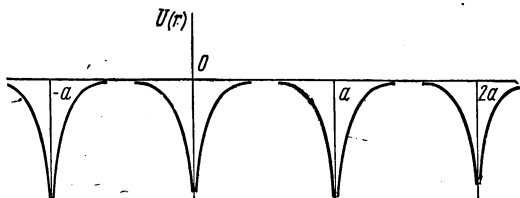


Рис. 21. Поле решетки «растянутого» кристалла

ядер (рис. 22), но при этом не учитывается взаимодействие атомов, другими словами, сумма не является самосогласованной. Пусть самосогласование потенциальной энергии поля решетки достигается посредством добавления некоторой энергии $W(r)$ таким образом, что

$$U(r) = \sum_n V_a(r-n) + W(r). \quad (18.7)$$

Функция $W(r)$ является периодической, так как $U(r)$ и $\sum_n V_a(r-n)$ — функции периодические; $W(r)$ не имеет особых точек; особые точки $U(r)$ передаются атомными потенциальными кривыми $V_a(r-n)$. Другими словами, $W(r)$ близка к нулю, $W(r)$ всюду ограничена. Аналитическое выражение для $W(r)$ находится элементарно:

$$W(r) = U(r) - \sum_n V_a(r-n) \quad (18.18)$$

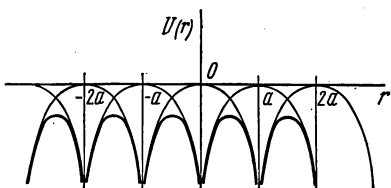


Рис. 22. Поле решетки кристалла без учета взаимодействия электронных оболочек

по «известным» $U(r)$ и $V_a(r-n)$.

Величина $W(r)$ будет меньше нуля, пока атомы притягиваются, и станет больше нуля, когда атомы начнут отталкиваться при очень сильном сжатии кристалла.

Перейдем к нахождению энергии в нулевом приближении ψ^0 позволяет получить значение энергии в первом приближении $E^{(1)}$:

$$\left[-\frac{\hbar^2}{2m} \Delta + \sum_n V_a(r-n) + W(r) \right] \psi^0(r) = E^{(1)} \psi^0(r). \quad (18.19)$$

Умножая уравнение (18.19) на $\psi^{0*}(r)$ и интегрируя его по объему кристалла, получим

$$E^{(1)} = \frac{\int \psi^{0*}(r) \left[-\frac{\hbar^2}{2m} \Delta + \sum_n V_a(r-n) + W(r) \right] \psi^0(r) dt}{\int \psi^{0*}(r) \psi^0(r) dt}. \quad (18.20)$$

Знаменатель выражения (18.20) был ранее найден (см. 18.15). Вычислим числитель, который обозначим через I:

$$\begin{aligned} I &= \int \psi^{0*}(r) \left[-\frac{\hbar^2}{2m} \Delta + \sum_n V_a(r-n) + W(r) \right] \psi^0(r) dt = \\ &= \int \left\{ \left[\sum_t e^{-i(\kappa t)} \psi_a^*(r-t) \right] \left[-\frac{\hbar^2}{2m} \Delta + \sum_n V_a(r-n) + W(r) \right] \times \right. \\ &\quad \times \left. \left[\sum_t e^{i(\kappa t)} \psi_a(r-t) \right] \right\} dt = \left\{ \sum_{1t} e^{i(\kappa, t-1)} \int \psi_a^*(r-1) \times \right. \\ &\quad \times \left. \left[-\frac{\hbar^2}{2m} \Delta + \sum_n V_a(r-n) + W(r) \right] \psi_a(r-t) \right\} dt. \quad (18.21) \end{aligned}$$

Положим

$$\begin{aligned} \mathbf{l} - \mathbf{t} = \mathbf{p}; \quad \mathbf{r} = \mathbf{r}' + \mathbf{l}; \quad \mathbf{r} - \mathbf{l} = \mathbf{r}'; \quad d\tau = d\tau'; \\ \mathbf{n}' = \mathbf{n} - \mathbf{l}; \quad W(\mathbf{r}) = W(\mathbf{r}'). \end{aligned} \quad (18.22)$$

В таком случае интеграл \mathbf{I} можно записать в виде одинарных сумм:

$$\begin{aligned} \mathbf{I} = N \left\{ \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} \int \psi_a^*(\mathbf{r}') \left[-\frac{\hbar^2}{2m} \Delta' + \sum_{\mathbf{n}'} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right] \times \right. \\ \left. \times \psi_a(\mathbf{r}' - \mathbf{p}) d\tau' \right\}, \end{aligned} \quad (18.23)$$

подобно тому, как это было сделано при вычислении нормировочного члена для $\psi^0(\mathbf{r})$.

Учитывая, что

$$\begin{aligned} \left[-\frac{\hbar^2}{2m} \Delta' + V_a(\mathbf{r}' - \mathbf{p}) + \left\{ \sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right\} \right] \psi_a(\mathbf{r}' - \mathbf{p}) = \\ = E_a \psi_a(\mathbf{r}' - \mathbf{p}) + \left\{ \sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right\} \psi_a(\mathbf{r}' - \mathbf{p}), \end{aligned} \quad (18.24)$$

можем записать

$$\begin{aligned} \mathbf{I} = N \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} \int \psi_a^*(\mathbf{r}') \left[E_a + \left\{ \sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}' - \mathbf{p}) \right\} \right] \times \\ \times \psi_a(\mathbf{r}' - \mathbf{p}) d\tau' = NE_a \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} S_{\mathbf{p}} + \\ + N \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} \int \psi_a^*(\mathbf{r}') \left\{ \sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right\} \psi_a(\mathbf{r}' - \mathbf{p}) d\tau'. \end{aligned} \quad (18.25)$$

Выделим во втором слагаемом член с $\mathbf{p} = 0$ и обозначим его через C :

$$C = \int \psi_a^*(\mathbf{r}') \left\{ \sum_{\mathbf{n}' \neq 0} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right\} \psi_a(\mathbf{r}') d\tau'. \quad (18.26)$$

Величина C представляет собой усредненную для некоторого атома потенциальную энергию электрона в поле всех остальных ядер с учетом самосогласованного характера поля. Теперь выражение для \mathbf{I} можно представить в следующей форме:

$$\begin{aligned} \mathbf{I} = NE_a \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} S_{\mathbf{p}} + NC + N \sum_{\mathbf{p}} e^{i(\kappa\mathbf{p})} \int \psi_a^*(\mathbf{r}') \times \\ \times \left[\sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right] \psi_a(\mathbf{r}' - \mathbf{p}) d\tau'. \end{aligned} \quad (18.27)$$

Обозначим через $A(\mathbf{p})$ интеграл в (18.27):

$$\int \psi_a^*(\mathbf{r}') \left[\sum_{\mathbf{n}' \neq \mathbf{p}} V_a(\mathbf{r}' - \mathbf{n}') + W(\mathbf{r}') \right] \psi_a(\mathbf{r}' - \mathbf{p}) d\tau' = A(\mathbf{p}). \quad (18.28)$$

$A(\mathbf{p})$ есть обменная энергия, обусловленная тем, что каждый электрон может быть найден с некоторой вероятностью у любого ядра. Это проявляется в том, что в образовании $A(\mathbf{p})$ участвуют волновые функции двух атомов, отстоящих на расстоянии $|\mathbf{p}|$. Другими словами, два ядра, отстоящие друг от друга на расстоянии $|\mathbf{p}|$, могут обмениваться электронами. Обмен происходит посредством поля всех ядер, кроме двух данных ядер, и посредством периодической самосогласующей части поля решетки $W(\mathbf{r})$. Очевидно, что $A(\mathbf{p})$ существенно только при малых расстояниях $|\mathbf{p}|$, так как в силу экспоненциального характера волновых функций $A(\mathbf{p}) \approx 0$ при больших \mathbf{p} . Физически это означает, что обмен в основном происходит только между ближайшими атомами. Два любых атома обмениваются электронами посредством последовательных обменов соседних атомов. Другими словами, электроны не локализируются у отдельных ядер, а «свободно» перемещаются по кристаллу, переходя от атома к атому благодаря обменному эффекту. Кроме того, следует заметить, что в образовании $A(\mathbf{p})$ основную роль должно играть периодическое поле $W(\mathbf{r})$, достигающее заметной величины в окрестности любого ядра, в то время как $V_\alpha(\mathbf{r}' - \mathbf{n}')$ при $\mathbf{n}' \neq 0$ и $\mathbf{n}' \neq \mathbf{p}$ — величина малая в окрестности «обменивающихся» ядер. Учитывая выражение для I можем записать выражение для энергии $E^{(1)}$:

$$E^{(1)} = \frac{N \left[E_\alpha \sum_{\mathbf{p}} e^{i(\mathbf{kp})} S_{\mathbf{p}} + C + \sum_{\mathbf{p}} e^{i(\mathbf{kp})} A(\mathbf{p}) \right]}{N \sum_{\mathbf{p}} e^{i(\mathbf{kp})} S_{\mathbf{p}}} =$$

$$= E_\alpha + C \left[\sum_{\mathbf{p}} e^{i(\mathbf{kp})} S_{\mathbf{p}} \right]^{-1} + \frac{\sum_{\mathbf{p}} e^{i(\mathbf{kp})} A(\mathbf{p})}{\sum_{\mathbf{p}} e^{i(\mathbf{kp})} S_{\mathbf{p}}}. \quad (18.29)$$

Мы видим, что в нулевом приближении энергия электронов системы атомов равна энергии электрона изолированного атома E_α . При сближении атомов (т. е. при учете взаимодействия электронов с ядрами и электронами других атомов) уровни энергии опускаются на величину C и расщепляются в полосу некоторой ширины. Учитывая, что $S_{\mathbf{p}} \approx 0$ при $\mathbf{p} \neq 0$, можно записать

$$E^{(1)} = E_\alpha + C + \sum_{\mathbf{p}} e^{i(\mathbf{kp})} A(\mathbf{p}). \quad (18.30)$$

В качестве первого примера рассмотрим простую кубическую решетку. Каждый атом имеет шесть ближайших атомов с координатами

$$\mathbf{p} = a \left\{ \begin{array}{l} (1, 0, 0); (-1, 0, 0) \\ (0, 1, 0); (0, -1, 0) \\ (0, 0, 1); (0, 0, -1) \end{array} \right\}. \quad (18.31)$$

Считая обменную энергию изотропной: $A(\mathbf{p}) = A(|\mathbf{p}|) = A$ и пренебрегая обменом электронами с более удаленными атомами, получим

$$E^{(1)} = E_a + C + 2A(\cos \kappa_x a + \cos \kappa_y a + \cos \kappa_z a) = E(\mathbf{k}). \quad (18.32)$$

Энергия зависит квазинепрерывно от волнового вектора \mathbf{k} и меняется в интервале от E_{\min} до E_{\max} :

$$\begin{aligned} E_{\min} &= E_a + C - 6|A|, \\ E_{\max} &= E_a + C + 6|A|. \end{aligned} \quad (18.33)$$

Таким образом, в результате взаимодействия атомов уровень энергии изолированного атома E_a опускается на величину C и рас-

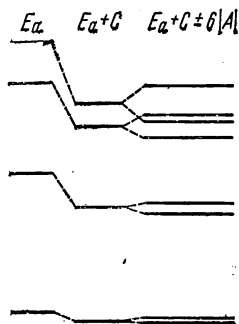


Рис. 23. Образование зон энергии из энергетических уровней при сближении атомов

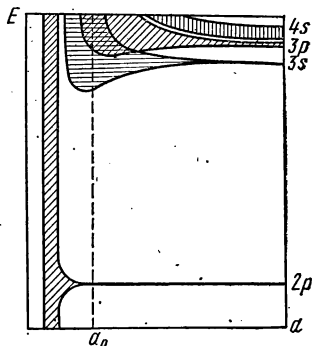


Рис. 24. Образование зон энергии при сближении атомов

цепляется в полосу, или зону, шириной $12|A|$ (для простой кубической решетки). Энергетические полосы разделены запрещенными интервалами энергии — запрещенными зонами. Ширина энергетической зоны зависит от величины обменной энергии A . Но обменная энергия зависит от области перекрытия волновых функций — чем сильнее перекрываются волновые функции атомов, тем больше обменная энергия. Отсюда следует, что уровни энергии внутренних электронных оболочек расщепляются меньше, чем наружных, поскольку электроны внутренних оболочек локализованы в меньших по размеру областях пространства. На рис. 23 представлена схема образования зон из атомных уровней.

Из рис. 23 видно, что с ростом энергии ширина зон энергии увеличивается, а ширина запрещенных зон уменьшается. Уровень опускается и расщепляется тем сильнее, чем выше он находится.

На рис. 24 показана схема образования зон энергии из некоторых энергетических уровней при сближении атомов, т. е. при изменении периода решетки. Через a_0 обозначен период решетки реального кристалла.

Для тетрагональной или ромбической кристаллических систем расстояния между атомами в различных направлениях разные, поэтому и обменный интеграл может зависеть от направления. Обозначив координаты соседних атомов \mathbf{p} в виде

$$\mathbf{p} = \left\{ \begin{array}{ll} (a_1, 0, 0); & (-a_1, 0, 0) \\ (0, a_2, 0); & (0, -a_2, 0) \\ (0, 0, a_3); & (0, 0, -a_3) \end{array} \right\} \quad (18.34)$$

и соответствующие этим направлениям обменные энергии через A_1, A_2, A_3 , можем записать

$$E(\mathbf{k}) = E_a + C + 2A_1 \cos \kappa_x a_1 + 2A_2 \cos \kappa_y a_2 + 2A_3 \cos \kappa_z a_3. \quad (18.35)$$

Найдем экстремальные точки, полагая $dE(\mathbf{k})/d\mathbf{k} = 0$. Легко видеть, что экстремумы энергии находятся в точках $\kappa_{0i} = 0$ и $\kappa_{0i} = \pm \pi/a_i$, т. е. в центре и вершинах зоны Бриллюэна.

Разлагая $E(\mathbf{k})$ в ряд Тейлора относительно точек \mathbf{k}_0 , получим в общем виде

$$E(\mathbf{k}) = E_{\text{extr}} + \frac{1}{2} \frac{d^2 E}{d\mathbf{k}^2} (\mathbf{k} - \mathbf{k}_0)^2 + \dots, \quad (18.36)$$

или

$$E(\mathbf{k}) = E_0 + C + 2 \sum_{i=1}^3 A_i + \frac{1}{2} \sum_{i=1}^3 \frac{\hbar^2}{m_i} \kappa_i^2 (\kappa_{0i} = 0) \quad (18.37)$$

и

$$E(\mathbf{k}) = E_0 + C - 2 \sum_{i=1}^3 A_i + \frac{1}{2} \sum_{i=1}^3 \frac{\hbar^2}{m_i} \left(\kappa_i \pm \frac{\pi}{a_i} \right)^2 \quad \left(\kappa_{0i} = \pm \frac{\pi}{a_i} \right). \quad (18.38)$$

Для эффективной массы \mathbf{m}^* имеем выражение

$$\mathbf{m}^{*-1} = \pm \begin{pmatrix} \frac{2A_1 a_1^2}{\hbar^2} & 0 & 0 \\ 0 & \frac{2A_2 a_2^2}{\hbar^2} & 0 \\ 0 & 0 & \frac{2A_3 a_3^2}{\hbar^2} \end{pmatrix}, \quad (18.39)$$

где минус соответствует точке $\mathbf{k}_0 = 0$, а плюс — точкам $(\pm \pi/a_1; \pm \pi/a_2; \pm \pi/a_3) = \pi(b_1, b_2, b_3)$.

Знак эффективной массы определяется знаком обменной энергии. При $A_i > 0$ в центре зоны Бриллюэна находится максимум энергии, минимум энергии находится в вершинах зоны Бриллюэна. При $A_i < 0$ максимум энергии будет находиться в вершинах, а минимум — в центре зоны Бриллюэна.

Энергия является действительно квадратичной функцией $(\mathbf{k} - \mathbf{k}_0)$ и периодична с периодом $2\pi b$, что непосредственно видно из аналитической зависимости $E(\mathbf{k})$.

Компоненты тензора эффективной массы в максимуме и минимуме отличаются только знаком, по модулю они равны. Другими словами, форма изоэнергетических поверхностей в окрестности максимума и минимума для одной и той же зоны одинакова. Компоненты тензора эффективной массы обратно пропорциональны обменной энергии:

$$m_i = \pm \frac{\hbar^2}{2a_i^2 A_i} \quad (18.40)$$

Так как обменная энергия определяет ширину зоны энергии, то можно сказать, что эффективная масса обратно пропорциональна ширине энергетической зоны. Это означает, что чем выше расположена зона энергии, тем меньше в ней эффективная масса носителей заряда.

Изоэнергетические поверхности в окрестности экстремумов являются эллипсоидами, например, при $A_i < 0$ ($\kappa_0 = 0$)

$$E(\kappa) = E_{\min} + \frac{\hbar^2}{2} \left(\frac{\kappa_x^2}{m_1} + \frac{\kappa_y^2}{m_2} + \frac{\kappa_z^2}{m_3} \right). \quad (18.41)$$

Максимумов для этого случая будет восемь в соответствии с числом вершин куба. Но так как каждая изоэнергетическая поверхность, построенная вокруг вершины куба, принадлежит первой зоне Бриллюэна на $1/8$, то на первую зону приходится только один полный эллипсоид энергии. Этот факт очень важен при подсчете плотности состояний в зоне энергии.

Для простой кубической решетки обменная энергия изотропна: $A(\mathbf{p}) = A(|\mathbf{p}|)$, поэтому эффективная масса является скаляром

$$m^* = \pm \hbar^2 / (2a^2 A). \quad (18.42)$$

Изоэнергетические поверхности в окрестности экстремума являются сферами. На рис. 25 представлена графическая зависимость $E(\kappa)$. Поверхности энергии имеют вид, который изображен на рис. 17.

Рассмотрим вопрос о ширине запрещенной зоны. Определим ширину запрещенной зоны ΔE_0 между n -й и $(n+1)$ -й энергетическими зонами как минимальное по энергии расстояние между ними:

$$\Delta E_0 = E_{(n+1)\min} - E_{n\max}. \quad (18.43)$$

Подставляя выражение (18.37) и (18.38) для экстремумов энергии в двух зонах, получим

$$\begin{aligned} \Delta E_0 &= \left[E_{a(n+1)} + C_{(n+1)} - 2 \sum_{i=1}^3 |A_{i(n+1)}| \right] - \left[E_{an} + C_n + 2 \sum_{i=1}^3 |A_{in}| \right] = \\ &= [E_{a(n+1)} - E_n] + [C_{(n+1)} - C_n] - 2 \sum_{i=1}^3 (|A_{i(n+1)}| + |A_{in}|). \end{aligned} \quad (18.44)$$

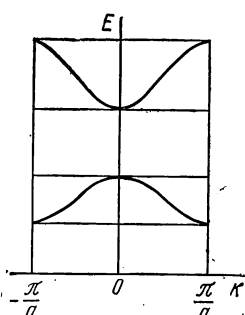


Рис. 25. Связь между энергией и волновым вектором при $A > 0$ и $A < 0$

Этот результат очень нагляден. Будем для простоты считать, что $C_{n+1} \approx C_n$. В таком случае ширина запрещенной зоны меньше расстояния между уровнями $E_{a(n+1)} - E_{an}$ на сумму полуширин соседних энергетических зон. Так как с ростом энергии расстояния между уровнями уменьшаются, а ширина зон энергии увеличивается, то *ширина запрещенных зон с ростом энергии уменьшается*. Этот результат является основой для объяснения зависимости ширины запрещенной зоны от порядкового номера элементов одной и той же группы. Например, ширина запрещенной зоны, отделяющей валентную зону от ближайшей свободной зоны для алмаза, кремния, германия и (серого) олова, должна уменьшаться в том же порядке, что соответствует экспериментальным данным.

Важное значение для объяснения многих явлений имеет знак A_i двух соседних зон. Если $A_{i(n+1)}$ и A_{in} имеют одинаковый знак, то максимум и минимум энергии для этих зон находятся в различных точках зоны Бриллюэна (центре и вершинах зоны). Если же $A_{i(n+1)}$ и A_{in} имеют разные знаки, то минимум и максимум находятся соответственно в одних и тех же точках зоны Бриллюэна (см. рис. 25). Каждой энергетической зоне соответствует своя зона Бриллюэна, которые считаем как бы наложенными одна на другую.

Косинусоидальная зависимость энергии E от \mathbf{k} для простой кубической решетки получена в результате учета обменной энергии $A(1, 0, 0)$ только для ближайших атомов. Если учесть более удаленные атомы, то выражение для $E(\mathbf{k})$ усложняется, однако основные качественные выводы при этом сохраняются. Учтем обменную энергию $A(1, 1, 0)$ с атомами, имеющими координаты $a(1, 1, 0)$; $(1, 0, 1)$ и т. д. Выражение для энергии будет при этом иметь вид

$$E(\mathbf{k}) = E_a + C + 2A(1, 0, 0)(\cos \kappa_x a + \cos \kappa_y a + \cos \kappa_z a) + 2A(1, 1, 0)[\cos(\kappa_x + \kappa_y) a + \cos(\kappa_x - \kappa_y) a + \cos(\kappa_x + \kappa_z) a + \cos(\kappa_x - \kappa_z) a + \cos(\kappa_y + \kappa_z) a + \cos(\kappa_y - \kappa_z) a]. \quad (18.45)$$

Если $|A(110)| \ll |A(100)|$, то характер изоэнергетических поверхностей изменится незначительно. Если же $|A(100)| \sim |A(110)|$, то изменения как в положении экстремумов, так и форме изоэнергетических поверхностей будут существенными.

Рассмотрим зависимость энергии от квазиимпульса для объемно-центрированной решетки, учитывая взаимодействие только ближайших соседних атомов. Если длину ребра куба обозначить через a , а начало координат совместить с атомом, находящимся в центре куба, то координаты атомов, лежащих в вершинах куба, равны

$$\mathbf{p} = \frac{a}{2} \left\{ \begin{array}{l} (1, 1, 1); (-1, -1, -1) \\ (-1, 1, 1); (1, -1, -1) \\ (1, -1, 1); (-1, 1, -1) \\ (1, 1, -1); (-1, -1, 1) \end{array} \right\} \quad (18.46)$$

Подставляя (18.46) в (18.30), получим

$$E(\mathbf{k}) = E_0 + C + 2A \left[\cos(\kappa_x + \kappa_y + \kappa_z) \frac{a}{2} + \cos(-\kappa_x + \kappa_y + \kappa_z) \frac{a}{2} + \right. \\ \left. + \cos(\kappa_x - \kappa_y + \kappa_z) \frac{a}{2} + \cos(\kappa_x + \kappa_y - \kappa_z) \frac{a}{2} \right] = \\ = E_a + C + 8A \cos \frac{\kappa_x a}{2} \cos \frac{\kappa_y a}{2} \cos \frac{\kappa_z a}{2}. \quad (18.47)$$

Как видно из выражения (18.47) для $E(\mathbf{k})$, экстремальные значения энергии определяются условием, что косинусы равны 1 по модулю, и

$$E_{\min} = E_a + C - 8|A|, \\ E_{\max} = E_a + C + 8|A|, \quad (18.48)$$

что дает для ширины зоны энергии величину $16|A|$.

Если построить в пространстве \mathbf{k} куб с ребром $4\pi/a$, то в нем будет находиться 27 экстремальных точек — в центре куба одна, в центре граней шесть, в вершинах куба восемь и в середине ребер двенадцать. Однако выделенный объем превосходит объем первой зоны Бриллюэна, представляющий собой додекаэдр.

Найдем зависимость энергии от волнового вектора для границентрированной решетки. Совмещая начало координат с атомом, находящимся в вершине куба с ребром a , запишем координаты 12 ближайших соседних атомов:

$$\mathbf{p} = \frac{a}{2} \left\{ \begin{array}{lll} (1, 1, 0); & (1, 0, 1); & (0, 1, 1) \\ (-1, 1, 0); & (1, 0, -1); & (0, -1, 1) \\ (1, -1, 0); & (-1, 0, 1); & (0, 1, -1) \\ (-1, -1, 0); & (-1, 0, -1); & (0, -1, -1) \end{array} \right\} \quad (18.49)$$

Считая, как и прежде, обменную энергию изотропной, можем записать:

$$E(\mathbf{k}) = E_a + C + 2A \left[\cos(\kappa_x + \kappa_y) \frac{a}{2} + \cos(\kappa_x - \kappa_y) \frac{a}{2} + \right. \\ \left. + \cos(\kappa_x + \kappa_z) \frac{a}{2} + \cos(\kappa_x - \kappa_z) \frac{a}{2} + \cos(\kappa_y + \kappa_z) \frac{a}{2} + \cos(\kappa_y - \kappa_z) \frac{a}{2} \right] = \\ = E_a + C + 4A \left[\cos \frac{a\kappa_x}{2} \cos \frac{a\kappa_y}{2} + \cos \frac{a\kappa_x}{2} \cos \frac{a\kappa_z}{2} + \cos \frac{a\kappa_y}{2} \cos \frac{a\kappa_z}{2} \right]. \quad (18.50)$$

Предположим для определенности, что $A > 0$. В таком случае в центре зоны Бриллюэна находится максимум энергии величиной

$$E_{\max} = E_a + C + 12A. \quad (18.51)$$

Минимум энергии не может быть равен $E_a + C - 12A$, поскольку для этого попарные произведения косинусов должны быть равны -1 , что невозможно. Записав условия экстремума для (18.51), найдем

$$E_{\min} = E_a + C - 4A, \quad (18.52)$$

т. е. ширина зоны энергии равна $16|A|$, а не $24|A|$.

В заключение этого параграфа остановимся на одном моменте, связанном с вырождением атомных уровней. s -уровень энергии атома является простым, все остальные уровни энергии вырожденные, фактор вырождения (без учета спина) равен при этом $g = (2l + 1)$, где l — орбитальное квантовое число, поэтому p -уровни имеют трехкратное вырождение, d -уровни — пятикратное вырождение и т. д. Благодаря взаимодействию орбитального и спинового магнитных моментов электрона уровни энергии имеют так называемую тонкую структуру.

Но если уровни энергии атома вырождены, то волновая функция электрона в кристалле должна строиться с учетом вырождения атомных состояний:

$$\psi^0(\mathbf{r}) = \sum_{\mathbf{n}, \alpha} e^{i(\mathbf{k}\mathbf{n})} c_{\alpha} \psi_{\alpha}(\mathbf{r} - \mathbf{n}), \quad (18.53)$$

где c_{α} — неизвестный коэффициент, α принимает g различных значений. Подставляя (18.53) в (18.19), можно вычислить энергию E . Отличие последнего случая от рассмотренного ранее состоит в том, что теперь интегралы, вычисленные с помощью различных атомных волновых функций, будут различными. В результате этого вырождение может быть частично или полностью снято, и для p -зоны можно получить три различные ветви энергии $E(\mathbf{k})$.

Резюме § 18

1. В теории квазисвязанного электрона в качестве нулевого приближения для гамильтониана поля решетки принимается гамильтониан электрона в изолированном атоме. Оператор возмущения представляет собой энергию взаимодействия атомов друг с другом и энергию электрона в поле всех остальных ядер. Волновая функция электрона в кристалле представляет собой линейную комбинацию атомных волновых функций, удовлетворяющую трансляционному условию.

2. «Возмущение», т. е. взаимодействие электрона данного атома со всеми ядрами и электронными оболочками, приводит к тому, что уровни энергии E_a электрона в атоме существенно видоизменяются: они опускаются вниз и расщепляются в полосу, или зону энергии.

3. Уровень E_a опускается тем сильнее, чем выше он расположен в изолированном атоме, поскольку с ростом энергии E_a размеры электронного облака увеличиваются.

4. Если атомы кристалла сближать неограниченно, то начиная с некоторых расстояний между электронными оболочками атомов возникнут силы отталкивания, потенциальная энергия $W(r)$ станет положительной, быстро возрастающей величиной, что приведет к подъему уровня энергии E_a ($C > 0$).

5. Ширина зоны энергии пропорциональна обменному интегралу, который возрастает с ростом E_a , поэтому чем выше расположена зона энергии, тем она шире (см. рис. 23).

6. Ширина запрещенной зоны тем меньше, чем выше расположены зоны энергии. Уменьшение ширины запрещенных зон с ростом энергии обусловлено уменьшением расстояний между соответствующими уровнями энергии и возрастанием ширины зоны энергии. При больших значениях энергии зоны энергии могут накладываться, перекрываться (см. рис. 23).

7. Эффективная масса электрона обратно пропорциональна обменному интегралу и тем самым ширине зоны энергии. Поскольку при прочих равных условиях увеличение ширины зоны энергии приводит к уменьшению ширины запрещенной зоны, можно ожидать, что в веществах с меньшей шириной запрещенной зоны эффективная масса электрона должна быть меньше.

8. Воздействия, приводящие к изменению расстояния между атомами (нагрев, сжатие или растяжение), меняют область перекрытия волновых функций, что приводит к изменению обменного интеграла, благодаря чему меняется ширина зоны энергии, эффективная масса и ширина запрещенных зон. Это в свою очередь приводит к изменению физических свойств полупроводников, зависящих от ширины запрещенной зоны или эффективной массы.

9. Зонная структура спектра энергии вытекает как из теории квазисвободного, так и из теории квазисвязанного электрона. При этом выводы первой теории применимы с большей точностью в области больших энергий, а выводы второй теории оправдываются при малых энергиях.

§ 19. МЕТОД ЭФФЕКТИВНОЙ МАССЫ. ВЛИЯНИЕ ВНЕШНИХ ПОЛЕЙ НА СПЕКТР ЭНЕРГИИ КРИСТАЛЛА

Перейдем к рассмотрению решений уравнения Шредингера для случая, когда на кристалл наложено внешнее поле $V(r)$:

$$\hat{H}\psi = \left[-\frac{\hbar^2}{2m} \Delta + U(r) + V(r) \right] \psi(r) = E\psi(r). \quad (19.1)$$

Чтобы решить это уравнение, необходимо знать поле $U(r)$, которое нам неизвестно. Однако существует удобный и достаточно точный метод решения уравнения (19.1), который получил название метода эффективной массы. Чтобы понять его, рассмотрим вновь решение уравнения Шредингера для идеального кристалла. Как известно, волновая функция электрона в идеальном кристалле пред-

ставляет собой функцию Блоха $\psi_{\mathbf{k}}(\mathbf{r})$, а энергия электрона в окрестности экстремума \mathbf{k}_0 является квадратичной функцией квазиимпульса или волнового вектора:

$$\hat{H}_0 \psi_{\mathbf{k}}(\mathbf{r}) = \left[-\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) \right] \psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k}) \psi_{\mathbf{k}}(\mathbf{r});$$

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2}{2m^*} (\mathbf{k} - \mathbf{k}_0)^2, \quad (19.2)$$

Рассмотрим гамильтониан \hat{H}_0 :

$$\hat{H}_0 = -\frac{\hbar^2}{2m^*} \Delta, \quad (19.3)$$

представляющий собой гамильтониан свободной частицы, масса которой совпадает с эффективной массой электрона. Решим уравнение Шредингера

$$\hat{H}_0 \tilde{\psi}_0 = \tilde{E}_0 \tilde{\psi}_0. \quad (19.4)$$

Подставляя вместо \hat{H}_0 его выражение, можем записать

$$\sum_{i=1}^3 \frac{1}{m_i} \frac{\partial^2}{\partial x_i^2} \tilde{\psi}_0 + \frac{2E_0}{\hbar^2} \tilde{\psi}_0 = 0. \quad (19.5)$$

Непосредственной подстановкой

$$\tilde{\psi}_0 = A e^{i(\alpha x + \beta y + \gamma z)} \quad (19.6)$$

в это уравнение можно убедиться, что $\tilde{\psi}_0$ является решением уравнения (19.4) и (19.5), причем

$$\tilde{E}_0 = \frac{\hbar^2}{2} \left(\frac{\alpha^2}{m_1} + \frac{\beta^2}{m_2} + \frac{\gamma^2}{m_3} \right). \quad (19.7)$$

Сравнивая (19.7) с (19.2), приходим к выводу, что уравнение (19.4) имеет тот же спектр энергии, что и уравнение (19.2), если выбрать α , β , γ в виде

$$\alpha = \kappa_x - \kappa_{0x}; \quad \beta = \kappa_y - \kappa_{0y}; \quad \gamma = \kappa_z - \kappa_{0z}, \quad (19.8)$$

а начало отсчета энергии совместить с точкой экстремума, т. е. положить $E(\mathbf{k}_0) = 0$. В общем случае гамильтониан

$$\hat{H}_0 = -\frac{\hbar^2}{2m^*} \Delta + E(\mathbf{k}_0) \quad (19.9)$$

имеет тот же спектр энергии, что и гамильтониан поля решетки в окрестности экстремума, а его собственные волновые функции имеют вид

$$\tilde{\psi}_0(\mathbf{r}) = \frac{1}{\sqrt{G}} e^{i(\mathbf{k} - \mathbf{k}_0, \mathbf{r})}. \quad (19.10)$$

Таким образом, движение частицы в кристалле подобно движению свободной частицы. Различие между действительно свободной части-

цей и частицей в кристалле заложено в различиях масс: вместо массы свободной частицы необходимо использовать эффективную массу m^* .

При наложении внешних полей $V(\mathbf{r})$ уравнение Шредингера (19.1) может быть записано в виде

$$\left[-\frac{\hbar^2}{2m^*} \Delta + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (19.11)$$

Решив это уравнение, найдем волновую функцию $\psi(\mathbf{r})$ и значение энергии E электрона в кристалле, находящемся во внешнем поле. Решение задач на основе уравнения (19.11) получило название метода эффективной массы. Его преимуществом по сравнению с общим уравнением (19.1) является то, что вместо неизвестного поля решетки $U(\mathbf{r})$ мы можем воспользоваться экспериментально определенной эффективной массой m^* частицы. Недостатком его является приближенный характер, поскольку оно имеет смысл лишь для состояний в окрестности экстремумов.

Используя функции Ванье, можно показать более строго, что уравнение (19.1) не приводит к существенному изменению спектра энергии в плавно меняющихся внешних полях, а волновая функция действительно имеет вид (19.10).

Чтобы найти изменения, возникающие в спектре энергии идеального кристалла под действием внешнего поля, воспользуемся не обычной теорией возмущений, а квантовым уравнением движения. Рассмотрим гамильтониан кристалла во внешнем поле (19.1). Квантовое уравнение движения для произвольной, не зависящей явно от времени физической величины с оператором \hat{L} , имеет вид

$$\frac{d\hat{L}}{dt} = [\hat{H}, \hat{L}] = \frac{1}{i\hbar} (\hat{L}\hat{H} - \hat{H}\hat{L}). \quad (19.12)$$

Применим это уравнение для гамильтониана идеального кристалла

$$\hat{H}_0 = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r});$$

$$\begin{aligned} \frac{d\hat{H}_0}{dt} &= [\hat{H}, \hat{H}_0] = [\hat{V}, \hat{H}_0] = [\hat{V}, \hat{T}] = \frac{1}{i\hbar} \left(-\frac{\hbar^2}{2m} \right) \{ \Delta \hat{V} - \hat{V} \Delta \} = \\ &= \frac{i\hbar}{2m} \{ (\Delta V) + 2(\nabla V, \nabla) \}. \end{aligned} \quad (19.13)$$

Остановимся на полях, слабо меняющихся в пространстве; для них величиной ΔV можно пренебречь по сравнению со вторым слагаемым. Учитывая, что $(\nabla V) = -\mathbf{F}_a$, а

$$-\frac{i\hbar}{m} \nabla = \frac{\hat{\mathbf{p}}}{m} = \hat{\mathbf{v}} = \frac{d\hat{\mathbf{r}}}{dt}, \quad (19.14)$$

запишем

$$\frac{d\hat{H}_0}{dt} = \left(\mathbf{F}_a, \frac{d\hat{\mathbf{r}}}{dt} \right). \quad (19.15)$$

Усредним это выражение по исходным состояниям с волновой функцией $\psi_{\mathbf{k}}(\mathbf{r})$. Предполагая, что F_a не зависит от координаты, можем вынести F_a из-под знака интеграла:

$$\begin{aligned} \left\langle \frac{d\hat{H}_0}{dt} \right\rangle &= \frac{d}{dt} \int \psi_{\mathbf{k}}^*(\mathbf{r}) \hat{H}_0 \psi_{\mathbf{k}}(\mathbf{r}) d\tau = \frac{dE_0(\mathbf{k})}{dt} = \\ &= \int \psi_{\mathbf{k}}^*(\mathbf{r}) (\hat{F}_a \hat{\mathbf{v}}) \psi_{\mathbf{k}}(\mathbf{r}) d\tau = (F_a \langle \mathbf{v} \rangle). \end{aligned} \quad (19.16)$$

Уравнение (19.16) является классическим, и оно показывает, что «полная» энергия $E_0(\mathbf{k})$ меняется со временем в результате работы внешней силы F_a . Его можно получить из уравнения движения

$$d\mathbf{P}/dt = F_a = -\nabla V. \quad (19.17)$$

Умножим (19.17) на скорость $\mathbf{v} = d\mathbf{r}/dt$:

$$\left(\mathbf{v} \frac{d\mathbf{P}}{dt} \right) = \left(\frac{\mathbf{P} - \mathbf{P}_0}{m^*} \frac{d\mathbf{P}}{dt} \right) = \frac{d}{dt} \frac{(\mathbf{P} - \mathbf{P}_0)^2}{2m^*} = \frac{d}{dt} E_0(\mathbf{P}) = -\frac{dV}{dr} \frac{dr}{dt} = -\frac{dV}{dt}, \quad (19.18)$$

или

$$E_0(\mathbf{P}) + V(\mathbf{r}) = H_0(\mathbf{P}, \mathbf{r}) = \text{const.} \quad (19.19)$$

Через $H_0(\mathbf{P}, \mathbf{r})$ мы обозначили сохраняющуюся во времени функцию Гамильтона электрона в кристалле, на который наложено внешнее потенциальное поле $V(\mathbf{r})$. Уравнению (19.19) как раз и соответствует уравнение (19.11), благодаря чему метод эффективной массы получает строгое квантовомеханическое обоснование.

Обратим внимание, что при движении частицы в кристалле под действием внешнего поля его полная энергия H_0 сохраняется, в то время как $E_0(\mathbf{P})$ и $V(\mathbf{r})$ меняются, причем в противоположных направлениях. Но если $V(\mathbf{r})$ может меняться в любых пределах, то $E_0(\mathbf{P})$ может меняться только в пределах зоны энергии: от E_{\min} до E_{\max} . Но отсюда вытекает, что изменение $V(\mathbf{r})$ ограничено тем же интервалом; действительно, возьмем дифференциалы от левой и правой частей уравнения (19.19), получим

$$dE_0 + dV = 0, \text{ или } \delta E_0 + \delta V = 0. \quad (19.20)$$

Полагая $\delta E_0 = E_{\max} - E_{\min}$, получим

$$\delta V = -\delta E_0 = E_{\min} - E_{\max}. \quad (19.21)$$

Из (19.21) также вытекает, что в однородном электрическом поле электрон должен совершать колебательное периодическое движение вдоль поля на участке длиной

$$\delta V = -(\delta r F_a) = -(E_{\max} - E_{\min}), \quad (19.22)$$

откуда

$$\delta r = \frac{E_{\max} - E_{\min}}{F_a} = \frac{E_{\max} - E_{\min}}{eE}. \quad (19.23)$$

Вернемся к уравнению (19.19). Зафиксируем значение $P=P'$ и рассмотрим возможные значения функции Гамильтона H_0 при изменении координаты. Поскольку $E_0(P)$ от координаты не зависит, то зависимость H_0 от r определяется внешним полем $V(r)$ (рис. 26, а). Рис. 26, б можно рассматривать как график зависимости уровня энергии $E_0(P')$ от координаты.

Если теперь задать все возможные значения энергии $E_0(P')$, то получим график зависимости положения зон энергии от координаты (рис. 27).

Таким образом, график $H_0(r)$ можно интерпретировать как искривление зон энергии под действием внешнего поля. Такая интерпретация зависимости $H_0(r)$ очень удобна, поскольку она позволяет рассматривать как «свободное» движение электрона, так и движение с рассеянием.

Рассмотрим «свободное» движение, которое описывается уравнением (19.19). При движении электрона под действием внешнего поля $V(r)$ полная энергия электрона сохраняется: $H_0 = \text{const}$. На рис. 27 это изображается прямой, параллельной оси абсцисс. Из рисунка видно, что электрон переходит с одного уровня зоны на

другие уровни, совершая периодическое движение в кристалле. Электрон с одной и той же полной энергией $H_0 = \text{const}$ может совершать колебательное периодическое движение между точками А—В, С—D и G—K двух зон. В соответствии с (19.23) размеры области δr будут тем меньше, чем больше наклон зон (сравните С—D и А—В) и чем уже зона энергии (сравните А—В и G—K). Однако электрон может перейти из точек отрезка G—K в точки отрезка А—В и тем самым он может перейти из второй зоны в третью, т. е. в более «высокую» зону энергии. Этот переход возможен благодаря туннельному эффекту — переход из точки К в точку А связан с переходом сквозь треугольный потенциальный барьер KLA. Как известно из квантовой механики, вероятность туннельного перехода связана экспоненциально с шириной и высотой потенциального барьера. Высота потенциального барьера KL представляет собой ширину запрещенной зоны ΔE_0 . Ширина барьера KA зависит в свою оче-

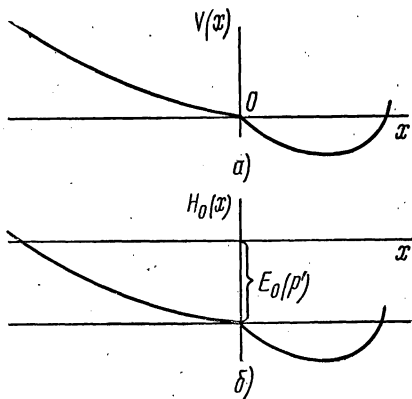


Рис. 26. Зависимость величины потенциальной энергии и функции Гамильтона от координат

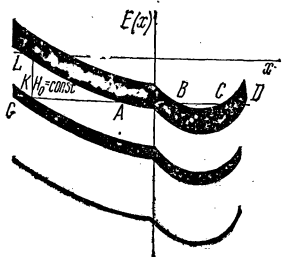


Рис. 27. Зависимость энергии электрона в зонах от координат

редь от KL и наклона зон, который определяется силой F_a , т. е. напряженностью поля E :

$$KA = KL/F_a = \Delta E_0 / |eE|. \quad (19.24)$$

Но так как высота барьера входит в показатель экспоненты в степени $1/2$, то приходим к выводу, что прозрачность треугольного барьера пропорциональна $\exp\left\{\gamma \frac{\Delta E_0^{3/2}}{|E|}\right\}$, т. е. вероятность туннельного перехода из зоны в зону растет экспоненциально с ростом напряженности внешнего электрического поля E , приводящего к наклону зон энергии. Вероятность перехода из точки K в точку A равна вероятности перехода из точки A в точку K , поэтому преимущественный переход будет иметь место из зоны, в которой концентрация электронов больше. Резкое возрастание числа электронов в верхней свободной зоне за счет нижней заполненной зоны под действием электрического поля носит название эффекта Зинера.

Туннельный переход имеет место и на участке BC . Теперь рассмотрим «несвободное» движение электрона, когда приобретаемая в поле $V(\mathbf{r})$ энергия передается, например, решетке. В этом случае электрон остается на одном и том же уровне зоны и может перемещаться вдоль кристалла на любые расстояния.

Резюме § 19

1. Для описания движения электрона в кристалле, на который наложено внешнее потенциальное поле $V(\mathbf{r})$, используется уравнение

$$\left[-\frac{\hbar^2}{2m^*} \Delta + V(\mathbf{r})\right] \psi(\mathbf{r}) = E\psi(\mathbf{r}), \quad (19.1p)$$

в которое переходит уравнение (19.1) при замене оператора периодического поля $\hat{H}_0 = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r})$ оператором свободной частицы $\hat{H}_0 = -\frac{\hbar^2}{2m^*} \Delta$ с массой, равной эффективной массе электрона в твердом теле. Этот метод решения уравнения Шредингера получил название метода эффективной массы.

2. Метод эффективной массы справедлив для невозмущенных состояний с энергией, близкой к экстремальному значению, при наложении на кристалл медленно меняющихся полей. Он позволяет исключить из уравнения (19.1) неизвестную величину $U(\mathbf{r})$ введением экспериментально измеримой величины эффективной массы m^* .

3. Внешние поля $V(\mathbf{r})$ искривляют зоны энергии идеального кристалла. Уровни энергии поднимаются в тех областях кристалла, где $V(\mathbf{r}) > 0$, и опускаются там, где $V(\mathbf{r}) < 0$. Смещение всех зон одинаково и равно $V(\mathbf{r})$ в каждой точке.

4. Наклон зон энергии в сильных электрических полях приводит к переходу электронов из «нижних» зон в «верхние» благодаря туннельному эффекту. Этот эффект носит название эффекта Зинера.

§ 20. ЛОКАЛИЗОВАННЫЕ СОСТОЯНИЯ

Предположим теперь, что на поле кристалла наложено достаточно сильное возмущение W , локализованное в небольшой области кристалла с «центром» в точке с координатой R . Обозначим его через $W(\mathbf{r}-R)$. Уравнение Шредингера имеет вид

$$\left[-\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) + W(\mathbf{r}) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r}), \quad (20.1)$$

или

$$[\hat{H}_0 + \hat{W}] \psi(\mathbf{r}) = E \psi(\mathbf{r}). \quad (20.2)$$

Разложив искомое решение $\psi(\mathbf{r})$ по функциям Блоха $\psi_{\mathbf{k}}(\mathbf{r})$:

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{r}) \quad (20.3)$$

и подставив (20.3) в уравнение (20.2), получим

$$\sum_{\mathbf{k}} c_{\mathbf{k}} [E_0(\mathbf{k}) + \hat{W}] \psi_{\mathbf{k}}(\mathbf{r}) = E \sum_{\mathbf{k}} c_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{r}). \quad (20.4)$$

Умножая на $\psi_{\mathbf{k}'}^*(\mathbf{r})$ и интегрируя по всему кристаллу, упростим уравнение, записав его в обычной матричной форме:

$$c_{\mathbf{k}'} E_0(\mathbf{k}') + \sum_{\mathbf{k}} W_{\mathbf{k}'\mathbf{k}} c_{\mathbf{k}} = E c_{\mathbf{k}'}, \quad (20.5)$$

или

$$[E - E_0(\mathbf{k}')] c_{\mathbf{k}'} - \sum_{\mathbf{k}} W_{\mathbf{k}'\mathbf{k}} c_{\mathbf{k}} = 0. \quad (20.6)$$

Вместо выписывания секулярного уравнения, из которого находится возможное значение энергии E , запишем уравнение (20.6), учитывая условие, наложенное на $W(\mathbf{r}-R)$: возмущение мало (или равно нулю) всюду, кроме малой области V_R вокруг точки R , поэтому при вычислении матричного элемента

$$W_{\mathbf{k}'\mathbf{k}} = \int_{(V_R)} \psi_{\mathbf{k}'}^*(\mathbf{r}) W(\mathbf{r}-R) \psi_{\mathbf{k}}(\mathbf{r}) d\tau \quad (20.7)$$

интегрирование производится по малому объему вокруг точки R . На основании теоремы о среднем вынесем волновые функции, взяв их в некоторой точке R' объема V_R :

$$W_{\mathbf{k}'\mathbf{k}} = \psi_{\mathbf{k}'}^*(R') \psi_{\mathbf{k}}(R') \int W(\mathbf{r}-R) d\tau = \psi_{\mathbf{k}'}^*(R') \psi_{\mathbf{k}}(R') W_0 V_R, \quad (20.8)$$

где

$$W_0 = \frac{1}{V_R} \int_{(V_R)} W(\mathbf{r}-R) d\tau \quad (20.9)$$

есть среднее значение энергии возмущения по объему V_R .

Подставляя выражение для $W_{\kappa'\kappa}$ в уравнение (20.6), получим

$$[E - E_0(\kappa')] c_{\kappa'} - \psi_{\kappa'}^*(R') W_0 V_R \sum_{\kappa} c_{\kappa} \psi_{\kappa}(R') = 0. \quad (20.10)$$

Но

$$\sum_{\kappa} c_{\kappa} \psi_{\kappa}(R') = \psi(R') \quad (20.11)$$

есть некоторая величина, не зависящая от κ и κ' , поэтому можем записать из (20.10) и (20.11) выражение для $c_{\kappa'}$:

$$c_{\kappa'} = \frac{\psi_{\kappa'}(R') W_0 V_R \psi(R')}{E - E_0(\kappa')}. \quad (20.12)$$

Подставляя теперь выражение для $c_{\kappa'}$ и c_{κ} в уравнение (20.10) и сокращая на $\psi_{\kappa'}$, $\psi(R')$ и $W_0 V_R$, запишем его в виде

$$1 - \sum_{\kappa} \frac{\psi_{\kappa}^*(R') \psi_{\kappa}(R') W_0 V_R}{E - E_0(\kappa)} = 0, \quad (20.13)$$

или

$$\sum_{\kappa} \frac{\psi_{\kappa}^*(R') \psi_{\kappa}(R')}{E - E_0(\kappa)} = \frac{1}{W_0 V_R}. \quad (20.14)$$

Мы видим, что собственные значения энергии возмущенной системы связаны с величиной возмущения посредством $W_0 V_R$ и его локализацией посредством значения функции в точке R' . Исследуем это уравнение. Оно является алгебраическим уравнением N -й степени относительно E , следовательно, оно имеет N корней. Пусть $W_0 V_R \rightarrow 0$, в этом случае правая часть стремится к ∞ . Так как $|\psi_{\kappa}(R')|^2$ есть постоянная величина, не зависящая от $W_0 V_R$, то, для того чтобы уравнение (20.14) удовлетворялось, необходимо, чтобы хотя бы одно из слагаемых (20.14) стремилось к бесконечности, или знаменатель одного из слагаемых стремился к нулю, например, при $\kappa = \kappa''$:

$$E - E_0(\kappa'') \rightarrow 0, \quad (20.15)$$

где κ'' может принимать одно из N возможных значений κ , благодаря чему E принимает одно из разрешенных значений энергии $E_0(\kappa'')$ невозмущенной системы. Другими словами, если возмущение W стремится к нулю, то решение «возмущенной» задачи, естественно, переходит в решение «невозмущенной» задачи.

Рассмотрим другой предельный случай: $W_0 V_R \rightarrow \pm \infty$. В этом случае правая часть (20.14) стремится к нулю. Для удобства исследования этого случая введем следующие обозначения: $\kappa = \kappa_1; \kappa_2; \dots; \kappa_N; E_0(\kappa_n) = a_n; E = x;$

$$|\psi_{\kappa_n}(R')|^2 = A_n; \frac{1}{W_0 V_R} = B \quad (20.16)$$

и рассмотрим функцию двух переменных x и B :

$$f(x, B) = \left[\sum_{n=1}^N \frac{A_n}{x-a_n} - B \right] \prod_{i=1}^N (x-a_i). \quad (20.17)$$

Возможные значения энергии E определяются условием $f(x, B) = 0$. Но $f(x, B) = 0$ удовлетворяется условиями:

$$\left[\sum_{n=1}^N \frac{A_n}{x-a_n} - B \right] = 0; \quad \prod_{i=1}^N (x-a_i) = 0. \quad (20.18)$$

Первое условие определяет корни уравнения возмущенной задачи, второе — невозмущенной, поэтому оно нас интересовать не будет, и будем считать в дальнейшем $\prod_{i=1}^N (x-a_i) \neq 0$. Представим $f(x, B)$ в другой форме:

$$\begin{aligned} f(x, B) &= -B \prod_{i=1}^N (x-a_i) + \sum_n \prod_{i \neq n}^N (x-a_i) A_n = \\ &= -Bx^N + \left[\sum_n A_n + B \sum_n a_n \right] x^{N-1} + \dots + \\ &+ \left[\sum_n \frac{A_n}{a_n} \prod_{i=1}^N a_i + B \prod_{i=1}^N a_i \right] x (-1)^{N+1}. \end{aligned} \quad (20.19)$$

Функция $f(x, B)$ является линейной по B и полиномом N -й степени по x , поэтому она при любом $B \neq 0$ имеет N корней.

При $x \rightarrow \infty$ основную роль будет играть член с высшей степенью, т. е. $-Bx^N$. Чтобы $f(x, B)$ оставалась конечной величиной, необходимо, чтобы $B \rightarrow 0$. Отсюда можно сделать другой вывод: если $B \rightarrow 0$, то $x \rightarrow \infty$, примерно как B^{-1} . Таким образом, $E \rightarrow \pm \infty$ при $W_0 V_R \rightarrow \pm \infty$. Возникает вопрос, все ли корни уравнения (20.19) уходят в бесконечность или не все?

Пусть $B = 0$, тогда

$$f(x, 0) = \left(\sum_{n=1}^N A_n \right) x^{N-1} + \dots + (-1)^{N+1} \prod_{i=1}^N a_i \sum_n \frac{A_n}{a_n}. \quad (20.20)$$

т. е. при $B = 0$ $f(x, 0)$ является полиномом $N-1$ степени по x и имеет $N-1$ корней, совпадающих с корнями невозмущенного уравнения.

Поэтому сделаем следующий вывод: при наложении на кристалл локального возмущения достаточной величины один из подуровней энергии невозмущенной системы отщепляется от совокупности уровней и опускается вниз или поднимается вверх. Совокупность из $N-1$ остальных уровней практически не меняет своего положения. При $W_0 V_R > 0$ вверх поднимается уровень максимального значения

энергии, и при $W_0 V_R < 0$ опускается вниз уровень минимальной энергии (рис. 28). Благодаря этому в запрещенной зоне появляется разрешенное состояние.

Рассмотрим волновую функцию состояния, находящегося в запрещенной зоне. Пусть энергия этого состояния $E_n > E_0(\kappa)$. Волновая функция $\psi_n(\mathbf{r})$ может быть записана следующим образом:

$$\begin{aligned} \psi_n(\mathbf{r}) &= \sum_{\kappa} c_{\kappa} \psi_{\kappa}(\mathbf{r}) = \sum_{\kappa} \frac{\psi_{\kappa}^*(\mathbf{R}') W_0 V_R \psi_n^*(\mathbf{R}')}{E_n - E_0(\kappa)} \psi_{\kappa}(\mathbf{r}) = \\ &= \psi_n(\mathbf{R}') W_0 V_R \sum_{\kappa} \frac{\psi_{\kappa}^*(\mathbf{R}')}{E_n - E_0(\kappa)} \psi_{\kappa}(\mathbf{r}). \end{aligned} \quad (20.21)$$

В образовании $\psi_n(\mathbf{r})$ участвуют все функции $\psi_{\kappa}(\mathbf{r})$ с весом $\frac{\psi_{\kappa}^*(\mathbf{R}')}{E_n - E_0(\kappa)}$.

Найдем предельный вид функции $\psi_n(\mathbf{r})$ при $W_0 V_R \rightarrow \infty$.

В этом случае

$$\frac{E_n}{W_0 V_R} \sim 1, \quad E_n \gg E_0(\kappa) \quad (20.22)$$

и

$$\begin{aligned} \psi_n(\mathbf{r}) &\approx \psi_n(\mathbf{R}') \sum_{\kappa} \psi_{\kappa}^*(\mathbf{R}') \psi_{\kappa}(\mathbf{r}) = \\ &= \psi_n(\mathbf{R}') \sum_{\kappa} e^{-i(\kappa \mathbf{R}')} \varphi_{\kappa}^*(\mathbf{R}') \psi_{\kappa}(\mathbf{r}). \end{aligned} \quad (20.23)$$

Рис. 28. Отщепление уровня энергии от зоны при наложении на кристалл локального возмущения

Если в последнем выражении заменить $\varphi_{\kappa}^*(\mathbf{R}')$ на некоторую максимальную величину $\varphi^*(\mathbf{R}')$, не зависящую от κ , то получим оценку суммы (20.23):

$$\psi_n(\mathbf{r}) \leq \psi_n(\mathbf{R}') \varphi^*(\mathbf{R}') \sum_{\kappa} e^{-i(\kappa \mathbf{R}')} \psi_{\kappa}(\mathbf{r}). \quad (20.24)$$

Введем обозначение

$$\sum_{\kappa} e^{-i(\kappa \mathbf{R}')} \psi_{\kappa}(\mathbf{r}) = \sqrt{N} \Phi(\mathbf{r} - \mathbf{R}'). \quad (20.25)$$

Функция $\Phi(\mathbf{r} - \mathbf{R}')$ есть функция Ванье. Она отлична от нуля в небольшой окрестности точки \mathbf{R}' . Действительно, $\Phi(\mathbf{r} - \mathbf{R}')$ можно представить в виде

$$\begin{aligned} \Phi(\mathbf{r} - \mathbf{R}') &= \frac{1}{\sqrt{N}} \sum_{\kappa} e^{i(\kappa, \mathbf{r} - \mathbf{R}')} \varphi_{\kappa}(\mathbf{r}) \approx \frac{1}{\sqrt{N}} \sum_{\kappa} e^{i(\kappa, \mathbf{r} - \mathbf{R}')} \varphi_{\kappa}(\mathbf{r} - \mathbf{R}') = \\ &= \frac{1}{\sqrt{N}} \sum_{\kappa} e^{i(\kappa \mathbf{r}')} \varphi_{\kappa}(\mathbf{r}') = \frac{1}{\sqrt{N}} \sum_{\kappa} \psi_{\kappa}(\mathbf{r}'). \end{aligned} \quad (20.26)$$

Но сумма очень большого числа периодических функций (плоских волн) дает величину, близкую к нулю всюду, кроме точки $\mathbf{r}' = 0$ и ее небольшой окрестности. Поэтому функция Ванье отлична от нуля в небольшой окрестности точки \mathbf{R}' .

Мы видим, что волновая функция $\psi_{\mathbf{l}}(\mathbf{r})$ пропорциональна функции Ванье $\Phi(\mathbf{r} - \mathbf{R}')$, т. е. $\psi_{\mathbf{l}}(\mathbf{r})$ отлична от нуля в малой области около \mathbf{R}' . Фактически же волновой пакет $\psi_{\mathbf{l}}(\mathbf{r})$ будет простирается на несколько большую область, чем функция Ванье в силу зависимости $\Phi_{\mathbf{k}}^*(\mathbf{R}')$ от \mathbf{k} , однако характер волновой функции остается тем же. Этот результат имеет очень важное значение — электрон с энергией $E_{\mathbf{l}}$, лежащей в запрещенной зоне, локализован в некотором объеме того же порядка, что и V_R . Другими словами, электрон локализован в области возмущения.

Для конечной величины $W_0 V_R$ область локализации электрона расширяется: чем меньше энергия возмущения, тем больше размеры области локализации. Это непосредственно видно из самого выражения для $\psi_{\mathbf{l}}(\mathbf{r})$: чем ближе $E_{\mathbf{l}}$ к $E_0(\mathbf{k})$, тем больший вес будут иметь соответствующие волновые функции $\psi_{\mathbf{k}}(\mathbf{r})$. При $W_0 V_R \rightarrow 0$ функция локализованного состояния $\psi_{\mathbf{l}}(\mathbf{r})$ переходит в одну из функций $\psi_{\mathbf{k}}(\mathbf{r})$, описывающую полностью делокализованное состояние электрона.

Полученный результат имеет важнейшее значение для физики полупроводников.

Резюме § 20

1. Локальные нарушения поля решетки в виде примесных атомов, вакансий, точечных дефектов, дислокаций обуславливают появление в запрещенной зоне разрешенных состояний, связанных с областью возмущения.

2. Волновая функция электрона отлична от нуля примерно в той же области кристалла, в которой существует возмущение, другими словами, электрон локализован в области возмущения.

3. Уровень энергии, соответствующий локализованному состоянию электрона, возникает в результате отщепления экстремального уровня зоны энергии и перехода его в запрещенную зону. Если возмущение отрицательно, то в запрещенной зоне возникает уровень энергии за счет отщепления уровня минимальной энергии. Если возмущение положительно, то в запрещенную зону переходит уровень максимальной энергии. Примером положительного возмущения является возмущение, существующее в области вакансии. Чем сильнее возмущение, тем дальше отстоит локальный уровень от соответствующей зоны (см. рис. 28).

4. Если концентрация локальных возмущений $N_{\mathbf{l}}$ возрастает, то среднее расстояние между ними уменьшается, поскольку оно равно $N_{\mathbf{l}}^{-1/3}$. С ростом концентрации $N_{\mathbf{l}}$ волновые функции локализованных состояний могут перекрываться, что приводит к обмен-

ному взаимодействию между локализованными состояниями, благодаря чему дискретный уровень энергии E_d , расщепляясь, образует зону энергии.

Если локальные возмущения обусловлены атомами примеси, то соответствующие уровни и зона называются примесными. Для превращения примесного уровня в примесную зону необходима достаточно большая концентрация атомов примеси.

§ 21. ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ПРИМЕСНЫХ СОСТОЯНИЙ

Наиболее важный случай локализованных состояний возникает при введении в полупроводник атомов примеси. Оценим положение примесного уровня в запрещенной зоне. Переход электрона с примесного уровня в зону энергии равносителен ионизации атома примеси, поэтому за начало отсчета энергии удобно взять дно зоны энергии. Расстояние между дном зоны и примесным уровнем должно быть равным энергии ионизации атома примеси. Энергия ионизации свободных атомов составляет величину от 4—5 эВ для щелочных металлов до 24 эВ для He, грубо можно считать ее равной десятку электронвольт. При введении атома в качестве примеси в полупроводниковое вещество он вступает во взаимодействие с атомами основного вещества, поэтому энергия связи электронов примесного атома с собственным ядром уменьшается по сравнению с энергией связи электрона в свободном атоме. Рассмотрим в качестве примера атом элемента пятой группы As, введенный в германий. Как уже говорилось, четыре валентных электрона мышьяка будут участвовать в образовании четырех парноэлектронных связей с ближайшими соседними атомами германия. Пятый электрон, не принимая участия в образовании ковалентных связей, вступит в сравнительно слабое взаимодействие с большим числом атомов германия. Его связь с атомом As уменьшится, он начнет двигаться по орбите большого радиуса (если говорить на языке наглядных представлений теории Бора), охватывающей десятки и сотни атомов основного вещества. Поэтому формально взаимодействие избыточного электрона с атомами основного вещества можно представить с помощью диэлектрической проницаемости ϵ вещества, т. е. предположим, что потенциальная энергия электрона примесного атома может быть записана в виде

$$V(r) = -Ze^2/(\epsilon r) \quad (\text{в системе Гаусса}), \quad (21.1)$$

что, вообще говоря, справедливо только для макроскопических точечных зарядов. Заряд иона полагаем равным $+Ze$, заряд электрона ($-e$). Полная энергия H электрона равна

$$H = \frac{p^2}{2m^*} - \frac{Ze^2}{\epsilon r} = T + V. \quad (21.2)$$

Поскольку квантовомеханическое решение задачи о движении электрона в кулоновском поле точечного заряда дает то же значе-

ние энергии, что и элементарная квантовая теория атома водорода Бора, ограничимся последней теорией для нахождения уровней энергии водородоподобного атома примеси.

Условие устойчивости движения по круговой орбите

$$m^*v^2/r = Ze^2/(\epsilon r^2) \quad (21.3)$$

позволяет найти связь между кинетической и потенциальной энергией:

$$T = m^*v^2/2 = Ze^2/(2\epsilon r) = -V/2, \quad (21.4)$$

откуда

$$H = \frac{1}{2} V = -T = -\frac{1}{2} \frac{Ze^2}{\epsilon r} = -\frac{p^2}{2m^*} = E. \quad (21.5)$$

Из условия квантования момента количества движения M по Бору

$$M = rp = rm^*v = n\hbar \quad (21.6)$$

совместно с условием устойчивости (21.3) получим

$$r^2 m^{*3} v^2 / (m^* r) = M^2 / (m^* r) = n^2 \hbar^2 / (m^* r) = Ze^2 / \epsilon; \quad (21.7)$$

$$1/r = m^* Ze^2 / (\epsilon \hbar^2 n^2) = 1/r_n.$$

Мы видим, что радиусы боровских орбит водородоподобной атомной системы возрастают в ϵ раз по сравнению со свободным атомом «водорода». Возможные значения энергии равны

$$E = E_n = -\frac{1}{2} \frac{Z^2 e^4 m^*}{\hbar^2 n^2 \epsilon^2}. \quad (21.8)$$

Энергия ионизации E_1 равна по модулю энергии основного состояния ($n=1$):

$$E_1 = |E_1| = \frac{1}{2} \frac{Z^2 e^4 m^*}{\hbar^2 \epsilon^2}. \quad (21.9)$$

Если выразить энергию ионизации E_1 в электронвольтах, подставив численное значение e , \hbar и m , то получим

$$E_1 = \frac{13,52 \cdot Z^2}{\epsilon^2} \left(\frac{m^*}{m} \right) \text{ (эВ)}, \quad (21.10)$$

где 13,52 соответствует величине (в эВ) энергии ионизации атома водорода. Проанализируем выражение (21.10) для E_1 . Прежде всего обратим внимание на зависимость E_1 от Z^2 . Она означает, что уровни однократно- и двукратнозаряженного иона примеси лежат на различных расстояниях от зоны энергии. Это естественно, поскольку возмущение, вносимое ионом с зарядом $2e$, больше, чем возмущение, вносимое ионом с зарядом e , что находится в полном соответствии с выводами общей теории, рассмотренной в предыдущем параграфе.

Далее, энергия ионизации примеси уменьшается в ϵ^2 раз по сравнению с энергией ионизации атома водорода. Для германия $\epsilon = 16$,

для кремния $\epsilon = 12$, поэтому энергия ионизации примеси в них в 256 и 144 раза соответственно меньше энергии ионизации атома водорода и составляет в электронвольтах порядка 0,05 и 0,1 от

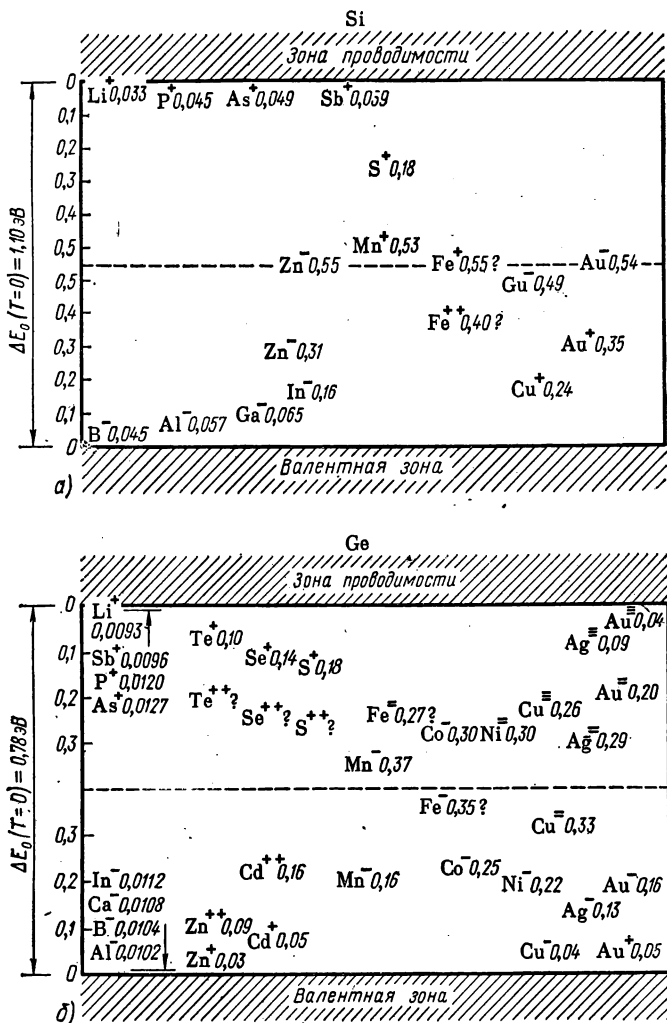


Рис. 29. Схема уровней энергии примеси в кремнии (а) и германии (б)

m^*/m . Поскольку m^*/m несколько меньше единицы, можем сказать, что энергия ионизации примеси в германии составляет величину менее 0,05 эВ, в кремнии — менее 0,10 эВ. На рис. 29 показаны уровни основных состояний различных примесей в кремнии (а) и германии (б). Различие в энергиях ионизации примеси можно каче-

ственно учесть, если вместо энергии ионизации атома водорода в (21.10) подставить энергию ионизации свободного атома, вводимого в качестве примеси. *Примесь, которая поставляет свободный электрон, называется донорной. Таким образом можем сказать, что донорный уровень возникает под дном зоны энергии.*

Таблица 3

Диэлектрическая проницаемость

Вещество	ϵ	Вещество	ϵ	Вещество	ϵ
Алмаз	5,7	ZnS	7,9	InAs	11,7
Кварц	4,3	CdTe	11,0	GaP	10,1
Сера	3,7	InSb	15,9	GaAs	12,5
Кремний	11,7	InP	10,8	GaSb	12,5
Германий	15,8				

Рассмотрим энергию ионизации доноров в соединениях $A^{III}B^V$. Их диэлектрическая проницаемость имеет величину того же порядка, что и диэлектрическая проницаемость германия и кремния (табл. 3), поэтому, казалось бы, энергия ионизации примеси должна быть для них того же порядка. Однако опыт показывает, что в ряде случаев она значительно меньше. Причину этого легко понять, если учесть, что для германия и кремния отношение m^*/m близко к единице, в то время как в соединениях $A^{III}B^V$ m^*/m (табл. 7), как правило, много меньше единицы, благодаря чему E_I может составлять сотые и даже тысячные доли электронвольта. В табл. 4 приведены E_I примеси в соединениях $A^{III}B^V$ и указаны радиусы первой боровской орбиты a_1 в ангстремах. Малый энергетический зазор между зоной и донорным уровнем является характерной особенностью некоторых соединений $A^{III}B^V$. В обычных условиях этот зазор в них экспериментально не обнаруживается. Так, например, измерения на InSb при $T=2^\circ\text{K}$ и концентрации электронов $n=10^{13}\text{ см}^{-3}$ не обнаружили энергетического зазора между зоной примеси и зоной энергии.

Таблица 4

«Средняя» энергия ионизации E_I примеси и радиус первой боровской орбиты a_1 в соединениях $A^{III}B^V$

Вещество	Акцепторы		Доноры	
	E_I , эВ	a_1 , Å	E_I , эВ	a_1 , Å
InSb	0,03	14	0,0007	640
InAs	0,05	12	0,002	310
InP	—	—	0,008	80
GaSb	0,03	15	0,003	150
GaAs	0,05	12	0,008	85

До сих пор говорилось о донорной примеси, отдающей избыточный электрон. Рассмотрим акцепторную примесь, например, атомы индия в германии. Нейтральный атом индия образует три из четырех парноэлектронные связи. Четвертая связь оказывается незавершенной. Перевод одного электрона от основного вещества завершает свободную связь и превращает атом индия в отрицательно заряженный ион индия.

Незавершенная связь между атомами основного вещества может перемещаться по кристаллу без подведения энергии извне. В то же время уход электрона от иона индия означает превращение его в нейтральный атом индия. Что представляет собой введение атома примеси с точки зрения нарушения периодичности поля решетки? Рассмотрим это на линейной модели. На рис. 22 дана потенциальная кривая цепочки атомов основного вещества. Предположим, что мы удалили один атом. Потенциальная яма, существовавшая на месте атома, должна исчезнуть. Но это равносильно наложению положительного возмущения на периодическое поле решетки, создаваемое удалением атома. Отсюда становится ясно, что вакантный узел должен нести себя как положительное локальное возмущение достаточно большой величины и распространенное на область, занимаемую несколькими соседними атомами, порождающее локальное состояние у потолка зоны. Другими словами, *вакантные узлы должны быть поставщиками дырок*. Однако вернемся к примесному атому индия. Замещение вакансии нейтральным атомом индия изменит слабо меняющееся поле вакансии только внутри самого атома индия, а вне атома поле решетки останется искаженным, характер искажения соответствует положительному возмущению. Расстояние между потолком зоны и локальным уровнем, которое можно интерпретировать как энергию образования из нейтрального атома отрицательного иона, можно записать в виде

$$E_1 = \frac{13,52}{\epsilon^2} \left(\frac{m^*}{m} \right), \quad (21.11)$$

где m^* — эффективная масса электрона у потолка зоны, взятая по модулю. В табл. 4 приведены энергии ионизации акцепторной примеси.

Моделирование возмущений, создаваемых примесными атомами водородоподобной модели, не достаточно, оно дает лишь порядок энергии ионизации примеси, приводящих к появлению дискретных уровней.

Из приведенного рассмотрения очевидно, что *точечные дефекты и дислокации могут быть поставщиками носителей заряда, чаще всего дырок*.

Однако помимо «мелких» уровней, определяемых соотношением (21.10) или (21.11), в полупроводнике имеются локальные уровни, лежащие на значительно больших расстояниях от зон энергии. Эти «глубокие» уровни энергии не могут быть объяснены водородоподобной моделью. Чтобы объяснить существование глубоких уровней энергии, нужно считать, что электроны в таких атомах примеси

слабо взаимодействуют с атомами основного вещества, орбита электрона атома примеси имеет малый радиус. Чтобы описать такое состояние электрона, выберем для $V(r)$ экранирующий потенциал типа

$$V(r) = -\frac{Ae^{-\alpha r}}{r}. \quad (21.12)$$

На малых расстояниях потенциал достигает больших значений, но уже на расстояниях порядка $(2-3)\alpha^{-1}$ возмущение обращается в нуль. Чем больше α , тем уже потенциальная яма, моделирующая примесь. Как известно из квантовой механики, чем уже потенциальная яма, тем больше расстояние между уровнями энергии и тем глубже лежат уровни энергии электрона в такой потенциальной яме. Оценим поведение уровня энергии электрона для возмущения (21.12). Для этого оценим среднюю энергию возмущения:

$$\int_{V_R} W(r) d\tau = -\int_0^R \frac{A4\pi e^{-\alpha r}}{r} r^2 dr \approx -\frac{4\pi A}{\alpha^2}. \quad (21.13)$$

Если за радиус потенциальной ямы принять величину c/α , где c — величина порядка $2-3$, то

$$V_R = 4\pi c^3/(3\alpha^3), \quad (21.14)$$

и для среднего возмущения W_0 получим

$$W_0 = -\frac{3A}{c^3} \alpha. \quad (21.15)$$

Таким образом, с ростом α диаметр потенциальной ямы уменьшается, а величина среднего возмущения и смещение локального уровня энергии увеличиваются. Кстати заметим, что для кулоновского потенциала $\alpha = 0$.

Глубокие уровни играют важную роль в протекании неравновесных процессов.

Вернемся теперь к выражению (21.8), которое перепишем с учетом (21.10) в виде

$$E_n = -\frac{E_I}{n^2} \quad (n = 1, 2, 3, \dots) \quad (21.16)$$

(энергия отсчитывается от экстремума соответствующей зоны). Это означает, что введение атома примеси приводит к возникновению в запрещенной зоне целой серии водородоподобных уровней энергии, которые сходятся к соответствующей зоне энергии, как это указано на рис. 30 для донорной примеси. Совокупность уровней энергии проявляется в оптических свойствах полупроводников (примесное поглощение света).

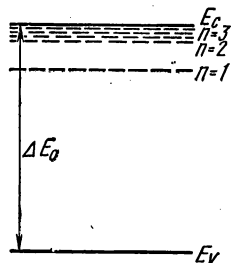


Рис. 30. Водородоподобная система уровней энергии примеси в запрещенной зоне

1. Примесь, вносящая отрицательное возмущение в поле решетки, приводит к появлению дискретного уровня энергии под дном зоны. Она является поставщиком электронов и называется донорной. Примесь, вносящая положительное возмущение, приводит к появлению дискретного уровня над потолком зоны; она является поставщиком дырок и называется акцепторной. Схема уровней примеси указана на рис. 31.

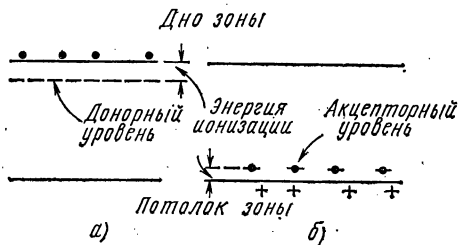


Рис. 31. Схема уровней донорной (а) и акцепторной (б) примесей

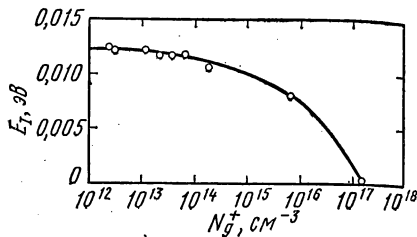


Рис. 32. Зависимость энергии ионизации примеси мышьяка от концентрации примеси в некотором интервале температур

2. Мелкие уровни энергии примеси качественно хорошо описываются водородоподобной моделью. Энергия ионизации примеси в этом случае обратно пропорциональна ϵ^2 и пропорциональна m^* . Глубокие уровни можно описать с помощью экранирующего потенциала.

3. С ростом концентрации примеси дискретный уровень вследствие обменного взаимодействия расщепляется в примесную зону. Энергия ионизации примеси уменьшается (рис. 32).

§ 22. ПОВЕРХНОСТНЫЕ СОСТОЯНИЯ

Реальные кристаллы имеют одно неизбежное нарушение периодичности решетки, связанное с конечными размерами тел и наличием границ. До сих пор мы исключали наличие границ условием периодического продолжения кристалла. Рассмотрим теперь, к каким изменениям спектра энергии приводит существование границ на одномерной модели потенциального поля. Поместим начало отсчета координаты на левой границе одномерного полубесконечного кристалла. В первой области ($x < 0$) $U(x) = U_0$, во второй области ($x > 0$) $U(x+a) = U(x)$.

Уравнение Шредингера для 1-й и 2-й областей имеет вид:

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_1(x)}{dx^2} + U_0\psi_1(x) = E\psi_1(x) \quad (1\text{-я область}). \quad (22.1)$$

$$-\frac{\hbar^2}{2m} \frac{d^2\psi_2(x)}{dx^2} + U(x)\psi_2(x) = E\psi_2(x) \quad (2\text{-я область}). \quad (22.2)$$

Каково состояние электрона при $E < U_0$? Запишем решение уравнения Шредингера для 1-й области:

$$\psi_1(x) = Ae^{\kappa x} + Be^{-\kappa x}; \quad \kappa = \sqrt{\frac{2m(U_0 - E)}{\hbar^2}}. \quad (22.3)$$

При $x \rightarrow -\infty$ второй член неограниченно возрастает, поэтому необходимо положить $B = 0$. Влияние границы на движение электрона при больших положительных значениях x будет мало, поэтому решение (22.2) должно иметь вид функции Блоха:

$$\psi_2(x) = e^{\pm i\kappa x} \varphi_{\pm \kappa}(x). \quad (22.4)$$

Волновой вектор, входящий в функцию Блоха для неограниченного кристалла — вещественный, в противном случае функция Блоха будет неограниченной на бесконечности. Но поскольку значения $x < 0$ в данном случае не реализуются, условие вещественности κ может быть нарушено. Более того, при вещественном κ не могут быть выполнены условия непрерывности волновой функции и ее первой производной в точке $x = 0$. Поэтому допустим, что комплексность волнового вектора обеспечивает необходимое требование, предъявляемое к волновой функции $\psi_2(x)$. В таком случае положим

$$\kappa = \kappa_1 + i\kappa_2 \quad (22.5)$$

и запишем общее решение для второй области:

$$\psi_2(x) = Ce^{i(\kappa_1 + i\kappa_2)x} \varphi_{\kappa}(x) + De^{-i(\kappa_1 + i\kappa_2)x} \varphi_{-\kappa}(x). \quad (22.6)$$

Необходимо выбрать $D = 0$, иначе $\psi_2 \rightarrow \infty$ при $x \rightarrow \infty$. Для нахождения A и C запишем условие непрерывности ψ и ψ' :

$$\psi_1(0) = A = \psi_2(0) = C\varphi_{\kappa}(0), \quad (22.7)$$

$$\psi_1'(0) = \kappa A = \psi_2'(0) = C[i\kappa\varphi_{\kappa}(0) + \varphi_{\kappa}'(0)]. \quad (22.8)$$

Разделив (22.8) на (22.7), исключим неизвестные коэффициенты A и C , в результате чего получим

$$\kappa\varphi_{\kappa}(0) = i\kappa\varphi_{\kappa}(0) + \varphi_{\kappa}'(0), \quad (22.9)$$

или

$$\kappa = i\kappa + \frac{\varphi_{\kappa}'(0)}{\varphi_{\kappa}(0)} = i\kappa + [\ln \varphi_{\kappa}(0)]'. \quad (22.10)$$

В общем случае функция $\varphi_{\kappa}(x)$ является функцией комплексной, поэтому и логарифм ее тоже является комплексной величиной, в то время как κ является величиной вещественной. Выделим вещественную и мнимую части производной логарифма

$$[\ln \varphi_{\kappa}(0)]' = \lambda + i\mu \quad (22.11)$$

и подставим в (22.10)

$$\kappa = i\kappa + \lambda + i\mu = i\kappa_1 - \kappa_2 + \lambda + i\mu, \quad (22.12)$$

откуда

$$\kappa = \lambda - \kappa_2; \quad \kappa_1 = -\mu. \quad (22.13)$$

Запишем выражение для энергии:

$$E(\kappa) = U_0 - \frac{\hbar^2 \kappa^2}{2m} = U_0 - \frac{\hbar^2 (\kappa_2 - \lambda)^2}{2m}. \quad (22.14)$$

При $\kappa_2 = 0$ энергия $E(\kappa) = E(\kappa_1)$ имеет определенную зонную структуру. Выражение для энергии ограничивает возможные значения κ условием $\kappa = \lambda$, где $\lambda = \text{Re}[\ln \varphi_\kappa(0)]'$, поскольку $\varphi_\kappa(x)|_{x=0}$ является некоторой функцией от κ . Если $\kappa_2 \neq 0$, то при фиксированном λ энергия оказывается квадратичной функцией κ_2 , причем это справедливо при любом значении λ . Но это означает, что при $\kappa_2 \neq 0$ появляются такие значения энергии, которые не содержались в спектре энергии электрона для неограниченного кристалла, поскольку значения энергии в неограниченном кристалле описываются условием: $\kappa_2 = 0$. Возникающие в результате ограничения кристалла уровни должны находиться в запрещенной зоне.

Рассмотрим состояния, соответствующие этим уровням:

$$\psi_2(x) = C e^{i\kappa_1 x} \varphi_\kappa(x) e^{-\kappa_2 x} \quad (x > 0); \quad (22.15)$$

$$\psi_1(x) = A e^{\kappa x} \quad (x < 0). \quad (22.16)$$

Мы видим, что волновые функции состояний, обусловленных границей, экспоненциально спадают в обе стороны от границы, т. е. эти состояния локализованы в пограничном слое толщиной порядка κ_2^{-1} , в связи с чем их называют *поверхностными состояниями*, а дополнительные уровни — *поверхностными уровнями*, или *уровнями Тамма*, который впервые (1932 г.) предсказал их существование. Конкретное выражение для положения уровней Тамма зависит от вида потенциала $U(x)$ и от положения границы по отношению к полю (к месту сечения) посредством $[\ln \varphi_\kappa(0)]'$.

Для трехмерного кристалла число поверхностных состояний достигает величины $\approx 10^{15} - 10^{16} \text{ см}^{-2}$.

Помимо уровней Тамма, в поверхностном слое имеется большое число локализованных состояний, связанных с многочисленными нарушениями поля решетки (дефектами и адсорбированными атомами).

Если число состояний, связанных с уровнями Тамма для данного кристалла, постоянно, то *число состояний, вызываемых дефектами решетки и адсорбированными атомами, может меняться в зависимости от обработки поверхности*. Между состояниями электрона на поверхности кристалла и в объеме возможен целый ряд различных взаимодействий, которые приводят к тому, что *состояния на поверхности оказывают сильное влияние на протекание физических явлений в объеме полупроводника*. Особенно сильно поверхностные состояния проявляются при работе полупроводниковых приборов. Некоторые явления, связанные с поверхностными состояниями, рассмотрим в § 67.

1. Ограниченность реального кристалла нарушает условие периодичности. Это приводит к тому, что в запрещенной зоне появляются разрешенные уровни энергии, которые называются поверхностными уровнями, или уровнями Тамма.

2. Волновые функции уровней Тамма затухают экспоненциально в обе стороны от границы, другими словами, граница приводит к возникновению локализованных состояний. Размеры области локализации электрона вне кристалла характеризуются величиной κ^{-1} и внутри кристалла — величиной κ_2^{-1} ; κ и κ_2 связаны с энергией E .

3. В приповерхностном слое кристалла имеется большое число локальных возмущений поля решетки, создаваемых дефектами и адсорбированными атомами, которые приводят согласно общей теории к возникновению локализованных состояний. Их также называют поверхностными состояниями.

4. Поверхностные состояния играют важную роль в работе полупроводниковых приборов. Число и вид поверхностных состояний зависят от обработки поверхности (шлифовка, полировка, травление, лакировка и т. д.) и окружающих условий (температура, влажность, газовая атмосфера и т. д.).

§ 23. КВАНТОВАНИЕ ЭНЕРГИИ ЭЛЕКТРОНА В МАГНИТНОМ ПОЛЕ. УРОВНИ ЛАНДАУ

Из предыдущего рассмотрения выпадает основа целого класса явлений, которые наблюдаются в сильном магнитном поле. Для их описания необходимо рассмотреть функцию Гамильтона частицы в магнитном поле, которое определяется векторным потенциалом $\mathbf{A}(\mathbf{r})$. Индукция магнитного поля \mathbf{B} связана с $\mathbf{A}(\mathbf{r})$ соотношением

$$\mathbf{B} = \text{rot } \mathbf{A}. \quad (23.1)$$

Если

$$\mathbf{p} = m\mathbf{v} \quad (23.2)$$

есть импульс частицы, то величина

$$\mathbf{P} = \mathbf{p} + e\mathbf{A} \quad (23.3)$$

носит название *обобщенного импульса* для частицы с зарядом e .

Функцию Гамильтона H можно выразить через обобщенный импульс, для этого необходимо учесть, что кинетическая энергия

$$T = \frac{p^2}{2m} = \frac{(\mathbf{P} - e\mathbf{A})^2}{2m} \quad (23.4)$$

и

$$H = T + U = \frac{(\mathbf{P} - e\mathbf{A})^2}{2m} + U(\mathbf{r}). \quad (23.5)$$

Чтобы получить оператор Гамильтона, необходимо вместо обобщенного импульса P подставить оператор $-i\hbar \nabla$.

$$\hat{P} = -i\hbar \nabla. \quad (23.6)$$

Это связано с тем, что при движении частицы в магнитном (электромагнитном) поле канонически сопряженными величинами являются координата (обобщенная в общем случае) и обобщенный импульс. Соотношение (23.6) позволяет записать оператор Гамильтона в следующем виде:

$$\hat{H} = \frac{(-i\hbar \nabla - eA)^2}{2m} + U(r). \quad (23.7)$$

Найдем оператор $(i\nabla + A)^2$;

$$(i\nabla + A)^2 = (i\nabla + A)(i\nabla + A) = -\Delta + i\nabla A + i(A\nabla) + A^2. \quad (23.8)$$

Но

$$\nabla A = (\nabla A) + (A\nabla) = \text{div } A + (A\nabla), \quad (23.9)$$

поэтому

$$(i\Delta + A)^2 = -\Delta + A^2 + 2i(A\nabla) + i \text{div } A \quad (23.10)$$

и

$$\hat{H} = -\frac{\hbar^2}{2m} \Delta + \frac{e^2 A^2}{2m} + \frac{i\hbar e}{m} (A\nabla) + \frac{i\hbar e}{2m} \text{div } A + U(r). \quad (23.11)$$

Рассмотрим свободный электрон $U(r) = 0$ в однородном магнитном поле $\mathbf{B} = (0, 0, B)$. Найдем вид вектора-потенциала \mathbf{A} :

$$\left. \begin{aligned} B_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0; \\ B_y &= \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = 0; \\ B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B. \end{aligned} \right\} \quad (23.12)$$

Уравнения (23.12) удовлетворяются, если положить, например,

$$A_y = A_z = 0; \quad A_x = -yB_z = -yB. \quad (23.13)$$

В этом случае

$$\hat{H} = -\frac{\hbar^2}{2m} \Delta + \frac{e^2 B^2}{2m} y^2 - \frac{i\hbar e B}{m} y \frac{\partial}{\partial x}, \quad (23.14)$$

а уравнение Шредингера будет иметь вид

$$-\frac{\hbar^2}{2m} \Delta \psi - \frac{i\hbar e B}{m} y \frac{\partial \psi}{\partial x} + \frac{e^2 B^2 y^2}{2m} \psi = E \psi(r). \quad (23.15)$$

В уравнении (23.15) переменные разделяются. Действительно, при $B = 0$ (23.15) представляет собой уравнение для волн де-Бройля

$$-\frac{\hbar^2}{2m} \Delta \psi_0 = E_0 \psi_0, \quad (23.16)$$

решение которого имеет вид

$$\psi_0(r) = \frac{1}{(2\pi)^{3/2}} e^{i(\kappa r)}; \quad E_0 = \frac{\hbar^2 \kappa^2}{2m}. \quad (23.17)$$

Поскольку в (23.15) зависимость от x и z подобна зависимости в (23.16), можем предположить, что

$$\psi(x, y, z) = e^{i(\kappa_x x + \kappa_z z)} \varphi(y). \quad (23.18)$$

Подставим (23.18) в (23.15), сокращая на экспоненту, получим

$$-\frac{\hbar^2}{2m} \left(-\kappa_x^2 - \kappa_z^2 + \frac{d^2}{dy^2} \right) \varphi(y) + \frac{\hbar \kappa_x e B y}{m} \varphi(y) + \frac{e^2 B^2 y^2}{2m} \varphi(y) = E \varphi(y). \quad (23.19)$$

Положим

$$y = y' + y_0, \quad (23.20)$$

тогда

$$y^2 = y'^2 + 2y_0 y' + y_0^2$$

и два члена левой части (23.19) дадут

$$\begin{aligned} \frac{\hbar \kappa_x B}{m} y' + \frac{\hbar \kappa_x B}{m} y_0 + \frac{e^2 B^2}{2m} y'^2 + \frac{2e^2 B^2}{2m} y' y_0 + \frac{e^2 B^2}{2m} y_0^2 &= \frac{e^2 B^2}{2m} y'^2 + \frac{e^2 B^2}{2m} y_0^2 + \\ + \frac{eB}{m} (\hbar \kappa_x + e B y_0) y' + \frac{\hbar \kappa_x e B}{m} y_0. \end{aligned} \quad (23.21)$$

Если выбрать $y_0 = -\frac{\hbar \kappa_x}{eB}$, то член с y' обращается в нуль, и уравнение Шредингера будет иметь вид

$$-\frac{\hbar^2}{2m} \left(-\kappa_x^2 - \kappa_z^2 + \frac{d^2}{dy'^2} \right) \varphi(y') + \left[\frac{e^2 B^2}{2m} y'^2 + \left(\frac{\hbar^2 \kappa_x^2}{2m} - \frac{\hbar^2 \kappa_z^2}{m} \right) \right] \varphi(y') = E \varphi(y'), \quad (23.22)$$

или

$$-\frac{\hbar^2}{2m} \frac{d^2 \varphi(y')}{dy'^2} + \frac{e^2 B^2}{2m} y'^2 \varphi(y') = \left(E - \frac{\hbar^2 \kappa_z^2}{2m} \right) \varphi(y'). \quad (23.23)$$

Полагая $\omega_0 = \frac{eB}{m}$, запишем

$$-\frac{\hbar^2}{2m} \frac{d^2 \varphi(y')}{dy'^2} + \frac{m \omega_0^2 y'^2}{2} \varphi(y') = \left(E - \frac{\hbar^2 \kappa_z^2}{2m} \right) \varphi(y') = E' \varphi(y'). \quad (23.24)$$

Уравнение (23.24) представляет собой уравнение гармонического осциллятора с массой m и частотой ω_0 . Энергия осциллятора

$$E' = E'_n = \hbar \omega_0 \left(n + \frac{1}{2} \right); \quad -n = 0, 1, 2, \dots \quad (23.25)$$

и, следовательно,

$$E = E_n = \frac{\hbar^2 \kappa_z^2}{2m} + \hbar \omega_0 \left(n + \frac{1}{2} \right). \quad (23.26)$$

Выражение (23.26) показывает, что на движение вдоль поля (по оси z) магнитное поле не оказывает никакого воздействия, энергия, связанная с этим движением, не квантована. Движение в плоскости xOy , перпендикулярной полю, носит совершенно другой характер — частица совершает круговое движение с частотой ω_0 — так называемой циклотронной частотой; энергия, связанная с этим движением, квантована, дискретна, расстояния между возможными значениями энергии равны $\hbar\omega_0$. Дискретность энергии вращательного движения отличает квантовую частицу от классической, для которой энергия вращательного движения непрерывна. На языке наглядных представлений можно сказать, что для классической частицы возможны любые радиусы окружности, в то время как для квантовой частицы радиусы орбит квантованы и равны



Рис. 33. Спектр энергии электрона в магнитном поле

$$r_n \approx \sqrt{\frac{2\hbar}{m\omega_0} \left(n + \frac{1}{2}\right)} = \sqrt{\frac{2\hbar}{eB} \left(n + \frac{1}{2}\right)}. \quad (23.27)$$

Для перехода частицы с одной орбиты на другую необходимо затратить энергию $\hbar\omega_0 \cdot \delta n$, где δn — целое число. Таким образом, процессы поглощения энергии, связанные с изменением вращательного движения, должны характеризоваться

дискретным линейчатым спектром (это явление носит название циклотронного резонанса).

Спектр энергии $E_n(k_z)$ частицы может быть изображен совокупностью квадратичных парабол, сдвинутых по оси энергии друг относительно друга на величину $\hbar\omega_0$ (рис. 33, а). Спектр энергии $E_n(k_z=0)$ изображается совокупностью дискретных уровней (рис. 33, б). Уровни энергии электрона в магнитном поле носят название уровней Ландау. Квантование энергии свободных электронов является физической основой диамагнетизма свободного электронного газа.

Рассмотрим теперь движение частицы в кристалле, на который наложено магнитное поле B . Гамильтониан кристалла в магнитном поле

$$\hat{H} = \frac{(P - eA)^2}{2m} + U(\mathbf{r}) \quad (23.28)$$

в приближении эффективной массы можно записать в виде

$$\hat{H} = \frac{(P - eA)^2}{2m^*}, \quad (23.29)$$

что совпадает с (23.4). Из (23.29) следует, что энергия электронов должна быть проквантована для каждой зоны в соответствии с выражением (23.26):

$$E = E_0 + \frac{\hbar^2 \kappa_z^2}{2m^*} + \hbar \omega_0 \left(n + \frac{1}{2} \right) = E(\kappa_z, n, \omega_0). \quad (23.30)$$

$$\omega_0 = \omega_c = \frac{eB}{m^*}, \quad n = 0, 1, 2, \dots$$

На рис. 34 наглядно изображено превращение квазинепрерывного спектра энергий в дискретный. Квантование энергии приводит к целому ряду магнетосцилляционных явлений: магнетооптическому поглощению, осцилляции магнитной восприимчивости, осцилляции магнетосопротивления и др. Эффекты квантования энергии могут быть заметны при условии, что расстояния между уровнями энергии превышают энергию kT , в противном случае функция распределения электронов по состояниям остается квазинепрерывной:

$$\hbar \omega_0 = \frac{e \hbar B}{m^*} > kT. \quad (23.31)$$

Подставляя значения констант e , \hbar , k , m , получим

$$1,6 \frac{m}{m^*} B \cdot 10^{-20} > 1,38 \cdot 10^{-16} T, \quad (23.32)$$

или

$$B \text{ (Гс)} > \frac{m^*}{m} T \cdot 10^4.$$

Если $T = 4,2^\circ$ (жидкий гелий), то

$$B \text{ (Гс)} > 42\,000 \left(\frac{m^*}{m} \right), \quad (23.33)$$

т. е. $B > 42\,000$ Гс при $\frac{m^*}{m} = 1$;

$$B > 4,2 \text{ кГс при } \frac{m^*}{m} = 0,1.$$

Очевидно, что для наблюдения квантования энергии в магнитном поле с использованием полей в десятки килогауссов необходимы температуры жидкого гелия.

Резюме § 23

При наложении магнитного поля на кристалл характер движения электрона меняется: электрон движется вдоль поля со скоростью

$$v_z = \frac{\hbar (\kappa_z - \kappa_{0z})}{m^*} \quad (23.1p)$$

и вращается в плоскости, перпендикулярной полю.

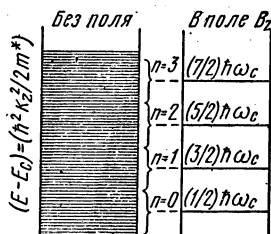


Рис. 34. Группировка квазинепрерывного спектра энергии в дискретные уровни Ландау

2. Частота вращения ω_0 , называемая циклотронной частотой, определяется индукцией B магнитного поля и эффективной массой m^* :

$$\omega_0 = eB/m^*. \quad (23.2p)$$

3. Энергия электрона, соответствующая вращательному движению, квантована, в результате чего в окрестности экстремума

$$E = E_0 + \frac{\hbar^2 \kappa_z^2}{2m^*} + \hbar \omega_0 \left(n + \frac{1}{2} \right) = E(\kappa_z, n, \omega_0). \quad (23.3p)$$

4. Гамильтониан свободного электрона в магнитном поле можно получить из выражения для кинетической энергии, которая зависит от импульса p , если ввести обобщенный импульс

$$P = p + eA, \quad (23.4p)$$

которому соответствует оператор $-i\hbar \nabla$.

5. Квантование энергии электрона в магнитном поле лежит в основе ряда магнетоосцилляционных явлений.

6. Полупроводник, находящийся в магнитном поле с индукцией B , может поглощать энергию излучения с частотой $\omega_0 = eB/m^*$. Это явление носит название циклотронного, или диамагнитного, резонанса. Электрон, поглотив квант энергии $\hbar \omega_0$, переходит на более высокий уровень Ландау.

§ 24. ПРИНЦИП ПАУЛИ. ПОНЯТИЕ МЕТАЛЛА, ПОЛУПРОВОДНИКА И ДИЭЛЕКТРИКА

Состояние классической частицы характеризуется набором шести физических величин, например (x, y, z, p_x, p_y, p_z) . Если не учитывать спин, то состояние квантовой частицы описывается тремя величинами, например (x, y, z) , (p_x, p_y, p_z) или (E, M^2, M_z) . Более точное описание должно в любом наборе содержать четвертую величину — проекцию спина на некоторое направление, обычно на ось z , — s_z . Таким образом, *полный набор, необходимый для описания квантовых состояний, состоит из четырех физических величин (L_1, L_2, L_3, s_z)* . Выбор трех физических величин (L_1, L_2, L_3) определяется обычно видом квантовой системы. В атоме наиболее употребителен набор из энергии E , модуля орбитального момента $|M|$ и проекции орбитального момента M_z . Обычно эти величины выражаются через соответствующие квантовые числа n , l и m_l . Для описания состояний электрона в кристалле наиболее удобно использовать набор из проекций квазиимпульса (P_x, P_y, P_z) или волнового вектора $(\kappa_x, \kappa_y, \kappa_z)$ и проекции спина s_z .

Напомним формулировку принципа Паули: *при измерении физических величин (L_1, L_2, L_3, s_z) , входящих в полный набор, каждый полный набор может быть получен не более чем один раз*. В более простой формулировке принцип Паули гласит: *в любом состоянии с полным набором (L_1, L_2, L_3, s_z) может находиться не более одного*

электрона. Применительно к кристаллу это означает, что в состоянии (k_x, k_y, k_z, s_z) может находиться не более одного электрона. Если учесть, что s_z принимает лишь два значения $+1/2$ и $-1/2$, то можно сказать, что в состоянии (k_x, k_y, k_z) может быть не более двух электронов. Набор (k_x, k_y, k_z) определяет величину энергии $E(k)$ поэтому, согласно принципу Паули, на данном энергетическом уровне может быть ограниченное число электронов.

Найдем число возможных состояний в зоне. Если кристалл имеет вид параллелепипеда с длиной ребер $N_1 a_1, N_2 a_2, N_3 a_3$, где N_i — число атомов вдоль i -го ребра, то полное число атомов в кристалле равно $N_1 \cdot N_2 \cdot N_3 = N$; i -ая проекция вектора k принимает следующие значения:

$$k_i = \frac{2\pi}{a_i} \frac{n_i}{N_i}; \quad n_i = 0; \pm 1; \pm 2; \dots \pm \frac{N_i - 1}{2}; -\frac{N_i}{2}; \quad (24.1)$$

число различных значений k_i равно N_i . Полное число различных состояний в зоне Бриллюэна равно $N_1 \cdot N_2 \cdot N_3 = N$, т. е. равно числу атомов в кристалле. Это означает, что в простой (невырожденной) энергетической зоне и соответствующей ей зоне Бриллюэна имеется $2N$ состояний и в ней может быть не более $2N$ электронов. Если зона g -кратно вырождена, то в ней может быть $2Ng$ электронов. Этот результат означает сохранение числа состояний при образовании кристалла из атомов. Действительно, пусть зона образована из g -кратно вырожденного уровня. В свободном атоме на нем может быть $2g$ электронов. Если кристалл образован из N свободных атомов, то у них имеется $2Ng$ состояний. Таким образом, число конечных состояний (в кристалле) равно числу исходных состояний (в атомах).

Если в некоторой зоне нет ни одного электрона, то при наложении внешнего электрического поля эта зона никакого вклада в проводимость внести не может. Точно такой же результат справедлив и для зоны, в которой все состояния заняты электронами. Действительно, электрон под действием внешнего поля переходит на соседние энергетические уровни. Но это возможно только в том случае, если соседние уровни свободны. Если же все состояния заняты, то переходы электронов в другие состояния согласно принципу Паули не возможны и, следовательно, электрическое поле не может создать направленного движения такой совокупности электронов. Электроны могут стать электронами проводимости только в том случае, если они находятся в незаполненной зоне. Этот простой вывод является основой для классификации веществ по классам проводимости.

К проводникам относятся вещества, имеющие при любой температуре не целиком заполненные зоны. К непроводникам относятся вещества, имеющие некоторое число целиком занятых зон, а их свободные зоны отделены от заполненных энергетическим зазором или запрещенной зоной. Непроводники могут быть полупроводниками или

диэлектриками. Деление непроводников на полупроводники и диэлектрики условно. К полупроводникам относят вещества, для которых ширина запрещенной зоны, отделяющей самую верхнюю из заполненных

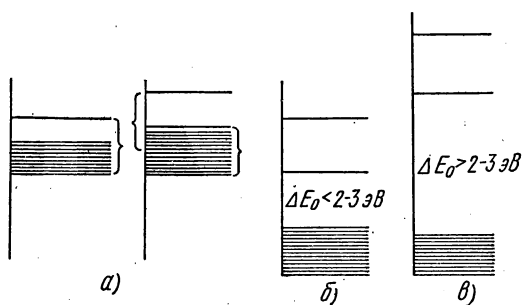


Рис. 35. Зоны энергии металла (а), полупроводника (б) и диэлектрика (в)

(она называется валентной) от самой нижней из свободных зон (она называется зоной проводимости), не более 2—3 эВ. К диэлектрикам относят вещества, ширина запрещенной зоны которых не менее 2—3 эВ. На рис. 35 дана энергетическая схема металла (а), полупроводника (б) и диэлектрика (в). Рассмотрим в качестве примера проводимость некоторых веществ с точки зрения расположения их зон.

Щелочные элементы. Щелочные элементы имеют один валентный электрон, находящийся в ns -состоянии. Остальные электроны занимают заполненные оболочки (или подоболочки). Для определенности в качестве примера рассмотрим натрий Na. Он имеет 11 электронов, состояние которых записывается следующим образом: $(1s^2)(2s^2)(2p^6)3s$. Видим, что первая $1s$ - и вторая $2s$ - и $2p$ -оболочки заняты. В $3s$ -оболочке находится один электрон. При образовании кристалла уровни атомов расщепляются в зоны, как это условно изображено на рис. 23. Приписывая зонам значки тех уровней, из которых они образовались, можем сказать, что зоны $1s$, $2s$, $2p$ заполнены целиком, так как в них находятся $2N$, $2N$ и $6N$ электронов при $2N$ -, $2N$ - и $6N$ -состояний. В зоне же $3s$ находятся N электронов при $2N$ возможных состояний. Таким образом, валентная зона, т. е. зона валентных электронов щелочных металлов, занята наполовину, половина зоны свободна. Щелочные элементы образуют хорошо проводящие кристаллы, они являются металлами.

Щелочноземельные элементы. Щелочноземельные элементы имеют два валентных электрона, находящихся в ns -состоянии. Например, для Mg $(1s^2)(2s^2)(2p^6)(3s^2)$. Если зоны образуются для Mg таким же образом, как и для Na, то Mg должен быть изолятором, поскольку $3s$ -зона целиком заполнена. Однако известно, что элементы второй группы являются хорошими металлами. Это можно объяснить, если предположить, что $3s$ - и $3p$ -зоны частично или полностью перекрываются, в результате чего электроны верхних уровней $3s$ -зоны занимают нижние уровни $3p$ -зоны.

Электроны s -зоны, принимающие участие в создании проводимости, занимают зону почти целиком, благодаря чему некоторые из них находятся в состояниях, в которых их эффективная масса отрицательна, вследствие чего они должны двигаться во внешних полях как частицы с положительным зарядом. Это проявляется в том,

что для многих из них коэффициент Холла больше нуля, в то время как для металлов с электронной проводимостью он должен быть меньше нуля.

Элементы третьей группы. Металлические свойства элементов третьей группы объясняются легко. Поскольку они имеют три валентных электрона, только один из которых находится в np -состоянии, то при образовании кристалла np -зона будет заполнена на $1/6$.

Элементы четвертой группы. Элементы четвертой группы имеют четыре валентных электрона, находящихся в состояниях $(ns^2)np^2$. Если бы при образовании кристалла характер вырождения зон совпадал с характером вырождения исходных атомных уровней, то элементы четвертой группы должны были бы быть металлами, поскольку валентная p -зона, содержащая два электрона на атом, должна быть заполнена на $1/3$. Однако такие элементы, как германий и кремний, являются классическими полупроводниками, алмаз — диэлектриком, а в некоторых случаях он выступает как полупроводник, α — Sp -полупроводник. Причина этого заключается в том, что при образовании кристалла происходит частичное снятие вырождения p -зоны. p -Зона расщепляется на две зоны — верхняя двукратно вырождена, нижняя невырождена (без учета спинового вырождения), при этом нижняя p -зона сливается с s -зоной, в результате чего при некоторых расстояниях между атомами из двух зон — s и p — образуются две зоны — верхняя образована из p -уровня, она двукратно вырождена, нижняя образована из s -уровня и p -уровня, и в верхней, и в нижней зонах содержится тем самым по $4N$ состояний, благодаря чему нижняя зона оказывается заполненной целиком, а верхняя — полностью свободна. Таким же образом ведут себя зоны при образовании кремния, германия, олова, только ширина запрещенной зоны для них меньше.

Аналогично можно объяснить полупроводниковые или диэлектрические свойства других простых веществ.

Из приведенных примеров видно, что для объяснения металлических свойств веществ необходимо предположить, что в некоторой зоне число электронов меньше числа состояний. Это может быть обусловлено или первоначальным неравенством числа исходных состояний и числа валентных электронов, или перекрытием зон, обеспечивающим излишек состояний по сравнению с числом валентных электронов. Для образования неметалла необходима такая комбинация зон, чтобы число состояний и число элементов в них были одинаковыми. Простейшим способом выполнения этого условия является образование вещества вида $A^x B^{8-x}$, например $A^{III} B^V$. Общее число валентных электронов на каждую пару атомов равно восьми. Если зоны элементов A и B не перекрываются между собой, то в s -и p -зонах одного из них (или s — одного и p — другого) как раз содержится восемь состояний на два атома. Рассмотрим в качестве примера AlP . Алюминий имеет 13 электронов, находящихся в состояниях $(1s^2)(2s^2)(2p^6)(3s^2)3p^2P_{1/2}$, фосфор имеет 15 электронов, распределенных следующим образом по состояниям: $(1s^2)(2s^2)(2p^6)(3s^2)$

$3p^3 4s_{3/2}$. Здесь через $^2P_{1/2}$ и $^4S_{3/2}$ указаны основные состояния атомов. В частности, отсюда видно, что единичные орбитальные моменты трех p -электронов дают суммарный момент, равный нулю, а спины всех трех p -электронов параллельны ($1 + 2S_z = 1 + 2 \cdot \frac{3}{2} = 4$).

Потенциалы ионизации алюминия и фосфора равны соответственно 5,99 и 11,98 В, т. е. основное состояние атома фосфора лежит ниже основного состояния атома алюминия на 6 электронвольт. В табл. 5 приведены потенциалы ионизации некоторых элементов. Так как второй потенциал ионизации атома фосфора определяет глубину залегания уровня s -электронов в нем, из табл. 5 видим, что s -уровни алюминия и фосфора лежат сравнительно глубоко под уровнями p -электронов фосфора. Поэтому можем предполагать, что s -зоны лежат сравнительно глубоко в соединении AlP , а валентная зона и зона проводимости образована из уровней p -электронов алюминия и фосфора, которые, частично расщепляясь и перекрываясь, дают трехкратно вырожденные валентную зону и зону проводимости.

Таблица 5

Потенциалы ионизации некоторых элементов

Элемент — Z	Состояние валентных электронов	E_I , эВ	Элемент — Z	Состояние валентных электронов	E_I , эВ
H — 1	1s	13,595	Ga — 31	(4s ²) 4p	16,11
Li — 3	2s	5,39	Ge — 32	(4s ²) 4p ²	7,88
B — 5	(2s ²) 2p	8,296	As — 33	(4s ²) 4p ³	9,81
C — 6	(2s ²) 2p ²	11,256	Se — 34	(4s ²) 4p ⁴	9,75
Na — 11	3s	5,138	Cd — 48	(5s ²)	8,991
Al — 13	(3s ²) 3p	5,984	In — 49	(5s ²) 5p	9,785
Si — 14	(3s ²) 3p ²	8,149	Sb — 51	(5s ²) 5p ³	8,639
P — 15	(3s ²) 3p ³	10,484	Te — 52	(5s ²) 5p ⁴	9,01
S — 16	(3s ²) 3p ⁴	10,357	Hg — 80	(6s ²)	10,43
Zn — 30	(4s ²)	5,391	Pb — 82	(6s ²) 6p ²	7,415

Интересно рассмотреть соединения типа $A^{II} B^{VI}$, которые имеют соответственно два и шесть валентных электронов, состояния которых можно записать в таком виде: $(ns^2)^1 S_0$ для элементов II группы и $(n's^2) n'p^4 {}^3 P_2$ для элементов IV группы.

Потенциалы ионизации этих элементов очень близки друг к другу, поэтому s -уровни A^{II} и p -уровни B^{VI} свободных атомов лежат примерно на одинаковой глубине. Если предположить, что валентная зона образована из уровней p -электронов, поскольку p -уровни должны опускаться ниже, так как волновая функция для них локализована в большей области, чем для s -электронов, то валентная зона должна быть трехкратно вырожденной, в ней должно быть 6 состояний на два атома, которые заняты 6 электронами атомов A^{II} и B^{VI} . Зона проводимости должна возникать из s -уровней элемента A^{II} и должна быть тем самым простой.

Таблица 6

Величины запрещенных зон (в эВ)

Тип соединений	Вещество	Ширина запрещенной зоны	Тип соединений	Вещество	Ширина запрещенной зоны		
Простые вещества	Si	1,10		CaTe	4,3		
	Ge	0,68		MgSe	5,6		
	Se	2,1		MgTe	4,7		
	Te	0,34		ZnO	3,2		
	α -Sn	0,08		ZnS	3,7		
I—V	KSb	0,9		ZnSe	2,6		
	K ₃ Sb	1,1		ZnTe	2,1		
	CsSb	0,8		SrO	5,8		
	Cs ₃ Bi	0,5		SrS	4,8		
I—VI	Cu ₂ O	2,0		SrSe	4,6		
	Ag ₂ S	0,9		SrTe	4,0		
I—VII	CuBr	2,9		CdS	2,4		
	AgI	2,8		CdSe	1,7		
II—IV	Mg ₂ Si	0,7		CdTe	1,5		
	Mg ₂ Ge	0,6		BaO	4,2		
	Mn ₂ Sn	0,3		BaS	4,0		
	Ca ₂ Si	1,9		BaSe	3,7		
	Ca ₂ Sn	0,9		BaTe	3,4		
	Ca ₂ Pb	0,46		HgS	2,0		
II—V	Zn ₃ P ₂	1,15		«красная» HgSe	0		
	CdSb	0,50		HgTe	0		
III—V	AlAs	2,4		IV—VI	TiO ₂	3,0	
	AlSb	1,5			SnO ₂	4,3	
	GaN	3,4			SnS	1,3	
	GaP	2,24			PbO	2,3	
	GaAs	1,4			PbS	0,40	
	GaSb	0,67			PbSe	0,25	
	InP	1,25			PbTe	0,31	
	InAs	0,33			V—VI	As ₂ Se ₃	1,7
	InSb	0,18				Sb ₂ O ₃	4,2
III—VI	Al ₂ O ₃	> 5		Sb ₂ S ₃		1,7	
	Al ₂ S ₃	4,1		Sb ₂ Se ₃		1,2	
	Al ₂ Se ₃	3,1		Sb ₂ Te ₃		0,3	
	Al ₂ Te ₃	2,5		Bi ₂ O ₃		3,2	
	GaSe	2,0		Bi ₂ Te ₃	0,15		
	Ga ₂ Se ₃	1,9		VI—VI	TeO ₂	1,5	
	GaTe	1,5			Другие соединения	ZnGeP ₂	2,2
	InSe	1,2		ZnSnP ₂		2,1	
	In ₂ Se ₃	1,2		CdGeP ₂		1,8	
	In ₂ Te ₃	1,0		CdSnP ₂		1,5	
IV—IV	SiC	2,3 (куб) 2,9 (гекс)		ZnIn ₂ Se		2,6	
	II—VI	CaS		5,4		ZnIn ₂ Te ₄	1,4
CaSe		5,0		CdIn ₂ Te ₄		0,9	
				HgIn ₂ Se ₄		0,6	

Аналогичным образом можно проводить анализ характера заполнения зон в других соединениях. Для этого необходимо учитывать структуру электронных оболочек атомов, образующих данное соединение. Однако такой анализ дает возможность лишь качественного суждения об общих закономерностях, например, изменения ширины запрещенной зоны при изменении состава.

Для объяснения большинства физических свойств полупроводников необходимо знать ширину запрещенной зоны ΔE_0 , которая отделяет валентную зону от зоны проводимости. В табл. 6 приведены экспериментальные данные о величине запрещенной зоны некоторых полупроводников.

Резюме § 24

1. Число состояний в зоне равно числу исходных состояний в атомах, из которых образован кристалл.
2. В каждой простой зоне может быть не более $2N$ электронов, если N — число атомов в кристалле.
3. К металлам относятся вещества, имеющие незаполненные зоны энергии. Наличие свободных состояний может быть обусловлено или тем, что число возможных состояний больше числа электронов, или тем, что перекрываются заполненные и свободные зоны.
4. К непроводникам относятся вещества, у которых имеется некоторое число целиком заполненных и целиком свободных зон. Самую верхнюю из заполненных зон называют валентной зоной, самую нижнюю из свободных зон называют зоной проводимости. Основную роль для протекания большинства физических процессов играет запрещенная зона, отделяющая валентную зону от зоны проводимости. К полупроводникам относятся вещества, для которых запрещенная зона имеет величину порядка 2—3 эВ. Вещества с большей шириной запрещенной зоны относятся к диэлектрикам.

§ 25. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ДЫРКИ

Рассматривая модельные представления о проводимости, мы ввели понятие дырочной проводимости как дефектной проводимости, обусловленной отсутствием некоторых электронов в связях. С точки зрения зонной теории дырочная проводимость обусловлена электронами почти заполненной зоны, т. е. электронами валентной зоны, в которой имеется некоторое число свободных состояний. Если обозначить концентрацию свободных состояний через p , то p можно считать концентрацией дырок. Другими словами, с точки зрения статистики дырку можно определить как свободное, незанятое электроном состояние $E(\mathbf{k})$. Однако динамические характеристики дырки этими простейшими понятиями не определяются. Их можно ввести, воспользовавшись понятием зон Бриллюэна.

Рассмотрим вначале зону Бриллюэна, в которой занято небольшое число n состояний в окрестности $\mathbf{P} = 0$. Предположим для простоты, что занятые состояния отделены от свободных резкой границей (рис. 36, а). Под действием внешнего поля \mathbf{E} все состояния сместятся на некоторую величину $\Delta\mathbf{P}$:

$$\Delta\mathbf{P} = e^{-}\mathbf{E}\tau, \quad (25.1)$$

где τ — среднее время свободного пробега электронов. Так как величина $\Delta\mathbf{P}$ не зависит от начального состояния $\mathbf{P}_i(0)$, то каждое занятое состояние смещается в зоне Бриллюэна на одну и ту же величину, что равносильно смещению занятых состояний как одного целого (рис. 36, б):

$$\mathbf{P}_i(t) = \mathbf{P}_i(0) + e^{-}\mathbf{E}\tau \quad (t \gg \tau). \quad (25.2)$$

Суммарный квазиимпульс \mathbf{P}_Σ всех электронов равен

$$\begin{aligned} \mathbf{P}_\Sigma &= \sum_{i=1}^n \mathbf{P}_i(t) = \sum_{i=1}^n \mathbf{P}_i(0) + ne^{-}\mathbf{E}\tau = \\ &= ne^{-}\mathbf{E}\tau = n\mathbf{P}_{cp}, \end{aligned} \quad (25.3)$$

так как

$$\sum_{i=1}^n \mathbf{P}_i(0) = 0 \quad (25.4)$$

в силу симметрии области занятых состояний, так что каждому состоянию $\mathbf{P}_i(0)$ соответствует состояние $-\mathbf{P}_i(0)$. Состоянию \mathbf{P}_i соответствует скорость электрона

$$\mathbf{v}_i(\mathbf{P}_i) = dE/d\mathbf{P}_i = \mathbf{P}_i/m^* \quad (25.5)$$

и элементарный ток

$$\mathbf{J}_i = e^{-}\mathbf{v}_i = e^{-}\mathbf{P}_i/m^*. \quad (25.6)$$

Полный ток

$$\mathbf{J} = \sum_{i=1}^n \mathbf{J}_i = \frac{e^{-}}{m^*} \sum_{i=1}^n \mathbf{P}_i = \frac{e^{-}}{m^*} ne^{-}\tau\mathbf{E} = e^{-}n\mu_n\mathbf{E}, \quad (25.7)$$

где

$$\mu_n = e^{-}\tau/m^* \quad (25.8)$$

есть подвижность электронов.

Если рассматривать зону Бриллюэна, соответствующую единичному кристаллу, то n — концентрация электронов, а \mathbf{J} — плотность тока, поэтому $e^{-}n\mu_n = \sigma$ есть удельная электрическая проводимость.

Рассмотрим теперь ток, создаваемый почти заполненной зоной, в которой имеется N' занятых и p свободных состояний в окрестности $\mathbf{k} = 0$ ($E(0) = E_{\max}$) (рис. 37, а). Под действием внешнего поля \mathbf{E} все состояния сместятся на одну и ту же величину $\Delta\mathbf{P} = e^{-}\mathbf{E}\tau$ и

$$\mathbf{P}_i(t) = \mathbf{P}_i(0) + e^{-}\mathbf{E}\tau.$$

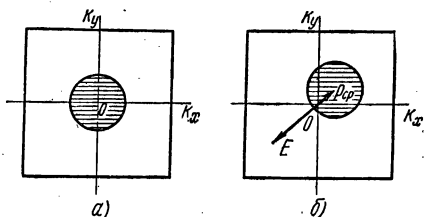


Рис. 36. Смещение электронов в почти свободной зоне Бриллюэна под действием электрического поля

Однако суммарный квазиимпульс не будет равен $N'e^-E\tau$, как это могло бы показаться при определении \mathbf{P}_Σ как суммы $\mathbf{P}_i(t)$. Это связано с тем, что в результате смещения каждого занятого состояния на величину $e^-E\tau$ вся совокупность N' состояний смещается на $e^-E\tau$. Вместе с занятыми состояниями на величину $e^-E\tau$ смещаются и свободные состояния.

В результате перемещения занятых состояний как целого часть состояний из первой зоны переходит во вторую зону, а в первой зоне появляется ряд дополнительных свободных состояний (рис. 37, б). Однако все состояния второй зоны необходимо перевести в первую зону добавлением к \mathbf{P}_i подходящего вектора $2\pi\hbar\mathbf{b}$, вследствие чего занятые состояния будут определяться так, как это указано на рис. 37, в. Но в таком случае \mathbf{P}_Σ не будет равно $N'e^-E\tau$. Чтобы

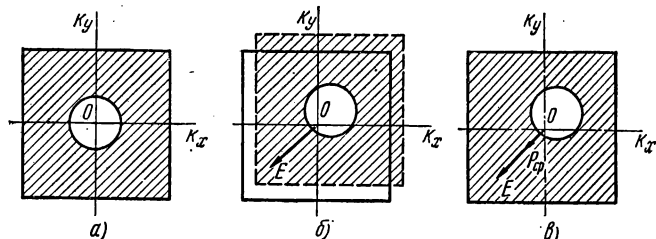


Рис. 37. Смещение электронов в почти заполненной зоне Бриллюэна под действием электрического поля

найти \mathbf{P}_Σ , необходимо учесть свойство зон Бриллюэна для заполненной зоны:

$$\sum_{i=1}^{N'} \mathbf{P}_i = 0, \quad (25.9)$$

поэтому

$$\mathbf{P}_\Sigma = \sum_{i=1}^{N'} \mathbf{P}_i = - \sum_{i=1}^p \mathbf{P}_i = -p e^- E \tau. \quad (25.10)$$

Таким образом, благодаря периодичности состояний в пространстве квазиимпульса суммарный квазиимпульс \mathbf{P}_Σ совокупности N' электронов смещается не вдоль силы e^-E , а против силы, т. е. в направлении $-(e^-E) = e^+E$. Средний квазиимпульс

$$\mathbf{P}_{\text{ср}} = \frac{\mathbf{P}_\Sigma}{N'} = - \frac{p e^- \tau}{N'} \mathbf{E}. \quad (25.11)$$

Средняя скорость

$$\mathbf{v}_{\text{ср}} = \mathbf{P}_{\text{ср}} / m^*. \quad (25.12)$$

Но так как $p \ll N' \approx N$, то $\mathbf{P}_{\text{ср}}$ находится вблизи $\mathbf{k} = 0$ и поэтому $m^* < 0$. Ток, создаваемый N' связанными электронами, равен

$$\mathbf{J}_{\text{связ}} = e^- N' \mathbf{v}_{\text{ср}} = \frac{e^- N' \mathbf{P}_{\text{ср}}}{m^*} = - e^- N' \frac{e^- \tau}{m^*} \frac{p}{N'} \mathbf{E}. \quad (25.13)$$

Если обозначить

$$-\frac{e^{-\tau p}}{m^* N'} = \mu_{\text{связ}} \quad (\mu_{\text{связ}} < 0), \quad (25.14)$$

то

$$\mathbf{J}_{\text{связ}} = -e N' \mu_{\text{связ}} \mathbf{E}. \quad (25.15)$$

Введем теперь понятие дырок на основе следующего условия: дырки — это квазичастицы, динамические свойства которых тождественны динамическим свойствам совокупности электронов, или

$$\mathbf{P}_{\Sigma p} \equiv \mathbf{P}_{\Sigma}, \quad (25.16)$$

где \mathbf{P}_{Σ} есть суммарный квазиимпульс N' занятых состояний, а $\mathbf{P}_{\Sigma p}$ — суммарный квазиимпульс совокупности дырок. Число дырок естественно определить как число свободных состояний. Тогда легко видеть, что квазиимпульс дырки определяется как

$$\mathbf{P}_p = -\mathbf{P}_0, \quad (25.17)$$

где \mathbf{P}_0 — квазиимпульс свободного состояния. Действительно, если в зоне имеется одно свободное состояние, то

$$\mathbf{P}_{\Sigma} = \sum_{i=1}^{N'} \mathbf{P}_i = \sum_{i=1}^{N'} \mathbf{P}_i + \mathbf{P}_0 - \mathbf{P}_0 = \sum_{i=1}^N \mathbf{P}_i - \mathbf{P}_0 = -\mathbf{P}_0. \quad (25.18)$$

Благодаря этому определению квазиимпульса дырки передача квазиимпульса дыркой решетке означает передачу решетке квазиимпульса совокупности электронов. Теперь легко найти закон изменения квазиимпульса дырки: так как

$$\frac{d\mathbf{P}_0}{dt} = e^{-} \mathbf{E}, \quad (25.19)$$

то

$$\frac{d\mathbf{P}_p}{dt} = -\frac{d\mathbf{P}_0}{dt} = -e^{-} \mathbf{E} = e^{+} \mathbf{E}, \quad (25.20)$$

т. е. приращение квазиимпульса дырки происходит таким образом, как если бы заряд дырки был положительным. Итак, дырка есть квазичастица с $e_p > 0$.

Выражение (25.13) для тока можно переписать в виде

$$\mathbf{J}_{\text{связ}} = -e N' \frac{e^{-\tau}}{m^*} \frac{p}{N'} \mathbf{E} = e^+ p \frac{e^{-\tau}}{m^*} \mathbf{E} = e^+ p \frac{e^{\tau}}{m_p^*} \mathbf{E}, \quad (25.21)$$

где

$$m_p^* = -m^* = -\frac{1}{\hbar^2} \frac{d^2 E}{dk^2}. \quad (25.22)$$

Другими словами, если определить массу дырки как массу электрона со знаком (—), то дырка имеет положительную массу, когда она находится у потолка валентной зоны, и отрицательную массу, когда она находится у дна зоны. Из (25.22) следует, что для дырок $E_p(\mathbf{k}) = -E(\mathbf{k})$, т. е. зависимость энергии от квазиимпульса имеет

тот же закон, что и для электронов, только направление отсчета энергии дырок противоположно направлению отсчета энергии для электрона. Именно в связи с этим изменился знак второй производной энергии по квазиимпульсу — там, где для электрона имеется максимум ($m^* < 0$), для дырки — минимум энергии ($m_p^* > 0$). Скорость дырки можно определить двояким способом: с одной стороны,

$$v_p = \frac{dE_p}{dP_p} = \frac{-dE}{-dP} = v, \quad (25.23)$$

с другой стороны,

$$v_p = \frac{P_p}{m_p^*} = \frac{-P_0}{-m^*} = v, \quad (25.24)$$

т. е. скорость дырки есть скорость свободного состояния.

Таким образом, мы определили дырку как квазичастицу, характеризующуюся следующими величинами:

заряд дырки

$$e_p = e^+ = -e > 0;$$

эффективная масса дырки

$$m_p^{*-1} = -m^{*-1} = -\frac{1}{\hbar^2} \frac{d^2E}{dk^2} = \frac{1}{\hbar^2} \frac{d^2E_p}{dk_p^2}; \quad (25.25)$$

квазиимпульс дырки

$$P_p = -P_0; \quad P_{\Sigma p} = \sum_{i=1}^p P_{ip} = \sum_{i=1}^{N'} P_i = P_{\Sigma};$$

скорость дырки

$$v_p = \frac{P_p}{m_p^*} = \frac{dE_p}{dP_p} = \frac{dE}{dP} = v;$$

волновой вектор дырки

$$k_p = -k;$$

энергия дырки

$$E_p(k_p) = -E(-k);$$

подвижность дырки

$$\mu_p = \frac{e_p \tau}{m_p^*} = \frac{-e \tau}{-m^*} = -\mu_n.$$

Таким образом, все характеристики дырок определяются через характеристики электрона в валентной зоне.

Дрейфовая скорость дырок направлена по полю:

$$v_{др} = \mu_p E, \quad (25.26)$$

в то время как дрейфовая скорость совокупности связанных электронов направлена против поля:

$$\mu_{связ} = -\frac{e \tau}{m^*} \frac{p}{N'} = -\frac{e^+ \tau}{m_p^*} \frac{p}{N'} = -\mu_p \frac{p}{N'}. \quad (25.27)$$

Определим функцию распределения дырки f_p :

$$f_p = 1 - f = 1 - \frac{1}{e^{\frac{E-F}{kT}} + 1} = \frac{1}{e^{\frac{F-E}{kT}} + 1}, \quad (25.28)$$

которую можно записать подобно функции распределения электронов, если определить уровень Ферми для дырок условием

$$F_p = -F. \quad (25.29)$$

Тогда

$$f_p = \frac{1}{e^{\frac{E_p - F_p}{kT}} + 1}. \quad (25.30)$$

В этом случае движение совокупности связанных электронов в электрическом поле полностью описывается движением совокупности дырок. Аналогично можно показать, что движение совокупности электронов валентной зоны в магнитном поле полностью описывается движением совокупности дырок.

Резюме § 25

1. Дырочная проводимость есть проводимость электронов валентной зоны или любой другой зоны, занятой более чем наполовину.
2. Дырка — это квазичастица, свойства которой определяются через свойства электрона в валентной зоне соотношениями (25.25—30). Концентрация дырок есть концентрация свободных состояний в валентной зоне.

§ 26. ЗОННАЯ СТРУКТУРА НЕКОТОРЫХ ПОЛУПРОВОДНИКОВ

Методы расчета

До сих пор мы рассматривали самые общие методы расчета энергетической структуры и в общем виде показали наличие зон энергии и их некоторые характеристики. Расчет зонной структуры конкретных полупроводниковых веществ крайне затруднен в силу целого ряда причин, и прежде всего потому, что отсутствует аналитическое выражение для $U(\mathbf{r})$. Поэтому при любых расчетах в формулах содержатся некоторые параметры, значение которых определяется на основе сравнения расчетов с экспериментальными данными. Например, ширина запрещенной зоны определяется только экспериментально.

В настоящее время существует много методов расчета энергетической структуры. Рассмотрим кратко некоторые из них.

а) Метод плоских волн (метод ПВ).

Представив периодическое поле решетки в виде комбинации фурье-компонентов, имеем разложение вида (17.10):

$$U(\mathbf{r}) = \sum_{\mathbf{b}} c_{\mathbf{b}} e^{i2\pi(\mathbf{b}\mathbf{r})}, \quad (26.1)$$

где $c_{\mathbf{b}}$ определяется соотношениями (17.12) и (17.9). Поскольку $\varphi_{\mathbf{k}}(\mathbf{r})$ — функция периодическая, ее можно разложить в ряд Фурье:

$$\varphi(\mathbf{r}) = \sum_{\mathbf{g}'} a_{\mathbf{g}'} e^{i2\pi(\mathbf{g}'\mathbf{r})}. \quad (26.2)$$

В таком случае волну Блоха можно представить в виде

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{g}'} a_{\mathbf{g}'} e^{i2\pi\left(\frac{\mathbf{k}}{2\pi} + \mathbf{g}', \mathbf{r}\right)}. \quad (26.3)$$

Подставив выражения для U и $\psi_{\mathbf{k}}$ в уравнение Шредингера, получим

$$a_{\mathbf{g}} \frac{\hbar^2 (\mathbf{k} + 2\pi\mathbf{g})^2}{2m} + \sum_{\mathbf{b}} c_{\mathbf{b}} a_{\mathbf{g}-\mathbf{b}} = E a_{\mathbf{g}}. \quad (26.4)$$

Система уравнений (26.4) относительно $a_{\mathbf{g}}$ позволяет найти $E(\mathbf{k})$ из секулярного уравнения в предположении, что $c_{\mathbf{b}}$ известны, т. е. известно выражение для поля решетки. При вычислении $E(\mathbf{k})$ необходимо брать большое число плоских волн, что существенно затрудняет практическое применение этого метода.

б) Метод ортогонализированных плоских волн (ОПВ), или метод Херринга.

Плохая сходимость рядов при вычислении $E(\mathbf{k})$ в методе ПВ связана с тем, что вблизи ядра волновая функция существенно отличается от плоской волны, поэтому для ее разложения требуется использование большого числа плоских волн с малыми длинами волн, или большими \mathbf{k} . Вблизи ядер состояние лучше описывается атомной волновой функцией $\psi_j^a(\mathbf{r} - \mathbf{n})$, поэтому можно составить из них сумму вида

$$\varphi_{\mathbf{k}j}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{n}} e^{i(\mathbf{k}\mathbf{n})} \psi_j^a(\mathbf{r} - \mathbf{n}), \quad (26.5)$$

где j — некоторое атомное состояние (s, p, d и т. п.); N — число атомов в кристалле, положение которых в решетке определяется вектором \mathbf{n} . Рассмотрим выражение вида:

$$\chi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{L^{3/2}} e^{i(\mathbf{k}\mathbf{r})} - \sum_j \mu_{\mathbf{k}j} \varphi_{\mathbf{k}j}(\mathbf{r}). \quad (26.6)$$

Вблизи некоторого ядра $\chi_{\mathbf{k}}(\mathbf{r})$ ведет себя в основном как соответствующая атомная волновая функция, вдали от ядер она подобна

плоской волне. Для упрощения вычислений можно потребовать, чтобы функции $\chi_k(\mathbf{r})$ были ортогональны всем «основным» состояниям $\varphi_{kj}(\mathbf{r})$:

$$\int \varphi_{kj}^*(\mathbf{r}) \chi_k(\mathbf{r}) d\tau = 0. \quad (26.7)$$

Это позволяет однозначно выбирать коэффициенты μ_{kj} . Метод ОПВ существенно упрощает вычисления, улучшая сходимость рядов при вычислении $E(k)$.

в) Метод присоединенных плоских волн (ППВ), или метод Слэтера.

В методе ППВ волновая функция для кристалла строится в виде комбинации плоских волн *вне атома* и атомных волновых функций $\psi_{lm}^a(\mathbf{r})$ *внутри атома*:

$$\Phi_k(\mathbf{r}) = \theta(r - r_i) a_0 e^{i(kr)} + \sum_{lm} a_{lm} \theta(r_i - r) \psi_{lm}^a(\mathbf{r}). \quad (26.8)$$

Атом при этом моделируется в виде шара радиуса r_i , $\theta(\xi)$ — единичная ступенчатая функция, включающая в действие либо плоскую волну при $r > r_i$, либо атомную функцию при $r < r_i$:

$$\theta(\xi) = \begin{cases} 1 & \text{при } \xi > 0 \\ 0 & \text{при } \xi < 0. \end{cases} \quad (26.9)$$

г) Метод псевдопотенциала.

На основе метода ОПВ был развит метод, получивший в дальнейшем название метода псевдопотенциала (Филлипс и Клейнман), широко используемый в настоящее время для расчета зонной структуры.

Пусть $\psi_k^{(\alpha)}(\mathbf{r})$ — искомая волновая функция, являющаяся решением уравнения Шредингера для кристалла, а $\varphi_{kj}(\mathbf{r})$ — сумма Блоха, составленная из атомных волновых функций таким образом, чтобы она была ортогональна функции $\psi_k^{(\alpha)}(\mathbf{r})$:

$$\int \psi_k^{(\alpha)*}(\mathbf{r}) \varphi_{kj}(\mathbf{r}) d\tau = 0. \quad (26.10)$$

Построим функцию $\Phi(\mathbf{r})$ следующим образом:

$$\Phi(\mathbf{r}) = \psi_k^{(\alpha)}(\mathbf{r}) - \sum_j a_{kj} \varphi_{kj}(\mathbf{r}). \quad (26.11)$$

Поскольку вблизи ядер $\psi_k^{(\alpha)}(\mathbf{r})$ ведет себя как одна из атомных функций, входящих в сумму Блоха, $\Phi(\mathbf{r})$ будет плавной функцией всюду, в том числе и вблизи ядер. Коэффициенты a_{kj} находятся из условия ортогональности $\psi_k^{(\alpha)}(\mathbf{r})$ и $\varphi_{kj}(\mathbf{r})$:

$$a_{kj} = - \int \varphi_{kj}^*(\mathbf{r}) \Phi(\mathbf{r}) d\tau = - (\varphi_{kj} | \Phi). \quad (26.12)$$

Найдем уравнение для $\Phi(\mathbf{r})$. Исходя из условий

$$\hat{H}\psi_{\mathbf{k}}^{(\alpha)}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}^{(\alpha)}(\mathbf{r}); \quad \hat{H}\varphi_{\mathbf{k}j} = E_j\varphi_{\mathbf{k}j}, \quad (26.13)$$

получим

$$\left\{ -\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}) \right\} \Phi(\mathbf{r}) + \sum_j a_{\mathbf{k}j} E_j \varphi_{\mathbf{k}j} = E(\mathbf{k}) \Phi(\mathbf{r}) + E(\mathbf{k}) \sum_j a_{\mathbf{k}j} \varphi_{\mathbf{k}j}. \quad (26.14)$$

Перепишем (26.14) следующим образом:

$$\left\{ \left(-\frac{\hbar^2}{2m}\Delta + U(\mathbf{r}) \right) + \sum_j \frac{a_{\mathbf{k}j} (E_j - E) \varphi_{\mathbf{k}j}}{\Phi(\mathbf{r})} \right\} \Phi(\mathbf{r}) = E\Phi(\mathbf{r}). \quad (26.15)$$

Член $\sum_j \frac{a_{\mathbf{k}j} (E_j - E) \varphi_{\mathbf{k}j}}{\Phi(\mathbf{r})} = V_R(\mathbf{r})$ называется потенциалом отталкивания, а $V_p = U + V_R$ называется псевдопотенциалом.

Теперь можно записать

$$\left\{ -\frac{\hbar^2}{2m}\Delta + V_p \right\} \Phi(\mathbf{r}) = E(\mathbf{k}) \Phi(\mathbf{r}). \quad (26.16)$$

Поскольку $\Phi(\mathbf{r})$ достаточно гладкая функция, ее можно образовать из небольшого числа плоских волн, которые обеспечивают быструю сходимость ряда для получения дисперсионной зависимости $E(\mathbf{k})$. Более того, псевдопотенциал можно рассматривать в качестве возмущения.

В разновидности метода, получившего название эмпирического метода псевдопотенциала, для нахождения V_p используются результаты экспериментальных исследований, прежде всего спектра отражения или вычисляемого из него спектра мнимой части диэлектрической проницаемости. Используя данные по положению пиков отражения, удастся вычислять значение $E(\mathbf{k})$ в произвольных точках зоны Бриллюэна, что является существенным достоинством метода по сравнению с другими. Его слабой стороной является необходимость подгонки параметров форм-фактора для получения численных значений межзонных расстояний.

д) Вариационный принцип, или метод ККР (метод Кона — Кориннга — Ростокера).

Величина I , определяемая соотношением

$$I = \int \psi^* (\hat{H} - E) \psi d\tau, \quad (26.17)$$

может вычисляться с произвольной функцией $\psi(\mathbf{r})$, поэтому I является функционалом, зависящим от вида $\psi(\mathbf{r})$: $I = I\{\psi(\mathbf{r})\}$. Если подставить под интеграл собственную функцию гамильтониана, то получим $I = 0$. Незначительное отклонение $\psi(\mathbf{r})$ от собственной должно мало сказаться на значении I . Другими словами, вариация собственной функции гамильтониана должна обеспечивать нулевое изменение функционала, т. е. $\delta I = 0$, при $\delta\psi \neq 0$.

Выбор пробных функций и модели потенциала позволяет получить расчет зонной структуры без дополнительной «подгонки» параметров.

Минимальную неопределенность в расчет вносит метод, основанный на релятивистском гамильтониане, однако результаты, получаемые на его основе, достаточно далеки от результатов, получаемых другими методами. Как видим, практически используемые методы расчета по существу представляют некоторую комбинацию методов квазисвободного и квазисвязанного электрона.

е) к·р-метод. Для описания многих явлений широко используется дисперсионное соотношение, полученное Кейном на основе модифицированного метода теории возмущения, суть которого можно понять, используя соотношение (11.19), записанное в виде:

$$\left\{ -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}) + \frac{\hbar}{m} (\mathbf{k}\mathbf{p}) \right\} \varphi_{\mathbf{k}n}(\mathbf{r}) = \left[E_n(\mathbf{k}) - \frac{\hbar^2 \mathbf{k}^2}{2m} \right] \varphi_{\mathbf{k}n}(\mathbf{r}). \quad (26.18)$$

В уравнении (26.18) у волновой функции и энергии добавлен второй индекс n , показывающий принадлежность состояний энергетической зоне n . Если считать, что для некоторого состояния \mathbf{k}_0 решение уравнения Шредингера известно, то можно искать решение в окрестности этого состояния:

$$\varphi_{\mathbf{k}n}(\mathbf{r}) = \sum_{n'} c_{n'n}(\mathbf{k} - \mathbf{k}_0) \varphi_{\mathbf{k}'n'}(\mathbf{r}), \quad (26.19)$$

$$\hat{H}_{\mathbf{k}_0} = \frac{\hat{\mathbf{p}}^2}{2m} + \frac{\hbar}{m} (\mathbf{k}_0 \hat{\mathbf{p}}) + \frac{\hbar^2 \mathbf{k}_0^2}{2m} + U(\mathbf{r}) \quad (26.20)$$

и

$$\hat{H}_{\mathbf{k}_0} \varphi_{\mathbf{k}_0 n} = E_n(\mathbf{k}_0) \varphi_{\mathbf{k}_0 n}. \quad (26.21)$$

Используя (26.20—21), можно записать (26.18) в виде

$$\left[\hat{H}_{\mathbf{k}_0} + \frac{\hbar}{m} (\mathbf{k} - \mathbf{k}_0, \mathbf{p}) + \frac{\hbar^2}{2m} (\mathbf{k}^2 - \mathbf{k}_0^2) \right] \varphi_{\mathbf{k}n} = E_n(\mathbf{k}) \varphi_{\mathbf{k}n}. \quad (26.22)$$

Подставим (26.19) в (26.22) и, умножив слева на $\varphi_{\mathbf{k}_0 n}$ и проинтегрировав по объему кристалла, получим уравнение (26.22) в матричной форме:

$$\sum_{n'} \left[\left\{ E_n(\mathbf{k}_0) + \frac{\hbar^2}{2m} (\mathbf{k}^2 - \mathbf{k}_0^2) \right\} \delta_{n'n} + \frac{\hbar}{m} (\mathbf{k} - \mathbf{k}_0, \mathbf{p}_{n'n}) \right] c_{n'n} = E_n(\mathbf{k}) c_{n'n}, \quad (26.23)$$

где

$$\mathbf{p}_{n'n} = \int \varphi_{n'\mathbf{k}_0}^*(\mathbf{r}) \hat{\mathbf{p}} \varphi_{\mathbf{k}_0 n}(\mathbf{r}) d\tau. \quad (26.24)$$

Член, содержащий матричный элемент импульса $\mathbf{p}_{n'n}$, можно принять в качестве возмущения, считая, что \mathbf{k} находится в окрестности

сти κ_0 . В этом случае во втором порядке обычной теории возмущения можно записать

$$E_n(\kappa) = E_n(\kappa_0) + \frac{\hbar}{m} (\kappa - \kappa_0, p_{nn}) + \frac{\hbar^2}{2m} (\kappa^2 - \kappa_0^2) + \frac{\hbar^2}{m^2} \sum_{n'} \frac{|(\kappa - \kappa_0, p_{nn'})|^2}{E_n(\kappa_0) - E_{n'}(\kappa_0)}. \quad (26.25)$$

Если состояние κ_0 соответствует экстремуму и при этом $p_{nn} + \hbar \kappa_0 = 0$, то выражение (26.25) описывает сфероидальные поверхности энергии, для которых компоненты эффективной массы по главным осям i имеют вид

$$\frac{1}{m_i} = \frac{1}{m} + \frac{2}{m^2} \sum_{n'} \frac{(ip_{nn'})^2}{E_n(\kappa_0) - E_{n'}(\kappa_1)} \quad (26.26)$$

и

$$E_n(\kappa) = E_n(\kappa_0) + \frac{\hbar^2}{2} \sum_i \frac{(\kappa_i - \kappa_{i0})^2}{m_i}. \quad (26.27)$$

Таким образом, κ - p -метод в этом случае приводит к известным результатам. Новые результаты получаются при учете спин-орбитального взаимодействия, физическая природа которого рассматривается в § 84, стр. 568. Оператор Гамильтона для спинорбитального взаимодействия в общем случае имеет вид

$$\hat{H}_{so} = \frac{\hbar^2}{4m^2c^2} ([\nabla U, \hat{p}] \sigma), \quad (26.28)$$

где σ — оператор спина (матрицы Паули). В κ - p -представлении (26.28) должен быть заменен двумя членами:

$$\hat{H}_{so} = \hat{H}_1 + \hat{H}_2, \quad (26.29)$$

где \hat{H}_1 совпадает с общим выражением (26.28), а второй член

$$\hat{H}_2 = \frac{\hbar^2}{4m^2c^2} ([\nabla U, \kappa] \sigma) \quad (26.30)$$

соответствует линейному члену в (26.18) и отражает особенность κ - p -представления. Включение других членов, определяющих спин-орбитальное взаимодействие, не приводит к существенным поправкам. Дальнейший расчет проводится с учетом симметрии \hat{H}_0 и \hat{H}_{so} в предположении, что взаимодействие имеет место только между зоной проводимости и валентной зоной, волновые функции которых соответствуют s - и трем p -волновым атомным функциям.

Ограничимся приведением результатов расчета дисперсионных соотношений. Для зоны проводимости можно получить

$$E_c(\kappa) = \frac{E_g}{2} + \frac{\hbar^2 \kappa^2}{2m} + \frac{E_g}{2} \left\{ 1 + 4 \frac{\hbar^2 \kappa^2}{2m c} \frac{f_1(E_c)}{E_g} \right\}^{1/2}. \quad (26.31)$$

Для трех валентных зон получено:

$$E_{1v}(k) = -\frac{\hbar^2 k^2}{2m} - \frac{\hbar^2 k^2}{2m_v} \left\{ 1 - \gamma' \frac{\kappa_x^2 \kappa_y^2 + \kappa_y^2 \kappa_z^2 + \kappa_z^2 \kappa_x^2}{\kappa^4/3} \right\};$$

$$E_{1v}(0) = 0; \quad (26.32)$$

$$E_{2v}(k) = \frac{E_g}{2} - \frac{\hbar^2 k^2}{2m} - \frac{E_g}{2} \left\{ 1 + 4 \frac{\hbar^2 k^2}{2m_c} \frac{2(E_g + \Delta)}{3E_g + 2\Delta} \frac{f_2(E_{2v})}{E_g} \right\}^{1/2};$$

$$E_{2v}(0) = 0; \quad (26.33)$$

$$E_{3v}(k) = -\frac{\Delta}{2} - \frac{\hbar^2 k^2}{2m} - \frac{\Delta}{2} \left\{ 1 + 4 \frac{\hbar^2 k^2}{2m_c} \frac{E_g}{3E_g + 2\Delta} \frac{f_3(E_{3v})}{\Delta} \right\}^{1/2};$$

$$E_{3v}(0) = -\Delta. \quad (26.34)$$

В выражениях (26.31—34) за начало отсчета энергии принята максимальная энергия в валентной зоне, которую имеют две зоны — зона тяжелых — E_{1v} — и зона легких — E_{2v} — дырок. Третья валентная зона — E_{3v} — опущена вниз на величину Δ , представляющую собой энергию спин-орбитального взаимодействия. Максимум энергии валентной зоны и минимум энергии зоны проводимости находятся в точке $k=0$, в так называемой точке Γ (гамма). Параметр Кейна E_g , определяющий положение дна зоны проводимости, представляет собой запрещенную зону, при этом допускается и отрицательное значение E_g . Через f_1 , f_2 , f_3 обозначены медленно меняющиеся функции параметров закона Кейна, принимающие значение, равное единице в точке Γ . Параметр m_c определяется соотношением

$$\frac{1}{m_c} = 2 \frac{P^2}{\hbar^2} \frac{E_g + \frac{2}{3} \Delta}{E_g(E_g + \Delta)}, \quad (26.35)$$

где параметр P связан с матричным элементом импульса для взаимодействующих зон:

$$P = -i \frac{\hbar}{m} P_{cv}. \quad (26.36)$$

Обычно параметр P измеряют в электронвольтах на сантиметр, и для многих полупроводников можно считать его равным примерно $6 \cdot 10^{-8}$ эВ·см.

Для «узкозонных» материалов при $kP \ll \Delta$ и $E_g \ll \Delta$ выражения (26.31—34) приобретают вид:

$$E_c(k) \approx \frac{E_g}{2} + \frac{\hbar^2 k^2}{2m} + \left\{ \frac{E_g^2}{4} + \frac{2}{3} P^2 k^2 \right\}^{1/2}; \quad (26.37)$$

$$E_{1v}(k) \approx -\frac{\hbar^2 k^2}{2m}; \quad (26.38)$$

$$E_{2v}(k) \approx \frac{E_g}{2} - \frac{\hbar^2 k^2}{2m} - \left\{ \frac{E_g^2}{4} + \frac{2}{3} P^2 k^2 \right\}^{1/2}; \quad (26.39)$$

$$E_{3v}(k) \approx -\Delta - \frac{\hbar^2 k^2}{2m} - \frac{P^2 k^2}{3(E_g + \Delta)}. \quad (26.40)$$

Таким образом, соотношение Кейна указывает на заметную роль линейного члена гамильтониана (26.18) для узкозонных материалов. При $E_g = 0$ зона проводимости и зона легких дырок имеют линейный дисперсионный закон, при этом в точке Γ имеет место трехкратное вырождение по энергии. При $E_g \neq 0$ отклонение от квадратичной зависимости, или, как говорят, от параболичности, будет наблюдаться при достаточно больших k .

При использовании любого метода приходится встречаться с большими математическими трудностями. Так, например, при вычислениях методом ортогонализированных плоских волн для германия и кремния уравнение для определения $E(k)$ в произвольной точке k

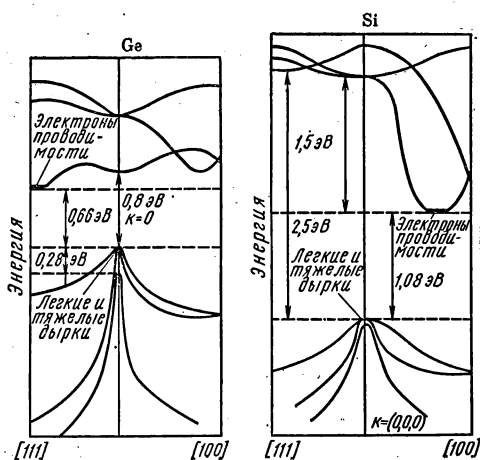


Рис. 38. Зонная структура германия и кремния

является уравнением 146-й степени. В этом случае необходимо использовать методы теории групп, учитывающие симметрию кристалла. Для некоторых симметричных точек уравнения существенно упрощаются, но тем не менее остаются уравнениями 16-й степени.

Имеющиеся к настоящему времени данные о зонной структуре полупроводников получены на основе как теоретических расчетов, так и экспериментальных работ. Рассмотрим структуру зон некоторых полупроводников.

Кремний. Прежде всего напомним, что атом кремния имеет 14 электронов, распределение их по состояниям таково:

$(1s^2) (2s^2) (2p^6) (3s^2) 3p^2 3p_{012}$, т. е. две оболочки заполнены, третья оболочка не заполнена. Спины двух p -электронов параллельны, поэтому основное состояние является триплетным. Если исходить из теории квазисвязанного электрона, то для построения нулевого приближения функции Блоха нужно использовать волновые функции p -состояний, которые *трижды вырождены* (без учета спина). При *учете взаимодействия частиц в кристалле вырождение снимается и образуются три полосы зависимости $E(k)$* , которые частично перекрываются, поэтому как валентная, так и зона проводимости представляют собой наложение трех различных зон. На рис. 38 это изображено тремя ветвями $E(k)$. Одна из ветвей $E(k)$ зоны проводимости лежит значительно ниже других. Положение абсолютного минимума и определяет дно зоны проводимости. Он лежит в направлении $[100]$, поэтому всего имеется 6 эквивалентных минимумов энергии. Минимум энергии для электронов и дырок называют иногда долинами, поэтому говорят, что зона проводимости кремния имеет шесть долин. Как

видно из рис. 38, зависимость $E(k)$ для разных направлений различна.

Изоэнергетические поверхности около абсолютных минимумов представляют собой эллипсоиды вращения относительно большой полуоси, которая совпадает с направлением $[100]$ (рис. 39). Зависимость энергии от k можно представить в виде

$$E(k) = E(k_0) + \frac{\hbar^2 [(\kappa_1 - \kappa_{01})^2 + (\kappa_2 - \kappa_{02})^2]}{2m_1} + \frac{\hbar^2 (\kappa_3 - \kappa_{03})^2}{2m_3},$$

$$m_1 = m_2 \neq m_3. \quad (26.41)$$

Опыты по циклотронному резонансу дают для компонентов тензора эффективной массы электрона в кремнии следующие значения:

$$m_1 = m_2 = m_t = 0,19m; \quad m_3 = m_l = 0,98m.$$

Отношение $m_3/m_1 = 5,16$ характеризует анизотропию свойств изоэнергетических поверхностей. Отношение полуосей эллипсоида равно

$$\sqrt{m_3/m_1} = 2,27.$$

Точка минимума находится недалеко от границы зоны Бриллюэна. На рис. 39 указаны изоэнергетические поверхности зоны проводимости кремния вокруг абсолютного минимума.

Для валентной зоны максимум энергии находится в центре зоны Бриллюэна $k_0 = 0$ для всех трех полос. При этом в точке k_0 все три зоны смыкаются, так что энергия в центре зоны Бриллюэна оказывается вырожденной. Учет спин-орбитального взаимодействия (тонкой структуры уровней) приводит к тому, что вырождение частично снимается — одна из зон опускается на величину $E_{s0} = 0,035$ эВ. Связь между энергией и волновым вектором задается соотношением

$$E_{1,2}(k) = E(0) - \frac{\hbar^2}{2m} [Ak^2 \pm \sqrt{B^2k^4 + C^2(\kappa_x^2\kappa_y^2 + \kappa_x^2\kappa_z^2 + \kappa_y^2\kappa_z^2)}], \quad (26.42)$$

где A, B, C — безразмерные константы, равные соответственно $4,1 \pm 0,2$; $1,6 \pm 0,2$; $3,3 \pm 0,5$. Изоэнергетические поверхности представляют собой деформированные сферы (гофрированные поверхности, см. рис. 39). В сферической системе координат выражение $E(k)$ можно представить в виде

$$E_{1,2}(k) = E(0) - \frac{\hbar^2 k^2}{2m} [A \pm \sqrt{B^2 + C^2 \sin^2 \theta (\sin^2 \varphi \cos^2 \varphi \sin^2 \theta + \cos^2 \theta)}]. \quad (26.43)$$

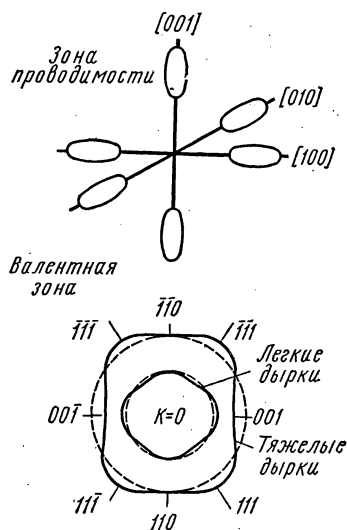


Рис. 39. Поверхности постоянной энергии в кремнии

Очевидно, что при заданном значении k подкоренное выражение может меняться от B^2 до $B^2 + \frac{5C^2}{16}$, а энергия будет меняться при этом от $A \pm B$ до $A \pm \sqrt{B^2 + \frac{5C^2}{16}}$ (в единицах $\frac{\hbar^2 \kappa^2}{2m}$), т. е. от 5,7 до 6,5 и от 2,5 до 2,7 соответственно. Если найдем вторую производную в точке $k=0$, то получим $\frac{1}{\hbar^2} \frac{d^2 E}{d \kappa^2} = \frac{A \pm B}{m}$, последнее соотношение равносильно пренебрежению членами при C , что невозможно. Если усреднить выражение для $E_{1,2}$ по углам, то можно записать

$$E_{1,2} = E(0) - \frac{\hbar^2 \kappa^2}{2m} \left[A \pm \sqrt{B^2 + \frac{C^2}{5}} \right],$$

т. е. гофрированная поверхность заменяется некоторой «средней» сферической поверхностью. В этом случае будем считать эффективную массу скалярной величиной

$$m_p^* = \frac{m}{A \pm \sqrt{B^2 + \frac{C^2}{5}}}, \quad (26.44)$$

при этом должны существовать два вида дырок:
тяжелые

$$m_T^* = \frac{m}{A - \sqrt{B^2 + \frac{C^2}{5}}}; \quad m_T^* = 0,52m \quad (26.45)$$

и *легкие*

$$m_L^* = \frac{m}{A + \sqrt{B^2 + \frac{C^2}{5}}}; \quad m_L^* = 0,16m \quad (26.46)$$

с отношением масс

$$\frac{m_T^*}{m_L^*} = 3,3. \quad (26.47)$$

Опыт дает $m_T^* = 0,49m$, $m_L^* = 0,16m$ и $m_T^*/m_L^* = 3,1$.

Для третьей ветви выражение

$$E_3(k) = E(0) - E_{so} - \frac{\hbar^2 \kappa^2}{2m} A; \quad E_{so} = 0,035 \text{ эВ}. \quad (26.48)$$

Эффективная масса m_c^* третьего типа дырок является скалярной величиной, она равна $m_c^* = \frac{m}{A}$; $m_c^* = 0,24m$, т. е. принимает среднее значение между массами легких и тяжелых дырок. Однако экспериментально дырки со средним значением массы не наблюдаются, поскольку соответствующая им полоса энергии опущена на величину 0,035 эВ по сравнению с энергией легких и тяжелых дырок.

Минимальное расстояние между дном зоны проводимости и потолком валентной зоны соответствует значениям энергии в различ-

ных точках зоны Бриллюэна, что имеет существенное значение для целого ряда физических явлений. Именно это минимальное расстояние и называется шириной запрещенной зоны, определяющей протекание всех процессов, связанных с термическим возбуждением.

Германий. Зона проводимости германия состоит из трех перекрывающихся полос энергии. Абсолютный минимум находится в точках зоны Бриллюэна, соответствующих направлению [111]. Следовательно, имеется восемь эквивалентных минимумов. Координаты минимумов имеют вид $(1/2, 1/2, 1/2)$ в единицах $\frac{2\pi}{a}$, т. е. они лежат на границе зоны Бриллюэна. Изоэнергетические поверхности имеют вид эллипсоида вращения с осью симметрии, совпадающей с кристаллографическими направлениями [111] (рис. 40). Выражение для энергии имеет вид, аналогичный соответствующему выражению в кремнии:

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2 (\kappa_1 - \kappa_{01})^2 + \hbar^2 (\kappa_2 - \kappa_{02})^2}{2m_1} + \frac{\hbar^2 (\kappa_3 - \kappa_{03})^2}{2m_3}. \quad (26.49)$$

Опыты по циклотронному резонансу дают

$$m_1 = m_2 = m_t = (0,082 \pm 0,001) m;$$

$$m_3 = m_l = (1,58 \pm 0,04) m.$$

Отношение квадратов полуосей равно

$$\frac{m_3}{m_1} = \frac{m_l}{m_t} = 19,3 \text{ и } \sqrt{\frac{m_3}{m_1}} = 4,4.$$

Зависимость энергии от \mathbf{k} для германия приведена на рис. 38 для двух разных направлений в зоне Бриллюэна. Поскольку минимум находится на границе зоны Бриллюэна, на первую зону Бриллюэна приходится половина эллипсоида энергии, в результате чего в германии имеется не восемь, а только четыре полных эллипсоида энергии.

Валентная зона германия имеет структуру, совершенно аналогичную структуре кремния (см. рис 38, 40). Величину усредненных эффективных масс легких и тяжелых дырок можно найти из значения констант: $A = 13,0 \pm 0,2$; $B = 8,9 \pm 0,1$; $C = 10,3 \pm 0,2$. Опыт дает $m_l^* = 0,04m$, $m_t^* = 0,34m$. Третья зона опущена на величину энергии спин-орбитального взаимодействия $E_{so} = 0,28$ эВ. Средние дырки экспериментально не наблюдаются.

Интерметаллы. Соединения элементов третьей и пятой групп, обозначаемые $A^{III}B^V$ и называемые интерметаллами, в последнее время имеют все большее значение благодаря целому ряду их свойств. Наиболее известны среди них GaAs; InSb; GaP; GaSb.

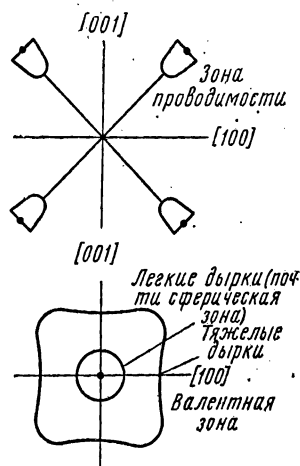


Рис. 40. Поверхности постоянной энергии в германии

Расчеты зонной структуры интерметаллов проводятся сравнением со структурой веществ IV группы. Это связано с тем, что интерметаллы имеют решетку типа цинковой обманки, в то время как элементы IV группы кристаллизуются в решетке типа алмаза. Как известно, эти решетки с точки зрения расположения узлов в пространстве совершенно подобны, но атомы типа А и В в решетке типа цинковой обманки чередуются, в то время как решетка типа алмаза построена из одинаковых атомов. Вследствие этого периодическое поле решетки интерметаллов не имеет центра инверсии, т. е. $U(\mathbf{r}) \neq U(-\mathbf{r})$. Предполагается, что потенциальное поле решетки соединения $A^{III}B^V$ можно представить в виде комбинации полей веществ IV группы и некоторой антисимметричной части, которую можно рассматривать в качестве возмущения, приводящего к изменению зон веществ IV группы, структура которых предполагается известной. Если обозначить потенциальную энергию соединений $A^{III}B^V$ в виде $U^{III-V}(\mathbf{r})$, а потенциальную энергию «соединений» $A^{IV}B^{IV}$ через $U^{IV-IV}(\mathbf{r})$, то

$$U^{III-V}(\mathbf{r}) = U_s^{IV-IV}(\mathbf{r}) + U_a^{IV-IV}(\mathbf{r}) + [\Delta U_s(\mathbf{r}) + \Delta U_a(\mathbf{r})], \quad (26.50)$$

$$\text{где } U_s^{IV-IV}(\mathbf{r}) = U_s^{IV-IV}(-\mathbf{r}); \quad U_a^{IV-IV}(\mathbf{r}) = -U_a^{IV-IV}(-\mathbf{r}),$$

т. е. симметричная U_s и антисимметричная U_a части поля «соединений» $A^{IV}B^{IV}$. Для простого вещества $U_a(\mathbf{r}) = 0$, ΔU_s и ΔU_a соответственно симметричная и антисимметричная части возмущения. Для соединений элементов, стоящих в одном периоде таблицы Менделеева, возмущение в основном должно быть антисимметричным, поэтому поправка к энергии «соединения» $A^{IV}B^{IV}$ в первом приближении должна быть пренебрежимо малой и необходимо использовать второе приближение теории возмущения. При этом для BN расчет основан на структуре C (алмаз), GaAs \rightarrow Ge, AlP \rightarrow Si, InSb \rightarrow α -Sn.

Анализ имеющихся экспериментальных данных совместно с теоретическими расчетами позволяет сделать следующие заключения об энергетической структуре интерметаллов. *Изоэнергетические поверхности как валентной зоны, так и зоны проводимости, являются сферическими, поэтому эффективная масса электронов и дырок является скалярной величиной. Зоны образованы наложением трех различных полос, поскольку они возникают из p-уровней.*

Валентная зона интерметаллов несколько отличается от валентной зоны веществ IV группы — она также состоит из трех полос, одна из которых вследствие спин-орбитального взаимодействия опущена на величину E_{so} , имеющую различное значение для разных соединений. *Две другие полосы, соответствующие легким и тяжелым дыркам, расщепляются в центре зоны Бриллюэна в силу наличия антисимметричной части потенциального поля. Благодаря этому максимумы энергии легких и тяжелых дырок смещены относительно зоны Бриллюэна. При этом может быть смещен максимум одной полосы, или обеих полос. Однако смещение, как правило, очень незна-*

чительно, так что энергия $E(0)$ меньше энергии в максимуме на сотые и даже тысячные доли электронвольта. Наименьшее смещение (возможно, равное нулю) имеет место для InSb и InAs. Максимумы в валентной зоне, как правило, смещены относительно центра зоны Бриллюэна в направлении $[111]$ или $[100]$ (рис. 41). Зона проводимости состоит из трех полос $E(k)$. Абсолютный минимум находится как в центре зоны Бриллюэна, так и в других точках (рис. 42).

Малая величина эффективной массы электронов приводит к быстрому росту энергии с ростом волнового вектора, при этом уже для небольших значений k энергия определяется не только квадратичным членом $\hbar^2 k^2 / (2m^*)$, но и более высокими степенями k , поэтому величина массы, определяемая в образцах с различной концентрацией электронов,

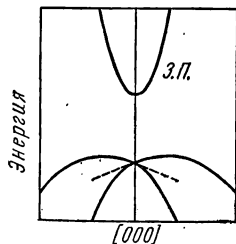


Рис. 41. Структура валентной зоны некоторых полупроводников $A^{III}B^V$

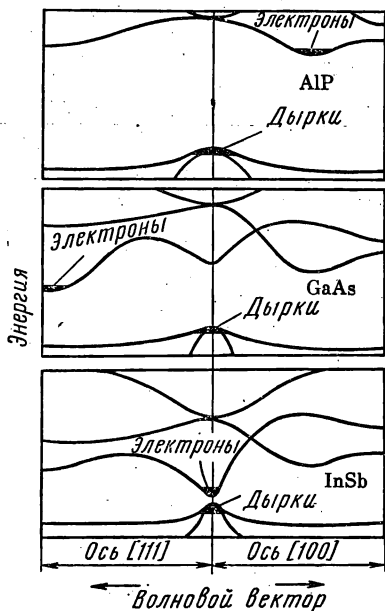


Рис. 42. Структура зон соединений $A^{III}B^V$

оказывается разной. Другими словами, можно сказать, что эффективная масса зависит от k , если определить некоторую «обобщенную» эффективную массу соотношением

$$E(k) = E(0) + \frac{\hbar^2 k^2}{2m^*(k)}. \quad (26.51)$$

Величина эффективной массы электронов и дырок в германии, кремнии и соединениях $A^{III}B^V$ приведена в табл. 7.

Многие тонкости структуры зон и характеристик носителей заряда еще требуют своего детального исследования.

В заключение сделаем несколько общих замечаний о ширине запрещенной зоны.

Эффективные массы электронов m_n^*/m и дырок m_p^*/m

Вещество	m_n^*/m	m_p^*/m	Вещество	m_n^*/m	m_p^*/m
Германий	$m_l = 1,3$ $m_t = 0,082$ при $\kappa = 0: 0,036$	0,34	Арсенид галлия	0,07	0,5
		0,04	Антимонид галлия	0,05	0,5
		0,07	Антимонид индия	0,013	0,6
Кремний	$m_l = 0,97$ $m_t = 0,19$	0,5	Арсенид индия	0,03	0,4
		0,16	Фосфид индия	0,07	0,4
		0,25			

Запрещенная зона в простых веществах IV группы уменьшается с ростом заряда ядра. В соединениях $A^{III}B^V$ ширина запрещенной зоны также уменьшается с ростом среднего заряда ядра, что изображено на рис. 43, а. Из рисунка видно, что ширина запрещенной зоны соединений $A^{III}B^V$ со средним зарядом ядра, равным заряду ядра элементов IV группы, примерно в два раза больше ширины запрещенной зоны соответствующего простого вещества. Это видно на примере ${}_{13}Al{}_{15}P$, средний заряд которого равен 14 и ему соответствует заряд кремния ${}_{14}Si$.

Этот факт легко объясняется как с точки зрения теории квазисвязанного, так и с точки зрения теории квазисвободного электрона. В теории квазисвязанного электрона это является следствием увеличения обменного интеграла, поскольку размеры электронного облака (электронной оболочки) больше для больших Z . В теории квазисвободного электрона этот факт можно объяснить более резким изменением потенциальной энергии при уменьшении заряда ядра и размеров электронного облака.

Большую величину запрещенной зоны соединений $A^{III}B^V$ по сравнению с шириной запрещенной зоны $A^{IV}B^{IV}$ с одинаковым средним зарядом можно объяснить тем, что потенциальная энергия в соединениях $A^{III}B^V$ изменяется более резко в области между атомами в связи с некоторой долей ионной составляющей химической связи атомов. Поскольку заряд валентных электронов в «соединениях» $A^{IV}B^{IV}$ распределен более равномерно в межатомном промежутке, чем в соединениях $A^{III}B^V$, то потенциальная энергия будет более плавно меняться с координатой, что приводит к меньшей ширине запрещенной зоны в них.

Рассмотрим связь между значениями ширины запрещенной зоны и среднего заряда для соединений $A^{III}B^V$. Из рис. 43, б видно, что значения ΔE_0 хорошо ложатся на прямую линию в пределах каждой подгруппы элементов второй группы.

Графики представляются в виде параллельных прямых, сдвинутых по оси ΔE_0 примерно на 2,2 эВ. Эта величина связана с энергией ионизации свободных атомов элементов первой (A_I^{II}) и второй

(A_{II}^{II}) подгрупп второй группы A^{II} . Если средний потенциал ионизации для A_{II}^{II} составляет величину порядка 9,6 эВ, которая сравнима со средним потенциалом ионизации элементов B^{VI} (вторая подгруппа), равным 10,5, то средний потенциал ионизации A_I^{II} составляет всего лишь 5,9 эВ.

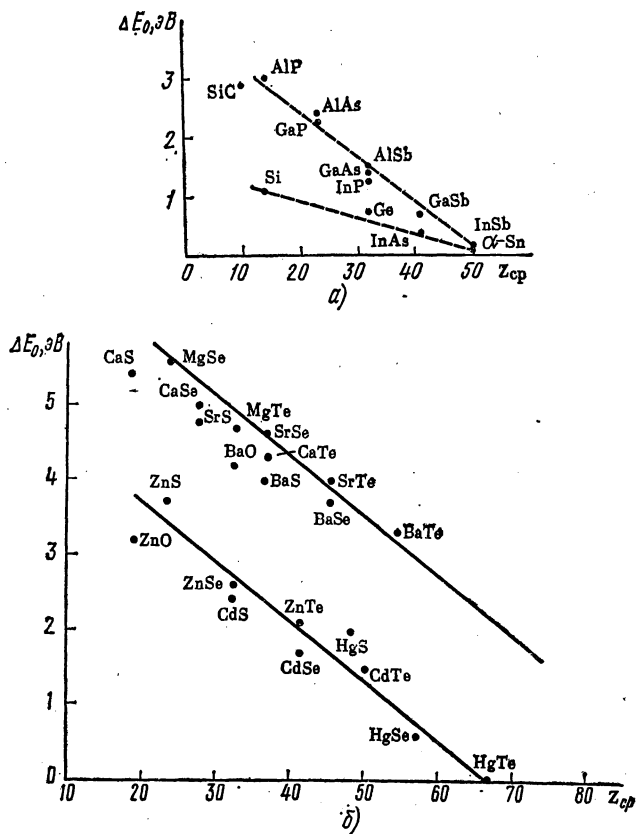


Рис. 43. Связь ширины запрещенной зоны со средним зарядом ядер соединений $A^{II}B^{VI}$ (а); $A^{II}B^{VI}$ (б)

Если предположить, что валентная зона $A^{II}B^{VI}$ возникает из уровня основного состояния B^{VI} , а зона проводимости возникает из основного состояния A^{II} , то становится совершенно понятным, что в $A_I^{II}B^{VI}$ ширина запрещенной зоны превосходит ΔE_0 в $A_{II}^{II}B^{IV}$, поскольку в A_{II}^{II} уровень основного состояния лежит выше уровня основного состояния B^{VI} в среднем на 1 эВ, в то время как для A_I^{II} уровень основного состояния лежит выше на величину 4,6 эВ. Вообще говоря, можно ожидать, что в $A_I^{II}B^{VI}$ зона проводимости

может образоваться из уровня возбужденного состояния элемента B^{VI} , в то время как в $A_{II}^{IV}B^{VI}$ она возникает, по-видимому, из основного состояния атомов A_{II}^{II} .

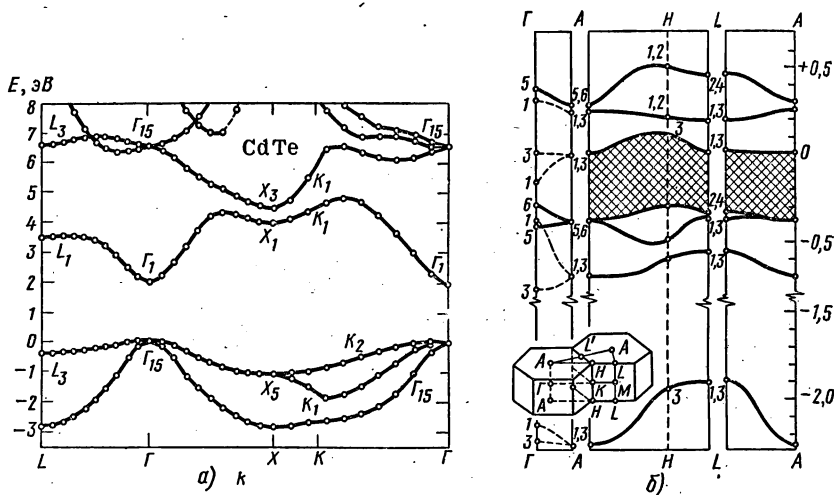


Рис. 44. Структура зон энергии некоторых соединений $A_{II}^{IV}B^{VI}$: теллурида кадмия (а) и селенида цинка (б), рассчитанная методом псевдопотенциала и ОПВ (сплошная линия—псевдопотенциал, кружком—ОПВ)

На рис. 44 представлена структура зон энергии теллурида кадмия и селенида цинка, рассчитанная различными методами. Экстремумы зон находятся в точке Γ , в которой валентная зона дважды вырождена, спин-орбитальное расщепление определяет положение третьей валентной зоны. В решетке вюрцита гексагональное кристаллическое поле снимает это вырождение.

§ 27. ПОНЯТИЕ О КВАЗИЧАСТИЦАХ

В предыдущем параграфе рассмотрена зонная структура, как следствие периодической структуры поля решетки, т. е. упорядочения в расположениях атомов на больших расстояниях — так называемый дальний порядок. Однако зонный характер энергии наблюдается у полупроводников и в том случае, когда дальний порядок в них нарушается, например, когда они находятся в расплавленном состоянии. Это говорит о том, что дальний порядок, т. е. периодический характер потенциальной энергии твердого тела, не является необходимым условием образования зон, это только одно из достаточных условий, которое оказалось существенным лишь в результате одноэлектронного приближения. По-видимому, достаточным условием

является уже структура поля на близких расстояниях (ближний порядок), но для этого необходимо рассматривать задачу многих частиц.

Для того чтобы показать, что дальний порядок не является необходимым для образования зон энергии, рассмотрим еще раз следующую задачу. Допустим, что имеем систему из N атомов, находящихся на различных расстояниях друг от друга. Предположим, что все расстояния достаточно велики, так что межатомным взаимодействием можно пренебречь. В этом случае система из N атомов имеет очевидное решение — волновая функция системы представляет собой произведение атомных функций, а энергия системы есть сумма энергий электронов в атоме. Такое состояние системы атомов является вырожденным, поскольку возможен обмен состояниями электронов различных атомов при сохранении энергии. Обменное вырождение имеет кратность $N!$. Сближим атомы так, чтобы между ними возникло взаимодействие. Учитывая попарное взаимодействие атомов, можем найти $(N-1)$ различных функций $W(r-n)$, дающих дополнительную потенциальную энергию для n -атома. Совершенно очевидно, что основную роль в величине $W(r-n)$ будут играть только ближайшие соседи n -атома. Совокупность функций $W(r-n)$ будет представлять возмущение для вырожденной системы. Но как известно, при наложении возмущения на вырожденную систему вырождение может быть полностью или частично снято. Это значит, что теперь энергия электронов в различных атомах будет разной в зависимости от вида $W(r-n)$, которая определяется только ближним порядком. Если вырождение снимается полностью, то каждый уровень энергии распадается на N подуровней. При большом N подуровни энергии будут практически сливаться, образуя полосу квазинепрерывных значений энергии. Если некоторые $W(r-n)$ будут одинаковыми, то соответствующее им решение останется вырожденным, но с другим по сравнению с исходным значением энергии.

Кратность вырождения какого-либо подуровня будет равна числу одинаковых функций $W(r-n)$. Вводя некоторый параметр k , описывающий распределение $W(r-n)$ по нему, получим зависимость энергии от k . Величину k можно будет связать с движением электронов по системе атомов в результате обмена атомов электронами. Такое качественное рассмотрение позволяет понять, что зонную структуру энергии может иметь для любой системы взаимодействующих атомов. Именно это соображение позволяет вводить понятие о примесных зонах, поскольку распределение атомов примеси по объему кристалла является нерегулярным.

Описание свойств твердого тела как системы частиц значительно упрощается введением квазичастиц, являющихся носителями определенных свойств, отражающих свойства тела. По существу такими квазичастицами являются электроны и дырки проводимости, свойства которых рассматривались выше. Помимо квазичастиц, являющихся носителями заряда (электроны и дырки), в твердом теле могут быть квазичастицы других типов.

К незаряженным, или бестоковым, квазичастицам относятся экситоны и поляроны.

Переход электрона из валентной зоны в зону проводимости равен носил «ионизации» атома основного вещества. Для этого атому необходимо сообщить энергию не меньшую ширины запрещенной зоны. Если атом может поглотить энергию меньше ширины запрещенной зоны, то электрон остается связанным с данным атомом, а атом должен быть в возбужденном состоянии. Этому состоянию атома (точнее всего кристалла) должен соответствовать уровень энергии, лежащий в запрещенной зоне. Такое возбужденное состояние атома Френкель назвал экситоном (excite — возбуждать, англ.). Энергия возбуждения может передаваться от атома к атому, поэтому экситон движется по кристаллу. Экситон — это движение в кристалле состояния возбуждения атомов, но не движение возбужденного атома. Если возможно несколько различных возбужденных состояний, то в запрещенной зоне должно находиться несколько различных уровней энергии. Экситонные уровни подобны системе уровней примеси.

Можно подойти к определению экситона с другой точки зрения. Пусть электрон из валентной зоны перешел в зону проводимости и имеет энергию E_n , а образовавшаяся дырка характеризуется энергией E_p . Между дыркой и электроном должно существовать притяжение, характеризующееся энергией:

$$V(\rho) = -\frac{e^2}{\epsilon |\mathbf{r}_n - \mathbf{r}_p|}; \quad \rho = |\mathbf{r}_n - \mathbf{r}_p| \quad (\text{в системе Гаусса}), \quad (27.1)$$

где \mathbf{r}_p — координата дырки, а \mathbf{r}_n — координата электрона. Движение системы из двух тел можно свести к движению одной частицы в силовом поле:

$$V(\rho) = -\frac{e^2}{\epsilon \rho}; \quad m_{np}^* = \frac{m_n^* m_p^*}{m_n^* + m_p^*}, \quad (27.2)$$

имеющей так называемую приведенную массу m_{np}^* , движение которой описывается в системе координат центра инерции уравнением

$$\left(-\frac{\hbar^2}{2m_{np}^*} \Delta - \frac{e^2}{\epsilon \rho} \right) \psi(\xi, \eta, \zeta) = \tilde{E} \psi(\xi, \eta, \zeta) = E^* \psi(\xi, \eta, \zeta), \quad (27.3)$$

где

$$\Delta = \frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial \zeta^2}; \quad \rho = (\xi^2 + \eta^2 + \zeta^2)^{1/2}. \quad (27.4)$$

Уравнение (27.3) есть не что иное, как уравнение для водородо-подобной атомной системы. Для энергии \tilde{E} можем записать выражение по аналогии с энергией для атома водорода. Для свободного атома водорода энергия имеет вид

$$E_n^H = -\frac{me^4}{2\hbar^2 n^2} = -\frac{E_1^H}{n^2}, \quad (27.5)$$

где E_I — энергия ионизации атома водорода, соответствующая энергии основного состояния ($n = 1$). При этом за нуль отсчета энергии принимается энергия покоящегося электрона, удаленного на бесконечность. Если за начало отсчета принять другую величину, например, считать за начало отсчета энергию нормального состояния, т. е. $E_I^H = 0$, то

$$E_n = E_I^H - \frac{E_I^H}{n^2} = E_I^H \left(1 - \frac{1}{n^2} \right). \quad (27.6)$$

Энергия E_n° свободной пары электрон—дырка может быть записана в аналогичном (27.5) виде

$$E_n^{\circ} = -\frac{E_1^{\circ}}{n^2}; \quad E_1^{\circ} = \frac{m_{np}^* e^4}{2\hbar^2 \epsilon^2}. \quad (27.7)$$

Чтобы указать положение уровней экситона относительно начала отсчета энергии для кристалла в целом, необходимо учесть, что при $n = \infty$ и $\rho = \infty$ $E_{\infty}^{\circ} = E_v + E_c$, а не нулю, поэтому

$$E_n^{\circ} = E_v + E_c - \frac{E_1^{\circ}}{n^2}. \quad (27.8)$$

Если отсчитывать энергию от потолка валентной зоны:

$E_v = 0$; $E_c = \Delta E_0$, то $E_n^{\circ} = \Delta E - \frac{E_1^{\circ}}{n^2}$. Основное состояние находится в запрещенной зоне и имеет величину энергии $E_1^{\circ} = \Delta E_0 - E_1^{\circ}$, которую можно оценить следующим образом:

$$E_1, \text{ эВ} = \frac{m_{np} e^4}{2\hbar^2 \epsilon^2} = \frac{13,5}{\epsilon^2} \left(\frac{m_{np}^*}{m} \right). \quad (27.9)$$

Если $m_n^* = m_p^* = m$, то $m_{np}^* = \frac{m}{2}$, и видим, что уровни экситона лежат близко к дну зоны проводимости. Радиус боровских орбит для экситона

$$a_n^{\circ}, \text{ \AA} = \frac{n^2 \hbar^2}{e^2 m_{np}^*} \epsilon = 0,528 \left(\frac{m}{m_{np}^*} \right) \epsilon. \quad (27.10)$$

Чем меньше эффективная масса носителей заряда и больше диэлектрическая проницаемость, тем больше радиус экситона и меньше энергия ионизации, поэтому в веществах с большим ϵ и малыми m_n^ , m_p^* наблюдать экситоны трудно. Экситоны большого радиуса называют экситонами Мотта.*

Мы получили для связанной пары электрон — дырка систему дискретных водородоподобных уровней энергии. Однако не было учтено

движение экситона в целом, т. е. движение его центра инерции. Это движение характеризуется кинетической энергией

$$T^э = \frac{\hbar^2 |K^э|^2}{2(m_n^* + m_p^*)}, \quad (27.11)$$

которую необходимо прибавить к $E_n^э$. Тем самым каждый уровень энергии $E_n^э$ должен превращаться в достаточно широкую полосу — *экситонную зону*. Однако экспериментально наблюдаются сравнительно узкие линии поглощения света, приписываемые экситонам. Причина этого связана с правилами отбора для волнового вектора.

В заключение укажем на существование еще одного вида возбужденного состояния, которое имеет место в веществах с большой долей ионной связи — *поляронов*. Полярон можно представить следующим образом. Электрон проводимости своим полем поляризует окружающее его пространство. В результате появления индуцированного положительного заряда возникает взаимодействие электрона с индуцированным зарядом, что приводит к возникновению дополнительной потенциальной энергии в виде потенциальной ямы, которая локализует электрон. При движении электрона область поляризации перемещается вместе с ним. Система, состоящая из электрона и поляризованной им области, носит название *полярона*.

Резюме § 26—27

1. Расчет зонной структуры конкретных полупроводников представляет собой сложную задачу. Найти зависимость $E(k)$ удастся только для некоторых областей зоны Бриллюэна. Имеются расчеты для германия и кремния. На их основе ведется расчет зонной структуры соединений $A^{III}B^V$ и $A^{II}B^{VI}$. Основные сведения о зонной структуре получают из анализа экспериментальных данных.

2. Экстремумы энергии электронов и дырок для германия и кремния находятся в различных точках зоны Бриллюэна. В соединениях $A^{III}B^V$ они могут быть как в различных точках, так и в одной и той же точке.

3. Валентная зона германия и кремния характеризуется тремя ветвями $E(k)$, которым соответствуют три типа дырок. Поверхности энергии имеют вид деформированных сфер (гофрированные поверхности).

4. Зона проводимости германия имеет восемь эквивалентных минимумов, лежащих на границе зоны Бриллюэна вдоль направлений типа $[111]$. Изоэнергетические поверхности являются эллипсоидами вращения, ось вращения совпадает с направлением типа $[111]$. На первую зону Бриллюэна приходится четыре полных эллипсоида энергии.

Зона проводимости кремния имеет шесть эквивалентных минимумов, лежащих в направлении типа $[100]$ недалеко от границы

зоны Бриллюэна. Изоэнергетические поверхности — эллипсоиды вращения, оси вращения направлены вдоль оси [100].

5. Все выводы зонной теории, касающиеся зависимости ширины запрещенной зоны от состава вещества, связи эффективной массы электрона с зоной и другие выводы, качественно находятся в хорошем согласии с экспериментом.

6. Зонная структура спектра энергии вытекает как из дальнего, так и из ближнего порядка.

7. Все результаты зонной теории относятся по существу к квазичастицам, которые являются носителями определенных свойств, отражающих свойства твердого тела.

Глава III

СТАТИСТИКА ЭЛЕКТРОНОВ И ДЫРОК В ПОЛУПРОВОДНИКАХ

§ 28. ПЛОТНОСТЬ СОСТОЯНИЙ

Проводимость вещества определяется концентрацией и подвижностью носителей заряда, поэтому, чтобы понять влияние внешних условий на проводимость, нужно выяснить зависимость концентрации и подвижности от внешних условий, и в первую очередь от температуры. Нас будут интересовать свободные носители заряда, т. е. электроны в зоне проводимости и дырки в валентной зоне. Чтобы подсчитать их число, необходимо знать число состояний и вероятность нахождения носителей заряда в этих состояниях.

Пусть вероятность нахождения электронов в единичном объеме фазового пространства, построенного вокруг точки этого пространства с координатой \mathbf{r} , \mathbf{k} в момент времени t , равна $f(\mathbf{r}, \mathbf{k}, t)$; \mathbf{r} и \mathbf{k} (или \mathbf{r} и \mathbf{P}) рассматриваем в качестве канонически сопряженных величин, удовлетворяющих обычным перестановочным соотношениям Гейзенберга:

$$r_l p_m - p_m r_l = i\hbar \delta_{lm}; \quad (l, m = x, y, z). \quad (28.1)$$

В элементе объема $d\Gamma$ фазового пространства имеется $d\Gamma/\hbar^3$ фазовых ячеек, в которых содержится $2d\Gamma/\hbar^3$ состояний (множитель 2 означает, что в каждой ячейке может быть 2 электрона с противоположно направленными спинами). Число электронов dn , находящихся в объеме $d\Gamma$, равно произведению числа состояний на вероятность нахождения электронов в этих состояниях:

$$dn = dn(\mathbf{r}, \mathbf{k}, t) = f(\mathbf{r}, \mathbf{k}, t) 2 \frac{d\Gamma}{\hbar^3}. \quad (28.2)$$

Если выразить $d\Gamma$ через объем обычного пространства $d\tau_r$ и пространства квазиимпульсов $d\tau_p = \hbar^3 d\tau_k$, то

$$d\Gamma = d\tau_r d\tau_p = dx dy dz dk_x dk_y dk_z \hbar^3. \quad (28.3)$$

Интегрируя по объему кристалла V и первой зоны Бриллюэна V_k , получим полное число электронов N , находящихся в некоторой энергетической зоне кристалла:

$$N = \frac{2}{\hbar^3} \int \int_{(V V_k)} f(\mathbf{r}, \mathbf{k}, t) d\Gamma. \quad (28.4)$$

Средняя концентрация электронов

$$\bar{n} = \frac{N}{V} = \frac{2}{V h^3} \int_{(V)} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}, t) d\Gamma = \frac{1}{4\pi^3 V} \int_{(V)} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}, t) d\tau_r d\tau_k. \quad (28.5)$$

Число электронов dn_r в элементе объема $d\tau_r$ равно

$$dn_r = n(\mathbf{r}, t) d\tau_r = d\tau_r \frac{2}{h^3} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}, t) d\tau_k. \quad (28.6)$$

Другими словами, концентрация электронов

$$n(\mathbf{r}, t) = \frac{dn_r}{d\tau_r} = \frac{1}{4\pi^3} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}, t) d\tau_k. \quad (28.7)$$

Через V_k мы обозначили объем первой зоны Бриллюэна независимо от того, в каком пространстве — \mathbf{k} или \mathbf{P} — она построена. Если кристалл однороден, то вероятность f от координаты не зависит, и интегрирование по $d\tau_r$ дает объем кристалла, в результате чего \bar{n} (28.5) равно истинной концентрации n (28.7):

$$\bar{n} = \frac{1}{4\pi^3 V} \int_{(V)} \int_{(V_k)} f(\mathbf{k}, t) d\tau_r d\tau_k = \frac{1}{4\pi^3} \int_{(V_k)} f(\mathbf{k}, t) d\tau_k = n(t). \quad (28.8)$$

Как было указано в первой главе, функция распределения есть функция от энергии и температуры, для стационарных состояний она не зависит и от времени. Так как энергия есть собственное значение оператора Гамильтона квантовой системы, то она не зависит от координаты, поэтому не будет зависеть от координаты и функция распределения $f_0 = f_0(E, T)$, где $f_0(E, T)$ — функция Ферми — Дирака (или Максвелла — Больцмана).

В интервале энергии dE может быть различное число состояний dS (без учета спина), отнесенное к единичному объему кристалла, в зависимости от величины энергии. Пусть между dS и dE существует равенство

$$dS = dS(E) = N(E) dE. \quad (28.9)$$

$N(E)$ равно числу состояний в единичном интервале энергии для единичного объема кристалла, $N(E)$ называют плотностью состояний:

$$N(E) = dS/dE. \quad (28.10)$$

Если вероятность заполнения состояний с энергией E равна $f_0(E, T)$, то число электронов dn в этих состояниях равно

$$dn = dn(E, T) = 2f_0(E, T) dS = 2f_0(E, T) N(E) dE. \quad (28.11)$$

Концентрация электронов будет определяться величиной

$$n = n(T) = 2 \int_{(E)} f_0(E, T) N(E) dE. \quad (28.12)$$

Интегрирование необходимо провести от дна зоны проводимости E_c до ее потолка. Учитывая резкую зависимость $f_0(E, T)$ от энергии, верхний предел можно положить равным бесконечности, так что

$$n = 2 \int_{E_c}^{\infty} f_0(E, T) N(E) dE. \quad (28.13)$$

Величина $N(E)$ тесным образом связана с формой изоэнергетических поверхностей. Действительно, построим в зоне Бриллюэна две изоэнергетические поверхности E и $E + dE$. Они выделяют некоторый слой пространства квазиимпульса (рис. 45). Пусть объем этого слоя $d\tau_p$, ему соответствует объем фазового пространства

$$d\Gamma = V d\tau_p; \quad (28.14)$$

число фазовых ячеек в нем $d\Gamma/h^3$, число состояний для единичного объема $dS = d\Gamma/Vh^3$. С другой стороны, $dS = N(E) dE$, поэтому

$$dS = N(E) dE = \frac{1}{V} \frac{d\Gamma}{h^3} = \frac{d\tau_p}{h^3} = \frac{d\tau_k}{4\pi^3}. \quad (28.15)$$

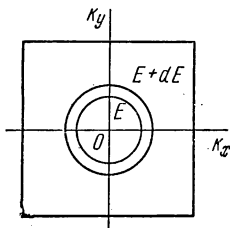


Рис. 45. Объем слоя в зоне Бриллюэна, заключенного между двумя изоэнергетическими поверхностями E и $E + dE$

Входящую в выражение (28.15) величину $d\tau_k(E, E + dE)$ можно найти, если известно уравнение изоэнергетических поверхностей. Заметим, что величины

$$1/h^3, 1/Vh^3 \text{ и } 1/V \cdot 8\pi^3 \quad (28.16)$$

можно рассматривать как плотности состояний соответственно в фазовом пространстве, в пространстве квазиимпульса и волнового вектора.

Рассмотрим некоторые частные случаи.

1. Сферические изоэнергетические поверхности с E_{\min} в центре зоны Бриллюэна. Пусть

$$E_k = E_c + \frac{\hbar^2 \kappa^2}{2m^*}. \quad (28.17)$$

Две изоэнергетические поверхности E и $E + dE$ выделяют сферический слой толщиной $d\kappa$ и объемом $d\tau_k(E, E + dE)$ (см. рис. 45):

$$d\tau_k = 4\pi \kappa^2 d\kappa. \quad (28.18)$$

Выразим κ через E :

$$\kappa^2 = \frac{2m^*}{\hbar^2} (E - E_c); \quad \kappa = \left(\frac{2m^*}{\hbar^2} \right)^{1/2} (E - E_c)^{1/2}. \quad (28.19)$$

Дифференцируя по κ первое выражение в (28.19), получим

$$2\kappa d\kappa = \frac{2m^*}{\hbar^2} dE. \quad (28.20)$$

Учитывая соотношение (28.18), (28.19) и (28.20), можем записать

$$d\tau_k(E, E + dE) = 4\pi\kappa^2 d\kappa = 2\pi\kappa^2 d\kappa = \\ = 2\pi \left(\frac{2m^*}{\hbar^2}\right)^{1/2} (E - E_c)^{1/2} \frac{2m^*}{\hbar^2} dE = 2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (E - E_c)^{1/2} dE. \quad (28.21)$$

Для dS из (28.15) и (28.21) получим

$$dS = \frac{d\tau_k}{8\pi^3} = N(E) dE = \frac{2\pi (2m^*)^{3/2}}{8\pi^3 \hbar^3} (E - E_c)^{1/2} dE, \quad (28.22)$$

или

$$N(E) = 2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (E - E_c)^{1/2}. \quad (28.23)$$

2. Сферические изоэнергетические поверхности с минимумом энергии в точках \mathbf{k}_0 . Пусть минимумы энергии лежат не в центре зоны Бриллюэна, а в некоторых точках \mathbf{k}_0 , и число их равно M . Построив изоэнергетические поверхности $E, E + dE$, получим M сфер (рис. 46). Уравнение одной из них имеет вид

$$E = E_c + \frac{\hbar^2 (\mathbf{k} - \mathbf{k}_0)^2}{2m^*}. \quad (28.24)$$

Радиусы сфер $|\mathbf{k} - \mathbf{k}_0|$, толщина сферического слоя $d\kappa = d|\mathbf{k} - \mathbf{k}_0|$; но число сфер равно теперь M , поэтому интервалу энергии dE соответствует объем M слоев:

$$d\tau_k(E, E + dE) = \\ = M \cdot 4\pi (\mathbf{k} - \mathbf{k}_0)^2 d|\mathbf{k} - \mathbf{k}_0|. \quad (28.25)$$

Выражая $|\mathbf{k} - \mathbf{k}_0|$ и $d|\mathbf{k} - \mathbf{k}_0|$ через энергию, получим

$$N(E) = M \cdot 2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (E - E_c)^{1/2}. \quad (28.26)$$

3. Эллипсоидальные изоэнергетические поверхности. Рассмотрим более общий случай, когда изоэнергетические поверхности имеют вид эллипсоида:

$$E(\mathbf{k}) = E_c + \frac{\hbar^2}{2} \left(\frac{\kappa_x^2}{m_1} + \frac{\kappa_y^2}{m_2} + \frac{\kappa_z^2}{m_3} \right), \quad (28.27)$$

или, в канонической форме,

$$\frac{\kappa_x^2}{a^2} + \frac{\kappa_y^2}{b^2} + \frac{\kappa_z^2}{c^2} = 1, \quad (28.28)$$

где полуоси эллипсоида

$$a_i = \left[\frac{2m_i (E - E_c)}{\hbar^2} \right]^{1/2}. \quad (28.29)$$

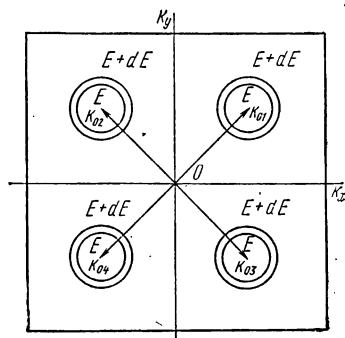


Рис. 46. Объем слоя в зоне Бриллюэна, соответствующий интервалу энергии dE при M экстремумах

Объем одного эллипсоида с полуосями a, b, c равен

$$\tau_{\mathbf{k}}^{(1)} = \frac{4\pi}{3} abc = \frac{4\pi}{3\hbar^3} (8m_1m_2m_3)^{1/2} (E - E_c)^{3/2}, \quad (28.30)$$

а объем одного слоя между двумя эллипсоидальными изоэнергетическими поверхностями

$$d\tau_{\mathbf{k}}^{(1)} = \frac{2\pi}{\hbar^3} (8m_1m_2m_3)^{1/2} (E - E_c)^{1/2} dE. \quad (28.31)$$

Подставляя $d\tau_{\mathbf{k}}^{(1)}$ в выражение для dS и учитывая, что минимум энергии может быть M , получим

$$dS = \frac{M d\tau_{\mathbf{k}}^{(1)}}{8\pi^3} = \frac{M \cdot 2\pi}{\hbar^3} (8m_1m_2m_3)^{1/2} (E - E_c)^{1/2} dE, \quad (28.32)$$

или

$$N(E) = \frac{M \cdot 2\pi}{\hbar^3} (8m_1m_2m_3)^{1/2} (E - E_c)^{1/2}. \quad (28.33)$$

Напомним, что, например, для германия $M=4$, а для кремния $M=6$.

Таким образом, *плотность состояний пропорциональна $(E - E_c)^{1/2}$ и $(m_1m_2m_3)^{1/2}$, где m_1, m_2, m_3 — компоненты тензора эффективной массы (в главных осях). При малых m_i $N(E)$ будет малой величиной.* Очевидно, $N(E)$ должно быть малой для интерметаллов.

Если положить

$$M^2 (m_1m_2m_3) = m_d^{*3}, \quad (28.34)$$

где m_d^* — *носит название эффективной массы для плотности состояний*, то в общем случае

$$N(E) = 2\pi (2m_d^*/\hbar^2)^{3/2} (E - E_c)^{1/2}, \quad (28.35)$$

что совершенно аналогично (28.23) для случая изотропной массы с одним минимумом.

Выражение (28.35) является наиболее общим, поскольку из него вытекают (28.23), (28.26) как частные случаи. Необходимо только помнить, что *зависимость $N(E)$ от энергии вида $N(E) \sim (E - E_c)^{1/2}$ справедлива до тех пор, пока энергия является квадратичной функцией квазиимпульса, другими словами, выражение (28.35) справедливо только для состояний вблизи минимума энергии, т. е. у дна зоны.*

Найдем выражение для плотности состояний у потолка зоны энергии, где энергия также является квадратичной функцией энергии. Если максимумы энергии находятся в точках \mathbf{k}_0 зоны Бриллюэна, а их число равно M , то $E(\mathbf{k})$ (для каждого максимума) можно представить в общем виде:

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \frac{\hbar^2}{2} \left[\frac{(\kappa_x - \kappa_{0x})^2}{m_1} + \frac{(\kappa_y - \kappa_{0y})^2}{m_2} + \frac{(\kappa_z - \kappa_{0z})^2}{m_3} \right]. \quad (28.36)$$

Учитывая, что тензор эффективной массы электронов в максимуме энергии отрицателен, и вводя вместо него тензор эффективной массы дырок условием $m_{ip} = -m_{in}$, можем записать

$$E(\mathbf{k}) = E(\mathbf{k}_0) - \frac{\hbar^2}{2} \left[\frac{(\kappa_x - \kappa_{0x})^2}{m_{1p}} + \frac{(\kappa_y - \kappa_{0y})^2}{m_{2p}} + \frac{(\kappa_z - \kappa_{0z})^2}{m_{3p}} \right], \quad (28.37)$$

или

$$\frac{\hbar^2}{2} \left[\frac{(\kappa_x - \kappa_{0x})^2}{m_{1p}} + \frac{(\kappa_y - \kappa_{0y})^2}{m_{2p}} + \frac{(\kappa_z - \kappa_{0z})^2}{m_{3p}} \right] = E_v - E(\mathbf{k}), \quad (28.38)$$

где $E_v = E(\mathbf{k}_0)$ — энергия в данном случае потолка валентной зоны. Опуская совершенно очевидные выкладки, найдем, что выражение для плотности состояний у потолка зоны имеет вид

$$N(E) = 2\pi \left(\frac{2m_{pd}^*}{\hbar^2} \right)^{3/2} (E_v - E)^{1/2}, \quad (28.39)$$

где через m_{pd}^* обозначена эффективная масса дырок для плотности состояний:

$$m_{pd}^* = (M^2 m_{p1} m_{p2} m_{p3})^{1/3}. \quad (28.40)$$

Плотность состояний у дна и потолка зон имеет совершенно аналогичный вид, поскольку найденные выражения для $N(E)$ являются следствием квадратичной зависимости энергии от квазимпульса в окрестности экстремумов, различный знак подкоренного выражения $-E - E_c$ и $E_v - E$ — является следствием различия в знаках эффективной массы в минимуме и максимуме энергии. Одно из выражений для $N(E)$ справедливо до тех пор, пока применима квадратичная зависимость энергии от квазимпульса. Вдали от границ зон выражение для $N(E)$ не может быть записано, пока неизвестна зависимость $E(\mathbf{k})$. Полное число состояний в зоне с учетом спина равно $2N$ или $2Ng$ (здесь g — кратность вырождения состояния), поэтому

$$2 \int_{E_{\min}}^{E_{\max}} N(E) dE = 2Ng. \quad (28.41)$$

На рис. 47 представлен ход функции $N(E)$ в общем виде. Для запрещенных зон $N(E) = 0$, если не считать локализованных состояний электрона.

Плотность состояния для локализованных состояний находится из следующих простых соображений. У каждого примесного атома может быть по одному электрону, поэтому полное число состояний будет равно N_d и N_a соответственно. Если предположить, что примесная зона не возникает и примесные уровни остаются дискретными, то для каждого уровня плотность должна быть равной бесконечности, с тем, однако, чтобы интеграл от плотности по всем воз-

можным состояниям (для одного уровня) был равен единице. Для этого необходимо выразить плотность состояний в виде δ -функции:

$$N_d(E) = N_d \delta(E - E_d), \quad (28.42)$$

$$N_a(E) = N_a \delta(E - E_a), \quad (28.43)$$

где через $N_d(E)$, $N_a(E)$ обозначена плотность состояний на донорном и акцепторном уровнях, характеризуемых энергией E_d и E_a соответственно. На рис. 47 плотность состояний $N_d(E)$ и $N_a(E)$ изображена вертикальной линией. Если при возрастании концентрации примеси образуется примесная зона, то в этом случае выражения (28.42) и (28.43) должны быть заменены выражением (28.35) или (28.39).

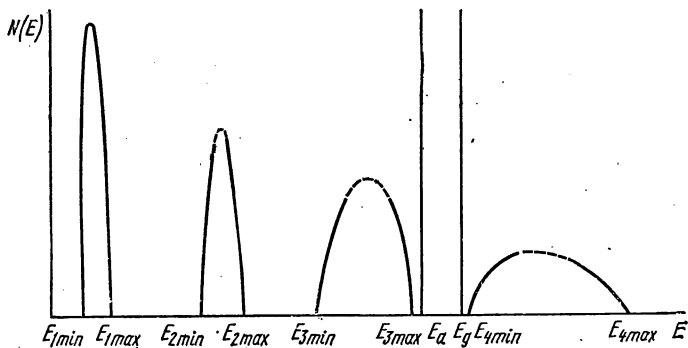


Рис. 47. Плотность состояний в различных зонах и на локальных уровнях энергии

Рассмотрим общий способ нахождения плотности состояний $N(E)$. Он состоит в том, что интегрирование по объему зоны Бриллюэна проводят в два приема: вначале по площади изоэнергетической поверхности, затем по энергии. Удобство этого метода проявляется в том случае, когда под интегралом стоят величины, зависящие от энергии, которые принимают постоянное значение на изоэнергетической поверхности. Элемент объема $d\tau_k$ можно выразить через элемент площади dS_E и нормальный компонент волнового вектора dk_n :

$$d\tau_k = dS_E dk_n. \quad (28.44)$$

Но dk_n выражается через

$$dE = \left(\frac{dE}{dk} dk \right) = |\nabla_k E| dk_n \quad (28.45)$$

и

$$dk_n = \frac{dE}{|\nabla_k E|}; \quad d\tau_k = \frac{dS_E dE}{|\nabla_k E|}, \quad (28.46)$$

после чего для n получим

$$n = \frac{1}{4\pi^3} \int \frac{dS_E dE}{|\nabla_k E|}. \quad (28.47)$$

Поскольку функция распределения зависит от энергии,

$$n = 2 \int_{(E)} f(E, T) dE \int_{(S_E)} \frac{dS_E}{8\pi^3 |\nabla_{\mathbf{k}} E|}. \quad (28.48)$$

Сравнивая (28.48) с (28.12), имеем для плотности состояний

$$N(E) = \frac{1}{8\pi^3} \int_{(S_E)} \frac{dS_E}{|\nabla_{\mathbf{k}} E|}. \quad (28.49)$$

Для сферических изоэнергетических поверхностей получаем из (28.49)

$$N(E) = \frac{1}{8\pi^3 |\nabla_{\mathbf{k}} E|} \cdot 4\pi k^2 = \frac{k^2}{2\pi^2 |\nabla_{\mathbf{k}} E|}. \quad (28.50)$$

Для квадратичного закона дисперсия (28.50) приводит к (28.23) и всем последующим соотношениям. Однако для неквадратичного закона дисперсии соотношение (28.49) является более подходящим. Найдем на его основе выражение для плотности состояний с использованием дисперсионного закона Кейна (26.37). Если для зоны проводимости за начало отсчета энергии принять дно зоны проводимости, то (26.37) примет вид после вычитания E_g :

$$E(\mathbf{k}) = -\frac{E_g}{2} + \frac{\hbar^2 k^2}{2m} + \sqrt{\frac{2P^2 k^2}{3} + \frac{E_g^2}{4}}. \quad (28.51)$$

Поскольку $P \sim 6 \cdot 10^{-8}$ эВ·см, при малых k можно отбросить второй член и записать

$$\left[E(\mathbf{k}) + \frac{E_g}{2} \right]^2 = \frac{2P^2 k^2}{3} + \frac{E_g^2}{4}, \quad (28.52)$$

или

$$\frac{2P^2 k^2}{3} = E(\mathbf{k}) [E_g + E(\mathbf{k})]. \quad (28.53)$$

Из (28.51) в общем случае для $|\nabla_{\mathbf{k}} E|$ найдем

$$|\nabla_{\mathbf{k}} E| = \frac{dE}{dk} = \frac{\hbar^2 k}{m} + \frac{(4P^2/3) k}{\sqrt{\frac{2}{3} P^2 k^2 + E_g^2/4}}, \quad (28.54)$$

или в пренебрежении членом $\hbar^2 k/m$:

$$|\nabla_{\mathbf{k}} E| = \frac{(4P^2 k/3)}{(2P^2 k^2/3 + E_g^2/4)^{1/2}}. \quad (28.55)$$

Подставив (28.55) в (28.50), получим

$$N(E) = \frac{3k \sqrt{\frac{2}{3} P^2 k^2 + \frac{E_g^2}{4}}}{2\pi^2 4P^2}. \quad (28.56)$$

Исключая k в соответствии с (28.53), для (28.56) получим

$$N(E) = \frac{E^{1/2}(E_g + E)^{1/2} \left(E + \frac{E_g}{2}\right)}{4\pi^2 \left(\frac{2P^2}{3}\right)^{3/2}}. \quad (28.57)$$

Введем эффективную массу на дне зоны проводимости m_n^{0*} соотношением

$$m_n^{0*} = \frac{3\hbar^2 E_g}{4P^2}. \quad (28.58)$$

Это позволяет записать (28.57) по аналогии с ранее использованными выражениями:

$$N(E) = \frac{1}{2} \left(\frac{2m_n^{0*}}{\hbar^2}\right)^{3/2} E^{1/2} \left(1 + \frac{E}{E_g}\right)^{1/2} \left(1 + 2\frac{E}{E_g}\right). \quad (28.59)$$

При $E_g = 0$ выражение для $N(E)$ проще получить из (28.57):

$$N(E) = \frac{E^2}{\left(\frac{4\pi P^2}{3}\right)^{3/2}}. \quad (28.60)$$

При $P \sim 6 \cdot 10^{-8}$ эВ·см имеем $N(E)$ [см⁻³] $\sim 5,4 \cdot 10^{20}$ [E(эВ)]².

§ 29. КОНЦЕНТРАЦИЯ ЭЛЕКТРОНОВ И ДЫРОК

Запишем теперь выражение (28.13) для концентрации электронов, в которое вместо $N(E)$ подставим выражение (28.35). Это можно сделать потому, что функция Ферми — Дирака очень быстро уменьшается с ростом энергии, поэтому вклад в концентрацию электронов от верхних энергетических состояний будет крайне мал, и использование (28.35) для $N(E)$ при всех значениях не приводит к заметной ошибке в n . В качестве иллюстрации этого на рис. 48 приведен график произведения функций $f_0(E, T)$ и $N(E)$ в виде

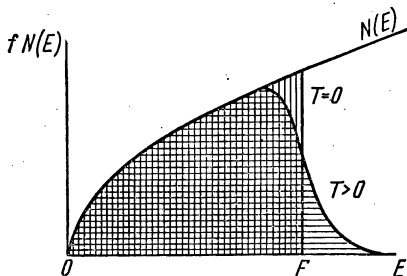


Рис. 48. График произведения $N(E)$ и $f(E, T)$ для металлов и вырожденных полупроводников

(28.35) для металла ($E_c = 0$; $F > 0$).

Итак, подставляя в (28.13) выражение (28.35) для $N(E)$ и функцию Ферми — Дирака для равновесных состояний

$$f_0(E, T) = \frac{1}{e^{\frac{E-F}{kT}} + 1}, \quad (29.1)$$

получим

$$n = 2 \int_{E_c}^{\infty} f_0(E, T) N(E) dE = 2 \int_{E_c}^{\infty} \frac{2\pi \left(\frac{2m_d^*}{\hbar^2}\right)^{3/2} (E - E_c)^{1/2} dE}{e^{\frac{E-F}{kT}} + 1}. \quad (29.2)$$

Если величина F известна, то выражение (29.2) позволяет вычислять концентрацию электронов.

Перепишем выражение (29.2) для n , перейдя к безразмерным переменным. Для этого положим

$$\frac{E - E_c}{kT} = x; \quad dx = \frac{dE}{kT}; \quad \frac{F - E_c}{kT} = \xi; \quad (29.3)$$

в таком случае

$$n = 4\pi \left(\frac{2m_d^* kT}{\hbar^2}\right)^{3/2} \int_0^{\infty} \frac{x^{1/2} dx}{e^{x-\xi} + 1}. \quad (29.4)$$

Введём следующие обозначения:

$$2 \left(\frac{2\pi m_d^* kT}{\hbar^2}\right)^{3/2} = N_c, \quad (29.5)$$

$$\int_0^{\infty} \frac{x^{1/2} dx}{e^{x-\xi} + 1} = \Phi_{1/2}(\xi). \quad (29.6)$$

Величина N_c носит название эффективного числа состояний в зоне проводимости, а $\Phi_{1/2}(\xi)$ называется интегралом Ферми порядка 1/2. В новых обозначениях концентрация электронов n имеет вид

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi). \quad (29.7)$$

Концентрация электронов является функцией температуры и уровня Ферми $n = n(T, F)$. Интеграл Ферми $\Phi_{1/2}(\xi)$ в общем виде в элементарных функциях не выражается, но для целого ряда важных в практическом отношении случаев существуют приближенные аналитические выражения, которые получим дальше.

Перейдем к нахождению выражения для концентрации свободных дырок в валентной зоне. Плотность состояний у потолка валентной зоны определяется соотношением (28.39). Функция распределения дырок по состояниям имеет вид

$$f_{op}(E, T) = 1 - f_0(E, T) = \frac{1}{e^{\frac{F-E}{kT}} + 1}. \quad (29.8)$$

В таком случае число дырок dp в интервале энергии dE равно числу состояний $2N(E) dE$, умноженных на вероятность того, что они заняты дырками — f_{op} :

$$dp = 2N(E) f_{op}(E, T). \quad (29.9)$$

Концентрация дырок в валентной зоне равна

$$p = 2 \int_{-\infty}^{E_v} f_{0p}(E, T) N(E) dE. \quad (29.10)$$

Нижний предел, равный $E_{v \min}$, заменили на $-\infty$, учитывая резкую зависимость f_0 от энергии. Подставляя выражение (29.8) для f_{0p} и (28.39) для $N(E)$, получим

$$p = 2 \int_{-\infty}^{E_v} 2\pi \left(\frac{2m_{pd}^*}{h^2} \right)^{3/2} \frac{(E_v - E)^{1/2} dE}{e^{\frac{F-E}{kT}} + 1}. \quad (29.11)$$

В выражениях (29.9–11) и дальнейших отсчет энергии для дырок совпадает с направлением отсчета энергии для электронов.

Введем обозначения:

$$\frac{E_v - E}{kT} = x; \quad dx = -\frac{dE}{kT}; \quad \frac{E_v - F}{kT} = \eta, \quad (29.12)$$

используя их, перепишем (29.11)

$$p = 4\pi \left(\frac{2m_{pd}^* kT}{h^2} \right)^{3/2} \int_0^{\infty} \frac{x^{1/2} dx}{e^{x-\eta} + 1} = \frac{2N_v}{\sqrt{\pi}} \Phi_{1/2}(\eta), \quad (29.13)$$

где

$$N_v = 2 \left(\frac{2\pi m_{pd}^* kT}{h^2} \right)^{3/2} \quad (29.14)$$

есть эффективное число состояний в валентной зоне, а

$$\Phi_{1/2}(\eta) = \int_0^{\infty} \frac{x^{1/2} dx}{e^{x-\eta} + 1} \quad (29.15)$$

есть интеграл Ферми, совпадающий по форме с $\Phi_{1/2}(\xi)$.

Согласно (29.7) и (29.13) концентрации электронов n и дырок p зависят от температуры T и уровня Ферми F .

Для интеграла Ферми $\Phi_{1/2}$ существуют различные приближенные выражения, справедливые в той или иной области изменения аргумента, а именно:

$$\Phi_{1/2}(\xi) = \begin{cases} \frac{\sqrt{\pi}}{2} e^{\xi} & \text{при } -\infty < \xi < -1, \\ \frac{\sqrt{\pi}}{2} \cdot \frac{1}{0,25 + e^{-\xi}} & \text{при } -1 < \xi < 5, \\ \frac{2}{3} \xi^{3/2} & \text{при } 5 < \xi < \infty. \end{cases} \quad (29.16)$$

Первое приближение, справедливое при $\xi < -1$, соответствует статистике Больцмана. Условием применимости классической статистики является неравенство $\xi < -1$ или $\frac{F - E_c}{kT} < -1$, откуда $F < E_c - kT$, т. е. полупроводник является невырожденным (подчиняется классической статистике), если уровень Ферми лежит ниже зоны проводимости не менее чем на kT . Если уровень Ферми лежит выше E_c более чем на $5kT$, то полупроводник полностью вырожденный. Если $E_c - kT < F < E_c + 5kT$, то свойства полупроводника являются переходными от невырожденного к полностью вырожденному. Как видим, в условие вырождения входит температура и положение уровня Ферми относительно дна зоны проводимости. Далее свяжем эти условия с другими характеристиками вещества и прежде всего с концентрацией примеси.

Покажем, что приближенные выражения для $\Phi_{1/2}(\xi)$, записанные выше, действительно имеют место.

Из выражения для $\Phi_{1/2}(\xi)$ непосредственно видно, что при $\xi < -1$ экспонента при любых $x > 0$ будет больше единицы; $e^{-\xi} > 1$. Но это значит, что при любых $\xi < -1$ функцию Ферми — Дирака можно заменить функцией Больцмана:

$$\frac{1}{e^{x-\xi} + 1} \approx e^{-x+\xi} \quad (\text{при } \xi < -1; x > 0). \quad (29.17)$$

В таком случае

$$\Phi_{1/2}(\xi) \approx \int_0^{\infty} x^{1/2} e^{-x+\xi} dx = e^{\xi} \int_0^{\infty} x^{1/2} e^{-x} dx. \quad (29.18)$$

Последний интеграл в (29.18) есть гамма-функция Эйлера $\Gamma(3/2)$ (он может быть сведен также к интегралу Пуассона):

$$\Gamma(3/2) = \int_0^{\infty} x^{1/2} e^{-x} dx = \frac{\sqrt{\pi}}{2}. \quad (29.19)$$

и

$$\Phi_{1/2}(\xi) \approx \frac{\sqrt{\pi}}{2} e^{\xi}. \quad (29.20)$$

в полном соответствии с (29.16).

Запишем выражение для концентрации электронов в невырожденном полупроводнике:

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi) \approx N_c e^{\xi} = N_c e^{-\frac{E_c - F}{kT}} = 2 \left(\frac{2\pi m_d^* kT}{h^2} \right)^{3/2} e^{-\frac{E_c - F}{kT}}; \quad (29.21)$$

$$F < E_c - kT.$$

Эффективное число состояний N_c зависит от температуры. Подставив численные значения универсальных констант, входящих в N_c , получим

$$N_c = 2 \left(\frac{2\pi m_d^* kT}{h^2} \right)^{3/2} = 4,82 \cdot 10^{15} \left(\frac{m_d^*}{m} \right)^{3/2} T^{3/2} = 2,5 \cdot 10^{19} \left(\frac{m_d^*}{m} \right)^{3/2} \left(\frac{T}{300} \right)^{3/2}, \quad (29.22)$$

$$n = 4,82 \cdot 10^{15} \left(\frac{m_d^*}{m} \right)^{3/2} T^{3/2} e^{-\frac{E_c - F}{kT}}. \quad (29.23)$$

Рассмотрим теперь случай полностью вырожденного полупроводника, для него:

$$\Phi_{1/2}(\xi) = \frac{2}{3} \xi^{3/2} = \frac{2}{3} \left(\frac{F - E_c}{kT} \right)^{3/2} \quad (F > F_c + 5kT). \quad (29.24)$$

Концентрация электронов при этом будет равна

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi) = \frac{8\pi}{3} \left(\frac{2m_d^*}{h^2} \right)^{3/2} (F - E_c)^{3/2}, \quad (29.25)$$

т. е. она не зависит от температуры. Напомним, что уровень Ферми находится в зоне проводимости выше ее дна не менее чем на $5kT$.

Найдем те условия, при которых интеграл Ферми может быть представлен в виде $\Phi_{1/2}(\xi) = \frac{2}{3} \xi^{3/2}$. Как известно, при малых температурах $\left(-\frac{\partial f_0}{\partial E} \right) \approx \delta(E - F)$. Это приближенное равенство позволяет легко вычислить интегралы от произведения двух функций $\varphi(E) f_0(E, T)$. Действительно,

$$\int_0^\infty \varphi(E) f_0(E, T) dE = \chi(E) f_0(E, T) \Big|_0^\infty - \int_0^\infty \chi(E) \frac{\partial f_0}{\partial E} dE = \\ = -\chi(0) f_0(0, T) + \chi(F) \approx \chi(F) - \chi(0), \quad (29.26)$$

где

$$\chi(E) = \int \varphi(E) dE, \quad \text{а} \quad f_0(0, T) \approx 1. \quad (29.27)$$

Пользуясь этим свойством функции распределения Ферми — Дирака, можем легко получить выражение для концентрации. В данном случае

$$\varphi(E) = (E - E_c)^{1/2} \quad \text{и} \quad \chi(E) = \frac{2}{3} (E - E_c)^{3/2},$$

поэтому

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi) = \frac{2N_c}{\sqrt{\pi} (kT)^{3/2}} \int_{E_c}^\infty \frac{(E - E_c)^{1/2} dE}{e^{\frac{E - F}{kT}} + 1} = \\ = \frac{2N_c}{\sqrt{\pi} (kT)^{3/2}} \cdot \frac{2}{3} (F - E_c)^{3/2}. \quad (29.28)$$

Из выражения (29.28) действительно видно, что

$$\Phi_{1/2}(\xi) = \frac{2}{3} \left(\frac{F - E_c}{kT} \right)^{3/2} = \frac{2}{3} \xi^{3/2}. \quad (29.29)$$

Условие независимости n от температуры равносильно условию, что $-\frac{\partial f_0}{\partial E} \approx \delta(E - F)$. Но для этого необходимо, чтобы $f_0(E, T)$ менялась в области $E \approx F$ как можно быстрее, при этом чем уже область $(E - F)$ изменения $f_0(E, T)$, тем ближе ее производная к δ -функции, δ -образный характер $-f'_0(E, T)$ лучше проявляется при низких температурах.

В переходной области от невырожденного к полностью вырожденному случаю концентрация электронов n будет зависеть от температуры:

$$n = N_c \frac{1}{0,25 + e^{\frac{E_c - F}{kT}}}. \quad (29.30)$$

Результаты, полученные для электронов, могут быть легко перенесены на дырки. С этой целью достаточно записать выражение для $\Phi_{1/2}(\eta)$ в виде (29.16):

$$\Phi_{1/2}(\eta) = \begin{cases} \frac{\sqrt{\pi}}{2} e^\eta & \text{при } -\infty < \eta < -1, \\ \frac{\sqrt{\pi}}{2} \frac{1}{0,25 + e^{-\eta}} & \text{при } -1 < \eta < 5, \\ \frac{2}{3} \eta^{3/2} & \text{при } 5 < \eta < \infty. \end{cases} \quad (29.31)$$

В невырожденном полупроводнике концентрация дырок определяется статистикой Больцмана, условием применимости которой является неравенство

$$\eta = \frac{E_v - F}{kT} < -1 \quad \text{или} \quad F > E_v + kT, \quad (29.32)$$

т. е. в невырожденном полупроводнике уровень Ферми должен лежать выше потолка валентной зоны по крайней мере на величину kT .

Для полностью вырожденного полупроводника

$$\eta = \frac{E_v - F}{kT} > 5 \quad \text{или} \quad F < E_v - 5kT, \quad (29.33)$$

т. е. уровень Ферми в полностью вырожденном полупроводнике должен находиться в валентной зоне ниже ее потолка на величину не менее $5kT$.

Если учесть замечания об изменении направления отсчета энергии для дырок на противоположное, то все результаты для дырок должны быть подобны результатам для электронов при соответствующей замене характеристик электронов характеристиками дырки.

Запишем выражения для концентрации дырок для двух предельных случаев: невырожденный полупроводник

$$p = N_v e^{-\frac{F - E_v}{kT}}; \quad N_v = 2 \left(\frac{2\pi m_{pd}^* kT}{h^2} \right)^{3/2} \quad (29.34)$$

и полностью вырожденный полупроводник

$$p = \frac{8\pi}{3} \left(\frac{2m_{pd}^*}{h^2} \right)^{3/2} (E_v - F)^{3/2}. \quad (29.35)$$

Резюме § 28, 29

1. Концентрация электронов n и дырок p определяется в виде интеграла по зоне Бриллюэна от функции распределения:

$$n = n(\mathbf{r}, t) = \frac{1}{4\pi^3} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}, t) d\tau_k, \quad (29.1p)$$

$$p = p(\mathbf{r}, t) = \frac{1}{4\pi^3} \int_{(V_k)} f_p(\mathbf{r}, \mathbf{k}, t) d\tau_k. \quad (29.2p)$$

2. Функция распределения Ферми—Дирака для электронов и дырок зависит от энергии и температуры, поэтому для вычисления концентрации электронов и дырок необходимо перейти от интегрирования по объему зоны Бриллюэна к интегрированию по энергии в пределах зоны энергии. Для этого вводится понятие плотности состояний $N(E)$, представляющее собой число состояний в единичном интервале энергии на единичный объем кристалла. $N(E)$ определяется условием

$$dS = N(E) dE = d\tau_k / 8\pi^3. \quad (29.3p)$$

Функция $N(E)$ определяется формой изоэнергетических поверхностей.

3. При квадратичной зависимости энергии от квазиимпульса для $N(E)$ у минимума энергии E_c имеем

$$N(E) = 2 \left(\frac{2\pi m_d^*}{h^2} \right)^{3/2} (E - E_c)^{1/2} \quad (29.4p)$$

и у максимума энергии E_v имеем

$$N(E) = 2 \left(\frac{2\pi m_{pd}^*}{h^2} \right)^{3/2} (E_v - E)^{1/2}. \quad (29.5p)$$

4. Эффективная масса для плотности состояний m_d^* связана с тензором эффективной массы m^* и числом M эквивалентных экстремумов энергии:

$$m_d^* = (M^2 m_1 m_2 m_3)^{1/3}. \quad (29.6p)$$

Если экстремум энергии лежит в центре зоны Бриллюэна, так что $V=1$, и поверхности энергии сферические, то $m_d^* = m^*$.

5. Концентрации электронов и дырок определяются соответственно следующими соотношениями:

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi); \quad (29.7p)$$

$$p = \frac{2N_v}{\sqrt{\pi}} \Phi_{1/2}(\eta), \quad (29.8p)$$

где через N_c и N_v обозначено эффективное число состояний для зоны проводимости и валентной зоны:

$$N_c = 2 \left(\frac{2\pi m_d^* kT}{h^2} \right) = 4,82 \cdot 10^{15} \left(\frac{m_d^*}{m} \right)^{3/2} T^{3/2} = 2,5 \cdot 10^{19} \left(\frac{m_d^*}{m} \right)^{3/2} \left(\frac{T}{300} \right)^{3/2}, \quad (29.9p)$$

$$\begin{aligned} N_v &= 2 \left(\frac{2\pi m_{pd}^* kT}{h^2} \right)^{3/2} = 4,82 \cdot 10^{15} \left(\frac{m_{pd}^*}{m} \right)^{3/2} T^{3/2} = \\ &= 2,5 \cdot 10^{19} \left(\frac{m_{pd}^*}{m} \right)^{3/2} \left(\frac{T}{300} \right)^{3/2}. \end{aligned} \quad (29.10p)$$

6. Концентрации электронов и дырок в невырожденном полупроводнике определяются выражениями:

$$n = N_c e^{-\frac{E_c - F}{kT}}, \quad (29.11p)$$

$$p = N_v e^{-\frac{F - E_v}{kT}}. \quad (29.12p)$$

Произведение концентраций электронов и дырок в невырожденном полупроводнике от положения уровня Ферми не зависит:

$$np = N_c N_v e^{-\frac{\Delta E_0}{kT}}. \quad (29.13p)$$

Полупроводник называется невырожденным, если уровень Ферми лежит ниже E_c или выше E_v не менее чем на kT . Другими словами, в невырожденном полупроводнике уровень Ферми находится в запрещенной зоне.

7. Полупроводник называется полностью вырожденным, если уровень Ферми лежит внутри зоны энергии на расстоянии не менее $5kT$ от экстремального значения энергии. Если уровень Ферми выше E_c не менее чем на $5kT$, то концентрация электронов не зависит от температуры

$$n = \frac{8\pi}{3} \left(\frac{2m_d^*}{h^2} \right)^{3/2} (F - E_c)^{3/2}, \quad (29.14p)$$

а концентрация дырок определяется выражением (28.12р). Если уровень Ферми ниже E_v не менее чем на $5kT$, то концентрация дырок не зависит от температуры:

$$p = \frac{8\pi}{3} \left(\frac{2m_{pd}^*}{h^2} \right)^{3/2} (E_v - F)^{3/2}, \quad (29.15p)$$

а концентрация электронов определяется выражением (29.11р).

8. Для вырожденных полупроводников независимо от закона дисперсии

$$n = \frac{1}{4\pi^3} \int_{(V_F)} f d\tau_k = \frac{1}{4\pi^3} V_F, \quad (29.16p)$$

где V_F — объем зоны Бриллюэна, ограниченный поверхностью Ферми. Для сферических поверхностей энергии $V_F = \frac{4\pi}{3} \kappa_F^3$,

$$n = \frac{1}{3\pi^2} \kappa_F^3, \quad (29.17p)$$

а радиус сферы Ферми

$$\kappa_F = (3\pi^2 n)^{1/3}. \quad (29.18p)$$

9. Для закона дисперсии Кейна при малых κ из (28.53) и (29.18р) следует:

$$n = \frac{1}{3\pi^2} \left(\frac{3}{2P^2} \right)^{3/2} F^{3/2} (E_g + F)^{3/2} \quad (29.19p)$$

и

$$n = \frac{8\pi}{3} \left(\frac{2m_{p2}^*}{h^2} \right)^{3/2} F^{3/2} \left(1 + \frac{F}{E_g} \right)^{3/2}. \quad (29.20p)$$

§ 30. УРАВНЕНИЕ ЭЛЕКТРОНЕЙТРАЛЬНОСТИ

Выражения для n и p позволяют вычислять концентрацию электронов и дырок при условии, что известно положение уровня Ферми. Однако уровень Ферми сам зависит от температуры и концентрации носителей заряда. Его положение может сильно меняться при введении примесей, создающих локализованные состояния. Это естественно, поскольку уровень Ферми определяет распределение электронов по состояниям. Вводя примесь, создаем в запрещенной зоне локализованные состояния, в которых могут находиться как электроны, так и дырки. Перераспределение электронов по состояниям при создании дискретных уровней в запрещенной зоне регулируется посредством изменения положения уровня Ферми.

Для вычисления величины F служит уравнение, которое обычно называется уравнением электронейтральности, имеющее простой и наглядный физический смысл. Прежде всего, предположим, что в полупроводнике имеются донорная и акцепторная примеси с концентрациями N_d и N_a . В результате термической ионизации в полу-

проводнике создается некоторое число свободных электронов и дырок. Свободные носители зарядов создаются в результате ионизации атомов примеси и атомов основного вещества; другими словами, в полупроводнике имеются свободные носители заряда и ионы. Как во всем кристалле, так и в любом физически малом объеме вещества суммарный заряд всех заряженных частиц должен быть равен нулю — это и есть условие электронейтральности, справедливое для незаряженного в целом тела.

Запишем условие электронейтральности для единичного объема вещества. Для этого необходимо подсчитать заряд положительных и отрицательных частиц.

Электроны возникают за счет ионизации донорной примеси и атомов основного вещества. Переход электронов из валентной зоны в зону проводимости или к атомам акцепторной примеси приводит к созданию свободных дырок.

Отрицательный заряд создают свободные электроны и ионы акцепторов, он равен $(n + N_a^-) e^-$. Положительный заряд создают свободные дырки и ионы доноров общей величиной $(p + N_d^+) e^+$.

Условие электронейтральности можно записать в виде

$$(p + N_d^+) e^+ + (n + N_a^-) e^- = 0. \quad (30.1)$$

Учитывая, что $e^- = -e^+$, получим уравнение электронейтральности

$$(n + N_a^-) - (p + N_d^+) = 0. \quad (30.2)$$

Если обозначить через n_d , p_d , n_a , p_a число электронов и дырок, находящихся на донорном и акцепторном уровнях, то можно записать ряд очевидных соотношений:

$$\begin{aligned} n_d &= N_d - N_d^+ = N_d - p_d; & N_d^+ &= N_d - n_d = p_d, \\ p_a &= N_a - N_a^- = N_a - n_a; & N_a^- &= N_a - p_a = n_a. \end{aligned} \quad (30.3)$$

В таком случае уравнение электронейтральности (30.1) или (30.2) можно записать в одной из таких форм:

$$(n + n_a) - (p + p_d) = 0, \quad (30.4)$$

или

$$n + n_d - p - p_a = N_d - N_a. \quad (30.5)$$

Чтобы составить уравнение, из которого можно определить положение уровня Ферми, необходимо выразить входящие в уравнение электронейтральности (30.5) величины через F . Выражения для n и p известны, необходимо найти n_d и p_d (или n_a и p_a). Чтобы найти число электронов, находящихся на примесном уровне, необходимо знать функцию распределения электронов по примесным состояниям.

Легко видеть, что функция Ферми — Дирака не может быть использована непосредственно для описания распределения электро-

нов по примесным состояниям, так как она справедлива в том случае, когда в данном энергетическом состоянии может быть два электрона с противоположно направленными спинами. Однако в состоянии с энергиями E_d или E_a может быть не более одного электрона. Если поместить в состояние с E_d или E_a еще один электрон, то в силу большого электростатического взаимодействия между электронами энергия E_d или E_a резко меняется. Другими словами уровни энергии однократно и двукратно ионизированной примеси отличны друг от друга. Невозможность нахождения двух и более электронов в примесном центре с энергией E_d или E_a должна отражать функция распределения электронов по состояниям. Расчет по методу Гиббса для систем с переменным числом частиц приводит к выводу, что для примесных состояний функция распределения электронов должна иметь вид

$$f = \frac{1}{\frac{1}{g_i} e^{\frac{E_i - F}{kT}} + 1}, \quad (30.6)$$

где g_i — степень вырождения i -го примесного состояния. Если $E_i = E_d$ принадлежит донорной примеси, то $g_i = 2$. Если $E_i = E_a$ принадлежит акцепторной примеси, то $g_i = 1/2$; таким образом распределение электронов по донорным уровням определяется выражением

$$f = \frac{1}{\frac{1}{2} e^{\frac{E_d - F}{kT}} + 1}, \quad (30.7)$$

а распределение электронов по акцепторным уровням имеет вид

$$f = \frac{1}{2e^{\frac{E_a - F}{kT}} + 1}. \quad (30.8)$$

Для дырок функция распределения имеет соответственно вид

$$f_p = \frac{1}{2e^{\frac{F - E_d}{kT}} + 1}; \quad f_p = \frac{1}{\frac{1}{2} e^{\frac{F - E_a}{kT}} + 1}. \quad (30.9)$$

Теперь легко найти число электронов или дырок, находящихся на примесных уровнях. Найдем, например, n_d :

$$n_d = \int N_d(E) f dE = N_d \int \delta(E - E_d) \frac{1}{\frac{1}{2} e^{\frac{E - F}{kT}} + 1} dE = \frac{N_d}{\frac{1}{2} e^{\frac{E_d - F}{kT}} + 1}. \quad (30.10)$$

Аналогично можно найти n_a, p_a, p_d :

$$n_a = \frac{N_a}{\frac{E_a - F}{2e} \frac{1}{kT} + 1}, \quad (30.11)$$

$$p_a = \frac{N_a}{\frac{1}{2} e \frac{1}{kT} + 1}; \quad p_d = \frac{N_d}{\frac{F - E_d}{2e} \frac{1}{kT} + 1}. \quad (30.12)$$

Запишем уравнение электронейтральности (30.5) с учетом явных выражений (29.1), (29.13), (30.10) и (30.12) для n, p, n_d и p_a :

$$4\pi \left(\frac{2m_d^*}{h^2}\right)^{3/2} \int_{E_c}^{\infty} \frac{(E - E_c)^{1/2} dE}{e \frac{1}{kT} + 1} + \frac{N_d}{\frac{1}{2} e \frac{1}{kT} + 1} - 4\pi \left(\frac{2m_{pd}^*}{h^2}\right)^{3/2} \int_{-\infty}^{E_v} \frac{(E_v - E)^{1/2} dE}{e \frac{1}{kT} + 1} - \frac{N_a}{\frac{1}{2} e \frac{1}{kT} + 1} = N_d - N_a. \quad (30.13)$$

Если рассматривать это уравнение относительно уровня Ферми, то оно может быть сведено к алгебраическому уравнению четвертой степени. Решать в общем виде его очень сложно, поэтому рассмотрим ряд частных случаев, имеющих важное практическое значение. Кроме того, необходимо помнить, что уравнение записано для случая, когда имеется всего лишь один донорный и один акцепторный уровень. В противном случае уравнение (30.13) будет еще более сложным.

§ 31. СОБСТВЕННЫЙ ПОЛУПРОВОДНИК

Полупроводник, в котором примесь $N_d = N_a = 0$ отсутствует, называется чистым или собственным полупроводником. Уравнение электронейтральности имеет вид

$$n - p = 0, \text{ или } n = p, \quad (31.1)$$

т. е. переход каждого электрона из валентной зоны порождает в ней дырку, поэтому число дырок в собственном полупроводнике всегда равно числу электронов (рис. 49). При условии (31.1) уравнение (30.13) для определения положения уровня Ферми имеет вид

$$\frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi) = \frac{2N_v}{\sqrt{\pi}} \Phi_{1/2}(\eta). \quad (31.2)$$

Предположим, что $N_v = N_c$, для этого необходимо, чтобы эффективные массы для плотности состояний были равны, т. е. $m_d^* = m_a^*$. В этом случае

$$\Phi_{1/2}(\xi) = \Phi_{1/2}(\eta), \quad (31.3)$$

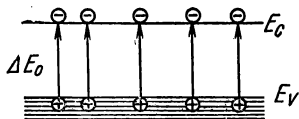


Рис. 49. Генерация электронов и дырок проводимости в собственном полупроводнике

следовательно, $\xi = \eta$ или, учитывая выражения (29.3) и (29.12), получим

$$\frac{F - E_c}{kT} = \frac{E_v - F}{kT}, \quad (31.4)$$

откуда

$$F = \frac{E_v + E_c}{2} = E^i, \quad (31.5)$$

т. е. уровень Ферми от температуры не зависит и лежит точно посредине запрещенной зоны (совпадает с E^i).

Поскольку ширина запрещенной зоны для полупроводников много больше, чем $2kT$, то $\xi \ll -1$ и $\eta \ll -1$; собственный полупроводник является невырожденным и интеграл Ферми равен $\frac{\sqrt{\pi}}{2} e^{\xi}$ и $\frac{\sqrt{\pi}}{2} e^{\eta}$ соответственно, поэтому для концентрации электронов и дырок получим следующие выражения:

$$n = N_c e^{\xi} = N_c e^{-\frac{E_c - F}{kT}} = N_c e^{-\frac{\Delta E_0}{2kT}} = N_c e^{-\frac{\delta E_a}{kT}}, \quad (31.6)$$

$$p = N_v e^{\eta} = N_v e^{-\frac{F - E_v}{kT}} = N_v e^{-\frac{\Delta E_0}{2kT}} = N_v e^{-\frac{\delta E_a}{kT}}. \quad (31.7)$$

Энергия активации δE_a для собственного полупроводника равна половине ширины запрещенной зоны.

Мы получили выражения (31.6) и (31.7) в предположении, что $N_v = N_c$. Пусть $N_v \neq N_c$. Но поскольку они являются величинами одного порядка, интеграл Ферми опять-таки можно представить в виде $\frac{\sqrt{\pi}}{2} e^{\xi}$ и $\frac{\sqrt{\pi}}{2} e^{\eta}$, сокращая (31.2) на $\frac{\sqrt{2}}{\pi}$, получим

$$n = N_c e^{\xi} = N_v e^{\eta} = p, \quad (31.8)$$

или

$$\xi - \eta = \ln \frac{N_v}{N_c} = \frac{F - E_c}{kT} - \frac{E_v - F}{kT}. \quad (31.9)$$

Из (31.9) следует

$$F = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} = \frac{E_c + E_v}{2} - kT \ln \left(\frac{N_c}{N_v} \right)^{1/2}. \quad (31.10)$$

Так как

$$\frac{N_c}{N_v} = \left(\frac{m_d^*}{m_{pd}^*} \right)^{3/2} = \left(\frac{m_{nd}^*}{m_{pd}^*} \right)^{3/2},$$

то

$$F = \frac{E_c + E_v}{2} + kT \ln \left(\frac{m_{pd}^*}{m_{nd}^*} \right)^{3/4}. \quad (31.11)$$

Через m_{nd}^* дана величина, которую раньше обозначали как m_d^* . Уровень Ферми при $T = 0$ лежит в середине запрещенной зоны; он

линейно зависит от температуры (рис. 50). С ростом температуры уровень Ферми приближается к той зоне, которая имеет меньшую плотность состояний и поэтому заполняется быстрее. Хотя на уровне Ферми в собственном полупроводнике и нет электронов, он сохраняет тот смысл, что величина $f = 1/2$ при $E = E_i$. Концентрация собственных носителей заряда определяется уравнением

$$n_i = n = p = \sqrt{N_c N_v} e^{-\frac{\Delta E_0}{2kT}}. \quad (31.12)$$

Если подставить численные значения констант, входящих в выражения N_c и N_v , то выражение для концентрации n_i будет иметь вид

$$n_i = 4,82 \cdot 10^{15} \left(\frac{m_{nd}^* m_{pd}^*}{m^2} \right)^{3/4} T^{3/2} e^{-\frac{\Delta E_0}{2kT}}, \quad (31.13)$$

$$np = n_i^2 = 2,31 \cdot 10^{31} \left(\frac{m_{nd}^* m_{pd}^*}{m^2} \right)^{3/2} T^3 e^{-\frac{\Delta E_0}{kT}}. \quad (31.14)$$

Если построить графическую зависимость $\ln n_i$ от обратной температуры, то она будет представлять собой практически прямую линию:

$$\ln n_i = \text{const} - \frac{3}{2} \ln \frac{1}{T} - \frac{\Delta E_0}{2k} \cdot \frac{1}{T}, \quad (31.15)$$

так как зависимость $\ln \frac{1}{T}$ по сравнению с линейным членом можно пренебречь. Угол наклона прямой определяется шириной запрещенной зоны:

$$\text{tg } \varphi = - \frac{\Delta E_0}{2k},$$

откуда

$$\Delta E_0 = 2k |\text{tg } \varphi|; \quad (31.16)$$

$\text{tg } \varphi$ измеряется по графику $\left(\ln n_i, \frac{1}{T} \right)$ (рис. 51).

Рис. 51. Зависимость $\ln n_i$ в собственном полупроводнике от обратной температуры

Оценим собственную концентрацию носителей заряда в некоторых полупроводниках. Для германия и кремния $\left(\frac{m_{nd}^* m_{pd}^*}{m^2} \right)^{3/2}$ равны

0,299 и 0,719, и при $T \approx 300^\circ \text{K}$ $n_i \approx 2 \cdot 10^{13} \text{ см}^{-3}$ и $n_i \approx 2 \cdot 10^{10} \text{ см}^{-3}$.

Согласно выражению для n_i концентрация носителей заряда при $T=0$ обращается в нуль, и сопротивление собственного полупроводника должно расти до бесконечности при $T \rightarrow 0$. Однако в реальных полупроводниках всегда остается примесь, которая обеспечивает конечную проводимость при любых температурах.

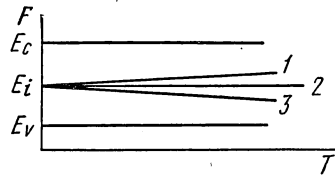


Рис. 50. Температурная зависимость уровня Ферми в собственном полупроводнике: 1 — $m_{nd}^* < m_{pd}^*$; 2 — $m_{nd}^* = m_{pd}^*$; 3 — $m_{nd}^* > m_{pd}^*$; $E_i = \frac{E_v + E_c}{2}$ — середина запрещенной зоны

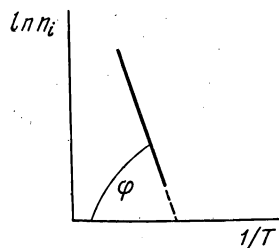


Рис. 51. Зависимость $\ln n_i$ в собственном полупроводнике от обратной температуры

Для германия и кремния $\left(\frac{m_{nd}^* m_{pd}^*}{m^2} \right)^{3/2}$ равны

0,299 и 0,719, и при $T \approx 300^\circ \text{K}$ $n_i \approx 2 \cdot 10^{13} \text{ см}^{-3}$ и $n_i \approx 2 \cdot 10^{10} \text{ см}^{-3}$.

Согласно выражению для n_i концентрация носителей заряда при $T=0$ обращается в нуль, и сопротивление собственного полупроводника должно расти до бесконечности при $T \rightarrow 0$. Однако в реальных полупроводниках всегда остается примесь, которая обеспечивает конечную проводимость при любых температурах.

В заключение этого параграфа рассмотрим вопрос о концентрации носителей заряда в полупроводниках, имеющих вырожденные зоны, подобно p -германию и p -кремнию.

Каждая зона характеризуется своей эффективной массой m_{p1}^* и m_{p2}^* . Плотность состояний для них может быть записана в виде:

$$N_1(E) = 2\pi \left(\frac{2m_{p1}^*}{h^2} \right)^{3/2} (E_v - E)^{1/2}, \quad (31.17)$$

$$N_2(E) = 2\pi \left(\frac{2m_{p2}^*}{h^2} \right)^{3/2} (E_v - E)^{1/2}. \quad (31.18)$$

Введем полную плотность состояний

$$N(E) = N_1(E) + N_2(E) = 2\pi \left(\frac{2m_{pd}^*}{h^2} \right)^{3/2} (E_v - E)^{1/2}, \quad (31.19)$$

где

$$(m_{pd}^*)^{3/2} = (m_{p1}^*)^{3/2} + (m_{p2}^*)^{3/2}, \quad (31.20)$$

после чего можно воспользоваться выражением (31.13) для концентрации носителей заряда. Если вырождение имеет кратность g , то необходимо сложить эффективные массы плотности состояний с учетом числа максимумов M для каждого сорта дырок (конечно, аналогично и для электронов) в степени $3/2$.

Легко видеть, что отношение концентраций дырок разного вида определяется только плотностями состояний для них:

$$\frac{p_1}{p_2} = \frac{N_{v1}}{N_{v2}} = \left(\frac{m_{p1}^*}{m_{p2}^*} \right)^{3/2}, \quad (31.21)$$

т. е. легких носителей заряда должно быть меньше, чем тяжелых. Для германия и кремния отношение масс $m_{п1}^*/m_{п2}^*$ легких и тяжелых дырок равно 0,13 и 0,3, поэтому концентрация легких дырок должна составлять примерно 0,047 и 0,165 от концентрации тяжелых дырок. Концентрация легких дырок в германии составляет 4,5% и в кремнии 14% от общего числа дырок.

В заключение этого параграфа рассмотрим случай, когда ширина зоны невелика: $\Delta E_0 \ll kT$. Если зона проводимости перекрывается с валентной зоной на величину Δ , то вещество называется полуметаллом. В данном случае рассмотрен чистый, или собственный, полуметалл с отрицательной запрещенной зоной

$$E_c - E_v = -\Delta. \quad (31.22)$$

При $kT \ll \Delta$ полуметалл может быть вырожденным одновременно как по электронам, так по дыркам, поэтому для концентрации электронов и дырок можно использовать выражения (29.14p) и (29.15p) и записать:

$$n = \frac{8\pi}{3} \left(\frac{2m_{nd}^*}{h^2} \right)^{3/2} \delta_n^{3/2}, \quad (31.23)$$

$$p = \frac{8\pi}{3} \left(\frac{2m_{pd}^*}{h^2} \right)^{3/2} \delta_p^{3/2}, \quad (31.24)$$

где

$$\delta_n = F - E_c; \delta_p = E_v - F; \delta_n + \delta_p = E_v - E_c = \Delta. \quad (31.25)$$

При повышении температуры вырождение одного из газов может быть снято. Наиболее вероятно снятие вырождения для дырочного газа, поскольку эффективная масса дырок больше эффективной массы электронов, и быстрее заполняется зона проводимости. Когда уровень Ферми перейдет из области перекрытия выше потолка валентной зоны, сохранится вырождение только электронного газа.

В области двойного вырождения из (31.23—24) с учетом $n = p$ имеем

$$m_{nd}^* \delta_n = m_{pd}^* \delta_p \quad (31.26)$$

и

$$\delta_n = \frac{m_{pd}^*}{m_{nd}^* + m_{pd}^*} \Delta; \delta_p = \frac{m_{nd}^*}{m_{nd}^* + m_{pd}^*} \Delta. \quad (31.27)$$

Для собственной концентрации полуметалла в области вырождения электронного и дырочного газа имеем

$$n = p = n_i = \frac{8\pi}{3} \left(\frac{2}{h^2} \cdot \frac{m_{nd}^* m_{pd}^*}{m_{nd}^* + m_{pd}^*} \right)^{3/2} \Delta^{3/2}. \quad (31.28)$$

Как видим из (31.28), собственная концентрация полуметалла в вырожденном состоянии определяется перекрытием зон энергии Δ при $\Delta = 0$ и $n_i = 0$.

Для полупроводника с положительной запрещенной зоной одно-временное вырождение электронов и дырок невозможно. Но если эффективная масса электронов значительно меньше эффективной массы дырок, то уровень Ферми может оказаться в зоне проводимости, в результате чего в собственном узкозонном полупроводнике электронный газ окажется вырожденным. Такое состояние реализуется, например, в теллуриде ртути, для которого термическая ширина запрещенной зоны (но не параметр Кейна E_g) равна нулю.

§ 32. ПРИМЕСНЫЙ ПОЛУПРОВОДНИК. ПРИМЕСЬ ОДНОГО ВИДА

Рассмотрим полупроводник, в который введена примесь одного вида, например донорная, так, что $N_d \neq 0$ и $N_a = 0$. Уравнение нейтральности (30.5) примет вид

$$n + n_d - p = N_d, \quad (32.1)$$

или

$$n = p + N_d^+. \quad (32.2)$$

Уравнение (32.2) имеет очень наглядный смысл. Свободные электроны возникают благодаря переходам из валентной зоны, что при-

водит к созданию p свободных дырок, и с уровня примеси, благодаря чему возникает N_d^+ -ионов донорной примеси (рис. 52). Уравнение (32.2) относительно F является уравнением третьей степени. Однако можно найти ряд сравнительно простых случаев, для которых можно легко определить положение уровня Ферми и тем самым концентрацию электронов и дырок.

Прежде всего отметим, что в невырожденном полупроводнике достаточно определить концентрацию носителей заряда только одного знака, концентрация носителей заряда другого знака легко вычисляется на основании соотношений (30.5) и (31.12):

$$np = n_i^2 = N_c N_v e^{-\frac{\Delta E_0}{kT}} \quad (32.3)$$

Выражение (32.3) показывает, что произведение концентраций электронов и дырок в невырожденном полупроводнике не зависит от положения уровня Ферми и тем самым от наличия в полупроводнике примеси и равно квадрату концентрации в собственном полупроводнике. Соотношение (32.3) позволяет находить концентрацию одного из компонентов по другому. Например, если концентрация электронов известна, то концентрация дырок находится на основании соотношения

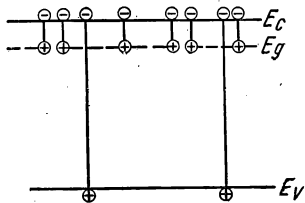


Рис. 52. Тепловая генерация носителей заряда в полупроводнике с донорной примесью

$$p = \frac{n_i^2}{n} = \frac{N_c N_v e^{-\frac{\Delta E_0}{kT}}}{n} \quad (32.4)$$

Вернемся к уравнению (32.2). Так как свободные электроны возникают благодаря ионизации примеси и основного вещества, то при данной температуре эти два процесса могут играть неодинаковую роль. Для перевода электрона из валентной зоны в зону проводимости необходима энергия, равная ширине запрещенной зоны ΔE_0 , в то время как для перевода электрона с уровня примеси в зону проводимости необходима энергия, равная энергии ионизации примеси ΔE_d , которая много меньше ширины запрещенной зоны. Поэтому при низкой температуре основную роль будут играть переходы электронов с примесного уровня, следовательно, $p \ll N_d^+$. Неравенство сохранится до тех пор, пока вся примесь не будет ионизована. Однако с ростом температуры произойдет ионизация примеси, и рост концентрации электронов n будет происходить вместе с ростом концентрации дырок p . При больших температурах $p \gg N_d^+ = N_d$, и полупроводник станет собственным. Понятие больших и малых температур должно быть связано с концентрацией примеси. Одна и та же температура может быть или большой, или малой в зависимости от концентрации примеси — при малой концентрации она может оказаться большой, при большой концентрации — малой.

1. Низкие температуры. Рассмотрим малые температуры, при которых основную роль играет ионизация примеси, в этом случае $p \ll N_d^+$ и уравнение (32.2) упрощается:

$$n = N_d^+, \text{ или } n = p_d. \quad (32.5)$$

Подставляя в (32.5) выражения для n и p_d , получим

$$n = N_c e^{-\frac{E_c - F}{kT}} = p_d = \frac{N_d}{2e^{\frac{F - E_d}{kT}} + 1}. \quad (32.6)$$

Обозначив $e^{\frac{F}{kT}} = x$, получим

$$N_c e^{-\frac{E_c}{kT}} x \left(1 + 2e^{-\frac{E_d}{kT}} x \right) = N_d, \quad (32.7)$$

или

$$2N_c e^{-\frac{E_c + E_d}{kT}} x^2 + N_c e^{-\frac{E_c}{kT}} x - N_d = 0, \quad (32.8)$$

$$x^2 + \frac{1}{2} e^{\frac{E_d}{kT}} x - \frac{N_d}{2N_c} e^{\frac{E_c + E_d}{kT}} = 0. \quad (32.9)$$

Решая уравнение (32.9) относительно x , получим

$$x = \frac{1}{4} e^{\frac{E_d}{kT}} \left(\pm \sqrt{1 + \frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}} - 1} \right). \quad (32.10)$$

Так как $x > 0$, минус перед радикалом необходимо отбросить, и из (32.10)

$$F = kT \ln \left\{ \frac{1}{4} e^{\frac{E_d}{kT}} \left(\sqrt{1 + \frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}} - 1} \right) \right\}. \quad (32.11)$$

Напомним, что выражение (32.11) соответствует столь низким температурам, при которых $p \ll N_d^+$ и $n = N_d^+$, т. е. электроны проводимости возникают в основном за счет ионизации донорной примеси.

Рассмотрим два предельных случая для F . С ростом температуры $\frac{\Delta E_d}{kT}$ стремится к единице, N_c возрастает и может стать больше N_d , однако при достаточно малых температурах может быть выполнено неравенство

$$\frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}} \gg 1, \quad (32.12)$$

поэтому для x получим

$$x = \frac{1}{4} e^{\frac{E_d}{kT}} \sqrt{\frac{8N_d}{N_c} e^{\frac{E_c - E_d}{2kT}}} = \sqrt{\frac{N_d}{2N_c}} e^{\frac{E_c + E_d}{2kT}} = e^{\frac{F}{kT}}. \quad (32.13)$$

Из (32.13) запишем выражение для положения уровня Ферми:

$$F = \frac{E_c + E_d}{2} + \frac{kT}{2} \ln \frac{N_d}{2N_c}. \quad (32.14)$$

При $T=0$

$$F = \frac{E_c + E_d}{2}, \quad (32.15)$$

т. е. уровень Ферми лежит посредине между дном зоны проводимости и примесным уровнем. При повышении температуры уровень Ферми повышается, проходит при некоторой температуре через максимум, затем опускается. При $2N_c = N_d$ он снова находится в середине между E_c и E_d . Однако для этой области температур справедливость формулы (32.14) определяется величиной концентрации N_d : для малых концентраций примеси N_c равно $N_d/2$ при таких температурах, для которых $e^{\frac{\Delta E_d}{2kT}} \gg 1$. Если же N_d велико, так что $e^{\frac{\Delta E_d}{2kT}} \approx 1$, то формула (32.14) для F не применима.

Найдем величину концентрации электронов:

$$n = N_c e^{\frac{E_c - F}{kT}} = \sqrt{\frac{N_c N_d}{2}} e^{-\frac{\Delta E_d}{2kT}}. \quad (32.16)$$

Обращает на себя внимание, что при малых температурах концентрация электронов определяется концентрацией примеси в степени $1/2$.

Если построим график $(\ln n, 1/T)$, то увидим, что он имеет практически вид прямой линии с тангесом угла наклона, равным $\Delta E_d/2k$, в то время как для собственной концентрации угол наклона определяется величиной $\Delta E_v/2k$.

Теперь рассмотрим противоположный случай:

$$\frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}} \ll 1; \text{ или } N_c \gg 8N_d. \quad (32.17)$$

Для этого необходима при той же концентрации доноров достаточно высокая температура с тем, чтобы $8N_d \ll N_c$. Разлагая радикал (32.12) в ряд и ограничиваясь первым членом, получим

$$x = \frac{1}{4} e^{\frac{E_d}{kT}} \left(1 + \frac{4N_d}{N_c} e^{\frac{\Delta E_d}{kT}} + \dots - 1 \right) = \frac{N_d}{N_c} e^{\frac{E_c}{kT}}. \quad (32.18)$$

Из (32.18) для уровня Ферми следует выражение

$$F = E_c + kT \ln \frac{N_d}{N_c}. \quad (32.19)$$

Так как оно справедливо при $N_c \gg N_d$, то логарифм в (32.19) меньше нуля, поэтому с ростом температуры уровень Ферми опускается. Найдем концентрацию электронов для этого случая:

$$n = N_c e^{-\frac{E_c - F}{kT}} = N_c e^{\ln \frac{N_d}{N_c}} = N_d, \quad (32.20)$$

т. е. концентрация электронов не зависит от температуры и равна концентрации примеси. Эта область температур носит название области истощения примеси. Напомним, что носители заряда называют основными, если их концентрация больше концентрации собственных носителей заряда n_i при данной температуре, если же их концентрация меньше n_i , то их называют неосновными носителями заряда. Мы видим, что в полупроводнике с донорной примесью основными носителями заряда являются электроны. Таким образом, можем сказать, что в области истощения примеси концентрация основных носителей заряда остается постоянной, концентрация же неосновных носителей заряда должна резко возрастать с температурой; действительно, из (32.4) можем записать:

$$p = \frac{n_i^2}{n} = \frac{n_i^2}{N_d} = \frac{N_c N_v}{N_d} e^{-\frac{\Delta E_0}{kT}}. \quad (32.21)$$

Это выражение будет справедливо до тех пор, пока концентрация дырок остается много меньше концентрации электронов:

$$p \ll n = N_d^+ = N_d. \quad (32.22)$$

2. Высокие температуры. С ростом температуры концентрация дырок возрастает и может стать сравнимой с концентрацией электронов. В этом случае уравнение (32.5) должно быть заменено общим уравнением (32.2), которое можем теперь существенно упростить. Действительно, запишем (32.2) с учетом (32.20):

$$n = p + N_d. \quad (32.23)$$

Оно справедливо для случая, когда вся примесь ионизована и необходимо учитывать ионизацию основного вещества.

Запишем уравнение (32.23) для невырожденного полупроводника:

$$n = \frac{n_i^2}{n} + N_d, \quad (32.24)$$

или

$$n^2 - nN_d - n_i^2 = 0. \quad (32.25)$$

Решая это уравнение, получим

$$n = \frac{N_d}{2} \left(1 \pm \sqrt{1 + \frac{4n_i^2}{N_d^2}} \right). \quad (32.26)$$

Поскольку подкоренное выражение больше единицы, а $n > 0$, минус перед корнем необходимо отбросить. Запишем теперь выражение для концентрации электронов и дырок:

$$n = \frac{N_d}{2} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_d^2}} \right), \quad (32.27)$$

$$p = \frac{2n_i^2}{N_d \left(1 + \sqrt{1 + \frac{4n_i^2}{N_d^2}} \right)}. \quad (32.28)$$

Учитывая связь между n и F , из (32.27) получим

$$\begin{aligned} F &= E_c + kT \ln \left\{ \frac{N_d}{2N_c} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_d^2}} \right) \right\} = \\ &= E_c + kT \ln \left\{ \frac{N_d}{2N_c} \left(1 + \sqrt{1 + \frac{4N_c N_v}{N_d^2} e^{-\frac{\Delta E_0}{kT}}} \right) \right\}. \end{aligned} \quad (32.29)$$

Рассмотрим два предельных случая. Если

$$\frac{4n_i^2}{N_d^2} \ll 1, \quad (32.30)$$

то из (32.27) и (32.29) следует

$$n = N_d; \quad p = \frac{n_i^2}{N_d}; \quad F = E_c + kT \ln \frac{N_d}{N_c}, \quad (32.31)$$

что находится в полном соответствии с полученными выше результатами для области истощения. Если же

$$\frac{4n_i^2}{N_d^2} \gg 1, \quad (32.32)$$

то

$$n = p = n_i; \quad F = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c}, \quad (32.33)$$

как и должно быть в случае собственного полупроводника.

Таким образом, положение уровня Ферми в невырожденном полупроводнике описывается двумя выражениями во всем возможном интервале температур. Выражение

$$F = E_d + kT \ln \left\{ \frac{1}{4} \left(\sqrt{1 + \frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}}} - 1 \right) \right\} \quad (32.34)$$

справедливо в области от $T=0$ до температуры истощения T_n . Область от T_n и выше описывается уравнением

$$F = E_c + kT \ln \left\{ \frac{N_d}{2N_c} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_d^2}} \right) \right\}. \quad (32.35)$$

Соотношения (32.34) и (32.35) позволяют описать температурную зависимость уровня Ферми в данном полупроводнике. При $T=0$ уровень Ферми лежит в середине между дном зоны проводимости и донорным уровнем. При повышении температуры уровень Ферми поднимается к E_c , затем в связи с ростом N_c он проходит через максимум и начинает опускаться.

Концентрация электронов при этом возрастает за счет ионизации примеси. При некоторой температуре $E=E_d$ — на уровне примеси находится $\frac{2}{3}N_d$ электронов, а $n=N_d/3$. При дальнейшем опускании уровня Ферми полупроводник переходит в область истощения: примесь вся ионизована, концентрация электронов остается постоянной, концентрация дырок возрастает, уровень Ферми приближается к середине запрещенной зоны. По мере приближения уровня Ферми к середине запрещенной зоны концентрация дырок возрастает при практически неизменной концентрации электронов. При дальнейшем росте концентрации дырок будет происходить и рост концентрации электронов, достигается равенство $n=p$, и полупроводник из примесного превращается в собственный полупроводник.

Температура перехода от истощения примеси к собственной проводимости зависит от концентрации примеси для данного полупроводника и от ширины запрещенной зоны при фиксированной концентрации примеси. Если определить переход от примесной к собственной концентрации некоторой условной границей $p=N_d$ или $n=2N_d$, то легко видеть, что температура перехода определится уравнением

$$pn = n_i^2 = 2N_d^2 = N_c N_v e^{-\frac{\Delta E_0}{kT_{ист}}},$$

или

$$T_{ист} = \frac{\Delta E_0}{k \ln \frac{N_c(T_{ист}) N_v(T_{ист})}{2N_d^2}}. \quad (32.36)$$

При фиксированном значении N_d температура перехода к собственной концентрации, тем выше, чем больше ΔE_0 . Для данного полупроводника температура перехода к собственной концентрации больше при большей концентрации примеси.

На рис. 53 показано положение уровня Ферми в зависимости от температуры при трех различных значениях концентрации примеси.

Запишем выражение для F и p в случае акцепторной примеси:

$$p = n + p_a = n + N_a^-. \quad (32.37)$$

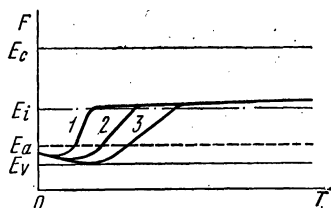


Рис. 53. Температурная зависимость уровня Ферми в полупроводнике с донорной примесью: 1 — $N_{д1}$; 2 — $N_{д2}$; 3 — $N_{д3}$ ($N_{д1} < N_{д2} < N_{д3}$)

Решая уравнение (32.7) таким же образом, как и в случае донорной примеси, получим при $n \ll N_a^-$

$$F = E_a - kT \ln \left\{ \frac{1}{4} \left(\sqrt{1 + \frac{8N_a}{N_v} e^{\frac{\Delta E_a}{kT}} - 1} \right) \right\}. \quad (32.38)$$

В случае $N_a^- = N_a$ и $n \gtrsim N_a$, аналогично получим

$$F = E_v - kT \ln \left\{ \frac{N_a}{2N_v} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_a^2}} \right) \right\};$$

$$p = \frac{N_a}{2} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_a^2}} \right). \quad (32.39)$$

На рис. 54 приведена температурная зависимость уровня Ферми в акцепторном полупроводнике.

В заключение этого параграфа оценим температуру, при которой наступает истощение примеси. Так как для области истощения должны выполняться уравнения (32.17) (нижняя граница) и (32.30) (верхняя граница), то, объединяя их вместе, получим

$$\frac{N_c(T_H)}{8} \gg N_a > 2n_i(T_{\text{верх}}). \quad (32.40)$$

Найдем нижнюю границу области истощения условием

$$N_c(T_H) = N_d. \quad (32.41)$$

Учитывая (28.9р), получим

$$(T_H/300)^{3/2} = (m/m_d^*)^{3/2} \cdot 4 \cdot 10^{-18} N_d, \quad (32.42)$$

$$\frac{T_H}{300} = \frac{m}{m_d^*} 2,5 \left(\frac{N_d}{10^{18}} \right)^{2/3}. \quad (32.43)$$

Пусть $m_d^* = 0,25 m$, в таком случае

$$(T_H/300) = 10 (N_d/10^{18})^{2/3}. \quad (32.44)$$

При $N_d = 10^{18} \text{ см}^{-3}$ $T_H = 3000^\circ \text{ К}$, а при $N_d = 10^{15} \text{ см}^{-3}$ $T_H = 30^\circ \text{ К}$, т. е. в первом случае истощение наступает только при 3000° К , а во втором случае уже при 30° К .

Верхнюю границу можно оценить на основании соотношения (32.36), необходимо только учесть, что $T_{\text{верх}}$ входит в N_c и N_v .

Для того чтобы оценить роль того или иного члена в выражении для n , p или F , необходимо сравнить величину kT с ΔE_0 или ΔE_d . Напомним, что постоянная Больцмана $k = 1,38 \cdot 10^{-23} \text{ Дж} \cdot \text{град}^{-1} = 1,38 \cdot 10^{-16} \text{ эрг} \cdot \text{град}^{-1} = 8,6167 \cdot 10^{-5} \text{ эВ} \cdot \text{град}^{-1}$. Обратная величина $1/k = 11605,4 \text{ град} \cdot \text{эВ}^{-1}$, другими словами, 1 эВ соответствует $11605,4 \text{ К}$. В табл. 8 приведены значения kT в электронвольтах при некоторых значениях температуры.

Таблица 8

Т, К	1	4,2	20	100	200	273	290	300	500
kT , эВ	$8,6 \cdot 10^{-5}$	$3,6 \cdot 10^{-4}$	$1,7 \cdot 10^{-3}$	$8,6 \cdot 10^{-3}$	0,017	0,0235	0,0250	0,0258	0,043

Напомним, что энергия ионизации примеси измеряется сотыми долями электронвольта.

Резюме § 30—32

1. Энергия Ферми представляет собой термодинамический потенциал Гиббса, отнесенный к одной частице, ее называют также химическим потенциалом. Она определяет приращение энергии системы частиц при увеличении числа частиц на единицу. Поэтому энергия Ферми определяется общим числом частиц. В полупроводниках энергия Ферми определяется условием распределения электронов по уровням энергии валентной зоны, зоны проводимости и дискретным уровням локализованных состояний. Уравнение, определяющее распределение электронов по состояниям, называют обычно уравнением электронейтральности:

$$n + n_d - p - p_a = N_d - N_a. \quad (32.1p)$$

Если выразить n , n_d , p , p_a через F , то получим уравнение (30.13), из которого можно найти положение уровня Ферми в общем случае.

2. Функция распределения электронов по дискретным уровням отличается от функции распределения электронов по состояниям в зоне, так как на уровне локализованного состояния может быть не более одного электрона на один примесный атом:

$$f(E_l, T) = \frac{1}{\frac{1}{g} e^{\frac{E_l - E}{kT}} + 1}. \quad (32.2p)$$

Величина $g=2$ для донорного уровня и $g=1/2$ для акцепторного уровня. Для дырок $f_p(E_l, T)$ находится из условия $f_p(E_l, T) = 1 - f(E_l, T)$.

3. Полупроводник называют собственным, если в нем нет примеси: $N_a = N_d = 0$. Для собственного полупроводника уравнение нейтральности имеет вид

$$n = p, \quad (32.3p)$$

что приводит к следующему выражению для F :

$$F = \frac{E_v + E_c}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} = \frac{E_c + E_v}{2} + \frac{3kT}{4} \ln \frac{m_{pd}^*}{m_{nd}^*}. \quad (32.4p)$$

При $T=0$ уровень Ферми лежит в середине запрещенной зоны и с ростом температуры он смещается к той зоне, в которой эффективная масса плотности состояний меньше.

4. В собственном полупроводнике концентрация носителей заряда определяется шириной запрещенной зоны:

$$n_i = n = p = \sqrt{N_c N_v} e^{-\frac{\Delta E_0}{2kT}}. \quad (32.5p)$$

5. В полупроводнике, содержащем примесь одного вида (например, донорную), уравнение нейтральности имеет вид

$$n = p + N_d^+. \quad (32.6p)$$

При низких температурах, когда $p \ll n$, полупроводник имеет примесную проводимость, положение уровня Ферми определяется уравнением

$$F = E_d + kT \ln \left\{ \frac{1}{4} \left(\sqrt{1 + \frac{8N_d}{N_c} e^{\frac{\Delta E_d}{kT}}} - 1 \right) \right\}. \quad (32.7p)$$

При $T=0$ $F = \frac{E_c + E_d}{2}$, с ростом температуры F приближается к зоне проводимости, проходит через максимум и начинает опускаться (см. рис. 53). Опускание происходит с наибольшей скоростью, когда вся примесь ионизована:

$$F = E_c + kT \ln \frac{N_d}{N_c}, \quad (32.8p)$$

$$n = N_d \quad (p \ll n). \quad (32.9p)$$

Эту область температур называют областью истощения примеси (иногда ее называют областью насыщения).

6. При температурах, когда вся примесь ионизована и происходит ионизация основного вещества, уравнение (32.6p) записывается в виде

$$n = N_d + p, \quad (32.10p)$$

что приводит к следующему выражению для F :

$$F = E_c + kT \ln \left\{ \frac{N_d}{2N_c} \left(1 + \sqrt{1 + \frac{4n_i^2}{N_c^2}} \right) \right\}, \quad (32.11p)$$

справедливому от области истощения до собственной концентрации включительно.

7. Чем шире запрещенная зона и чем больше концентрация примеси, тем при большей температуре происходит переход к собственной проводимости.

§ 33. ПОЛУПРОВОДНИК, СОДЕРЖАЩИЙ АКЦЕПТОРНУЮ И ДОНОРНУЮ ПРИМЕСЬ

Рассмотрим общий случай, когда в полупроводнике имеется как донорная, так и акцепторная примесь. Предположим, что $T=0$. В этом случае система электронов должна занимать все наиболее низкие энергетические состояния. Зона проводимости будет полностью свободна, а валентная зона целиком занята (т. е. $n=p=0$). Так как имеется N_a свободных состояний и N_d электронов, то электроны от доноров перейдут к акцепторам. Если $N_d=N_a$, то в полупроводнике образуются в равном числе ионы N_d^+ и N_a^- . Пусть теперь температура растет. Так как на донорном уровне нет электронов, то в зону проводимости возможны переходы только из валентной зоны и с уровня E_a , но расстояние E_c-E_a почти равно ширине запрещенной зоны. Поэтому концентрация n будет возрастать с ростом температуры почти так же, как в собственном полупроводнике. Уровень Ферми будет лежать почти в середине запрещенной зоны, как и в собственном полупроводнике:

$$F = \frac{E_d + E_a}{2} \quad (T=0). \quad (33.1)$$

Такой полупроводник носит название *скомпенсированного*, поскольку происходит полная взаимная компенсация примесей, которые не являются поставщиками свободных электронов и дырок. Из уравнения нейтральности можем записать $n=p$, так как $N_d=N_a$ и $n_d=p_a$. Будучи собственным полупроводником по величине концентрации носителей заряда, скомпенсированный полупроводник в других отношениях ведет себя иначе, и прежде всего это проявляется в различии подвижностей носителей заряда. Это и понятно, поскольку в скомпенсированном полупроводнике нарушения периодичности поля решетки значительно больше, чем в собственном.

Если концентрации примеси не равны, то компенсация будет не полной. Пусть $N_d > N_a$. В таком случае величина $N'_d = N_d - N_a$ будет играть роль примеси одного вида, поскольку часть $N_d - N'_d$ пойдет на компенсацию акцепторной примеси.

Однако имеется некоторое отличие в поведении частично скомпенсированного и не скомпенсированного полупроводников. Для исследования этого различия вернемся к уравнению электронейтральности:

$$n + n_d - p - p_a = N_d - N_a = N'_d. \quad (33.2)$$

При $T \rightarrow 0$ n и p обращаются в нуль, и уравнение электронейтральности будет иметь вид

$$n_d - p_a = N'_d, \quad (33.3)$$

или

$$\frac{N_d}{\frac{1}{2} e^{\frac{E_d - F}{kT}} + 1} - \frac{N_a}{\frac{1}{2} e^{\frac{F - E_a}{kT}} + 1} = N_d - N_a = N'_d. \quad (33.4)$$

Из физических соображений можно предположить, что $F > E_a$, это приводит к $p_a = 0$ при $T = 0$, отсюда

$$n_d = N_d - N_a = N'_d. \quad (33.5)$$

Теперь легко найти положение уровня Ферми для этого предельного случая:

$$\frac{N_d}{\frac{E_d - F}{\frac{1}{2} e^{\frac{E_d - F}{kT}} + 1}} = N'_d, \quad (33.6)$$

или

$$e^{\frac{E_d - F}{kT}} = 2 \frac{N_d - N'_d}{N'_d} = 2 \frac{N_a}{N_d - N_a}, \quad (33.7)$$

откуда

$$F = E_d + kT \ln \frac{N_d - N_a}{2N_a}. \quad (33.8)$$

Если $T = 0$, то $F = E_d$, т. е. уровень Ферми совпадает с донорным уровнем. Для концентрации электронов можем записать

$$n = N_c \frac{N_d - N_a}{2N_a} e^{-\frac{\delta E_a}{kT}}, \quad (33.9)$$

или

$$n = \frac{(N_d - N_a) N_c}{2N_a} e^{-\frac{\delta E_a}{kT}}.$$

Энергия активации δE_a равна энергии ионизации ΔE_d донорной примеси, в то время как в чисто донорном полупроводнике энергия активации равна $\Delta E_d/2$. Формулы (33.8) и (33.9) применимы при самых низких температурах вплоть до $T = 0$. Однако можно пользоваться ими и при $T \neq 0$. В выражениях (33.8—9) N_a должно быть отлично от нуля и не равно N_d : $0 < N_a < N_d$. Если $N_a > N_d > 0$, то можем записать совершенно аналогичное (33.9) выражение:

$$F = E_a - kT \ln \frac{N_a - N_d}{2N_d} \quad (33.10)$$

и

$$p = \frac{(N_a - N_d) N_v}{2N_d} e^{-\frac{\Delta E_a}{kT}}; \quad \delta E_a = \Delta E_a. \quad (33.11)$$

Выражение (33.8) для F показывает, что с ростом температуры уровень Ферми смещается от $F = E_d$ при $T = 0$ вверх или вниз в зависимости от соотношения между N_a и N_d . Если $N_d = 3N_a$, то $F = E_d$ и F не зависит от температуры (конечно, пока справедливы исходные уравнения). При $N_d > 3N_a$ уровень Ферми поднимается

вверх и тем быстрее, чем меньше N_a . При $N_d < 3N_a$ уровень Ферми с ростом температуры опускается, как это представлено на рис. 55.

Зная предельное выражение для F в частично скомпенсированном полупроводнике, можно получить более общее выражение, справедливое в большем интервале температур, если исходить из следующего уравнения электронейтральности:

$$n + n_d = N_d - N_a = N'_d; \quad n = N'_d - n_d = N'_d{}^+ \quad (33.12)$$

Учитывая решение этого уравнения для малых температур, можем записать

$$F = \frac{E_c + E_d}{2} + \frac{kT}{2} \ln \frac{N_d}{2N_c} = \frac{E_c + E_d}{2} + \frac{kT}{2} \ln \frac{N_d - N_a}{2N_c} \quad (33.13)$$

$$n = \sqrt{\frac{(N_d - N_a) N_c}{2}} e^{-\frac{\Delta E_d}{2kT}} \quad (33.14)$$

В отличие от уравнений (33.8), (33.9), справедливых и при $T=0$, уравнения (33.13), (33.14) при $T=0$ не применимы, так как n сравнимо с p_a , $N_c \approx 0$ при $N_a \neq 0$. Нижний предел применимости этой формулы можно установить соотношением

$$n = p_a; \quad (p \ll p_a) \text{ при } F \approx E_d. \quad (33.15)$$

Отсюда легко получить выражение для T ; при котором (33.8) заменяется на (33.13):

$$kT = \frac{\Delta E_0 - \Delta E_d}{\ln N_a - \ln 2N_c} \approx \frac{\Delta E_0}{\ln N_a} \quad (33.16)$$

Мы видим, что чем больше N_d , тем больше температурный интервал $0 - T$, в котором ограничено действие формул (33.15) и в котором применимо выражение (33.8).

Таким образом, наличие компенсирующей примеси приводит к более быстрому изменению положения уровня Ферми.

Рассмотрим в общем виде ход уровня Ферми в области низких температур в зависимости от соотношений концентрации N_d и N_a . Особый интерес будет представлять случай, когда $N_a \approx N_d$.

Уравнение электронейтральности

$$n + n_d - p - p_a = N_d - N_a \quad (33.17)$$

для низких температур можно упростить, опустив в нем n и p . Это будет тем более допустимо, чем меньше T и меньше разность

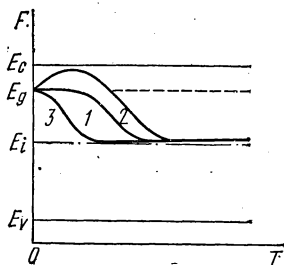


Рис. 55. Температурная зависимость уровня Ферми в полупроводнике, имеющем донорную и акцепторную примесь: 1 — $N_d = 3N_a$; 2 — $N_d > 3N_a$; 3 — $N_d < 3N_a$

$N_d - N_a$. Случай $N_d \gg N_a$ уже был разобран. Уравнение (33.17) можно в таком случае записать в виде

$$N_d - n_d = N_d^+ = p_d = N_a - p_a = N_a^- = n_a, \quad (33.18)$$

или

$$\frac{N_d}{e^{\frac{F-E_d}{kT}} + 1} = \frac{N_a}{e^{\frac{E_a-F}{kT}} + 1}. \quad (33.19)$$

Введем обозначения

$$e^{\frac{F}{kT}} = x; \quad \frac{1}{2} e^{\frac{E_d}{kT}} = D; \quad 2e^{\frac{E_a}{kT}} = A; \quad \frac{N_d}{N_a} = r; \quad r - 1 = s, \quad (33.20)$$

после чего запишем уравнение (33.19) в новых обозначениях:

$$r \left(\frac{A}{x} + 1 \right) = \left(\frac{x}{D} + 1 \right), \quad (33.21)$$

или

$$x^2 - sDx - rAD = 0, \quad (33.22)$$

откуда

$$x = \frac{sD}{2} \left(1 \pm \sqrt{1 + \frac{4rA}{s^2D}} \right) = \frac{D}{2} \left(s \pm \sqrt{s^2 + 4r \frac{A}{D}} \right). \quad (33.23)$$

Так как F — действительная величина, то $x > 0$, поэтому необходимо брать оба решения в зависимости от знака s .

При $s > 0$, т. е.

$$r = \frac{N_d}{N_a} > 1, \quad \text{или} \quad N_d > N_a,$$

$$x = \frac{sD}{2} \left(1 + \sqrt{1 + \frac{4rA}{s^2D}} \right); \quad (33.24)$$

при $s < 0$, т. е.

$$r = \frac{N_d}{N_a} < 1, \quad \text{или} \quad N_d < N_a,$$

$$x = \frac{sD}{2} \left(1 - \sqrt{1 + \frac{4rA}{s^2D}} \right); \quad (33.25)$$

при $s = 0$, т. е.

$$r = 1, \quad \text{или} \quad N_d = N_a,$$

$$x = \sqrt{AD} = \left(2e^{\frac{E_a}{kT}} \cdot \frac{1}{2} e^{\frac{E_d}{kT}} \right)^{1/2} = e^{\frac{E_a + E_d}{2kT}} = e^{\frac{F}{kT}}. \quad (33.26)$$

В последнем случае выражение для F имеет вид

$$F = \frac{E_d + E_a}{2}, \quad (33.27)$$

что было записано ранее, на основе наглядных представлений.

В случае положительных значений s из (33.24) следует

$$F = kT \ln \left\{ \frac{sD}{2} \left(1 + \sqrt{1 + \frac{4rA}{s^2D}} \right) \right\}. \quad (33.28)$$

Учитывая, что $\frac{A}{D} = 4e^{-\frac{E_d - E_a}{kT}} \rightarrow 0$ при $T \rightarrow 0$, получим:

для больших s при $4rA/s^2D \ll 1$

$$F = kT \ln \frac{sD}{2} \cdot 2 = kT \ln \left\{ \frac{N_d - N_a}{N_a} \cdot \frac{1}{2} e^{\frac{E_d}{kT}} \right\} = E_d + kT \ln \frac{N_d - N_a}{2N_a}; \quad (33.29)$$

для малых s при $\frac{4rA}{s^2D} \gg 1$

$$F = \frac{E_a + E_d}{2}. \quad (33.30)$$

Если $s < 0$, то согласно (33.25)

$$F = kT \ln \left\{ \frac{|s|D}{2} \left[\sqrt{1 + \frac{4rA}{s^2D}} - 1 \right] \right\}; \quad (33.31)$$

для больших $|s|$ при $\frac{4rA}{s^2D} \ll 1$

$$\begin{aligned} F &= kT \ln \left\{ \frac{|s|D}{2} \left[1 + \frac{2rA}{s^2D} + \dots - 1 \right] \right\} \approx \\ &\approx kT \ln \frac{rA}{|s|} = E_a + kT \ln \frac{N_a - N_d}{2N_d}, \end{aligned} \quad (33.32)$$

что находится в полном соответствии с (33.29) для $s > 0$;
для малых $|s|$ при $4rA/s^2D \gg 1$

$$F = kT \ln \left\{ \frac{|s|D}{2} \sqrt{\frac{4rA}{s^2D}} \right\} = kT \ln \sqrt{AD} = \frac{E_a + E_d}{2} \quad (33.33)$$

в полном соответствии с (33.30).

Если рассматривать ход уровня Ферми как функцию s при неизменной температуре, то выражения (33.28) и (33.31) полностью описывают его.

Для концентрации электронов и дырок из (33.29) и (33.32) следует:

$$n = \frac{(N_d - N_a) N_c}{2N_a} e^{-\frac{\Delta E_d}{kT}}, \quad (33.34)$$

$$p = \frac{(N_a - N_d) N_v}{2N_d} e^{-\frac{\Delta E_d}{kT}}. \quad (33.35)$$

Тангенс угла наклона графика $\ln n$ как функции обратной температуры определяется величиной $\Delta E_d/k$ и $\Delta E_a/k$, что в два раза больше тангенса угла наклона в случае, когда в полупроводнике содержится примесь только одного вида.

1. Меняя концентрацию донорной N_d и акцепторной N_a примеси, можно в широких пределах менять концентрацию свободных электронов n и дырок p . Произведение же концентрации n и p не зависит от вида и количества примесей, пока полупроводник остается невырожденным, при этом

$$np = n_i^2. \quad (33.1p)$$

2. Полупроводник, в котором $N_d = N_a$, называется скомпенсированным. Уровень Ферми в нем лежит вблизи середины запрещенной зоны.

3. Если концентрации примеси несколько различны, то уровень Ферми совпадает при $T=0$ с уровнем той примеси, которая содержится в большом количестве. При повышении температуры происходит переход к собственному полупроводнику при тем меньшей температуре, чем ближе друг к другу концентрации примеси.

4. Если концентрации N_d и N_a различаются значительно, то полупроводник ведет себя как полупроводник с одним типом примеси.

§ 34. ВЫРОЖДЕННЫЙ ПОЛУПРОВОДНИК

Мы определили в § 29 вырожденный полупроводник, в котором уровень Ферми лежит внутри зоны энергии на расстоянии не менее $5kT$ от ее границы. Концентрация свободных носителей заряда при этом не зависит от температуры. Покажем теперь, что вырождение наступает в результате сильного легирования полупроводника.

Как было показано выше, с ростом температуры уровень Ферми в примесном полупроводнике приближается к зоне энергии. Найдем положение максимума F при изменении температуры. Из (32.14) получим

$$\frac{dF}{dT} = \frac{k}{2} \ln \frac{N_d}{2N_c} - \frac{kT}{2} \frac{N_d}{2N_c} \frac{2}{N_d} \frac{dN_c}{dT} = 0, \quad (34.1)$$

или

$$\ln \frac{N_d}{2N_c} - \frac{T}{N_c} \frac{dN_c}{dT} = 0. \quad (34.2)$$

Но так как

$$\frac{dN_c}{dT} = \frac{3}{2} \frac{N_c}{T}, \quad (34.3)$$

то условие экстремума (34.1) или (34.2) примет вид

$$\ln \frac{N_d}{2N_c} = \frac{3}{2}; \quad \frac{N_d}{2N_c} = e^{3/2}, \quad (34.4)$$

т. е. температура T_{\max} , при которой F достигает максимального значения $F = F_{\max}$, определяется из условия

$$N_c(T_{\max}) = \frac{N_d}{2e^{3/2}}. \quad (34.5)$$

Учитывая (29.22) для N_c , получим следующее выражение для T_{\max} :

$$T_{\max} = 8,15 \left(\frac{m}{m_d^*} \right) \left(\frac{N_d}{10^{18}} \right)^{2/3}. \quad (34.6)$$

Выражение (34.6) показывает, что температура T_{\max} *возрастает с ростом концентрации примеси как $N_d^{2/3}$* . Найдем величину F_{\max} :

$$F_{\max} = \frac{E_c + E_d}{2} + \frac{kT_{\max}}{2} \ln \frac{N_d}{2N_c(T_{\max})} = \frac{E_c + E_d}{2} + \frac{3}{4} kT_{\max}, \quad (34.7)$$

или

$$F_{\max} = \frac{E_c + E_d}{2} + 5,3 \cdot 10^{-4} \left(\frac{m}{m_d^*} \right) \left(\frac{N_d}{10^{18}} \right)^{2/3}. \quad (34.8)$$

В (34.8) и ниже k выражена в электронвольтах. Концентрация $N_d^{(кр)}$, при которой F совпадает с дном зоны проводимости, называется критической. Она может быть найдена из условия

$$F_{\max} = E_c = \frac{E_c + E_d}{2} + 5,3 \cdot 10^{-4} \left(\frac{m}{m_d^*} \right) \left(\frac{N_d^{(кр)}}{10^{18}} \right)^{2/3}, \quad (34.9)$$

или

$$N_d^{(кр)} (\text{см}^{-3}) = 10^{22,5} \left(\frac{m_d^*}{m} \right)^{3/2} [\Delta E_d (\text{эВ})]^{3/2}. \quad (34.10)$$

Оценим критическую концентрацию N_d^* при $m_d^* = m$ и $\Delta E_d = 0,03$ эВ. Подставляя значения m_d^* и ΔE_d в (34.10), получим: $N_d^{(кр)} = 1,6 \cdot 10^{20} \text{ см}^{-3}$. Если принять для m_d^*/m величину 0,3, то $N_d^{(кр)}$ станет равной $2,5 \cdot 10^{19} \text{ см}^{-3}$.

Таким образом, критическая концентрация очень чувствительна к энергии ионизации примеси и величине эффективной массы. В соединениях $A^{III}B^V$ критическая концентрация может иметь значения много меньше 10^{19} см^{-3} . Действительно, если положить $m_d^*/m = 10^{-3}$ и $\Delta E_d = 0,0001$ эВ, то для $N_d^{(кр)}$ получим значение $N_d^{(кр)} = 10^{12} \text{ см}^{-3}$, что наблюдается для антимонида индия. Критическая концентрация $N_d^{(кр)}$ позволяет оценить концентрацию, необходимую для начала вырождения полупроводника, поскольку при $F = E_c$ полупроводник перестает быть невырожденным, однако это еще не есть вырождение в том смысле, что концентрация носителей заряда не зависит от температуры в каком-то интервале. Для этого необходима большая концентрация примеси. В принципе можно оценить положение уровня Ферми и концентрацию носителей заряда, исходя из уравнения нейтральности, в котором при $N_a = 0$ можно отбросить p и p_a ; получим $n + n_d = N_d$ при $n = p_d = N_d^+$. Подставляя выражения для n и p_d , запишем

$$\frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi) = \frac{N_d}{2e \frac{F - E_d}{kT} + 1}. \quad (34.11)$$

Если полупроводник полностью вырожден, то можем пренебречь единицей по сравнению с экспонентой в выражении для ρ_d , и подставляя вместо $\Phi_{1/2}(\xi)$ его выражение, запишем:

$$\frac{4N_c}{3\sqrt{\pi}} \left(\frac{F - E_c}{kT} \right)^{3/2} = \frac{N_d}{2} e^{\frac{F - E_d}{kT}}. \quad (34.12)$$

Из (34.12) можно записать следующее равенство:

$$\xi^{3/2} e^{\xi} = \left(\frac{3\sqrt{\pi}}{8} \frac{e^{\frac{\Delta E_d}{kT}}}{N_c} \right) N_d. \quad (34.13)$$

Определив из (34.13) ξ как функцию N_d , можно найти n .

Однако необходимо заметить, что соотношения (34.12) и (34.13) имеют мало смысла, поскольку при столь больших концентрациях, которые необходимы для вырождения, примесной уровень превращается в зону, накладывающуюся на зону проводимости. При этом примесная зона оказывается не заполненной. Это приводит к тому, что вырождение не снимается и при очень низких температурах, поскольку остается механизм проводимости посредством примесной зоны. Благодаря слиянию зон вырождение наблюдается в широком интервале температур, например, у некоторых интерметаллов от комнатных до температур жидкого водорода. Кроме того, необходимо отметить, что в силу образования примесной зоны энергия ионизации примеси с ростом ее концентрации уменьшается и согласно (34.10) необходимая для вырождения концентрация в свою очередь уменьшается.

Для полностью вырожденного полупроводника бывает важно вычислить положение уровня Ферми по известной концентрации носителей заряда:

$$F - E_c = (h^2/2m_{nd}^*) (3/8\pi)^{2/3} n^{2/3}, \quad (34.14)$$

$$E_v - F = (h^2/2m_{pd}^*) (3/8\pi)^{2/3} p^{2/3}. \quad (34.15)$$

Концентрация же электронов или дырок может быть определена экспериментально.

Вырожденные полупроводники используют для изготовления таких приборов, как туннельные диоды и полупроводниковые квантовые генераторы. Кроме того, они имеют важное теоретическое значение. Сильнолегированные полупроводники трудно анализировать теоретически. Действительно, большая концентрация электронов приводит к сильному взаимодействию электронов с ионом донора, к его экранированию. Экранировка поля донора приводит к уменьшению энергии ионизации вплоть до нуля, поэтому говорить о размытии уровней примеси в этом случае не имеет смысла. Для исследования сильнолегированных полупроводников необходимо использовать такие методы, как метод функции Грина.

Отметим в заключение, что в сильнолегированном полупроводнике плотность состояния в зоне сильно искажается примесью. Это иска-

жение сохраняется и при уменьшении концентрации примеси. Оно приводит к тому, что исчезает резкая граница зоны энергии, про-является «хвост» у плотности состояния, экспоненциально затухаю-щий и распространяющийся вплоть до другой зоны.

Резюме § 34

1. Увеличение концентрации примеси приводит к повышению концентрации носителей заряда, что вызывает уменьшение расстоя-ния между зоной энергии и уровнем Ферми. При некоторой кон-центрации примеси $N^{(кр)}$ уровень Ферми совпадает с экстремумом энергии в зоне. Концентрацию $N^{(кр)}$ называют критической. Если концентрация примеси превосходит $N^{(кр)}$, то полупроводник стано-вится частично или полностью вырожденным.

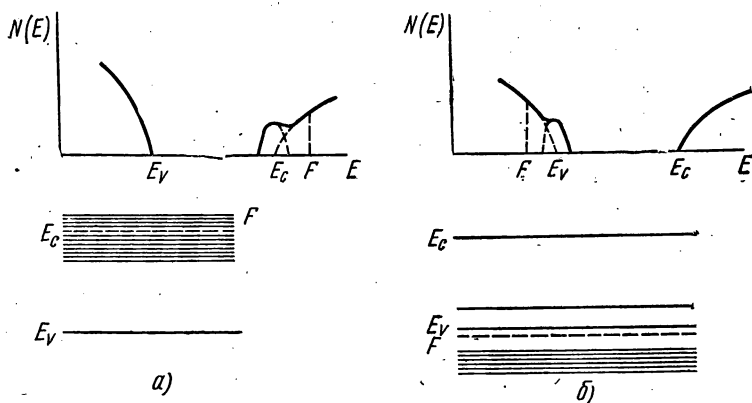


Рис. 56. Плотность состояний и положение уровня Ферми в вы-рожденном (а) и дырочном (б) полупроводниках

2. Нижней температурой вырождения T_B^H называют температуру, при которой уровень Ферми с ростом температуры от $T=0$ входит в зону энергии. При $T < T_B^H$ происходит «вымораживание» носителей заряда из зоны энергии на уровень примеси.

3. Верхней температурой вырождения T_B^B электронного газа называют температуру, при которой электронный газ становится классическим, она определяется обычным условием:

$$kT_B^B = F - E_c; \quad kT_B^B = E_v - F. \quad (34.1p)$$

Однако концентрация носителей заряда в этом случае может оставаться постоянной, если температура еще недостаточна для перехода к собственной концентрации.

4. Нижняя температура вырождения может отсутствовать в связи с тем, что с ростом концентрации примеси уровень примеси расши-

рется и превращается в зону примеси. При этом зона примеси может оказаться заполненной лишь частично, вследствие чего носители заряда могут быть свободными при любой температуре, включая $T=0$, если даже между зоной энергии и зоной примеси имеется энергетический зазор (энергетическая щель). При достаточно большой концентрации примеси зона примеси может перекрываться с основной зоной. В любом случае резкая граница зоны энергии размывается при введении примеси в полупроводник (рис. 56).

§ 35. ПЛОТНОСТЬ СОСТОЯНИЙ В МАГНИТНОМ ПОЛЕ

Как было показано в § 23, в сильном магнитном поле спектр энергии электронов резко меняется — появляются уровни Ландау:

$$E = \frac{\hbar^2 \kappa_z^2}{2m^*} + \hbar \omega_0 \left(n + \frac{1}{2} \right); \quad E_c = 0, \quad (35.1)$$

где $n=0, 1, 2, \dots$ — квантовое число, $\omega_0 = \frac{eB}{m^*}$ — циклотронная частота. В этом случае теряет смысл говорить об изоэнергетических поверхностях в зоне Бриллюэна, поскольку E не зависит от κ_x и κ_y . Однако необходимо иметь в виду, что при $B \rightarrow 0$ выражение (35.1) должно перейти в выражение

$$E = \frac{\hbar^2 \kappa_x^2}{2m^*} + \frac{\hbar^2 \kappa_y^2}{2m^*} + \frac{\hbar^2 \kappa_z^2}{2m^*}. \quad (35.2)$$

Другими словами, символически

$$\lim_{B \rightarrow 0} \hbar \omega_0 \left(n + \frac{1}{2} \right) = \frac{\hbar^2 (\kappa_x^2 + \kappa_y^2)}{2m^*}. \quad (35.3)$$

В случае, когда полупроводник находится в магнитном поле, за динамические величины, определяющие состояние электрона, можно взять κ_z , n , ω_0 , s_z .

Найдем выражение для плотности состояний по энергии $N_B(E)$. В интервале энергии E , $E + dE$ содержится

$$2dS = 2N_B(E) dE \quad (35.4)$$

состояний. Очевидно, что полное число состояний должно быть равно $2N_g$, где N — число атомов в кристалле, g — фактор вырождения зоны, 2 — множитель, связанный со спиновым вырождением:

$$2 \int_{E_{\min}}^{E_{\max}} N_B(E) dE = 2Ng. \quad (35.5)$$

Число состояний (по κ_z) в единичном объеме кристалла и в интервале $d\kappa_z$ при фиксированных n и ω_0 равно

$$dS_{\kappa_z} = \frac{d\kappa_z}{2\pi}. \quad (35.6)$$

Из (35.1) найдем связь между dk_z и dE :

$$\kappa_z = \left(\frac{2m^*}{\hbar^2} \right)^{1/2} [E - \hbar\omega_0 (n + 1/2)]^{1/2}, \quad (35.7)$$

$$dk_z = \frac{1}{2} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} [E - \hbar\omega_0 (n + 1/2)]^{-1/2} dE. \quad (35.8)$$

Согласно (35.6) можно записать

$$dS_{\kappa_z} = \frac{dk_z}{2\pi} = \frac{1}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} [E - \hbar\omega_0 (n + 1/2)]^{-1/2} dE. \quad (35.9)$$

Чтобы записать теперь выражение для плотности состояний по энергии, необходимо учесть следующее соображение. При фиксированном интервале энергии E , $E + dE$ существует несколько различных интервалов dk_z , что очевидно из выражения (35.8) и рис. 57.

Действительно, энергии E соответствует ν парабол, величину ν можно найти из соотношений

$$\nu \leq \frac{E - \hbar\omega_0}{2\hbar\omega_0} < \nu + 1. \quad (35.10)$$

Величина n может меняться от нуля до ν , следовательно, интервалу энергии E , $E + dE$ соответствуют ν интервалов dk_z , поэтому для нахождения dS_{κ_z} необходимо просуммировать состояния ν интервалов dk_z :

$$dS_{\kappa_z} = \left\{ \frac{1}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \sum_{n=0}^{\nu} [E - \hbar\omega_0 (n + 1/2)]^{-1/2} \right\} dE. \quad (35.11)$$

Кроме того, поскольку в выражение для энергии входит $\hbar\omega_0 = \frac{e\hbar}{m^*} B$, плотность состояний должна зависеть и от B . Пусть она определяется некоторой функцией $g(B)$, вид которой найдем из предельного перехода $B \rightarrow 0$.

Сравнивая (35.11) и (35.4) и учитывая функцию $g(B)$, можем записать выражение для плотности состояний:

$$N_B(E) = \frac{g(B)}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \sum_{n=0}^{\nu} [E - \hbar\omega_0 (n + 1/2)]^{-1/2} = \sum_{n=0}^{\nu} N_B^{(n)}(E). \quad (35.12)$$

Рассмотрим одно из слагаемых в (35.12):

$$N_B^{(n)}(B) = \frac{g(B)}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} [E - \hbar\omega_0 (n + 1/2)]^{-1/2}. \quad (35.13)$$

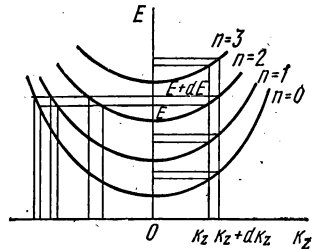


Рис. 57. Связь между интервалом энергии dE и интервалом значений dk_z .

Так как n фиксировано, то при $E \gg \hbar\omega_0(n+1/2)$ плотность $N_B^{(n)}(E)$ мала и ведет себя как $1/\sqrt{E}$.

Поскольку величина $N_B^{(n)}(E)$ должна быть вещественной, $N_B^{(n)}(E)$ определено только для энергии $E > \hbar\omega_0(n+1/2)$, при $E \rightarrow \hbar\omega_0(n+1/2)$ $N_B^{(n)}(E) \rightarrow \infty$. На рис. 58 сплошной и пунктирной линиями представлена графическая зависимость одной из величин $N_B^{(n)}(E)$, соответствующая $n=0$.

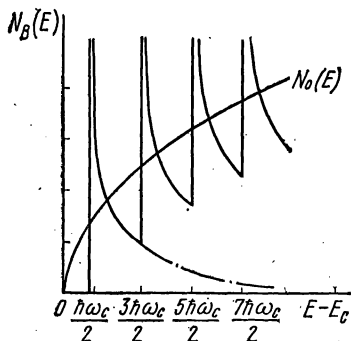


Рис. 58. Плотность состояний в магнитном поле

$\{\hbar\omega_0(n+1/2), E_0\}$ действительно содержит конечное число состояний:

$$\int_{\hbar\omega_0(n+1/2)}^{E_0} N_B^{(n)}(E) dE = \frac{g(B)}{4\pi} \left(\frac{2m^*}{\hbar^2}\right)^{1/2} \int_{\hbar\omega_0(n+1/2)}^{E_0} [E - \hbar\omega_0(n+1/2)]^{-1/2} dE = \frac{g(B)}{2\pi} \left(\frac{2m^*}{\hbar^2}\right)^{1/2} [E_0 - \hbar\omega_0(n+1/2)]^{1/2}. \quad (35.14)$$

Таким образом, в любом конечном интервале энергии $\{\hbar\omega_0(n+1/2); \hbar\omega_0(n+1/2) + \Delta E\}$ содержится конечное число состояний

$$\frac{g(B)}{2\pi} \left(\frac{2m^*}{\hbar^2}\right)^{1/2} (\Delta E)^{1/2}. \quad (35.15)$$

Бесконечно большая плотность состояний соответствует значению $\kappa_z=0$; без магнитного поля в точке $\kappa_z=0$ и $\kappa_x=\kappa_y=0$ $N(E)=0$. (Будем считать для определенности, что в точке $\mathbf{k}=0$ лежит экстремум энергии.)

Если нарисовать графически $N_B(E)$, то ход $N_B(E)$ в окрестности n -ых особых точек определяется только членом $N_B^{(n)}(E)$, вдали от них необходимо учитывать сумму ν гипербол. Очевидно, что чем больше E , тем больше ν и тем больше будет $N_B(E)$ для значений E внутри интервала между особыми точками. Это наглядно представлено на рис. 58, который показывает, что с ростом E величина $N_B(E)$ слева от каждой особой точки находится тем выше, чем больше ν . Если $B \rightarrow 0$, то число ν для любого конечного значения стремится к бесконечности. Однако стремление ν к бесконечности происходит

таким образом, что $\lim_{\nu \rightarrow \infty} \nu \hbar \omega_0 = \lim_{B \rightarrow 0} \nu \hbar \omega_0 = E$ для произвольного значения энергии. При $B \rightarrow 0$

$$N_B(E) \rightarrow N_0(E) = 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2} = N(E). \quad (35.16)$$

В таком случае можем найти предельное значение $N_0(E)$, заменив сумму по n интегралом:

$$\begin{aligned} \sum_{n=0}^{\nu} [E - \hbar \omega_0 (n + 1/2)]^{-1/2} &\rightarrow \int_0^{\nu} [E - \hbar \omega_0 (n + 1/2)]^{-1/2} dn = \\ &= -\frac{1}{\hbar \omega_0} \int_{E - \frac{\hbar \omega_0}{2}}^0 x^{-1/2} dx = \frac{2}{\hbar \omega_0} \left(E - \frac{\hbar \omega_0}{2} \right)^{1/2}. \end{aligned} \quad (35.17)$$

При малом поле $\sum_{n=0}^{\nu} [E - \hbar \omega_0 (n + 1/2)]^{-1/2}$ переходит в $\frac{2}{\hbar \omega_0} \times \left(E - \frac{\hbar \omega_0}{2} \right)^{1/2}$. При $B \rightarrow 0$, $\omega_0 \rightarrow 0$ и $\left(E - \frac{\hbar \omega_0}{2} \right)^{1/2} \rightarrow E^{1/2}$, однако $\frac{1}{\hbar \omega_0} \rightarrow \infty$. Учтем теперь функцию $g(B)$. Если положим

$$g(B) = G \hbar \omega_0, \quad (35.18)$$

то расходимость суммы (35.17) будет устранена. Найдем G :

$$\begin{aligned} N_B(E) &= \frac{G \hbar \omega_0}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \sum_{n=0}^{\nu} [E - \hbar \omega_0 (n + 1/2)]^{-1/2} \rightarrow \\ &\rightarrow \frac{G \hbar \omega_0}{4\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \cdot \frac{2}{\hbar \omega_0} \left(E - \frac{\hbar \omega_0}{2} \right)^{1/2} \rightarrow N_0(E) = 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} E^{1/2}. \end{aligned} \quad (35.19)$$

Сравнивая предельные выражения $N_B(E)$ и $N(E)$, можем записать:

$$\frac{G}{2\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} = 2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2}. \quad (35.20)$$

Из (35.20) и (35.18) следует

$$G = \frac{4\pi m^*}{\hbar^2}, \quad (35.21)$$

$$g(B) = G \hbar \omega_0 = \frac{2m^*}{\hbar} \omega_0 = \frac{2eB}{\hbar}, \quad (35.22)$$

$$N_B(E) = \frac{\hbar \omega_0}{2} \left[2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \right] \sum_{n=0}^{\nu} [E - \hbar \omega_0 (n + 1/2)]^{-1/2}. \quad (35.23)$$

Выражение (35.23) представляет собой плотность состояний при наложении на кристалл магнитного поля. Выводя его, мы не учитывали собственный магнитный момент электронов $\frac{e}{m^*} \mathbf{S} = \boldsymbol{\mu}^*$, благодаря которому электрон обладает дополнительной энергией

$$-(\boldsymbol{\mu}^* \mathbf{B}) = -\left(\frac{e\hbar}{m^*} \mathbf{S}_z\right) = -(\mathbf{S}\omega_0) = \pm \mu_0^* B = \pm \frac{\hbar\omega_0}{2}, \quad (35.24)$$

где

$$\mu_0^* = \frac{e\hbar}{2m^*} = \frac{m}{m^*} \mu_0 \quad (35.25)$$

есть магнетон Бора электрона с эффективной массой m^* , а $\mu_0 = \frac{e\hbar}{2m}$ — магнетон Бора; \mathbf{S} — вектор спина, проекция которого на \mathbf{B} равна $\pm \frac{\hbar}{2}$. Пренебрегая расщеплением уровней энергии, мы удваиваем каждое энергетическое состояние.

Обозначая энергию Ферми $F(B)$, запишем функцию Ферми — Дирака:

$$f = \frac{1}{e^{\frac{E-F(B)}{kT}} + 1}. \quad (35.26)$$

Для концентрации электронов n_n получим выражение

$$\begin{aligned} n_n &= 2 \int_0^\infty N_B(E) f(E, T) dE = \\ &= \hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \right] \int_0^\infty \sum_{n=0}^{\nu} [E - \hbar\omega_0(n + 1/2)]^{-1/2} \times \left[e^{\frac{E-F(B)}{kT}} + 1 \right]^{-1} dE. \end{aligned} \quad (35.27)$$

Вычислим это выражение для двух предельных случаев — вырожденного и невырожденного полупроводника.

В случае *вырожденного полупроводника* функцию распределения заменяем прямоугольной ступенькой, и выражение для n_n примет вид

$$\begin{aligned} n_n &= \hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \right] \int_0^{F(B)} \sum_{n=0}^{\nu} [E - \hbar\omega_0(n + 1/2)]^{-1/2} dE = \\ &= \hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \right] \sum_{n=0}^{\nu} \int_{\hbar\omega_0(n+1/2)}^{F(B)} [E - \hbar\omega_0(n + 1/2)]^{-1/2} dE = \\ &= 2\hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \right] \sum_{n=0}^{\nu} [F(B) - \hbar\omega_0(n + 1/2)]^{1/2}. \end{aligned} \quad (35.28)$$

В слабом магнитном поле сумму можно заменить на интеграл:

$$\begin{aligned} \hbar\omega_0 \sum_{n=0}^{\nu} [F(B) - \hbar\omega_0(n+1/2)]^{1/2} &\approx \\ \approx \hbar\omega_0 \int_0^{\nu} [F(B) - \hbar\omega_0(n+1/2)]^{1/2} dn &\approx \frac{2}{3} F^{3/2}(B), \end{aligned} \quad (35.29)$$

$$n_n = \frac{8\pi}{3} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} F^{3/2}(B). \quad (35.30)$$

Из (35.30) запишем выражение для $F(B)$:

$$F(B) = \frac{\hbar^2}{2m^*} \left(\frac{3n_n}{8\pi} \right)^{2/3} = F_0. \quad (35.31)$$

В сильном магнитном поле величина ν будет малой. Рассмотрим такое поле, когда остается один невозбужденный уровень $n=0$ (квантовый предел), тогда от суммы в (35.28) остается один член:

$$n_n = 2\hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \right] \left[F(B) - \frac{\hbar\omega_0}{2} \right]^{1/2}, \quad (35.32)$$

откуда

$$F(B) = \frac{\hbar\omega_0}{2} + \frac{n_n^2}{16\pi^2 \left(\frac{2m^*}{\hbar^2} \right) (\hbar\omega_0)^2} = \frac{\hbar\omega_0}{2} \left[1 + \frac{1}{18} \left(\frac{F_0}{\frac{\hbar\omega_0}{2}} \right)^3 \right]. \quad (35.33)$$

где F_0 определяется выражением (35.31).

Рассмотрим невырожденный полупроводник, для него концентрация электронов дается выражением

$$n_n = \hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \right] e^{\frac{F(B)}{kT}} \sum_{n=0}^{\nu} \int_{\hbar\omega_0(n+1/2)}^{\infty} [E - \hbar\omega_0(n+1/2)]^{-1/2} e^{-\frac{E}{kT}} dE. \quad (35.34)$$

Особенность подынтегральной функции в точке $E = \hbar\omega_0(n+1/2)$ носит δ -образный характер и не приводит к расходимости интеграла, что легко проверяется непосредственным вычислением. Действительно, рассмотрим интеграл

$$\begin{aligned} \int_a^{\infty} (x-a)^{-1/2} e^{-\frac{x}{b}} dx &= 2(x-a)^{1/2} e^{-\frac{x}{b}} \Big|_a^{\infty} + \frac{2}{b} \int_a^{\infty} (x-a)^{1/2} e^{-\frac{x}{b}} dx = \\ &= \frac{2}{b} b^{3/2} e^{-\frac{a}{b}} \int_a^{\infty} \frac{(x-a)^{1/2}}{\sqrt{b}} e^{-\frac{x-a}{b}} d\left(\frac{x}{b}\right) = \\ &= 2b^{1/2} e^{-\frac{a}{b}} \int_0^{\infty} y^{1/2} e^{-y} dy = 2b^{1/2} e^{-\frac{a}{b}} \Gamma(3/2) = \sqrt{\pi b} e^{-\frac{a}{b}}. \end{aligned} \quad (35.35)$$

Из (35.35) видим, что в результате интегрирования мы получили значение функции $e^{-\frac{x}{b}}$ в точке $x=a$, т. е. $N_B^{(n)}(E)$ действительно ведет себя подобно δ -функции. Запишем выражение для n_n , учитывая, что $b=kT$; $a=\hbar\omega_0(n+1/2)$:

$$n_n = \hbar\omega_0 \left[2\pi \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \right] \sqrt{\pi kT} e^{\frac{F(B)}{kT}} \sum_{n=0}^{\infty} e^{-\frac{\hbar\omega_0(n+1/2)}{kT}}. \quad (35.36)$$

Но

$$\sum_{n=0}^{\infty} e^{-\frac{\hbar\omega_0}{kT} n} = \frac{1}{1 - e^{-\frac{\hbar\omega_0}{kT}}}, \quad (35.37)$$

так как $(e^{-\frac{\hbar\omega_0}{kT}})^n$ образует геометрическую прогрессию. Если учтем, что

$$\frac{e^{-\frac{\hbar\omega_0}{2kT}}}{1 - e^{-\frac{\hbar\omega_0}{kT}}} = \frac{1}{e^{\frac{\hbar\omega_0}{2kT}} - e^{-\frac{\hbar\omega_0}{2kT}}} = \left[2 \operatorname{sh} \left(\frac{\hbar\omega_0}{2kT} \right) \right]^{-1}, \quad (35.38)$$

то окончательно для n можем записать

$$n_n = \left(\frac{\hbar\omega_0}{2kT} \right) N_c \frac{e^{\frac{F(B)}{kT}}}{\operatorname{sh} \left(\frac{\hbar\omega_0}{2kT} \right)}. \quad (35.39)$$

При известной концентрации n_n можно найти выражение для уровня Ферми $F(B)$:

$$F(B) = kT \ln \left\{ \frac{n_n \operatorname{sh} \left(\frac{\hbar\omega_0}{2kT} \right)}{N_c \left(\frac{\hbar\omega_0}{2kT} \right)} \right\}. \quad (35.40)$$

В слабом магнитном поле

$$\operatorname{sh} \left(\frac{\hbar\omega_0}{2kT} \right) \approx \left(\frac{\hbar\omega_0}{2kT} \right), \quad (35.41)$$

$$F = kT \ln \frac{n_n}{N_c}. \quad (35.42)$$

Если учесть, что F в (35.42) отсчитывается от E_c (или $E_c=0$), то (35.42) примет вид

$$n_n = N_c e^{\frac{E_c - F}{kT}}, \quad (35.43)$$

что в точности совпадает с выражением (29.11р).

В сильном магнитном поле

$$\operatorname{sh}\left(\frac{\hbar\omega_0}{2kT}\right) \approx e^{\frac{\hbar\omega_0}{2kT}}, \quad (35.44)$$

$$n_n = \left(\frac{\hbar\omega_0}{2kT}\right) N_c e^{\frac{F(B) - \frac{\hbar\omega_0}{2}}{kT}}, \quad (35.45)$$

или

$$n_n = \left(\frac{\hbar\omega_0}{2kT}\right) N_c e^{-\frac{E_c + \frac{\hbar\omega_0}{2} - F(B)}{kT}}. \quad (35.46)$$

Величина

$$E_c + \frac{\hbar\omega_0}{2} = E_c(B) \quad (35.47)$$

представляет собой наименьший ($n=0$) уровень энергии при наложении магнитного поля, поэтому можно сказать, что магнитное поле способствует снятию вырождения. Оно приводит к увеличению ширины запрещенной зоны.

Резюме § 35

1. В магнитном поле плотность состояний существенно видоизменяется. Она обращается в бесконечность в «вершинах» парабол Ландау и уменьшается с ростом энергии для каждой параболы Ландау как $\frac{1}{\sqrt{E}}$. Именно это свойство плотности состояний в магнитном поле позволяет говорить о дискретных уровнях Ландау.

2. Выражение для концентрации носителей заряда и для уровня Ферми в слабом магнитном поле не отличается от соответствующих выражений в полупроводнике без магнитного поля. В сильном магнитном поле соответствующие выражения существенно видоизменяются.

3. Рассмотренная в этой главе концентрация носителей заряда обусловлена тепловой генерацией, т. е. носители заряда возникают за счет энергии тепловых колебаний решетки, с которой электронный (или дырочный) газ находится в термодинамическом равновесии, поэтому эти носители заряда называют равновесными, а соответствующую им концентрацию называют равновесной концентрацией носителей заряда.

КИНЕТИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

§ 36. КИНЕТИЧЕСКОЕ УРАВНЕНИЕ БОЛЬЦМАНА

Физические явления, обусловленные движением носителей заряда под действием внешних и внутренних полей или разности температур, называют кинетическими явлениями, или явлениями переноса. К ним относятся электропроводность и теплопроводность, гальваномагнитные, термомагнитные и термоэлектрические явления. Кинетические явления лежат в основе фотоэлектрических и фотомагнитных эффектов. Для качественного описания кинетических явлений достаточно использовать представление о движении частиц под действием внешних сил. Однако для их количественного описания модельных представлений о движении частиц недостаточно. Чтобы получить правильные выражения для величин, характеризующих кинетические явления, необходимо использовать более общие методы их описания, учитывающие различную роль носителей заряда, находящихся в различных состояниях. Таким, достаточно мощным, методом теоретического исследования кинетических эффектов является метод кинетического уравнения Больцмана, которое описывает изменение состояния частиц в результате действия различных факторов.

В идеальном кристалле состояние $\psi_k(\mathbf{r})$ остается неизменным в течение сколь угодно большого времени.

Тем самым функция распределения электронов по состояниям $f(\mathbf{r}, \mathbf{k})$ также остается неизменной. Если наложить на систему электронов в идеальном кристалле внешнее силовое поле $V(\mathbf{r})$, то состояние каждого электрона в зоне Бриллюэна изменяется в соответствии с уравнением

$$\frac{d\mathbf{P}}{dt} = -\nabla V; \quad \frac{d\mathbf{k}}{dt} = \frac{1}{\hbar} \mathbf{F}_a, \quad (36.1)$$

или

$$\mathbf{P}(t) = \int_0^t \mathbf{F}_a(\xi) d\xi + \mathbf{P}_0(0). \quad (36.2)$$

Изменение квазимпульса за время t не зависит от \mathbf{P}_0 , а определяется только импульсом силы

$$\Delta\mathbf{P} = \mathbf{P}(t) - \mathbf{P}_0(0) = \int_0^t \mathbf{F}_a(\xi) d\xi. \quad (36.3)$$

Соотношение (36.3) показывает, что если в некоторый момент $t=0$ электроны занимали состояния, определяемые функцией $f(\mathbf{r}, \mathbf{k}) = f\left(\mathbf{r}, \frac{\mathbf{p}_0}{\hbar}\right)$, то в момент времени t распределение электронов по состояниям должно определяться другой функцией, имеющей вид

$$f\left[\mathbf{r}, \mathbf{k}(t) - \frac{1}{\hbar} \int_0^t \mathbf{F}_a(\xi) d\xi\right] = \tilde{f}(\mathbf{r}, \mathbf{k}). \quad (36.4)$$

Другими словами, внешние (по отношению к периодическому полю) силы приводят к изменению функции распределения электронов по состояниям.

Запишем уравнение, определяющее изменение функции распределения во времени, т. е. запишем ее полную производную по времени:

$$\frac{df(\mathbf{r}, \mathbf{k}, t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{r}} \cdot \frac{d\mathbf{r}}{dt} + \frac{\partial f}{\partial \mathbf{k}} \cdot \frac{d\mathbf{k}}{dt} = \frac{\partial f}{\partial t} + (\nabla_{\mathbf{r}} f \cdot \mathbf{v}) + \frac{1}{\hbar} (\nabla_{\mathbf{k}} f \mathbf{F}_a). \quad (36.5)$$

Мы учли (36.1) и то, что $d\mathbf{r}/dt = \mathbf{v}$ — скорость электрона. В соответствии с теоремой Лиувилля о неизменности фазового объема при движении системы вдоль фазовых траекторий или учитывая сохранение числа состояний, можем записать

$$df/dt = 0, \quad (36.6)$$

или

$$-\frac{\partial f}{\partial t} = (\nabla_{\mathbf{r}} f \cdot \mathbf{v}) + (\nabla_{\mathbf{k}} f \frac{\mathbf{F}_a}{\hbar}). \quad (36.7)$$

Уравнение (36.7) показывает, что изменение функции распределения со временем в каждой точке фазового пространства (\mathbf{r}, \mathbf{k}) обусловлено движением частиц в обычном пространстве и в пространстве волнового вектора.

Сила \mathbf{F}_a определяется как внешними микроскопическими полями, так и любыми нарушениями идеальности поля решетки — вакансиями, атомами и ионами примеси, тепловыми колебаниями решетки.

Для большинства практически важных случаев необходимо знать поведение твердого тела в результате воздействия внешних макроскопических полей. В таком случае все силы, обусловленные любыми локальными нарушениями периодического поля, являются внутренними для данного кристалла, они должны быть выделены в особый класс сил. Разделим силы \mathbf{F}_a на два класса — силы, порождаемые внешними макроскопическими полями, которые обозначим через \mathbf{F} , и силы, обусловленные локальными нарушениями периодичности поля решетки, для которых используем обозначение \mathbf{F}_d . Действия внешних \mathbf{F} и внутренних \mathbf{F}_d сил противоположны. Внешние силы \mathbf{F} приводят к возникновению направленного движения частиц в пространстве квазиимпульса и в пространстве координат. Действительно, для каждой частицы

$$\Delta \mathbf{P} = \int_0^t \mathbf{F}(\xi) d\xi = \mathbf{F}t, \quad (36.8)$$

если F не зависит от времени. Можем записать точно такое же выражение и для F_d :

$$\Delta P = \int_0^t F_d(\xi) d\xi. \quad (36.9)$$

Но F_d в каждой точке есть результирующая огромного числа локальных полей, имеющая очень резкую и сложную зависимость от координаты электрона или дырки. Это легко понять из рассмотрения действия полей двух ионов на три электрона, имеющих одинаковый первоначальный квазиимпульс. Из рис. 59 видно, что изменения квазиимпульсов трех электронов совершенно различны, поскольку силы, действующие на электроны, зависят от их положения. Поэтому *учесть действие внутренних сил F_d на функцию распределения на основании динамических закономерностей невозможно, их необходимо заменить статистическими закономерностями.*

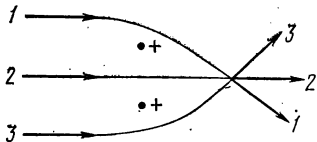


Рис. 59. Влияние полей ионов на изменение квазиимпульса электронов

Если внешние поля F приводят к «медленным» изменениям состояния частицы, то внутренние поля F_d могут привести к резким изменениям состояния за короткое время, в течение которого электрон проходит малую область локального возмущения. Действительно, если область действия локального возмущения имеет раз-

меры в несколько периодов решетки, т. е. порядка 10^{-7} см, а скорость частицы — около 10^7 см/с (тепловая скорость), то время взаимодействия с локальным центром составит всего лишь 10^{-14} с. Такое *кратковременное взаимодействие приводит к значительному изменению скорости и квазиимпульса электрона*, что равносильно удару в механике, поэтому его называют *соударением или столкновением*. В связи с тем что соударения приводят к изменению числа частиц, имеющих направленное движение, *процессы соударения называют также процессами рассеяния.*

Чтобы описать различный характер действия внешних и внутренних сил, перепишем уравнение (36.7) с учетом двух классов сил F и F_d :

$$-\frac{\partial f}{\partial t} = (\mathbf{v}, \nabla_r f) + \frac{1}{\hbar} (F, \nabla_k f) + \frac{1}{\hbar} (F_d, \nabla_k f). \quad (36.10)$$

Уравнение (36.10) показывает, что функция распределения изменяется вследствие движения частиц со скоростью \mathbf{v} и в результате действия внешних F и внутренних F_d сил.

Обозначим изменение функции распределения вследствие движения частиц и действия внешних сил через $(\partial f / \partial t)_{\text{пол}}$, очевидно, что

$$-(\partial f / \partial t)_{\text{пол}} = (\mathbf{v}, \nabla_r f) + \frac{1}{\hbar} (F, \nabla_k f). \quad (36.11)$$

Здесь $(\partial f/\partial t)_{\text{пол}}$ носит название *полевого члена* уравнения Больцмана.

Обозначим через $(\partial f/\partial t)_{\text{ст}}$ изменение функции распределения в результате соударений:

$$-\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = \frac{1}{\hbar} (\mathbf{F}_\mathbf{r}, \nabla_\mathbf{k} f). \quad (36.12)$$

Величину $(\partial f/\partial t)_{\text{ст}}$ называют *интегралом столкновений* (или *соударений*), последнее название связано с интегральным представлением $(\partial f/\partial t)_{\text{ст}}$. Таким образом, изменение функции распределения во времени представлено в виде двух членов — полевого члена $(\partial f/\partial t)_{\text{пол}}$ и интеграла соударения $(\partial f/\partial t)_{\text{ст}}$:

$$(\partial f/\partial t) = (\partial f/\partial t)_{\text{пол}} + (\partial f/\partial t)_{\text{ст}}. \quad (36.13)$$

Для нахождения $(\partial f/\partial t)_{\text{ст}}$ используем статистические методы описания физических явлений.

Предположим, что в результате столкновений частицы переходят из состояний (\mathbf{r}, \mathbf{k}) в состояния $(\mathbf{r}', \mathbf{k}')$. Пусть вероятность этого перехода в единицу времени равна $\omega(\mathbf{k}, \mathbf{k}')$. Очевидно, что в результате соударений координата не испытывает резкого изменения, поэтому вероятность перехода не будет зависеть от \mathbf{r} и \mathbf{r}' . Возьмем два элементарных объема $d\tau_\mathbf{k}$ и $d\tau_{\mathbf{k}'}$ вокруг точек \mathbf{k} и \mathbf{k}' . Число состояний в них с учетом спина равно соответственно $d\tau_\mathbf{k}/4\pi^3$ и $d\tau_{\mathbf{k}'}/4\pi^3$ для единичного объема кристалла. Число занятых состояний равно $f(\mathbf{r}, \mathbf{k}) d\tau_\mathbf{k}/4\pi^3$ и $f(\mathbf{r}, \mathbf{k}') d\tau_{\mathbf{k}'}/4\pi^3$; а свободных состояний $[1 - f(\mathbf{r}, \mathbf{k})] d\tau_\mathbf{k}/4\pi^3$ и $[1 - f(\mathbf{r}, \mathbf{k}')] d\tau_{\mathbf{k}'}/4\pi^3$. В результате столкновений электроны переходят из $d\tau_\mathbf{k}$ в $d\tau_{\mathbf{k}'}$ и из $d\tau_{\mathbf{k}'}$ в $d\tau_\mathbf{k}$. Число переходов будет определяться, помимо вероятности $\omega(\mathbf{k}, \mathbf{k}')$, числом занятых исходных состояний и свободных конечных состояний (необходимо учитывать принцип Паули). Поэтому в результате прямых переходов электронов из $d\tau_\mathbf{k}$ в $d\tau_{\mathbf{k}'}$ и обратных из $d\tau_{\mathbf{k}'}$ в $d\tau_\mathbf{k}$ число занятых состояний за время dt изменится на величину

$$\begin{aligned} & - dt \left\{ \omega(\mathbf{k}, \mathbf{k}') \frac{d\tau_\mathbf{k}}{4\pi^3} f(\mathbf{r}, \mathbf{k}) [1 - f(\mathbf{r}, \mathbf{k}')] \frac{d\tau_{\mathbf{k}'}}{4\pi^3} \right\} + \\ & + dt \left\{ \omega(\mathbf{k}', \mathbf{k}) \frac{d\tau_{\mathbf{k}'}}{4\pi^3} f(\mathbf{r}, \mathbf{k}') [1 - f(\mathbf{r}, \mathbf{k})] \frac{d\tau_\mathbf{k}}{4\pi^3} \right\}. \end{aligned} \quad (36.14)$$

Первое слагаемое определяет уменьшение числа частиц в элементе объема $d\tau_\mathbf{k}$ в результате прямых переходов из $d\tau_\mathbf{k}$ в $d\tau_{\mathbf{k}'}$, второе слагаемое определяет увеличение числа частиц в $d\tau_\mathbf{k}$ в результате обратных переходов из $d\tau_{\mathbf{k}'}$ в $d\tau_\mathbf{k}$ с вероятностью $\omega(\mathbf{k}', \mathbf{k})$. Полное число изменений занятых состояний за время dt в элементе объема $d\tau_\mathbf{k}$ найдем интегрированием выражения (36.14) по всей области изменений \mathbf{k}' , т. е. по объему зоны Бриллюэна $V_\mathbf{k}$:

$$\begin{aligned} & dt \frac{d\tau_\mathbf{k}}{4\pi^3} \int_{(V_\mathbf{k})} \left\{ \omega(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}') [1 - f(\mathbf{r}, \mathbf{k})] - \right. \\ & \left. - \omega(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k}) [1 - f(\mathbf{r}, \mathbf{k}')] \right\} \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \end{aligned} \quad (36.15)$$

С другой стороны, число занятых состояний в каждый момент времени равно $f(\mathbf{r}, \mathbf{k}) d\tau_{\mathbf{k}}/4\pi^3$ и изменение его за время dt в результате соударений можно выразить в виде

$$\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} dt \frac{d\tau_{\mathbf{k}}}{4\pi^3}. \quad (36.16)$$

Приравнивая выражения (36.15) и (36.16), получим после сокращения на $dt d\tau_{\mathbf{k}}/4\pi^3$:

$$\begin{aligned} \left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = & -\frac{1}{\hbar} (\mathbf{F}_{\mathbf{r}}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k})) = \int_{(V_{\mathbf{k}})} \{ \omega(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}') [1 - f(\mathbf{r}, \mathbf{k})] - \\ & - \omega(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k}) [1 - f(\mathbf{r}, \mathbf{k}')] \} \frac{d\tau_{\mathbf{k}'}}{4\pi^3} = \int_{(V_{\mathbf{k}})} \{ [\omega(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}') - \\ & - \omega(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k})] - [\omega(\mathbf{k}', \mathbf{k}) - \omega(\mathbf{k}, \mathbf{k}')] f(\mathbf{r}, \mathbf{k}) f(\mathbf{r}, \mathbf{k}') \} \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \end{aligned} \quad (36.17)$$

В том случае, когда вероятности прямых и обратных переходов равны:

$$\omega(\mathbf{k}, \mathbf{k}') = \omega(\mathbf{k}', \mathbf{k}), \quad (36.18)$$

интеграл столкновений (36.17) упрощается:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [f(\mathbf{r}, \mathbf{k}') - f(\mathbf{r}, \mathbf{k})] \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \quad (36.19)$$

Учитывая (36.13), (36.11) и (36.17), получим

$$\frac{\partial f}{\partial t} = -(\mathbf{v}, \nabla_{\mathbf{r}} f) - \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f) + \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [f(\mathbf{r}, \mathbf{k}') - f(\mathbf{r}, \mathbf{k})] \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \quad (36.20)$$

Уравнение (36.20) носит название *кинетического уравнения Больцмана*.

Поскольку под интегралом в (36.20) стоит искомая функция, это уравнение является интегро-дифференциальным. Его решение в общем виде не получено.

Для стационарного состояния $\partial f/\partial t = 0$, и кинетическое уравнение имеет вид

$$(\partial f/\partial t)_{\text{пол}} = -(\partial f/\partial t)_{\text{ст}}, \quad (36.21)$$

или

$$(\mathbf{v}, \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k})) + \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k})) = \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [f(\mathbf{r}, \mathbf{k}') - f(\mathbf{r}, \mathbf{k})] \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \quad (36.22)$$

Выражения (36.21) или (36.22) показывают, что в стационарном состоянии изменения функции распределения, создаваемые внешними полями и движением частиц, компенсируются столкновениями носителей заряда с локальными нарушениями периодичности поля решетки. Если $(\partial f/\partial t)_{\text{пол}} \neq (\partial f/\partial t)_{\text{ст}}$, то $df/dt \neq 0$ и функция распределения меняется во времени в ту или другую сторону в зависимости от того, какой процесс превалирует — изменение f под действием полей или в результате рассеяния носителей заряда.

Условия применимости кинетического уравнения можно сформулировать следующим образом.

1. Внешнее воздействие не должно менять спектр энергии электрона в кристалле. Это накладывает ограничения на величины полей.

2. Поскольку кинетическое уравнение является квазиклассическим, оно не применимо для описания процессов малой длительности, происходящих в малых объемах, так как это приводит к большой неопределенности в энергии и квазиимпульсе.

Резюме § 36

1. Физические явления, обусловленные движением носителей заряда в полупроводнике, в котором создан градиент температуры или действуют внутренние и внешние поля, называют кинетическими явлениями. Для их описания используют кинетическое уравнение Больцмана.

2. Из того факта, что полное число состояний в кристалле есть постоянная величина, следует, что полная производная по времени от функции распределения $f(\mathbf{r}, \mathbf{k}, t)$ равна нулю:

$$df/dt = 0. \quad (36.1p)$$

Дифференцируя $f(\mathbf{r}, \mathbf{k}, t)$ по времени как сложную функцию, т. е. считая, что \mathbf{r} и \mathbf{k} являются функциями времени, и учитывая (36.1p), получим

$$-\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} = (\mathbf{v}, \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t)) + \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t)). \quad (36.2p)$$

Изменение функции распределения со временем представляется в виде суммы двух членов — полевого и столкновений:

$$\partial f/\partial t = (\partial f/\partial t)_{\text{пол}} + (\partial f/\partial t)_{\text{ст}}, \quad (36.3p)$$

где

$$\left(\frac{\partial f}{\partial t}\right)_{\text{пол}} = -(\mathbf{v}, \nabla_{\mathbf{r}} f) - \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f) \quad (36.4p)$$

и

$$\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = -\frac{1}{\hbar} (\mathbf{F}_{\text{д}}, \nabla_{\mathbf{k}} f). \quad (36.5p)$$

3. Соударения приводят к переходу частиц из одного состояния в другие с вероятностью $\omega(\mathbf{k}, \mathbf{k}')$. Это позволяет выразить действие соударений $(\partial f / \partial t)_{\text{ст}}$ в виде

$$\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = \int_{(V_{\mathbf{k}})} \{ \omega(\mathbf{k}', \mathbf{k}) f(\mathbf{r}, \mathbf{k}') [1 - f(\mathbf{r}, \mathbf{k})] - \omega(\mathbf{k}, \mathbf{k}') f(\mathbf{r}, \mathbf{k}) [1 - f(\mathbf{r}, \mathbf{k}')] \} \frac{d\tau_{\mathbf{k}'}}{4\pi^3}. \quad (36.6p)$$

4. Кинетическое уравнение Больцмана имеет вид

$$\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} = -(\mathbf{v}, \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k}, t)) - \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k}, t)) + \left(\frac{\partial f}{\partial t}\right)_{\text{ст}}. \quad (36.7p)$$

Для стационарного состояния с учетом $\omega(\mathbf{k}, \mathbf{k}') = \omega(\mathbf{k}', \mathbf{k})$ уравнение Больцмана имеет вид

$$\begin{aligned} & \left((\mathbf{v}, \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k})) + \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k})) \right) = \\ & = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [f(\mathbf{r}, \mathbf{k}') - f(\mathbf{r}, \mathbf{k})] d\tau_{\mathbf{k}'}. \end{aligned} \quad (36.8p)$$

В стационарном состоянии изменения, вносимые полем, компенсируются соударениями.

§ 37. ВРЕМЯ РЕЛАКСАЦИИ

Решение кинетического уравнения в общем виде представляет собой очень сложную задачу, которая значительно упрощается для того случая, когда можно ввести так называемое время релаксации.

Предположим, что в некоторый момент времени $t=0$ полевой член обращается в нуль (выключаются поля):

$$(\partial f / \partial t)_{\text{пол}} = 0. \quad (37.1)$$

Из (36.13) следует, что функция распределения изменяется в результате действия соударений:

$$\partial f / \partial t = (\partial f / \partial t)_{\text{ст}}. \quad (37.2)$$

В момент выключения поля система частиц находилась в стационарном неравновесном состоянии, после выключения полей процессы соударений должны привести систему частиц в равновесное состояние, другими словами, *соударения восстанавливают нарушенное полями равновесное состояние*. Простейшее предположение относительно протекания процесса релаксации состоит в том, что *скорость восстановления равновесия пропорциональна величине отклонения* $[f(\mathbf{r}, \mathbf{k}, t) - f_0(\mathbf{r}, \mathbf{k})]$ *от равновесия:*

$$\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = -\frac{f(\mathbf{r}, \mathbf{k}, t) - f_0(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})}. \quad (37.3)$$

В (37.3) f_0 есть функция распределения в равновесном состоянии, $f(\mathbf{r}, \mathbf{k}, t)$ — в неравновесном состоянии, $1/\tau(\mathbf{k})$ — коэффициент пропорциональности, зависящий, вообще говоря, от \mathbf{k} и \mathbf{r} , однако, поскольку в дальнейшем рассмотрении будет важна зависимость τ от \mathbf{k} , опустим зависимость τ от \mathbf{r} . Считая τ величиной положительной, мы ввели в (37.3) знак минус, который обеспечивает возвращение системы к равновесному состоянию. Уравнение (37.3) решается элементарно:

$$f(\mathbf{r}, \mathbf{k}, t) = f_0(\mathbf{r}, \mathbf{k}) = [f(\mathbf{r}, \mathbf{k}, 0) - f_0(\mathbf{r}, \mathbf{k})] e^{-\frac{t}{\tau}}. \quad (37.4)$$

Величина $\tau(\mathbf{k})$ показывает, насколько быстро восстанавливается нарушенное полем равновесное состояние, поэтому она носит название времени релаксации. Время релаксации позволяет выразить интеграл соударений в более простом виде, действительно, из (37.3) и (36.19) следует

$$\left(\frac{\partial f}{\partial t}\right)_{\text{ст}} = \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [f(\mathbf{r}, \mathbf{k}', t) - f(\mathbf{r}, \mathbf{k}, t)] \frac{d\tau_{\mathbf{k}'}}{4\pi^3} = -\frac{f - f_0}{\tau(\mathbf{k})}. \quad (37.5)$$

Из (37.5) вытекает «явное» выражение для времени релаксации через вероятность рассеяния и функцию распределения:

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') \frac{[f(\mathbf{r}, \mathbf{k}, t) - f(\mathbf{r}, \mathbf{k}', t)]}{[f(\mathbf{r}, \mathbf{k}, t) - f_0(\mathbf{r}, \mathbf{k})]} d\tau_{\mathbf{k}'}. \quad (37.6)$$

Дальнейшее, необходимое для решения уравнения Больцмана, предположение состоит в следующем: время релаксации $\tau(\mathbf{k})$ является однозначной характеристикой процессов соударений как во время релаксации, так и при действии внешних полей, другими словами, предположим, что время релаксации не зависит от внешних полей. Позже обсудим это предположение, а теперь используем его для решения уравнения Больцмана, описывающего стационарное состояние. Из (36.8) и (37.5) следует

$$(\mathbf{v} \cdot \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k})) + \frac{1}{\hbar} (\mathbf{F} \cdot \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k})) = -\frac{f(\mathbf{r}, \mathbf{k}) - f_0(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})}. \quad (37.7)$$

Уравнение (37.7) является основным уравнением, описывающим кинетические явления в стационарном состоянии в том случае, когда можно ввести время релаксации для учета действия процессов соударений. Решение уравнения (37.7) будем искать в виде ряда

$$f(\mathbf{r}, \mathbf{k}) = f_0(\mathbf{r}, \mathbf{k}) + f^{(1)}(\mathbf{r}, \mathbf{k}) + f^{(2)}(\mathbf{r}, \mathbf{k}) + \dots, \quad (37.8)$$

где $f^{(1)}(\mathbf{r}, \mathbf{k})$; $f^{(2)}(\mathbf{r}, \mathbf{k})$ есть поправка в первом, втором и т. д. приближениях к равновесной функции распределения. Производные от $f^{(1)}(\mathbf{r}, \mathbf{k})$ по \mathbf{r} и \mathbf{k} должны быть того же порядка, что и $f^{(2)}(\mathbf{r}, \mathbf{k})$, поэтому, чтобы найти решение (37.7) в первом приближении, можем

отбросить все величины более высокого порядка. Итак, ищем решение уравнения (37.7) в первом приближении:

$$f(\mathbf{r}, \mathbf{k}) = f_0(\mathbf{r}, \mathbf{k}) + f^{(1)}(\mathbf{r}, \mathbf{k}). \quad (37.9)$$

Подставим (37.9) в (37.7), получим уравнение для $f^{(1)}(\mathbf{r}, \mathbf{k})$:

$$\begin{aligned} & (\mathbf{v}, \nabla_{\mathbf{r}} f_0(\mathbf{r}, \mathbf{k}) + \nabla_{\mathbf{r}} f^{(1)}(\mathbf{r}, \mathbf{k})) + \\ & + \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f_0(\mathbf{r}, \mathbf{k}) + \nabla_{\mathbf{k}} f^{(1)}(\mathbf{r}, \mathbf{k})) = - \frac{f^{(1)}(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})}. \end{aligned} \quad (37.10)$$

Вычислим $\nabla_{\mathbf{r}} f_0(\mathbf{r}, \mathbf{k})$ и $\nabla_{\mathbf{k}} f_0(\mathbf{r}, \mathbf{k})$, учитывая, что от \mathbf{r} могут зависеть F и T , а от \mathbf{k} зависит E :

$$\nabla_{\mathbf{r}} f_0(\mathbf{r}, \mathbf{k}) = \frac{e^{\frac{E-F}{kT}} \nabla_{\mathbf{r}} \left(\frac{E-F}{kT} \right)}{\left[e^{\frac{E-F}{kT}} + 1 \right]^2} = - \frac{\partial f_0}{\partial E} \left\{ \nabla_{\mathbf{r}} F + (E-F) \frac{\nabla_{\mathbf{r}} T}{T} \right\} \quad (37.11)$$

и

$$\nabla_{\mathbf{k}} f_0(\mathbf{r}, \mathbf{k}) = - \frac{e^{\frac{E-F}{kT}}}{\left[e^{\frac{E-F}{kT}} + 1 \right]^2} \frac{\nabla_{\mathbf{k}} E}{kT} = \frac{\partial f_0}{\partial E} \mathbf{v}. \quad (37.12)$$

Мы видим, что в выражения для градиентов $\nabla_{\mathbf{r}} f_0$ и $\nabla_{\mathbf{k}} f_0$ входит величина

$$\frac{\partial f_0}{\partial E} = - \frac{e^{\frac{E-F}{kT}}}{\left[e^{\frac{E-F}{kT}} + 1 \right]^2} \frac{1}{kT}. \quad (37.13)$$

Учитывая (37.11) и (37.12), уравнение (37.10) для $f^{(1)}$ приведем к виду

$$\begin{aligned} - \frac{f^{(1)}(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})} = & - (\mathbf{v}, \nabla_{\mathbf{r}} F + (E-F) \nabla_{\mathbf{r}} \ln T) \frac{\partial f_0}{\partial E} + \\ & + (\mathbf{v}, \nabla_{\mathbf{r}} f^{(1)}(\mathbf{r}, \mathbf{k})) + (e\mathbf{E} + e[\mathbf{v}\mathbf{B}], \mathbf{v}) \frac{\partial f_0}{\partial E} + \frac{1}{\hbar} (e\mathbf{E} + e[\mathbf{v}\mathbf{B}], \nabla_{\mathbf{k}} f^{(1)}), \end{aligned} \quad (37.14)$$

где вместо \mathbf{F} мы поставили выражение для силы Лоренца. Полагая, что производные от $f^{(1)}$ имеют порядок малости, сравнимый с $f^{(2)}$, и пренебрегая вторым и четвертым членами в (37.14), получим

$$- \frac{f^{(1)}}{\tau(\mathbf{k})} = \frac{\partial f_0}{\partial E} \{ (e\mathbf{E} - \nabla_{\mathbf{r}} F - (E-F) \nabla_{\mathbf{r}} \ln T), \mathbf{v} \}. \quad (37.15)$$

Из (37.15) мы видим, что в первом приближении $f^{(1)}(\mathbf{r}, \mathbf{k})$ не зависит от магнитного поля. Чтобы получить зависимость от \mathbf{B} , необходимо оставить в (37.14) член с $\nabla_{\mathbf{k}} f^{(1)}(\mathbf{r}, \mathbf{k})$.

Вычислим его, используя (37.15). Однако прежде чем искать $\nabla_{\mathbf{k}} f^{(1)}(\mathbf{r}, \mathbf{k})$, запишем $f^{(1)}(\mathbf{r}, \mathbf{k})$, исходя из выражения (37.15),

$$f^{(1)}(\mathbf{r}, \mathbf{k}) = -\frac{\partial f_0}{\partial E}(\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k}), \mathbf{v}), \quad (37.16)$$

где $\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k})$ — неизвестная векторная функция, к нахождению которой сводится задача нахождения $f^{(1)}(\mathbf{r}, \mathbf{k})$. Из сравнения (37.16) и (37.15) следует, что для $\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k})$ без учета магнитного поля можем записать

$$\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k}) = \tau(\mathbf{k}) \{e\mathbf{E} - \nabla_{\mathbf{r}} F - (E - F) \nabla_{\mathbf{r}} \ln T\}. \quad (37.17)$$

Учитывая, что $\mathbf{E} = -\nabla_{\mathbf{r}} \varphi(\mathbf{r})$, получим

$$\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k}) = -\tau(\mathbf{k}) \{ \nabla_{\mathbf{r}} (e\varphi + F) + (E - F) \nabla_{\mathbf{r}} \ln T \}. \quad (37.18)$$

Теперь вычислим $\nabla_{\mathbf{k}} f^{(1)}(\mathbf{r}, \mathbf{k})$, учитывая (37.16):

$$\nabla_{\mathbf{k}} f^{(1)}(\mathbf{r}, \mathbf{k}) = -\frac{d}{d\mathbf{k}} \left\{ \frac{\partial f_0}{\partial E} \left(\boldsymbol{\kappa} \frac{dE}{\hbar d\mathbf{k}} \right) \right\} = -\frac{\partial f_0}{\partial E} \left(\boldsymbol{\kappa} \frac{\hbar}{m^*} \right) - \left(\mathbf{v}, \frac{d}{d\mathbf{k}} \cdot \frac{\partial f_0}{\partial E} \boldsymbol{\kappa} \right). \quad (37.19)$$

В выражении (37.19) $\frac{1}{m^*} = \frac{d^2 E}{\hbar^2 d^2 \mathbf{k}^2}$ — обобщенный тензор обратной эффективной массы. Если подставить (37.19) в (37.14), в котором необходимо сохранить только член с \mathbf{B} , получим из последнего слагаемого в (37.14)

$$\begin{aligned} \frac{e}{\hbar} ([\mathbf{v}\mathbf{B}], \nabla_{\mathbf{k}} f^{(1)}) &= -\frac{e}{\hbar} ([\mathbf{v}, \mathbf{B}], \left(\boldsymbol{\kappa} \frac{\hbar}{m^*} \right)) \frac{\partial f_0}{\partial E} - \frac{e}{\hbar} ([\mathbf{v}, \mathbf{B}], \\ & \left(\mathbf{v}, \nabla_{\mathbf{k}} \frac{\partial f_0}{\partial E} \boldsymbol{\kappa} \right)) = e \left(\left[\frac{\boldsymbol{\kappa}}{m^*}, \mathbf{B} \right], \mathbf{v} \right) \frac{\partial f_0}{\partial E}, \end{aligned} \quad (37.20)$$

а второе слагаемое обращается в нуль в силу $[\mathbf{v}\mathbf{v}] = 0$. Следовательно, (37.14) имеет вид

$$\begin{aligned} -\frac{f^{(1)}(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})} &= \frac{1}{\tau(\mathbf{k})} \cdot \frac{\partial f_0}{\partial E}(\boldsymbol{\kappa}\mathbf{v}) = \frac{\partial f_0}{\partial E}(\mathbf{v}, e\mathbf{E} - \nabla_{\mathbf{r}} F - (E - F) \nabla_{\mathbf{r}} \ln T + \\ & + e \left[\frac{\boldsymbol{\kappa}}{m^*}, \mathbf{B} \right]). \end{aligned} \quad (37.21)$$

Для $\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k})$ мы получили уравнение

$$\boldsymbol{\kappa}(\mathbf{r}, \mathbf{k}) = -\tau(\mathbf{k}) \left\{ \nabla_{\mathbf{r}} (F + e\varphi) + (E - F) \nabla_{\mathbf{r}} \ln T - e \left[\frac{\boldsymbol{\kappa}}{m^*}, \mathbf{B} \right] \right\}. \quad (37.22)$$

Чтобы решить уравнение (37.22), введем обозначения для сокращения записи:

$$\nabla_{\mathbf{r}} (F + e\varphi) + (E - F) \nabla_{\mathbf{r}} \ln T = -\mathbf{L}, \quad (37.23)$$

$$\tau(\mathbf{k}) \mathbf{L} = \mathbf{A}. \quad (37.24)$$

В новых обозначениях уравнение (37.22) имеет вид

$$\boldsymbol{\kappa} = \mathbf{A} + e\tau \left[\frac{\boldsymbol{\kappa}}{m^*}, \mathbf{B} \right]. \quad (37.25)$$

Решение уравнения (37.25) проводится несколько различно для случаев, когда m^* является скалярной и тензорной величиной.

1. Скалярная эффективная масса. Так как m^* — скалярная величина, то мы можем ввести вектор $\varphi(\mathbf{k})$:

$$\varphi(\mathbf{k}) = \frac{e\tau(\mathbf{k})}{m^*} \mathbf{B}. \quad (37.26)$$

Тогда (37.25) можно записать в виде

$$\kappa = \mathbf{A} + [\kappa\varphi]. \quad (37.27)$$

Умножим (37.27) скалярно на φ :

$$(\kappa\varphi) = (\mathbf{A}\varphi) + ([\kappa\varphi]\varphi) = (\mathbf{A}\varphi), \quad (37.28)$$

так как

$$([\kappa\varphi]\varphi) = ([\varphi\varphi]\kappa) = 0. \quad (37.29)$$

Умножим (37.27) векторно на φ :

$$\begin{aligned} [\kappa\varphi] &= [\mathbf{A}\varphi] + [[\kappa\varphi]\varphi] = [\mathbf{A}\varphi] + \varphi(\kappa\varphi) - \kappa(\varphi\varphi) = \\ &= [\mathbf{A}\varphi] + \varphi(\mathbf{A}\varphi) - \kappa\varphi^2 = \kappa - \mathbf{A}. \end{aligned} \quad (37.30)$$

При упрощении выражений в цепи равенств (37.30) мы учли правило разложения двойного векторного произведения по сомножителям одинарного векторного произведения и заменили $(\kappa\varphi)$ и $[\kappa\varphi]$ на $(\mathbf{A}\varphi)$ и $\kappa - \varphi$ согласно (37.28) и (37.27). Из (37.30) для κ следует выражение

$$\kappa(1 + \varphi^2) = \mathbf{A} + [\mathbf{A}\varphi] + \varphi(\mathbf{A}\varphi), \quad (37.31)$$

или

$$\kappa = \frac{\mathbf{A} + [\mathbf{A}\varphi] + \varphi(\mathbf{A}\varphi)}{1 + \varphi^2}. \quad (37.32)$$

Подставляя вместо \mathbf{A} и φ их выражения, получим

$$\begin{aligned} \kappa(\mathbf{r}, \mathbf{k}) &= \frac{-\tau \{ \nabla_{\mathbf{r}}(e\varphi + F) + (E - F) \nabla_{\mathbf{r}} \ln T \} -}{1 + \frac{e^2\tau^2}{m^{*2}} B^2} \rightarrow \\ &\rightarrow \frac{-\frac{e\tau^2}{m^*} [\nabla_{\mathbf{r}}(e\varphi + F) + (E - F) \nabla_{\mathbf{r}} \ln T, \mathbf{B}]}{1 + \frac{e^2\tau^2}{m^{*2}} B^2} \rightarrow \\ &\rightarrow \frac{-\frac{e^2\tau^3}{m^{*3}} \mathbf{B} (\nabla_{\mathbf{r}}(e\varphi + F) + (E - F) \nabla_{\mathbf{r}} \ln T, \mathbf{B})}{1 + \frac{e^2\tau^2}{m^{*2}} B^2}. \end{aligned} \quad (37.33)$$

Эффекты, зависящие от $f^{(1)}(\mathbf{r}, \mathbf{k})$ и, следовательно, от $\kappa(\mathbf{r}, \mathbf{k})$, носят название поперечных, если $\mathbf{B} \perp \mathbf{L}$, и продольных, если $\mathbf{B} \parallel \mathbf{L}$. Для поперечных эффектов $(\mathbf{L}, \mathbf{B}) = 0$ и

$$\kappa = \frac{\mathbf{A} + [\mathbf{A}\varphi]}{1 + \varphi^2}. \quad (37.34)$$

Для продольных эффектов $[LB] = 0$ и

$$\kappa = \frac{A + \varphi(A\varphi)}{1 + \varphi^2} = \frac{A + A\varphi^2}{1 + \varphi^2} = A = -\tau \{ \nabla_r (e\varphi + F) + (E - F) \nabla_r \ln T \}, \quad (37.35)$$

т. е. в продольном магнитном поле явления протекают так же, как и при $\mathbf{B} = 0$, если m^* — скалярная величина. Если же продольные эффекты наблюдаются, то это свидетельствует о тензорном характере эффективной массы.

2. Тензорная эффективная масса. Предположим, что m^{*-1} имеет диагональный вид, в таком случае тензор m^* также имеет диагональный вид. Рассмотрим теперь векторное произведение $[m^{*-1}\kappa, \mathbf{B}] = \mathbf{D}$. Обозначим $m^{*-1}\kappa = \eta$. Легко видеть, что

$$\kappa = m^* \eta. \quad (37.36)$$

Найдем компоненты вектора \mathbf{D} :

$$\begin{aligned} D_x &= \eta_y B_z - \eta_z B_y; & D_y &= \eta_z B_x - \eta_x B_z, \\ D_z &= \eta_x B_y - \eta_y B_x, \end{aligned} \quad (37.37)$$

но

$$\eta_i = \{m^{*-1}\kappa\}_i = \sum_j m_{ij}^{*-1} \kappa_j = m_i^{-1} \kappa_i = \frac{\kappa_i}{m_i}, \quad (37.38)$$

поэтому, например, D_x имеет вид

$$D_x = \frac{\kappa_y B_z}{m_y} - \frac{\kappa_z B_y}{m_z}. \quad (37.39)$$

Рассмотрим $m^{*-1}\mathbf{D}$, определяемый своими компонентами,

$$\{m^{*-1}\mathbf{D}\}_i = \sum_j m_{ij}^{*-1} D_j = \frac{D_i}{m_i}, \quad (37.40)$$

или, с учетом (37.37)

$$\begin{aligned} \{m^{*-1}\mathbf{D}\}_x &= \frac{D_x}{m_x} = \frac{\kappa_y B_z}{m_x m_y} - \frac{\kappa_z B_y}{m_x m_z} = \\ &= \frac{1}{m_x m_y m_z} (\kappa_y m_z B_z - \kappa_z m_y B_y) = |m^{*-1}| \left(\kappa_y \sum_j m_{zj} B_j - \kappa_z \sum_j m_{yj} B_j \right), \end{aligned} \quad (37.41)$$

т. е.

$$m^{*-1}\mathbf{D} = m^{*-1} [m^{*-1}\kappa, \mathbf{B}] = |m^{*-1}| [\kappa, m^*\mathbf{B}]. \quad (37.42)$$

Через $|m^*|$ обозначен определить матрицы тензора m^* :

$$|m^*| = \text{Det} \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix} = m_1 m_2 m_3. \quad (37.43)$$

Вернемся к уравнению (37.25) для κ . Умножим его скалярно на m^{*-1} и, учитывая (37.42), получим

$$m^{*-1}\kappa = m^{*-1}A + e\tau m^{*-1} [m^{*-1}\kappa, B] = m^{*-1}A + \frac{e\tau}{|m^*|} [\kappa, m^*B]. \quad (37.44)$$

Умножая (37.44) векторно на B , будем иметь

$$[m^{*-1}\kappa, B] = [m^{*-1}A, B] + \frac{e\tau}{|m^*|} [[\kappa, m^*B] B]. \quad (37.45)$$

Разложим двойное векторное произведение

$$[[\kappa, m^*B] B] = (\kappa B) m^*B - (B, m^*B) \kappa \quad (37.46)$$

и подставим его в (37.45); получим

$$[m^{*-1}\kappa, B] = [m^{*-1}A, B] + \frac{e\tau}{|m^*|} m^*B (\kappa B) - \frac{e\tau}{|m^*|} (B, m^*B) \kappa. \quad (37.47)$$

Но из (37.25) следует, что

$$(\kappa B) = (AB), \quad (37.48)$$

$$e\tau [m^{*-1}\kappa, B] = \kappa - A, \quad (37.49)$$

поэтому, сравнивая (37.49), (37.47), запишем

$$\kappa - A = e\tau [m^{*-1}A, B] + \frac{e^2\tau^2}{|m^*|} (AB) m^*B - \frac{e^2\tau^2}{|m^*|} (B, m^*B) \kappa. \quad (37.50)$$

Собирая одинаковые члены, получим

$$\kappa = \frac{A + e\tau [m^{*-1}A, B] + \frac{e^2\tau^2}{|m^*|} (AB) m^*B}{1 + \frac{e^2\tau^2}{|m^*|} (B, m^*B)}. \quad (37.51)$$

Если эффективная масса — скалярная величина, т. е. все диагональные элементы равны m^* , то $|m^*| = m^{*3}$ и уравнение (37.51) переходит в уравнение (37.32).

Резюме § 37

1. Процессы соударения приводят к восстановлению нарушаемого полями равновесного распределения электронов и дырок. Их действие можно описать величиной времени релаксации $\tau(\kappa)$, которая определяется выражением (37.6) через интеграл соударений. Время релаксации τ есть величина, равная среднему времени существования неравновесного состояния после выключения полей, вызвавших это отклонение.

2. Предполагая, что время релаксации не зависит от внешних полей, приводим уравнение Больцмана для стационарных состояний к виду

$$(\mathbf{v}, \nabla_{\mathbf{r}} f(\mathbf{r}, \mathbf{k})) + \frac{1}{\hbar} (\mathbf{F}, \nabla_{\mathbf{k}} f(\mathbf{r}, \mathbf{k})) = - \frac{f(\mathbf{r}, \mathbf{k}) - f_0(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})}. \quad (37.1p)$$

3. Решая уравнение (37.1p) методом разложения в ряд по величине полей, получаем уравнение для поправки первого приближения:

$$-\frac{f^{(1)}(\mathbf{r}, \mathbf{k})}{\tau(\mathbf{k})} = \frac{\partial f_0}{\partial E} (\mathbf{v}, e\mathbf{E} + \nabla_{\mathbf{r}}(E - F) - (E - F) \nabla_{\mathbf{r}} \ln T) + \frac{e}{\hbar} [(\mathbf{v}\mathbf{B}) \nabla_{\mathbf{k}} f^{(1)}]. \quad (37.2p)$$

4. Записав $f^{(1)}(\mathbf{r}, \mathbf{k})$ в виде

$$f^{(1)}(\mathbf{r}, \mathbf{k}) = - \frac{\partial f_0}{\partial E} (\kappa(\mathbf{r}, \mathbf{k}), \mathbf{v}), \quad (37.3p)$$

для $\kappa(\mathbf{r}, \mathbf{k})$ получаем уравнение

$$\kappa(\mathbf{r}, \mathbf{k}) = \mathbf{A} + e\tau \left[\frac{\kappa}{m^*}, \mathbf{B} \right], \quad (37.4p)$$

где

$$\mathbf{A} = \tau \mathbf{L} = -\tau(\mathbf{k}) \{ \nabla_{\mathbf{r}}(e\varphi + F) + (E - F) \nabla_{\mathbf{r}} \ln T \}. \quad (37.5p)$$

5. Решение для $\kappa(\mathbf{r}, \mathbf{k})$ в случае, когда эффективная масса является тензором, имеет вид

$$\kappa(\mathbf{r}, \mathbf{k}) = \frac{\mathbf{A} + e\tau [m^{*-1}\mathbf{A}, \mathbf{B}] + \frac{e^2\tau^2}{|m^*|} (\mathbf{A}\mathbf{B}) m^*\mathbf{B}}{1 + \frac{e^2\tau^2}{|m^*|} (\mathbf{B}, m^*\mathbf{B})}. \quad (37.6p)$$

В случае скалярной эффективной массы

$$\kappa(\mathbf{r}, \mathbf{k}) = \frac{\mathbf{A} + \frac{e\tau}{m^*} [\mathbf{A}\mathbf{B}] + \frac{e^2\tau^2}{m^{*2}} (\mathbf{A}\mathbf{B}) \mathbf{B}}{1 + \frac{e^2\tau^2}{m^{*2}} \mathbf{B}^2}. \quad (37.7p)$$

Подставляя (37.7p) или (37.6p) с учетом (37.5p) в (37.3p), получим решение кинетического уравнения Больцмана для стационарного состояния.

6. Все уравнения, полученные в этом параграфе, одинаково справедливы как для электронов, так и для дырок, если учесть, что во всех выражениях необходимо иметь в виду величины, характеризующие данные частицы. Это относится не только к заряду e , эффективной массе m^* , времени релаксации τ , но и функции распределения $f_0(\mathbf{r}, \mathbf{k})$ в соответствии с условиями, изложенными в § 25. Если в полупроводнике имеются носители заряда нескольких типов, то каждый тип носителей заряда может быть описан своей функцией распределения.

§ 38. ПЛОТНОСТЬ ЭЛЕКТРИЧЕСКОГО ТОКА И ПЛОТНОСТЬ ПОТОКА ЭНЕРГИИ

Как указывалось неоднократно, для создания направленного потока носителей заряда необходимо нарушить симметрию функции распределения, нарушить равновесное состояние системы зарядов. Направленный поток заряженных частиц создает ток, его можно характеризовать вектором плотности тока, выражение для которого легко найти.

В элементе объема $d\tau_k$ и единичном объеме кристалла содержится $2 \frac{d\tau_k}{8\pi^3}$ состояний, в которых находится

$$dn = \frac{d\tau_k}{4\pi^3} f(\mathbf{r}, \mathbf{k}) \quad (38.1)$$

носителей заряда *, имеющих скорость $\mathbf{v} = \frac{1}{\hbar} \frac{dE}{d\mathbf{k}}$ и создающих элементарную плотность тока

$$d\mathbf{j} = e\mathbf{v} \frac{d\tau_k}{4\pi^3} f(\mathbf{r}, \mathbf{k}). \quad (38.2)$$

Полная плотность тока

$$\mathbf{j} = \frac{e}{4\pi^3} \int_{(V_k)} \mathbf{v} f(\mathbf{r}, \mathbf{k}) d\tau_k = \frac{e}{4\pi^3} \int_{(V_k)} \mathbf{v}(\mathbf{k}) f^{(1)}(\mathbf{r}, \mathbf{k}) d\tau_k. \quad (38.3)$$

В (38.3) мы учли, что

$$\int_{(V_k)} \mathbf{v} f_0(\mathbf{r}, \mathbf{k}) d\tau_k = 0, \quad (38.4)$$

потому что $f_0(\mathbf{r}, \mathbf{k})$ — четная функция \mathbf{k} , а $\mathbf{v} f_0(\mathbf{r}, \mathbf{k})$ — соответственно нечетная функция, которая при интегрировании в симметричных пределах дает нуль. Физически это означает, что *в веществе, находящемся в состоянии термодинамического равновесия, тока нет.* Таким образом, \mathbf{j} можно выразить в виде

$$\mathbf{j} = -\frac{e}{4\pi^3} \int_{(V_k)} \mathbf{v} \frac{\partial f_0}{\partial E}(\mathbf{x}\mathbf{v}) d\tau_k = -\frac{e}{4\pi^3 \hbar^2} \int_{(V_k)} \frac{\partial f_0}{\partial E} \left(\mathbf{x} \frac{dE}{d\mathbf{k}} \right) \frac{dE}{d\mathbf{k}} d\tau_k. \quad (38.5)$$

Направленный поток частиц создает не только электрический ток, но и приводит к переносу энергии, так как каждая частица несет энергию E , вследствие чего *плотность потока энергии \mathbf{W} , т. е. количество энергии, проходящей в единицу времени через нормальное единичное сечение, определяется выражением*

$$\mathbf{W} = \frac{1}{4\pi^3} \int_{(V_k)} E \mathbf{v} f^{(1)}(\mathbf{r}, \mathbf{k}) d\tau_k = -\frac{1}{4\pi^3} \int_{(V_k)} E \mathbf{v} \frac{\partial f_0}{\partial E}(\mathbf{x}\mathbf{v}) d\tau_k. \quad (38.6)$$

* Через dn обозначено число электронов или дырок.

Если вместо κ в (38.5) и (38.6) подставить (37.52), то получим выражение для \mathbf{j} и \mathbf{W} :

$$\mathbf{j} = -\frac{e}{4\pi^3} \int_{(V_\kappa)} \frac{\partial f_0}{\partial E} \frac{\left(\left\{ \tau \mathbf{L} + e\tau^2 [m^{*-1} \mathbf{L}, \mathbf{B}] + \frac{e^2 \tau^3}{|m^*|} (\mathbf{L}\mathbf{B}) m^* \mathbf{B} \right\} \mathbf{v} \right) v d\tau_\kappa}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})}, \quad (38.7)$$

$$\mathbf{W} = -\frac{1}{4\pi^3} \int_{(V_\kappa)} \frac{\partial f_0}{\partial E} E \frac{\left(\left\{ \tau \mathbf{L} + e\tau^2 [m^{*-1} \mathbf{L}, \mathbf{B}] + \frac{e^2 \tau^3}{|m^*|} (\mathbf{L}\mathbf{B}) m^* \mathbf{B} \right\} \mathbf{v} \right) v d\tau_\kappa}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})}. \quad (38.8)$$

Прежде чем вычислять выражения (38.7) и (38.8), рассмотрим соотношение вида

$$\mathbf{M} = -\frac{1}{4\pi^3} \int_{(V_\kappa)} E r^{-1} \tau^s \frac{\partial f_0}{\partial E} \frac{(\mathbf{G}\mathbf{v}) \mathbf{v} d\tau_\kappa}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})}, \quad (38.9)$$

где вектор \mathbf{G} не зависит от κ . Для i -компонента M_i можно записать

$$M_i = -\frac{1}{4\pi^3} \int_{(V_\kappa)} \frac{E r^{-1} \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})} \frac{\partial f_0}{\partial E} \left(\sum_j G_j v_j \right) v_i d\tau_\kappa. \quad (38.10)$$

Если обозначить

$$-\frac{1}{4\pi^3} \int_{(V_\kappa)} \frac{E r^{-1} \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})} \frac{\partial f_0}{\partial E} v_i v_j d\tau_\kappa = K'_{rs}{}^{ij}, \quad (38.11)$$

то (38.10) можно записать в виде

$$M_i = \sum_j K'_{rs}{}^{ij} G_j \quad (38.12)$$

и

$$\mathbf{M} = K'_{rs} \mathbf{G}, \quad (38.13)$$

где K'_{rs} является тензором второго ранга, а $K'_{rs}{}^{ij}$ — его ij -элемент, определяемый выражением (38.11).

Из (38.7) и (38.8) видно, что \mathbf{j} и \mathbf{W} можно выразить через тензор типа K'_{rs} , который назовем *тензором (обобщенного) кинетического коэффициента*. Запишем \mathbf{j} и \mathbf{W} в случае скалярной эффективной массы.

Учитывая, что \mathbf{L} можно представить в виде

$$\begin{aligned} \mathbf{L} &= -\nabla_r (e\varphi + F) - (E - F) \nabla_r \ln T = \\ &= e\mathbf{E} - \nabla_r F - E \frac{\nabla_r T}{T} + \frac{F}{T} \nabla_r T = e\mathbf{E} - E \frac{\nabla_r T}{T} - T \nabla_r \frac{F}{T}, \end{aligned} \quad (38.14)$$

опустив значок γ у оператора набла, запишем

$$\begin{aligned}
 \mathbf{j} = & -\frac{e}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} \tau v^2 \left\{ eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T} \right\} d\tau_K - \\
 & -\frac{e}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} \tau^2 v^2 \frac{e}{m^*} \left[eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T}, \mathbf{B} \right] d\tau_K - \\
 & -\frac{e}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} \tau^3 v^2 \frac{e^2}{m^{*2}} \mathbf{B} \left(eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T}, \mathbf{B} \right) d\tau_K
 \end{aligned} \quad (38.15)$$

$$\begin{aligned}
 \mathbf{W} = & -\frac{1}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} E \tau v^2 \left\{ eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T} \right\} d\tau_K - \\
 & -\frac{1}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} E \tau^2 v^2 \frac{e}{m^*} \left[eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T}, \mathbf{B} \right] d\tau_K - \\
 & -\frac{1}{12\pi^3} \int_{(V_K)} \frac{\partial f_0}{\partial E} E \tau^3 v^2 \frac{e^2}{m^{*2}} \mathbf{B} \left(eE - T\nabla \frac{F}{T} - E \frac{\nabla T}{T}, \mathbf{B} \right) d\tau_K
 \end{aligned} \quad (38.16)$$

Соотношения (38.15) и (38.16) могут быть выражены через кинетические коэффициенты:

$$\begin{aligned}
 \mathbf{j} = & \left(eE - T\nabla \frac{F}{T} \right) eK'_{11} - e \frac{\nabla T}{T} K'_{21} + \frac{e^2}{m^*} \left[eE - T\nabla \frac{F}{T}, \mathbf{B} \right] K'_{12} - \\
 & - \frac{e^2}{m^*} \left[\frac{\nabla T}{T}, \mathbf{B} \right] K'_{22} + \frac{e^3}{m^{*2}} \mathbf{B} \left(eE - T\nabla \frac{F}{T}, \mathbf{B} \right) K'_{13} - \frac{e^3}{m^{*2}} \mathbf{B} \left(\frac{\nabla T}{T}, \mathbf{B} \right) K'_{23}.
 \end{aligned} \quad (38.17)$$

Выражение для \mathbf{W} имеет аналогичный вид, только первый индекс у кинетического коэффициента на единицу больше, а показатель на единицу меньше:

$$\begin{aligned}
 \mathbf{W} = & \left(eE - T\nabla \frac{F}{T} \right) K'_{21} - \frac{\nabla T}{T} K'_{31} + \frac{e}{m^*} \left[eE - T\nabla \frac{F}{T}, \mathbf{B} \right] K'_{22} - \\
 & - \frac{e}{m^*} \left[\frac{\nabla T}{T}, \mathbf{B} \right] K'_{32} + \frac{e^2}{m^{*2}} \mathbf{B} \left(eE - T\nabla \frac{F}{T}, \mathbf{B} \right) K'_{23} - \frac{e^2}{m^{*2}} \mathbf{B} \left(\frac{\nabla T}{T}, \mathbf{B} \right) K'_{33}.
 \end{aligned} \quad (38.18)$$

Перепишем выражение (38.17) для \mathbf{j} , объединяя члены:

$$\begin{aligned} \mathbf{j} = & \left\{ e^2 K'_{11} \mathbf{E} - e K'_{11} T \nabla \frac{F}{T} - e K'_{21} \frac{\nabla T}{T} \right\} + \\ & + \left[\frac{e^3}{m^*} K'_{12} \mathbf{E} - \frac{e}{m^*} K'_{12} T \nabla \frac{F}{T} - \frac{e^2}{m^*} K'_{22} \frac{\nabla T}{T}, \mathbf{B} \right] + \\ & + \left(\frac{e^4}{m^{*2}} K'_{13} \mathbf{E} - \frac{e^3}{m^{*2}} K'_{13} T \nabla \frac{F}{T} - \frac{e^3}{m^{*2}} K'_{23} \frac{\nabla T}{T}, \mathbf{B} \right) \mathbf{B}. \end{aligned} \quad (38.19)$$

Первый член в (38.19) определяет омический ток, второй и третий члены представляют токи, обусловленные градиентом химического потенциала и температуры, т. е. связаны с диффузионным и термоэлектрическим токами, четвертый член обуславливает поперечные гальвано- и термомагнитные токи, пятый член определяет изменение продольных токов, вызываемое магнитным полем.

Резюме § 38.

1. Плотность тока \mathbf{j} и плотность потока энергии \mathbf{W} зависят от градиентов электростатического потенциала ϕ , химического потенциала F , температуры T , магнитного поля \mathbf{B} и характеристик вещества, определяемых эффективной массой и тензором кинетического коэффициента K'_{rs} :

$$\begin{aligned} \mathbf{j} = & \left\{ e^2 K'_{11} \mathbf{E} - e K'_{11} T \nabla \frac{F}{T} - e K'_{21} \frac{\nabla T}{T} \right\} + \\ & + \left[\frac{e^3}{m^*} K'_{12} \mathbf{E} - \frac{e^2}{m^*} K'_{12} T \nabla \frac{F}{T} - \frac{e^2}{m^*} K'_{22} \frac{\nabla T}{T}, \mathbf{B} \right] + \\ & + \left(\frac{e^4}{m^{*2}} K'_{13} \mathbf{E} - \frac{e^3}{m^{*2}} K'_{13} T \nabla \frac{F}{T} - \frac{e^3}{m^{*2}} K'_{23} \frac{\nabla T}{T}, \mathbf{B} \right) \mathbf{B} \end{aligned} \quad (38.1p)$$

и

$$\begin{aligned} \mathbf{W} = & \left\{ e K'_{21} \mathbf{E} - K'_{21} T \nabla \frac{F}{T} - K'_{31} \frac{\nabla T}{T} \right\} + \\ & + \left[\frac{e^2}{m^*} K'_{22} \mathbf{E} - \frac{e}{m^*} K'_{22} T \nabla \frac{F}{T} - \frac{e}{m^*} K'_{32} \frac{\nabla T}{T}, \mathbf{B} \right] + \\ & + \left(\frac{e^3}{m^{*2}} K'_{23} \mathbf{E} - \frac{e^2}{m^{*2}} K'_{23} T \nabla \frac{F}{T} - \frac{e^2}{m^{*2}} K'_{33} \frac{\nabla T}{T}, \mathbf{B} \right) \mathbf{B}. \end{aligned} \quad (38.2p)$$

2. Тензор кинетического коэффициента K'_{rs} определяется своими элементами K'^{ij}_{rs} в виде

$$K'^{ij}_{rs} = -\frac{1}{4\pi^3} \int_{(V_K)} \frac{E r^{-1} \tau^s}{1 + \frac{e^2 \tau^2}{m^*}} \frac{\partial f_0}{\partial E} v_i v_j d\tau_K. \quad (38.3p)$$

Физический смысл K'_{rs} в зависимости от значков, различный.

§ 39. КИНЕТИЧЕСКИЕ КОЭФФИЦИЕНТЫ

Скалярная эффективная масса. Поскольку j и W выражаются через тензоры кинетических коэффициентов K'_{rs} , необходимо уметь вычислять их для различных случаев. Предположим, что эффективная масса является скалярной величиной. Для этого случая время релаксации из соображений симметрии можно считать изотропной величиной: $\tau(\mathbf{k}) = \tau(|\mathbf{k}|)$, другими словами, *время релаксации должно зависеть только от энергии* $\tau = \tau(E)$. Энергию представим в виде квадратичной функции квазимпульса во всех точках зоны Бриллюэна, учитывая, что в выражение для K'_{rs} входит $\frac{\partial f_0}{\partial E}$, которая быстро уменьшается с ростом энергии. Скорость можно выразить через \mathbf{k} или \mathbf{P} обычным соотношением

$$\mathbf{v} = \frac{\mathbf{P}}{m^*} = \frac{\hbar \mathbf{k}}{m^*} \quad (39.1)$$

и

$$E = E_0 + \frac{\hbar^2 \mathbf{k}^2}{2m^*} = E_0 + \frac{m^* \mathbf{v}^2}{2}. \quad (39.2)$$

Запишем выражение для K'^{ij}_{rs} :

$$K'^{ij}_{rs} = -\frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \frac{E r^{-1} \tau^s}{1 + \frac{e^2 \tau^2}{m^{*2}} B^2} \frac{\partial f_0}{\partial E} v_i v_j d\tau_{\mathbf{k}}. \quad (39.3)$$

Интегрирование по объему зоны Бриллюэна сведем к интегрированию по энергии. Для этого выразим $d\tau_{\mathbf{k}}$ через энергию $dE(\mathbf{k})$:

$$dE = (\nabla_{\mathbf{k}} E, d\mathbf{k}) = |\nabla_{\mathbf{k}} E| d\kappa_n = \hbar |\mathbf{v}| d\kappa_n, \quad (39.4)$$

где $d\kappa_n$ есть проекция вектора $d\mathbf{k}$ на направление нормали к изоэнергетической поверхности. Введем элемент поверхности энергии dS_E , через который элемент объема можно выразить в виде

$$d\tau_{\mathbf{k}} = dS_E d\kappa_n = \frac{dS_E dE}{\hbar v}. \quad (39.5)$$

Подставим (39.5) в (39.3):

$$K'^{ij}_{rs} = -\frac{1}{4\pi^3} \int_{(E)} \frac{E r^{-1} \tau^s}{1 + \mu^2 B^2} \frac{\partial f_0}{\partial E} dE \int_{(S_E)} \frac{v_i v_j}{\hbar v} dS_E. \quad (39.6)$$

В (39.6) для упрощения записи введено обозначение

$$\mu = \frac{e\tau(E)}{m^*} = \mu(E), \quad (39.7)$$

где $\mu(E)$ есть подвижность носителей заряда с энергией E . Но

интеграл по поверхности можно представить в виде интеграла по пространственному углу; так как

$$dS_E = \kappa^2 d\Omega, \quad (39.8)$$

то

$$\int_{(S_E)} \frac{v_i v_j}{\hbar v} dS_E = M \int_{(S_{1E})} \frac{\hbar \kappa_i \hbar \kappa_j}{m^* \hbar^2 \frac{\kappa}{m^*}} \kappa^2 d\Omega = \frac{M}{m^*} \kappa^3 \int_{(S_{1E})} \frac{\kappa_i \kappa_j}{\kappa^2} d\Omega. \quad (39.9)$$

В (39.9) интегрирование по полной поверхности энергии сведено к интегрированию по одной S_{1E} из M поверхностей энергии. Считая, что в зоне Бриллюэна введена сферическая система координат с полярной осью, направленной вдоль κ_z , запишем

$$\frac{\kappa_x}{\kappa} = \sin \theta \cos \varphi; \quad \frac{\kappa_y}{\kappa} = \sin \theta \sin \varphi; \quad \frac{\kappa_z}{\kappa} = \cos \theta. \quad (39.10)$$

Вычислим (39.9) для двух случаев:

$$\int_{(S_{1E})} \frac{\kappa_x \kappa_x}{\kappa^2} d\Omega = \int_0^\pi \int_0^{2\pi} \sin^2 \theta \cos^2 \varphi \sin \theta d\theta d\varphi = \frac{4\pi}{3}, \quad (39.11)$$

$$\int_{(S_{1E})} \frac{\kappa_x \kappa_y}{\kappa^2} d\Omega = \int_0^\pi \int_0^{2\pi} \sin^2 \theta \cos \varphi \sin \varphi \sin \theta d\theta d\varphi = 0. \quad (39.12)$$

Проводя аналогичные вычисления, получим в общем виде

$$\int_{(S_{1E})} \frac{\kappa_i \kappa_j}{\kappa^2} d\Omega = \frac{4\pi}{3} \delta_{ij}, \quad (39.13)$$

т. е. в случае сферических поверхностей энергии тензор K'_{rs} имеет диагональный вид. Выразим интеграл по поверхности энергии через величину энергии на этой поверхности:

$$\int_{(S)} \frac{v_i v_j}{\hbar v} dS_E = \frac{M}{m^*} \kappa^3 \frac{4\pi}{3} \delta_{ij} = \frac{4\pi}{3} \frac{M}{m^*} \left[\frac{2m^* (E - E_0)}{\hbar^2} \right]^{\frac{3}{2}} \delta_{ij}. \quad (39.14)$$

Но

$$2\pi M \left(\frac{2m^*}{\hbar^2} \right)^{\frac{3}{2}} (E - E_0)^{\frac{1}{2}} = N(E) \quad (39.15)$$

есть плотность состояний по энергии, поэтому (39.14) представим в виде

$$\int_{(S_E)} \frac{v_i v_j}{\hbar v} dS_E = \frac{2}{3m^*} (E - E_0) 8\pi^3 N(E) \delta_{ij}. \quad (39.16)$$

Подставив (39.16) в (39.15), получим (при $E_0 = 0$)

$$K'_{rs}{}^{ij} = -\frac{4\delta_{ij}}{3m^*} \int_0^\infty \frac{E r \tau^s}{1 + \mu^2 B^2} N(E) \frac{\partial f_0}{\partial E} dE. \quad (39.17)$$

Выражение (39.17) показывает, что все диагональные элементы K'_{rs} одинаковы; это означает, что в случае сферических изоэнергетических поверхностей кинетические коэффициенты являются скалярными величинами: $K'_{rs} = K'_{rs}$.

1. Невырожденный полупроводник. В невырожденном полупроводнике функция Ферми — Дирака переходит в функцию Больцмана, для которой

$$\frac{\partial f_0}{\partial E} = -\frac{f_0}{kT} \quad (39.18)$$

и

$$K'_{rs} = \frac{4}{3m^*} \int_0^{\infty} \frac{Er\tau^s}{1+\mu^2 B^2} f_0(E, T) N(E) \frac{dE}{kT}. \quad (39.19)$$

Положим $r=s=0$ и $B=0$; интеграл для K_{00} вычисляется элементарно:

$$K_{00} = \frac{4}{3m^*} \int_0^{\infty} f_0(E, T) N(E) \frac{dE}{kT} = \frac{2}{3m^* kT} n, \quad (39.20)$$

где n — концентрация носителей заряда (электронов или дырок):

$$n = 2 \int_0^{\infty} f_0(E, T) N(E) dE, \quad (39.21)$$

т. е. K_{00} выражается через концентрацию носителей заряда. Но из (39.19) видно, что K'_{rs} должно быть пропорционально концентрации носителей заряда, поскольку $2 \int_0^{\infty} f_0(E, T) N(E) dE$ есть число частиц в интервале энергии dE , так что

$$K'_{rs} = \frac{2}{3m^* kT} \int_0^{\infty} \frac{Er\tau^s}{1+\mu^2 B^2} dn(E). \quad (39.22)$$

Умножим и разделим (39.19) на концентрацию частиц, получим

$$K'_{rs} = \frac{n}{m^*} \frac{2}{3kT} \frac{\int_0^{\infty} \frac{Er\tau^s}{1+\mu^2 B^2} dn(E)}{\int_0^{\infty} dn(E)}. \quad (39.23)$$

Обозначим множитель при $\frac{n}{m^*}$ (39.23) в виде

$$\frac{2}{3kT} \frac{\int_0^{\infty} \frac{Er\tau^s}{1+\mu^2 B^2} dn(E)}{\int_0^{\infty} dn(E)} = \left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle; \quad (39.24)$$

он представляет собой усредненное с весом $\frac{Er}{1+\mu^2 B^2}$ по носителям заряда время релаксации в степени s .

Выражение (39.24) можно записать в более удобной для вычисления форме, если перейти к безразмерной переменной $x = \frac{E}{kT}$:

$$\begin{aligned} \left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle &= \frac{2(kT)^{r+\frac{1}{2}} \int_0^\infty \frac{\tau^s e^{-x}}{1+\mu^2 B^2} x^{r+\frac{1}{2}} dx}{3(kT)^{\frac{3}{2}} \int_0^\infty e^{-x} x^{\frac{1}{2}} dx} = \\ &= \frac{2(kT)^{r-1} \int_0^\infty \frac{\tau^s}{1+\mu^2 B^2} e^{-x} x^{r+\frac{1}{2}} dx}{3 \int_0^\infty e^{-x} x^{\frac{1}{2}} dx}. \end{aligned} \quad (39.25)$$

Время релаксации в степени s , усредняемое с весом $\frac{Er}{1+\mu^2 B^2}$, обозначают также в виде среднего значения τ^s с числом скобок, равным r :

$$\left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle = \left\langle \underbrace{\dots \left\langle \frac{\tau^s}{1+\mu^2 B^2} \right\rangle \dots}_r \right\rangle. \quad (39.26)$$

Таким образом, кинетические коэффициенты можно выразить в виде

$$K_{rs} = \frac{n}{m^*} \left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle. \quad (39.27)$$

Если в полупроводнике имеются носители заряда различных видов, то для каждого из них можно вычислить кинетические коэффициенты в соответствии с (39.27).

Рассмотрим один частный, но практически очень важный случай, когда время релаксации является степенной функцией энергии:

$$\tau(E) = \tau_0 E^p = \frac{\tau_0}{(kT)^p} x^p = \tau_0' x^p, \quad (39.28)$$

где τ_0 и τ_0' — некоторые константы. Подставим (39.28) в (39.25):

$$\left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle = \frac{2}{3} (kT)^{r-1} \tau_0'^s \frac{\int_0^\infty \frac{x^{sp+r+1/2}}{1+\mu^2 B^2} e^{-x} dx}{\int_0^\infty x^{\frac{1}{2}} e^{-x} dx}. \quad (39.29)$$

Если положить $B=0$, то интегралы (39.29) сводятся к гамма-функции Эйлера

$$\langle Er\tau^s \rangle = (kT)^{r-1} \tau_0^s \frac{\Gamma\left(sp+r+\frac{3}{2}\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (39.30)$$

Появление $\Gamma(5/2)$ вместо $\Gamma(3/2)$ в знаменателе связано с тем, что мы вместо $2/3$ в числителе записали $3/2$ в знаменателе и учли, что $3/2 \cdot \Gamma(3/2) = \Gamma(5/2)$. Выражение (39.30) будет использовано в дальнейшем.

2. Вырожденный полупроводник. Для вырожденного полупроводника $-\frac{\partial f_0}{\partial E} \approx \delta(E-F)$, поэтому интеграл (39.17) вычисляется элементарно:

$$\begin{aligned} K'_{rs} &= \frac{4}{3m^*} \frac{Fr\tau^s(F)}{1+\mu^2(F)B^2} N(F) = \frac{8\pi}{3m^*} \left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}} F^{\frac{3}{2}} \frac{Fr^{-1}\tau^s(F)}{1+\mu^2(F)B^2} = \\ &= \frac{n}{m^*} \frac{Fr^{-1}\tau^s}{1+\mu^2(F)B^2}. \end{aligned} \quad (39.31)$$

Если записать K'_{rs} для вырожденных полупроводников в виде (39.27), то среднее значение времени релаксации можно определить соотношением

$$\left\langle \frac{Er\tau^s}{1+\mu^2 B^2} \right\rangle = \frac{Fr^{-1}\tau^s(F)}{1+\mu^2(F)B^2}, \quad (39.32)$$

т. е. усредненное значение времени релаксации определяется только временем релаксации носителей заряда, лежащих на поверхности Ферми.

Эффективная масса — тензор второго ранга. Перейдем к вычислению кинетических коэффициентов для случая, когда эффективная масса является тензорной величиной, имеющей (для простоты) диагональный вид

$$K'^{ij}_{rs} = -\frac{1}{4\pi^3} \int_{(V_k)} \frac{Er^{-1}\tau^s v_i v_j}{1 + \frac{e^2 \tau^2}{|m^*|} (B, m^* B)} \frac{\partial f_0}{\partial E} d\tau_k. \quad (39.33)$$

Предположим по-прежнему, что τ является функцией только энергии, т. е. теперь τ будем считать *анизотропной величиной*: $\tau = \tau(E)$. Выразим элемент объема $d\tau_k$ через элемент энергии dE и элемент площади поверхности энергии:

$$d\tau_k = dS_E dk_n = \frac{dS_E dE}{\hbar |\mathbf{v}|} = \frac{dS_E dE}{\hbar \sqrt{\sum_{l=1}^3 \frac{\hbar^2 \kappa_l^2}{m_l^2}}}. \quad (39.34)$$

Для вычисления интеграла по поверхности введем различный масштаб по различным направлениям в зоне Бриллюэна, т. е. введем новые переменные

$$\omega_l = \frac{\hbar \kappa_l}{\sqrt{2m_l}}; \quad \hbar \kappa_l = \sqrt{2m_l} \omega_l, \quad (39.35)$$

или

$$\kappa_l = \frac{\sqrt{2m_l}}{\hbar} \omega_l. \quad (39.36)$$

Найдем связь между энергией и ω_l :

$$E = E_0 + \sum_{l=1}^3 \frac{\hbar^2 \kappa_l^2}{2m_l} = E_0 + \sum_{l=1}^3 \omega_l^2, \quad (39.37)$$

Мы видим из (39.37), что в переменных ω_l изоэнергетическая поверхность является сферической.

Выразим K_{rs} через w . Для этого запишем

$$d\tau_{\mathbf{k}} = d\kappa_x d\kappa_y d\kappa_z = \frac{(8m_1 m_2 m_3)^{\frac{1}{2}}}{\hbar^3} d\tau_w. \quad (39.38)$$

Согласно (39.5)

$$d\tau_w = dS(w) dw_n, \quad (39.39)$$

с другой стороны,

$$dE = \left(\frac{dE}{dw} dw \right) = (wdw) = wd\omega_n = \sqrt{E} d\omega_n, \quad (39.40)$$

поэтому

$$d\tau_w = dS(w) \frac{dE}{\sqrt{E}}, \quad (39.41)$$

или

$$d\tau_{\mathbf{k}} = \frac{(8m_1 m_2 m_3)^{\frac{1}{2}}}{\hbar^3} \frac{dE dS(w)}{\sqrt{E}}. \quad (39.42)$$

Учитывая, что

$$v_l = \frac{\hbar \kappa_l}{m_l} = \sqrt{\frac{2}{m_l}} \omega_l, \quad (39.43)$$

запишем.

$$\begin{aligned} K_{rs}^{ij} &= -\frac{1}{4\pi^3} \int_{(\dot{E})} \frac{\partial f_0}{\partial E} \frac{E r^{-1} \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})} \frac{dE}{\sqrt{E}} \int_{(S)} \frac{(8|m^*|)^{\frac{1}{2}}}{\hbar^3} \sqrt{\frac{4}{m_i m_j}} \omega_i \omega_j dS = \\ &= -\frac{2(8|m^*|)^{\frac{1}{2}} M}{\hbar^3 \sqrt{m_i m_j}} \int_{(\dot{E})} \frac{\partial f_0}{\partial E} \frac{E r^{-1} \tau^s \omega^4}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, m^* \mathbf{B})} \frac{dE}{\sqrt{E}} \int_{(S_i)} \frac{\omega_i \omega_j}{\omega^2} d\Omega. \quad (39.44) \end{aligned}$$

Но согласно (39.13)

$$\int_{(S_1)} \frac{\omega_i \omega_j}{\omega^2} d\Omega = \frac{4\pi}{3} \delta_{ij}, \quad (39.45)$$

и, следовательно, K'_{rs} является диагональным тензором. Вводя выражение для плотности состояний

$$N(E) = \frac{M2\pi}{h^3} (8m_1 m_2 m_3)^{\frac{1}{2}} E^{\frac{1}{2}} \quad (E_0 = 0), \quad (39.46)$$

можем представить кинетический коэффициент в виде

$$K'_{rs} = -\frac{4\delta_{ij}}{3\sqrt{m_i m_j}} \int_0^\infty \frac{\partial f_0}{\partial E} \frac{N(E) E r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, \mathbf{m}^* \mathbf{B})} dE. \quad (39.47)$$

Для изотропной (скалярной) эффективной массы выражение (39.47) переходит в (39.17).

Выражение (39.47) для K'_{rs} можно записать в форме

$$K'_{rs} = \frac{n\delta_{ij}}{m_i} \left\langle \frac{E r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, \mathbf{m}^* \mathbf{B})} \right\rangle, \quad (39.48)$$

поскольку при $i = j$ $m_i = m_j$. Мы видим, что все кинетические коэффициенты анизотропны, при этом тензор K'_{rs} пропорционален тензору \mathbf{m}^{*-1} :

$$K'_{rs} = n \left\langle \frac{E r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, \mathbf{m}^* \mathbf{B})} \right\rangle \mathbf{m}^{*-1}. \quad (39.49)$$

Это означает, что на протекание кинетических явлений существенное влияние оказывает форма изоэнергетических поверхностей.

Резюме § 39

1. Вычисление кинетических коэффициентов удобно свести к интегрированию по поверхности энергии с последующим интегрированием по энергии. Вычисления имеют более простой вид для сферических поверхностей энергии, поэтому для упрощения расчетов эллипсоидальные поверхности энергии преобразуются в сферические соответствующим выбором масштаба вдоль осей эллипсоида.

2. Тензор кинетического коэффициента пропорционален тензору обратной эффективной массы.

3. Тензор кинетического коэффициента

$$K'_{rs} = n m^{*-1} \left\langle \frac{E r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (\mathbf{B}, \mathbf{m}^* \mathbf{B})} \right\rangle. \quad (39.1p)$$

Если магнитное поле отсутствует, то, опуская штрих, запишем

$$K_{rs} = nm^{*-1} \langle E^r \tau^s \rangle = \frac{n \langle \dots \langle \tau^s \rangle_{(r)} \dots \rangle}{m^*}. \quad (39.2p)$$

4. Усредненное время релаксации определяется для невырожденных полупроводников выражением

$$\left\langle \frac{E^r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (B, m^* B)} \right\rangle = (kT)^{r-1} \frac{\int_0^\infty \frac{\tau^s(x) x^{r+\frac{1}{2}}}{1 + \frac{e^2 \tau^2}{|m^*|} (B, m^* B)} e^{-x} dx}{\int_0^\infty x^{\frac{3}{2}} e^{-x} dx} \quad (39.3p)$$

и для вырожденных полупроводников

$$\left\langle \frac{E^r \tau^s}{1 + \frac{e^2 \tau^2}{|m^*|} (B, m^* B)} \right\rangle = \frac{F^{r-1} \tau^s(F)}{1 + \frac{e^2 \tau^2}{|m^*|} (B, m^* B)}. \quad (39.4p)$$

§ 40. ЭЛЕКТРОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

Для описания электропроводности полупроводников необходимо найти связь между плотностью тока \mathbf{j} и электрическим полем \mathbf{E} , вызывающим этот ток. Для этого необходимо положить $\mathbf{V} = 0$ и $\nabla F = \nabla T = 0$. В таком случае из (38.21) можем записать

$$\mathbf{j} = e^2 K_{11} \mathbf{E} = \sigma \mathbf{E}. \quad (40.1)$$

Из (40.1) следует, что тензор проводимости σ связан с тензором K_{11} соотношением

$$e^2 K_{11} = \sigma, \quad (40.2)$$

или

$$\sigma = e^2 \frac{n \langle E \tau \rangle}{m^*} = en \frac{e \langle \tau \rangle}{m^*} = en \mu_d. \quad (40.3)$$

Таким образом, кинетическое уравнение Больцмана не только приводит к закону Ома, но и раскрывает более глубоко смысл проводимости. Удельная проводимость и подвижность определяются усредненным с весом E временем релаксации

$$\langle \tau \rangle = \langle E \tau \rangle = \frac{\int_0^\infty x^{\frac{3}{2}} e^{-x} \tau(x) dx}{\int_0^\infty x^{\frac{3}{2}} e^{-x} dx}. \quad (40.4)$$

При степенной зависимости времени релаксации от энергии $\tau = \tau_0 E^p$, или $\tau = \tau_0 \chi^p$, (40.4) позволяет записать для невырожденного полупроводника

$$\langle \tau \rangle = \tau_0 \frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (40.5)$$

В табл. 8 приведено отношение $\frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}$ для некоторых значений p .

Таблица 8

p	0	$1/2$	1	$3/2$	2	$5/2$	$-1/2$	-1	$-3/2$	-2
$\Gamma\left(\frac{5}{2} + p\right)$	$\frac{3\sqrt{\pi}}{4}$	2	$\frac{15\sqrt{\pi}}{8}$	6	$\frac{105\sqrt{\pi}}{16}$	24	1	$\frac{\sqrt{\pi}}{2}$	1	$\sqrt{\pi}$
$\frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}$	1	$\frac{8}{3\sqrt{\pi}}$	32/5	$8/\sqrt{\pi}$	35/4	$\frac{32}{\sqrt{\pi}}$	$\frac{4}{3\sqrt{\pi}}$	2/3	$\frac{4}{3\sqrt{\pi}}$	4/3

Для вырожденного полупроводника

$$\langle \tau \rangle = \tau(F), \quad (40.6)$$

поэтому дрейфовая подвижность определяется только временем релаксации электронов или дырок, находящихся в состояниях вблизи поверхности Ферми, в то время как в проводимость входит концентрация электронов. Это противоречие легко понять, если рассмотреть выражение для функции распределения

$$f = f_0 + f^{(1)} = f_0 - \frac{\partial f_0}{\partial E}(\chi v) = f_0 - \frac{\partial f_0}{\partial E} e(Ev) \tau. \quad (40.7)$$

Для вырожденных полупроводников $\frac{\partial f_0}{\partial E} \approx 0$ при $E \neq F$ и поэтому поправка к невозмущенной функции f_0 равна нулю. Другими словами, внешнее поле E возмущает только состояния в окрестности $\pm kT$ у поверхности Ферми. Поскольку релаксация должна иметь место только для этих состояний, усредненное время релаксации $\langle \tau \rangle$ определяется величиной $\tau(F)$.

Для невырожденных полупроводников

$$-\frac{\partial f_0}{\partial E} = \frac{f_0}{kT} \quad (40.8)$$

и

$$f = f_0 + f^{(1)} = f_0 + \frac{f_0}{kT} e(Ev) \tau, \quad (40.9)$$

т. е.

$$f^{(1)} = f_0 \frac{e}{kT} (Ev) \tau \quad (40.10)$$

— различные энергетические состояния возмущаются по-разному — чем меньше энергия, тем больше f_0 и тем больше должна быть $f^{(1)}$. Однако $v = 0$ при $E = E_c$, поэтому состояния, лежащие у E_c , возмущаются меньше, чем состояния, лежащие несколько выше E_c (порядка kT). Рассмотрим относительное изменение функции распределения

$$\frac{f - f_0}{f_0} = \frac{f^{(1)}}{f_0} = \frac{e}{kT} (Ev) \tau. \quad (40.11)$$

Мы видим, что относительное изменение функции распределения с ростом энергии E , и тем самым v , возрастает. Кроме того, относительное изменение $\frac{f^{(1)}}{f_0}$ будет тем больше, чем ниже температура. Это имеет наглядный смысл. Электрическое поле E изменяет f в результате работы dA сил электрическое поле eE при движении в нем электронов. За время dt электрон проходит путь $ds = vdt$.

Разность потенциалов $d\varphi$ между точками, лежащими на расстоянии ds , равна

$$d\varphi = - (E ds) = - (Ev) dt, \quad (40.12)$$

и разность потенциальных энергий

$$dV = ed\varphi = - e(Ev) dt. \quad (40.13)$$

Если учесть, что полная энергия системы электрон — поле остается неизменной, то

$$dE + dV = 0, \quad (40.14)$$

$$dE = - dV = e(Ev) dt = dA \quad dA = - dV. \quad (40.15)$$

т. е., если электрон движется по полю $(Ev) > 0$, то его энергия уменьшается ($e < 0$), и электрон переходит на более низкие уровни энергии. Если электрон движется против поля, то $(Ev) < 0$ и $dE > 0$ — энергия такого электрона увеличивается, и электрон переходит на верхние уровни. Таким образом, электрическое поле увеличивает энергию тех электронов, которые имеют компонент скорости против поля, и уменьшает энергию электронов, имеющих компонент скорости вдоль поля. Это приводит согласно (40.11) к увеличению числа частиц, имеющих скорость против поля, и к уменьшению числа частиц, имеющих скорость (или компонент скорости) вдоль поля.

При этом изменение числа частиц будет тем больше, чем меньше f , т. е. чем свободнее уровни. Если уровни энергии заняты, то переход электронов с уровня на уровень невозможен, поэтому $f^{(1)} = 0$ для $E < F$ в вырожденных полупроводниках. Но это значит, что в вырожденных полупроводниках (и металлах) в создании тока участвуют электроны в узком энергетическом интервале порядка $F \pm kT$.

Поскольку число электронов у поверхности Ферми будет тем больше, чем больше площадь поверхности, а S_F в свою очередь будет тем больше, чем больше концентрация электронов n , число эффективных носителей заряда оказывается пропорционально полному числу электронов n , что и объясняет противоречие между приведенными выше рассуждениями и формулой (40.16)

$$\sigma = en\mu(F) = e^2\tau(F)nm^{*-1}, \quad (40.16)$$

где фигурирует полное число электронов.

Рассмотрим теперь проводимость полупроводника, имеющего M долин, или, точнее, M эллипсоидов энергии.

Если обозначить номер долины буквой ν и плотность тока, создаваемого носителями заряда этой долины, буквой $\mathbf{j}^{(\nu)}$, то полная плотность тока

$$\mathbf{j} = \sum_{\nu=1}^M \mathbf{j}^{(\nu)}. \quad (40.17)$$

Чтобы найти $\mathbf{j}^{(\nu)}$, необходимо выразить вектор \mathbf{E} в главных осях тензора m^{*-1} для ν -долины; пусть \mathbf{E} имеет вид

$$\mathbf{E} = (E_1^{(\nu)}; E_2^{(\nu)}; E_3^{(\nu)}). \quad (40.18)$$

В этом случае

$$\mathbf{j}^{(\nu)} = (j_1^{(\nu)}; j_2^{(\nu)}; j_3^{(\nu)}) = (\sigma_1^{(\nu)}E_1^{(\nu)}; \sigma_2^{(\nu)}E_2^{(\nu)}; \sigma_3^{(\nu)}E_3^{(\nu)}), \quad (40.19)$$

где

$$\sigma_i^{(\nu)} = \frac{e^2 n^{(\nu)}}{m_i} \langle \tau \rangle \quad (40.20)$$

имеет одинаковый вид для всех долин. Через $n^{(\nu)}$ мы обозначили количество носителей заряда в ν -долине (в ν -эллипсоиде). Поскольку все эллипсоиды эквивалентны, то из условия

$$\sum_{\nu=1}^M n^{(\nu)} = n \quad (40.21)$$

следует, что

$$n^{(\nu)} = \frac{n}{M}. \quad (40.22)$$

Ток \mathbf{j} можно выразить, преобразовав все тензоры проводимости $\sigma^{(\nu)}$ к одним и тем же осям координат, тогда

$$\mathbf{j} = \sum_{\nu=1}^M \sigma^{(\nu)} \mathbf{E} = \sigma \mathbf{E}, \quad (40.23)$$

где тензор проводимости σ есть сумма тензоров $\sigma^{(v)}$. Как известно, ij -компонент суммы тензоров равен сумме ij -компонентов слагаемых, которые должны быть записаны в одной и той же системе координат. Для того чтобы преобразовать тензор от одной системы координат к другой, необходимо знать взаимосвязь соответствующих координат. Напомним, что если известен тензор T в некоторой системе координат $\{x_i\}$, то тот же тензор T' в другой системе координат $\{x'_i\}$ может быть получен следующим образом:

$$T'_{ij} = \sum_{lm} \frac{\partial x'_i}{\partial x_l} \frac{\partial x'_j}{\partial x_m} T_{lm}. \quad (40.24)$$

Если

$$x'_i = \sum_l a_{il} x_l, \quad (40.25)$$

то

$$\frac{\partial x'_i}{\partial x_l} = a_{il} = \cos(x'_i, x_l) \quad (40.26)$$

и

$$T'_{ij} = \sum_{lm} a_{il} a_{jm} T_{lm}. \quad (40.27)$$

Чтобы найти тензор T' в осях координат $\{x'_i\}$, необходимо задать положение двух систем координат. Рассмотрим в качестве примера кремний как более простой случай.

За оси координат (x', y', z') для кристалла в целом возьмем три направления типа $[100]$ (рис. 60). Для 1 и 4 эллипсоидов в качестве главных осей (x, y, z) возьмем соответственно:

$$\begin{aligned} x' &= z^{(1)} = z^{(4)}, \\ y' &= x^{(1)} = x^{(4)}, \\ z' &= y^{(1)} = y^{(4)}. \end{aligned} \quad (40.28)$$

Матрицу преобразований осей можно записать в виде

$$A^{(1)} = A^{(4)} = \{a_{ij}^{(1)}\} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (40.29)$$

Для 2 и 5 эллипсоидов можно записать следующие преобразования осей:

$$\begin{aligned} x' &= y^{(2)} = y^{(5)}; \quad y' = z^{(2)} = z^{(5)}; \quad z' = x^{(2)} = x^{(5)}; \\ A^{(2)} &= A^{(5)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \tilde{A}^{(1)}. \end{aligned} \quad (40.30)$$

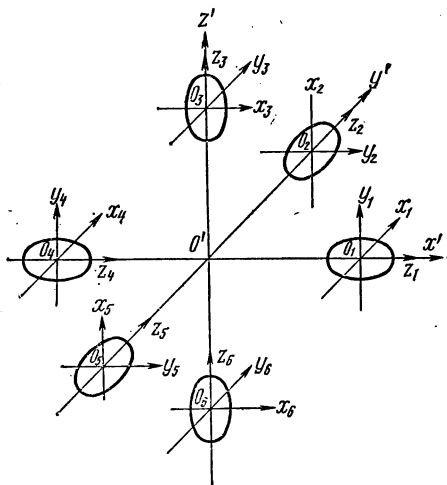


Рис. 60. Связь между осями координат изоэнергетических поверхностей зоны проводимости кремния

Для 3 и 6 эллипсоидов штрихованные и нештрихованные оси совпадают

$$x' = x^{(3)} = x^{(6)}; \quad y' = y^{(3)} = y^{(6)}; \quad z' = z^{(3)} = z^{(6)};$$

$$A^{(3)} = A^{(6)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (40.31)$$

Выбор матрицы преобразований координат определяется следующими очевидными соображениями: третья главная ось соответствует оси вращения эллипсоида, поэтому

$$z^{(1)} = z^{(4)} = x'; \quad z^{(2)} = z^{(5)} = y'; \quad z^{(3)} = z^{(6)} = z'. \quad (40.32)$$

Первая и вторая оси лежат произвольно в плоскости, перпендикулярной оси вращения; для упрощения преобразований их удобно совместить с осями штрихованными, сохраняя лишь правовинтовую систему координат. Запишем тензоры $\sigma^{(v)}$ в осях (x', y', z') , используя закон преобразования (40.27):

$$\sigma'_{ij}{}^{(v)} = \sum_{lm} a_{il} a_{jm} \sigma_{lm}^{(v)} = \sum_{lm} a_{il} a_{jm} \sigma_l^{(v)} \delta_{lm} = \sum_l a_{il} a_{jl} \sigma_l^{(v)}. \quad (40.33)$$

Опуская выкладки, запишем тензоры $\sigma'^{(v)}$ (учитывая (40.33) и (40.29–31):

$$\begin{aligned} \sigma'^{(1)} = \sigma'^{(4)} &= \begin{pmatrix} \sigma_3^{(1)} & 0 & 0 \\ 0 & \sigma_1^{(1)} & 0 \\ 0 & 0 & \sigma_2^{(1)} \end{pmatrix}; & \sigma'^{(2)} = \sigma'^{(5)} &= \begin{pmatrix} \sigma_2^{(2)} & 0 & 0 \\ 0 & \sigma_3^{(2)} & 0 \\ 0 & 0 & \sigma_1^{(2)} \end{pmatrix}; \\ \sigma'^{(3)} = \sigma'^{(6)} &= \begin{pmatrix} \sigma_1^{(3)} & 0 & 0 \\ 0 & \sigma_2^{(3)} & 0 \\ 0 & 0 & \sigma_3^{(3)} \end{pmatrix}. \end{aligned} \quad (40.34)$$

Вид тензоров $\sigma'^{(v)}$ легко получить, если учесть, что произведения элементов двух разных ($i \neq j$) строк и одного и того же столбца (l) для всех трех матриц A равны нулю, поэтому σ' остаются диагональными. Вид, например, $\sigma'^{(1)}$ легко понять из простых физических соображений: если приложить электрическое поле вдоль оси x' , то для первого и четвертого эллипсоидов оно будет направлено вдоль оси вращения, поэтому ток должен определяться компонентом σ_3 , для двух других пар эллипсоидов поле лежит в плоскости, перпендикулярной оси вращения, поэтому их вклад в проводимость определяется компонентами σ_1 и σ_2 . Найдем тензор полной проводимости, который будет также диагональным:

$$\sigma_{ij} = \sigma_i \delta_{ij},$$

где

$$\begin{aligned} \sigma_1 &= 2(\sigma_1^{(3)} + \sigma_2^{(2)} + \sigma_3^{(1)}), \\ \sigma_2 &= 2(\sigma_1^{(1)} + \sigma_2^{(3)} + \sigma_3^{(2)}), \\ \sigma_3 &= 2(\sigma_1^{(2)} + \sigma_2^{(1)} + \sigma_3^{(3)}). \end{aligned} \quad (40.35)$$

Мы видим, что диагональные компоненты полной проводимости получаются суммированием диагональных элементов разных эллипсоидов проводимости. Если эллипсоиды эквивалентны, то все $\sigma^{(v)}$ одинаковы, и

$$\begin{aligned}\sigma_1 &= 2(\sigma_1^{(v)} + \sigma_2^{(v)} + \sigma_3^{(v)}) = \sigma, \\ \sigma_2 &= 2(\sigma_1^{(v)} + \sigma_2^{(v)} + \sigma_3^{(v)}) = \sigma, \\ \sigma_3 &= 2(\sigma_1^{(v)} + \sigma_2^{(v)} + \sigma_3^{(v)}) = \sigma\end{aligned}\quad (40.36)$$

— удельная электрическая проводимость описывается диагональным тензором с равными компонентами, т. е. σ является скалярной величиной. Таким образом, *симметричное распределение анизотропных проводимостей долин приводит к изотропной полной проводимости*. Выразим проводимость σ через концентрацию и подвижность носителей заряда:

$$\begin{aligned}\sigma &= 2(\sigma_1^{(v)} + \sigma_2^{(v)} + \sigma_3^{(v)}) = 2e^2 n^{(v)} \langle \tau \rangle \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) = \\ &= 2e^2 n \langle \tau \rangle \left(\frac{2}{m_l} + \frac{1}{m_t} \right).\end{aligned}\quad (40.37)$$

Мы учли, что $m_1 = m_2 = m_t$ и $m_3 = m_l$. Введем «изотропную» эффективную массу проводимости \tilde{m}^* соотношением

$$\frac{1}{\tilde{m}^*} = \frac{1}{3} \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) = \frac{1}{3} \left(\frac{2}{m_t} + \frac{1}{m_l} \right).\quad (40.38)$$

Используя \tilde{m}^* , можем записать выражение для σ в виде

$$\sigma = \frac{6e^2 n^{(v)} \langle \tau \rangle}{\tilde{m}^*} = \frac{ene \langle \tau \rangle}{\tilde{m}^*} = en\mu_d,\quad (40.39)$$

где

$$\mu_d = \frac{e \langle \tau \rangle}{\tilde{m}^*} = \frac{e \langle \tau \rangle}{3} \left(\frac{2}{m_t} + \frac{1}{m_l} \right).\quad (40.40)$$

Мы видим, что *подвижность определяется «изотропной» эффективной массой \tilde{m}^* , связанной с формой изоэнергетических поверхностей*. \tilde{m}^{*-1} равна среднему арифметическому значению m_i^{-1} , в то время как m_d^{-1} связана со среднегеометрическим значением компонентов m_i^{-1} :

$$m_d^3 = M^2 8 m_1 m_2 m_3 = M^2 8 m_t^2 m_l.\quad (40.41)$$

Аналогичным образом можно рассмотреть проводимость электронного германия.

Рассмотрим вопрос о проводимости вещества, в котором имеются носители заряда различных типов: электроны и дырки с различными эффективными массами. Если обозначить плотность тока, соз-

даваемого носителями заряда α -типа, через j_α , то полная плотность тока j может быть записана в виде

$$j = \sum_{\alpha} j_{\alpha} = \sigma E = \left(\sum_{\alpha} \sigma_{\alpha} \right) E, \quad (40.42)$$

или

$$\sigma = \sum_{\alpha} \sigma_{\alpha} = \sum_{\alpha} e_{\alpha}^2 K_{11}(\alpha) = e^2 \sum_{\alpha} \frac{n_{\alpha} \langle \tau_{\alpha} \rangle}{m_{\alpha}^*} = \sum_{\alpha} e_{\alpha} n_{\alpha} \mu_{d\alpha}. \quad (40.43)$$

Обратимся теперь к уравнению (38.2р) для плотности потока энергии. При тех же предположениях, которые были сделаны при выводе выражения (40.1), получаем

$$W = e K_{21} E. \quad (40.44)$$

Выразим плотность потока энергии W через плотность тока j :

$$W = e K_{21} E = e K_{21} \frac{j}{e^2 K_{11}} = \Pi j. \quad (40.45)$$

Поток энергии, обусловленный направленным движением носителей зарядов, называют потоком Пельтье, а величину Π называют коэффициентом Пельтье. Из (40.45) следует, что коэффициент Пельтье выражается через кинетические коэффициенты:

$$\Pi = \frac{K_{21}}{e K_{11}} = \frac{1}{e} \frac{\langle E^2 \tau \rangle}{\langle E \tau \rangle}. \quad (40.46)$$

Резюме § 40

1. Плотность тока в однородном полупроводнике, все точки которого имеют одну и ту же температуру, без магнитного поля определяется выражением

$$j = e^2 K_{11} E, \quad (40.1p)$$

т. е. кинетический коэффициент K_{11} равен тензору удельной проводимости σ , деленной на квадрат заряда.

2. Усредненное время релаксации $\langle \tau \rangle$ выражается в невырожденном полупроводнике в случае, когда $\tau = \tau_0 E^p = \tau_0' x^p$, через гамма-функцию:

$$\langle \tau \rangle = \tau_0' \frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (40.2p)$$

3. В вырожденном полупроводнике усредненное время релаксации совпадает со временем релаксации носителей заряда, находящихся на поверхности Ферми:

$$\langle \tau \rangle \approx \tau(F). \quad (40.3p)$$

4. Из кинетического уравнения следует, что дрейфовая подвижность носителей заряда определяется усредненным по энергии временем релаксации:

$$\mu_d = \frac{e \langle \tau \rangle}{m^*}. \quad (40.4p)$$

5. Если полупроводник имеет M эквивалентных долин, то полная проводимость σ равна сумме проводимостей каждой долины:

$$\sigma = \sum_{v=1}^M \sigma^{(v)}, \quad (40.5p)$$

при этом тензоры $\sigma^{(v)}$ необходимо записать в одной и той же системе координат. Если проводимость, создаваемая каждой долиной, анизотропна, то при симметричном расположении долин в зоне Бриллюэна полная проводимость σ становится изотропной. Ее можно характеризовать изотропной эффективной массой проводимости \tilde{m}^* , которая в случае кремния (зона проводимости) связана с компонентами тензора эффективной массы в виде

$$\frac{1}{\tilde{m}^*} = \frac{1}{3} \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) = \frac{1}{3} \left(\frac{2}{m_t} + \frac{1}{m_l} \right). \quad (40.6p)$$

§ 41. ГАЛЬВАНОМАГНИТНЫЕ ЭФФЕКТЫ

Физические явления, возникающие в веществе, находящемся в магнитном поле, при прохождении через вещество электрического тока под действием электрического поля, называют гальваномагнитными эффектами. Другими словами, гальваномагнитные явления наблюдаются в веществе при совместном действии электрического и магнитного полей. К гальваномагнитным явлениям относятся: 1) эффект Холла; 2) магнитнорезистивный эффект, или магнетосопротивление; 3) эффект Эттингсгаузена, или поперечный гальванотермомагнитный эффект; 4) эффект Нернста, или продольный гальванотермомагнитный эффект. Эффект Холла в узком смысле слова называют также гальваномагнитным эффектом. Указанные выше названия «поперечный» и «продольный» гальванотермомагнитные эффекты отражают направления градиентов температуры относительно тока; по отношению к магнитному полю они могут быть поперечными или продольными.

Гальваномагнитные эффекты можно представить на основе рассмотрения движения заряженной частицы в электрическом и магнитном полях под действием силы Лоренца:

$$\mathbf{F} = e\mathbf{E} + e[\mathbf{v}\mathbf{B}] = m^* \ddot{\mathbf{r}}. \quad (41.1)$$

Напомним, что в параллельных электрическом и магнитном полях частица движется по винтовой линии с непрерывно возрастающим

шагом. Это легко понять, если учесть, что в одном магнитном поле частица, имеющая скорость v_{\parallel} вдоль поля и v_{\perp} — перпендикулярно полю, — вращается по окружности радиуса

$$r = \frac{m^* v_{\perp}}{eB} = \frac{v_{\perp}}{\omega_c} \quad (41.2)$$

с угловой скоростью $\omega_c = \frac{eB}{m^*}$ и перемещается вдоль поля со скоростью v_{\parallel} .

Поскольку электрическое поле не влияет на v_{\perp} , но меняет v_{\parallel} , становится очевидным, что движение происходит по винтовой линии с переменным шагом.

В поперечных (или скрещенных) полях E и B частица, не имеющая начальной скорости, движется по циклоиде: частица вращается по окружности радиуса

$$r = \frac{m^* E}{eB^2}, \quad (41.3)$$

центр которой движется равномерно в направлении, перпендикулярном электрическому и магнитному полям со скоростью дрейфа u_d :

$$u_d = \frac{[EB]}{B^2}. \quad (41.4)$$

Если частица имеет начальную скорость v_0 , лежащую в плоскости, перпендикулярной магнитному полю, то траекторией частицы является трахоида (удлиненная или укороченная циклоида).

Чтобы представить движение частицы в том случае, когда начальная скорость имеет составляющую вдоль магнитного поля, необходимо иметь в виду, что на эту составляющую скорости не оказывает воздействие ни электрическое, ни магнитное поля.

При движении частиц в твердом теле необходимо учесть соударения, которые нарушают направленное движение частиц под действием полей. *После каждого соударения частица будет двигаться по винтовой линии или трахоиде, которые характеризуются новыми параметрами.*

Для характеристики величины поля необходимо сравнить время релаксации с периодом вращения частицы под действием магнитного поля. Если время релаксации значительно превосходит период $\frac{2\pi}{\omega_c}$, то за время τ частица совершает несколько оборотов, двигаясь по циклоиде или винтовой линии. Это возможно при больших магнитных полях. Если частица не совершает даже одного оборота за время τ , то магнитные поля считаются малыми. Таким образом, в сильных полях

$$\frac{\omega_c \tau}{2\pi} = \frac{e\tau}{m^*} \frac{B}{2\pi} \gg 1, \quad (41.5)$$

в слабых полях

$$\frac{\omega_c \tau}{2\pi} = \frac{e\tau}{m^*} \frac{B}{2\pi} \ll 1. \quad (41.6)$$

Как видим из (41.5) и (41.6), понятие «сильные» или «слабые» поля зависит не только от величины индукции магнитного поля B , но и от подвижности носителей заряда. Условия (41.5) и (41.6) можно связать с радиусом окружности r , по которой движется частица, и длиной свободного пробега l :

$$l = v\tau; \quad r = \frac{v}{\omega_c}; \quad \frac{l}{r} = \omega_c\tau. \quad (41.7)$$

Следовательно, в слабых магнитных полях $r \gg l$ — траектория частицы искривляется незначительно, в сильных магнитных полях траектория изменяется очень сильно.

Для понимания одних явлений достаточно учесть только скорость дрейфового движения $v_d = \mu_d E$, в то время как для понимания других эффектов важно иметь в виду разброс скоростей электронов. Все это учитывается кинетическим уравнением, поэтому оно позволяет получить значительно более точное описание кинетических эффектов.

1. Эффект Холла. Рассмотрим качественно действие магнитного поля на полупроводник, по которому течет ток. Пусть полупроводник имеет вид параллелепипеда сечением $a \cdot c$ (рис. 61). Электрическое поле направлено вдоль оси x : $E = (E, 0, 0)$, магнитное поле вдоль оси y : $B = (0, B, 0)$. При включении электрического поля возникает электрический ток

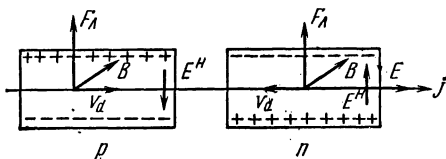


Рис. 61. Возникновение поля Холла в электронном и дырочном полупроводниках

$$j = \sigma E. \quad (41.8)$$

Носители заряда получают скорость направленного движения v_d — дрейфовую скорость — по полю для дырок и против поля для электронов. При включении магнитного поля на электроны и дырки действует сила

$$F = e [v_d B], \quad (41.9)$$

перпендикулярная v_d и B . Но

$$v_d = \mu_d E = \frac{e \langle \tau \rangle}{m^*} E, \quad (41.10)$$

поэтому

$$F = \frac{e^2 \langle \tau \rangle}{m^*} [EB], \quad (41.11)$$

т. е. сила Лоренца не зависит от знака носителей заряда, а определяется только направлением полей E и B , или j и B . Для случая, представленного на рис. 61, F направлена вверх. Носители заряда — электроны и дырки — отклоняются в одну и ту же сторону, если их скорость определяется электрическим полем.

Таким образом, в результате действия полей E , B и столкновений электроны и дырки будут двигаться по траектории в виде прямой линии, усредняющей отрезки циклоид, под углом φ к полю E . Другими словами, вектор j будет повернут на угол φ относительно вектора E , причем направление поворота зависит от знака носителей заряда (рис. 62, а, б), именно в силу того, что и электроны, и дырки отклоняются в одну и ту же сторону.

Таким образом явление должно протекать в неограниченном веществе. Если же полупроводник имеет конечные в направлении оси z размеры, то в результате того, что компонент $j_z \neq 0$, прои-

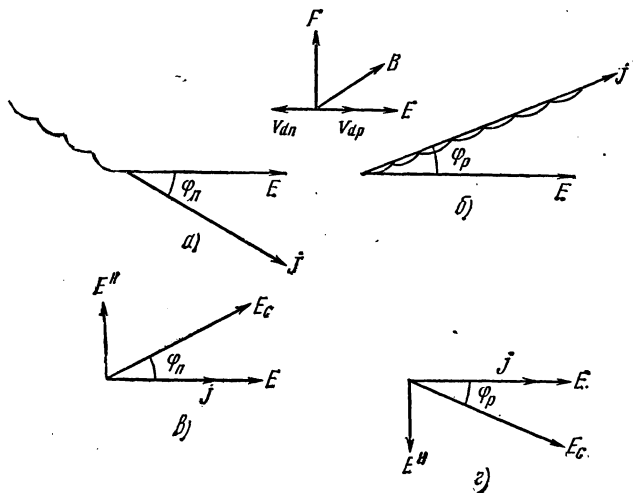


Рис. 62. Угол Холла в неограниченном (а, б) и ограниченном (в, г) полупроводниках (а, в — n -тип; б, г — p -тип)

зойдет накопление носителей заряда на верхней (в нашем случае) стороне образца и возникнет их дефицит на нижней. Противоположные стороны образца заряжаются, и возникает поперечное по отношению к F электрическое поле. Это поле носит название поля Холла, а явление возникновения поперечного поля под действием магнитного поля называют эффектом Холла. Направление поля Холла E^H зависит от знака носителей заряда, в нашем случае E^H направлено вверх в n -образце и вниз в p -образце. До наложения на образец магнитного поля эквипотенциальные поверхности представляли собой плоскости, перпендикулярные оси x , т. е. вектору j . Величина E^H будет расти до тех пор, пока поперечное поле не скомпенсирует силу Лоренца (41.9). После этого носители заряда будут двигаться как бы только под действием одного поля E , и траектория носителей заряда будет представлять собой снова прямую линию вдоль оси x (напомним, что речь идет только о направленном движении), тем самым вектор j будет направлен по полю E . Но суммарное электри-

ческое поле $E' = E + E^H$ будет повернуто на некоторый угол φ относительно оси x или j (рис. 62; в, г).

Таким образом, в неограниченном полупроводнике поворачивается вектор тока, в ограниченном — вектор электрического поля и в любом случае между j и E' (или E) возникает угол φ , называемый углом Холла. Эквипотенциальные поверхности в ограниченном образце повернуты на угол φ относительно их первоначального положения, поэтому в точках, лежащих в одной плоскости, перпендикулярной j , появляется разность потенциалов $V^H = cE^H$, где E^H — напряженность поля Холла, c — размер образца в направлении, перпендикулярном E и B ; V^H носит название холловой разности потенциалов.

Холл экспериментально нашел, что E^H определяется плотностью тока j и индукцией магнитного поля B , а также свойствами образца. Свойства образца определяются некоторой величиной R , называемой коэффициентом Холла. Четыре величины: E^H , j , B и R связаны эмпирическим соотношением

$$E^H = R [Bj] = -R [jB]. \quad (41.12)$$

Легко найти величину R , если учесть, что холлово поле должно компенсировать силу Лоренца:

$$eE^H + F = 0. \quad (41.13)$$

Из (41.13) следует

$$E^H = -\frac{1}{e} F = -\frac{1}{e} e [v_d B] = -[v_d B] = -\mu_d [EB], \quad (41.14)$$

с другой стороны, согласно (41.12)

$$E^H = -R [jB] = -R\sigma [EB]. \quad (41.15)$$

Сравнивая (41.14) и (41.15), видим, что

$$R\sigma = \mu_d \quad (41.16)$$

и

$$R = \frac{\mu_d}{\sigma} = \frac{1}{en}. \quad (41.17)$$

В данном случае через n обозначена концентрация носителей заряда (электронов или дырок). Мы видим, что коэффициент Холла обратно пропорционален концентрации носителей заряда и его знак совпадает со знаком носителей заряда.

Определив знак R , можно найти тем самым знак носителей заряда или тип проводимости. Знак же R определяется по знаку E^H , или V^H , если соответствующим образом определить знак V^H . Угол Холла φ можно определить из соотношения

$$\operatorname{tg} \varphi = \frac{E^H}{E} = -\frac{RB\sigma E}{E} = -R\sigma B = -\mu_d B. \quad (41.18)$$

При заданных значениях E и B поле Холла определяется только подвижностью носителей заряда.

В образце со смешанной проводимостью коэффициент Холла должен зависеть от характеристик носителей заряда. Легко видеть, что если в собственном образце подвижности носителей заряда одинаковы, то холлово поле будет равно нулю.

Коэффициент Холла должен зависеть от температуры. Температурная зависимость коэффициента Холла позволяет получить тем самым экспериментально температурную зависимость концентрации носителей заряда:

$$\ln n = -\ln(e_n R_n); \quad \ln p = -\ln(e_p R_p). \quad (41.19)$$

Оценим порядок величины R . Пусть $n = 10^{16} \text{ см}^{-3}$, тогда

$$R = \frac{1}{1,6 \cdot 10^{-19} \cdot 10^{16}} \frac{\text{см}^3}{\text{Кл}} \approx 600 \frac{\text{см}^3}{\text{Кл}} = 6 \cdot 10^{-4} \frac{\text{м}^3}{\text{Кл}}.$$

В табл. 9 приведены значения R (в $\text{см}^3/\text{Кл}$) для некоторых значений концентрации носителей заряда (в см^{-3}).

Таблица 9

$\frac{n}{\text{см}^{-3}}$	10^{10}	10^{13}	10^{15}	$6,25 \cdot 10^{15}$	10^{16}	$6,25 \cdot 10^{16}$	10^{17}	$6,25 \cdot 10^{17}$	
$\text{см}^3 \text{Кл}^{-1} R$	$6,25 \cdot 10^8$	$6,25 \cdot 10^5$	$6,25 \cdot 10^3$	10^3	625	100	62,5	10	
$\frac{n}{\text{см}^{-3}}$	10^{18}	$6,25 \cdot 10^{18}$	10^{19}	$6,25 \cdot 10^{19}$	10^{20}	$6,25 \cdot 10^{20}$	10^{21}	$6,25 \cdot 10^{21}$	10^{22}
$\text{см}^3 \text{Кл}^{-1} R$	6,25	1	0,625	0,1	0,063	0,01	0,006	0,001	0,0006

Мы определили R через поле E^H , однако экспериментально определяют холлову разность потенциалов $V^H = E^H c$, а вместо плотности тока измеряют ток $I = j a c$, поэтому R можно выразить в виде

$$R = \frac{E^H}{jB} = \frac{V^H}{\frac{I}{ac} B} = \frac{V^H a}{IB}. \quad (41.20)$$

Если для измерений входящих в формулу (41.20) величин использовать *вольт, метр, ампер, тесла*, то получим R в $\text{м}^3/\text{Кл}$. Если использовать несистемные единицы — *вольт, сантиметр, ампер, гаусс*, а R выражать в $\text{см}^3/\text{Кл}$, то, учитывая связь между единицами: $1 \text{ м} = 10^3 \text{ см}$; $1 \text{ Т} = 10^4 \text{ Гс}$, получим

$$R \left(\frac{\text{см}^3}{\text{Кл}} \right) = 10^8 \frac{V^H (\text{В}) a (\text{см})}{I (\text{А}) B (\text{Гс})}. \quad (41.21)$$

2. Магнетосопротивление, эффект Гаусса. Магнитное поле приводит не только к появлению холлова угла между j и E , но и влияет на величину электропроводности. Без магнитного

поля частица движется прямолинейно и между двумя столкновениями проходит путь, равный длине свободного пробега l .

Если включить магнитное поле, то траектория будет представлять собой в неограниченном образце участок циклоиды длиной l , и за время свободного пробега вдоль поля E частица пройдет путь, меньший чем l , а именно,

$$l_x \approx l \cos \varphi \approx l \left(1 - \frac{\varphi^2}{2}\right) \approx l \left(1 - \frac{\mu^2 B^2}{2}\right). \quad (41.22)$$

Поскольку за время τ частица проходит меньший путь вдоль поля E , то это равносильно уменьшению дрейфовой скорости, или подвижности, а тем самым и проводимости, т. е. *сопротивление должно возрасть*. Очевидно, что

$$\frac{\sigma_0 - \sigma}{\sigma_0} = \frac{l_0 - l}{l_0} = \frac{\mu^2 B^2}{2}, \quad (41.23)$$

или

$$\frac{\rho - \rho_0}{\rho_0} = \frac{\mu^2 B^2}{2}. \quad (41.24)$$

Если учесть статистический разброс времен (и длин) свободного пробега, то получим

$$\frac{\Delta \rho}{\rho_0} = \mu^2 B^2. \quad (41.25)$$

Таким образом, *сопротивление в магнитном поле возрастает*.

Если рассмотреть ограниченный полупроводник, то холлово поле компенсирует действие магнитного поля, в результате чего носители заряда движутся прямолинейно, поэтому магнетосопротивление с этой точки зрения должно отсутствовать. На самом деле оно имеет место и в этом случае, *поскольку холлово поле компенсирует действие магнитного поля лишь в среднем, как если бы все носители заряда двигались с одной и той же (дрейфовой) скоростью*. Однако скорости электронов (и дырок) различны, поэтому на частицы, движущиеся со скоростями, большими средней скорости, сильнее действует магнитное поле, чем холлово. Наоборот, более медленные частицы отклоняются под действием преобладающего холлово поля. В результате разброса частиц по скоростям уменьшается вклад в проводимость быстрых и медленных носителей заряда, что приводит к увеличению сопротивления, но в значительно меньшей степени, чем в неограниченном полупроводнике. Эффект магнетосопротивления оказывается чувствительным к форме образца. Неограниченный образец можно моделировать в виде диска (диск Корбино). Так как ток имеет радиальный характер, то отклонение носителей заряда под действием магнитного поля происходит в перпендикулярном к радиусу направлении, поэтому не происходит разделения и накопления зарядов, и холлово поле не возникает.

Если магнитное поле направлено вдоль j , то в этом случае изменения сопротивления не должно было бы быть. Однако в ряде

веществ наблюдается магнетосопротивление, что объясняется сложной формой изоэнергетических поверхностей. В некоторых веществах при определенных условиях увеличивается проводимость при наложении магнитного поля; это явление называют *отрицательным магнетосопротивлением*.

3. Эффект Эттингсгаузена. Если в среднем действие силы Лоренца и поля Холла компенсируют друг друга, то *вследствие разброса скоростей носителей заряда отклонение «более горячих» и «более холодных» происходит по-разному — они отклоняются к противоположным граням полупроводника*. Электроны, сталкиваясь с решеткой, приходят с ней в термодинамическое равновесие. Если они при этом отдают энергию, то полупроводник нагревается; если они отбирают энергию у решетки, то полупроводник охлаждается, в результате чего *возникает градиент температуры в направлении, перпендикулярном полю \mathbf{B} и току \mathbf{j} (эффект Эттингсгаузена)*. Для характеристики эффекта Эттингсгаузена служит коэффициент Эттингсгаузена A^E :

$$\nabla T^E = -A^E [\mathbf{Bj}] = A^E [\mathbf{jB}], \quad (41.26)$$

или

$$\frac{\partial T^E}{\partial z} = A^E B_y j_x; \quad A^E = \frac{\nabla_z T^E}{B_y j_x}. \quad (41.27)$$

При изменении направления магнитного поля (так же, как и тока) знак $\nabla_z T^E$ меняется подобно тому, как меняется знак поля Холла. *Эффекты Холла и Эттингсгаузена, зависящие от направления \mathbf{B} , называют нечетными*.

4. Эффект Нернста, или продольный гальванотермомангнитный эффект. Этот эффект состоит в том, что *вдоль тока \mathbf{j} возникает градиент температур, который не зависит от направления магнитного поля, но меняет знак при изменении направления тока*. Эффект Нернста возникает в результате того, что в направлении тока уменьшается поток «горячих» и «холодных» электронов. *Эффекты Гаусса и Нернста являются четными*.

В заключение этого параграфа укажем, что *гальваномагнитные явления делят на адиабатические и изотермические. Явления называют адиабатическими, если образец не обменивается энергией с окружающей средой, и изотермическими, если в результате обмена энергией с окружающей средой в направлении, перпендикулярном полю \mathbf{B} и току \mathbf{j} , не возникает градиента температур*.

Резюме § 41

1. Явления, возникающие в полупроводнике с током при помещении его в магнитное поле, называют гальваномагнитными. Их называют адиабатическими, если полупроводник не обменивается тепловой энергией через боковую поверхность с окружающей средой. Если в результате обмена энергией с окружающей средой

в полупроводнике поддерживается одинаковая температура в плоскости, перпендикулярной току, то эффекты называют изотермическими.

2. Эффект Холла состоит в появлении поперечного электрического поля E^H :

$$E^H = R [Bj] = -R [jB]. \quad (41.1p)$$

Величину

$$R = -\frac{E^H}{[jB]} \quad (41.2p)$$

называют коэффициентом Холла.

3. Магнетосопротивлением, или магнитнорезистивным эффектом, называют изменение сопротивления полупроводника под действием магнитного поля. Для характеристики магнетосопротивления служит коэффициент магнетосопротивления H , определяемый соотношением

$$H = \frac{1}{B^2} \frac{\rho - \rho_0}{\rho_0}. \quad (41.3p)$$

4. Эффект Эттингсгаузена, или поперечный гальванотермомагнитный эффект, состоит в возникновении поперечного градиента температуры

$$\nabla T^E = A^E [jB]. \quad (41.4p)$$

Величину A^E называют коэффициентом Эттингсгаузена. Он может быть определен соотношением

$$A^E = \frac{\nabla_z T^E}{B_y j_x}. \quad (41.5p)$$

5. Эффект Нернста, или продольный гальванотермомагнитный эффект, состоит в появлении продольного относительно тока градиента температур

$$\frac{\partial T^N}{\partial x} = A^N B_y^2 j_x. \quad (41.6p)$$

6. При изменении направления тока все гальваномагнитные эффекты меняют знак, т. е. меняется направление полей и градиентов температуры.

7. При изменении знака магнитного поля знаки эффектов Холла и Эттингсгаузена меняются на противоположные (нечетные, или поперечные, эффекты), а знаки магнитнорезистивного эффекта и эффекта Нернста сохраняются (четные, или продольные, эффекты).

8. Движение (дрейф) электронов и дырок при заданном направлении тока противоположно, а их отклонение в магнитном поле, обусловленное дрейфовой скоростью, одинаково. Благодаря этому поле Холла в электронном и дырочном полупроводниках имеет противоположный знак, а эффекты Эттингсгаузена, Нернста и магнетосопротивление от знака носителей заряда не зависят.

9. Эффект Эттингсгаузена может быть только адиабатическим, остальные могут быть как адиабатическими, так и изотермическими.

10. Знак эффектов Эттингсгаузена и Нернста зависит от механизма рассеяния, так как

$$\varphi = \omega_c \tau = \omega_c \tau_0 E^p \quad (41.7p)$$

и (при малых φ)

$$\frac{\Delta l}{l} \approx \frac{\varphi^2}{2} = \frac{\omega_c^2 \tau_0^2}{2} E^{2p}, \quad (41.8p)$$

поэтому

$$\frac{d}{dE} \left(\frac{\Delta l}{l} \right) = 2p \omega_c^2 \tau_0^2 E^{2p-1}. \quad (41.9p)$$

При $p=0$ относительное изменение длины свободного пробега носителей заряда не зависит от энергии, следовательно, разброс в скоростях не должен сказываться на составе тока при наложении магнитного поля, поэтому эффекты Эттингсгаузена и Нернста должны отсутствовать. Если $p > \frac{1}{2}$, то относительно уменьшается роль «горячих» электронов, и полупроводник должен охлаждаться вдоль направления движения носителей заряда: $\nabla_x T$ совпадает с током j_x в электронном полупроводнике и противоположен току в дырочном полупроводнике. При $p < 0$ увеличивается доля «горячих» электронов в составе тока, и знак градиента температуры изменится по сравнению с предыдущим случаем.

§ 42. ЭФФЕКТ ХОЛЛА В ОБЛАСТИ ПРИМЕСНОЙ ПРОВОДИМОСТИ

Перейдем к описанию эффекта Холла на основе кинетического уравнения. Пусть на вещество наложено электрическое \mathbf{E} и магнитное \mathbf{B} поля. Будем считать, что полупроводник однороден и изотермичен: $\nabla F = \nabla T = 0$.

Согласно (38.21) выражение для тока можно представить в виде

$$\mathbf{j} = e^2 K'_{11} \mathbf{E} + \frac{e^3}{m^*} K'_{12} [\mathbf{E}\mathbf{B}] + \frac{e^4}{m^{*2}} K'_{13} \mathbf{B} (\mathbf{E}\mathbf{B}). \quad (42.1)$$

Ограничимся случаем поперечных полей: $(\mathbf{E}\mathbf{B}) = 0$, поэтому (42.1) упростится

$$\mathbf{j} = e^2 K'_{11} \mathbf{E} + \frac{e^3}{m^*} K'_{12} [\mathbf{E}\mathbf{B}]. \quad (42.2)$$

Для решения уравнения (42.2) необходимо задать дополнительные условия, аналогичные граничным условиям при решении дифференциальных уравнений.

Рассмотрим вначале неограниченный полупроводник. Наложим на него поля $\mathbf{E} = (E_x, 0, 0)$ и $\mathbf{B} = (0, B_y, 0) = (0, B, 0)$. В этом случае в полупроводнике возникает ток \mathbf{j} , имеющий отличные от нуля компоненты j_x и j_z :

$$j_x = e^2 K'_{11} E_x; \quad j_y = 0; \quad j_z = \frac{e^3}{m^*} K'_{12} B E_x, \quad (42.3)$$

Наличие компонента j_z обусловлено силой Лоренца, действующей на носители заряда вследствие их дрейфовой скорости. Механизм рассеяния проявляется посредством кинетического коэффициента K_{12} , дрейфовая скорость проявляется посредством зависимости j_z от E_x . Угол φ между полным током \mathbf{j} и полем \mathbf{E} назовем углом Холла. Он вызывается поворотом вектора тока \mathbf{j} относительно поля \mathbf{E} . Его можно определить условием

$$\operatorname{tg} \varphi = \frac{j_z}{j_x} = \frac{e}{m^*} \frac{K'_{12}}{K'_{11}} B = \mu^H B. \quad (42.4)$$

Величина μ^H определяет угол Холла и имеет размерность подвижности; она называется холловой подвижностью:

$$\mu^H = \frac{e}{m^*} \frac{K'_{12}}{K'_{11}}. \quad (42.5)$$

Ее знак определяется знаком носителей заряда. При $E_x > 0$ и $B_y > 0$ компонент $j_z > 0$ для дырок и $j_z < 0$ для электронов. Условие $j_z \neq 0$ означает соответственно поворот вектора \mathbf{j} против или по часовой стрелке.

Если полупроводник ограничен, то явление протекает иначе. Предположим, что образец имеет вид параллелепипеда, ребра которого направлены вдоль осей координат. Очевидно, что ток будет идти только вдоль оси x в соответствии со схемой подсоединения полупроводника в электрическую цепь. Компоненты тока j_y и j_z равны нулю. Но это возможно только в том случае, если поле \mathbf{E} будет иметь отличными от нуля не только E_x , но и другие компоненты. Действительно, если положим $\mathbf{E} = (E_x, 0, 0)$ и одновременно $\mathbf{B} = (0, B, 0)$ и $\mathbf{j} = (j_x, 0, 0)$, то из (42.3) следует, что при $j_z = 0$ должно быть $E_x = 0$ и $j_x = 0$. Это означает, что три наложенных выше условия для ограниченного полупроводника являются несовместными. Поэтому будем считать, что

$$\mathbf{E} = (E_x, E_y, E_z); \quad \mathbf{B} = (0, B, 0); \quad \mathbf{j} = (j_x, 0, 0), \\ [\mathbf{E}\mathbf{B}] = (-E_z B, 0, E_x B).$$

Запишем выражение (42.2) для тока:

$$j_x = e^2 K'_{11} E_x - \frac{e^3}{m^*} K'_{12} B E_z \neq 0; \quad (42.6)$$

$$j_y = e^2 K'_{11} E_y = 0; \quad E_y = 0; \quad (42.7)$$

$$j_z = e^2 K'_{11} E_z + \frac{e^3}{m^*} K'_{12} B E_x = 0. \quad (42.8)$$

Найдем E_z из (42.8):

$$E_z = - \frac{e}{m^*} \frac{K'_{12}}{K'_{11}} B E_x = - \mu^H B E_x. \quad (42.9)$$

Возникновение поля E_z представляет собой эффект Холла, т. е. E_z есть холлово поле. Полное поле E поворачивается на угол φ , определяемый условием

$$\operatorname{tg} \varphi = \frac{E_z}{E_x} = -\mu^H B. \quad (42.10)$$

Условие (42.10) совпадает с условием (42.4), но имеет противоположный знак. Это естественно, ибо если j поворачивается по часовой стрелке по отношению к E , то вектор E должен повернуться против часовой стрелки по отношению к j . Угол φ , определяемый условием (42.10), будем также называть углом Холла. Выразим теперь j_x через E_x , для этого подставим в (42.6) вместо E_z его выражение (42.9):

$$j_x = e^2 K'_{11} E_x + \frac{e^4}{m^{*3}} \frac{K'_{12} B^2}{K'_{11}} E_x = \sigma_B E_x, \quad (42.11)$$

где

$$\sigma_B = e^2 K'_{11} + \frac{e^4}{m^{*3}} \frac{K'_{12} B^2}{K'_{11}} = e^2 K'_{11} [1 + \mu^{H^2} B^2] \quad (42.12)$$

есть удельная электрическая проводимость в направлении поля E_x при наложении магнитного поля.

Введем теперь коэффициент Холла. Согласно (42.12) R вводится условием

$$R = -\frac{E^H}{[jB]}. \quad (42.13)$$

Но

$$E^H = (0, 0, E_z); \quad [jB] = (0, 0, j_x B), \quad (42.14)$$

поэтому из (42.13) и (42.14) следует

$$E_z = -R j_x B; \quad R = -\frac{E_z}{j_x B}; \quad (42.15)$$

с другой стороны, из (42.9) и (42.11)

$$E_z = -\mu^H B E_x = -\mu^H B \frac{j_x}{\sigma_B}. \quad (42.16)$$

Сравнивая (42.15) и (42.16), можем записать

$$R = \frac{\mu^H}{\sigma_B}, \quad (42.17)$$

или

$$\mu^H = R \sigma_B. \quad (42.18)$$

Подставляя μ^H из (42.5) и σ_B из (42.12), получим:

$$R = \frac{\mu^H}{\sigma_B} = \frac{\mu^H}{e^2 K'_{11} [1 + \mu^{H^2} B^2]} = \frac{\frac{e}{m^*} \frac{K'_{12}}{K'_{11}}}{e^2 K'_{11} \left[1 + \frac{e^2}{m^{*3}} \left(\frac{K'_{12}}{K'_{11}} \right)^2 B^2 \right]} \quad (42.19)$$

Выражение (42.19) для R справедливо при любых магнитных полях (если не учитывать изменения спектра энергии) при наличии носителей заряда одного знака, т. е. в области примесной проводимости. Так как

$$K'_{11} = \frac{n}{m^*} \left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle, \quad (42.20)$$

$$K'_{12} = \frac{n}{m^*} \left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle, \quad (42.21)$$

то

$$R = \frac{\frac{e}{m^*} \left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle}{\frac{e^2 n}{m^*} \left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle \left[1 + \frac{e^2}{m^{*2}} \frac{\left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle^2}{\left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle^2} B^2 \right]}. \quad (42.22)$$

Это выражение показывает, что коэффициент Холла R зависит сложным образом от магнитного поля, он обратно пропорционален концентрации носителей заряда и его знак совпадает со знаком носителей заряда.

Рассмотрим два предельных случая — случай слабых и сильных полей.

Учитывая, что $\mu = \mu(E) = \frac{e\tau(E)}{m^*}$, перепишем (42.44) в таком виде:

$$R = \frac{1}{en} \frac{\left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle^2 + \left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle^2 B^2}. \quad (42.23)$$

Слабое магнитное поле определяется условием

$$\mu^2 B^2 \ll 1. \quad (42.24)$$

Сильное магнитное поле определяется противоположным условием

$$\mu^2 B^2 \gg 1. \quad (42.25)$$

По сравнению с условиями (41.5) и (41.6) условия (42.24) и (42.25) более слабые, они удовлетворяются значительно легче.

1. Слабое магнитное поле. В слабом магнитном поле

$$K'_{11} = \frac{n}{m^*} \left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle \approx \frac{n}{m^*} \langle \tau \rangle = K_{11};$$

$$K'_{12} \approx \frac{n}{m^*} \langle \tau^2 \rangle = K_{12}. \quad (42.26)$$

Кроме того,

$$\frac{e^2}{m^{*2}} \left(\frac{K'_{12}}{K'_{11}} \right)^2 B^2 \approx \frac{e^2}{m^{*2}} \left(\frac{K_{12}}{K_{11}} \right)^2 B^2 = \frac{e^2}{m^{*2}} \left(\frac{\langle \tau^2 \rangle}{\langle \tau \rangle} \right)^2 B^2 =$$

$$= \frac{\langle \tau^2 \rangle^2}{\langle \tau \rangle^4} \langle \mu^2 B^2 \rangle \ll 1, \quad (42.27)$$

так как $\frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}$ порядка единицы. Учитывая (42.27), выражение для R можем записать в виде

$$R = \frac{e K_{12}}{m^* K_{11}} = \frac{1}{em^*} \frac{K_{12}}{K_{11}^2} = \frac{1}{en} \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = \frac{A}{en}. \quad (42.28)$$

Выражение (42.28) для R в слабом магнитном поле не зависит от B и совпадает с точностью до константы A (холл-фактор)

$$A = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} \quad (42.29)$$

с выражением (41.17), полученным на основе наглядных упрощенных представлений об эффекте Холла. Величина A зависит от механизма рассеяния. Если $\tau = \tau_0 E^p = \tau_0' x^p$, то для невырожденных полупроводников

$$A = \frac{\Gamma\left(\frac{5}{2} + 2p\right)}{\Gamma\left(\frac{5}{2}\right)} \frac{\Gamma\left(\frac{5}{2} + 2p\right) \Gamma\left(\frac{5}{2}\right)}{\left[\frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}\right]^2} = \frac{\Gamma\left(\frac{5}{2} + 2p\right) \Gamma\left(\frac{5}{2}\right)}{\left[\Gamma\left(\frac{5}{2} + p\right)\right]^2}. \quad (42.30)$$

В табл. 10 дано значение холл-фактора A для разных p .

Таблица 10

p	0	1/2	1	3/2	-1/2	-1	2
A	1	$\frac{45\pi}{128} = 1,13$	$\frac{7}{5} = 1,40$	$\frac{315\pi}{512} = 1,93$	$\frac{3\pi}{8} = 1,18$	$\frac{27\pi}{16} = 5,30$	$\frac{99}{35} = 2,83$

Как будет показано в следующей главе, при рассеянии на ионах примеси $p = 3/2$ и $A = \frac{315\pi}{512} = 1,93$; при рассеянии на тепловых колебаниях решетки $p = -1/2$ и $A = \frac{3\pi}{8} = 1,18$. Поскольку тот или другой механизм рассеяния определяется температурным интервалом, можно сказать, что в области низких температур, когда основную роль играет рассеяние на ионах примеси, необходимо полагать $A = 1,93$. В области температур, когда основную роль в рассеянии играют тепловые колебания решетки, $A = 1,18$.

Холлова подвижность в слабом поле

$$\mu^H = \frac{e K_{12}}{m^* K_{11}} = \frac{e}{m^*} \frac{\langle \tau^2 \rangle}{\langle \tau \rangle}. \quad (42.31)$$

Так как дрейфовая подвижность в слабом поле имеет вид

$$\mu_d = \frac{e}{m^*} \langle \tau \rangle, \quad (42.32)$$

то

$$\mu^H = \frac{e \langle \tau \rangle \langle \tau^2 \rangle}{m^* \langle \tau \rangle^2} = A \mu_d, \quad (42.33)$$

т. е. холлова подвижность, определяющая угол Холла, пропорциональна дрейфовой подвижности.

Рассмотрим величину σ_B . С учетом (42.27) (42.12) ее можно записать в виде (слабое поле!)

$$\sigma_B = e^2 K'_{11} [1 + \mu^H B^2] \approx e^2 K_{11} = \sigma_0. \quad (42.34)$$

Из (42.18) следует

$$\mu^H = R \sigma_0. \quad (42.35)$$

Определив экспериментально R и σ_0 и взяв их произведение, получим μ^H . Если известен механизм рассеяния, то по μ^H можно определить $\mu_d = \frac{\mu^H}{A}$, а по R — концентрацию носителей заряда и их знак; благодаря этому эффект Холла является одним из важнейших методов исследования полупроводников.

Для вырожденных полупроводников и металлов

$$A = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = \frac{\tau^2 (F)}{[\tau (F)]^2} = 1, \quad (42.36)$$

поэтому

$$\mu^H = \mu_d; \quad R = \frac{1}{en}. \quad (42.37)$$

2. Сильные поля. В сильных полях

$$\frac{K'_{12}}{K'_{11}} = \frac{\left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle} \approx \frac{\left\langle \tau^2 \left(\frac{e^2 \tau^2}{m^{*2}} \right)^{-1} \right\rangle \frac{1}{B^2}}{\left\langle \tau \left(\frac{e^2 \tau^2}{m^{*2}} \right)^{-1} \right\rangle \frac{1}{B^2}} = \frac{1}{\left\langle \frac{1}{\tau} \right\rangle} = \langle \tau^{-1} \rangle^{-1} \quad (42.38)$$

$$\text{и} \quad \left(\frac{K'_{12}}{K'_{11}} \right)^2 \approx \langle \tau^{-1} \rangle^{-2}. \quad (42.39)$$

Учитывая (42.39), перепишем выражение (42.19) для сильного поля

$$R \approx \frac{\langle \tau^{-1} \rangle^{-1}}{en \cdot \langle \tau^{-1} \rangle \left(\frac{m^{*2}}{e^2} \cdot \frac{1}{B^2} + \langle \tau^{-1} \rangle^{-2} \right)}. \quad (42.40)$$

При сколь угодно больших полях первое слагаемое знаменателя (42.40) может быть много меньше второго, поэтому

$$R = \frac{1}{en}, \quad (42.41)$$

т. е. в сильном магнитном поле коэффициент Холла не зависит от механизма рассеяния ($A=1$), как в вырожденном полупроводнике. Сравнивая выражения (42.41) и (42.28) для R в сильных и слабых полях, мы видим, что R зависит от B : с ростом поля величина R уменьшается от $\frac{A}{en}$ до $\frac{1}{en}$. Переход происходит в области полей, для которых

$$\mu^2 B^2 \approx 1. \quad (42.42)$$

Рассматривая отношение величин R при малых полях — R_0 — и при больших полях — R_∞ —, легко получить величину A :

$$A = \frac{R_0}{R_\infty}. \quad (42.43)$$

В заключение заметим, что в понятие сильного или слабого поля входит подвижность носителей заряда с данной энергией. Но если $\tau = \tau_0 E^p$ и $p > 0$, то условие $\mu^2 B^2 \ll 1$ формально не может быть выполнено ни при каких полях для всех возможных значений энергии. Однако это не так. Как число, так и вклад носителей заряда больших энергий в образование усредненных значений времен релаксаций будут незначительными. Поэтому условия (42.24) и (42.25) необходимо понимать в том смысле, что они должны выполняться, если под μ понимать дрейфовую (или холлову) подвижность, т. е. оценкой величины поля должны быть условия

$$\mu_d^2 B^2 \ll 1; \mu_d^2 B^2 \gg 1. \quad (42.44)$$

Рассмотрим пример. В InSb $\mu_d \approx 8 \text{ м}^2/(\text{В} \cdot \text{с})^{-1}$, поэтому слабое поле должно удовлетворять условию $B^2 \ll 1/64 = 0,015$ и $B < 0,1 \text{ Т} = 1000 \text{ Гс}$. В $p\text{-Si}$ $\mu_d \sim 0,05 \frac{\text{м}^2}{\text{В} \cdot \text{с}}$ и практически достижимые поля являются слабыми.

Резюме § 42

1. Из уравнения для плотности тока

$$\mathbf{j} = e^2 K'_{11} \mathbf{E} + \frac{e^3}{m^*} K'_{12} [\mathbf{E} \mathbf{B}] \quad (42.1p)$$

с «граничными» условиями

$$\mathbf{j} = (j_x, 0, 0) \quad (42.2p)$$

вытекает выражение для поперечного поля E_z — поля Холла:

$$E_z = -\frac{\mu^H}{\sigma_B} B_y j_x = -R B_y j_x, \quad (42.3p)$$

откуда для коэффициента Холла

$$R = \frac{\mu^H}{\sigma_B}. \quad (42.4p)$$

2. Подвижность Холла μ^H определяет напряженность холлова поля через напряженность электрического поля E_x , вызывающего ток, и индукцию магнитного поля $B = B_y$:

$$\mu^H = \frac{E_z}{B_y E_x}, \quad (42.5p)$$

угол Холла φ определяется холловой подвижностью μ^H :

$$\operatorname{tg} \varphi = \frac{E_z}{E_x} = \mu^H B. \quad (42.6p)$$

Холлова подвижность выражается через кинетические коэффициенты:

$$\mu^H = \frac{e}{m^*} \frac{K'_{12}}{K'_{11}} = \frac{e}{m^*} \frac{\left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle}. \quad (42.7p)$$

3. В слабом магнитном поле $\mu_d^2 B^2 \ll 1$

$$\sigma_B \approx \sigma_0 = en\mu_d; \quad \mu^H = \frac{e}{m^*} \frac{\langle \tau^2 \rangle}{\langle \tau \rangle} = A\mu_d; \quad R = \frac{A}{en}. \quad (42.8p)$$

4. В сильном магнитном поле

$$R = \frac{1}{en}; \quad \mu^H = \frac{e}{m^*} \langle \tau^{-1} \rangle^{-1}, \quad (42.9p)$$

т. е. коэффициент Холла не зависит от механизма рассеяния.

5. Механизм рассеяния влияет на значение R посредством множителя A ; в слабом поле холл-фактор

$$A = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}, \quad (42.10p)$$

что дает $\frac{3\pi}{8}$ при рассеянии на тепловых колебаниях решетки, $\frac{315\pi}{512}$ при рассеянии на ионах примеси и 1 в вырожденных полупроводниках и металлах. В сильном поле $A = 1$.

6. Концентрация носителей заряда

$$n = \frac{A}{eR}. \quad (42.11p)$$

§ 43. ЭФФЕКТ ХОЛЛА В ВЕЩЕСТВЕ С НОСИТЕЛЯМИ ЗАРЯДА НЕСКОЛЬКИХ ТИПОВ

В этом параграфе рассмотрим эффект Холла в однородном полупроводнике в изотермических условиях в предположении, что в нем имеется несколько типов носителей заряда. Это может быть собственный полупроводник, полупроводник в области смешанной проводимости или полупроводник в области примесной проводимости при

наличии носителей заряда нескольких масс, например, p -Si или p -Ge, в которых имеются легкие и тяжелые дырки.

Для носителей заряда α -вида можно записать обычное выражение для тока:

$$j_{\alpha} = e_{\alpha}^2 K'_{11}(\alpha) E + \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) [EB]. \quad (43.1)$$

Полный ток

$$j = \sum_{\alpha} j_{\alpha} = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) E + \left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right) [EB]. \quad (43.2)$$

Все выражения предыдущего параграфа должны быть справедливы, если вместо K'_{11} и K'_{12} подставить соответствующие суммы. Рассмотрим решение уравнения (43.2) для ограниченного полупроводника:

$$V = (0, B, 0); \quad E = (E_x, E_y, E_z); \quad [EB] = (-E_z B, 0, E_x B); \quad j = (j_x, 0, 0). \quad (43.3)$$

Коэффициент Холла R определяется условием (42.15). Найдем

$$j_x = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) E_x - \left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right) B E_z; \quad (43.4)$$

$$j_y = 0, \quad E_y = 0; \quad (43.5)$$

$$j_z = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) E_z + \left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right) B E_x = 0. \quad (43.6)$$

Из (43.6) определим E_z :

$$E_z = - \frac{\left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right) B E_x}{\left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right)}. \quad (43.7)$$

Это выражение можно представить аналогично (42.9), если определить подвижность Холла μ^H тем же условием:

$$\operatorname{tg} \varphi = \frac{E_z}{E_x} = - \mu^H B. \quad (43.8)$$

Тогда

$$\mu^H = \frac{\left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right)}{\left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right)} \quad (43.9)$$

и

$$E_z = - \mu^H B E_x. \quad (43.10)$$

Выразим E_z через j_x :

$$j_x = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) E_x + \frac{\left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right)^2}{\left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right)} B^2 E_x = \sigma_B E_x, \quad (43.11)$$

где

$$\sigma_B = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) \{1 + \mu^H B^2\} \quad (43.12)$$

есть проводимость в направлении тока при наложении магнитного поля. Из (42.15), (43.10) и (43.11) можем записать

$$R = -\frac{E_z}{B j_x} = -\frac{E_z}{B \sigma_B E_x} = \frac{\mu^H B E_x}{\sigma_B B E_x} = \frac{\mu^H}{\sigma_B}, \quad (43.13)$$

что совершенно аналогично (42.17), однако выражения для σ_B и μ^H в данном случае другие. Подставляя их выражения (43.12) и (43.9), получим

$$R = \frac{\mu^H}{\sigma_B} = \frac{\mu^H}{\left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) \{1 + \mu^H B^2\}} = \frac{\left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right)}{\left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right)^2 + \left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right)^2 B^2}. \quad (43.14)$$

Подставляя в (43.14) выражения (29.27) для кинетических коэффициентов, запишем

$$R = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}^2}{1 + \mu_{\alpha}^2 B^2} \right\rangle}{\left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)^2 + \left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}^2}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)^2 B^2}. \quad (43.15)$$

Если в веществе имеется один тип носителей заряда, то (43.15) переходит в (42.23).

1. Слабые поля. Рассмотрим слабые поля: $\mu_{\alpha}^2 B^2 \ll 1$, при этом неравенство должно удовлетворяться для всех типов носителей заряда, и прежде всего для носителей заряда с максимальной подвижностью. Пренебрегая величиной $\mu_{\alpha}^2 B^2$ по сравнению с единицей во всех членах и вторым слагаемым в знаменателе по сравнению с первым слагаемым, получим

$$R = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2}, \quad (43.16)$$

но

$$\langle \mu_\alpha \rangle = \frac{e_\alpha}{m_\alpha^*} \langle \tau_\alpha \rangle = \mu_{d\alpha} \quad (43.17)$$

и

$$\langle \mu_\alpha^2 \rangle = \frac{e_\alpha^2}{m_\alpha^{*2}} \langle \tau_\alpha^2 \rangle = \left(\frac{e_\alpha \langle \tau_\alpha \rangle}{m_\alpha^*} \right)^2 \frac{\langle \tau_\alpha^2 \rangle}{\langle \tau_\alpha \rangle^2} = \mu_{d\alpha}^2 A_\alpha, \quad (43.18)$$

поэтому

$$R = \frac{\sum_\alpha e_\alpha n_\alpha A_\alpha \mu_{d\alpha}^2}{\left(\sum_\alpha e_\alpha n_\alpha \mu_{d\alpha} \right)^2} = \frac{\sum_\alpha \sigma_\alpha \mu_\alpha^H}{\left(\sum_\alpha \sigma_\alpha \right)^2} = \frac{\mu^H}{\sigma}, \quad (43.19)$$

где

$$\sigma = \sum_\alpha \sigma_\alpha; \quad \mu^H = \frac{\sum_\alpha \mu_\alpha^H \sigma_\alpha}{\sum_\alpha \sigma_\alpha}. \quad (43.20)$$

Рассмотрим выражение (43.19) для двух типов носителей заряда:

$$R = \frac{e_1 n_1 A_1 \mu_{d1}^2 + e_2 n_2 A_2 \mu_{d2}^2}{(e_1 n_1 \mu_{d1} + e_2 n_2 \mu_{d2})^2}. \quad (43.21)$$

Если полупроводник имеет смешанную проводимость, то, полагая

$$n_1 = n; \quad n_2 = p; \quad e_1 = e_n; \quad e_2 = e_p; \quad \mu_{d1} = \mu_n; \quad \mu_{d2} = \mu_p; \quad \frac{\mu_n}{\mu_p} = -b; \\ (b > 0),$$

можем записать

$$R = \frac{e_n n A_n \mu_n^2 + e_p p A_p \mu_p^2}{(e_n n \mu_n + e_p p \mu_p)^2} = \frac{1}{e_p} \frac{A_p \rho - A_n n b^2}{(\rho + n b)^2}. \quad (43.22)$$

Предположив, что $A_p = A_n = A$; $\frac{n}{p} = x$, получим

$$R = \frac{A}{e_p} \frac{\rho - n b^2}{(\rho + n b)^2} = \frac{A}{e_p \rho} \frac{1 - x b^2}{(1 + x b)^2}. \quad (43.23)$$

Рассмотрим три частных случая:
дырочный полупроводник

$$n = 0; \quad R = \frac{A}{e_p \rho}; \quad R > 0; \quad (43.24)$$

электронный полупроводник

$$p = 0; \quad R = -\frac{A}{e_p n} = \frac{A}{e_n n}; \quad R < 0; \quad (43.25)$$

собственный полупроводник

$$p = n; \quad x = 1; \quad R = \frac{A}{e_p \rho} \frac{1 - b^2}{(1 + b)^2} = \frac{A}{e_p \rho} \frac{1 - b}{1 + b}. \quad (43.26)$$

Обычно $b > 1$, поэтому в собственном полупроводнике $R < 0$. Если $b = 1$, то $R = 0$ и $E_z = 0$ — это соответствует тому, что отклоняемые в одну и ту же сторону электроны и дырки не создают поля Холла — их заряды компенсируют друг друга и поле не возникает. Если же $b \neq 0$, то $R \neq 0$, и знак R определяется знаком носителей заряда, имеющих большую подвижность.

Так как и n и p являются функциями температуры, то R также зависит от температуры. В области собственной проводимости $x = 1$, R является отрицательной величиной, уменьшающейся по модулю с ростом T . Если при малом T полупроводник был акцепторным, то при этих температурах $R > 0$ и, следовательно, с ростом T коэффициент Холла должен пройти через нулевое значение (рис. 63). Эта температура носит название температуры инверсии $T_{инв}$. Она определяется условием

$$R(T_{инв}) = 0 = p - nb^2, \quad (43.27)$$

откуда

$$b^2 = \frac{p}{n}; \quad b = \sqrt{\frac{p}{n}}. \quad (43.28)$$

Если известны n и p , то можно вычислить b .

В интерметаллах, например, $b \gg 1$, поэтому проводимость может быть дырочной — $p \gg n$, а R при этом может оказаться отрицательной величиной. В общем случае знак R определяется знаком величины

$$1 - xb^2 = 1 - \frac{n \mu_n^2}{p \mu_p^2}. \quad (43.29)$$

Коэффициент Холла в полупроводнике в области смешанной проводимости определяется четырьмя величинами: n , p , μ_n , μ_p . Проводимость σ определяется теми же величинами.

Если носители заряда имеют одинаковый знак, как в p -Si или p -Ge, то

$$R = \frac{1}{e_p} \frac{A_1 p_1 \mu_{p1}^2 + A_2 p_2 \mu_{p2}^2}{(p_1 \mu_{p1} + p_2 \mu_{p2})^2}. \quad (43.30)$$

Обозначив

$$\frac{\mu_{p2}}{\mu_{p1}} = b; \quad \frac{p_2}{p_1} = x; \quad A_1 = A_2 = A, \quad \text{получим}$$

$$R = \frac{A}{e_p p_1} \frac{1 + xb^2}{(1 + xb)^2}. \quad (43.31)$$

Выражение (43.31), как и (43.23), показывает, что относительный вклад быстрых носителей заряда в образование поля Холла опре-

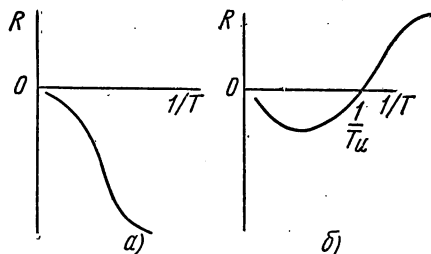


Рис. 63. Зависимость коэффициента Холла от обратной температуры в электронном (а) и дырочном (б) полупроводниках (тип проводимости определен при низких температурах)

деляется величиной xb^2 , в то время как их вклад в проводимость определяется величиной xb , т. е. в эффекте Холла носители заряда с большей подвижностью играют большую роль, чем в проводимости.

2. Сильные поля. Перейдем к рассмотрению сильных полей: $\mu_\alpha^2 B^2 \gg 1$; из (43.15) можно записать

$$R = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha}}{\left(\sum_{\alpha} e_{\alpha} n_{\alpha}\right)^2 + \left(\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^{-1} \rangle\right)^2 \frac{1}{B^2}}. \quad (43.32)$$

Для собственного полупроводника числитель (43.32) обращается в нуль, в то время как знаменатель отличен от нуля при любом конечном B , поэтому для собственного полупроводника $R = 0$.

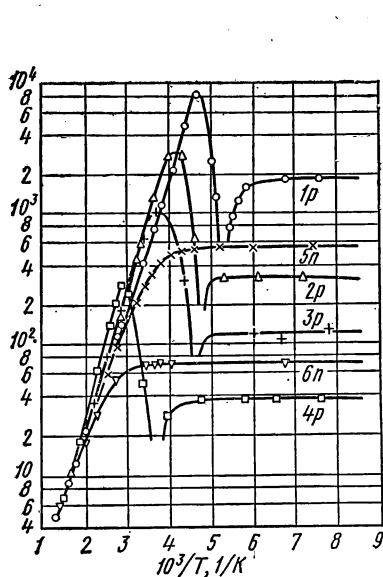


Рис. 64. Зависимость $\lg |R|$ от обратной температуры в электронном и дырочном атимониде индия

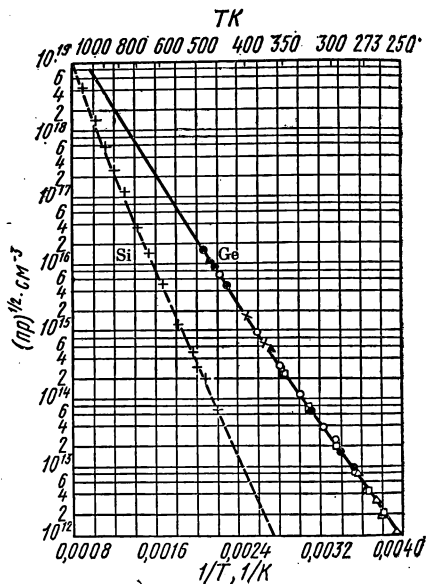


Рис. 65. Зависимость $\ln n_i$ от обратной температуры в германии и кремнии

Если полупроводник не собственный, то

$$R = \frac{1}{\sum_{\alpha} e_{\alpha} n_{\alpha}} = \frac{1}{\sum R_{\alpha}^{-1}}, \quad (43.33)$$

или

$$R^{-1} = \sum_{\alpha} R_{\alpha}^{-1}. \quad (43.34)$$

На рис. 64 в качестве примера приведен график зависимости $|R|$ от T^{-1} для нескольких образцов с электронной (n) и дырочной

(р) проводимостью. Из рис. 64 видно, что в области низких температур (от 120°K до 200°K) концентрация носителей заряда остается постоянной, это соответствует области истощения примеси. Концентрация свободных носителей заряда равна $3 \cdot 10^{15} \text{ см}^{-3}$ в 1р и $1,5 \times 10^{17} \text{ см}^{-3}$ в 4р. Температура инверсии имеет тем большую величину, чем больше концентрация примеси, как это видно из кривых для образцов 1р, 2р, 3р и 4р. Переход к собственной проводимости происходит при температуре тем меньшей, чем меньше концентрация примеси. В области собственной проводимости $\lg |R|$ меняется линейно с обратной температурой. На рис. 65 приведен график зависимости n_i в германии и кремнии от обратной температуры, полученный на основе изучения зависимости коэффициента Холла от обратной температуры.

§ 44. ЗАВИСИМОСТЬ КОЭФФИЦИЕНТА ХОЛЛА ОТ МАГНИТНОГО ПОЛЯ

Коэффициент Холла, вводимый в выражение для E^H , как показывает расчет на основе кинетического уравнения, зависит сложным образом от магнитного поля. Соотношения (42.28), (42.41) и (43.19), (43.33) дают по существу выражения для R , справедливые лишь в пределах $B \rightarrow 0$ и $B \rightarrow \infty$, эти предельные выражения обозначим через R_0 и R_∞ . Рассмотрим теперь выражения, справедливые при $B \neq 0$ и $B \neq \infty$. Остановимся вначале на примесном полупроводнике с носителями заряда одного типа, для которого R согласно (42.23) равен

$$R = \frac{1}{en} \frac{\left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle^2 + \left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle^2 B^2}. \quad (44.1)$$

Рассмотрим вначале вырожденный полупроводник. Согласно (39.32) усреднение по энергии равносильно подстановке значений соответствующих величин на поверхности Ферми. В таком случае (44.1) примет вид:

$$R = \frac{1}{en} \cdot \frac{\frac{\mu^2(F)}{1 + \mu^2(F) B^2}}{\frac{\mu^2(F)}{[1 + \mu^2(F) B^2]^2} + \frac{\mu^4(F) B^2}{[1 + \mu^2(F) B^2]^2}} = \frac{1}{en}, \quad (44.2)$$

т. е. для вырожденного полупроводника коэффициент Холла не зависит от индукции магнитного поля во всем возможном интервале значений.

Рассмотрим теперь невырожденный полупроводник, разложим величину $\frac{1}{1 + \mu^2 B^2}$ в ряд по $\mu^2 B^2$:

$$\frac{1}{1 + \mu^2 B^2} = \sum_m (-1)^m (\mu^2 B^2)^m = 1 - \mu^2 B^2 + \mu^4 B^4 - \mu^6 B^6 + \dots \quad (44.3)$$

и в случае слабого поля ограничимся двумя первыми членами. Ошибка при этом не будет превосходить $\mu^4 B^4$. В таком случае

$$\left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle = \langle \mu^2 - \mu^4 B^2 + \dots \rangle \approx \langle \mu^2 \rangle - \langle \mu^4 \rangle B^2$$

(44.4)

и

$$\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle = \langle \mu - \mu^3 B^2 + \dots \rangle \approx \langle \mu \rangle - \langle \mu^3 \rangle B^2.$$

Из (44.3) и (44.4) получим

$$R = \frac{1}{en} \frac{\langle \mu^2 \rangle - \langle \mu^4 \rangle B^2}{[\langle \mu \rangle - \langle \mu^3 \rangle B^2]^2 + [\langle \mu^2 \rangle - \langle \mu^4 \rangle B^2]^2 B^2} =$$

$$= \frac{1}{en} \frac{\langle \mu^2 \rangle}{\langle \mu \rangle^2} \left\{ 1 - B^2 \left(\frac{\langle \mu^4 \rangle}{\langle \mu^2 \rangle} - 2 \frac{\langle \mu^3 \rangle}{\langle \mu \rangle} + \frac{\langle \mu^2 \rangle^2}{\langle \mu \rangle^2} \right) \right\}. \quad (44.5)$$

Так как $\frac{1}{en} \frac{\langle \mu^2 \rangle}{\langle \mu \rangle^2} = R_0$, то

$$R = R_B = R_0 \left\{ 1 - B^2 \left(\frac{\langle \mu^2 \rangle^2 - 2 \langle \mu \rangle \langle \mu^3 \rangle}{\langle \mu \rangle^2} + \frac{\langle \mu^4 \rangle}{\langle \mu^2 \rangle} \right) \right\}, \quad (44.6)$$

т. е. с ростом поля коэффициент Холла уменьшается пропорционально B^2 . Коэффициент пропорциональности зависит от квадрата подвижности и механизма рассеяния. Действительно, обозначим

$$\frac{\langle \mu^2 \rangle}{\langle \mu \rangle^2} = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = A; \quad \frac{\langle \mu^3 \rangle}{\langle \mu \rangle^3} = \frac{\langle \tau^3 \rangle}{\langle \tau \rangle^3} = C; \quad \frac{\langle \mu^4 \rangle}{\langle \mu \rangle^4} = \frac{\langle \tau^4 \rangle}{\langle \tau \rangle^4} = D, \quad (44.7)$$

тогда

$$R_B = R_0 \left\{ 1 - \mu_d^2 B^2 \left(A^2 - 2C + \frac{D}{A} \right) \right\}. \quad (44.8)$$

Найдем величину коэффициентов А, С и D, когда действует один механизм рассеяния $-\tau = \tau_0 E^p$:

$$\left(A^2 - 2C + \frac{D}{A} \right) = \left\{ \frac{\Gamma\left(\frac{5}{2} + 2p\right) \Gamma\left(\frac{5}{2}\right)^2}{\left[\Gamma\left(\frac{5}{2} + p\right)\right]^2} - \right.$$

$$\left. - 2 \frac{\Gamma\left(\frac{5}{2} + 3p\right) \left[\Gamma\left(\frac{5}{2}\right)\right]^2}{\left[\Gamma\left(\frac{5}{2} + p\right)\right]^3} + \frac{\Gamma\left(\frac{5}{2} + 4p\right) \left[\Gamma\left(\frac{5}{2}\right)\right]^2}{\left[\Gamma\left(\frac{5}{2} + p\right)\right]^2 \Gamma\left(\frac{5}{2} + 2p\right)} \right\}. \quad (44.9)$$

В табл. 11 приведены значения множителя (44.9) для трех случаев: $p = -1/2$; $p = 0$ и $p = 3/2$.

Т а б л и ц а 11

p	$-1/2$	0	$3/2$
$A^2 - 2C + \frac{D}{A}$	$\frac{9\pi^2}{64} = 1,39 = \left(\frac{3\pi}{8}\right)^2$	0	$615 = (24,8)^2$

Переход от рассеяния на тепловых колебаниях решетки к рассеянию на ионах примеси приводит к резкому изменению множителя при B^2 :

$$R_B = R_0 (1 - 1,39\mu_d^2 B^2) \quad \text{для } p = -\frac{1}{2},$$

$$R_B = R_0 (1 - 615\mu_d^2 B^2) \quad \text{для } p = \frac{3}{2}. \quad (44.10)$$

Таким образом, в области слабых полей коэффициент Холла уменьшается пропорционально $\mu_d^2 B^2$.

Рассмотрим область сильных полей $\mu_d^2 B^2 \gg 1$. Разложение (44.3) здесь справедливо, но при этом ряд расходится, необходимо перейти к разложению в ряд по $(\mu_d^2 B^2)^{-1}$. Можно записать таким образом:

$$\frac{1}{1+x} = \frac{1}{x \left(1 + \frac{1}{x}\right)} = \frac{1}{x} \sum_{m=0}^{\infty} (-1)^m \left(\frac{1}{x}\right)^m; \quad |x| > 1, \quad (44.11)$$

или

$$\frac{1}{1+\mu^2 B^2} = \frac{1}{\mu^2 B^2} \sum_m (-1)^m \left(\frac{1}{\mu^2 B^2}\right)^m = \frac{1}{\mu^2 B^2} \left(1 - \frac{1}{\mu^2 B^2} + \dots\right). \quad (44.12)$$

Учитывая разложение (44.12), запишем

$$\left\langle \frac{\mu^2}{1+\mu^2 B^2} \right\rangle = \left\langle \mu^2 \left(\frac{1}{\mu^2 B^2} - \frac{1}{\mu^4 B^4} + \dots \right) \right\rangle \approx \frac{1}{B^2} (1 - \langle \mu^{-2} \rangle B^{-2}) \quad (44.13)$$

и

$$\left\langle \frac{\mu}{1+\mu^2 B^2} \right\rangle = \left\langle \mu \left(\frac{1}{\mu^2 B^2} - \frac{1}{\mu^4 B^4} + \dots \right) \right\rangle = \frac{1}{B^2} (\langle \mu^{-1} \rangle - \langle \mu^{-3} \rangle B^{-2}). \quad (44.14)$$

Подставив (44.13) и (44.14) в (44.1) и оставляя члены не ниже B^{-4} , получим

$$R = \frac{1}{en} \frac{\frac{1}{B^2} (1 - \langle \mu^{-2} \rangle B^{-2})}{\langle \mu^{-1} \rangle^2 B^{-4} + \frac{1}{B^4} (1 - 2 \langle \mu^{-2} \rangle B^{-2}) B^2} =$$

$$= \frac{1}{en} \{1 + (\langle \mu^{-2} \rangle - \langle \mu^{-1} \rangle^2) B^{-2}\} = R_B. \quad (44.15)$$

Выражение (44.15) можно представить в виде

$$R_B = \left\{ 1 + \frac{\langle \mu^{-2} \rangle - \langle \mu^{-1} \rangle^2}{B^2} \right\} R_\infty. \quad (44.16)$$

Полагая $\tau = \tau_0 E^p$, получим, например, при $p = -1/2$

$$R_B = R_\infty \left(1 + \frac{1,76}{\mu_0^2 B^2} \right), \quad (44.17)$$

где

$$\mu_0 = \frac{e\tau'_0}{m^*}, \quad \tau'_0 = \tau_0 (kT)^p.$$

Таким образом, в области слабых и сильных полей

$$R_B = \frac{A}{en} (1 - a\mu_d^2 B^2) \quad \text{и} \quad R_B = \frac{1}{en} \left(1 + \frac{a'}{\mu_d^2 B^2} \right), \quad (44.18)$$

где a и a' зависят от механизма рассеяния. Общий ход $R(B)$ представлен на рис. 66.

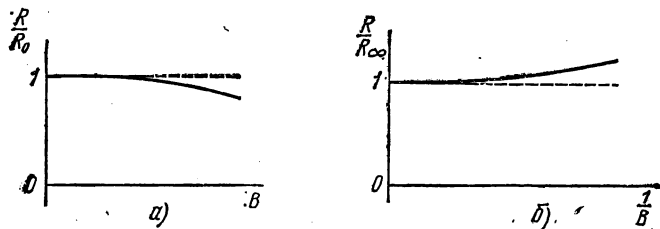


Рис. 66. Зависимость коэффициента Холла от магнитного поля в области слабых (а) и сильных (б) полей

Перейдем теперь к исследованию зависимости $R(B)$ для общего случая, когда имеются носители заряда нескольких типов:

$$R = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}^2}{1 + \mu_{\alpha}^2 B^2} \right\rangle}{\left[\left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)^2 + \left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}^2}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)^2 B^2 \right]}. \quad (44.19)$$

Для слабых полей с учетом (44.4) получим

$$R = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} |\langle \mu_{\alpha}^2 \rangle - \langle \mu_{\alpha} \rangle^2 B^2|}{\left\{ \left[\sum_{\alpha} e_{\alpha} n_{\alpha} (\langle \mu_{\alpha} \rangle - \langle \mu_{\alpha} \rangle B^2) \right]^2 + \left[\sum_{\alpha} e_{\alpha} n_{\alpha} (\langle \mu_{\alpha}^2 \rangle - \langle \mu_{\alpha} \rangle^2 B^2) \right]^2 B^2 \right\}}. \quad (44.20)$$

Оставим в (44.20) только члены не выше B^2 :

$$\begin{aligned}
 R &= \frac{\left\{ \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right] - \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^4 \rangle \right] B^2 \right\}}{\left\{ \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2 + \left(\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right] \right)^2 \right\}} = \\
 &= 2 \frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right] \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right] B^2}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2} = \\
 &= \frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right]}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2} \left\{ 1 - B^2 \left[\frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^4 \rangle}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right]} + \right. \right. \\
 &+ \left. \left. \frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right]^2 - 2 \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right] \left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^3 \rangle \right]}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2} \right\}, \quad (44.21)
 \end{aligned}$$

где

$$\frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right]}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]^2} = R_0. \quad (44.22)$$

Выражение (44.21) переходит в (44.6) при $\alpha = 1$.

Для собственного полупроводника $n_1 = n = n_2 = p$. Полагаем $\langle \mu_1 \rangle = -b \langle \mu_2 \rangle$; тогда

$$R_B = R_0 \left\{ 1 - \mu_{pd}^2 B^2 \left[\frac{D}{A} (1 + b^2) + A^2 (1 - b^2) - 2C \frac{1 + b^2}{1 + b} \right] \right\}. \quad (44.23)$$

Если $b \ll 1$, то

$$R_B = R_0 \left\{ 1 - \mu_{pd}^2 B^2 \left[\frac{D}{A} + A^2 - 2C \right] \right\} \quad (44.24)$$

в полном соответствии с (44.9) для одного типа носителей заряда. Если $b \gg 1$, то

$$R_B = R_0 \left\{ 1 - \mu_{nd}^2 B^2 \left[\frac{D}{A} + A^2 - 2C \right] \right\}. \quad (44.25)$$

Выражения (44.24) и (44.25) показывают, что изменение коэффициента Холла в слабом магнитном поле обуславливается в первую очередь более подвижными носителями заряда.

Тот же результат получили и для случая полупроводника с несколькими типами носителей заряда одного знака.

Подобное положение наблюдается, например, в p -Ge. С ростом поля из-за вклада легких дырок R уменьшается, поскольку одно и то же поле оказывается сильным для легких дырок и слабым для тяжелых дырок. В электронном германии R практически не зависит от поля в области «нуль — несколько килогауссов».

Можно представить, что если подвижности нескольких типов носителей заряда различаются значительно, то на кривой $R(B)$

можно заметить столько перегибов, сколько типов носителей заряда участвуют в образовании проводимости и поля Холла. Эффекты резкого изменения будут тем более заметны, чем ближе концентрации носителей зарядов разного типа.

Рассмотрим, как проявляется вырождение на зависимости коэффициента Холла от поля в случае, когда проводимость осуществляется несколькими типами носителей заряда. Не анализируя (44.1), можно сказать, что в этом случае R должно зависеть от B . В подтверждение этого сравним (43.23) и (43.33), из которых видно, что $R_0 \neq R_\infty$.

Для смешанной проводимости при вырождении электронов и дырок, что возможно в полуметаллах при перекрытии зон энергии, (44.1) примет вид:

$$R(B) = \frac{\frac{\sigma_1 \mu_1}{1 + \mu_1^2 B^2} + \frac{\sigma_2 \mu_2}{1 + \mu_2^2 B^2}}{\left(\frac{\sigma_1}{1 + \mu_1^2 B^2} + \frac{\sigma_2}{1 + \mu_2^2 B^2} \right)^2 + \left(\frac{\sigma_1 \mu_1 B}{1 + \mu_1^2 B^2} + \frac{\sigma_2 \mu_2 B}{1 + \mu_2^2 B^2} \right)^2}. \quad (44.26)$$

Введем обозначение

$$\frac{1 + \mu_2^2 B^2}{1 + \mu_1^2 B^2} = f(b, B). \quad (44.27)$$

Из (44.27) видно, что $f(b, 0) = 1$ и $f(b, \infty) = b^2$. Используя $f(b, B)$ перепишем (44.26):

$$R = \frac{1 + \mu_2^2 B^2}{e_1 n_1} \cdot \frac{1 + \frac{e_2}{e_1} x b^2 \frac{1}{f(b, B)}}{\left(1 + \frac{x b}{f(b, B)} \right)^2 + \mu_1^2 B^2 \left(1 + \frac{e_2}{e_1} \cdot \frac{x b^2}{f(b, B)} \right)^2}. \quad (44.28)$$

Из (44.28) следует (43.23) при $B \rightarrow 0$ и (44.33) при $B \rightarrow \infty$. Для собственного полуметалла имеем:

$$R = \frac{1 + \mu_1^2 B^2}{e_p n_i} \cdot \frac{1 - \frac{b^2}{f(b, B)}}{\left(1 + \frac{b}{f(b, B)} \right)^2 + \mu_p^2 B^2 \left(1 - \frac{b^2}{f(b, B)} \right)^2}. \quad (44.29)$$

Хотя поле влияет на величину коэффициента Холла и в вырожденных полупроводниках, однако это влияние при прочих равных условиях меньше, чем в невырожденных полупроводниках. Это видно, например, из соотношения (44.25) — при доминировании электронов в образовании поля Холла вырождение электронного газа приводит к независимости R от B . С точки зрения наглядных представлений об эффекте Холла зависимость R от B следует ожидать в тех случаях, когда магнитное поле меняет вклад в образование поля Холла носителей заряда различной энергии в невырожденных полупроводниках с одним типом носителей заряда или различных типов носителей заряда независимо от степени вырождения. В полупроводниках или полуметаллах с большим отношением подвижно-

стей b в области малых полей могут доминировать более подвижные носители заряда даже при их небольшой концентрации, но с ростом поля их вклад может существенно уменьшаться, что позволяет наблюдать изменение знака поля Холла. На рис. 67 в качестве примера приведена полевая инверсия знака коэффициента Холла $Cd_xHg_{1-x}Te$, у которого $b \approx 60 \div 100$.

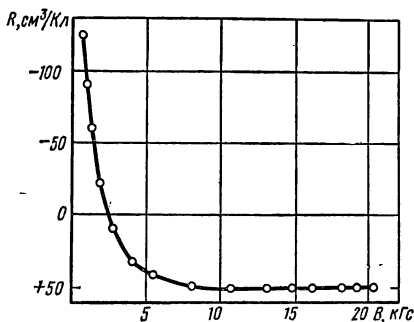


Рис. 67. Полевая инверсия знака коэффициента Холла в $Cd_xHg_{1-x}Te$ при $77^\circ K$

Изучение явления Холла имеет важнейшее значение при определении концентрации и подвижности носителей заряда, при изучении поведения различных примесей в полупроводниках. Зависимость $\ln|R|$ от T^{-1} позволяет определить энергию ионизации примеси или ширину запрещенной зоны, другими словами, получение большинства сведений относительно зависимости концентраций носителей заряда от внешних условий или состава полупроводника возможно при исследовании явления Холла.

Резюме § 43, 44

1. Уравнения для определения коэффициента Холла R и подвижности Холла μ^H в случае, когда в полупроводнике содержатся носители заряда нескольких типов, имеют тот же вид, что и в случае одного типа. Между R и μ^H имеется связь:

$$R\sigma_B = \mu^H. \quad (44.1p)$$

2. Подвижность Холла

$$\mu^H = \frac{\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha)}{\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha)} = \frac{\sum_{\alpha} \sigma'_{\alpha} \mu_{\alpha}^H}{\sum_{\alpha} \sigma'_{\alpha}} \quad (44.2p)$$

связана с подвижностью Холла μ_{α}^H каждого типа носителей заряда и проводимостью, создаваемой каждым типом носителей заряда. Подвижность Холла μ^H определяет напряженность поля Холла через напряженность поля E_x и индукцию магнитного поля B :

$$E_z = -\mu^H B E_x = \operatorname{tg} \varphi E_x. \quad (44.3p)$$

В собственном полупроводнике подвижность Холла

$$\mu^H = \frac{\sigma_n \mu_n^H + \sigma_p \mu_p^H}{\sigma_n + \sigma_p} = \frac{\mu_p^H + b \mu_n^H}{1 + b}. \quad (44.4p)$$

Если $A_n = A_p$, то

$$\mu^H = \frac{A(1-b^2)\mu_p}{1+b} = (1-b)\mu_p^H. \quad (44.5p)$$

3. Выражение для коэффициента Холла имеет вид (43.15). В слабых полях

$$R = \frac{\mu^H}{\sigma} = \frac{\sum_{\alpha} \sigma_{\alpha} \mu_{\alpha}^H}{\left(\sum_{\alpha} \sigma_{\alpha}\right)^2}. \quad (44.6p)$$

Для собственного полупроводника

$$R = \frac{A}{e_p p} \frac{1-b}{1+b}. \quad (44.7p)$$

В полупроводниках, имеющих вырожденные зоны, более подвижные носители заряда играют большую роль в образовании поля Холла, чем в проводимости.

4. В сильных магнитных полях в собственном полупроводнике $R=0$, а в полупроводнике, имеющем смешанную проводимость,

$$R^{-1} = \sum_{\alpha} R_{\alpha}^{-1} = \sum_{\alpha} e_{\alpha} n_{\alpha}. \quad (44.8p)$$

5. В области слабых магнитных полей и в области сильных магнитных полей коэффициент Холла уменьшается пропорционально $(\mu_d B)^2$ и $(\mu_d B)^{-2}$.

6. Изучение эффекта Холла используется для определения концентрации и подвижности носителей заряда, для изучения их температурной зависимости, для определения энергии активации.

§ 45. МАГНИТНОРЕЗИСТИВНЫЙ ЭФФЕКТ

Рассмотрим магнитнорезистивный эффект на основе кинетического уравнения. Пусть ограниченный полупроводник в виде параллелепипеда, ребра которого направлены вдоль осей координат, имеет сферические поверхности энергии и один тип носителей заряда. Ток идет вдоль оси x . Проводимость вдоль поля E_x имеет вид (42.12):

$$\sigma_B = e^2 K'_{11} + \frac{e^4}{m^{*2}} \frac{(K'_{13})^2}{K'_{11}} B^2, \quad (45.1)$$

или

$$\sigma_B = en \left\langle \frac{\mu}{1+\mu^2 B^2} \right\rangle + en \frac{\left\langle \frac{\mu^2}{1+\mu^2 B^2} \right\rangle^2 B^2}{\left\langle \frac{\mu}{1+\mu^2 B^2} \right\rangle}. \quad (45.2)$$

Коэффициент магнетосопротивления H принято вводить соотношением

$$H = \frac{\rho_B - \rho_0}{\rho_0} \cdot \frac{1}{B^2} = \frac{1}{B^2} \frac{\Delta \rho_B}{\rho_0} = \frac{\sigma_0 - \sigma_B}{\sigma_B} \cdot \frac{1}{B^2}. \quad (45.3)$$

Если H известен, то сопротивление ρ_B в поле B можно записать в виде

$$\rho_B = \rho_0 (1 + HB^2); \quad \sigma_B = \frac{\sigma_0}{1 + HB^2}. \quad (45.4)$$

Так как $\sigma_0 = e^2 K_{11}$, то из (45.3), (45.1) и (45.2) для H запишем выражение

$$H = \frac{1}{B^2} \frac{e^2 (K_{11} - K'_{11}) - \frac{e^4}{m^{*2}} \frac{(K'_{12})^2}{K'_{11}} B^2}{e^2 K'_{11} + \frac{e^4}{m^{*2}} \frac{(K'_{12})^2}{K'_{11}} B^2}, \quad (45.5)$$

или

$$H = \frac{1}{B^2} \frac{\langle \mu \rangle - \left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle - \frac{\left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle^2 B^2}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle}}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle + \frac{\left\langle \frac{\mu^2}{1 + \mu^2 B^2} \right\rangle^2 B^2}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle}}. \quad (45.6)$$

Рассмотрим область полей, в которой согласно (44.4) можно записать

$$H = \frac{1}{B^2} \frac{\langle \mu^3 \rangle B^2 - B^2 \frac{[\langle \mu^2 \rangle - \langle \mu^4 \rangle B^2]^2}{[\langle \mu \rangle - \langle \mu^3 \rangle B^2]}}{\langle \mu \rangle - \langle \mu^3 \rangle B^2 + B^2 \frac{[\langle \mu^2 \rangle - \langle \mu^4 \rangle B^2]^2}{[\langle \mu \rangle - \langle \mu^3 \rangle B^2]}} =$$

$$= \frac{\langle \mu^3 \rangle \langle \mu \rangle - \langle \mu^2 \rangle^2}{\langle \mu \rangle^2} \left\{ 1 + B^2 \left[\frac{\langle \mu^4 \rangle}{\langle \mu^2 \rangle} - \frac{\langle \mu^2 \rangle^2}{\langle \mu \rangle^2} \right] \right\}. \quad (45.7)$$

В области слабых полей (в пределе $B=0$)

$$H = H_0 = \frac{\langle \mu^3 \rangle \langle \mu \rangle - \langle \mu^2 \rangle^2}{\langle \mu \rangle^2} = \frac{e^2}{m^{*2}} \frac{\langle \tau^3 \rangle \langle \tau \rangle - \langle \tau^2 \rangle^2}{\langle \tau \rangle^2} = \mu_d^2 [C - A^2], \quad (45.8)$$

т. е. коэффициент магнетосопротивления постоянен, сопротивление растет пропорционально B^2 :

$$\rho_B = \rho_0 (1 + H_0 B^2). \quad (45.9)$$

С ростом поля коэффициент магнетосопротивления (45.7) уменьшается, что означает замедление роста сопротивления.

Рассмотрим *сильные поля* (в пределе $B = \infty$)

$$H = \frac{1}{B^2} \frac{\langle \mu \rangle - \langle \mu^{-1} \rangle^{-1}}{\langle \mu^{-1} \rangle^{-1}} = \frac{1}{B^2} (\langle \mu \rangle \langle \mu^{-1} \rangle - 1). \quad (45.10)$$

Коэффициент магнетосопротивления в области сильных полей уменьшается как B^{-2} . Следовательно, сопротивление при больших полях имеет насыщение:

$$\rho_B \approx \rho_0 (1 + \langle \mu \rangle \langle \mu^{-1} \rangle - 1) = \rho_0 \langle \mu \rangle \langle \mu^{-1} \rangle = \rho_\infty. \quad (45.11)$$

Величина насыщения магнетосопротивления

$$\Delta \rho_\infty = \rho_\infty - \rho_0 = \rho_0 (\langle \mu \rangle \langle \mu^{-1} \rangle - 1) = \rho_0 (\langle \tau \rangle \langle \tau^{-1} \rangle - 1) \quad (45.12)$$

зависит от механизма рассеяния.

Если действует один какой-либо механизм рассеяния, так что $\tau = \tau_0 E^p$, то

$$\langle \tau \rangle \langle \tau^{-1} \rangle = \frac{\Gamma\left(\frac{5}{2} + p\right) \Gamma\left(\frac{5}{2} - p\right)}{\left[\Gamma\left(\frac{5}{2}\right)\right]^2}. \quad (45.13)$$

В табл. 12 приведена величина $[\langle \tau \rangle \langle \tau^{-1} \rangle]^{-1}$ и $H B^2$ для некоторых значений p .

Таблица 12

p	0	$\pm 1/2$	± 1	$\pm 3/2$	± 2
$[\langle \tau \rangle \langle \tau^{-1} \rangle]^{-1}$	1	$\frac{9\pi}{32} = 0,884$	$\frac{3}{5} = 0,600$	$\frac{3\pi}{32} = 0,295$	$\frac{3}{35} = 0,86$
$H B^2$	0	0,116	0,400	0,705	0,914

Мы видим, что положительное магнетосопротивление достигает насыщения, величина которого зависит от p . При $p = 1/2$ отношение $\frac{\Delta \rho_\infty}{\rho_0}$ составляет примерно 12%.

В вырожденных полупроводниках $H = 0$ при любых полях. Это означает, что в вырожденных полупроводниках и металлах магнетосопротивление в первом приближении равно нулю. Чтобы получить отличные от нуля члены, необходимо учесть следующие члены в разложении (3.42).

Вычисления дают

$$H = \frac{H_0}{1 + (\sigma R B)^2}, \quad (45.14)$$

где

$$H_0 = \frac{\pi^2}{12} \left[\frac{ekT\tau(F)}{m^*F} \right]^2, \quad (45.15)$$

Т с. в слабых полях магнетосопротивление квадратично по B ,
 в сильных полях магнетосопротивление достигает насыщения.

Рассмотрим ограниченный полупроводник с несколькими типами носителей заряда; согласно (43.12) и (43.15)

$$\sigma_B = \sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle + \frac{\left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}^2}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)^2 B^2}{\left(\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right)}. \quad (45.16)$$

В области слабых полей

$$\sigma_B = \sum_{\alpha} e_{\alpha} n_{\alpha} (\langle \mu_{\alpha} \rangle - \langle \mu_{\alpha}^3 \rangle B^2) + \frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} (\langle \mu_{\alpha}^2 \rangle - \langle \mu_{\alpha}^4 \rangle B^2) \right]^2 B^2}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} (\langle \mu_{\alpha} \rangle - \langle \mu_{\alpha}^3 \rangle B^2) \right]}. \quad (45.17)$$

Оставляя только члены при нулевой и второй степени B , получим

$$H_0 = \frac{1}{B^2} \frac{\sigma_0 - \sigma_B}{\sigma_B} = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^3 \rangle - \frac{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^2 \rangle \right]^2}{\left[\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle \right]}}{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle}. \quad (45.18)$$

Рассмотрим собственный полупроводник:

$$\begin{aligned} \langle \mu_1 \rangle &= \mu_{nd}; & \langle \mu_2 \rangle &= \mu_{pd}; & \langle \mu_{\alpha}^2 \rangle &= \langle \mu_{\alpha} \rangle^2 A_{\alpha}; & A_1 &= A_2 = A; \\ \langle \mu_{\alpha}^3 \rangle &= \langle \mu_{\alpha} \rangle^3 C_{\alpha}; & C_1 &= C_2 = C; & n_1 &= n = n_2 = p. \end{aligned} \quad (45.19)$$

Подставляя (45.19) в (45.18), получим

$$H_0 = \frac{C(\mu_p^3 - \mu_n^3) - \frac{A^2(\mu_p^2 - \mu_n^2)^2}{\mu_p - \mu_n}}{\mu_p - \mu_n} = \mu_p^2 \left[C \frac{1+b^3}{1+b} - A^2(1-b)^2 \right]. \quad (45.20)$$

Соотношение (45.20) имеет наглядный смысл. Если мы имеем дело с носителями заряда одного знака, то согласно (45.8) $H_0 = \mu_p^2 [C - A^2]$. Величина C определяет магнетосопротивление вследствие «закручивания» носителей заряда магнитным полем; член с A^2 определяет компенсирующее действие поля Холла (учитывает ограниченность образца).

Если в полупроводнике концентрации электронов и дырок не равны, то, обозначая $\frac{n}{p} = x$, из (45.20) получим

$$H_0 = C \mu_p^2 \frac{(1+xb^3)}{(1+xb)} - A^2 \mu_p^2 \frac{(1-xb^2)^2}{(1+xb)^2}. \quad (45.21)$$

Если носители заряда имеют одинаковый знак, то

$$H_0 = \mu_{p1}^2 \left[C \frac{(1+xb^3)}{(1+xb)} - A^2 \frac{(1+xb^2)^2}{(1+xb)^2} \right], \quad (45.21a)$$

где μ_{p1} — подвижность носителей заряда с концентрацией p_1 (или n_1),
 $x = \frac{p_2}{p_1}$; $b = \frac{\mu_{p2}}{\mu_{p1}}$.

Магнетосопротивление в сильных полях можно получить обычным способом из общего выражения (45.16).

Рассмотрим теперь магнетосопротивление в неограниченном веществе. Как уже говорилось неоднократно, поле Холла в этом случае не возникает, ток поворачивается на угол φ , появляется отличный от нуля компонент тока j_z . Выражение для тока согласно (42.2) имеет вид

$$\mathbf{i} = e^2 K'_{11} \mathbf{E} + \frac{e^3}{m^*} K'_{12} [\mathbf{E} \mathbf{B}] \quad (45.22)$$

с «граничными» условиями: $\mathbf{E} = (E_x, 0, 0)$ при $\mathbf{B} = (0, B, 0)$. Запишем (45.22) для компонентов тока:

$$\begin{aligned} i_x &= e^2 K'_{11} E_x = \sigma_{Bx} E_x; \\ j_y &= 0; \\ j_z &= \frac{e^3}{m^*} K'_{12} B E_x = \sigma_{Bz} E_x. \end{aligned} \quad (45.23)$$

Проводимость σ_{Bx} определяет ток вдоль электрического поля при наложении магнитного поля, σ_{Bx} зависит от B , так как от B зависит K'_{11} . Величина σ_{Bz} определяет проводимость вдоль оси z , в направлении вектора $[\mathbf{E}, \mathbf{B}]$; ее называют *проводимостью Холла*, σ_{Bz} отлична от σ_{Bx} . Говоря о магнетосопротивлении, имеют в виду изменение сопротивления вдоль поля E_x (или j_x). Найдем коэффициент магнетосопротивления

$$\begin{aligned} H &= \frac{1}{B^2} \frac{\sigma_0 - \sigma_B}{\sigma_B} = \frac{1}{B^2} \frac{\langle \mu \rangle - \left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle} = \\ &= \frac{1}{B^2} \frac{\left\langle \frac{\mu^3 B^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle} = \frac{\left\langle \frac{\mu^3}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\mu}{1 + \mu^2 B^2} \right\rangle}. \end{aligned} \quad (45.24)$$

В слабых полях (с учетом членов порядка $\mu^2 B^2$)

$$H = \frac{\langle \mu^3 \rangle - \langle \mu^5 \rangle B^2}{\langle \mu \rangle - \langle \mu^3 \rangle B^2} = \frac{\langle \mu^3 \rangle}{\langle \mu \rangle} \left[1 + \left(\frac{\langle \mu^3 \rangle}{\langle \mu \rangle} - \frac{\langle \mu^5 \rangle}{\langle \mu^3 \rangle} \right) B^2 \right]. \quad (45.25)$$

В полях, когда $\mu^2 B^2 \approx 0$ ($B \rightarrow 0$),

$$H_0 = \frac{\langle \mu^3 \rangle}{\langle \mu \rangle} = \mu_d^2 \frac{\langle \tau^3 \rangle}{\langle \tau \rangle^3} = C \mu_d^2. \quad (45.26)$$

Таблица 13

	0	-1/2	1/2	+1	3/2	2	5/2
$\frac{H}{\mu_d^2}$ (огр)	0	0,38	0,11	0,56	2,98	7,7	9,2
$\frac{H}{\mu_d^2}$ (неогр)	1	$\frac{9\pi}{16} = 1,77$	$\frac{27}{64} = 1,33$	$\frac{63}{25} = 2,52$	$\frac{15\pi}{8} = 5,90$	15,7	$\frac{945\pi}{64} = 46,4$
$\frac{\langle \tau^2 \rangle^2}{\langle \tau \rangle^4}$	1	$\frac{9\pi^2}{64} = 1,39$	$\frac{9\pi^2 \cdot 15}{64 \cdot 16} = 1,22$	$\frac{49}{25} = 1,96$	3,72	8	37,2

Сравнивая выражение (45.26) с (45.8), видим, что H_0 отличается множителем при μ_d^2 , который равен C в случае неограниченного вещества, и $C - A^2$ в случае ограниченного. Другими словами, в ограниченном образце магнетосопротивление выражено слабее, поскольку оно компенсируется полем Холла. Из (45.26) следует, что

$$\mu_d = \left(\frac{H_0}{C} \right)^{\frac{1}{2}}. \quad (45.27)$$

В сильном магнитном поле

$$H_\infty = \frac{\langle \mu \rangle}{\langle \mu^{-1} \rangle} = \frac{e^2 \langle \tau \rangle^2}{m^* a^2} \frac{1}{\langle \tau \rangle \langle \tau^{-1} \rangle} = \mu_d^2 \frac{1}{\langle \tau \rangle \langle \tau^{-1} \rangle}, \quad (45.28)$$

в то время как в ограниченном полупроводнике H_∞ определяется соотношением (45.10). Мы видим, что в неограниченном веществе H не зависит от B как в слабых, так и в сильных полях. Это значит, что $\rho_B = \rho_0 (1 + HB^2)$ при $B \rightarrow 0$ и $B \rightarrow \infty$. Последний результат понятен — с ростом поля B носители заряда начинают вращаться по окружностям все меньшего и меньшего радиуса $r = p/eB$ (p — импульс), и дрейфовать в направлении оси z со скоростью $u = B/E$, вследствие чего сопротивление возрастает до бесконечности. В табл. 13 приведены значения H/μ_d^2 в неограниченном и ограниченном веществе.

Из табл. 13 видно, что коэффициент магнетосопротивления в неограниченном полупроводнике в слабом поле превосходит H в ограниченном образце в зависимости от p в несколько раз, например,

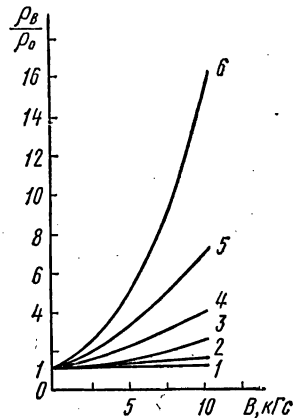


Рис. 68. Зависимость магнитно-резистивного эффекта от величины отношения длины образца к ширине: 1) ∞ ; 2) 5; 3) 2; 4) 1; 5) 0,4; 6) 0 (диск Корбино)

при $\rho = 1/2$ в 11 раз. Это различие увеличивается по мере возрастания поля. На рис. 68 приведена зависимость магнетосопротивления в InSb от формы образца. Рисунок показывает, что компенсирующее холлово поле играет тем большую роль, чем меньше отношение ширины к длине образца. Переход к образцу в виде диска (диск Корбино) равносителен переходу к «неограниченному образцу».

Рассмотрим в заключение этого параграфа магнетосопротивление в неограниченном веществе при наличии носителей заряда нескольких типов. Ток определяется выражением

$$\mathbf{j} = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) \mathbf{E} + \left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12}(\alpha) \right) [\mathbf{E}\mathbf{B}], \quad (45.29)$$

или

$$j_x = \left(\sum_{\alpha} e_{\alpha}^2 K'_{11}(\alpha) \right) E_x = \sigma_{Bx} E_x, \quad (45.30)$$

$$j_y = 0;$$

$$j_z = \sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12} B E_x = \sigma_{Bz} E_x. \quad (45.31)$$

Коэффициент магнетосопротивления

$$H = \frac{1}{B^2} \cdot \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \left[\langle \mu_{\alpha} \rangle - \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle \right]}{\sum_{\alpha} e_{\alpha} n_{\alpha} \left\langle \frac{\mu_{\alpha}}{1 + \mu_{\alpha}^2 B^2} \right\rangle}; \quad (45.32)$$

в слабых полях

$$H = H_0 = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^3 \rangle}{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle} = \frac{\sum_{\alpha} \sigma_{\alpha} H_{0\alpha}}{\sum_{\alpha} \sigma_{\alpha}}; \quad (45.33)$$

в сильных полях

$$H = H_{\infty} = \frac{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha} \rangle}{\sum_{\alpha} e_{\alpha} n_{\alpha} \langle \mu_{\alpha}^{-1} \rangle} = \frac{\sum_{\alpha} \sigma_{\alpha}}{\sum_{\alpha} \frac{\sigma_{\alpha}}{H_{\infty\alpha}}};$$

$$H^{-1} = \frac{\sum_{\alpha} \sigma_{\alpha} H_{\infty\alpha}^{-1}}{\sum_{\alpha} \sigma_{\alpha}}. \quad (45.34)$$

Рассмотрим полупроводник в области смешанной проводимости (слабое поле):

$$H_0 = \frac{C_p \rho \mu_p^3 - C_n n \mu_n^3}{\rho \mu_p - n \mu_n} = \frac{C_p \mu_p^3}{\rho \mu_p} \frac{1 + x b^3}{1 + x b} = H_p \frac{1 + x b^3}{1 + x b}, \quad (45.35)$$

где

$$H_p = \mu_p^2 C_p = \mu_p^2 \frac{\langle \tau_p^3 \rangle}{\langle \tau_p \rangle^3}. \quad (45.36)$$

Аналогичное соотношение можно получить и для сильного поля. Разобранная теория магнетосопротивления относится к простейшему полупроводнику со скалярной эффективной массой.

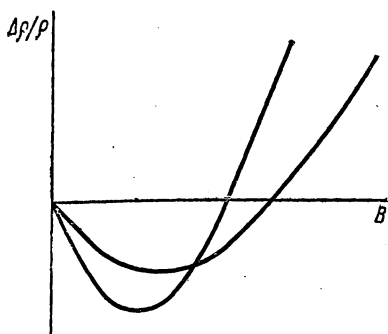


Рис. 69. Отрицательное магнетосопротивление в теллуриде ртути и $\text{Cd}_{0,25}\text{Hg}_{0,75}\text{Te}$

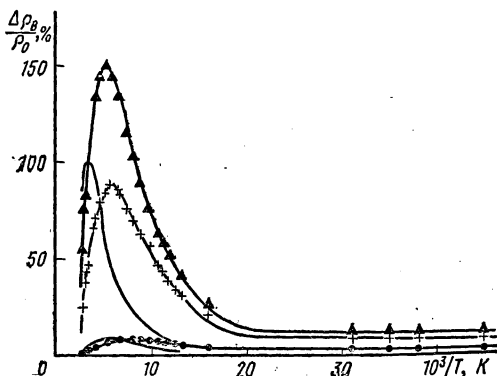


Рис. 70. Температурная зависимость магнетосопротивления в твердом растворе теллуридов кадмия и ртути $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ ($x \approx 0,25$)

В полупроводниках, имеющих тензорную эффективную массу, возможна анизотропия магнетосопротивления. К числу аномалий относится отрицательное магнетосопротивление, которое объясняется в настоящее время как результат проявления примесных состояний.

На рис. 69 приведена зависимость магнетосопротивления от поля в образцах твердого раствора теллуридов кадмия и ртути $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ при $x \approx 0,25$. На рисунке хорошо видна область отрицательного магнетосопротивления. Температура влияет на величину магнетосопротивления, поскольку она меняет концентрацию носителей заряда.

На рис. 70 дана в качестве примера температурная зависимость магнетосопротивления образца твердого раствора теллуридов кадмия и ртути. Видно, что имеет место максимум у зависимости $\frac{\Delta\rho}{\rho_0}(T)$.

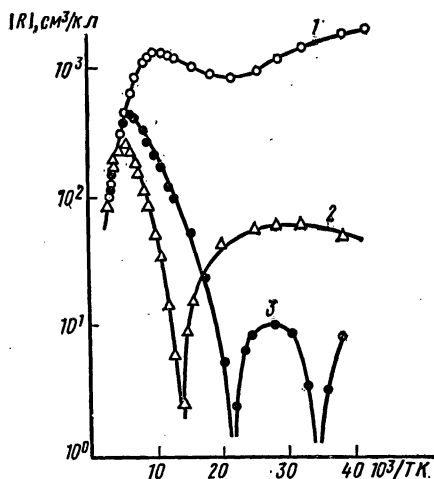


Рис. 71. Температурная зависимость коэффициента Холла образца $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$ ($x \approx 0,25$): 1—2,68; 2—10,78; 3—19,2 кГс

На рис. 71 изображена температурная зависимость коэффициента Холла того же образца. Сравнение рисунков показывает, что максимум магнетосопротивления находится в области инверсии знака

коэффициента Холла, когда происходит смена типа носителей заряда, доминирующих в гальваномагнитных явлениях. В сильном поле (19,2 кГс) наблюдается двойная инверсия коэффициента Холла. Уменьшение магнетосопротивления при комнатной температуре связано с тем, что при этой температуре образец становится почти собственным при вырожденном состоянии электронного газа.

Резюме § 45

1. Изменение сопротивления полупроводника в магнитном поле обусловлено закручивающим действием магнитного поля на движущиеся носители заряда. Магнитнорезистивный эффект характеризуется коэффициентом H :

$$H = \frac{1}{B^2} \frac{\rho_B - \rho_0}{\rho_0} = \frac{1}{B^2} \frac{\sigma_0 - \sigma_B}{\sigma_B}, \quad (45.1p)$$

который позволяет записать сопротивление ρ_B в виде

$$\rho_B = \rho_0 (1 + HB^2). \quad (45.2p)$$

2. В ограниченном полупроводнике в области слабых полей величина H постоянна, в области сильных полей $H \sim B^{-2}$, что приводит к насыщению сопротивления.

3. В неограниченном полупроводнике H постоянно как в области слабых, так и в области сильных полей, но H имеет при этом разное значение.

§ 46. ТЕПЛОПРОВОДНОСТЬ ПОЛУПРОВОДНИКОВ

Если в веществе создан градиент температуры ∇T , то в нем возникает поток энергии W в направлении, противоположном ∇T :

$$W = -\kappa \nabla T, \quad (46.1)$$

$\kappa = |W|/|\nabla T|$ носит название коэффициента теплопроводности, он равен численно количеству энергии, проходящей в единицу времени через поперечное единичное сечение в образце, на концах которого создана разность температур в один градус. В СИ κ измеряется Дж/(с·м²·град/м) = Вт/(м·град). При определении W предполагается, что какие-либо другие процессы в веществе не имеют места. Благодаря теплопроводности происходит передача энергии от «нагревателя» к «холодильнику». В системе, в которой отсутствует источник тепла и поглотитель, но был создан градиент температур, теплопроводность приводит к выравниванию температур, т. е. приводит к установлению теплового равновесия.

Передача тепла осуществляется двумя различными механизмами. Теплопроводность, обусловленную движением носителей заряда, называют электронной или дырочной. Она характеризуется величиной коэффициента теплопроводности κ_e . Второй механизм связан

с тепловыми колебаниями решетки. Атомы (ионы) решетки, совершая колебания относительно положения равновесия, обмениваются энергией. Если в веществе создан градиент температуры, то обмен энергией происходит таким образом, что энергия в большей степени передается от атома, совершающего большие колебания, к атому, совершающему меньшие колебания, т. е. в сторону меньших температур. Теплопроводность, обусловленную колебаниями решетки, называют решеточной, или фононной, теплопроводностью; она характеризуется величиной κ_L . Полная теплопроводность тем самым может характеризоваться величиной κ :

$$\kappa = \kappa_e + \kappa_L. \quad (46.2)$$

Величина κ_L должна быть связана с упругими свойствами твердого тела, κ_e — с концентрацией носителей заряда. В диэлектриках $\kappa_L \gg \kappa_e$. В металлах возможно $\kappa_e \gg \kappa_L$, в полупроводниках величина κ_e должна в сильнейшей степени зависеть от их состава и температуры. Ниже рассмотрим только теплопроводность, обусловленную носителями заряда. Для ее описания запишем выражения для плотности тока (38.20) и плотности потока энергии (38.21) в предположении, что магнитного поля нет ($\mathbf{B} = 0$):

$$\mathbf{j} = \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) eK_{11} - eK_{21} \frac{\nabla T}{T} \quad (46.3)$$

и

$$\mathbf{W} = \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) K_{21} - K_{31} \frac{\nabla T}{T}. \quad (46.4)$$

Найдем выражение для коэффициента теплопроводности κ_e , который определяется условием

$$\mathbf{W} = -\kappa_e \nabla T. \quad (46.5)$$

в предположении, что

$$\mathbf{j} = 0. \quad (46.6)$$

Из условия (46.6) и выражения (46.3) определим величину $\left(e\mathbf{E} - T\nabla \frac{F}{T} \right)$, значение которой подставим в (46.4):

$$e\mathbf{E} - T\nabla \frac{F}{T} = \frac{K_{21}}{K_{11}T} \nabla T, \quad (46.7)$$

$$\mathbf{W} = \frac{K_{21}^2}{K_{11}T} \nabla T - K_{31} \frac{\nabla T}{T} = -\frac{K_{31}K_{11} - K_{21}^2}{K_{11}T} \nabla T. \quad (46.8)$$

Из (46.8) и (46.5) следует

$$\kappa_e = \frac{K_{31}K_{11} - K_{21}^2}{K_{11}T}, \quad (46.9)$$

или

$$\kappa_e = \frac{n}{m^*} \frac{\langle E^3 \tau \rangle \langle E \tau \rangle - \langle E^2 \tau \rangle^2}{\langle E \tau \rangle T}. \quad (46.10)$$

Для невырожденного полупроводника в предположении, что время релаксации является степенной функцией энергии, согласно (39.30) получим

$$\kappa_e = \frac{n}{m^*} \frac{5k^2 T \tau'_0}{4} \frac{\left\{ 7 \frac{\Gamma\left(\frac{9}{2} + p\right) \Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{9}{2}\right) \Gamma\left(\frac{5}{2}\right)} - 5 \frac{\Gamma^2\left(\frac{7}{2} + p\right)}{\Gamma^2\left(\frac{7}{2}\right)} \right\}}{\Gamma\left(\frac{5}{2} + p\right)}. \quad (46.11)$$

В выражение для κ_e входит неизвестная постоянная величина τ'_0 . Ее можно исключить, если рассмотреть отношение κ_e и σ . Точнее, рассмотрим отношение $\frac{\kappa_e}{\sigma T} = L$, оно не зависит от τ'_0 и его называют числом Лоренца:

$$L = \frac{\kappa_e}{\sigma T} = \frac{K_{31} K_{11} - K_{31}^2}{e^2 K_{11}^2 T^2}. \quad (46.12)$$

Для числа Лоренца согласно (46.11) имеем выражение (невырожденный полупроводник):

$$L = \frac{K_{31} K_{11} - K_{31}^2}{e^2 K_{11}^2 T^2} = \frac{k^2}{e^2} \left(\frac{5}{2} + p \right). \quad (46.13)$$

Мы видим, что число Лоренца зависит только от p . Если записать κ_e через σ и L , то получим

$$\frac{\kappa_e}{\sigma} = LT = \frac{k^2 T}{e^2} \left(\frac{5}{2} + p \right), \quad (46.14)$$

$$\kappa_e = \sigma LT = \frac{k^2}{e^2} \left(\frac{5}{2} + p \right) \sigma T. \quad (46.15)$$

Отношение (46.14) выражает закон Видемана — Франца. Для вырожденного полупроводника число Лоренца

$$L = \frac{\kappa_e}{\sigma T} = \frac{\pi^2 k^2}{3 e^2}, \quad (46.16)$$

что получается только с учетом последующих членов разложения (3.42), т. е. во втором приближении.

Если в веществе имеется несколько типов носителей заряда, то выражение для κ_e можно получить, исходя из уравнений

$$\begin{aligned}
 j &= \sum_{\alpha} \left(e_{\alpha}^2 K_{11(\alpha)} \mathbf{E} - T \nabla \frac{F_{\alpha}}{T} e_{\alpha} K_{11(\alpha)} \right) - \sum_{\alpha} e_{\alpha} K_{21(\alpha)} \frac{\nabla T}{T} = \\
 &= \left(\sum_{\alpha} e_{\alpha}^2 K_{11(\alpha)} \right) \mathbf{E} - \left(\sum_{\alpha} e_{\alpha} K_{11(\alpha)} \nabla F_{\alpha} \right) + \\
 &+ \left[\sum_{\alpha} e_{\alpha} (K_{11(\alpha)} F_{\alpha} - K_{21(\alpha)}) \right] \frac{\nabla T}{T} = 0 \quad (46.17)
 \end{aligned}$$

и

$$\begin{aligned}
 W &= \left(\sum_{\alpha} e_{\alpha}^2 K_{21(\alpha)} \right) \mathbf{E} - \left(\sum_{\alpha} K_{21(\alpha)} \nabla F_{\alpha} \right) + \\
 &+ \left[\sum_{\alpha} (K_{21(\alpha)} E_{\alpha} - K_{31(\alpha)}) \right] \frac{\nabla T}{T}. \quad (46.18)
 \end{aligned}$$

В выражениях (46.17) и (46.18) уровень Ферми имеет значок α , его необходимо указывать, поскольку при вычислении кинетических коэффициентов за начало отсчета энергии мы принимали E_0 (E_c или E_v), другими словами, начало отсчета энергии для носителей заряда разного типа различно, и поэтому значение энергии Ферми для каждого типа необходимо брать в той же системе отсчета. Поэтому выражение F для электронов и дырок различно. Их сумма равна

$$F_n + F_p = (F - E_c) + (E_v - F) = -\Delta E_0, \quad (46.19)$$

где F — «истинное» положение уровня Ферми в некоторой общей для электронов и дырок системе отсчета энергии.

Если F_{α} различны, то ∇F_{α} могут быть одинаковы с точностью до знака, поскольку ∇F_{α} определяет силу, действующую на α -тип носителей заряда, обусловленную неравномерностью распределения примеси. Положим

$$\nabla F_{\alpha} = -e_{\alpha} \mathbf{E}^i. \quad (46.20)$$

Выразим величину $\mathbf{E} + \mathbf{E}^i$ из (46.17):

$$\mathbf{E} + \mathbf{E}^i = \frac{\sum_{\alpha} (K_{21(\alpha)} - F_{\alpha} K_{11(\alpha)}) e_{\alpha}}{\left(\sum_{\alpha} e_{\alpha}^2 K_{11(\alpha)} \right)} \frac{\nabla T}{T} \quad (46.21)$$

и подставим в (46.18), после чего получим

$$\begin{aligned}
 W &= -\kappa_e \nabla T = \left(\sum_{\alpha} e_{\alpha} K_{21(\alpha)} \right) \frac{\left[\sum_{\alpha} e_{\alpha} (K_{21(\alpha)} - F_{\alpha} K_{11(\alpha)}) \right]}{\left(\sum_{\alpha} e_{\alpha}^2 K_{11(\alpha)} \right)} + \\
 &+ \sum_{\alpha} (F_{\alpha} K_{21(\alpha)} - K_{31(\alpha)}) \frac{\nabla T}{T}. \quad (46.22)
 \end{aligned}$$

Рассмотрим полупроводник, имеющий смешанную проводимость. Согласно (46.22) для κ_e можно записать

$$\begin{aligned}
 -\kappa_e &= \frac{(K_{21p} - K_{21n}) [K_{21p} - F_p K_{11p} - K_{21n} + F_n K_{11n}]}{(K_{11p} + K_{11n}) T} + \\
 &\quad + \frac{F_p K_{21p} - K_{31p} + F_n K_{21n} - K_{31n}}{T} = \\
 &= \frac{K_{21p}^2 \left(1 - \frac{K_{21n}}{K_{21p}}\right)^2 - K_{11p} \left(F_p - F_n \frac{K_{11n}}{K_{11p}}\right) \left(1 - \frac{K_{21n}}{K_{21p}}\right) K_{21p}}{K_{11p} \left(1 + \frac{K_{11n}}{K_{11p}}\right) T} + \\
 &\quad + \frac{K_{21p}}{T} \left(F_p + F_n \frac{K_{21n}}{K_{21p}}\right) - \frac{K_{31p}}{T} \left(1 + \frac{K_{31n}}{K_{31p}}\right). \quad (46.23)
 \end{aligned}$$

Проводимость можно представить в виде

$$\sigma = e_p^2 (K_{11p} + K_{11n}) = e^2 K_{11p} \left(1 + \frac{K_{11n}}{K_{11p}}\right). \quad (46.24)$$

Найдем значение числа Лоренца или, точнее, $Le^2 T^2$:

$$\begin{aligned}
 Le^2 T^2 &= -\frac{K_{21p}^2 \left(1 - \frac{K_{21n}}{K_{21p}}\right)^2}{K_{11p}^2 \left(1 + \frac{K_{11n}}{K_{11p}}\right)^2} + \frac{K_{21p} \left(1 - \frac{K_{21n}}{K_{21p}}\right) \left(F_p - F_n \frac{K_{11n}}{K_{11p}}\right)}{K_{11p} \left(1 + \frac{K_{11n}}{K_{11p}}\right)^2} - \\
 &\quad - \frac{K_{21p}}{K_{11p}} \frac{\left(F_p + F_n \frac{K_{21n}}{K_{21p}}\right)}{\left(1 + \frac{K_{11n}}{K_{11p}}\right)} + \frac{K_{31p}}{K_{11p}} \frac{\left(1 + \frac{K_{31n}}{K_{31p}}\right)}{\left(1 + \frac{K_{11n}}{K_{11p}}\right)}. \quad (46.25)
 \end{aligned}$$

Рассмотрим отношение кинетических коэффициентов:

$$\frac{K_{11n}}{K_{11p}} = xb = -\frac{n\mu_n}{p\mu_p}; \quad (46.26)$$

$$\frac{K_{21n}}{K_{21p}} = xb \frac{\langle\tau_p\rangle \langle\langle\tau_n\rangle\rangle}{\langle\langle\tau_p\rangle\rangle \langle\tau_n\rangle}; \quad (46.27)$$

$$\frac{K_{31n}}{K_{31p}} = xb \frac{\langle\langle\langle\tau_n\rangle\rangle\rangle \langle\tau_p\rangle}{\langle\tau_n\rangle \langle\langle\langle\tau_p\rangle\rangle\rangle}; \quad (46.28)$$

$$\frac{K_{21p}}{K_{11p}} = \frac{\langle\langle\tau_p\rangle\rangle}{\langle\tau_p\rangle}; \quad \frac{K_{31p}}{K_{11p}} = \frac{\langle\langle\langle\tau_p\rangle\rangle\rangle}{\langle\tau_p\rangle}. \quad (46.29)$$

Для случая, когда рассеяние обусловлено одним механизмом, отношения кинетических коэффициентов можно выразить через гамма-

функцию:

$$\frac{K_{21p}}{K_{11p}} = kT \frac{\Gamma\left(\frac{7}{2} + p\right)}{\Gamma\left(\frac{5}{2} + p\right)} = \left(\frac{5}{2} + p\right) kT; \quad (46.30)$$

$$\frac{K_{31p}}{K_{11p}} = (kT)^2 \frac{\Gamma\left(\frac{9}{2} + p\right)}{\Gamma\left(\frac{5}{2} + p\right)} = \left(\frac{5}{2} + p\right) \left(\frac{7}{2} + p\right) (kT)^2. \quad (46.31)$$

Из полученных выше выражений (46.26—29) видно также, что

$$\frac{K_{31n}}{K_{31p}} = xb \frac{\Gamma\left(\frac{9}{2} + p_n\right)}{\Gamma\left(\frac{9}{2} + p_p\right)}; \quad \frac{K_{21n}}{K_{21p}} = xb \frac{\Gamma\left(\frac{7}{2} + p_n\right)}{\Gamma\left(\frac{7}{2} + p_p\right)}, \quad (46.32)$$

где p_n и p_p — определяют зависимость времени релаксации от энергии для электронов и дырок.

Отношение кинетических коэффициентов $K_{n1}(\alpha)$ равно отношению концентраций и подвижностей, т. е. проводимостей, так что

$$\frac{K_{r1(n)}}{K_{r1(p)}} = xb \frac{\Gamma\left(\frac{3}{2} + r + p_n\right)}{\Gamma\left(\frac{3}{2} + r + p_p\right)} = \frac{\sigma_n}{\sigma_p} \frac{\Gamma\left(\frac{3}{2} + r + p_n\right)}{\Gamma\left(\frac{3}{2} + r + p_p\right)}. \quad (46.33)$$

Учитывая выражения для отношений кинетических коэффициентов, найдем величину L (при $p_n = p_p$):

$$\begin{aligned} Le^2 T^2 = & - \left(\frac{5}{2} + p\right)^2 (kT)^2 \frac{\left(1 - \frac{\sigma_n}{\sigma_p}\right)^2}{\left(1 + \frac{\sigma_n}{\sigma_p}\right)^2} + \left(\frac{5}{2} + p\right) (kT) \frac{\left(1 - \frac{\sigma_n}{\sigma_p}\right) \left(F_p - F_n \frac{\sigma_n}{\sigma_p}\right)}{\left(1 + \frac{\sigma_n}{\sigma_p}\right)^2} - \\ & - \left(\frac{5}{2} + p\right) (kT) \frac{\left(F_p + F_n \frac{\sigma_n}{\sigma_p}\right)}{\left(1 + \frac{\sigma_n}{\sigma_p}\right)} + \left(\frac{5}{2} + p\right) \left(\frac{7}{2} + p\right) (kT)^2. \quad (46.34) \end{aligned}$$

Рассмотрим случай $\sigma_n/\sigma_p = 1$. Из (46.34) запишем

$$Le^2 T^2 = \left(\frac{5}{2} + p\right) \left[\left(\frac{7}{2} + p\right) + \frac{\Delta E_0}{2kT} \right] (kT)^2 \quad (46.35)$$

$$L = \frac{k^2}{e^2} \left(\frac{5}{2} + p\right) \left[\left(\frac{7}{2} + p\right) + \frac{\Delta E_0}{2kT} \right]. \quad (46.36)$$

Если сравнить (46.36) с (46.13), то увидим, что в случае носителей заряда двух типов коэффициент Лоренца, а следовательно, и теплопроводность возрастают в $\left[\frac{7}{2} + p + \frac{\Delta E_0}{2kT} \right]$ раз по сравнению со случаем, когда имеются носители заряда одного типа.

Этот результат имеет очень наглядное физическое объяснение. При создании градиента температур возникает диффузионный поток носителей заряда. Но его возникновение приводит к разделению зарядов и, следовательно, к появлению электрического поля, которое препятствует движению носителей заряда, приводящему к разделению зарядов. По исходному условию (46.6) полный ток $j = 0$. Следовательно, *потоки носителей заряда вдоль и против ∇T должны быть равны*, но так как носители заряда, идущие вдоль ∇T (к горячему концу), имеют меньшую энергию, чем носители заряда, идущие против ∇T (к холодному концу), то *происходит перенос энергии от горячего конца к холодному без переноса заряда*. Теплопроводность при этом оказывается сравнительно небольшой, и она не зависит от положения уровня Ферми, поскольку весь эффект основан на разности энергий потоков носителей заряда, идущих от горячего и холодного концов образца. Если же в полупроводнике имеются носители заряда *двух типов разного знака*, то диффузионные потоки носителей заряда приводят к возникновению электрических полей, направленных в противоположные стороны. Поэтому *суммарное электрическое поле будет небольшим* — оно должно удовлетворять условию $j = 0$, но одинаково направленные потоки носителей заряда разного знака дают токи, противоположно направленные. Поскольку *электрическое поле мало, оно не препятствует движению носителей заряда, поэтому поток частиц будет большим*. При этом он обусловлен градиентом концентрации, однако выравнивания концентраций не произойдет, поскольку в области горячего конца происходит *непрерывное преимущественное рождение пар носителей заряда*, а в области холодного конца *рекомбинация носителей заряда превалирует над их генерацией*. Но при каждой рекомбинации пары носителей заряда выделяется энергия ΔE_0 , затраченная на их генерацию на горячем конце. Другими словами, *каждая пара носителей заряда переносит дополнительную энергию ΔE_0 , что и приводит к столь резкому возрастанию числа Лоренца и, следовательно, коэффициента теплопроводности*.

Для собственного полупроводника $\frac{\sigma_n}{\sigma_p} = b$ и

$$L = \frac{k^2}{e^2} \left(\frac{5}{2} + p \right) \left\{ \left(\frac{7}{2} + p \right) + \frac{2b\Delta E_0}{(1+b)^2 kT} - \left(\frac{5}{2} + p \right) \left(\frac{1-b}{1+b} \right)^2 \right\}. \quad (46.37)$$

Из (46.37) видно, что при $b = 1$ выражение для L совпадает с (46.36), поскольку при этом $\sigma_n = \sigma_p$. Если же $b \neq 1$, то L уменьшается по сравнению со случаем $b = 1$ (или $\sigma_n = \sigma_p$), уменьшение будет тем больше, чем больше отличаются подвижности.

При $b \gg 1$ выражение (46.37) можно представить в виде

$$L = \left(\frac{5}{2} + p \right) \left[1 + \frac{2\Delta E_0}{bkT} \right] \frac{k^2}{e^2}. \quad (46.38)$$

При $b \rightarrow \infty$ выражение (46.38) переходит в выражение (46.42), поскольку это равносильно переходу к проводимости одного типа.

§ 47. ТЕРМОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ

К термоэлектрическим явлениям относятся три эффекта: эффект Зеебека, эффект Пельтье и эффект Томсона. Рассмотрим качественно эти явления.

1. Явление Зеебека, или термоэлектрический эффект. Пусть имеются два образца 1 и 2 различных веществ, находящихся в контакте (рис. 72). Если температура контактов различна — $T + dT$ и T , то в замкнутой цепи возникает ток, который носит название термоэлектрического. Если цепь разорвать в произвольном месте, то на концах разомкнутой цепи появляется разность потенциалов, носящая название термоэлектродвижущей силы (термо-э. д. с.). Зеебек, открывший это явление, установил, что разность потенциалов $d\mathcal{E}_{12}$ в разомкнутой цепи зависит от разности температур и вида материалов:

$$d\mathcal{E}_{12} = \alpha_{12} dT. \quad (47.1)$$

$d\mathcal{E}_{12}$ (и α_{12}) принято считать положительной величиной, если потенциал «горячего» контакта выше потенциала «холодного» контакта, как это изображено на рис. 72. Кроме того, $d\mathcal{E}_{12} = -d\mathcal{E}_{21}$, или $\alpha_{12} = -\alpha_{21}$. Согласно рис. 72, если $d\mathcal{E}_{12} > 0$, то ток течет по часовой стрелке; если $d\mathcal{E}_{12} < 0$, то ток течет против часовой стрелки. Другими словами, $d\mathcal{E}_{12} > 0$ означает переход положительного заряда от 1 к 2 в «горячем» контакте, а $d\mathcal{E}_{12} < 0$ — в «холодном» контакте. Величина α_{12} характеризует пару веществ, она называется дифференциальной термо-э. д. с. Зависеть α_{12} может от температуры T , поэтому термо-э. д. с., возникающая в цепи, контакты в которой имеют конечную разность температур $T_2 - T_1$, равна

$$\mathcal{E}_{12} = \int_{T_1}^{T_2} \alpha_{12}(T) dT. \quad (47.2)$$

Рассмотрим картину возникновения термо-э. д. с. на контакте металлов. Как известно, в любой системе, находящейся в состоянии термодинамического равновесия, электрохимический потенциал (уровень Ферми) один и тот же в любой части системы. Если приведены в контакт два металла, то их уровни Ферми должны совпасть. Но если контактирующие металлы имеют различную концентрацию электронов, то энергии Ферми, отсчитанные от дна зоны проводимости в каждом металле, будут различны, и тем самым E_{c1} не совпадет с E_{c2} , поскольку энергии Ферми разны. Разность $E_{c2} - E_{c1}$ представляет собой потенциальный барьер, возникающий на контакте. Она получила название внутренней контактной разности потен-

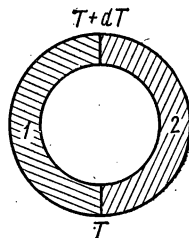


Рис. 72. Схема наблюдения термоэлектрического эффекта

циалов U^i . Определяется U^i разностью энергий Ферми в исходных металлах:

$$U^i = \frac{F_1 - F_2}{e}. \quad (47.3)$$

На контакте тем самым существует электрическое поле, локализованное в тонком приконтактном слое.

Если составить замкнутую цепь из двух металлов, то U^i возникает на обоих контактах.

Очевидно, что электрическое поле будет направлено одинаковым образом в обоих контактах — от большего F к меньшему. Это значит, что если совершить обход по замкнутому контуру, то в одном контакте обход будет происходить по полю, а в другом — против поля. *Циркуляция вектора E тем самым будет равна нулю.*

Предположим теперь, что температура одного из контактов изменилась на dT . Так как энергия Ферми зависит от температуры, U^i изменится.

Но если изменилась внутренняя контактная разность потенциалов, то изменилось электрическое поле в одном из контактов, и поэтому циркуляция вектора E будет отлична от нуля, т. е. *появляется э. д. с. в замкнутой цепи.*

2. Явление Пельтье, или электротермический эффект Пельтье. *Если через контакт двух веществ пропустить электрический ток, то в контакте происходит или выделение, или поглощение тепла в зависимости от направления тока. Это явление называют явлением Пельтье.* Если изменить направление тока, то меняется знак эффекта. Величина выделяемого тепла и его знак зависят от вида контактирующих веществ, силы тока и времени прохождения тока, т. е. количество выделяемого тепла пропорционально количеству прошедшего через контакт заряда $dq = I dt$:

$$dQ_{12} = \Pi_{12} I dt. \quad (47.4)$$

dQ_{12} (и Π_{12}) означает, что ток идет от первого вещества ко второму, в то время как dQ_{21} (и Π_{21}) относится к случаю противоположному — ток идет от второго к первому веществу. Очевидно,

$$dQ_{21} = \Pi_{21} I dt = -dQ_{12} = -\Pi_{12} I dt. \quad (47.5)$$

Если теплота выделяется, то она считается больше нуля. Таким образом, $\Pi_{ij} > 0$, если при прохождении тока из i образца в j теплота выделяется.

Было найдено, что *если внешний ток совпадает по направлению с термотоком, который возникает при нагревании данного контакта, то этот контакт охлаждается.* Это легко понять на основе закона сохранения энергии. Если нагревать некоторый контакт, что приводит к возникновению термотока, то его направление должно быть таким, чтобы подводимая теплота в нем поглощалась. Поэтому если внешний ток имеет такое же направление, что и термоток, то этот контакт должен охлаждаться.

Причину возникновения явления Пельтье легко понять, если обратиться к рис. 73, на котором изображена энергетическая картина контакта металла и полупроводника. Как известно, в металле проводимость осуществляется электронами с энергией вблизи уровня Ферми. В полупроводнике (в данном случае электронном) ток переносится электронами зоны проводимости. Из рис. 73 видно, что средняя энергия электронов проводимости в полупроводнике больше, чем в металле, на величину не меньшую, чем $E_c - F$. Чтобы электроны из металла могли перейти в полупроводник, они должны преодолеть потенциальный барьер высотой по крайней мере $E_c - F$; для этого они должны получить энергию от решетки, что приводит к охлаждению металла в области контакта. Ясно, что количество необходимой энергии зависит от числа прошедших через контакт электронов, т. е. от прошедшего заряда. Если изменить направление тока, то теперь электроны, переходя из полупроводника в металл, будут иметь избыток энергии по сравнению с электронами проводимости металла. Приходя в равновесие с ними, электроны, прошедшие через контакт, отдадут избыток энергии (не менее $E_c - F$) решетке, что и приведет к выделению тепла на контакте. Из механизма возникновения эффекта Пельтье следует, что для контакта металл — металл коэффициенты Пельтье должны иметь значительно меньшую величину, чем на контакте металла и полупроводника или двух полупроводников.

Можно подойти к пониманию явления Пельтье несколько иначе. На контакте двух веществ имеется внутреннее контактное поле. Если через контакт идет ток, то контактное поле будет либо способствовать прохождению тока, либо препятствовать. Если ток идет против контактного поля, то внешний источник должен затратить дополнительную энергию, которая выделяется в контакте, что приведет к его нагреву. Если же ток идет по направлению контактного поля, то он может поддерживаться этим полем, которое и совершает работу по перемещению зарядов. Необходимая для этого энергия отбирается у вещества, что приводит к охлаждению его в месте контакта.

Эффекты Зеебека и Пельтье являются не только контактными явлениями, но и объемными. Они могут наблюдаться в объеме неоднородного полупроводника.

3. Явление Томсона, или электротермический эффект Томсона. Если полупроводник нагрет неравномерно, то концентрация носителей заряда в нем будет больше там, где выше температура, поэтому градиент температуры приводит к градиенту концентрации, вследствие чего возникает диффузионный поток носителей заряда. Это нарушает электронейтральность. Разделение зарядов порождает электрическое поле, препятствующее разделению.

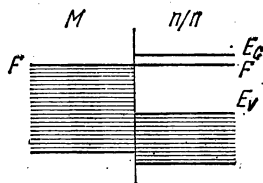


Рис. 73. Энергетическая схема контакта металла с электронным полупроводником (с равными работами выхода).

Таким образом, если в полупроводнике имеется градиент температуры, то в нем имеется объемное электрическое поле E^i .

Предположим теперь, что через такой образец пропускается электрический ток под действием внешнего электрического поля E . Если ток идет против внутреннего поля E^i , то внешнее поле должно совершать дополнительную работу при перемещении зарядов относительно поля E^i , что приведет к выделению тепла, дополнительного к ленц-джоулевым потерям. Если ток (или внешнее поле E) направлен по E^i , то E^i само совершает работу по перемещению зарядов для создания тока. В этом случае внешний источник тратит энергию для поддержания тока меньшую, чем в том случае, когда внутреннего поля E^i нет. Работа поля E^i может совершаться только за счет тепловой энергии самого проводника, поэтому он охлаждается. Явление выделения или поглощения тепла в проводнике, обусловленное градиентом температуры, при прохождении тока носит название эффекта Томсона. Таким образом, вещество нагревается, когда поля E и E^i противоположно направлены, и охлаждается, когда их направления совпадают. Томсон нашел, что выделение тепла в объеме dV определяется соотношением

$$dQ^T = -\tau (\nabla T) dt dV; \quad dV = Sdl, \quad (47.6)$$

τ называют коэффициентом Томсона.

Между коэффициентами α_{12} , Π_{12} и τ существует определенная связь, которую можно установить на основе законов термодинамики. Термоэлектрические явления лежат в основе работы термогенераторов, термопар и других приборов. Они играют роль при любых измерениях в физике полупроводников. Если при измерениях контакты измерительной цепи будут находиться при различной температуре, то термо-э. д. с. может существенно исказить результаты измерений.

Перейдем к теоретическому описанию термоэлектрических явлений на основе кинетического уравнения. Запишем выражения для тока и потока тепла в общем случае, но без магнитного поля: $\mathbf{B} = 0$

$$\mathbf{j} = e^2 K_{11} \mathbf{E} - e K_{11} T \nabla \frac{F}{T} - e K_{21} \frac{\nabla T}{T}, \quad (47.7)$$

$$\mathbf{W} = e K_{21} \mathbf{E} - K_{21} T \nabla \frac{F}{T} - K_{31} \frac{\nabla T}{T}. \quad (47.8)$$

Выразим поле \mathbf{E} из (47.7) и подставим в (47.8):

$$\begin{aligned} \mathbf{E} &= \frac{1}{e^2 K_{11}} \left[\mathbf{j} + e K_{11} T \nabla \frac{F}{T} + e K_{21} \frac{\nabla T}{T} \right] = \\ &= \frac{\mathbf{j}}{e^2 K_{11}} + \frac{1}{e} \nabla F + \frac{K_{21} - F K_{11}}{e K_{11} T} \nabla T. \end{aligned} \quad (47.9)$$

Поле \mathbf{E} состоит из трех частей: первый член обусловлен током \mathbf{j} , т. е. омическим падением напряжения, второй член связан с неоднородностью

родностью вещества и третий — с неизотермичностью. Найдем W , подставив (47.9) в (47.8):

$$W = \frac{eK_{21}}{e^2K_{11}} \left\{ \mathbf{j} + eK_{11}T\nabla \frac{F}{T} + eK_{21} \frac{\nabla T}{T} \right\} - K_{21}T\nabla \frac{F}{T} - K_{31} \frac{\nabla T}{T} = \\ = \frac{K_{21}}{eK_{11}} \mathbf{j} - \frac{K_{31}K_{11} - K_{21}^2}{TK_{11}} \nabla T. \quad (47.10)$$

Выражение (47.10) показывает, что энергия переносится в результате направленного движения носителей заряда, приводящего к возникновению тока \mathbf{j} , и хаотического движения носителей заряда, обуславливающего теплопроводность κ_e :

$$W = \frac{K_{21}}{eK_{11}} \mathbf{j} - \kappa_e \nabla T = \Pi \mathbf{j} - \kappa_e \nabla T. \quad (47.11)$$

Вернемся к выражению для поля \mathbf{E} . Если в веществе тока нет ($\mathbf{j} = 0$), то

$$\mathbf{E} = \frac{1}{e} \nabla F + \frac{K_{21} - FK_{11}}{eTK_{11}} \nabla T. \quad (47.12)$$

Поле, обусловленное градиентом температуры, носит название термоэлектрического:

$$\mathbf{E}^\alpha = \frac{K_{21} - FK_{11}}{eTK_{11}} \nabla T = \alpha \nabla T. \quad (47.13)$$

Величину α называют абсолютной дифференциальной термоэлектродвижущей силой, она равна

$$\alpha = \frac{K_{21} - FK_{11}}{eTK_{11}}. \quad (47.13')$$

Введем абсолютную интегральную термо-э. д. с. V_T условием

$$V_T = \int_0^T \alpha(\xi) d\xi, \quad (47.14)$$

V_T зависит от координаты, поскольку от координаты зависит температура. Найдем градиент V_T :

$$\nabla V_T = \frac{dV_T}{dT} \nabla T = \alpha(T) \nabla T = \mathbf{E}^\alpha. \quad (47.15)$$

Следовательно, V_T есть потенциал данной точки, взятый с обратным знаком:

$$V_T(T(\mathbf{r})) = -\varphi(T(\mathbf{r})). \quad (47.16)$$

Определяя разность потенциалов между точками 1 и 2 обычным соотношением

$$\varphi_{12} = \varphi(1) - \varphi(2), \quad (47.17)$$

можем записать

$$\varphi_{12} = V_T(2) - V_T(1). \quad (47.18)$$

Поскольку электрическое поле E^α определяется через градиент некоторой функции, т. е. является потенциальным, то циркуляция E^α по замкнутому контуру должна быть равна нулю:

$$\mathcal{E} = - \oint (E^\alpha dl) = 0, \quad (47.19)$$

т. е. э. д. с. \mathcal{E} равна нулю. Отсюда следует невозможность определения абсолютной термоэлектродвижущей силы α или V_T . Определенное соотношением (47.12) поле E^α не может привести к возникновению э. д. с. и тока в замкнутой цепи, образованной одним и тем же веществом. Однако если цепь составлена из двух или более различных веществ, то циркуляция E^α может быть отлична от нуля, и тем самым в замкнутой цепи может течь ток. Рассмотрим с этой целью цепь, состоящую из двух различных веществ 1 и 2 (рис. 72). Вычислим циркуляцию вектора E^α :

$$\begin{aligned} \mathcal{E} = \mathcal{E}_{12} &= - \oint (E^\alpha dl) = - \int_A^B (E^\alpha dl)_1 - \int_B^A (E^\alpha dl)_2 = \\ &= - \int_A^B \alpha_1 (\nabla T dl) + \int_A^B \alpha_2 (\nabla T dl) = \int_{T(A)}^{T(B)} (\alpha_2 - \alpha_1) dT = \int_{T(A)}^{T(B)} \alpha_{12} dT. \end{aligned} \quad (47.20)$$

Выражение (47.20) совпадает с (47.2). Отсюда следует, что относительная дифференциальная термо-э. д. с. α_{12} представляет собой разность абсолютных дифференциальных термо-э. д. с.:

$$\alpha_{12} = \alpha_2 - \alpha_1. \quad (47.21)$$

В вырожденных полупроводниках и металлах $K_{21} = FK_{11}$, поэтому $\alpha = 0$. Чтобы получить отличные от нуля значения, необходимо учесть в разложении $f_0(E)$ последующие члены. Расчет дает в этом случае

$$\alpha = \frac{\pi^2}{3} \frac{k^2 T}{e} \frac{\partial [\ln \sigma(E)]}{\partial E} \Big|_{E=F}, \quad (47.22)$$

где

$$\frac{\partial [\ln \sigma(E)]}{\partial E} \Big|_{E=F} \approx F^{-1}, \quad (47.23)$$

так что

$$\alpha \approx \frac{\pi^2}{3} \frac{k^2 T}{eF}. \quad (47.24)$$

При $T \approx 300$ К и $F \sim 10$ эВ $\alpha \approx 10^{-6}$ В/град. Такого же порядка должна быть и относительная термо-э. д. с. Выражение (47.24) вытекает из температурной зависимости энергии Ферми. Действительно, если учесть, что $\frac{dF}{dT} \neq 0$, и определить α соотношением

$$\alpha = \frac{1}{e} \left[\frac{dF}{dT} + \frac{K_{21} - FK_{11}}{TK_{11}} \right], \quad (47.25)$$

то при $K_{21} = FK_{11}$ получим

$$\alpha = \frac{1}{e} \frac{dF}{dT} = - \frac{\pi^2}{e} \frac{k^2 T}{F}. \quad (47.26)$$

Рассмотрим теперь невырожденный полупроводник. Так как согласно (46.28) $\frac{K_{21}}{K_{11}} = \frac{\langle\langle\tau\rangle\rangle}{\langle\tau\rangle}$, то

$$\alpha = \frac{1}{eT} \left[\frac{\langle\langle\tau\rangle\rangle}{\langle\tau\rangle} - (F - E_c) \right]. \quad (47.27)$$

В выражении (47.27) вместо F мы подставили $F - E_c$, поскольку энергия отсчитывается от дна зоны проводимости (или потолка валентной зоны для дырок). Если рассеяние определяется каким-либо одним механизмом, то согласно (46.31)

$$\frac{K_{21}}{K_{11}} = \left(\frac{5}{2} + p \right) kT$$

и

$$\alpha = \frac{k}{e} \left[\frac{5}{2} + p - \frac{F - E_c}{kT} \right] = \frac{k}{e} \left[\frac{5}{2} + p + \ln \frac{N_c}{n} \right]. \quad (47.28)$$

Мы видим, что α сложным образом зависит от состава и температуры полупроводника. Так как $n < N_c$, то выражение в скобках больше нуля и знак α определяется знаком носителя заряда.

Оценим величину α , используя соотношение (47.28). Полагая, что F находится в середине между E_c и E_d , и считая, что $\Delta E_d \approx \approx 0,05$ эВ, а $2kT \approx 0,01$ эВ (что соответствует $T \approx 60$ К), получим $\alpha \approx \frac{k}{e} \left[\frac{5}{2} + p + 5 \right]$; $\alpha \approx 9 \frac{k}{e}$ при $p = 3/2$. Учитывая, что $\frac{k}{e} = = 8,62 \cdot 10^{-5}$ В/град = 86,2 мкВ/град, получим $\alpha \approx 0,7$ мВ/град, что почти на три порядка превосходит термо-э. д. с. в металлах.

Выражение (47.28) не применимо при очень низких температурах, поскольку $\alpha \rightarrow \infty$ при $T \rightarrow 0$, в то время как при $T \rightarrow 0$ согласно термодинамике α должна также стремиться к нулю.

Перейдем теперь к более высоким температурам, соответствующим области истощения, в которой $N_c > N_d$ и $n \approx N_d$. В этой области температур

$$\alpha = \frac{k}{e} \left(\frac{5}{2} + p + \ln \frac{N_c}{N_d} \right), \quad (47.29)$$

т. е. α растет с ростом температуры.

При контакте полупроводника и металла относительная термо-э. д. с. будет равна фактически термо-э. д. с. полупроводника.

Если определять α соотношением (47.25), то выражения для термо-э. д. с. в вышеперезообразованных случаях будут несколько иными.

На совпадении знака α со знаком носителя заряда основан простейший способ определения типа проводимости, или знака носителя заряда, методом термозонда. На металлической пластине располагается исследуемый материал, к которому прижимается нагретый зонд, соединенный с пластиной через гальванометр. В контакте зонда с образцом температура повышается, концентрация носителей заряда возрастает, они диффундируют в глубь образца. Возникает электрическое поле, зонд заряжается знаком неосновных носителей

заряда — в материале n -типа он заряжается положительно по отношению к пластине, в образце p -типа — отрицательно. В цепи возникает ток, направление которого определяется гальванометром. Установив заранее связь между типом носителей заряда и направлением отклонения стрелки гальванометра, можно легко определить знак носителей заряда.

Чтобы найти значение α , соответствующее переходу от истощения примеси к собственной концентрации, необходимо помнить, что в области собственной концентрации мы имеем дело с носителями заряда двух типов, и поэтому выражение, полученное для примесного полупроводника, необходимо обобщить на случай собственной проводимости, что будет сделано ниже.

Чтобы получить описание эффектов Пельтье и Томсона, необходимо рассмотреть баланс энергии в полупроводнике. Прежде всего отметим, что с точки зрения механизма выделения или поглощения тепла эффекты Пельтье и Томсона не отличаются друг от друга. И в том, и в другом случае энергия выделяется в результате работы внутренних полей при перемещении заряда. При этом эффект Томсона представляет собой выделение энергии в термоэлектрическом поле $E^{\alpha} = \alpha \nabla T$, а эффект Пельтье — выделение энергии в поле $E^{\Pi} = -\frac{1}{e} \nabla F$, обусловленном неоднородностью вещества.

Итак, пусть в веществе имеется ∇F и создан градиент температуры, что приведет к возникновению термоэлектрического поля E^{α} . Пусть в веществе идет ток с плотностью j . В единице объема в единицу времени будет выделяться энергия \dot{Q} , обусловленная работой \dot{A} внешних и внутренних сил при перемещении заряда. Кроме того, необходимо учесть расходимость теплового потока W в каждой точке, поэтому

$$\dot{Q} = -\operatorname{div} W + \dot{A}. \quad (47.30)$$

Для того чтобы найти \dot{A} , необходимо учесть следующее. Ток вызывается внешним источником, который создает в веществе внешнее поле $E^{\text{стор}}$. В каждой точке полное поле E равно сумме внутренних полей $E^i = -\frac{1}{e} \nabla F + \alpha \nabla T$ и поля $E^{\text{стор}}$:

$$E = E^{\text{стор}} + E^i. \quad (47.31)$$

Ток j в каждой точке должен быть равен

$$j = \sigma (E^{\text{стор}} + E^i) = e^2 K_{11} (E^{\text{стор}} + E^i). \quad (47.32)$$

Выделяемая в единичном объеме мощность \dot{A} равна

$$\dot{A} = (Ej) = \sigma (E^{\text{стор}} + E^i)^2 = \sigma (E^{\text{стор}})^2 + \sigma E^{i^2} + 2\sigma (E^i E^{\text{стор}}). \quad (47.33)$$

Внешний же источник затрачивает энергию

$$(jE^{\text{стор}}) = (E^{\text{стор}}, \sigma [E^{\text{стор}} + E^i]) = \sigma (E^{\text{стор}})^2 + \sigma (E^i E^{\text{стор}}). \quad (47.34)$$

Величина $\sigma E^{i^2} + \sigma (E^i E^{\text{стор}})$ представляет собой работу в единицу времени внутренних сил по перемещению заряда при прохождении

тока. Эта работа может совершаться только за счет тепловой энергии тела. Если эта работа положительна, то тело должно охлаждаться, а если отрицательна, то тело должно нагреваться.

Все эти процессы должны описываться кинетическим уравнением. Найдем величину \dot{Q} . Так как

$$\dot{A} = (Ej), \quad (47.35)$$

то, учитывая выражение (47.7) для j и (47.9) для E , можем записать

$$\begin{aligned} \dot{A} = (jE) &= \left(j, \frac{j}{e^2 K_{11}} + \frac{1}{e} \nabla F + \frac{K_{21} - FK_{11}}{e K_{11} T} \nabla T \right) = \\ &= \frac{j^2}{\sigma} + \frac{1}{e} (j \nabla F) + \alpha (j \nabla T). \end{aligned} \quad (47.36)$$

Из (47.36) мы видим, что выделяемая энергия обусловлена ленц-джоулевым теплом $\frac{j^2}{\sigma}$ и работой сил, определяемых градиентом уровня Ферми и температуры. Учитывая, что

$$\begin{aligned} \nabla F + \frac{K_{21} - FK_{11}}{TK_{11}} \nabla T &= T \nabla \frac{F}{T} + \frac{K_{21}}{K_{11}} \frac{\nabla T}{T} = T \left(\nabla \frac{F}{T} - \frac{K_{21}}{K_{11}} \nabla \frac{1}{T} \right) = \\ &= -T \nabla \left(\frac{K_{21} - FK_{11}}{TK_{11}} \right) = -Te \nabla \alpha, \end{aligned} \quad (47.37)$$

получим для \dot{A} :

$$\dot{A} = \frac{j^2}{\sigma} - (j, T \nabla \alpha) \quad (47.38)$$

и

$$\dot{Q} = \frac{j^2}{\sigma} + \text{div } \kappa_e \nabla T - (j, T \nabla \alpha). \quad (47.39)$$

Обозначим теплоту Томсона \dot{Q}^T соотношением

$$\dot{Q}^T = -\tau (j \nabla T). \quad (47.40)$$

Сравнивая (47.40) и (47.39), можем записать

$$\tau \nabla T = T \nabla \alpha. \quad (47.41)$$

Так как

$$\nabla \alpha = \frac{\partial \alpha}{\partial T} \nabla T, \quad (47.42)$$

то

$$\tau = T \frac{\partial \alpha}{\partial T}, \quad (47.43)$$

или

$$\alpha(T) = \int_0^T \frac{\tau(\xi)}{\xi} d\xi. \quad (47.44)$$

Таким образом, выражение (47.44), полученное на основе кинетического уравнения, связывает α с τ .

Вернемся к коэффициенту Пельтье, который определили, исходя из уравнения

$$W = \Pi j - \kappa_e \nabla T. \quad (47.45)$$

В феноменологической теории явлений переноса принято вводить коэффициент Пельтье π соотношением

$$W = \left(\pi + \frac{\tilde{\mu}}{e} \right) j - \kappa_e \nabla T, \quad (47.46)$$

где $\tilde{\mu}$ — электрохимический потенциал:

$$\tilde{\mu} = F + e\varphi, \quad (47.47)$$

и

$$j = -\frac{1}{e} \sigma \nabla \tilde{\mu} - \sigma \alpha \nabla T. \quad (47.48)$$

Сравнивая (47.46) и (47.45), можно записать

$$\Pi = \pi + \frac{F}{e} + \varphi. \quad (47.49)$$

Мы видим, что π определяет поток кинетической энергии, в то время как Π определяет поток полной энергии, включая электрохимический потенциал каждого заряда.

Положим $\varphi = 0$, в этом случае для π имеем

$$\pi = \Pi - \frac{F}{e} = \frac{K_{21}}{eK_{11}} - \frac{F}{e} = \frac{1}{e} \left(\frac{K_{21} - FK_{11}}{K_{11}} \right). \quad (47.50)$$

Учитывая (47.13), можем записать

$$\pi = \alpha T. \quad (47.51)$$

Рассмотрим теплоту Пельтье. Поскольку

$$\dot{Q} = -\operatorname{div} W + \dot{A}, \quad (47.52)$$

а работа внешних и внутренних сил должна компенсироваться в объеме тела только расходимостью теплового потока (в стационарных условиях), то рассмотрим $\operatorname{div} W$:

$$\operatorname{div} W = \operatorname{div} \Pi j - \operatorname{div} \kappa_e \nabla T. \quad (47.53)$$

Так как нас интересует теплота, обусловленная потоком Пельтье, то

$$\operatorname{div} \Pi j = (\nabla \Pi, j) - \Pi \operatorname{div} j = (j, \nabla \Pi); \operatorname{div} j = 0. \quad (47.54)$$

Найдем

$$\begin{aligned} \nabla \Pi &= \nabla \left(\pi + \frac{F}{e} \right) = \nabla \left(\alpha T + \frac{F}{e} \right) = T \nabla \alpha + \alpha \nabla T + \frac{1}{e} \nabla F = \\ &= \tau \nabla T + \alpha \nabla T + \frac{1}{e} \nabla F. \end{aligned} \quad (47.55)$$

Мы видим, что расходимость потока Пельтье действительно связана с работой полей, обусловленных градиентом уровня Ферми. Объемный эффект Пельтье обычно не играет большой роли, поэтому рассмотрим только контактный эффект. При прохождении тока через контакт двух веществ расходимость потока будет равна разности потоков:

$$Q_{12}^{\Pi} = W_2 - W_1 = \Pi_{12} j_{12}. \quad (47.56)$$

Но

$$W_1 = \Pi_1 j; \quad W_2 = \Pi_2 j$$

и

$$\Pi_{12} = \Pi_2 - \Pi_1. \quad (47.57)$$

Относительный коэффициент Пельтье, т. е. коэффициент, определяющий выделение тепла на контакте двух веществ при прохождении тока из первого во второе вещество, равен разности абсолютных коэффициентов Пельтье. Таким образом,

$$\begin{aligned} \Pi_{12} = \Pi_2 - \Pi_1 &= \left(\pi_2 + \frac{F_2}{e} \right) - \left(\pi_1 + \frac{F_1}{e} \right) = \\ &= \pi_2 - \pi_1 + \frac{1}{e} (F_2 - F_1). \end{aligned} \quad (47.58)$$

Для идеального контакта $F_2 = F_1$ и

$$\Pi_{12} = \pi_2 - \pi_1 = (\alpha_2 - \alpha_1) T = \alpha_{12} T, \quad (47.59)$$

т. е.

$$\Pi_{12} = \alpha_{12} T = \pi_{12}. \quad (47.60)$$

Таким же образом можно определить остальные соотношения Томсона для относительных термоэлектрических коэффициентов.

В заключение этого параграфа рассмотрим термо-э. д. с. в проводнике, имеющем смешанную проводимость. Так как

$$j = \sum_i e_i K_{11(i)} (e_i E - \nabla F_i) + \sum_i e_i (K_{11(i)} F_i - K_{21(i)}) \frac{\nabla T}{T}, \quad (47.61)$$

то из условия $j = 0$ и $\nabla F_i = 0$ можем определить термоэлектрическое поле E^α :

$$E^\alpha = \frac{\sum_i e_i (K_{21(i)} - F_i K_{11(i)})}{T \sum_i e_i^2 K_{11(i)}} \nabla T = \alpha \nabla T. \quad (47.62)$$

Пусть полупроводник имеет электроны и дырки, тогда

$$\alpha = \frac{(K_{21p} - F_p K_{11p}) - (K_{21n} - F_n K_{11n})}{e_p T (K_{11p} + K_{11n})}. \quad (47.63)$$

Учитывая выражение (47.13) для термо-э. д. с. в случае носителей заряда одного типа, можем записать для (47.63):

$$\alpha = \frac{\alpha_p K_{11p} + \alpha_n K_{11n}}{K_{11p} + K_{11n}} = \frac{\alpha_p \sigma_p + \alpha_n \sigma_n}{\sigma_p + \sigma_n}. \quad (47.64)$$

Аналогично можно получить

$$\pi = \frac{\pi_p \sigma_p + \pi_n \sigma_n}{\sigma_p + \sigma_n}. \quad (47.65)$$

Для собственного полупроводника

$$\alpha = \frac{\alpha_p + \alpha_n b}{1 + b}; \quad \pi = \frac{\pi_p + \pi_n b}{1 + b}. \quad (47.66)$$

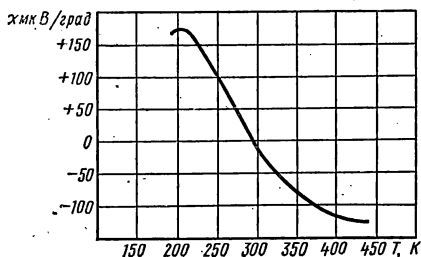
Уменьшение термо-э. д. с. и потока энергии в полупроводнике, имеющем собственную или смешанную проводимость по сравнению с примесным, объясняется тем, что направление термоэлектрического поля электронов и дырок и их потоков энергии имеют противоположный знак и, следовательно, взаимно ослабляют друг друга.

Уменьшение этого поля увеличивает диффузионные потоки электронов и дырок, что приводит к резкому возрастанию теплопереноса из-за рекомбинации носителей заряда на холодном конце, о чем было сказано выше.

Выбор материала для термопар, термоэлектрических генераторов и холодильников определяется по величине параметра эффективности Z , который задается соотношением

$$Z = \frac{\alpha^2 \sigma}{\kappa} = \frac{\alpha^2}{\rho \kappa}. \quad (47.67)$$

Рис. 74. Температурная зависимость дифференциальной термоэлектродвижущей силы в контакте медь — теллурид ртути



Смысл введения величины Z можно понять на основе следующих соображений. Термоэлектродвижущая сила \mathcal{E} при разности температур ΔT тем больше, чем больше α . Для того чтобы термопреобразователь давал большую мощность, в нем должен быть создан возможно больший ток, который определяется проводимостью σ и термо-э. д. с. α , в результате чего выделяемая мощность будет пропорциональна $\alpha \sigma \alpha = \sigma \alpha^2$. Но создание необходимой разности температур приводит к возникновению потока тепла, величина которого при заданной разности температур ΔT определяется полной теплопроводностью κ . Чем меньше κ , тем больше должен быть к. п. д. преобразователя. Более точно параметр эффективности должен определяться относительной термо-э. д. с., поэтому для Z вводится выражение

$$Z = \frac{\alpha_2^2}{\left[\left(\frac{\kappa_1}{\sigma_1} \right)^2 + \left(\frac{\kappa_2}{\sigma_2} \right)^2 \right]} = \left[\frac{\alpha_2 - \alpha_1}{\left(\frac{\kappa_1}{\sigma_2} \right)^2 + \left(\frac{\kappa_2}{\sigma_2} \right)^2} \right]^2. \quad (47.68)$$

На рис. 74 приведена температурная зависимость термо-э. д. с. образца теллурида ртути, легированного медью. Медь, будучи

акцептором в соединениях типа $A^{IV}B^{VI}$, обеспечивает дырочную проводимость теллурида ртути в низкотемпературной области, что дает положительное значение термо-э. д. с. При повышении температуры начинают доминировать более подвижные электроны, что приводит к смене знака α .

Резюме § 47

1. К термоэлектрическим явлениям относятся явления Зеебека, Пельтье и Томсона.

2. Эффект Зеебека, или термоэлектрический эффект, состоит в возникновении электрического поля E^α в веществе, в котором создан градиент температуры ∇T :

$$E^\alpha = \alpha \nabla T. \quad (47.1p)$$

Объемное термоэлектрическое поле E^α препятствует разделению зарядов, обусловленного тем, что концентрация носителей заряда в точках с различной температурой различна. Термоэлектрическое поле E^α , вызывая ток проводимости $j_E = \sigma E^\alpha$, уравнивает диффузионный ток $j_D = -eD\nabla n$, обусловленный градиентом концентрации ∇n :

$$j = j_E + j_D = 0 \text{ и } \alpha = \frac{eD}{\sigma} \frac{dn}{dT}, \quad (47.2p)$$

где D — коэффициент диффузии.

Контактная термоэлектродвижущая сила обусловлена зависимостью внутренней контактной разности потенциалов от температуры.

3. В области примесной проводимости абсолютная дифференциальная термо-э. д. с. равна в невырожденном полупроводнике

$$\alpha = \frac{k}{e} \left(\frac{5}{2} + p + \ln \frac{N_c}{n} \right). \quad (47.3p)$$

Знак α совпадает со знаком носителя заряда. В вырожденном полупроводнике

$$\alpha = \frac{\pi^2 k^2 T}{e F}. \quad (47.4p)$$

В полупроводнике, имеющем смешанную проводимость $\sigma = \sigma_n + \sigma_p$:

$$\alpha = \frac{\alpha_n \sigma_n + \alpha_p \sigma_p}{\sigma_n + \sigma_p}. \quad (47.5p)$$

4. Абсолютная термо-э. д. с. α не может быть экспериментально определена, поскольку в замкнутом контуре из одного и того же вещества поле E^α не может вызвать ток, так как вследствие потенциальности поля E^α его циркуляция равна нулю.

5. Термоэлектрический ток может существовать в цепи, состоящей по крайней мере из двух различных веществ. Термо-э. д. с. \mathcal{E}_{12} характеризуется относительной дифференциальной термо-э. д. с.

$$\alpha_{12} = \alpha_2 - \alpha_1 \quad (47.6p)$$

и разностью температур

$$\mathcal{E}_{12} = \int_{T_1}^{T_2} \alpha_{12} dT. \quad (47.7p)$$

6. В неоднородной среде возникает электрическое поле E^i , обусловленное разделением зарядов вследствие различия концентрации носителей заряда:

$$E^i = -\frac{1}{e} \nabla F. \quad (47.8p)$$

При прохождении электрического тока поле E^i совершает работу по перемещению зарядов. Если E^i совпадает по направлению с внешним полем $E^{\text{стоп}}$, вызывающим ток j , то E^i способствует прохождению тока, в результате чего источник тока совершает меньшую работу, необходимую для перемещения зарядов. Работа поля E^i по перемещению зарядов совершается за счет тепловой энергии тела, что приводит к охлаждению тела. Если E^i направлено против $E^{\text{стоп}}$, то это приводит к дополнительному по отношению к ленц-джоулеву выделению тепла и к нагреву тела. Количество выделяемого тепла зависит от прошедшего через объем (или контакт) заряда, а знак тепла меняется при изменении направления тока.

Выделение тепла в результате работы поля E^i (47.8p) носит название эффекта Пельтье. За единицу времени на единицу площади контакта выделяется энергия

$$\dot{Q}_{12} = \dot{Q}_{12}^{\text{П}} = \Pi_{12} j = (\Pi_2 - \Pi_1) j, \quad (47.9p)$$

представляющая собой разность потоков энергии Пельтье. Между Π и α существует связь:

$$\Pi_{12} = \alpha_{12} T = \pi_{12}, \quad (47.10p)$$

$$\pi = \Pi - \frac{F}{e} = \alpha T. \quad (47.11p)$$

7. Выделение или поглощение тепла \dot{Q}^T , дополнительного к ленц-джоулеву теплу, при прохождении электрического тока через вещество, в котором имеется градиент температуры ∇T , называется эффектом Томсона

$$\dot{Q}^T = -\tau (j \nabla T), \quad (47.12p)$$

где \dot{Q}^T представляет собой количество тепла, выделяемого в единицу времени в единице объема, τ — коэффициент Томсона. Эффект Томсона возникает в результате работы термоэлектрического поля E^α по

перемещению заряда при прохождении тока. Между α и τ имеется связь:

$$\tau = T \frac{\partial \alpha}{\partial T}; \quad \alpha(T) = \int_0^T \frac{\tau(\xi)}{\xi} d\xi. \quad (47.13p)$$

Из (47.10p) и (47.13p) следует

$$\frac{d\pi_{12}}{dT} = a_{12} + (\tau_2 - \tau_1) = \frac{\pi_{12}}{T} + (\tau_2 - \tau_1). \quad (47.14p)$$

8. Эффективность работы термоэлектрического преобразователя энергии определяется параметром эффективности Z :

$$Z = \left[\frac{\alpha_2 - \alpha_1}{\left(\frac{\kappa_1}{\sigma_1} \right)^{\frac{1}{2}} + \left(\frac{\kappa_2}{\sigma_2} \right)^{\frac{1}{2}}} \right]^2. \quad (47.15p)$$

§ 48. ТЕРМОМАГНИТНЫЕ ЯВЛЕНИЯ

Явления, которые наблюдаются в полупроводниках, имеющих градиент температуры, в отсутствие электрического тока при наложении на полупроводник магнитного поля называют термомагнитными явлениями. Физической основой их возникновения является взаимодействие носителей заряда с магнитным полем, или, проще говоря, линейная зависимость силы Лоренца от скорости.

К числу термомагнитных явлений относятся:

эффект Риги—Ледюка — возникновение поперечного градиента температуры (тепловой аналог эффекта Холла);

эффект Маджи—Риги—Ледюка — изменение теплопроводности в направлении градиента температуры, или возникновение продольного градиента температуры;

поперечный эффект Нернста—Эттингсгаузена — возникновение поперечного электрического поля;

продольный эффект Нернста—Эттингсгаузена — изменение термоэ. д. с., или возникновение продольного электрического поля.

Рассмотрим кратко механизм возникновения термомагнитных явлений.

Поперечный эффект Нернста—Эттингсгаузена, *термогальваномагнитный эффект*, состоит в появлении поперечного электрического поля E_z^{NE} (или поперечной разности потенциалов V_z^{NE}) при наложении магнитного поля $\mathbf{B} = (0, B, 0)$ на вещество, в котором имеется направленный поток энергии \mathbf{W} , созданный градиентом температуры $\nabla T = (\nabla_x T, 0, 0)$. Величина поля E_z^{NE} зависит от ∇T , \mathbf{B} и свойств вещества, которые характеризуются коэффициентом Нернста—Эттингсгаузена A_{\perp}^{NE} :

$$E_z^{NE} = A_{\perp}^{NE} \nabla_x T \cdot B_y. \quad (48.1)$$

Если размер образца вдоль оси z есть a , то, измерив разность потенциалов V_z^{NE} , можно определить A_{\perp}^{NE} из соотношения

$$A_{\perp}^{NE} = \frac{E_z^{NE}}{\nabla_x T \cdot B} = \frac{V_z^{NE}}{aB\nabla_x T} \quad (48.2)$$

при условии $\mathbf{j} = 0$. Равенство тока нулю обусловлено тем, что диффузионный ток компенсируется дрейфовым током, вызванным движением носителей заряда в термоэлектрическом поле $\alpha\nabla T$. На дрейфовый и диффузионный токи действует магнитное поле. Так как дрейфовый и диффузионный потоки частиц направлены в противоположные стороны, то их отклонение в магнитном поле происходит в противоположных направлениях, z -ая составляющая тока j_z будет представлять собой сумму j_{Dz} -диффузионной составляющей и j_{Ez} -дрейфовой составляющей тока:

$$j_z^{NE} = j_{Dz} + j_{Ez}. \quad (48.3)$$

Однако из условия

$$j_x = j_{Dx} + j_{Ex} = 0 \quad (48.4)$$

не вытекает обращение j_z в нуль.

Но как только появилась составляющая j_z^{NE} , то она приводит к накоплению заряда на боковых гранях образца, что в свою очередь приводит к возникновению поля E_z^{NE} , которое компенсирует ток j_z^{NE} :

$$j_z^{NE} + \sigma_{Bz} E_z^{NE} = 0, \quad (48.5)$$

откуда

$$E_z^{NE} = -\frac{j_z^{NE}}{\sigma_{Bz}} = -\frac{j_{Ez} + j_{Dz}}{\sigma_{Bz}}. \quad (48.6)$$

Если изменить направление \mathbf{B} , то E_z^{NE} меняет знак подобно полю Холла. Величина A_{\perp}^{NE} должна зависеть от знака заряда. Действительно, при заданном ∇T направление потока носителей заряда не зависит от знака их заряда, но направление токов будет противоположным для электронов и дырок. В данном полупроводнике A_{\perp}^{NE} может иметь различный знак в зависимости от того, какая из составляющих — j_{Dz} или j_{Ez} — больше. Как показывает расчет, знак E_z^{NE} зависит от механизма рассеяния носителей заряда, поэтому если при изменении температуры меняется вид рассеяния, то E_z^{NE} может изменить знак.

Поясним механизм возникновения термогальваномагнитного эффекта несколько иначе. Наличие потока энергии $\mathbf{W}_x = -\kappa_e \nabla_x T$ в отсутствие тока ($\mathbf{j} = 0$) означает, что в среднем более быстрые частицы перемещаются против градиента температур, в то время как более медленные (более «холодные») частицы движутся вдоль $\nabla_x T$. Поскольку сила Лоренца пропорциональна скорости, то «горячие»

частицы испытывают большую силу, чем «холодные». Предположим, что время релаксации не зависит от энергии. В этом случае отклонение «горячих» частиц происходит на больший угол, чем «холодных», поэтому противоположно направленные потоки «горячих» и «холодных» частиц не могут скомпенсировать друг друга. Это приводит к накоплению «горячих» частиц у боковой грани полупроводника, что выравнивает потоки «горячих» и «холодных» частиц. Поле E_z^{NE} , обеспечивающее условие $j_z = 0$, и представляет собой поле поперечного эффекта Нернста—Эттингсгаузена. Направление этого поля должно зависеть от знака заряда частиц. Проведенное выше рассуждение справедливо и в том случае, когда время релаксации растет с ростом энергии: $\tau = \tau_0 E^p$ при $p > 0$. Но если $p < 0$, то отклонение «холодных» частиц должно быть больше, чем «горячих», поэтому знак поля E_z^{NE} должен быть противоположным ранее разобранному случаю. Другими словами, величина и знак поля E_z^{NE} зависят от механизма рассеяния.

Продольный эффект Нернста—Эттингсгаузена, или продольный термогальваномагнитный эффект, состоит в возникновении продольного электрического поля, или разности потенциалов, вдоль градиента температур $\nabla_x T$. Но поскольку вдоль $\nabla_x T$ есть термоэлектрическое поле $E_x^\alpha = \alpha \nabla_x T$, то возникновение дополнительного поля вдоль $\nabla_x T$ равносильно изменению термоэлектрического поля E_x^α при наложении магнитного поля. Продольный эффект Нернста—Эттингсгаузена характеризуется коэффициентом A_{\parallel}^{NE} , который определяется соотношением

$$E_x(B) - E_x(0) = [\alpha(B) - \alpha(0)] \nabla_x T = A_{\parallel}^{NE} \alpha(0) B^2 \nabla_x T. \quad (48.7)$$

Из (48.7) можно определить A_{\parallel}^{NE} :

$$A_{\parallel}^{NE} = \frac{\alpha(B) - \alpha(0)}{\alpha(0) B^2} = \frac{1}{B^2} \frac{\Delta \alpha}{\alpha_0}. \quad (48.8)$$

Объяснение продольного эффекта Нернста—Эттингсгаузена очень простое. Поскольку магнитное поле меняет роль медленных и быстрых частиц в переносе энергии, а термоэлектрическое поле обеспечивает отсутствие переноса заряда при переносе энергии, тем самым термоэлектрическое поле должно измениться при наложении магнитного поля. Величина A_{\parallel}^{NE} должна зависеть от механизма рассеяния, но не зависит от направления магнитного поля.

Изменение роли «горячих» и «холодных» носителей заряда приводит к тому, что изменится поток энергии вдоль оси x , т. е. появится дополнительный градиент температуры вдоль оси x , что можно рассматривать как изменение теплопроводности. Изменение теплопроводности κ_e в магнитном поле вдоль оси x — в направлении первичного потока энергии — называют эффектом Маджи—Риги—Ледюка. Он характеризуется величиной λ , определяемой следующим образом:

$$\lambda = \frac{1}{B^2} \frac{\kappa_e(0) - \kappa_e(B)}{\kappa_e(0)}. \quad (48.9)$$

Эффект Риги—Ледюка, или термомагнитный эффект (в узком смысле слова), состоит в появлении поперечного градиента температуры $\nabla_z T$ при наложении магнитного поля на полупроводник, в котором первоначально создан градиент температуры $\nabla_x T$. Эффект Риги—Ледюка определяется коэффициентом Риги—Ледюка A^{RL} :

$$\nabla_z T = A^{RL} \nabla_x T \cdot B_y, \quad (48.10)$$

или

$$A^{RL} = \frac{\nabla_z T}{B \nabla_x T}. \quad (48.11)$$

Рассмотрим механизм возникновения эффекта Риги—Ледюка. Для этого достаточно вернуться к объяснению поперечного эффекта Нернста—Эттингсгаузена. При его объяснении мы нашли, что более «горячие» и более «холодные» носители заряда отклоняются в противоположные стороны. Но это означает, что грань, к которой отклоняются более «горячие» электроны, должна нагреваться, в то время как противоположная грань должна охлаждаться; другими словами, возникает поперечная разность температур, или поперечный градиент температуры $\nabla_z T$, который вызовет поток энергии W_z , стремящийся скомпенсировать различие в потоках энергии, переносимой «горячими» и «холодными» частицами к боковым граням. При полной компенсации потоков энергии вдоль оси z , вызванных полной теплопроводностью $\kappa = \kappa_e + \kappa_L$ и магнитным полем, поток энергии вдоль оси z обращается в нуль: $W_z = 0$, поэтому $W = (W_x, 0, 0)$, в то время как $\nabla T = (\nabla_x T, 0, \nabla_z T)$, т. е. между ∇T и W возникает угол φ_{RL} ; согласно (48.11)

$$\operatorname{tg} \varphi_{RL} = \frac{\nabla_z T}{\nabla_x T} = A^{RL} B. \quad (48.12)$$

Из (48.12) следует, что A^{RL} по размерности совпадает с подвижностью, он должен зависеть от знака заряда частиц и механизма рассеяния.

Вычисление коэффициентов A_{\perp}^{NE} , A_{\parallel}^{NE} , λ и A^{RL} возможно на основе кинетического уравнения, в котором учитывается распределение частиц по состояниям.

Резюме § 48

1. В веществе, в котором имеется поток энергии W , вызванный градиентом температур ∇T , в отсутствие тока ($j = 0$) под действием магнитного поля $B = (0, B, 0)$ возникает ряд явлений, называемых термомагнитными. К ним относятся:

а) эффект Риги—Ледюка, или просто термомагнитный эффект, — возникновение поперечного градиента температур:

$$\nabla_z T = A^{RL} B \nabla_x T; \quad (48.1p)$$

б) поперечный эффект Нернста — Эттингсгаузена, или поперечный термогальваномагнитный эффект, — возникновение поперечного электрического поля E_z^{NE} :

$$E_z^{NE} = A_{\perp}^{NE} B \nabla_x T; \quad (48.2p)$$

в) продольный эффект Нернста — Эттингсгаузена, или продольный термогальваномагнитный эффект, — возникновение электрического поля вдоль $\nabla_x T$, или изменение термо-э. д. с. $\Delta\alpha = \alpha(B) - \alpha_0$:

$$A_{\parallel}^{NE} = \frac{\Delta\alpha}{B^2\alpha_0}, \quad (48.3p)$$

г) эффект Риги — Ледюка — Маджи — возникновение продольного градиента температур, или изменение теплопроводности:

$$\lambda = \frac{1}{B^2} \cdot \frac{\kappa_0 - \kappa(B)}{\kappa_0}. \quad (48.4p)$$

Два первых эффекта являются нечетными, два последних — четными по магнитному полю.

§ 49. ОБЩИЙ АНАЛИЗ КИНЕТИЧЕСКИХ ЯВЛЕНИЙ

Описание любого кинетического явления заложено в кинетическом уравнении и вытекающих из него выражениях (38.1p) и (38.2p) для \mathbf{j} и \mathbf{W} , которые должны рассматриваться совместно, поскольку \mathbf{j} и \mathbf{W} определяются через одни и те же величины, поэтому, меняя условия, накладываемые, например, на \mathbf{W} , мы автоматически меняем величины, входящие в \mathbf{j} , и наоборот. Проиллюстрируем это положение в дальнейшем конкретными примерами. Не проводя вычислений для описания всех термомагнитных явлений и двух оставшихся неразобранными гальваномагнитных явлений, ограничимся общим анализом уравнений для \mathbf{j} и \mathbf{W} и покажем, каким образом из них можно получить описание любого конкретного эффекта. С этой целью преобразуем выражения для плотности тока и плотности потока энергии к другому виду.

Представим выражения для плотности тока \mathbf{j} и плотности потока энергии \mathbf{W} в случае поперечного магнитного поля¹ в виде суммы

¹ Как было показано в § 37, продольные эффекты в веществах со скалярной эффективной массой отсутствуют. Разобранные выше продольные эффекты Нернста — Эттингсгаузена, Маджи — Риги — Ледюка, Нернста, магнетосопротивления принадлежат к поперечным эффектам, поскольку они наблюдаются при $(\nabla T \mathbf{V}) = (\mathbf{J} \mathbf{V}) = 0$. Поэтому названия «поперечный» или «продольный» эффекты Нернста — Эттингсгаузена с этой точки зрения неудачны. Название «продольный» или «поперечный» означает в данном случае не направление \mathbf{V} , а направление дополнительного электрического поля или градиента температуры по отношению к исходным полям или градиентам температуры, вызывающим ток или поток энергии.

двух компонентов — «продольных» \mathbf{j}_1 , \mathbf{W}_1 и «поперечных» \mathbf{j}_2 , \mathbf{W}_2 соответственно:

$$\mathbf{j} = \mathbf{j}_1 + \mathbf{j}_2; \quad \mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2, \quad (49.1)$$

где

$$\mathbf{j}_1 = eK'_{11} \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) - eK'_{21} \frac{\nabla T}{T}; \quad (49.2)$$

$$\mathbf{j}_2 = \left[\frac{e^2}{m^*} K'_{12} \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) - \frac{e^2}{m^*} K'_{22} \frac{\nabla T}{T}, \mathbf{B} \right]; \quad (49.3)$$

$$\mathbf{W}_1 = K'_{21} \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) - K'_{31} \frac{\nabla T}{T}; \quad (49.4)$$

$$\mathbf{W}_2 = \left[\frac{e}{m^*} K'_{22} \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) - \frac{e}{m^*} K'_{32} \frac{\nabla T}{T}, \mathbf{B} \right]. \quad (49.5)$$

Рассмотрим компонент \mathbf{j}_1 :

$$\begin{aligned} \mathbf{j}_1 &= eK'_{11} (e\mathbf{E} - \nabla F) + \frac{eK'_{11}F}{T} \nabla T - \frac{eK'_{21}}{T} \nabla T = \\ &= e^2 K'_{11} \left(\mathbf{E} - \nabla \frac{F}{e} \right) - e^2 K'_{11} \left(\frac{K'_{21} - FK'_{11}}{eK'_{11}T} \right) \nabla T. \end{aligned} \quad (49.6)$$

Обозначим для краткости

$$e^2 K'_{11} = \sigma_B. \quad (49.7)$$

Учитывая, что

$$\frac{K'_{21} - FK'_{11}}{eK'_{11}T} = \alpha(\mathbf{B}) = \alpha', \quad (49.8)$$

компонент \mathbf{j}_1 можно представить в виде

$$\mathbf{j}_1 = \sigma_B \left\{ \left(\mathbf{E} - \nabla \frac{F}{e} \right) - \alpha' \nabla T \right\}. \quad (49.9)$$

Первый сомножитель \mathbf{j}_2 преобразуем таким образом:

$$\begin{aligned} &\frac{e^2}{m^*} K'_{12} \left(e\mathbf{E} - T\nabla \frac{F}{T} \right) - \frac{e^2}{m^*} K'_{22} \frac{\nabla T}{T} = \\ &= e^2 K'_{11} \frac{eK'_{12}}{m^* K'_{11}} \left(\mathbf{E} - \nabla \frac{F}{e} \right) + \frac{e^2}{m^*} \frac{K'_{12}F - K'_{22}}{T} \nabla T. \end{aligned} \quad (49.10)$$

Учитывая, что $\frac{e}{m^*} \frac{K'_{12}}{K'_{11}} = \mu^H$ — холлова подвижность, и обозначая по аналогии с (49.8)

$$\frac{K'_{22} - FK'_{12}}{eTK'_{12}} = \beta(\mathbf{B}) = \beta', \quad (49.11)$$

можем записать

$$\frac{e^2}{m^*} \frac{K'_{12}F - K'_{22}}{T} = e^2 K'_{11} \frac{e}{m^*} \frac{K'_{12}}{K'_{11}} \cdot \frac{K'_{12}F - K'_{22}}{eTK'_{12}} = -\sigma_B \mu^H \beta' \quad (49.12)$$

и

$$\mathbf{j}_2 = \sigma_B \mu^H \left[\left(\mathbf{E} - \nabla \frac{F}{e} \right) - \beta' \nabla T, \mathbf{B} \right]. \quad (49.13)$$

Преобразуем аналогично выражение для потока энергии. Первый член W_1 запишем таким образом:

$$\begin{aligned} W_1 &= K'_{21} (eE - \nabla F) + \frac{FK'_{21}}{T} \nabla T - K'_{31} \frac{\nabla T}{T} = \\ &= e^2 K'_{11} \frac{K'_{21}}{eK'_{11}} \left(\mathbf{E} - \nabla \frac{F}{e} \right) - \frac{K'_{31} - FK'_{21}}{T} = \\ &= \sigma_B \left(\mathbf{E} - \nabla \frac{F}{e} \right) \Pi - \frac{K'_{31} - FK'_{21}}{T} \nabla T. \end{aligned} \quad (49.14)$$

Но

$$\begin{aligned} K'_{31} - FK'_{21} &= K'_{31} - \frac{K'^2_{21}}{K'_{11}} + \frac{K'^2_{21}}{K'_{11}} - FK'_{21} = \\ &= \frac{K'_{31}K'_{11} - K'^2_{21}}{K'_{11}} + K'_{21} \left(\frac{K'_{31} - FK'_{11}}{K'_{11}} \right) = \kappa'_e T + K'_{21} e T \alpha' = \\ &= \kappa'_e T + \Pi e^2 K'_{11} T \alpha' = \kappa'_e T + \sigma_B \Pi \alpha' T, \end{aligned} \quad (49.15)$$

поэтому

$$W_1 = \sigma_B \left(\mathbf{E} - \nabla \frac{F}{e} - \alpha \nabla T \right) \Pi - \kappa'_e \nabla T. \quad (49.16)$$

Рассмотрим первый сомножитель в W_2 :

$$\begin{aligned} \frac{e}{m^*} K'_{22} \left(e\mathbf{E} - T \nabla \frac{F}{T} \right) - \frac{e}{m^*} K'_{32} \frac{\nabla T}{T} &= \frac{e^2}{m^*} K'_{22} \left(\mathbf{E} - \nabla \frac{F}{e} \right) + \\ + \frac{e}{m^*} \left(\frac{FK'_{22} - K'_{32}}{T} \right) \nabla T &= e^2 K'_{11} \frac{K'_{21}}{eK'_{11}} \cdot \frac{eK'_{22}}{m^* K'_{21}} \left(\mathbf{E} - \nabla \frac{F}{e} \right) + \frac{e}{m^*} \left(\frac{FK'_{22} - K'_{32}}{T} \right) \nabla T = \\ &= \sigma_B \Pi \mu^E \left(\mathbf{E} - \nabla \frac{F}{e} \right) + \frac{e}{m^*} \frac{FK'_{22} - K'_{32}}{T} \nabla T. \end{aligned} \quad (49.17)$$

Через μ^E мы обозначили величину

$$\mu^E = \frac{e}{m^*} \frac{K'_{22}}{K'_{21}} = \frac{e}{m^*} \frac{\left\langle \frac{\tau^2}{1 + \mu^2 B^2} \right\rangle}{\left\langle \frac{\tau}{1 + \mu^2 B^2} \right\rangle}. \quad (49.18)$$

Преобразуем второе слагаемое в выражении (49.17):

$$\begin{aligned} K'_{32} - FK'_{22} &= K'_{32} - \frac{K'^2_{22}}{K'_{12}} + \frac{K'^2_{22}}{K'_{12}} - K'_{22} F = \\ &= \frac{K'_{32}K'_{12} - K'^2_{22}}{K'_{12}} + \frac{K'_{22}(K'_{22} - FK'_{12})}{K'_{12}}. \end{aligned} \quad (49.19)$$

Второй член (49.19) можно записать так:

$$\begin{aligned} \frac{K'_{22}(K'_{22} - FK'_{12})}{K'_{12}} &= K'_{22} e T \beta' = \\ &= \frac{e}{m^*} \frac{K'_{22}}{K'_{21}} \frac{K'_{21}}{eK'_{11}} e m^* K'_{11} T \beta' = \sigma_B \mu^E \Pi T \beta' \frac{m^*}{e}. \end{aligned} \quad (49.20)$$

Первое слагаемое в последнем члене (49.19) с учетом $\frac{e}{m^*T}$ запишем в форме

$$\frac{e}{m^*T} \frac{K'_{32}K'_{12} - K'_{22}{}^2}{K'_{12}} = \frac{eK'_{32}}{m^*K'_{21}} \frac{K'_{21}K'_{32}K'_{12} - K'_{22}{}^2}{TK'_{12}} = \mu^E \lambda'. \quad (49.21)$$

Учитывая (49.17, 18, 20, 21), запишем

$$W_2 = \left[\sigma_B \mu^E \left(\mathbf{E} - \nabla \frac{F}{e} - \beta' \nabla T \right) \Pi - \mu^E \lambda' \nabla T, \mathbf{B} \right]. \quad (49.22)$$

Обозначив

$$\mathbf{E} - \nabla \frac{F}{e} = \mathbf{E}^*, \quad (49.23)$$

перепишем выражения для тока и потока энергии:

$$\mathbf{j} = \sigma_B (\mathbf{E}^* - \alpha' \nabla T) + \sigma_B \mu^H [\mathbf{E}^* - \beta' \nabla T, \mathbf{B}] \quad (49.24)$$

и

$$\mathbf{W} = \sigma_B (\mathbf{E}^* - \alpha' \nabla T) \Pi - \kappa'_e \nabla T + \mu^E [\sigma_B (\mathbf{E}^* - \beta' \nabla T) \Pi - \lambda' \nabla T, \mathbf{B}]. \quad (49.25)$$

Уравнения (49.24) и (49.25) описывают все явления, связанные с движением носителей заряда, а именно:

*электропроводность и теплопроводность,
термоэлектрические явления,
гальваномагнитные явления,
термомагнитные явления.*

Все эти эффекты имеют место одновременно, накладываясь один на другой. Для выделения каждого эффекта в отдельности — в «чистом виде» — необходимо создать определенные экспериментальные условия, в противном случае одновременно могут проявиться несколько эффектов. Ниже мы укажем условия, при которых можно наблюдать то или иное явление. Уравнения (49.24) и (49.25) содержат описание как изотермических, так и адиабатических эффектов. Укажем смысл отдельных членов в уравнениях (49.24) и (49.25).

Член $\sigma_B (\mathbf{E}^* - \alpha' \nabla T)$ описывает ток проводимости в полном электрическом поле, обусловленном градиентами электростатического потенциала и энергии Ферми и термо-э. д. с.; $\sigma_B \mu^H [\mathbf{E}^* \mathbf{B}]$ выражает ток Холла, а член

$$- \sigma_B \mu^H \beta' [\nabla T, \mathbf{B}] \quad (49.26)$$

дает выражения для тока, который естественно назвать током Нернста — Эттингсгаузена, так как именно этот ток приводит к появлению поперечного электрического поля Нернста — Эттингсгаузена. Поскольку величины проводимости и термо-э. д. с. зависят от \mathbf{B} , в уравнениях (49.24) и (49.25) заложено описание магнетосопротивления и продольного эффекта Нернста — Эттингсгаузена.

Выражение для потока энергии состоит из членов двух типов: один тип членов содержит в качестве множителя коэффициент

Пельтье; очевидно, эти члены описывают явления, связанные с переносом энергии при прохождении тока, и приводят к гальваномагнитным эффектам; члены, содержащие ∇T и \mathbf{B} , обуславливают термомагнитные явления.

Покажем, как использовать уравнения (49.24) и (49.25) для описания отдельных эффектов.

1. *Изотермические эффекты*: $\nabla T = 0$.

1. Магнитного поля нет: $\mathbf{B} = 0$. Уравнения (49.24) и (49.25) имеют вид

$$\mathbf{j} = \sigma_0 \mathbf{E}^*; \quad \mathbf{W} = \sigma_0 \mathbf{E}^* \Pi = \Pi \mathbf{j}, \quad (49.27)$$

они описывают ток проводимости и поток Пельтье под действием электрического поля $\mathbf{E}^* = -\nabla \left(\varphi + \frac{F}{e} \right)$. В однородном веществе $\nabla F = 0$ ток вызывается градиентом электростатического потенциала, в неоднородном — градиентом электрохимического потенциала. Если $\mathbf{j} = 0$, то в веществе существует градиент электростатического потенциала

$$-\nabla \varphi = \nabla \frac{F}{e}; \quad \mathbf{E} = \frac{1}{e} \nabla F. \quad (49.28)$$

Уравнение (49.28) приводит нас к новому явлению: *в неоднородном полупроводнике существует объемное электрическое поле*.

2. Полупроводник в магнитном поле. Уравнения для \mathbf{j} и \mathbf{W} запишутся в виде

$$\mathbf{j} = \sigma_B \mathbf{E}^* + \sigma_B \mu^H [\mathbf{E}^* \mathbf{B}], \quad (49.29)$$

$$\mathbf{W} = \sigma_B \mathbf{E}^* \Pi + \sigma_B \mu^E [\mathbf{E}^* \mathbf{B}] \Pi. \quad (49.30)$$

Первый член в \mathbf{j} описывает омический ток, второй — холлов ток, с ними связаны потоки тепла Пельтье. Накладывая определенные граничные условия, можно прийти к изотермическому эффекту Холла. Для этого достаточно положить

$$\mathbf{j} = (j_x, 0, 0); \quad \mathbf{B} = (0, B, 0); \quad \mathbf{E}^* = (E_x^*, 0, E_z^*). \quad (49.31)$$

Выражение для потока \mathbf{W} не может быть выбрано произвольно, например, нельзя положить $\mathbf{W} = (W_x, 0, 0)$, потому что уравнения для \mathbf{j} и \mathbf{W} будут несовместимыми.

Положим $\mathbf{W} = (W_x, 0, W_z)$ и запишем

$$W_z = \sigma_B E_z^* \Pi + \sigma_B \mu^E E_x^* B \Pi, \quad (49.32)$$

в то время как

$$j_z = \sigma_B E_z^* + \sigma_B \mu^H E_x^* B. \quad (49.33)$$

Действительно, из $j_z = 0$ не вытекает $W_z = 0$. Физически это легко объяснить. Обращение в нуль j_z вызвано тем, что на боковых гранях накапливается заряд, который приводит к появлению поля Холла, оно в свою очередь компенсирует действие магнитного поля. Иначе обстоит с потоком тепла. Поток тепла W_z не компенсируется никакими источниками тепла, поэтому между образцом и окружаю-

щей средой существует идеальный тепловой контакт, обеспечивающий непрерывность потока W_z , что в свою очередь приводит к выполнению условия $\nabla T = 0$.

Если $W_z = 0$, то поток тепла в направлении оси z , вызванный магнитным полем, может быть скомпенсирован теплопроводностью, следовательно, должен возникнуть градиент температуры в направлении оси z . Мы видим, что эффект Эттингсгаузена возможен, если не происходит обмена энергией с окружающей средой, т. е. эффект Эттингсгаузена является адиабатическим. Прежде чем перейти к нему, рассмотрим другие изотермические эффекты. Уравнения (49.29) и (49.30) содержат в себе описание изотермического магнетосопротивления как для ограниченного образца (с учетом 49.31), так и для неограниченного ($E_z = 0$; $j_z \neq 0$).

II. Неизотермические эффекты: $\nabla T \neq 0$.

1. Магнитного поля нет: $\mathbf{B} = 0$. Уравнения (49.24) и (49.25) имеют вид

$$\mathbf{j} = \sigma_0 (\mathbf{E}^* - \alpha \nabla T), \quad (49.34)$$

$$\mathbf{W} = \sigma_0 (\mathbf{E}^* - \alpha \nabla T) \Pi - \kappa_e \nabla T = \Pi \mathbf{j} - \kappa_e \nabla T. \quad (49.35)$$

Если $\mathbf{j} = 0$, то это обуславливает поток тепла в результате теплопроводности:

$$\mathbf{W} = -\kappa_e \nabla T. \quad (49.36)$$

В веществе возникает термоэлектрическое поле

$$\mathbf{E}^* = \alpha \nabla T = \mathbf{E}^\alpha. \quad (49.37)$$

Если $\mathbf{j} \neq 0$, то осуществляются условия для всех термоэлектрических явлений (Зеебека, Пельтье и Томсона), которые мы разобрали ранее.

2. Полупроводник в магнитном поле: $\mathbf{B} \neq 0$.

Ток \mathbf{j} и поток энергии \mathbf{W} определяются общими выражениями (49.24) и (49.25).

Если до наложения магнитного поля ток шел в направлении оси x и при этом $\nabla T = 0$, то магнитное поле может изменить условие $\nabla T = 0$ двояким образом:

$$\nabla T = (0; 0; \nabla_z T), \quad (49.38)$$

$$\nabla T = (\nabla_x T; 0; 0). \quad (49.39)$$

Первый случай (49.38) есть эффект Эттингсгаузена, второй — (49.39) — эффект Нернста. Проиллюстрируем метод применения уравнений (49.34) и (49.35) на примере эффекта Эттингсгаузена.

Эффект Эттингсгаузена. Для его наблюдения необходимо положить

$$\mathbf{j} = (j_x; 0; 0); \quad \mathbf{W} = (W_x; 0; 0). \quad (49.40)$$

Как уже упоминали, для одновременного выполнения условий $j_z = W_z = 0$ необходимо, чтобы $\nabla_z T \neq 0$, поэтому будем считать

$$\nabla T = (0; 0; \nabla_z T). \quad (49.41)$$

Запишем выражения (49.24) и (49.25) для j и W с учетом (49.40), (49.41); $V = (0; B; 0)$:

$$j_x = \sigma_B E_x^* - \sigma_B \mu^H (E_z^* - \beta' \nabla_z T) B; \quad (49.42)$$

$$j_z = \sigma_B (E_z^* - \alpha' \nabla_z T) + \sigma_B \mu^H E_x^* B = 0; \quad (49.43)$$

$$W_x = \sigma_B E_x^* \Pi - \mu^E (\sigma_B E_z^* \Pi - \sigma_B \beta' \nabla_z T \Pi - \lambda' \nabla_z T) B; \quad (49.44)$$

$$W_z = \sigma_B (E_z^* - \alpha' \nabla_z T) \Pi + \mu^E \sigma_B E_x^* \Pi B - \kappa'_e \nabla_z T = 0. \quad (49.45)$$

Найдем E_z^* и $\nabla_z T$ из условий $j_z = W_z = 0$. Из (49.43) запишем

$$\sigma_B (E_z^* - \alpha' \nabla_z T) = -\sigma_B \mu^H E_x^* B. \quad (49.46)$$

Подставим (49.46) в (49.45), получим

$$W_z = -\sigma_B \mu^H E_x^* \Pi B + \mu^E \sigma_B E_x^* \Pi B - \kappa'_e \nabla_z T = 0. \quad (49.47)$$

Из (49.47) найдем

$$\nabla_z T = \frac{(\mu^E - \mu^H)}{\kappa'_e} \sigma_B E_x^* B \Pi. \quad (49.48)$$

Подставив (49.48) в (49.46), найдем

$$E_z^* = \alpha' \nabla_z T - \mu^H E_x^* B = \left(\frac{\alpha' \sigma_B \Pi (\mu^E - \mu^H)}{\kappa'_e} - \mu^H \right) E_x^* B. \quad (49.49)$$

Выражение (49.49) показывает, что в адиабатическом случае ($W_z = 0$) к изотермическому полю Холла ($-\mu^H E_x^* B$) добавляется термоэлектрическое поле $\alpha' \nabla_z T$, поэтому адиабатический коэффициент Холла R_a будет отличаться от изотермического R_I . Чтобы ввести коэффициенты Эттингсгаузена A^E и Холла R_a , надо выразить $\nabla_z T$ и E_z^* через j_x и B . Для этого подставим найденные значения E_z^* и $\nabla_z T$ в (49.42):

$$j_x = \sigma_B E_x^* - \sigma_B \mu^H \left\{ \frac{\sigma_B \alpha' (\mu^E - \mu^H) \Pi}{\kappa'_e} - \mu^H - \frac{\beta' \sigma_B (\mu^E - \mu^H) \Pi}{\kappa'_e} \right\} B E_x^* = \\ = \left\{ \sigma_B - \sigma_B \mu^H B \left[\frac{\sigma_B \Pi (\mu^E - \mu^H) (\alpha' - \beta')}{\kappa'_e} - \mu^H \right] \right\} E_x^* = \sigma_{Bx} E_x^*. \quad (49.50)$$

Найдем коэффициент A^E и R_a из (49.48) и (49.49), учитывая (49.50):

$$A^E = \frac{\nabla_z T}{j_x B} = \frac{\sigma_B \Pi (\mu^E - \mu^H) E_x^* B}{\kappa'_e B j_x} = \frac{\sigma_B \Pi (\mu^E - \mu^H)}{\kappa'_e \sigma_{Bx}} \quad (49.51)$$

$$R_a = -\frac{E_z^*}{j_x B} = \frac{\left[\mu^H - \frac{\alpha' \sigma_B \Pi (\mu^E - \mu^H)}{\kappa'_e} \right] E_x^* B}{B j_x} = \left[\frac{\mu^H}{\sigma_{Bx}} - \frac{\alpha' \sigma_B \Pi (\mu^E - \mu^H)}{\kappa'_e \sigma_{Bx}} \right] = \\ = \frac{\mu^H}{\sigma_{Bx}} - \alpha' A^E = R_I - \alpha' A^E. \quad (49.52)$$

Величина σ_{Bx} , определенная из (49.51), описывает *адиабатическое магнетосопротивление*. Найдем A^E и R_a в области слабых полей (при $B \rightarrow 0$), в этом случае $\sigma_B \approx \sigma_0$; $\sigma_{Bx} \approx \sigma_0$ и

$$A^E = \frac{\Pi (\mu^E - \mu^H)}{\kappa_e}, \quad (49.53)$$

$$R_a = R_I - \alpha A^E. \quad (49.54)$$

Различие в R_a и R_I имеет наглядный физический смысл. Поле E_z^* в адиабатических условиях меньше E_z^* в изотермических условиях на величину термоэлектрического поля $\alpha \nabla_z T$. Величина же $\nabla_z T$, соответствующая единичному току и единичной индукции, как раз равна A^E .

Вычислим величину A^E . Прежде всего заметим, что поток W_z будет определяться не только электронной теплопроводностью, но и фононной, поэтому вместо κ_e необходимо подставить полную теплопроводность $\kappa = \kappa_e + \kappa_L$:

$$\begin{aligned} A^E &= \frac{\Pi (\mu^E - \mu^H)}{\kappa} = \frac{1}{\kappa} \frac{K_{21}}{eK_{11}} \left(\frac{e}{m^*} \frac{K_{22}}{K_{21}} - \frac{e}{m^*} \frac{K_{12}}{K_{11}} \right) = \\ &= \frac{e}{m^*} \frac{1}{\kappa e} \frac{\langle \tau \rangle}{\langle \tau \rangle} \left(\frac{\langle \tau^2 \rangle}{\langle \tau \rangle} - \frac{\langle \tau^2 \rangle}{\langle \tau \rangle} \right) = \frac{\mu_d}{e\kappa} \frac{\langle \tau \rangle}{\langle \tau \rangle} \left(\frac{\langle \tau^2 \rangle}{\langle \tau \rangle \langle \tau \rangle} - \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} \right). \quad (49.55) \end{aligned}$$

Выражение (49.55) показывает, что эффект Эттингсгаузена не зависит от знака носителей заряда, но зависит от механизма рассеяния. Действительно, пусть $\tau = \tau_0 E^p$, в этом случае

$$\begin{aligned} \langle \tau \rangle &= \tau_0' \frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}; \quad \langle \tau^2 \rangle = \tau_0'^2 \frac{\Gamma\left(\frac{5}{2} + 2p\right)}{\Gamma\left(\frac{5}{2}\right)}; \\ \langle \langle \tau \rangle \rangle &= \frac{5}{2} (kT) \tau_0' \frac{\Gamma\left(\frac{7}{2} + p\right)}{\Gamma\left(\frac{7}{2}\right)}; \quad \langle \langle \tau^2 \rangle \rangle = \frac{5}{2} (kT) \tau_0'^2 \frac{\Gamma\left(\frac{7}{2} + 2p\right)}{\Gamma\left(\frac{7}{2}\right)}. \quad (49.56) \end{aligned}$$

Подставим значения усредненных времен релаксации (49.56) в (49.55), получим

$$\begin{aligned} A^E &= \frac{\mu_d}{e\kappa} \frac{5}{2} (kT) \frac{\Gamma\left(\frac{7}{2} + p\right) \Gamma\left(\frac{5}{2}\right)}{\Gamma\left(\frac{7}{2}\right) \Gamma\left(\frac{5}{2} + p\right)} \left[\frac{5}{2} (kT) \frac{\Gamma\left(\frac{7}{2} + 2p\right) \Gamma\left(\frac{7}{2}\right) \Gamma\left(\frac{5}{2}\right)}{\Gamma\left(\frac{7}{2}\right) \frac{5}{2} (kT) \Gamma\left(\frac{7}{2} + p\right) \Gamma\left(\frac{5}{2} + p\right)} - \right. \\ &\quad \left. - \frac{\Gamma\left(\frac{5}{2} + 2p\right) \left[\Gamma\left(\frac{5}{2}\right) \right]^2}{\Gamma\left(\frac{5}{2}\right) \left[\Gamma\left(\frac{5}{2} + p\right) \right]^2} \right] = p \frac{kT \mu_d}{\kappa e} \frac{\Gamma\left(\frac{5}{2} + 2p\right) \Gamma\left(\frac{5}{2}\right)}{\left[\Gamma\left(\frac{5}{2} + p\right) \right]^2} = \\ &= p \frac{kT}{e\kappa} \mu^H = p \frac{kT}{\kappa} \frac{R_I \sigma_0}{e}. \quad (49.57) \end{aligned}$$

При $p > 0$ коэффициент Эттингсгаузена $A^E > 0$, т. е. $\nabla_z T > 0$ при $B > 0$ и $j_x > 0$ — верхняя грань нагревается, нижняя охлаждается. Это означает, что сильнее отклоняются вверх «горячие» носители заряда. При $p = 0$, когда время релаксации не зависит от энергии, носители заряда отклоняются в одинаковой мере, селекции в отклонении не происходит и градиент температуры не возникает. При $p < 0$ величины $A^E < 0$ горячие носители заряда отклоняются вниз (при выбранных нами направлениях полей и тока). Если при изменении температуры или состава и структуры образца происходит изменение знака эффекта Эттингсгаузена, то это означает изменение роли различных механизмов рассеяния, переход от одного вида к другому.

Оценим величину A^E . Пусть $\kappa = 1$ Вт·см/(см²·град) = $= 10^2$ Вт(м·град)⁻¹; $kT = 0,03$ эВ; $\mu^H = 10\,000$ см²(В·с)⁻¹, тогда A^E равно $(-)$ 150 см³·град/Дж $= (-)$ $1,5 \cdot 10^{-4}$ м³·град/Дж и $4,5 \cdot 10^{-4}$ м³·град/Дж при $p = (-)$ $1/2$ и $3/2$ соответственно. При плотности тока $j_x = 1$ А/см² = 10^4 А/м² и $B = 1$ Т = 10^4 Гс $\nabla_z T = A^E j_x B$ имеет величину $(-)$ $1,5$ град/м и $4,5$ град/м = $0,05$ град/см. Как видим, градиент температуры сравнительно мал. Оценим αA^E : при $\alpha = 1$ мВ/град = 10^{-3} В/град; $p = 3/2$ величина $\alpha A^E = 5 \cdot 10^{-7}$ м³·град × В/(Дж·град) = $5 \cdot 10^{-7}$ м³/Кл = $0,5$ см³/Кл. Для слабо легированных веществ этой поправкой можно пренебречь, но для *сильно легированных веществ необходимо учитывать различие между R_α и R_I* . Если предположить, что $\nabla_x T \neq 0$, то можно получить выражение для эффекта Нернста. При этом возможно два эффекта — изотермический, если $\nabla_z T \neq 0$, и адиабатический, если $W_z \neq 0$. Нахождение A^N проводится аналогично вычислению A^E .

III. Термомагнитные явления.

Для описания термомагнитных явлений необходимо положить

$$j = 0; \quad \nabla T \neq 0, \quad (49.58)$$

следовательно,

$$j_x = \sigma_B (E_z^* - \alpha' \nabla_x T) - \sigma_B \mu^H (E_z^* - \beta' \nabla_z T) B = 0, \quad (49.59)$$

$$j_z = \sigma_B (E_x^* - \alpha' \nabla_z T) + \sigma_B \mu^H (E_x^* - \beta' \nabla_x T) B = 0, \quad (49.60)$$

$$W_x = \sigma_B (E_x^* - \alpha' \nabla_x T) \Pi - \kappa_e' \nabla_x T - \mu^E \sigma_B (E_z^* - \beta' \nabla_z T) B \Pi + \mu^E \lambda' \nabla_z T \cdot B, \quad (49.61)$$

$$W_z = \sigma_B (E_z^* - \alpha' \nabla_z T) \Pi - \kappa_e' \nabla_z T + \mu^E \sigma_B (E_x^* - \beta' \nabla_x T) B \Pi - \mu^E \lambda' \nabla_x T \cdot B. \quad (49.62)$$

Если найти из уравнений (49.59—62) величины E_x^* , E_z^* , $\nabla_z T$ и выразить их через B и $\nabla_x T$, то получим выражения для коэффициентов продольного и поперечного эффектов Нернста — Эттингсгаузена и Риги — Ледюка. Очевидно, что выражения для E_x^* и E_z^* можно искать при двух различных предположениях: или $\nabla_z T = 0$ — *изотермические эффекты*, или $W_z = 0$ — *адиабатические эффекты*. Эффект Риги — Ледюка может быть только адиабатическим. Найдя E_x^* , E_z^* и $\nabla_z T$ и подставив их в выражение для W_x , сведем его к виду

$W_x = -\kappa'_e \nabla_x T$, откуда получим выражение эффекта Маджи — Риги — Ледюка (при этом $\kappa'_e \nabla_x T$ необходимо заменить на $(\kappa'_e + \kappa_L) \nabla_x T$).

Учитывая зависимость кинетических коэффициентов от магнитного поля, можно найти коэффициенты гальвано- и термомагнитных явлений как в области слабых, так и сильных полей, как в области примесной, так и смешанной проводимости.

Резюме § 49

1. Выражения (49.24) и (49.25) для j и W содержат описания всех кинетических явлений в полупроводниках со скалярной эффективной массой. В табл. 14 указаны условия наблюдения отдельных явлений.

Таблица 14

Условия наблюдения поперечных гальваномагнитных и термомагнитных эффектов

Гальваномагнитные эффекты	Определение коэффициента	Термомагнитные эффекты	Определение коэффициента
Эффект Холла — поперечное поле E_z ($j_z = \nabla_x T = 0$)	$R = -\frac{E_z}{j_x B_y}$	Эффект Риги — Ледюка — поперечный градиент температур $\nabla_x T$ ($j = 0$)	$A^{RL} = \frac{\nabla_x T}{B_y \nabla_x T}$
Магнетосопротивление — изменение сопротивления, продольная разность потенциалов ΔV_x ($j_z = \nabla_x T = 0$)	$H = \frac{1}{B^2} \frac{\rho(B) - \rho(0)}{\rho(0)}$	Эффект Маджи — Риги — Ледюка — изменение теплопроводности, продольная разность температур $\Delta(\nabla_x T)$ ($j = 0$)	$\lambda = \frac{1}{B^2} \frac{\kappa(B) - \kappa(0)}{\kappa(0)}$
Эффект Эттингсгаузена — поперечный градиент температур $\nabla_x T$ ($j_z = \nabla_x T = 0$)	$A^E = \frac{\nabla_x T}{j_x B_y}$	Поперечный эффект Нернста — Эттингсгаузена — поперечное электрическое поле E_z ($j = 0$)	$A_{\perp}^{NE} = \frac{E_z}{B_y \nabla_x T}$
Эффект Нернста — продольная разность температур, продольный градиент температуры $\nabla_x T$ ($j_x = 0$)	$A^N = \frac{\nabla_x T}{j_x B_y^2}$	Продольный эффект Нернста — Эттингсгаузена — появление продольного поля и разности потенциалов, изменение термо-э. д. с. $\Delta(E_x)$; ($j = 0$)	$A_{\parallel}^{NE} = \frac{1}{B^2} \frac{\alpha(B) - \alpha(0)}{\alpha(0)}$

Изотермические эффекты: $\nabla_x T = 0$; адиабатические эффекты: $W_z = 0$. Магнитное поле $\mathbf{B} = (0, B_y, 0)$.

§ 50. О КИНЕТИЧЕСКИХ ЯВЛЕНИЯХ В ПОЛУПРОВОДНИКАХ
С ТЕНЗОРНОЙ ЭФФЕКТИВНОЙ МАССОЙ

В § 40 мы разобрали явление электропроводности в полупроводниках, имеющих тензорную эффективную массу. При наложении магнитного поля выражение для тока усложняется существенно. Прежде всего рассмотрим величину $(\mathbf{V}, \mathbf{m}^* \mathbf{V})$. Пусть компоненты \mathbf{V} в главных осях тензора \mathbf{m}^* (или \mathbf{m}^{*-1} , или $\sigma^{(v)}$) равны $(B_1, B_2, B_3) = \mathbf{V}$, в таком случае вектор $\mathbf{m}^* \mathbf{V}$ имеет компоненты

$$\mathbf{m}^* \mathbf{V} = (m_1 B_1; m_2 B_2; m_3 B_3) \quad (50.1)$$

и, следовательно, скалярное произведение $(\mathbf{V}, \mathbf{m}^* \mathbf{V})$ зависит от направления \mathbf{V} относительно главных осей эллипсоида \mathbf{m}^* :

$$(\mathbf{V}, \mathbf{m}^* \mathbf{V}) = m_1 B_1^2 + m_2 B_2^2 + m_3 B_3^2 = m_t (B_1^2 + B_2^2) + m_l B_3^2. \quad (50.2)$$

Оно может меняться от $m_l B^2$ при ориентации поля \mathbf{V} вдоль оси вращения эллипсоида энергии до $m_t B^2$, когда \mathbf{V} лежит в плоскости, перпендикулярной оси вращения. Но это значит, что кинетический коэффициент K'_{rs} зависит от направления магнитного поля \mathbf{V} . Если определить направление \mathbf{V} относительно оси вращения углом θ , то

$$(\mathbf{V}, \mathbf{m}^* \mathbf{V}) = B^2 (m_t \sin^2 \theta + m_l \cos^2 \theta) \quad (50.3)$$

является периодической функцией угла θ .

Запишем выражение (38.7) для $\mathbf{j}^{(v)}$:

$$\mathbf{j}^{(v)} = e^2 K'_{11}{}^{(v)} \mathbf{E} + e^3 K'_{12}{}^{(v)} [\mathbf{m}^{*-1} \mathbf{E}, \mathbf{V}] + \frac{e^4}{|\mathbf{m}^*|} K'_{13}{}^{(v)} (\mathbf{E} \mathbf{V}) \mathbf{m}^* \mathbf{V}. \quad (50.4)$$

Мы видим, что выражение (50.4) для тока имеет значительно более сложный вид по сравнению со случаем $\mathbf{V} = 0$. Не будем анализировать это выражение и рассматривать кинетические явления на основе уравнения (50.4). Принципиальных трудностей это не представляет. Запишем только уравнение (50.4) в тензорном виде. Первый член определяет, как мы уже видели, тензор второго ранга $\sigma_{ij}^{(v)}$. Второй член можно представить через тензор третьего ранга $\sigma_{ijk}^{(v)}$, и третий член позволяет ввести тензор четвертого ранга $\sigma_{ijkl}^{(v)}$, компоненты которых зависят от магнитного поля. Для слабого поля \mathbf{V} , когда $\frac{e^2 \tau^2}{|\mathbf{m}^*|} (\mathbf{V}, \mathbf{m}^* \mathbf{V}) \ll 1$, введенные тензоры не будут зависеть от магнитного поля и $\mathbf{j}^{(v)}$ можно представить в виде

$$\mathbf{j}_i^{(v)} = \sum_{l=1}^3 \sigma_{il}^{(v)} E_l + \sum_{l, k=1}^3 \sigma_{ilk}^{(v)} E_l B_k + \sum_{l, m, k=1}^3 \sigma_{ilmk}^{(v)} E_l B_m B_k. \quad (50.5)$$

Тензор второго ранга был найден ранее в § 40:

$$\sigma_{il}^{(v)} = \sigma_i^{(v)} \delta_{il} = \frac{e^2 n_v \langle \tau \rangle}{m_i} \delta_{il}. \quad (50.6)$$

Тензор третьего ранга имеет вид

$$\sigma_{ilk}^{(v)} = e^3 n \frac{\langle \tau^2 \rangle}{v m_l m_i} \varepsilon_{ilk}, \quad (50.7)$$

где ε_{ilk} — так называемый *единичный антисимметричный тензор*, определяемый условием

$$\varepsilon_{ilk} = \begin{cases} \pm 1 & \text{при } i \neq l \neq k, \\ 0 & \text{при } i = l; l = k; k = i. \end{cases} \quad (50.8)$$

Плюс берется, если ilk можно получить из (1, 2, 3) четным числом перестановок, и минус — если нечетным, именно

$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = 1; \quad \varepsilon_{132} = \varepsilon_{321} = \varepsilon_{213} = -1, \quad (50.9)$$

остальные члены равны нулю. Выражение (50.8) легко получить, раскрывая член $K_{12}^{(v)} [m^* E, B]$. Найдем i -компонент $j_i^{(v)}$, обусловленный вторым слагаемым

$$j_i^{(v)} = \sum_l K_{12}^{(v)} [m^{*-1} E, B]_l. \quad (50.10)$$

Но, например,

$$[m^{*-1} E, B]_x = \frac{E_y B_z}{m_y} - \frac{E_z B_y}{m_z} \quad (50.11)$$

и

$$K_{12}^{(v)} = \frac{n_v \langle \tau^2 \rangle}{m_x} \delta_{ix}. \quad (50.12)$$

Мы видим, что в сумме (50.10) фигурируют члены вида $\frac{n_v \langle \tau^2 \rangle}{m_i m_l} E_l B_k$, причем $i \neq l \neq k$, с различным знаком в соответствии с четностью перестановок. Используя единичный антисимметричный тензор третьего ранга, приходим к выражению (50.7) и (50.5). Таким же образом можно прийти к представлению третьего слагаемого в виде (50.5), где $\sigma_{ijkl}^{(v)}$ можно записать в виде

$$\begin{aligned} \sigma_{ijkl}^{(v)} &= n_v e^4 \langle \tau^3 \rangle \left[\frac{1}{3} \left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3} \right) \right]^{-1} \times \\ &\times \left[\frac{1}{3} \left(\frac{1}{m_1 m_2} + \frac{1}{m_2 m_3} + \frac{1}{m_3 m_1} \right) \right]^2 F_{ijkl}^{(v)} = \\ &= n_v e^4 \langle \tau^3 \rangle \frac{(2m_l + m_i)^2}{3m_i^3 m_l (m_l + 2m_i)} F_{ijkl}^{(v)}. \end{aligned} \quad (50.13)$$

Компоненты тензора анизотропии $F_{ijkl}^{(v)}$ представлены в табл. 15.

Для того чтобы найти соответствующие значения компонентов полных (суммарных) тензоров, необходимо найти сумму соответст-

вующих компонентов тензоров ν -эллипсоида, записав их в одной и той же системе координат. Если провести соответствующие расчеты, то получим:

для проводимости

$$\begin{aligned}\sigma_{11} &= \sigma_{22} = \sigma_{33} = \sigma_0 = en\mu_d; \\ n &= Mn_\nu; \\ \mu &= \frac{e\langle\tau\rangle}{\tilde{m}^*}; \\ \frac{1}{\tilde{m}^*} &= \frac{1}{3} \left(\frac{2}{m_t} + \frac{1}{m_l} \right); \end{aligned} \quad (50.14)$$

для проводимости Холла

$$\sigma_{ijk} = \frac{e^3 n \langle\tau^2\rangle}{\tilde{m}^{*2}} \varepsilon_{ijk}; \quad \frac{1}{\tilde{m}^{*2}} = \frac{1}{3} \left(\frac{1}{m_l^2} + \frac{2}{m_l m_t} \right) \quad (50.15)$$

и

$$\sigma_{ijkl} = \frac{\langle\tau^3\rangle}{\langle\tau\rangle^3} \sigma_0 \mu^2 \left(\frac{\tilde{m}}{\tilde{m}^*} \right)^4 F_{ijkl}. \quad (50.16)$$

Значения коэффициентов F_{ijkl} приведены в табл. 15 для двух случаев: экстремумы находятся вдоль направлений [100], как в кремнии, и вдоль [111], как в германии. Тензор σ_{ij} определяет ток под

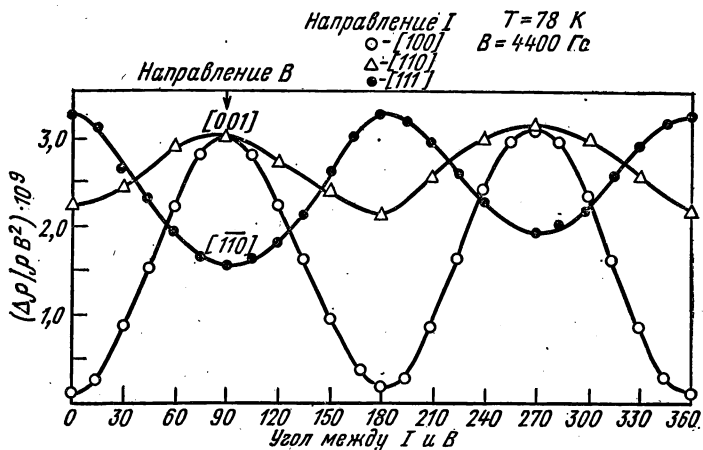


Рис. 75. Зависимость магнетосопротивления от угла между током и магнитным полем

действием электрического поля; σ_{ijk} определяет холлов ток и эффект Холла. Величина σ_{ijkl} связана с магнетосопротивлением. Гальваномагнитные эффекты можно описать, используя уравнение (50.5) с соответствующими значениями компонентов тензоров второго, третьего и четвертого ранга. На рис. 75 в качестве иллюстрации

приведена зависимость коэффициента магнетосопротивления $H = \frac{1}{B^2} \left(\frac{\Delta \rho}{\rho_0} \right)$ от угла между током \mathbf{j} и магнитным полем \mathbf{B} для германия n -типа.

§ 51. ТЕНЗОРЕЗИСТИВНЫЙ ЭФФЕКТ. ТЕНЗОЧУВСТВИТЕЛЬНОСТЬ

Тензорезистивным эффектом, или тензосопротивлением¹, называют изменение электрического сопротивления полупроводника в результате действия нагрузки, создающей деформацию. Непосредственно тензосопротивление не вытекает из кинетического уравнения, однако его необходимо отнести к кинетическим явлениям, поскольку оно представляет собой изменение одного из важнейших кинетических явлений — сопротивления, или проводимости, — под действием напряжений в теле, создаваемых внешней нагрузкой.

Для описания тензорезистивного эффекта необходимы некоторые понятия из теории упругости, которые кратко напомним. Деформация тела приводит к тому, что координаты точек тела меняются. Если до деформации координаты точек тела определяются радиусом-вектором \mathbf{r} , а в деформированном теле \mathbf{r}' , то

$$\mathbf{u} = \mathbf{r}' - \mathbf{r} \quad (51.1)$$

называется вектором деформации, или вектором смещения. Задание вектора \mathbf{u} как функции точек кристалла — $\mathbf{u}(\mathbf{r})$ — полностью определяет состояние деформированного тела. Деформация тела может быть описана симметричным тензором деформации ϵ_{ik} :

$$\epsilon_{ik} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) = \epsilon_{ki}. \quad (51.2)$$

Тензор деформации, записанный в своих главных осях, имеет диагональный вид

$$\epsilon_{ik} = \epsilon^{(i)} \delta_{ik}. \quad (51.3)$$

Тензор деформации ϵ_{ik} определяет изменение расстояний между точками тела при деформации.

Если в недеформированном теле расстояние между двумя точками dl , а в деформированном теле расстояние между теми же точками равно dl' , то связь между dl' и dl можно выразить через тензор деформации ϵ_{ik} . Действительно,

$$dl'^2 = dx^2 + dy^2 + dz^2 = \sum_{i=1}^3 dx_i^2 \quad (51.4)$$

и

$$dl'^2 = \sum_{i=1}^3 dx_i'^2 = \sum_{i=1}^3 (dx_i + du_i)^2 = \sum_{i=1}^3 (dx_i^2 + 2dx_i du_i + du_i^2). \quad (51.5)$$

¹ Тензосопротивление в некоторых случаях называют пьезосопротивлением.

Пренебрегая величиной du_i^2 как величиной второго порядка малости, выражая du_i через dx_k :

$$du_i = \sum_{k=1}^3 \frac{\partial u_i}{\partial x_k} dx_k \quad (51.6)$$

и учитывая (51.4), получим

$$\begin{aligned} dl'^2 &= \sum_{i=1}^3 dx_i'^2 = dl^2 + 2 \sum_{i,k=1}^3 \frac{\partial u_i}{\partial x_k} dx_i dx_k = dl^2 + \sum_{i,k} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) dx_i dx_k = \\ &= dl^2 + 2 \sum_{i,k} u_{ik} dx_i dx_k. \end{aligned} \quad (51.7)$$

Если тензор деформации имеет диагональный вид (51.3), то (51.7) упрощается:

$$dl'^2 - dl^2 = 2 \sum_i u^{(i)} \delta_{ik} dx_i dx_k = 2 \sum_i u^{(i)} dx_i^2. \quad (51.8)$$

Из (51.8) можем записать

$$dl' = \sqrt{dl^2 + 2 \sum_{i,k} u_{ik} dx_i dx_k} = dl \sqrt{1 + 2 \sum_{i,k} u_{ik} n_i n_k}. \quad (51.9)$$

Относительное удлинение вдоль направления $\mathbf{n} = (n_1, n_2, n_3)$ можно выразить через тензор деформации следующим образом:

$$u = \frac{dl' - dl}{dl} = \sqrt{1 + 2 \sum_{i,k} u_{ik} n_i n_k} - 1 \cong \sum u_{ik} n_i n_k. \quad (51.10)$$

Найдем изменение объема при деформации:

$$\begin{aligned} d\tau' &= dx' dy' dz' = dx (1 + u^{(1)}) dy (1 + u^{(2)}) dz (1 + u^{(3)}) = \\ &= d\tau (1 + u^{(1)} + u^{(2)} + u^{(3)}), \end{aligned} \quad (51.11)$$

или

$$\frac{d\tau' - d\tau}{d\tau} = \Delta = \sum_i u^{(i)} = \sum_i u_{ii} = \sum_i \frac{\partial u_i}{\partial x_i} = \operatorname{div} \mathbf{u}. \quad (51.12)$$

Таким образом, главные значения тензора деформации определяют относительные удлинения вдоль главных осей тензора, а их сумма — относительное изменение объема.

Тензор деформации $\{u_{ik}\}$ можно представить в несколько другом виде, для этого запишем тождество:

$$u_{ik} = \left[u_{ik} - \left(\frac{1}{3} \sum_{l=1}^3 u_{ll} \right) \delta_{ik} \right] + \left(\frac{1}{3} \sum_{l=1}^3 u_{ll} \right) \delta_{ik}. \quad (51.13)$$

Второй тензор, имеющий диагональный вид с одинаковыми диагональными элементами, равными одной трети относительного изменения объема Δ , называют тензором всестороннего сжатия.

Тензор, определенный членом, стоящим в квадратной скобке (51.13), называют *тензором сдвига*. Так как для него сумма диагональных элементов равна нулю, то *при чистом сдвиге изменения объема не происходит*.

При деформации тела в нем возникают внутренние напряжения, стремящиеся вернуть тело в равновесное, недеформированное состояние. Они могут быть описаны некоторым симметричным тензором второго ранга — *тензором напряжений* $\{p_{ik}\}$. Величина p_{ik} представляет собой проекцию силы на ось i , действующую на единичную площадку, перпендикулярную оси k . Сила $F^{(1)}$, действующая на единичный объем тела, может быть выражена через тензор напряжений p следующим образом:

$$F^{(1)} = \operatorname{div} p, \quad (51.14)$$

$$F_i^{(1)} = \sum_k \frac{\partial p_{ik}}{\partial x_k} \quad (i = 1, 2, 3). \quad (51.15)$$

Между тензором деформации и тензором напряжений должна быть определенная связь. Действительно, с ростом деформации напряжения должны возрастать. В пределах упругих деформаций между деформацией и напряжением на основании закона Гука должна существовать линейная зависимость. Величину, связывающую деформацию с напряжением, называют обычно *модулем упругости*. Поскольку в общем случае деформация и напряжение являются тензорами второго ранга, и модуль упругости должен быть тензором, причем более высокого ранга, именно четвертого. Обозначим его элементы через λ_{ijkl} , тензор λ называют *тензором модулей упругости*, или просто *тензором упругости*. В соответствии с законом Гука запишем

$$p_{ik} = \sum_m \lambda_{iklm} u_{lm}. \quad (51.16)$$

Тензор модулей упругости λ симметричен по парам своих индексов:

$$\lambda_{ijkl} = \lambda_{jikl} = \lambda_{ijlk} = \lambda_{klij}, \quad (51.17)$$

так как он связывает между собой два симметричных тензора p_{ik} и u_{lm} . С учетом (51.17) в общем случае из 81 элемента тензора λ отличными друг от друга элементами могут быть не более 21. С учетом свойств симметрии решетки число независимых модулей упругости может быть уменьшено. Например, в триклинной системе число независимых модулей равно 18, в ромбоэдрической — 12, в гексагональной — 5 и в кубической только три, которые можно обозначить как

$$\lambda_{xxxx}; \lambda_{xxyy}; \lambda_{xyxy}. \quad (51.18)$$

Изотропные тела характеризуются всего двумя модулями упругости — модулями сдвига и всестороннего сжатия.

Если тело находится в равновесии, то полная сила, действующая на единичный объем тела, равна нулю, следовательно, условием равновесия является

$$F_i^{(v)} = \sum_k \frac{\partial p_{ik}}{\partial x_k} = 0. \quad (51.19)$$

Пусть p есть внешнее давление, действующее на тело. Возьмем элемент поверхности dS . Внешняя сила уравнивается внутренними напряжениями:

$$pdS_i = p_i dS = \sum_k p_{ik} dS_k. \quad (51.20)$$

Обозначим $dS = n dS$, где n — внешняя нормаль к поверхности тела. В таком случае

$$p_i dS = \sum_k p_{ik} n_k dS, \quad (51.21)$$

или

$$p_i = \sum_k p_{ik} n_k, \quad (51.22)$$

где n_k — проекция нормали на ось k . При всестороннем сжатии $p_i = -p$ и, учитывая, что давление носит нормальный характер, получим

$$p_{ik} = -p \delta_{ik}. \quad (51.23)$$

Если тело подвергнуть *одностороннему сжатию* $p = (0, 0, p)$, то

$$\{p_{ik}\} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -p \end{pmatrix}. \quad (51.24)$$

В случае, когда давление определяется вдоль направлений $n = (n_1, n_2, n_3)$, тензор напряжения имеет вид

$$p_{ik} = -p n_i n_k. \quad (51.25)$$

Учитывая связь (51.16) между p_{ik} и ϵ_{lm} , можно выразить деформацию через внешнее давление. Поскольку тензор модулей упругостей имеет значительно меньше 81 компонента, для него используют обозначения, указывая не четыре индекса, а только два. Например, в случае кристаллов с кубической решеткой вводят следующие обозначения:

$$\lambda_{1111} \equiv \lambda_{xxxx} = c_{11}; \quad \lambda_{1122} = c_{12}; \quad \lambda_{1212} = c_{44}.$$

Компоненты c_{ik} являются симметричными относительно своих индексов.

Перейдем к описанию тензорезистивного эффекта. Пусть вещество характеризуется тензором удельного сопротивления ρ^0 с компонентами ρ_{ik}^0 .

Если полупроводник деформирован, то его удельное сопротивление изменилось, оно равно ρ^T или ρ_{ik}^T . Величина $\rho^T - \rho^0$, или $\rho_{ik}^T - \rho_{ik}^0$, представляет собой изменение удельного сопротивления в результате действия нагрузки, вызывающей деформацию и напряжения в полупроводнике. Изменение удельного сопротивления можно выразить двояким образом: либо через напряжения, либо через деформацию. Поскольку между напряжением и деформацией существует вполне определенная связь, оба способа описания являются эквивалентными. Выразим изменение сопротивления $\rho^T - \rho^0$ через тензор напряжения ρ_{ik} в виде

$$\rho_{ik}^T - \rho_{ik}^0 = \rho_{ik}^0 \sum_{lm} \pi_{iklm} p_{lm}, \quad (51.26)$$

или

$$\rho_{ik}^T = \rho_{ik}^0 \left(1 + \sum_{lm} \pi_{iklm} p_{lm} \right), \quad (51.27)$$

$$\frac{\rho_{ik}^T - \rho_{ik}^0}{\rho_{ik}^0} = \sum_{lm} \pi_{iklm} p_{lm}. \quad (51.28)$$

Тензор четвертого ранга π_{iklm} называют обычно тензором коэффициентов пьезосопротивления, или тензором пьезосопротивления.

Для кубического кристалла тензор пьезосопротивления имеет всего три различные величины, которые для простоты изображаются величинами с двумя индексами:

$$\pi_{1111} = \pi_{11}; \quad \pi_{1122} = \pi_{12}; \quad \pi_{1212} = \pi_{44}. \quad (51.29)$$

Если учесть (51.16), то ρ_{ik}^T можно выразить через тензор деформации и тензор модулей упругости в виде

$$\rho_{ik}^T = \rho_{ik}^0 \left(1 + \sum_{lm} \pi_{iklm} p_{lm} \right) = \rho_{ik}^0 \left(1 + \sum_{lm, st} \pi_{iklm} \lambda_{lmst} u_{st} \right). \quad (51.30)$$

При экспериментальном исследовании тензорезистивного эффекта выбирают образец в виде стержня прямоугольного сечения, который подвергают одностороннему сжатию или растяжению. Если ток направлен вдоль оси, по которой направлено растяжение, то мы имеем продольный эффект, если ток направлен под углом 90° к оси, вдоль которой приложена внешняя нагрузка, то мы имеем поперечный эффект.

Запишем выражение для сопротивления ρ_{ik}^T при одностороннем сжатии ($p > 0$), учитывая (51.25):

$$\rho_{ik}^T = \rho_{ik}^0 \left(1 + \sum_{lm} \pi_{iklm} p_{lm} \right) = \rho_{ik}^0 \left(1 - p \sum_{lm} \pi_{iklm} n_l n_m \right). \quad (51.31)$$

Пусть ток j течет вдоль оси стержня, тогда $j_k = j n_k$. Найдем компоненты поля E_i на основании закона Ома:

$$E_i = \sum_k \rho_{ik}^T j_k = \sum_k \rho_{ik}^0 \left(1 - p \sum_{lm} \pi_{iklm} n_l n_m \right) j n_k. \quad (51.32)$$

Напряженность поля E вдоль оси стержня обозначим через E^{\parallel} :

$$E^{\parallel} = (En) = \sum_{i=1}^3 E_i n_i. \quad (51.33)$$

Назовем *продольным удельным сопротивлением* ρ^{\parallel} величину, равную отношению E^{\parallel} к j :

$$\rho^{\parallel} = \frac{E^{\parallel}}{j} = \frac{\sum_i E_i n_i}{j} = \sum_{ik} \rho_{ik}^0 \left(1 - p \sum_{lm} \pi_{iklm} n_l n_m \right) n_k n_i. \quad (51.34)$$

При $p=0$

$$\rho^{\parallel} = \sum_{ik} \rho_{ik}^0 n_i n_k = \rho_0^{\parallel}. \quad (51.35)$$

Продольным коэффициентом тензосопротивления, или тензочувствительностью по напряжению, называют величину π_l (или π_{long}):

$$\pi_l = \frac{\rho^{\parallel} - \rho_0^{\parallel}}{\rho_0^{\parallel} (-p)}. \quad (51.36)$$

Тензосопротивление вдоль оси образца можно выразить с помощью коэффициента π_l :

$$\rho^{\parallel} = \rho_0^{\parallel} (1 - \pi_l p). \quad (51.37)$$

Выразим *продольный коэффициент тензосопротивления π_l через тензор пьезосопротивления*, исходя из (51.36), (51.35) и (51.34):

$$\pi_l = \sum_{iklm} \pi_{iklm} n_i n_k n_l n_m. \quad (51.38)$$

Сумма (51.38) легко вычисляется, если учесть (51.29):

$$\begin{aligned} \pi_l &= \pi_{11} (n_1^4 + n_2^4 + n_3^4) + 2 (\pi_{12} + \pi_{44}) (n_1^2 n_2^2 + n_1^2 n_3^2 + n_2^2 n_3^2) = \\ &= \pi_{11} + 2 (\pi_{44} + \pi_{12} - \pi_{11}) (n_1^2 n_2^2 + n_1^2 n_3^2 + n_2^2 n_3^2). \end{aligned} \quad (51.39)$$

Если образец (стержень) вырезан так, что его ось совпадает с направлением типа [100], то вектор \mathbf{n} имеет только одну ненулевую проекцию $\mathbf{n} = (1, 0, 0)$, вследствие чего

$$\pi_{l[100]} = \pi_x = \pi_y = \pi_z = \pi_{11}. \quad (51.40)$$

Если образец вырезан вдоль направления типа [110], то $\mathbf{n} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right)$ и

$$\pi_{l[110]} = \frac{\pi_{11} + \pi_{12} + \pi_{44}}{2}. \quad (51.41)$$

Аналогично можно получить

$$\pi_{l[111]} = \frac{2}{3} (\pi_{12} + \pi_{44}) + \frac{\pi_{11}}{3}. \quad (51.42)$$

Таким образом, коэффициент тензочувствительности π_i по напряжению зависит от направления оси образца относительно осей кристалла согласно (51.39).

Можно ввести коэффициент тензочувствительности S_i по деформации u вдоль оси. Модулем Юнга E^0 , как известно, называют величину, определяемую соотношением

$$-p = E^0 u, \quad (51.43)$$

где согласно (51.10)

$$u = \sum_{ik} u_{ik} n_i n_k. \quad (51.44)$$

Подставляя (51.43) в (51.37), получим

$$\rho^{\parallel} = \rho_0^{\parallel} (1 - p \pi_i) = \rho_0^{\parallel} (1 + E^0 \pi_i u) = \rho_0^{\parallel} (1 + S_i u), \quad (51.45)$$

или

$$S_i = \frac{\rho^{\parallel} - \rho_0^{\parallel}}{u \rho_0} = \pi_i E^0. \quad (51.46)$$

Необходимо только помнить, что определенный в (51.43) модуль Юнга E^0 зависит от направления. Его можно выразить через тензор модулей упругости согласно (51.25) и (51.10) следующим образом:

$$-p = \sum p_{ik} n_i n_k = E_n^0 \sum u_{lm} n_l n_m. \quad (51.47)$$

Выражая p_{ik} через u_{lm} согласно (51.16), получим

$$\frac{1}{E_n^0} = \frac{u}{-p} = \frac{\sum_{lm} u_{lm} n_l n_m}{\sum_{iklm} \lambda_{iklm} u_{lm} n_i n_k}, \quad (51.48)$$

или

$$\frac{1}{E_n^0} = \frac{c_{11} + c_{12}}{(c_{11} + 2c_{12})(c_{11} - c_{12})} + \left(\frac{1}{c_{44}} - \frac{2}{c_{11} - c_{12}} \right) (n_1^2 n_2^2 + n_1^2 n_3^2 + n_2^2 n_3^2). \quad (51.49)$$

Например, для направления вдоль оси [100], $\mathbf{n} = (1, 0, 0)$ получим

$$E^0 = \frac{c_{11} - c_{12}}{c_{11} - c_{22}} (c_{11} + 2c_{22}) = E_{[100]}^0. \quad (51.50)$$

Тензочувствительность полупроводников в десятки раз превосходит тензочувствительность металлов. Например, для p -кремния с удельным сопротивлением $0,1 \text{ Ом} \cdot \text{см}$ S_i имеет значение около 125 (тензочувствительность по относительной деформации S_i безразмерная величина), что примерно в 60 раз больше соответствующей величины для проволочных тензометров.

**§ 52. ТЕНЗОРРЕЗИСТИВНЫЙ ЭФФЕКТ.
КОЭФФИЦИЕНТЫ ПЬЕЗОСОПРОТИВЛЕНИЯ**

В табл. 16 приведены некоторые экспериментальные данные по германию и кремнию для коэффициентов пьезосопротивления.

Таблица 16

Адиабатические коэффициенты пьезосопротивления.
($T = 20^\circ\text{C}$)

Материал	ρ_0 Ом·см	π_{11}	π_{12}	π_{44}	$\pi_l = \pi_{[111]}$	$E_{[111]}^0$ дин/см	$S_{[111]} =$ $= E_{[111]}^0 \times$ $\times \pi_{[111]}$
		(в $10^{-12}\text{см}^2/\text{дин}$)					
Германий: <i>n</i> -тип	1,5	-2,3	-3,2	-138,1	-94,9	$1,55 \cdot 10^{12}$	-147
	5,7	-2,7	-3,9	-136,8	-94,7		-147
	9,9	-4,7	-5,0	-137,9	-96,9		-150
	16,6	-5,2	-5,5	-138,7	-101,2		-157
	1,1	-3,7	3,2	96,7	65,4		101,5
<i>p</i> -тип	15,0	10,6	5,0	46,5	31,4	1,87 · 10 ¹²	48,7
	7,8	6,6	-1,1	138,1	93,6	$1,87 \cdot 10^{12}$	—
Кремний: <i>n</i> -тип	11,7	-102,2	53,4	-13,6	-81,3		175
<i>p</i> -тип							-142

Табл. 16 показывает, что как коэффициенты пьезосопротивления, так и тензочувствительность S_l и π_l зависят от величины и типа проводимости, а также от температуры и даже от деформации. Чтобы понять эти зависимости, необходимо рассмотреть физическую природу тензорезистивного эффекта. В общем случае тензор удельного сопротивления можно выразить через концентрацию носителей заряда и тензор подвижности μ :

$$\rho^{-1} = en\mu = \sigma. \quad (52.1)$$

Изменение ρ может быть связано с изменением концентрации и подвижности.

1. Объемное сжатие. Простейшим случаем тензосопротивления является случай объемного, или всестороннего, сжатия. Тензор деформации вырождается в скаляр $\epsilon_{ii} = \epsilon$. Постоянная решетки a_0 кристалла уменьшается

$$a' = a_0(1 - \epsilon). \quad (52.2)$$

Так как расстояние между атомами уменьшается, то увеличивается перекрытие волновых функций электронов, а потенциальная энергия $W(a)$, описывающая взаимодействие атомов решетки, изменяется. Если считать, что минимум $W(a)$ соответствует $a = a_0$, то $W(a)$ возрастает как при сжатии, так и при растяжении кристалла

(рис. 76). Учитывая изменение W (a) и области перекрытия волновых функций, мы можем прийти к выводу, что *меняется как обменный интеграл A (s), так и величина опускания уровней энергии C* , которые были введены в теории квазисвязанного электрона. Это приводит в свою очередь к изменению ширины запрещенных зон и зон энергии.

Но изменение ширины запрещенной зоны должно привести к изменению концентрации электронов и дырок. Обозначим через β коэффициент относительного изменения ширины запрещенной зоны при изменении давления:

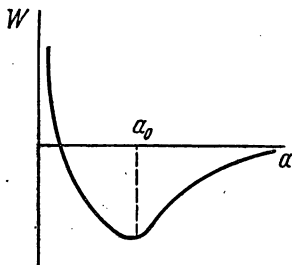


Рис. 76. Зависимость энергии взаимодействия атомов решетки от расстояния между ними

$$\beta = -\frac{1}{\Delta E_0} \frac{\partial \Delta E_0}{\partial p};$$

$$\Delta E_0(p) = \Delta E_0(1 - \beta p). \quad (52.3)$$

Если через γ обозначить коэффициент относительного изменения ширины запрещенной зоны с деформацией:

$$\gamma = \frac{1}{\Delta E_0} \frac{\partial \Delta E_0}{\partial u};$$

$$\Delta E_0(u) = \Delta E_0(1 + \gamma u), \quad (52.4)$$

то из (52.4), (52.3) и (51.43) можно записать

$$\beta = \frac{\gamma}{E^0}, \quad (52.5)$$

где E^0 — модуль Юнга.

Изменение ширины запрещенной зоны обусловлено тем, что смещается дно зоны проводимости и потолок валентной зоны. При этом необходимо иметь в виду, что смещение E_c и E_v может происходить по-разному. В общем случае положение E_c и E_v является функцией относительного изменения объема или тензора деформации:

$$\tilde{E}_c(\mathbf{u}) = E_c + \frac{\partial E_c}{\partial \mathbf{u}} \mathbf{u} + \frac{1}{2} \frac{\partial^2 E_c}{\partial u^2} \mathbf{u}^2 + \dots \quad (52.6)$$

Ограничиваясь малыми деформациями, можно отбросить все более высокие степени относительной деформации, кроме первой:

$$E_c(\mathbf{u}) = E_c(0) + \Delta_c \mathbf{u}. \quad (52.7)$$

Дополнительная потенциальная энергия электрона в деформированной решетке носит название потенциала деформации, а Δ_c называют постоянной потенциала деформации. Аналогично можно представить для потолка валентной зоны

$$E_v(\mathbf{u}) = E_v(0) + \Delta_v \mathbf{u}. \quad (52.8)$$

Изменение ширины запрещенной зоны при деформации можно записать в виде

$$\Delta E_0(\mathbf{u}) = E_c(\mathbf{u}) - E_v(\mathbf{u}) = \Delta E(0) + (\Delta_c - \Delta_v) \mathbf{u}, \quad (52.9)$$

другими словами величину γ можно выразить через Δ_c и Δ_v . Выразим произведение концентраций носителей заряда через тензор деформации:

$$\begin{aligned} np &= n_i^2 = N_c N_v e^{-\frac{\Delta E_0(u)}{kT}} = n_i^2(0) e^{-\frac{(\Delta_c - \Delta_v)u}{kT}} = \\ &= n_i^2(0) \left(1 - \frac{\Delta_c - \Delta_v}{kT} u + \dots \right). \end{aligned} \quad (52.10)$$

Таким образом, приращение произведений концентраций электронов и дырок связано с постоянными потенциала деформации и величиной деформации u . В собственном полупроводнике $n = p$ и

$$\frac{\delta p}{p(0)} = \frac{\delta n_i}{n_i(0)} = \frac{\Delta_v - \Delta_c}{2kT} u = \frac{\delta p}{p(0)}. \quad (52.11)$$

Учитывая выражение для проводимости, можем записать

$$\frac{\delta \rho}{\rho_0} = -\frac{\delta n}{n} = \frac{\Delta_c - \Delta_v}{2kT} u = \frac{\Delta_v - \Delta_c}{2kT} \frac{\rho}{E^{10}}. \quad (52.12)$$

Сравнивая (52.12) с (51.36), можем записать выражение для продольного коэффициента пьезосопротивления

$$\pi_l = \frac{\Delta_c - \Delta_v}{2kTE^{10}}; \quad S_l = \frac{\Delta_c - \Delta_v}{2kT}. \quad (52.13)$$

Оценка π_l на основании (52.13) дает сравнительно небольшое изменение сопротивления при всестороннем сжатии. Так как согласно (51.13) всестороннее сжатие в общем случае входит как составная часть в любую деформацию, кроме чистого сдвига, то изменение концентрации свободных носителей заряда, обусловленное всесторонним сжатием, должно наблюдаться во всех случаях.

Однако в области истощения примеси всестороннее сжатие может изменить общую концентрацию частиц только на удвоенное изменение концентрации неосновных носителей заряда, поэтому тензосопротивление проявится слабо. В действительности же в целом ряде случаев тензосопротивление проявляется значительно сильнее, что может быть объяснено только сложной структурой зон энергии.

2. Одностороннее сжатие или растяжение. Значительно большее тензосопротивление наблюдается при одностороннем сжатии или растяжении, когда тензор напряжения имеет вид (51.24) или (51.25). Деформация тела определяется тензором деформации u_{ik} . Необходимо отметить следующее обстоятельство: если в направлении сжатия расстояние между атомами уменьшается, то в поперечном направлении расстояние между атомами увеличивается. Это по-разному меняет характер перекрытия волновых функций атомов вдоль разных направлений.

Выразим положение дна зоны проводимости в виде

$$E_c(u) = E_c(0) + \sum_{ik} \Delta_{ik}^{(c)} u_{ik}, \quad (52.14)$$

где u_{ik} — компонент тензора деформации, а $\Delta_{ik}^{(c)}$ — компонент тензора коэффициентов потенциала деформации. Выражение (52.14) справедливо

ливо для малых деформаций. Обобщим поставленную задачу на случай полупроводника, имеющего M долин. Предположим, что положение дна зоны проводимости в каждой долине (ν -долине) описывается своим тензором $\Delta_{ik}^{(\nu)} = \Delta_{ik}^{(\nu)}$:

$$E_c^{(\nu)}(\mathbf{u}) = E_c(0) + \sum_{ik} \Delta_{ik}^{(\nu)} u_{ik}. \quad (52.15)$$

Поскольку тензор деформации одинаков для всех долин, а $\Delta^{(\nu)}$, вообще говоря, различны, то *дно зоны проводимости сместится по-разному в каждой долине*. Положение уровня Ферми от номера долины не зависит, поэтому расстояния между уровнем Ферми и дном зоны становятся различными, что приводит к различной величине концентраций электронов в каждой долине: если дно долины поднимается, то число электронов в ней уменьшается; если дно долины опускается, то число электронов в ней увеличивается:

$$n^{(\nu)} = N_c e^{\frac{F' - E_c^{(\nu)}(\mathbf{u})}{kT}} = n^{(\nu)}(\mathbf{u}). \quad (52.16)$$

Через F' обозначен уровень Ферми в деформированном кристалле:

$$F' = F + \delta F; \quad \delta F = \sum_{ik} D_{ik} u_{ik}. \quad (52.17)$$

Наиболее интересен случай примесной проводимости. Рассматривая для простоты область истошения примеси, мы можем записать условие сохранения числа электронов проводимости

$$\sum_{\nu=1}^M n^{(\nu)}(\mathbf{u}) = n^{(\nu)}(0) M = n. \quad (52.18)$$

Разлагая (52.16) в ряд по малой величине тензора деформации и ограничиваясь линейным членом, получим

$$n^{(\nu)}(\mathbf{u}) = n^{(\nu)}(0) \left(1 - \frac{\delta F - \sum_{ik} \Delta_{ik} u_{ik}}{kT} + \dots \right). \quad (52.19)$$

Подставляя (52.19) в (52.18), запишем

$$n^{(\nu)}(0) M = n^{(\nu)}(0) M - \frac{n^{(\nu)}(0)}{kT} \left[\sum_{\nu=1}^M \left(\delta F - \sum_{ik} \Delta_{ik}^{(\nu)} u_{ik} \right) \right]. \quad (52.20)$$

Приравнявая члены в квадратной скобке (52.20) к нулю и вместо δF подставляя (52.17), получим

$$M \delta F = \sum_{\nu=1}^M \sum_{i, k=1}^3 \Delta_{ik}^{(\nu)} u_{ik} = M \sum_{i, k=1}^3 D_{ik} u_{ik}. \quad (52.21)$$

Из (52.21) имеем

$$D_{ik} = \frac{1}{M} \sum_{\nu=1}^M \Delta_{ik}^{(\nu)} = \langle \Delta \rangle_{ik}, \quad (52.22)$$

т. е. тензор D_{ik} , определяющий смещение уровня Ферми, обеспечивающее сохранение числа электронов проводимости, представляет собой усредненный по минимумам тензор $\Delta_{ik}^{(\nu)}$. Учитывая (52.22), перепишем выражение для $n^{(\nu)}(\mathbf{u})$:

$$n^{(\nu)}(\mathbf{u}) = n^{(\nu)}(0) \left[1 - \frac{\sum_{ik} (\langle \Delta \rangle_{ik} - \Delta_{ik}^{(\nu)}) u_{ik}}{kT} \right] = n^{(\nu)}(0) + \delta n^{(\nu)}. \quad (52.23)$$

Так как концентрация электронов в различных минимумах при деформации меняется по-разному, то минимумы перестают быть эквивалентными. Запишем теперь выражение для полной проводимости

$$\sigma(\mathbf{u}) = \sum_{\nu=1}^M \sigma^{(\nu)}(\mathbf{u}) = e \sum_{\nu=1}^M n^{(\nu)}(\mathbf{u}) \mu^{(\nu)}. \quad (52.24)$$

Если поверхности энергии в минимумах сферические, то $\mu^{(\nu)} = \mu$ не зависит от номера долины и в этом случае согласно (52.18) $\sigma(\mathbf{u}) = \sigma(0)$, т. е. в этом приближении тензосопротивление отсутствует. Если же $\mu^{(\nu)}$ анизотропна, то проводимость становится при наложении деформации анизотропной:

$$\sigma(\mathbf{u}) = \sigma(0) - e \sum_{\nu=1}^M \mu^{(\nu)} \delta n^{(\nu)}. \quad (52.25)$$

При фиксированной деформации, определяющей значения $\delta n^{(\nu)}$, анизотропия проводимости связана с анизотропией подвижности, т. е. с анизотропией обратной эффективной массы. Рассмотрим в качестве примера кремний n -типа.

Если сжать кристалл кремния вдоль оси [100], то расстояния между атомами в этом направлении уменьшатся, обменный интеграл для этого направления возрастет, а для направлений [010] и [001] уменьшится. Дно зоны проводимости в направлении [100] опустится, а в перпендикулярных направлениях поднимется. Так как

$$\sum_{\nu=1}^M \delta n^{(\nu)} = 0, \quad (52.26)$$

то область занятых состояний в минимумах $(\pm \kappa_{0x}, 0, 0)$ увеличится, а в остальных уменьшится (рис. 77). Обозначим через δn_1 увеличение концентрации электронов в первом и четвертом минимумах, очевидно, что

$$-4\delta n_2 = 2\delta n_1. \quad (52.27)$$

Запишем изменение тензора проводимости $\delta\sigma$:

$$\delta\sigma_{xx} = -e^2 \delta n_1 (\mu_{xx}^{(1)} - 2\mu_{xx}^{(2)}) = -e^2 \langle \tau \rangle 2\delta n_1 \left(\frac{1}{m_l} - \frac{2}{m_t} \right), \quad (52.28)$$

$$\delta\sigma_{yy} = -e^2 \langle \tau \rangle \left(\frac{2\delta n_1}{m_t} + \frac{2\delta n_2}{m_t} + \frac{2\delta n_3}{m_l} \right) = -e^2 \langle \tau \rangle \delta n_1 \times \\ \times \left(\frac{1}{m_t} - \frac{1}{m_l} \right) = \delta\sigma_{zz}. \quad (52.29)$$

Таким образом, изменение проводимости для разных осей различно как по знаку, так и по модулю. Если наложить давление вдоль оси [110], то минимумы 1, 2, 4, 5 опустятся, а 3, 6 поднимутся; $4\delta n_1 = -2\delta n_3$. Выражение для изменения проводимости имеет вид

$$\delta\sigma_{xx} = \\ = -2e^2 \delta n_1 \langle \tau \rangle \left(\frac{1}{m_t} - \frac{1}{m_l} \right) = \\ = \delta\sigma_{yy}, \quad (52.30)$$

$$\delta\sigma_{zz} = -4e^2 \langle \tau \rangle \delta n_1 \left(\frac{1}{m_t} - \frac{1}{m_l} \right). \quad (52.31)$$

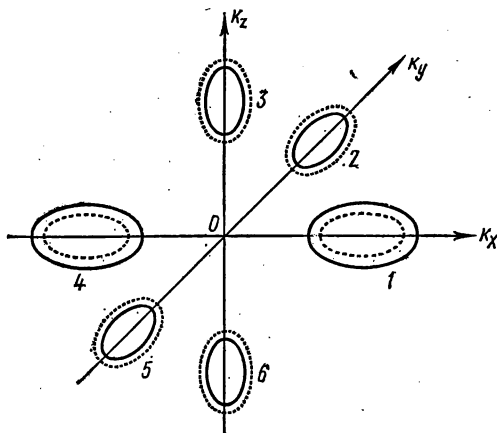


Рис. 77. Перераспределение электронов между минимумами энергии в кремнии при деформации

Если же сжать кристалл в направлении [111], то все экстремумы останутся эквивалентными, перераспределение электронов не произойдет, и в этом приближении тензосопротивление наблюдаться не должно. В действительности оно будет иметь место вследствие изменения концентрации носителей заряда, о котором говорили в первом пункте.

Перейдем к рассмотрению тензосопротивления в германии или кремнии p -типа. В них наблюдается большое изменение сопротивления, которое не может быть объяснено указанным выше изменением концентрации носителей заряда. Вспомним, что валентная зона германия и кремния обладает почти сферическими поверхностями энергии, поэтому объяснить anomalно большое тензосопротивление анизотропией проводимостей подобно тому, как это имеет место в зоне проводимости, нельзя. Объяснение было найдено при учете вырождения валентной зоны — две ветви энергии — для легких и тяжелых дырок — сливаются в точке $k=0$. При наложении анизотропной деформации нарушается симметрия поля решетки, что приводит к снятию вырождения — потолок валентной зоны легких и тяжелых дырок смещается на различную величину и в противоположных

направлениях. Величина смещения уровней энергии $\delta E(\kappa)$ зависит от κ следующим образом:

$$\delta E(\kappa) = \pm \frac{bB}{2\bar{B}} \left(\frac{3}{\kappa^2} \sum_{ij} u_{ij} \kappa_i \kappa_j - \sum_i u_{ii} \right), \quad (52.32)$$

где $\bar{B} = \left(B^2 + \frac{1}{5} C^2 \right)^{\frac{1}{2}}$; B, C — константы, входящие в выражение (26.2), b — константа деформационного потенциала. Верхний знак относится к легким дыркам, нижний — к тяжелым. Смещение зон энергии легких и тяжелых дырок меняет их концентрацию при сохранении полного числа дырок проводимости (в случае примесной проводимости в области истощения) $p_n + p_t = N_a$. Но перераспределение концентраций легких и тяжелых дырок вследствие различия в их подвижности меняет проводимость и сопротивление:

$$\delta\sigma = e_p (\delta p_n \mu_{p,n} + \delta p_t \mu_{p,t}) \quad (52.33)$$

при $\delta p_n = -\delta p_t$.

Таким образом, аномально большое тензосопротивление германия и кремния p -типа связано с различием масс и тем самым подвижностей легких и тяжелых дырок.

Если поместить деформированный кристалл в магнитное поле, то все гальвано- и термомагнитные явления должны протекать иначе, чем в недеформированном кристалле. Аналогичным образом изменятся и термоэлектрические явления. Изменение всех кинетических явлений обусловлено тем, что деформация кристалла меняет его энергетическую структуру. При более точном описании явлений в деформированных полупроводниках необходимо учитывать изменение механизма рассеяния, эффективной массы и других величин.

Выражение для плотности тока типа (50.5) изменится при наложении деформации. На основании закона Ома выражение для плотности тока можно записать в виде

$$j_i = \sum_{k=1}^3 \sigma_{ik}(\mathbf{u}, \mathbf{B}) E_k, \quad (52.34)$$

где проводимость σ_{ik} можно представить для слабых полей в виде

$$\begin{aligned} \sigma_{ik}(\mathbf{u}, \mathbf{B}) = \sigma_{ik}^0 \left[1 + \sum_{lm} (\pi^{-1})_{iklm} u_{lm} \right] + \left\{ \sum_l \sigma_{ikl} + \sum_{lmn} \sigma_{iklmn} u_{mn} \right\} B_l + \\ + \left\{ \sum_{lm} \sigma_{iklm} + \sum_{imnp} \sigma_{iklmnp} u_{np} \right\} B_l B_m. \end{aligned} \quad (52.35)$$

Если деформации отсутствуют ($u_{st} = 0$), то (52.35) переходит в выражение типа (50.5); при $\mathbf{B} = 0$ оно переходит в (51.30), если учтем, что тензоры ρ_{ik} и σ_{ik} являются взаимобратными, через $(\pi^{-1})_{iklm}$ обозначен элемент $iklm$ тензора, обратного тензору пьезосопротивления. Тензоры σ_{ikl} и σ_{iklm} описывают проводимость Холла и магнитно-

резистивный эффект. Тензоры σ_{ijklmn} и $\sigma_{ijklmnp}$ описывают соответственно поперечный и продольный тензогальваномагнитные эффекты.

Если полупроводник имеет M экстремумов, то необходимо выражение (52.35) просуммировать по всем экстремумам.

Резюме § 51—52

1. Тензорезистивный эффект состоит в изменении сопротивления (проводимости) полупроводника или металла в результате его деформации. Физической причиной тензосопротивления является изменение энергетической структуры полупроводника. Изменение ширины запрещенной зоны приводит к изменению концентрации носителей заряда и тем самым к изменению сопротивления.

2. В веществах, имеющих сложную структуру зон подобно германию или кремнию n -типа, деформация кристалла, вызванная односторонним сжатием или растяжением, приводит к большому изменению сопротивления, которое не может быть объяснено изменением общей концентрации носителей заряда. Объяснение состоит в том, что в результате анизотропной деформации экстремумы энергии становятся неэквивалентными, происходит перераспределение электронов по экстремумам. Минимумы, дно которых опустится, дадут больший вклад в проводимость, чем минимумы, дно которых поднимается. Изменение проводимости при этом наблюдается только в том случае, если поверхности энергии отличны от сферических.

3. В полупроводниках кремния p -типа большая величина изменения сопротивления обусловлена снятием вырождения зон энергии при наложении анизотропной деформации. В результате снятия вырождения меняется число легких и тяжелых дырок, обладающих различной подвижностью и дающих благодаря этому различный вклад в проводимость, что приводит к изменению сопротивления даже при сохранении общего числа дырок.

4. Тензорезистивный эффект описывается тензором коэффициентов тензосопротивления (пьезосопротивления). Он является тензором четвертого ранга.

Если на полупроводник наложить одновременно магнитное поле и деформацию, то его проводимость можно представить в случае слабых магнитных полей в виде (52.35).

Глава V

ТЕОРИЯ РАССЕЙЯНИЯ НОСИТЕЛЕЙ ЗАРЯДА

§ 53. ЭФФЕКТИВНОЕ СЕЧЕНИЕ РАССЕЙЯНИЯ

В идеальном кристалле направленное движение электронов и дырок может существовать без внешнего электрического поля сколь угодно долго. В реальном кристалле электрический ток после выключения поля уменьшается по экспоненте

$$j(t) = j_0 e^{-\frac{t}{\langle \tau \rangle}}, \quad (53.1)$$

где $\langle \tau \rangle$ — некоторая характеристика вещества, имеющая величину порядка 10^{-13} с, т. е. ток обращается в нуль практически мгновенно. *Механизмом, обеспечивающим возврат системы в равновесное состояние, является рассеяние носителей заряда, их столкновение с различными нарушениями периодичности поля решетки.*

Характеристикой процесса рассеяния служит величина, называемая обычно *эффективным сечением*. Ее можно ввести следующим образом.

Пусть через единичное сечение в единицу времени проходит одна частица со скоростью v , она образует *одночастичный поток*. Определяя поток в общем случае как число частиц, проходящих за единицу времени через единичное сечение, мы можем выразить n через концентрацию частиц в пучке n_1 и их скорость v : $n = n_1 v$. При $n = 1 \text{ см}^{-2} \text{ с}^{-1}$ концентрация частиц в пучке равна $n_1 = \frac{1}{v} \text{ см}^{-2} \text{ с}^{-1} / (\text{см} \cdot \text{с}^{-1}) = \frac{1}{v} \text{ см}^{-3}$. Поместим в пределах этого сечения некоторый непрозрачный для частиц экран с площадкой S . Вероятность попадания частицы в экран равна отношению $S : l = S$ при условии, что частица с достоверностью проходит через единичное сечение (1 частица на 1 см^2 за 1 с). Если каждое попадание в площадку S выводит частицы из пучка, то S равно вероятности столкновения или рассеяния. Таким образом, зная площадь площадки S , мы знаем вероятность столкновения частицы с площадкой. Если на единичное сечение приходится N площадок, то вероятность столкновения увеличится в N раз и будет равна NS . Если через единичное сечение в единицу времени проходит не одна, а n частиц, то число рассеянных в единицу времени частиц $\Delta n = nNS$.

При движении частиц в среде, в которой имеются рассеивающие центры с концентрацией N , на участке пути длиной dx вдоль движения частиц для пучка единичного сечения находится $N \cdot l \cdot dx$

рассеивающих центров, вероятность частице испытать рассеяние на них равна $SNdx$. Число рассеянных в единицу времени частиц $dn = nNSdx$. Если в точке x поток частиц имеет величину $n(x)$, то в точке $x + dx$ он станет меньше на $dn = nNSdx$, так что $n(x + dx) = n(x) - dn$. Поскольку происходит уменьшение потока частиц, то $n(x + dx) < n(x)$ при $dx > 0$, и, следовательно, $dn < 0$, поэтому можно записать

$$-dn = n(x) SNdx. \quad (53.2)$$

Уравнение (53.2) легко решается, если S и N не зависят от x :

$$n(x) = n_0 e^{-SNx}. \quad (53.3)$$

Мы видим, что, зная вероятность рассеяния S , можно вычислить поток частиц в любой точке x . Число частиц в потоке, или сам поток, уменьшается по экспоненте. На расстояниях $\Delta x = \frac{1}{NS}$ число частиц уменьшается в e раз. Можно придать несколько другой смысл величине $\frac{1}{NS}$. Для этого возьмем интервал $\{x, x + dx\}$, на нем рассеивается $-dn(x)$ частиц, все они прошли путь от x до $x + dx$, или, с точностью до dx , путь x . Их суммарный путь равен

$$-dn(x)x. \quad (53.4)$$

Если просуммируем (53.4) по всем значениям x , то получим сумму длин отрезков, которые прошли все n_0 частиц без рассеяния. Разделив сумму длин всех отрезков на первоначальное число частиц в пучке, получим средний отрезок l , который проходят частицы без рассеяния:

$$\begin{aligned} l &= \frac{1}{n_0} \int_0^{\infty} x(-dn) = -\frac{n_0}{n_0} \int_0^{\infty} x e^{-SNx} SN dx = \\ &= \frac{-1}{SN} \int_0^{\infty} y e^{-y} dy = \frac{1}{SN}. \end{aligned} \quad (53.5)$$

Выражение (53.5) связывает длину свободного пробега l с эффективным сечением рассеяния. Но из (53.2) следует, что вероятность рассеяния на участке пути длиной dx равна

$$SNdx = \frac{dx}{l}. \quad (53.6)$$

Следовательно, величина $SN = l^{-1}$ есть *вероятность рассеяния на единичном интервале пути*. Это соотношение совершенно аналогично соотношению между временем свободного пробега и вероятностью рассеяния за единицу времени, полученному в расчетах § 2 на основании законов теории вероятности. Можно связать эти вели-

чины между собой. Если v есть средняя скорость по модулю, то ее можно определить отношением

$$v = \frac{l}{\tau}. \quad (53.7)$$

В таком случае

$$\tau = \frac{l}{v} = \frac{1}{vSN}. \quad (53.8)$$

Величина S была введена как площадь сечения «экрана», на котором происходит рассеяние частиц потока. Для электронов, движущихся в твердом теле, необходимо ввести некоторые более сложные понятия, описывающие процессы рассеяния. *Рассеивающим центром называют любое нарушение идеального поля решетки, приводящее к изменению квазиимпульса и скорости электрона или дырки.* Процесс взаимодействия носителей заряда с ними характеризуется некоторой вероятностью W .

Построим вокруг рассеивающего центра экран с площадью $S^* = W$. Вероятности рассеяния экраном и рассеивающим центром в этом случае равны, поэтому *вероятность рассеяния называют часто эффективным сечением рассеяния.*

Если возможно протекание нескольких различных процессов с эффективными сечениями S_i^* и числом рассеивающих центров i -сорта N_i , то полная вероятность рассеяния W является суммой отдельных вероятностей рассеяния:

$$W = S^* = \sum_i N_i S_i^*. \quad (53.9)$$

Величину W можно определить через длину свободного пробега $l = \frac{1}{W}$, в то же время $S_i N_i = \frac{1}{l_i}$ и

$$\frac{1}{l} = \sum_i \frac{1}{l_i}. \quad (53.10)$$

Другими словами, если существует несколько механизмов рассеяния и i -й механизм, действуя в «чистом» виде, обуславливает длину свободного пробега величиной l_i , то при их одновременном действии длина свободного пробега l должна быть, естественно, меньше любого из них, что подтверждается соотношением (53.10).

Оно выражает собой аддитивность вероятностей независимых событий. Из (53.10) следует, что полная длина свободного пробега всегда меньше самой малой парциальной длины свободного пробега. Чтобы оценить вклад некоторого механизма рассеяния в полную величину l , необходимо учитывать концентрацию соответствующих рассеивающих центров и эффективное сечение рассеяния на них.

Эффективное сечение S^* может быть функцией энергии, эффективной массы или каких-либо других параметров рассеиваемых

частиц. Задачей теории рассеяния является установление зависимости S^* от энергии, массы частиц для различных видов рассеивающих центров.

Эффективное сечение S^* мы определили как вероятность того, что частица рассеется, отклонится на любой произвольный угол от направления первоначального движения. Но поскольку процесс рассеяния является случайным процессом, то различные частицы будут отклоняться от направления своего движения на различные углы θ , φ . Пусть в элемент телесного угла $d\Omega = \sin\theta d\theta d\varphi$ на участке dx в единицу времени рассеется dn' частиц. Это число можно представить в виде

$$dn' = dn'(\theta, \varphi) = nN dx \sigma(\theta, \varphi) d\Omega, \quad (53.11)$$

откуда

$$\sigma(\theta, \varphi) = \frac{dn'}{nN dx d\Omega}. \quad (53.12)$$

Величина $\sigma(\theta, \varphi)$ численно равна числу частиц dn' , рассеянных за единицу времени в единичный телесный угол $d\Omega = 1$ на единичном участке пути $dx = 1$ для единичного потока $n = n_1 v = 1$ на одном рассеивающем центре $N = 1$. Величина

$$\sigma(\theta, \varphi) d\Omega = \frac{dn'}{1} \quad (n_1 v = n = 1; N dx = 1) \quad (53.13)$$

представляет собой вероятность рассеяния частицы из одночастичного потока на одном рассеивающем центре на углы θ и φ в телесный угол $d\Omega$, поэтому $\sigma(\theta, \varphi)$ называют дифференциальным эффективным сечением. Размерность $\sigma(\theta, \varphi)$ можно найти из (53.13): так как dn' есть вероятность рассеяния одной частицы в единицу времени ($[dn'] = T^{-1}$) для одночастичного потока ($n_1 v = 1 \text{ см}^{-2}\text{с}^{-1}$, т. е. $[n_1 v] = L^{-2}T^{-1}$), то

$$[\sigma(\theta, \varphi)] = \frac{T^{-1}}{L^{-2}T^{-1}} = L^2. \quad (53.14)$$

Мы видим, что дифференциальное эффективное сечение имеет размерность площади, так же как и эффективное сечение S^* . Если просуммировать величины (53.13) по всем углам θ , φ , то получим вероятность рассеяния на произвольные, любые углы, т. е. получим полную вероятность S^* :

$$\int_{(4\pi)} \sigma(\theta, \varphi) d\Omega = S^*. \quad (53.15)$$

Уравнение (53.15) представляет собой условие нормировки $\sigma(\theta, \varphi)$, величину S^* можно назвать интегральным эффективным сечением. Если дифференциальное эффективное сечение не зависит от углов θ и φ , т. е. рассеяние на любые углы равновероятно, то такое рас-

сеяние называют *изотропным*. Очевидно, что для изотропного рассеяния σ связано с S^* простым соотношением

$$S^* = \int_{(4\pi)} \sigma(\theta, \varphi) d\Omega = \sigma \int_{(4\pi)} d\Omega = 4\pi\sigma, \quad (53.16)$$

$$\sigma = \frac{S^*}{4\pi}. \quad (53.17)$$

Угол θ представляет собой угол между первоначальным направлением движения частицы и ее движением после рассеяния. Угол θ отсчитывается от некоторой плоскости, проходящей через полярную ось (ось z), совпадающую с направлением первоначального движения частицы. Если рассеивающий центр обладает осевой симметрией, то σ зависит только от полярного угла θ и определяет вероятность рассеяния в телесный угол, заключенный между двумя конусами $\theta + d\theta$ и θ (рис. 78).

Полагая, что σ не зависит от φ , и интегрируя (53.13) по φ , получим

$$\begin{aligned} \sigma(\theta) \sin \theta d\theta \int_0^{2\pi} d\varphi &= \sigma(\theta) 2\pi \sin \theta d\theta = \\ &= \sigma(\theta) d\Omega, \end{aligned} \quad (53.18)$$

где $d\Omega = 2\pi \sin \theta d\theta$ есть угол между двумя конусами.

Как известно, соударения могут быть упругими и неупругими.

Соударения называют упругими, если кинетическая энергия соударяющихся частиц сохраняется. Если же кинетическая энергия частиц после удара будет больше или меньше, чем до удара, то соударение носит название *неупругого*.

При столкновении легкой частицы с тяжелым неподвижным рассеивающим центром энергия частицы при упругом соударении практически не меняется. Проверим это утверждение, для этого запишем законы сохранения импульса и энергии:

$$\mathbf{p}_1 + \mathbf{0} = \mathbf{p}_2 + \mathbf{P}, \quad (53.19)$$

$$T_1 + 0 = \frac{p_1^2}{2m} + 0 = \frac{p_2^2}{2m} + \frac{P^2}{2M} = T_2 + T, \quad (53.20)$$

где $(0, \mathbf{P})$ и $(\mathbf{p}_1, \mathbf{p}_2)$ — импульсы тяжелой и легкой частиц до соударения и после соударения.

Выражая \mathbf{P} из (53.19) и подставляя в (53.20), получим

$$\mathbf{P} = \mathbf{p}_1 - \mathbf{p}_2; \quad P^2 = p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta. \quad (53.21)$$

Из (53.20) можно записать

$$P^2 = \frac{M}{m} (p_1^2 - p_2^2). \quad (53.22)$$

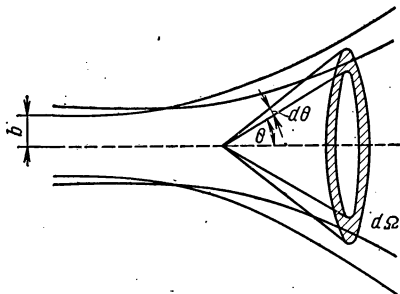


Рис. 78. Телесный конический угол рассеяния

Сравнивая (53.22) и (53.21), получим

$$p_1^2 + p_2^2 - 2p_1 p_2 \cos \theta - \frac{M}{m} p_1^2 + \frac{M}{m} p_2^2 = 0, \quad (53.23)$$

или

$$p_2^2 - \frac{2p_1 \cos \theta}{\left(1 + \frac{M}{m}\right)} p_2 + \frac{1 - \frac{M}{m}}{1 + \frac{M}{m}} p_1^2 = 0. \quad (53.24)$$

Решим уравнение (53.24)

$$\begin{aligned} p_2 &= \frac{p_1 \cos \theta}{1 + \frac{M}{m}} \pm \sqrt{\frac{p_1^2 \cos^2 \theta}{\left(1 + \frac{M}{m}\right)^2} - \frac{\left(1 - \frac{M}{m}\right)}{\left(1 + \frac{M}{m}\right)} p_1^2} = \\ &= p_1 \frac{\cos \theta \pm \sqrt{\frac{M^2}{m^2} - \sin^2 \theta}}{1 + \frac{M}{m}}. \end{aligned} \quad (53.25)$$

Для случая, когда $\frac{m}{M} \ll 1$, сохраняя в (53.25) только члены первого порядка малости по $\frac{m}{M}$ и отбрасывая знак минус перед корнем ($p_2 > 0$), получим

$$p_2 = p_1 \frac{\cos \theta + \frac{M}{m}}{1 + \frac{M}{m}} = p_1 \left[1 - \frac{m}{M} (1 - \cos \theta) \right]. \quad (53.26)$$

Мы видим, что импульс рассеянной частицы по модулю несколько меньше импульса падающей частицы:

$$p_1 - p_2 = p_1 \frac{m}{M} (1 - \cos \theta). \quad (53.27)$$

Найдем изменение энергии рассеиваемой частицы:

$$\begin{aligned} \Delta T &= T_1 - T_2 = T = \frac{1}{2m} (p_1^2 - p_2^2) = \frac{1}{2m} (p_1 + p_2) (p_1 - p_2) = \\ &= \frac{1}{2m} 2p_1 \frac{p_1 m}{M} (1 - \cos \theta) = \frac{p_1^2}{2m} \frac{2m}{M} (1 - \cos \theta) = T_1 \frac{2m}{M} (1 - \cos \theta). \end{aligned} \quad (53.28)$$

Доля энергии, передаваемой падающей частицей рассеиваемому центру, зависит от угла рассеяния: $\Delta T = T = 0$ при $\theta = 0$ (рассеяния нет). Максимальное изменение энергии падающей частицы происходит при рассеянии на угол $\theta = \pi$, когда частица отскакивает назад:

$$\Delta T = T = T_1 \frac{4m}{M}. \quad (53.29)$$

Найдем среднюю энергию, передаваемую рассеиваемой частицей центру рассеяния за одно столкновение. Для этого учтем, что вероятность рассеяния в угловой интервал $d\theta$ согласно (53.18) равна $\sigma(\theta) 2\pi \sin \theta d\theta$; в то же время это есть вероятность того, что при рассеянии частицей будет потеряна энергия $\Delta T(\theta)$. Умножая $\Delta T(\theta)$ на вероятность рассеяния и учитывая, что функция $\sigma(\theta)$ нормирована на S^* , запишем

$$\begin{aligned} \langle \Delta T \rangle &= \frac{\int_0^\pi \Delta T(\theta) \sigma(\theta) 2\pi \sin \theta d\theta}{\int_0^\pi \sigma(\theta) 2\pi \sin \theta d\theta} = T_1 \frac{2m}{M} \frac{2\pi \int_0^\pi \sigma(\theta) (1 - \cos \theta) \sin \theta d\theta}{S^*} = \\ &= T_1 \frac{2m}{M} \frac{\sigma_c}{S^*}, \end{aligned} \quad (53.30)$$

где

$$\sigma_c = 2\pi \int_0^\pi \sigma(\theta) (1 - \cos \theta) \sin \theta d\theta \quad (53.31)$$

есть *усредненное дифференциальное эффективное сечение*; усреднение по углам θ производится с весовой функцией $(1 - \cos \theta)$; σ_c называют *эффективным сечением проводимости или подвижности, или транспортным эффективным сечением*. Весовая функция уменьшает вклад от $\sigma(\theta)$ в σ_c при рассеянии на малые углы и увеличивает почти в два раза роль отклонений на большие углы. Это понятно, поскольку при рассеянии на большие углы доля передаваемой энергии велика. Рассмотрим простейший случай, когда рассеяние изотропно. Вычислим транспортное эффективное сечение при изотропном рассеянии:

$$\sigma_c = 2\pi \int_0^\pi (1 - \cos \theta) \sin \theta d\theta = 4\pi\sigma = S^*. \quad (53.32)$$

Следовательно, при изотропном рассеянии $\langle \Delta T \rangle = \frac{2m}{M} T_1$.

Величину σ_c можно интерпретировать несколько иначе. Так как при рассеянии на угол θ скорость и импульс, связанные с направленным движением, меняются, можем найти долю скорости направленного движения v_z , теряемую при соударении. Учитывая, что

$$v_{z2} \approx v_{z1} \cos \theta, \quad (53.33)$$

можем записать

$$\Delta v_z = v_1 - v_{z2} \cong v_z (1 - \cos \theta), \quad (53.34)$$

или

$$\frac{\Delta v_z}{v_z} = 1 - \cos \theta. \quad (53.35)$$

Усредняя величину Δv_z по углам θ , получим

$$\langle \Delta v_z \rangle = v_z \langle 1 - \cos \theta \rangle = \frac{v_z}{S^*} 2\pi \int_0^\pi (1 - \cos \theta) \sigma(\theta) \sin \theta d\theta = v_z \frac{\sigma_c}{S^*}. \quad (53.36)$$

Отношение $\frac{\langle \Delta v_z \rangle}{v_z}$ равно отношению $\frac{\sigma_c}{S^*}$. Если обозначить $\frac{\sigma_c}{S^*}$ через q^{-1} , то

$$\langle \Delta v_z \rangle = \frac{v_z}{q}; \quad q = \frac{v_z}{\langle \Delta v_z \rangle} = \frac{S^*}{\sigma_c}. \quad (53.37)$$

Величину q можно рассматривать как *число столкновений, в результате которых скорость направленного движения теряется полностью*. Так как для изотропного рассеяния $\sigma_c = S^*$, при изотропном рассеянии скорость направленного движения полностью теряется при одном соударении.

В заключение этого параграфа перечислим основные типы центров рассеяния:

- 1) ионы примеси,
- 2) тепловые колебания решетки,
- 3) атомы примеси,
- 4) вакансии и точечные дефекты,
- 5) дислокации,
- 6) границы кристаллитов, плоскости спайности, границы кристалла,
- 7) носители зарядов.

Очевидно, что роль указанных рассеивающих центров различна. Для ее оценки необходимо найти вероятность рассеяния на каждом из перечисленных видов рассеивающих центров. Для этого достаточно оценить эффективное сечение величиной некоторой площадки, в пределах которой возможно взаимодействие между носителем заряда и рассеивающим центром.

Рассеяние носителей заряда друг на друге не должно играть большой роли, поскольку их взаимодействие было учтено введением самосогласованного поля. Самосогласованное поле позволило рассматривать носители заряда как невзаимодействующие. Это не означает, конечно, что электроны не подчиняются, например, закону Кулона. Взаимодействие происходит на основе обычных законов электродинамики, однако в силу большого числа взаимодействующих частиц и их волновых свойств реальное движение протекает так, как если бы фактически взаимодействия не было, т. е. имеет место своеобразная *динамическая устойчивость самого движения*, что заложено в функциях Блоха и в *сохранении квазиимпульса*. Из сказанного можно сделать вывод, что рассеяние на носителях заряда не может являться *основным* механизмом, определяющим характер протекания кинетических явлений.

Рассеяние на границах кристалла, на границах кристаллитов или зерен должно играть важную роль в монокристаллических и поликристаллических пленках. В объеме *монокристалла* этот фактор несуществен.

Дислокации захватывают большие области кристалла, поэтому они должны иметь очень большое эффективное сечение. Если считать, что линейная дислокация имеет длину 1 мм и диаметр ее измеряется сотней периодов решетки, то площадь ее осевого сечения составит величину порядка $5 \cdot 10^{-8} \cdot 10^2 \cdot 10^{-1} = 5 \cdot 10^{-7} \text{ см}^2$. Если объемная плотность дислокации равна $N_D \approx 10^8 \text{ см}^{-3}$, то $l_D = S_D^* N_D^{-1} = 2 \cdot 10^{-2} \text{ см}$. Как будет видно из последующих оценок, эта величина достаточно большая, и для ряда явлений кристалл со столь большой плотностью дислокаций не может быть использован.

Для дефектов и примесных атомов за величину S^* можно принять площадь порядка площади грани элементарной ячейки, т. е. $S_a^* = 3 \cdot 10^{-15} \text{ см}^2$. Если предположить, что $N_a \approx 10^{16} \text{ см}^{-3}$, то $l_a \approx 3 \cdot 10^{-2} \text{ см}$.

Для ионов примеси S_I^* , вероятно, надо положить больше S_a^* по крайней мере на два порядка, в таком случае при $N_I \approx 10^{16} \text{ см}^{-3}$ и $S_I \approx 3 \cdot 10^{-13} \text{ см}^2$ для l_I получим $l_I \approx 3 \cdot 10^{-4} \text{ см} = 3 \text{ мкм}$. Оценивая роль тепловых колебаний решетки, необходимо иметь в виду, что под S_I^* надо понимать не площадь сечения атома основного вещества, а *площадь сечения той области, которую он занимает, совершая колебания*, т. е. надо учесть амплитуду колебаний. Очевидно, что амплитуда колебаний должна быть тем больше, чем больше температура. Если считать, что амплитуда колебаний порядка 1 \AA , то площадь сечения области, занимаемой атомом вследствие колебаний (за вычетом площади сечения самого атома), будет равна произведению диаметра атома на удвоенную амплитуду, т. е. $S_T \approx 10^{-16} \text{ см}^2$. Мы видим, что эффективное сечение рассеяния на тепловых колебаниях меньше, чем для всех предыдущих рассмотренных видов рассеивающих центров. Однако число колеблющихся атомов велико: $N_T \approx 10^{22} \text{ см}^{-3}$, и следовательно, $l_T \approx 10^{-6} \text{ см} = 100 \text{ \AA}$.

Приведенные оценки являются сугубо качественными, однако они достаточно верно передают соотношение между длинами свободного пробега при рассеянии на центрах различной природы.

Действительно, как показывает опыт, *при высоких температурах основную роль в рассеянии играют тепловые колебания решетки. С понижением температуры подвижность определяется рассеянием на ионах примеси*. Если концентрация ионов примеси мала, то основное значение имеет рассеяние на примесных атомах или дефектах. В дальнейшем сможем сделать эти выводы на основании математических выражений для времен релаксации при действии различных видов рассеивающих центров.

§ 54. СВЯЗЬ ВРЕМЕНИ РЕЛАКСАЦИИ С ЭФФЕКТИВНЫМ СЕЧЕНИЕМ

Все кинетические явления связаны с усредненными по энергиям временами релаксации. Время релаксации определяется через интеграл столкновений, величина которого зависит от вероятностей переходов электронов и дырок из одного состояния в другое под действием соударений. Поэтому естественно предположить, что должна быть связь между эффективным сечением рассеяния и временем релаксации.

В квантовой механике и физической кинетике доказывается, что *элементарные вероятности прямых и обратных переходов равны*. Это значит, что $\omega(\mathbf{k}, \mathbf{k}') = \omega(\mathbf{k}', \mathbf{k})$. Учитывая это равенство, запишем выражение для времени релаксации:

$$\left(\frac{\partial f}{\partial t}\right)_c = -\frac{f-f_0}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \{ \omega(\mathbf{k}', \mathbf{k}) f(\mathbf{k}') [1-f(\mathbf{k})] - \omega(\mathbf{k}, \mathbf{k}') f(\mathbf{k}) [1-f(\mathbf{k}')] \} d\tau_{\mathbf{k}'}. \quad (54.1)$$

Учитывая, что равновесная функция $f=f_0$ обращает интеграл столкновений в нуль, полагая $f=f_0+f^{(1)}$ и пренебрегая членами вида $f^{(1)2}$, получим

$$\frac{1}{\tau(\mathbf{k})} = -\frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \left\{ \omega(\mathbf{k}', \mathbf{k}) \left[\frac{f^{(1)}(\mathbf{k}')}{f^{(1)}(\mathbf{k})} [1-f_0(\mathbf{k})] - f_0(\mathbf{k}') \right] - \omega(\mathbf{k}, \mathbf{k}') \left[1-f_0(\mathbf{k}') - \frac{f^{(1)}(\mathbf{k}')}{f^{(1)}(\mathbf{k})} f_0(\mathbf{k}) \right] \right\} d\tau_{\mathbf{k}'}, \quad (54.2)$$

что совпадает с выражением (36.6р).

Учитывая, что $\omega(\mathbf{k}, \mathbf{k}') = \omega(\mathbf{k}', \mathbf{k})$, сокращая одинаковые члены в (54.2), выражение для $\tau(\mathbf{k})$ представим в виде

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int \omega(\mathbf{k}, \mathbf{k}') \left[1 - \frac{f^{(1)}(\mathbf{k}')}{f^{(1)}(\mathbf{k})} \right] d\tau_{\mathbf{k}'}. \quad (54.3)$$

Согласно (37.16), $f^{(1)}$ имеет вид

$$f^{(1)}(\mathbf{k}) = -\left. \frac{\partial f_0}{\partial E} \right|_{\mathbf{k}} (\boldsymbol{\kappa}(\mathbf{k}) \mathbf{v}(\mathbf{k})); \quad f^{(1)}(\mathbf{k}') = -\left. \frac{\partial f_0}{\partial E} \right|_{\mathbf{k}'} (\boldsymbol{\kappa}(\mathbf{k}') \mathbf{v}(\mathbf{k}')). \quad (54.4)$$

Если подставить вместо $f^{(1)}(\mathbf{k})$ и $f^{(1)}(\mathbf{k}')$ их выражение (54.4), то получим уравнение для $\tau(\mathbf{k})$:

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') \left[1 - \frac{\left. \frac{\partial f_0}{\partial E} \right|_{\mathbf{k}'} (\boldsymbol{\kappa}(\mathbf{k}') \mathbf{v}(\mathbf{k}'))}{\left. \frac{\partial f_0}{\partial E} \right|_{\mathbf{k}} (\boldsymbol{\kappa}(\mathbf{k}) \mathbf{v}(\mathbf{k}))} \right] d\tau_{\mathbf{k}'}. \quad (54.5)$$

Величина $\boldsymbol{\kappa}$ должна определяться по уравнениям (37.52), из которых видно, что $\boldsymbol{\kappa}$ в свою очередь зависит сложным образом от времени релаксации, поэтому уравнение (54.5) является нелинейным инте-

гральным уравнением, решить которое в общем виде не представляется возможным. Введение времени релаксации позволило получить решение кинетического уравнения, однако трудность, существовавшая при решении уравнения Больцмана, введением времени релаксации не ликвидирована, а только перенесена с уравнения для $f^{(1)}$ на τ . Кроме того, решая уравнение для $f^{(1)}$, мы считали, что время релаксации не зависит от внешних полей. Точнее, можно сказать, что процессы рассеяния при наложении полей \mathbf{E} , \mathbf{B} и без них протекают иначе. Дальше увидим, что зависимость процессов соударений от внешних полей выступает в виде определенных физических явлений. Однако поскольку мы считали, что $\tau(\mathbf{k})$ не зависит от внешних полей, то формально можно рассмотреть уравнение (54.5) для случая $\mathbf{E} \rightarrow 0$; $\mathbf{B} \rightarrow 0$; тогда

$$-\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') \left[1 - \frac{\frac{\partial f_0}{\partial E} \Big|_{\mathbf{k}'} (v(\mathbf{k}') E^0) \tau(\mathbf{k}')}{\frac{\partial f_0}{\partial E} \Big|_{\mathbf{k}} (v(\mathbf{k}) E^0) \tau(\mathbf{k})} \right] d\tau_{\mathbf{k}'}, \quad (54.6)$$

где E^0 — единичный вектор, вообще говоря, произвольного направления.

Для большинства практически важных случаев выражение (54.6) существенно упрощается. Мы предполагали ранее, что τ зависит только от энергии. Направим ось z вдоль вектора \mathbf{k} , тогда \mathbf{k}' определяет направление движения частицы после рассеяния. Выбирая E^0 вдоль оси z , получим

$$(v(\mathbf{k}) E^0) = v(\mathbf{k}); \quad (v(\mathbf{k}') E^0) = v(\mathbf{k}') \cos(\mathbf{k}', \mathbf{k}) = v(\mathbf{k}') \cos \theta \quad (54.7)$$

и, следовательно,

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}'})} \omega(\mathbf{k}, \mathbf{k}') \left[1 - \frac{\frac{\partial f_0}{\partial E} \Big|_{\mathbf{k}'} \tau(E') v(\mathbf{k}') \cos(\mathbf{k}', \mathbf{k})}{\frac{\partial f_0}{\partial E} \Big|_{\mathbf{k}} \tau(E) v(\mathbf{k})} \right] d\tau_{\mathbf{k}'}. \quad (54.8)$$

Рассмотрим простейший случай — сферические зоны энергии. Как сказано в § 53, при упругом столкновении частицы с рассеивающим центром ее энергия E меняется очень незначительно, не более чем на $\frac{4m}{M} E$. Следовательно, можно считать, что при упругом рассеянии состояние частицы в зоне Бриллюэна остается на одной и той же изоэнергетической поверхности. При этих предположениях выражение (54.6) может быть записано в виде

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{V_{\mathbf{k}}} \omega(\mathbf{k}, \mathbf{k}') [1 - \cos(\mathbf{k}', \mathbf{k})] d\tau_{\mathbf{k}'}, \quad (54.9)$$

так как $E \approx E'$, то модуль скорости сохраняется

$$\tau(E') \approx \tau(E) \quad \text{и} \quad v(\mathbf{k}') \approx v(\mathbf{k}).$$

В сферической системе координат с полярной осью, направленной вдоль вектора κ , $\cos(\kappa', \kappa) = \cos \theta$ и

$$\begin{aligned} \frac{1}{\tau(\kappa)} &= \frac{1}{4\pi^3} \int_{(V_\kappa)} \omega(\kappa, \kappa') [1 - \cos \theta] d\tau_{\kappa'} = \\ &= \frac{1}{4\pi^3} \int \omega(\kappa, \kappa', \theta, \varphi) (1 - \cos \theta) \kappa'^2 d\Omega_{\kappa'} \end{aligned} \quad (54.10)$$

Представим $\omega(\kappa, \kappa', \theta, \varphi)$ в виде двух функций: радиальной $\tilde{\omega}(\kappa, \kappa')$ и угловой $\tilde{\omega}(\theta, \varphi)$. Пусть $\tilde{\omega}(\theta, \varphi)$ есть *вероятность рассеяния одной частицы в единичный телесный угол в направлении (θ, φ)* . Дифференциальное эффективное сечение представляет собой вероятность рассеяния одной частицы на одном рассеивающем центре из единичного потока, следовательно,

$$\tilde{\omega}(\theta, \varphi) = Nv(\kappa) \sigma(\theta, \varphi), \quad (54.11)$$

так как одной частице, имеющей скорость $v(\kappa)$, соответствует поток $1 \cdot v(\kappa)$.

Пусть $\tilde{\omega}(\kappa, \kappa')$ определяет *вероятность перехода частицы из единичного объема $1(\kappa)$ в шаровой слой радиуса κ' единичной толщины: $d\kappa' = 1$* .

Рассмотрим произведение $\tilde{\omega}(\kappa, \kappa') \tilde{\omega}(\theta, \varphi)$; оно определяет *вероятность перехода в объем, выделяемый единичным телесным углом из шарового слоя единичной толщины, площадь его основания равна, очевидно, κ'^2* . Если мы разделим $\tilde{\omega}(\kappa, \kappa') \tilde{\omega}(\theta, \varphi)$ на κ'^2 , то получим вероятность перехода в слой единичной толщины, опирающийся на единичную площадь, т. е. в единичный объем с центром в точке κ' , поэтому можем записать

$$\omega(\kappa, \kappa') = \frac{\tilde{\omega}(\kappa, \kappa') \tilde{\omega}(\theta, \varphi)}{\kappa'^2} = \frac{\tilde{\omega}(\kappa, \kappa') Nv(\kappa) \sigma(\theta, \varphi)}{\kappa'^2}. \quad (54.12)$$

Выразим время релаксации через эффективное сечение:

$$\frac{1}{\tau(\kappa)} = Nv(\kappa) \int_{(4\pi)} \sigma(\theta, \varphi) (1 - \cos \theta) d\Omega \frac{1}{4\pi^3} \int_0^\infty \tilde{\omega}(\kappa, \kappa') d\kappa'. \quad (54.13)$$

Если рассеяние упругое, то состояние рассеиваемой частицы остается на изоэнергетической поверхности, поэтому $\tilde{\omega}(\kappa, \kappa')$ должно быть пропорционально $\delta(\kappa - \kappa')$. В качестве коэффициента пропорциональности возьмем $4\pi^3$:

$$\tilde{\omega}(\kappa, \kappa') = 4\pi^3 \delta(\kappa - \kappa') \quad (54.14)$$

$$\frac{1}{4\pi^3} \int_0^\infty 4\pi^3 \delta(\kappa - \kappa') d\kappa' = 1, \quad (54.15)$$

поскольку рассеяние во все состояния κ' есть достоверное событие.

Учитывая (54.15), запишем окончательно

$$\frac{1}{\tau(\kappa)} = Nv(\kappa) \int_{(4\pi)} \sigma(\theta, \varphi) (1 - \cos \theta) d\Omega = Nv(\kappa) \sigma_c, \quad (54.16)$$

или

$$\tau(\kappa) = \frac{1}{Nv(\kappa) \sigma_c}; \quad l(\kappa) = \tau(\kappa) v(\kappa) = \frac{1}{N\sigma_c}, \quad (54.17)$$

что аналогично (53.8).

Резюме § 53, 54

1. Дифференциальное эффективное сечение рассеяния $\sigma(\theta, \varphi)$ есть вероятность того, что одна частица, имеющая скорость, равную единице, при соударении с одним рассеивающим центром попадает в единичный телесный угол $d\Omega = 1$, построенный вокруг направления, определяемого углами θ, φ при совмещении начала отсчета с рассеивающим центром. Зная $\sigma(\theta, \varphi)$, можем найти число частиц $dn'(\theta, \varphi, x)$, попадающих в телесный угол $d\Omega(\theta, \varphi)$ при рассеянии частиц, имеющих концентрацию в пучке n_1 и скорость v при рассеянии на центрах с концентрацией N , при прохождении частицами отрезка пути dx :

$$-dn'(\theta, \varphi, x) = \sigma(\theta, \varphi) n_1(x) vN dx d\Omega(\theta, \varphi). \quad (53.1p)$$

2. Интегральным эффективным сечением рассеяния S^* называется вероятность рассеяния на произвольные углы, S^* связано с $\sigma(\theta, \varphi)$ условием нормировки:

$$S^* = \int_{(4\pi)} \sigma(\theta, \varphi) d\Omega. \quad (53.2p)$$

3. Транспортным эффективным сечением, или эффективным сечением подвижности или проводимости, называют величину σ_c , получаемую из дифференциального эффективного сечения $\sigma(\theta, \varphi)$ интегрированием по всем углам с весовой функцией $(1 - \cos \theta)$:

$$\sigma_c = \int_0^\pi \int_0^{2\pi} \sigma(\theta, \varphi) (1 - \cos \theta) \sin \theta d\theta d\varphi. \quad (53.3p)$$

Если дифференциальное эффективное сечение не зависит от углов θ и φ , то

$$S^* = 4\pi\sigma = \sigma_c. \quad (53.4p)$$

В общем случае отношение S^* к σ_c определяет среднее число соударений q , при которых полностью теряется скорость направленного движения:

$$q = \frac{v_z}{\langle \Delta v_z \rangle} = \frac{S^*}{\sigma_c}. \quad (53.5p)$$

При изотропном рассеянии скорость направленного движения теряется в среднем при одном соударении.

4. Длина свободного пробега l_i направленного движения связана с σ_{ci} и концентрацией рассеивающих центров N_i :

$$l_i = \frac{1}{N_i \sigma_{ci}}; \quad (53.6p)$$

$$l^{-1} = \sum_i N_i \sigma_{ci} = \sum_i l_i^{-1}, \quad (53.7p)$$

5. Время релаксации $\tau(\mathbf{k})$ связано с σ_c :

$$\frac{1}{\tau(\mathbf{k})} = N v(\mathbf{k}) \sigma_c. \quad (53.8p)$$

6. Изменение энергии частицы с массой m^* при рассеянии на центре с массой M равно в среднем

$$\langle \Delta E \rangle = 2 \frac{m^*}{M} E \frac{\sigma_c}{S^*}. \quad (53.9p)$$

7. Рассеяние электронов и дырок может происходить: 1) на ионах примеси; 2) на атомах примеси; 3) на вакансиях и точечных дефектах; 4) на дислокациях; 5) на границах кристалла, плоскостях спайности, границах зерен; 6) на электронах и дырках; 7) на тепловых колебаниях решетки.

8. Задачей теории рассеяния является вычисление эффективного сечения рассеяния на центрах различной природы, что позволяет по σ_c вычислить $\tau(\mathbf{k})$ и $\langle E \tau^s \rangle$ и тем самым дать описание кинетических явлений, зависящих от механизма рассеяния.

§ 55. ПОНЯТИЕ О ТЕОРИИ КВАНТОВЫХ ПЕРЕХОДОВ

Как было указано в предыдущем параграфе, для вычисления кинетических коэффициентов необходимо знать дифференциальное эффективное сечение, которое может быть вычислено методами квантовой механики.

В связи с тем что некоторые результаты теории квантовых переходов необходимы для понимания не только теории рассеяния, но и теории оптических явлений, напомним ее основные положения.

Пусть некоторая квантовая система, описываемая гамильтонианом \hat{H}^0 , имеет определенную совокупность стационарных состояний, характеризуемых энергией $E^0(\alpha)$ и волновыми функциями $\psi_\alpha^0(\mathbf{r})$, удовлетворяющими уравнению

$$\hat{H}^0 \psi_\alpha^0(\mathbf{r}) = E^0(\alpha) \psi_\alpha^0(\mathbf{r}) \quad (55.1)$$

(α — полный набор физических величин). Волновая функция $\psi_\alpha^0(\mathbf{r})$

является решением *стационарного* уравнения Шредингера. Решение *временного* уравнения Шредингера

$$i\hbar \frac{\partial \psi_{\alpha}^0(\mathbf{r}, t)}{\partial t} = \hat{H}^0 \psi_{\alpha}^0(\mathbf{r}, t) \quad (55.2)$$

связано с решением стационарного уравнения простым соотношением

$$\psi_{\alpha}^0(\mathbf{r}, t) = \psi_{\alpha}^0(\mathbf{r}) e^{-i \frac{E^0(\alpha)}{\hbar} t}, \quad (55.3)$$

что легко проверяется непосредственной подстановкой (55.3) в (55.2), откуда вытекает (55.1).

Пусть при $t=0$ на квантовую систему наложено некоторое поле $\hat{W} = \hat{W}(\mathbf{r}, t)$. Гамильтониан системы имеет теперь вид

$$\hat{H} = \hat{H}^0 + \hat{W}. \quad (55.4)$$

Волновая функция возмущенной системы $\psi(\mathbf{r}, t)$ определяется из уравнения

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi = (\hat{H}^0 + \hat{W}) \psi. \quad (55.5)$$

Для решения этого уравнения согласно Дираку разложим $\psi(\mathbf{r}, t)$ о функциям $\psi_{\alpha}^0(\mathbf{r}, t)$,

$$\psi(\mathbf{r}, t) = \int c_{\alpha}(t) \psi_{\alpha}^0(\mathbf{r}, t) d\alpha. \quad (55.6)$$

Мы записали разложение $\psi(\mathbf{r}, t)$ по $\psi_{\alpha}^0(\mathbf{r}, t)$, считая, что α имеет непрерывный спектр, в противном случае необходимо заменить интеграл суммой по всем значениям α . Подставим (55.6) в (55.5):

$$\begin{aligned} i\hbar \int \left\{ \frac{dc_{\alpha}(t)}{dt} \psi_{\alpha}^0(\mathbf{r}, t) + c_{\alpha}(t) \frac{\partial \psi_{\alpha}^0(\mathbf{r}, t)}{\partial t} \right\} d\alpha = \\ = \int c_{\alpha}(t) \left\{ \hat{H}^0 \psi_{\alpha}^0(\mathbf{r}, t) + \hat{W} \psi_{\alpha}^0(\mathbf{r}, t) \right\} d\alpha. \end{aligned} \quad (55.7)$$

Учитывая (55.2), получим

$$i\hbar \int \frac{dc_{\alpha}(t)}{dt} \psi_{\alpha}^0(\mathbf{r}, t) d\alpha = \int c_{\alpha}(t) \hat{W} \psi_{\alpha}^0(\mathbf{r}, t) d\alpha. \quad (55.8)$$

Умножим (55.8) слева на $\psi_{\beta}^{0*}(\mathbf{r}, t)$ и проинтегрируем по всему пространству. Условие нормировки функций $\psi_{\alpha}^0(\mathbf{r})$

$$\int \psi_{\beta}^{0*}(\mathbf{r}) \psi_{\alpha}^0(\mathbf{r}) d\tau = \delta(\beta - \alpha) \quad (55.9)$$

позволяет записать

$$i\hbar \frac{dc_{\beta}}{dt} = \int c_{\alpha}(t) e^{\frac{i}{\hbar} [E^0(\beta) - E^0(\alpha)] t} W_{\beta\alpha} d\alpha. \quad (55.10)$$

Уравнение (55.10) — это уравнение (55.5) в матричной форме, оно является точным уравнением. Решение его в общей форме крайне

затруднительно, поэтому оно решается обычными методами теории возмущений. Пусть \hat{W} мало, тогда его можно считать возмущением. За меру малости величин принимается малость матричного элемента

$$W_{\beta\alpha} = \int \psi_{\beta}^{0*}(\mathbf{r}) \hat{W}(\mathbf{r}, t) \psi_{\alpha}^0(\mathbf{r}) d\tau = W_{\beta\alpha}(t). \quad (55.11)$$

Будем искать $c_{\alpha}(t)$ в виде ряда

$$c_{\alpha}(t) = c_{\alpha}^0(t) + c_{\alpha}^{(1)}(t) + c_{\alpha}^{(2)}(t) + \dots \quad (55.12)$$

$c_{\alpha}^0(t)$ есть нулевое приближение, соответствующее $\hat{W} = 0$. В нулевом приближении, когда $\hat{W} = 0$, правая часть (55.10) обращается в нуль, следовательно, в нулевом приближении $c_{\alpha}^0(t)$ являются константами — в полном соответствии с тем, что при $\hat{W} = 0$ система находится в стационарном состоянии. Рассмотрим состояние, которое соответствует энергии $E^0(\gamma)$. Для него все $c_{\alpha}^0 = 0$ при $\alpha \neq \gamma$. Так как вероятность нахождения системы во всех состояниях α равна 1, то при $c_{\alpha} = 0$ ($\alpha \neq \gamma$) необходимо положить $c_{\gamma} = \infty$, т. е.

$$c_{\alpha}^0 = \delta(\alpha - \gamma). \quad (55.13)$$

Чтобы получить уравнение для $c_{\beta}^{(1)}(t)$, необходимо взять нулевое приближение для $c_{\alpha}(t)$, поскольку в уравнение (55.10) входит $W_{\beta\alpha}$:

$$\begin{aligned} i\hbar \frac{dc_{\beta}^{(1)}(t)}{dt} &= \int \delta(\alpha - \gamma) e^{\frac{i}{\hbar} [E^0(\beta) - E^0(\alpha)] t} W_{\beta\alpha}(t) d\alpha = \\ &= e^{\frac{i}{\hbar} [E^0(\beta) - E^0(\gamma)] t} W_{\beta\gamma}. \end{aligned} \quad (55.14)$$

Уравнение (55.14) интегрируется элементарно:

$$c_{\beta}^{(1)}(t) = \frac{1}{i\hbar} \int_0^t W_{\beta\gamma}(t) e^{i\omega_{\beta\gamma} t} dt, \quad (55.15)$$

где

$$\omega_{\beta\gamma} = \frac{E^0(\beta) - E^0(\gamma)}{\hbar}.$$

Величина $|c_{\beta}^{(1)}(t)|^2$ есть вероятность нахождения квантовой системы в состоянии $\psi_{\beta}^0(\mathbf{r}, t)$ в момент времени t . В момент времени $t=0$ система находилась в состоянии $\psi_{\gamma}^0(\mathbf{r}, 0)$. Используя выражение (55.15), получим для вероятности $|c_{\beta}^{(1)}(t)|^2$:

$$|c_{\beta}^{(1)}(t)|^2 = \frac{1}{\hbar^2} \left| \int_0^t W_{\beta\gamma}(t) e^{i\omega_{\beta\gamma} t} dt \right|^2. \quad (55.16)$$

Эту величину можно рассматривать как вероятность перехода системы из состояния $\psi_{\gamma}^0(\mathbf{r}, t)$ в состояние $\psi_{\beta}^0(\mathbf{r}, t)$ под действием возмущения \hat{W} в течение времени t .

Вероятность перехода в единицу времени, которую обозначим как $\omega(\gamma, \beta)$, можно определить соотношением

$$\omega(\gamma, \beta) = \frac{d}{dt} |c_{\beta}^{\alpha}(t)|^2 = \frac{d}{dt} \frac{1}{\hbar^2} \left| \int_0^t W_{\beta\gamma} e^{i\omega_{\beta\gamma}t} dt \right|^2. \quad (55.17)$$

Рассмотрим возмущение, не зависящее от времени. Вынося матричный элемент возмущения из-под знака интеграла и интегрируя по времени, получим

$$\begin{aligned} \omega(\gamma, \beta) &= \frac{|W_{\beta\gamma}|^2}{\hbar^2 \omega_{\beta\gamma}^2} \frac{d}{dt} |e^{i\omega_{\beta\gamma}t} - 1|^2 = \frac{|W_{\beta\gamma}|^2}{\hbar^2 \omega_{\beta\gamma}^2} \frac{d}{dt} 2(1 - \cos \omega_{\beta\gamma}t) = \\ &= \frac{2|W_{\beta\gamma}|^2}{\hbar^2 \omega_{\beta\gamma}^2} \omega_{\beta\gamma} \sin \omega_{\beta\gamma}t = \frac{2}{\hbar^2} |W_{\beta\gamma}|^2 \frac{\sin \omega_{\beta\gamma}t}{\omega_{\beta\gamma}}. \end{aligned} \quad (55.18)$$

Как показывает выражение (55.18), вероятность перехода из состояния γ в состояние β имеет осциллирующий характер. Частота осцилляций зависит от разности энергий. Если рассматривать $\omega(\gamma, \beta)$ как функцию $\omega_{\beta\gamma}$, то мы видим, что эта функция для каждого момента времени монотонно уменьшается с ростом частоты. Другими словами, если рассматривать вероятность перехода как функцию разности энергий между исходным и конечным состояниями, то вероятность с ростом указанной разности падает.

Рассмотрим величину $I = \frac{\sin \omega_{\beta\gamma}t}{\omega_{\beta\gamma}} = t \frac{\sin x}{x}$ как функцию частоты. Если $\omega_{\beta\gamma} \rightarrow 0$, то

$$\lim_{\omega_{\beta\gamma} \rightarrow 0} I = \lim_{\omega_{\beta\gamma} \rightarrow 0} \frac{\sin \omega_{\beta\gamma}t}{\omega_{\beta\gamma}} = t. \quad (55.19)$$

График функции I приведен на рис. 79.

Чем больше время t , тем больше величина максимума и меньше интервал частот, охватываемый первым максимумом:

$$\Delta\omega_{\beta\gamma} = \frac{2\pi}{t}. \quad (55.20)$$

При $t \rightarrow \infty$ интервал частот основного максимума $\Delta\omega_{\beta\gamma} \rightarrow 0$ и $I \rightarrow \infty$.

Рассмотрим интеграл от I по частотам:

$$\int_{-\infty}^{\infty} I d\omega_{\beta\gamma} = \int_{-\infty}^{\infty} \frac{\sin \omega_{\beta\gamma}t}{\omega_{\beta\gamma}} d\omega_{\beta\gamma} = \pi. \quad (55.21)$$

Таким образом, при больших интервалах t

$$\frac{1}{\pi} \frac{\sin \omega_{\beta\gamma}t}{\omega_{\beta\gamma}} \approx \delta(\omega_{\beta\gamma}). \quad (55.22)$$

Равенство (55.22) справедливо, если время действия возмущения достаточно велико; в этом случае можем записать

$$\omega(\gamma, \beta) = \frac{2\pi}{\hbar^2} |W_{\beta\gamma}|^2 \delta(\omega_{\beta\gamma}) = \frac{2\pi}{\hbar} |W_{\beta\gamma}|^2 \delta[E^0(\beta) - E^0(\gamma)]. \quad (55.23)$$

Выражение (55.23) показывает, что вероятность перехода пропорциональна квадрату модуля соответствующего матричного элемента возмущения. Она отлична от нуля, если $E^0(\beta) = E^0(\gamma)$, т. е. если энергия системы сохраняется.

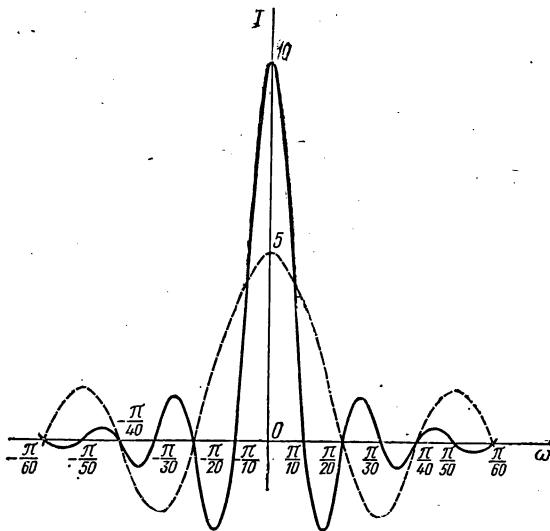


Рис. 79. График $I = \frac{\sin \omega t}{\omega}$ как функции частоты

Используем выражение (55.23) для вероятности перехода из одного состояния в другое под действием возмущения при рассеянии частиц на некотором силовом центре, который описывается потенциальной энергией взаимодействия частицы с центром $V(\mathbf{r}) = \hat{W}(\mathbf{r})$. Для вычисления матричного элемента оператора возмущения необходимо взять функцию Блоха для начального и конечного состояний, получим

$$W_{\mathbf{k}'\mathbf{k}} = \int \psi_{\mathbf{k}'}^0(\mathbf{r}) \hat{W}(\mathbf{r}) \psi_{\mathbf{k}}^0(\mathbf{r}) d\tau = \int e^{-i(\mathbf{k}' - \mathbf{k}, \mathbf{r})} V(\mathbf{r}) \varphi_{\mathbf{k}'}^*(\mathbf{r}) \varphi_{\mathbf{k}}(\mathbf{r}) d\tau. \quad (55.24)$$

В приближении эффективной массы волна Блоха может быть заменена волной де-Бройля, т. е. периодические функции $\varphi_{\mathbf{k}'}(\mathbf{r})$ и $\varphi_{\mathbf{k}}(\mathbf{r})$ заменяются на постоянные величины. Выберем амплитуды падающей и рассеянной волны из следующих соображений. Рассеянная волна $\varphi_{\mathbf{k}'}(\mathbf{r})$ описывает нахождение частицы во всем пространстве, поэтому

она должна быть нормирована на δ -функцию, вследствие чего запишем ее в виде

$$\psi_{k'}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} e^{i(\mathbf{k}'\mathbf{r})}. \quad (55.25)$$

Падающую волну пронормируем таким образом, чтобы она описывала поток вероятности единичной величины. Так как плотность потока вероятности определяется величиной

$$\mathbf{j}_w = \frac{i\hbar}{2m^*} (\psi \nabla \psi^* - \psi^* \nabla \psi), \quad (55.26)$$

полагая $\psi_k(\mathbf{r}) = A e^{i(\mathbf{k}\mathbf{r})}$, получим

$$\mathbf{j}_w = \frac{i\hbar}{2m^*} [A^* (-i\mathbf{k}) A - A^* (i\mathbf{k}) A] = \frac{|A|^2}{m^*} \hbar \mathbf{k} = |A|^2 \mathbf{v}. \quad (55.27)$$

Таким образом, единичной плотности потоку вероятности соответствует условие

$$|A|^2 v = 1; \quad A = \sqrt{\frac{1}{v}} = \sqrt{\frac{m^*}{\hbar k}} \quad (55.28)$$

или

$$\psi_k(\mathbf{r}) = \left(\frac{m^*}{\hbar k}\right)^{1/2} e^{i(\mathbf{k}\mathbf{r})}. \quad (55.29)$$

Подставляя $\psi_k(\mathbf{r})$ и $\psi_{k'}(\mathbf{r})$ из (55.29) и (55.25) в (55.24), получим для матричного элемента оператора возмущения

$$\begin{aligned} W_{k'k} &= \int \frac{1}{(2\pi)^3} e^{-i(\mathbf{k}'\mathbf{r})} V(\mathbf{r}) \left(\frac{m^*}{\hbar k}\right)^{1/2} e^{i(\mathbf{k}\mathbf{r})} d\tau = \\ &= \frac{m^*}{(2\pi)^2} \frac{1}{(\hbar k)^{1/2}} \int e^{i(\mathbf{k}\mathbf{r})} V(\mathbf{r}) d\tau, \end{aligned} \quad (55.30)$$

где

$$\mathbf{k} = \mathbf{k} - \mathbf{k}'.$$

Обозначим через $V_{\mathbf{k}}$ интеграл, входящий в матричный элемент (55.30),

$$V_{\mathbf{k}} = \int e^{i(\mathbf{k}\mathbf{r})} V(\mathbf{r}) d\tau. \quad (55.31)$$

Рассмотрим случай сферически симметричного возмущения: $V(\mathbf{r}) = V(r)$. Для вычисления (55.31) удобно перейти к сферической системе координат, направив полярную ось вдоль вектора \mathbf{k} :

$$\begin{aligned} V_{\mathbf{k}} &= \int_0^\infty \int_0^\pi \int_0^{2\pi} e^{i\mathbf{k}r \cos \theta'} V(r) r^2 dr \sin \theta' d\theta' d\varphi = \\ &= 2\pi \int_0^\infty r^2 V(r) dr \int_0^\pi e^{i\mathbf{k}r \cos \theta'} \sin \theta' d\theta'. \end{aligned} \quad (55.32)$$

Интеграл по θ' вычисляется элементарно:

$$\int_0^\pi e^{i\kappa r \cos \theta'} \sin \theta' d\theta' = -\frac{1}{i\kappa r} e^{i\kappa r \cos \theta'} \Big|_0^\pi = \frac{2 \sin \kappa r}{\kappa r}. \quad (55.33)$$

Подставив (55.33) в (55.30), будем иметь

$$V_\kappa = 4\pi \int_0^\infty r V(r) \frac{\sin \kappa r}{\kappa} dr. \quad (55.34)$$

Запишем выражение (55.23) для вероятности перехода, используя (55.34):

$$\omega(\kappa, \kappa') = \frac{2\pi}{\hbar} \frac{m^*}{(2\pi)^3 \hbar \kappa} |V_\kappa|^2 \delta[E(\kappa) - E(\kappa')]. \quad (55.35)$$

Если проинтегрировать (55.35) по всем состояниям κ' , то получим вероятность перехода $\omega(\kappa)$ из данного состояния κ во все возможные состояния для одной частицы, создающей поток единичной плотности, следовательно, $\omega(\kappa)$ тождественно интегральному эффективному сечению:

$$\begin{aligned} \omega(\kappa) &\equiv S^* = \int \omega(\kappa, \kappa') d\tau_{\kappa'} = \\ &= \frac{m^*}{4\pi^2 \hbar^2 \kappa} \int |V_\kappa|^2 \delta[E(\kappa) - E(\kappa')] d\tau_{\kappa'}. \end{aligned} \quad (55.36)$$

Интеграл по κ' вычисляется элементарно, если учесть, что

$$\begin{aligned} \delta\left[\frac{\hbar^2 \kappa^2}{2m^*} - \frac{\hbar^2 \kappa'^2}{2m^*}\right] &= \frac{2m^*}{\hbar^2} \delta(\kappa^2 - \kappa'^2); \\ d\tau_{\kappa'} &= d\Omega \frac{\kappa'}{2} d(\kappa'^2). \end{aligned} \quad (55.37)$$

Интегрирование по κ' сводится к замене κ' на κ :

$$S^* = \frac{m^* \kappa}{4\pi^2 \hbar^2 \kappa} \int_{(4\pi)} |V_\kappa|^2_{\kappa=\kappa'} d\Omega. \quad (55.38)$$

С другой стороны,

$$S^* = \int \sigma(\theta, \varphi) d\Omega. \quad (55.39)$$

Сравнивая (55.39) с (55.38), запишем

$$\begin{aligned} \sigma(\theta, \varphi) &= \frac{m^* \kappa}{4\pi^2 \hbar^2} \left| 4\pi \int_0^\infty r V(r) \frac{\sin \kappa r}{\kappa} dr \right|^2 = \\ &= \left| \frac{2m^*}{\hbar^2} \int_0^\infty r V(r) \frac{\sin \kappa r}{\kappa} dr \right|^2. \end{aligned} \quad (55.40)$$

Таким образом, найдя V_κ , мы можем выразить дифференциальное эффективное сечение через матричный элемент возмущения.

Резюме § 55

1. Возмущение, наложенное на квантовую систему, находившуюся в некотором стационарном состоянии, вызывает переход из этого состояния в другие стационарные состояния. Вероятность перехода за единицу времени из состояния α в состояние β равна

$$w(\alpha, \beta) = \frac{d}{dt} \frac{1}{\hbar^2} \left| \int_0^t W_{\beta\alpha}(t) e^{-i \frac{E(\beta) - E(\alpha)}{\hbar} t} dt \right|^2. \quad (55.1p)$$

Если возмущение не зависит от времени, то

$$w(\alpha, \beta) = \frac{2\pi}{\hbar} |W_{\beta\alpha}|^2 \delta[E(\beta) - E(\alpha)]. \quad (55.2p)$$

Матричный элемент $W_{\beta\alpha}$ вычисляется с помощью волновых функций стационарных состояний $\psi_\beta^0(\mathbf{r})$ и $\psi_\alpha^0(\mathbf{r})$.

2. Для сферически симметричного возмущения вероятность перехода (55.2p) позволяет выразить дифференциальное эффективное сечение рассеяния:

$$\sigma(\theta, \varphi) = \left| \frac{2m^*}{\hbar^2} \int_0^\infty r V(r) \frac{\sin \kappa r}{\kappa} dr \right|^2, \quad (55.3p)$$

где $V(r)$ характеризует энергию электрона в поле сферически симметричного рассеивающего центра, а $\kappa = |\mathbf{k} - \mathbf{k}'|$. Так как согласно (55.37) $\kappa = \kappa'$, то

$$\kappa = 2\kappa \sin \frac{\theta}{2}, \quad (55.4p)$$

где θ — угол рассеяния (рис. 80).

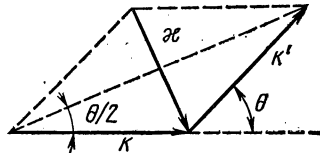


Рис. 80. Связь между κ и углом рассеяния θ

§ 56. РАССЕЯНИЕ НА ИОНАХ ПРИМЕСИ

Применим общую теорию квантовых переходов, изложенную в предыдущем параграфе, для определения эффективного сечения рассеяния электронов и дырок на ионах примеси. Для этого необходимо записать энергию возмущения в виде

$$V(r) = \pm \frac{Ze^2}{\epsilon r}. \quad (56.1)$$

Плюс относится к случаю, когда заряды рассеиваемой частицы и иона одинаковые, минус означает, что носитель заряда и ион имеют различные знаки. Заряд иона обозначен через Ze , расстояние между ионом и частицей равно r . Чтобы обобщить результаты, рассмотрим рассеяние частиц на центрах с экранирующим потенциалом:

$$V(r) = \pm \frac{Ze^2}{\epsilon r} e^{-\kappa_0 r}. \quad (56.2)$$

Размер области действия поля центра рассеяния определяется величиной $R = \frac{1}{\kappa_0}$. При $\kappa_0 = 0$ $R = \infty$, что соответствует кулоновскому полю заряда Ze . Потенциальная энергия (56.2) достаточно хорошо описывает электрон (или дырку) в поле нейтрального атома, при этом $\kappa_0 \approx 10^8 \text{ см}^{-1}$.

Чтобы найти дифференциальное эффективное сечение рассеяния, необходимо знать V_{κ} :

$$\begin{aligned} \frac{V_{\kappa}}{4\pi} &= \pm \int_0^{\infty} \frac{Ze^2}{er} e^{-\kappa_0 r} \frac{\sin \kappa r}{\kappa} dr = \\ &= \pm \frac{Ze^2}{2i\epsilon\kappa} \left\{ \int_0^{\infty} [e^{-\kappa_0 r + i\kappa r} - e^{-\kappa_0 r - i\kappa r}] dr \right\} = \pm \frac{Ze^2}{\epsilon} \frac{1}{\kappa_0^2 + \kappa^2}. \end{aligned} \quad (56.3)$$

Подставляя выражение (56.3) в (55.40), запишем выражение для $\sigma(\theta, \varphi)$:

$$\sigma(\theta, \varphi) = \frac{4m^{*2} Z^2 e^4}{\hbar^4 \epsilon^2 |\kappa_0^2 + \kappa^2|^2}. \quad (56.4)$$

Угловая зависимость $\sigma(\theta, \varphi)$ содержится в κ :

$$\sigma(\theta, \varphi) = 4 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2} \right)^2 \frac{1}{[\kappa_0^2 + 4\kappa^2 \sin^2 \frac{\theta}{2}]^2} = \sigma(\theta). \quad (56.5)$$

Выражение (56.5) для рассеяния на ионах упрощается: $\kappa_0 = 0$ и

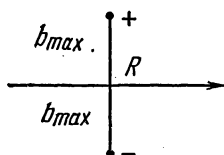


Рис. 81. Максимальное прицельное расстояние при движении частицы в поле двух ионов

$$\sigma(\theta) = \frac{1}{4} \left(\frac{Ze^2}{\epsilon m^* v^2} \right)^2 \frac{1}{\sin^4 \frac{\theta}{2}} \quad (56.6)$$

— известная формула Резерфорда, полученная им при описании рассеяния α -частиц полем ядер.

Если вычислить S^* и σ_c из (56.6), то получим бесконечность. Это связано с тем, что малым углам отклонения соответствует большое эффективное сечение, поскольку малым углам отклонения соответствуют большие расстояния между ионом и рассеиваемой частицей — большие *прицельные расстояния* b . При движении же частицы в твердом теле нет необходимости рассматривать изменение b от нуля до бесконечности. Можно установить верхний предел значения прицельного расстояния из следующих простых соображений. Предположим, что два иона находятся на расстоянии R друг от друга (рис. 81). Очевидно, что отклонение носителя заряда определяется ближайшим ионом. Естественно выбрать в качестве верхнего предела для прицельного расстояния величину $R/2$. Если концентрация ионов N_I , то среднее расстояние между ними $N_I^{-\frac{1}{3}}$, поэтому

верхний предел для b_{\max} равен $\frac{1}{2} N_I^{-\frac{1}{3}}$. Ему соответствует некоторый минимальный угол отклонения θ_{\min} , определяемый из условия

$$\operatorname{tg} \frac{\theta_{\min}}{2} = \frac{Ze^2}{\varepsilon m^* v^2} \frac{2}{N_I^{-\frac{1}{3}}}. \quad (56.7)$$

Введение θ_{\min} позволяет получить конечное значение S^* :

$$\begin{aligned} S^* &= 2\pi \int_{\theta_{\min}}^{\pi} \sigma(\theta) \sin \theta \, d\theta = \frac{\pi}{2} \frac{1}{4} \left(\frac{Ze^2}{\varepsilon m^* v^2} \right)^2 \int_{\theta_{\min}}^{\pi} \frac{\sin \theta \, d\theta}{\sin^4 \frac{\theta}{2}} = \\ &= \frac{\pi}{2} \left(\frac{Ze^2}{\varepsilon m^* v^2} \right)^2 \int_{\theta_{\min}}^{\pi} \frac{2 \cos \frac{\theta}{2} \cdot 2d \frac{\theta}{2}}{\sin^3 \frac{\theta}{2}} = \pi \left(\frac{Ze^2}{\varepsilon m^* v^2} \right)^2 \left(\frac{1}{\sin^2 \frac{\theta_{\min}}{2}} - 1 \right). \end{aligned} \quad (56.8)$$

Найдем эффективное сечение проводимости σ_c :

$$\sigma_c = 2\pi \int_{\theta_{\min}}^{\pi} \sigma(\theta) (1 - \cos \theta) \sin \theta \, d\theta = \frac{\pi}{2} \left(\frac{Ze^2}{\varepsilon m^* v^2} \right)^2 \int_{\theta_{\min}}^{\pi} \frac{(1 - \cos \theta) \sin \theta \, d\theta}{\sin^4 \frac{\theta}{2}}. \quad (56.9)$$

Учитывая, что

$$1 - \cos \theta = 2 \sin^2 \frac{\theta}{2}; \quad \sin \theta = 2 \sin \frac{\theta}{2} \cos \frac{\theta}{2}, \quad (56.10)$$

получим

$$\int_{\theta_{\min}}^{\pi} \frac{(1 - \cos \theta) \sin \theta \, d\theta}{\sin^4 \frac{\theta}{2}} = 8 \int_{\theta_{\min}}^{\pi} \frac{\cos \frac{\theta}{2} \cdot d \frac{\theta}{2}}{\sin \frac{\theta}{2}} = -8 \ln \sin \frac{\theta_{\min}}{2} \quad (56.11)$$

и

$$\sigma_c = -4\pi \left(\frac{Ze^2}{\varepsilon m^* v^2} \right)^2 \ln \sin \frac{\theta_{\min}}{2}. \quad (56.12)$$

Выразим θ_{\min} через b_{\max} в соответствии с (56.7):

$$\frac{1}{\sin^2 \frac{\theta_{\min}}{2}} = 1 + \operatorname{ctg}^2 \frac{\theta_{\min}}{2} = 1 + \frac{1}{2} \left(\frac{\varepsilon m^* v^2}{Ze^2} \right)^2. \quad (56.13)$$

В таком случае, учитывая (56.13), (56.8) и (56.9), получим для S^* и σ_c

$$S^* = \pi \left(\frac{N_I^{-\frac{1}{3}}}{2} \right)^2 = \pi b_{\max}^2 \quad (56.14)$$

и

$$\begin{aligned} \sigma_c &= -4\pi \left(\frac{Ze^2}{\epsilon m^* v^2} \right)^2 \ln \left[1 + \frac{N_I^{-\frac{2}{3}}}{4} \left(\frac{\epsilon m^* v^2}{Ze^2} \right)^2 \right]^{-\frac{1}{2}} = \\ &= 2\pi \left(\frac{Ze^2}{\epsilon m^* v^2} \right)^2 \ln \left[1 + \frac{1}{4N_I^{\frac{2}{3}}} \left(\frac{\epsilon m^* v^2}{Ze^2} \right)^2 \right]. \end{aligned} \quad (56.15)$$

Выражение для S^* имеет очень наглядный смысл: свободный носитель заряда взаимодействует с ионом на любом расстоянии, поэтому интегральное эффективное сечение рассеяния на ионе равно бесконечности. Но если учесть, что при большом числе ионов достаточно

ограничиться прицельным расстоянием $b_{\max} = \frac{N_I^{-\frac{1}{3}}}{2}$, то для эффективного сечения получим величину, равную площади круга радиуса b_{\max} . Отметим, что чем больше концентрация ионов примеси, тем сильнее они взаимно компенсируют поля, тем меньше интегральное сечение рассеяния носителей заряда на них. Пусть $N_I = 10^{16} \text{ см}^{-3}$, в этом случае $b_{\max} \approx 10^{-5} \text{ см}$ и $S^* \approx 10^{-10} \text{ см}^2$, что намного превосходит величину S_I , взятую нами в § 53 для оценки роли ионов примеси в рассеянии. Если оценить длину свободного пробега, то получим для $N_I = 10^{16} \text{ см}^{-3}$, $l_I = 10^{-6} \text{ см}$. Это означает, что ионы примеси играют значительно большую роль в рассеянии и определении длины свободного пробега, чем предполагалось в § 53.

Зависимость l_I от концентрации ионов примеси сравнительно слабая:

$$l_I = \frac{1}{N_I S_I^*} = \frac{1}{N_I \frac{\pi}{4} N_I^{-\frac{2}{3}}} = \frac{4}{\pi N_I^{\frac{1}{3}}} = \frac{8}{\pi} b_{\max}. \quad (56.16)$$

В табл. 17 указано значение l_I , подсчитанное по формуле (56.16) для некоторых значений N_I .

Таблица 17

$N_I, \text{ см}^{-3}$	10^{21}	10^{18}	10^{15}	10^{12}
$l_I, \text{ см}$	$1,3 \cdot 10^{-7}$	$1,3 \cdot 10^{-6}$	$1,3 \cdot 10^{-5}$	$1,3 \cdot 10^{-4}$

Интегральное эффективное сечение не зависит от энергии и скорости носителей заряда, но дифференциальное эффективное сечение сильно зависит от скорости носителей заряда:

$$\sigma(\theta) \sim \frac{1}{v^4}. \quad (56.17)$$

Кроме того, как показывает выражение для $\sigma(\theta)$, рассеяние на ионах резко анизотропно: $\sigma(\theta) \sim \frac{1}{\sin^4 \frac{\theta}{2}}$. Эффективное сечение проводимости σ_c имеет двойную зависимость от скорости — степенную и логарифмическую. Введем обозначение:

$$x = \frac{N_I^{-\frac{2}{3}}}{4} \left(\frac{\epsilon m^* v^2}{Ze^2} \right)^2; \quad \sigma_c = \frac{\pi}{2} N_I^{-\frac{2}{3}} \frac{\ln(1+x)}{x}. \quad (56.18)$$

При $x \gg 1$

$$\sigma_c \sim \frac{1}{v^4}, \quad (56.19)$$

поскольку логарифмическая зависимость значительно слабее степенной. Если же $x \ll 1$, то выражение (56.15) для σ_c упрощается:

$$\sigma_c \approx \frac{\pi}{2} N_I^{-\frac{2}{3}} \frac{\ln(1+x)}{x} \approx \frac{2\pi N_I^{-\frac{2}{3}}}{4}. \quad (56.20)$$

Эффективное сечение проводимости не зависит от скорости и энергии рассеиваемых частиц. Рассмотрим условия, необходимые для выполнения неравенства (56.19) или (56.20); для этого введем значение энергии в выражения для x :

$$x = \left(\frac{\epsilon m^* v^2}{Ze^2} \right)^2 \frac{1}{N_I^{\frac{2}{3}}} = \left(\frac{E}{\frac{Ze^2}{\epsilon N_I^{\frac{1}{3}}}} \right)^2. \quad (56.21)$$

Но $\frac{Ze^2}{\epsilon N_I^{\frac{1}{3}}} = V(2b_{\max})$ есть потенциальная энергия на расстоянии

$2b_{\max}$ от рассеивающего центра. При $E \ll V(2b_{\max})$, т. е. для «медленных» носителей заряда, σ_c не зависит от их скорости и энергии. При $E \gg V(2b_{\max})$ $\sigma_c \sim \frac{1}{v^4}$, или $\sigma_c \sim \frac{1}{E^2}$; при $E \approx V(2b_{\max})$ зависимость σ_c от v остается, но значительно более слабая, чем при больших энергиях. Значения $V(2b_{\max})$ для различных N_I при $\epsilon = 10$ и $Z = 1$ даны в табл. 18.

В табл. 18, кроме того, приведены значения температур T_K , вычисленные из условия $V(2b_{\max}) = kT$. Интервал энергий (0 — E)

$N_I, \text{ см}^{-3}$	10^{21}	10^{18}	10^{15}	10^{12}	10^9
$2b_{\text{max}}, \text{ см}$	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}
$V (2b_{\text{max}}), \text{ эВ}$	0,143	0,0143	0,0014	0,0001	0,00001
$T_{\text{max}} = \frac{V}{k}, \text{ К}$	1670	167	16,7	1,7	0,17

и соответствующий ему интервал температур, в которых σ_c не зависит от скорости, тем больше, чем выше концентрация ионов примеси. При $N_I \approx 10^{16}$ и ниже интервал значений энергий, в котором σ_c не зависит от E , очень мал, поэтому можно считать, что $\sigma_c \sim E^{-2}$. Перейдем к рассмотрению зависимости времени релаксации от энергии. Согласно (54.17) и (56.15)

$$\tau_I(\kappa) = \frac{1}{N_I \sigma_c v(\kappa)} = \frac{\varepsilon^2 m^* v^3}{2\pi N_I Z^2 e^4 \ln \left[1 + \left(\frac{\varepsilon m^* v^2}{2N_I^{2/3} Z e^2} \right)^2 \right]}. \quad (56.22)$$

Это выражение для τ_I называют *формулой Конуэлл — Вайскопфа*. Из (56.22) видно, что $\tau_I \sim v^3$. Выражая скорость через энергию:

$$v = \left(\frac{2E}{m^*} \right)^{1/2}, \text{ получим}$$

$$\tau_I = \frac{\sqrt{2} \varepsilon^2 m^{*2} E^{3/2}}{N_I \pi Z^2 e^4 \ln \left[1 + \left(\frac{vE}{N_I^{2/3} Z e^2} \right)^2 \right]} \approx \tau_0 E^{3/2}. \quad (56.23)$$

Мы видим, что при рассеянии носителей заряда на ионах примеси время релаксации зависит от энергии в виде $\tau(E) = \tau_0 E^p$ при $p = 3/2$. Релаксационный процесс длится тем дольше, чем больше энергия носителей заряда. Применимость этой формулы была уже рассмотрена. При $E \rightarrow 0$ выражение (56.24) не применимо, оно должно быть заменено выражением

$$\tau_I = \frac{1}{N_I \sigma_c v} = \frac{1}{N_I \frac{\pi}{2} N_I^{-2/3} v} = \frac{2}{\pi N_I^{1/3} v} = \frac{V m^*}{\sqrt{2} \pi N_I^{1/3}} E^{-1/2}, \quad (56.24)$$

что вытекает непосредственно из (56.23).

Если определить длину свободного пробега носителей заряда с энергией E соотношением $l_I(E) = v\tau_I(E)$, то

$$l_I(E) \approx \frac{1}{N_I \sigma_c}; \quad l_I \sim E^2; \quad l_I(E) = \text{const} \quad (56.25)$$

при больших и соответственно малых энергиях электронов и дырок. Вернемся к выражению (56.5), с помощью которого вычислим транспортное эффективное сечение:

$$\begin{aligned} \sigma_c &= 2\pi \int_0^\pi 4 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2} \right)^2 \frac{(1 - \cos \theta) \sin \theta \, d\theta}{\left[\kappa_0^2 + 4\kappa^2 \sin^2 \frac{\theta}{2} \right]^2} = \\ &= 32\pi \left(\frac{Ze^2 m^*}{\epsilon \hbar^2} \right)^2 \int_0^\pi \frac{\sin^2 \frac{\theta}{2} \, d \sin^2 \frac{\theta}{2}}{\left[\kappa_0^2 + 4\kappa^2 \sin^2 \frac{\theta}{2} \right]^2}. \end{aligned} \quad (56.26)$$

Вынося $4\kappa^2$ и вводя новую переменную $y = \sin^2 \frac{\theta}{2}$, получим

$$\sigma_c = 2 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2 \kappa^2} \right)^2 \int_0^1 \frac{y \, dy}{\left[\frac{\kappa_0^2}{4\kappa^2} + y \right]^2}. \quad (56.27)$$

Интеграл в (56.27) вычисляется, например, по частям, в результате чего имеем

$$\sigma_c = 2\pi \left(\frac{Ze^2 m^*}{\epsilon \hbar^2 \kappa^2} \right)^2 \left[\ln \left(1 + \frac{4\kappa^2}{\kappa_0^2} \right) - \frac{4\kappa^2}{\kappa_0^2 + 4\kappa^2} \right]. \quad (56.28)$$

При $\kappa_0 \rightarrow 0$ выражение для $\sigma_c \rightarrow \infty$. Параметр κ_0 можно связать с концентрацией свободных носителей заряда, учитывая экранирующее действие роя электронов или дырок, притягивающихся полем иона. Но поскольку принципиального отличия от выражения для σ_c и τ_I , даваемого формулой Конуэлл — Вайскопфа, не возникает, исследовать выражение (56.28) не будем.

Резюме § 56

1. Дифференциальное эффективное сечение для рассеяния носителей заряда в кулоновском поле иона представляет собой формулу Резерфорда

$$\sigma(\theta, \varphi) = \frac{1}{4} \left(\frac{Ze^2}{\epsilon m^* v^2} \right)^2 \frac{1}{\sin^4 \frac{\theta}{2}}. \quad (56.1p)$$

Рассеяние резко анизотропно, превалирует рассеяние на малые углы, которым соответствуют большие прицельные расстояния. Выражения для S^* и σ_c расходятся.

2. Обрезая прицельные расстояния на некотором значении b_{\max} или не принимая во внимание отклонения на углы меньше некоторого θ_{\min} , получим конечные значения для S^* и σ_c . Физической основой, позволяющей ввести b_{\max} , является экранирующее действие роя свободных носителей заряда, превращающего кулоновский потенциал в экранирующий потенциал вида (56.2), или компенсирующее действие полей соседних ионов. В первом случае выражения для σ и σ_c имеют вид

$$\sigma(\theta) = 4 \left(\frac{Ze^2 m^*}{\epsilon \hbar^2} \right)^2 \frac{1}{\left[\kappa_0^2 + 4\kappa^2 \sin^2 \frac{\theta}{2} \right]^2}; \quad (56.2p)$$

$$\sigma_c(\theta) = 2\pi \left(\frac{Ze^2 m^*}{\epsilon \hbar^2 \kappa^2} \right)^2 \left[\ln \left(1 + \frac{4\kappa^2}{\kappa_0^2} \right) - \frac{4\kappa^2}{\kappa_0^2 + 4\kappa^2} \right], \quad (56.3p)$$

где κ_0 связано с концентрацией свободных носителей заряда. Во втором случае для σ_c получаем

$$\sigma_c = 2\pi \left(\frac{Ze^2}{\epsilon m^* v^2} \right)^2 \ln \left[1 + \left(\frac{\epsilon m^* v^2}{2 N_I^{1/3} Ze^2} \right)^2 \right]. \quad (56.4p)$$

3. При достаточно больших энергиях носителей заряда эффективное сечение проводимости σ_c обратно пропорционально квадрату энергии, в то время как для малых энергий σ_c не зависит от энергии.

4. Время релаксации носителей заряда при рассеянии на ионах примеси имеет вид

$$\tau_I = \frac{(2m^*)^{1/2} \epsilon^2}{N_I \pi Z^2 e^4} \frac{E^{3/2}}{\ln \left[1 + \left(\frac{\epsilon E}{N_I^{1/3} Ze^2} \right)^2 \right]}. \quad (56.5p)$$

Оно пропорционально эффективной массе в степени $1/2$. Для больших энергий $\tau_I \sim E^{3/2}$, для малых энергий $\tau_I \sim E^{-1/2}$. Интервал энергий, при котором $\tau_I \sim E^{-1/2}$, играет заметную роль при малых температурах, близких к нулю.

§ 57. РАССЕЯНИЕ НА НЕЙТРАЛЬНЫХ АТОМАХ ПРИМЕСИ

Рассеяние носителей заряда на нейтральных примесях можно описать выражениями (56.5) и (56.28) соответственно. Поскольку для атома примеси $\kappa_0 \approx 10^8 \text{ см}^{-1}$, что соответствует энергии $\frac{m}{m^*} 8 \text{ эВ}$, практически для всех носителей заряда величиной κ^2 по сравнению

с κ_0^2 можно пренебречь, в связи с чем выражения для $\sigma(\theta)$ и σ_c упрощаются:

$$\sigma(\theta) = 4 \left(\frac{Ze^2 m^*}{\varepsilon \hbar^2 \kappa_0^2} \right)^2 \quad (57.1)$$

и

$$\sigma_c = 8\pi \left(\frac{Ze^2 m^*}{\varepsilon \hbar^2 \kappa_0^2} \right)^2, \quad (57.2)$$

т. е. рассеяние изотропно, а σ и σ_c не зависят от энергии рассеиваемых частиц.

Оценим величину σ ; положив $m^* = m$; $\varepsilon = 10$; $Z = 1$; $\kappa_0 = 10^8 \text{ см}^{-1}$, получим $\sigma \approx 2 \cdot 10^{-17} \text{ см}^2$.

Время релаксации при рассеянии на нейтральных атомах примеси согласно (57.4) равно

$$\tau_{\text{H}} = \frac{1}{N_{\text{H}} \sigma_c} = \frac{1}{N_{\text{H}} V \sqrt{2} \pi} \left(\frac{\varepsilon \hbar^2 \kappa_0^2}{Ze^2} \right)^2 \frac{E^{-\frac{1}{2}}}{m^{*\frac{3}{2}}}. \quad (57.3)$$

Выражения (57.2) и (57.3) для σ_c и τ_{H} справедливы постольку, поскольку нейтральный атом или точечный дефект можно описать экранирующим потенциалом (56.2). Укажем на одну особенность в зависимости σ_c от диэлектрической проницаемости вещества. При оценке σ_c для κ_0 мы использовали значение, справедливое для изолированного атома. Если атом примеси описывается водородоподобной моделью, то размер электронного облака атома примеси возрастает в ε раз, как это было показано в § 21, поэтому необходимо заменить κ_0 на $\kappa'_0 = \frac{\kappa_0}{\varepsilon}$, вследствие чего изменится характер зависимости σ и σ_c от ε :

$$\sigma_c = 8\pi \left(\frac{Ze^2 m^* \varepsilon}{\hbar^2 \kappa_0^2} \right)^2, \quad (57.4)$$

что при вышеуказанных предположениях дает $\sigma \approx 2 \cdot 10^{-13} \text{ см}^2$.

Полученные выше выражения справедливы в том случае, если рассеиваемая частица не влияет на поле рассеивающего центра. В действительности же это не так. Свободный носитель заряда своим полем поляризует атом примеси, вследствие чего у атома примеси появляется индуцированный дипольный момент. Его величина d определяется напряженностью поля, создаваемого рассеиваемой частицей и поляризуемостью χ атома:

$$|d| = \frac{\chi e}{\varepsilon r^3}. \quad (57.5)$$

Индукцированный дипольный момент в свою очередь создает электрическое поле, действующее на свободный заряд. Как известно из электростатики, напряженность поля диполя в направлении его оси равна

$$E_d = \frac{2d}{\varepsilon r^3} = \frac{2\chi e}{\varepsilon^2 r^5}. \quad (57.6)$$

Для потенциальной энергии свободного носителя заряда в поле индуцированного им диполя из (57.8) получим

$$V = -\frac{\chi e^2}{2\epsilon^2 r^4}. \quad (57.7)$$

Выражение (57.7) справедливо до тех пор, пока расстояние между рассеивающим центром и носителем заряда велико по сравнению с κ_0^{-1} , поскольку при меньших расстояниях не применимо выражение (57.6), справедливое лишь для точечного диполя. При меньших расстояниях необходимо найти распределение заряда рассеивающего центра с учетом поля рассеиваемой частицы. Однако при достаточно больших расстояниях, при которых экранирующий потенциал практически равен нулю, будет иметь место рассеяние носителей заряда нейтральным рассеивающим центром за счет *поляризационного потенциала* (57.7). Найдем дифференциальное эффективное сечение рассеяния на потенциале (потенциальной энергии) (57.7).

Для этого вычислим V_κ :

$$\frac{V_\kappa}{4\pi} = \int_{R_0}^{\infty} r V(r) \frac{\sin \kappa r}{\kappa} dr = -\frac{\chi e^2}{2\epsilon^2} \int_{R_0}^{\infty} \frac{\sin \kappa r}{\kappa r^3} dr. \quad (57.8)$$

В отличие от кулоновского поля выражение (57.8) сходится при $r \rightarrow \infty$, но расходится при $r \rightarrow 0$. Однако расходимость в последнем случае снимается введением некоторого минимального расстояния R_0 , поскольку при $r \rightarrow 0$ выражение (57.7), как уже говорили, не применимо.

Вычисление V_κ из (57.8) можно провести, переходя к комплексным переменным. Однако проведем его оценку иначе. Прежде всего обратим внимание, что при $\kappa = 0$ интеграл вычисляется элементарно:

$$\frac{V_{\kappa=0}}{4\pi} = -\frac{\chi e^2}{2\epsilon^2} \int_{R_0}^{\infty} \frac{dr}{r^2} = \frac{\chi e^2}{2\epsilon^2 R_0}. \quad (57.9)$$

Это же выражение будет справедливым при $\theta = 0$ и любой энергии носителя заряда или при $\kappa = 0$. Можно считать применимым выражение (57.9) и в том случае, когда κR_0 мало. В таком случае приходим к выводу, что *при малой энергии носителей заряда, когда $\kappa R_0 \ll 1$, дифференциальное эффективное сечение рассеяния на поляризационном потенциале изотропно и не зависит явно от энергии:*

$$\sigma = \frac{4m^*{}^2}{\hbar^4} \frac{\chi^2 e^4}{4\epsilon^4 R_0^3}. \quad (57.10)$$

Если κ величина большая, то $\sin \kappa r$ — быстро осциллирующая функция, поэтому интеграл (57.9) практически равен нулю при больших θ . Но при малом θ , когда можно считать, что $\frac{\sin \kappa r}{\kappa} \sim r$, интеграл вновь вычисляется, он равен $\frac{e^2 \chi}{2\epsilon^2 R_0}$. Таким образом, *рассеяние на нейтральных центрах вследствие их поляризации резко*

анизотропно — рассеяние происходит только внутри конуса с угловым раствором $\Delta\theta \sim \frac{1}{\kappa R_0}$, вне этого конуса рассеяние практически отсутствует. Эффективное сечение рассеяния внутри конуса слабо зависит от энергии. Но так как при изменении κ угловой раствор $\Delta\theta$ меняется, то изменяется и дифференциальное эффективное сечение примерно как $(\kappa R_0)^{-2}$. Вычисляя транспортное эффективное сечение, получим $\sigma_c \sim (\kappa R_0)^{-4}$. Таким образом, с ростом энергии рассеиваемой частицы изотропное рассеяние переходит в резко анизотропное. Однако в действительности все обстоит еще более сложно. При рассеянии электронов вследствие тождественности рассеиваемого электрона и электрона атома возможен обменный эффект. Кроме того, необходимо учитывать, что водородоподобный нейтральный рассеивающий центр имеет целый ряд возбужденных состояний, вследствие чего рассеяние может быть неупругим. Другими словами, задача о рассеянии электронов и дырок нейтральными центрами является значительно сложнее задачи о рассеянии частиц заряженными примесями. Решая задачу о рассеянии носителей заряда в поле водородоподобного атома примеси, Эргинсой показал, что σ_c примерно пропорционально v^{-1} ; из этого следует, что время релаксации не зависит от энергии рассеиваемых частиц. Для оценки времени релаксации обычно используют приближенное соотношение, полученное при решении уравнений численным методом:

$$\tau_H = \frac{m^{*2} e^2}{20 \epsilon \hbar^3} \frac{1}{N_H}, \quad (57.11)$$

где N_H — концентрация нейтральной примеси при данной температуре. Необходимо отметить, что электрически активная примесь может быть нейтральной только при очень низких температурах.

В заключение этого параграфа отметим, что при рассеянии на дислокациях эффективное сечение рассеяния существенно зависит от угла между начальной скоростью носителей заряда и осью дислокаций. Можно найти σ , исходя из общего уравнения (55.30).

Опуская выкладки, укажем, что для времени релаксации при рассеянии на линейных дислокациях обычно пользуются выражением вида

$$\tau_d = \frac{3}{8 R_d v N_d}, \quad (57.12)$$

где R_d — радиус, N_d — поверхностная плотность дислокаций.

Резюме § 57

1. Рассеяние носителей заряда на нейтральных рассеивающих центрах изотропно. Анизотропия наблюдается лишь для очень больших энергий электронов и дырок.

2. При рассеянии носителей заряда на атомах примеси или точечных дефектах время релаксации практически не зависит от скорости частиц; при рассеянии на линейных дислокациях $\tau \sim v^{-1}$.

Для понимания многих физических явлений в полупроводниках необходимо иметь представление о методе описания колебаний решетки. Из общего курса физики известно, что в нагретом теле атомы совершают хаотические тепловые колебания относительно положений равновесия. Амплитуда тепловых колебаний зависит от температуры. Электроны и дырки, двигаясь в веществе, атомы которого совершают хаотические движения, могут обмениваться с ними энергией. Благодаря колебаниям решетки устанавливается термодинамическое равновесие между решеткой и электронным газом, оно характеризуется некоторой температурой, одинаковой как для атомов (ионов) решетки, так и для электронного газа. Перейдем к математическому описанию колебаний решетки.

Обозначим через R_{nj}^0 равновесную координату атома, находящегося в j -узле элементарной ячейки, положение которой в кристалле определяется вектором $\mathbf{n} = (n_1 a_1; n_2 a_2; n_3 a_3)$. Пусть атом R_{nj}^0 смещен из положения равновесия и его координата при этом равна R'_{nj} . Вектор

$$\mathbf{u}_{nj} = R_{nj} - R_{nj}^0 \quad (58.1)$$

называют вектором смещения j -атома ячейки \mathbf{n} . Пусть $U_0 = U_0(\dots, R_{nj}^0, \dots)$ есть равновесное значение потенциальной энергии решетки, а U — потенциальная энергия решетки со смещенными атомами, ее можно выразить в виде функции от смещения атомов:

$$U = U(\dots, R_{nj}, \dots) = U(\dots, \mathbf{u}_{nj}, \dots). \quad (58.2)$$

Разложим U в ряд Тейлора по вектору смещения относительно равновесного состояния кристалла:

$$\begin{aligned} U(\dots, \mathbf{u}_{nj}, \dots) = & U_0 + \sum_{nj} \frac{\partial U}{\partial u_{nj}} u_{nj} + \frac{1}{2} \sum_{nj, n'j'} \frac{\partial^2 U}{\partial u_{nj} \partial u_{n'j'}} u_{nj} u_{n'j'} + \\ & + \frac{1}{6} \sum_{nj, n'j', n''j''} \frac{\partial^3 U}{\partial u_{nj} \partial u_{n'j'} \partial u_{n''j''}} u_{nj} u_{n'j'} u_{n''j''} + \dots \end{aligned} \quad (58.3)$$

Так как разложение ведется относительно равновесного состояния, в котором тело имеет экстремальное значение энергии, то $\frac{\partial U}{\partial u_{nj}} = 0$ и, следовательно, ряд начинается с квадратичных членов.

Обозначим элементы тензоров второго и третьего ранга соответственно

$$A_{nj, n'j'} = \frac{\partial^2 U}{\partial u_{nj} \partial u_{n'j'}} = A_{n'j', nj}, \quad (58.4)$$

$$B_{nj, n'j', n''j''} = \frac{\partial^3 U}{\partial u_{nj} \partial u_{n'j'} \partial u_{n''j''}}. \quad (58.5)$$

Учитывая (58.4) и (58.5), запишем

$$U = U_0 + \frac{1}{2} \sum_{n_j, n'_j} A_{n_j, n'_j} u_{n_j} u_{n'_j} + \dots \quad (58.6)$$

Если смещения малы, то, пренебрегая кубическими членами, представим U в виде билинейной функции от смещения атомов. Сила, действующая на атом mi , равна производной по u_{mi} от U со знаком минус:

$$\begin{aligned} F_{mi} &= -\frac{\partial U}{\partial u_{mi}} = -\frac{1}{2} \sum_{n_j, n'_j} A_{n_j, n'_j} (u_{n'_j} \delta_{mi, n_j} + u_{n_j} \delta_{mi, n'_j}) = \\ &= -\frac{1}{2} \left(\sum_{n'_j} A_{mi, n'_j} u_{n'_j} + \sum_{n_j} A_{n_j, mi} u_{n_j} \right) = -\sum_{n_j} A_{mi, n_j} u_{n_j}. \end{aligned} \quad (58.7)$$

Величина $A_{n_j, mi}$ зависит не от m и n в отдельности, а от расстояния между атомами, т. е. от $|n - m|$:

$$A_{n_j, mi} = A_{ji} (n - m). \quad (58.8)$$

Отсюда следует, что трансляция на целое число периодов решетки не меняет «коэффициента упругости»:

$$A_{n+p, j; m+p, i} = A_{ji} (n - m) = A_{n_j, mi}. \quad (58.9)$$

Запишем уравнение движения атома mi с массой M_i , учитывая, что F_{mi} определяется соотношением (58.7):

$$M_i \ddot{u}_{mi} = -\sum_{n_j} A_{mi, n_j} u_{n_j}. \quad (58.10)$$

Уравнение (58.10) содержит в явном виде массу атома M_i , что в ряде случаев вызывает дополнительные трудности при решении задачи. Введем так называемые *приведенные смещения* w_{n_j} , определяемые условием

$$w_{n_j} = \sqrt{M_i} u_{n_j}, \quad (58.11)$$

а вместо матрицы $\{A_{n_j, n'_j}\}$ введем матрицу $\{D_{n_j, n'_j}\}$:

$$D_{n_j, n'_j} = \frac{A_{n_j, n'_j}}{\sqrt{M_i M_{j'}}}. \quad (58.12)$$

Матрица $\{D_{n_j, n'_j}\}$ носит название динамической. Размерность элементов матрицы D равна T^{-2} . Действительно, так как размерность A совпадает с размерностью коэффициента упругости

$$[A_{n_j, mi}] = \left[\frac{F}{L} \right] = \left[\frac{ML}{T^2 L} \right] = \left[\frac{M}{T^2} \right], \quad (58.13)$$

то размерность $D_{nj, ml}$ совпадает с размерностью квадрата частоты. Из очевидного равенства

$$A_{nj, ml} u_{nj} u_{ml} = \frac{A_{nj, ml}}{\sqrt{M_j M_l}} \sqrt{M_j} u_{nj} \sqrt{M_l} u_{ml} \quad (58.14)$$

следует, что потенциальная энергия в приведенных смещениях имеет вид

$$U = U_0 + \frac{1}{2} \sum_{nj, ml} D_{nj, ml} w_{nj} w_{ml}. \quad (58.15)$$

Уравнение движения (58.10) можно теперь представить в таком виде:

$$M_l \ddot{u}_{ml} = \sqrt{M_l} \ddot{w}_{ml} = - \sum_{nj} A_{ml, nj} u_{nj} = - \sum_{nj} D_{ml, nj} w_{nj} \sqrt{M_j}, \quad (58.16)$$

или

$$\ddot{w}_{ml} = - \sum_{nj} D_{ml, nj} w_{nj}. \quad (58.17)$$

Система уравнений (58.10) или (58.17) показывает, что смещения различных атомов u_{nj} (или w_{nj}) взаимосвязаны. Решение системы уравнений (58.17) в общем случае очень сложно. Чтобы упростить систему уравнений (58.17), перейдем к другим переменным, описывающим состояние кристалла. Обозначим через N число элементарных ячеек в кристалле, а через s — число атомов в элементарной ячейке, число степеней свободы кристалла равно $3sN$; они описываются sN векторами u_{nj} , или $3sN$ их компонентами. Вместо w_{nj} можно ввести любую другую совокупность переменных, пригодных для описания состояния кристалла. Наиболее просто состояние кристалла описывается в так называемых нормальных, или главных, координатах.

Рассмотрим с этой целью произвольное линейное преобразование координат (сохраним векторное выражение для переменных).

Пусть имеется sN новых переменных q_α ($\alpha = 1, \dots, sN$), которые связаны со старыми переменными посредством матрицы S , задающей линейное преобразование координат в виде

$$q_\alpha = \sum_{nj} S_{\alpha, nj} w_{nj}. \quad (58.18)$$

Будем считать, что матрица S имеет отличный от нуля детерминант, в связи с чем у нее есть обратная матрица, которую мы обозначим через T :

$$T = S^{-1}; \quad ST = TS = I, \quad (58.19)$$

где I — единичная матрица. Уравнение (58.19) можно записать в таком виде

$$\sum_{nj} S_{\alpha, nj} T_{nj, \beta} = \delta_{\alpha\beta},$$

$$\sum_{\alpha} T_{nj, \alpha} S_{\alpha, ml} = \delta_{nj, ml} = \delta_{nm} \delta_{jl}. \quad (58.20)$$

Поскольку выбор матрицы S произволен, рассмотрим случай, когда S унитарна*. Для унитарной матрицы можно записать

$$T_{nj, \beta} = S_{\beta, nj}^*, \quad S_{\beta, nj} = T_{nj, \beta}^* \quad (58.21)$$

$$\sum_{nj} S_{\alpha, nj} S_{\beta, nj}^* = \delta_{\alpha\beta}, \quad (58.22)$$

$$\sum_{\alpha} S_{\alpha, nj}^* S_{\alpha, ml} = \delta_{nj, ml} = \delta_{nm} \delta_{jl}. \quad (58.23)$$

Из (58.18) можно выразить w_{nj} через q_{α} ; для этого умножим (58.18) слева на $T_{ml, \alpha}$, просуммируем по α и получим

$$\begin{aligned} \sum_{\alpha} T_{ml, \alpha} q_{\alpha} &= \sum_{\alpha} \sum_{nj} T_{ml, \alpha} S_{\alpha, nj} w_{nj} = \sum_{nj} \left(\sum_{\alpha} T_{ml, \alpha} S_{\alpha, nj} \right) w_{nj} = \\ &= \sum_{nj} \delta_{ml, nj} w_{nj} = w_{ml}, \end{aligned} \quad (58.24)$$

т. е.

$$w_{ml} = \sum_{\alpha} T_{ml, \alpha} q_{\alpha} = \sum_{\alpha} S_{\alpha, ml}^* q_{\alpha}. \quad (58.25)$$

Найдем уравнение движения в новых переменных:

$$\ddot{w}_{ml} = \sum_{\alpha} T_{ml, \alpha} \ddot{q}_{\alpha} = - \sum_{nj, \beta} D_{ml, nj} w_{nj} = - \sum_{nj, \beta} D_{ml, nj} T_{nj, \beta} q_{\beta}. \quad (58.26)$$

Умножив уравнение (58.26) слева на $S_{\gamma, ml}$ и просуммировав по ml , получим:

для левой части уравнения

$$\sum_{ml} S_{\gamma, ml} \sum_{\alpha} T_{ml, \alpha} \ddot{q}_{\alpha} = \sum_{\alpha} \left(\sum_{ml} S_{\gamma, ml} T_{ml, \alpha} \right) \ddot{q}_{\alpha} = \sum_{\alpha} \delta_{\gamma\alpha} \ddot{q}_{\alpha} = \ddot{q}_{\gamma}, \quad (58.27)$$

и для правой части уравнения

$$\begin{aligned} - \sum_{ml} S_{\gamma, ml} \sum_{nj, \beta} D_{ml, nj} T_{nj, \beta} q_{\beta} &= - \sum_{\beta} \left(\sum_{ml, nj} S_{\gamma, ml} D_{ml, nj} T_{nj, \beta} \right) q_{\beta} = \\ &= - \sum_{\beta} (SDT)_{\gamma\beta} q_{\beta} = - \sum_{\beta} (SDS^{-1})_{\gamma\beta} q_{\beta}. \end{aligned} \quad (58.28)$$

Учитывая (58.27) и (58.28), запишем

$$\ddot{q}_{\gamma} = - \sum_{\beta} (SDS^{-1})_{\gamma\beta} q_{\beta}. \quad (58.29)$$

* Матрица S называется унитарной, если обратная к ней матрица совпадает с сопряженной: $S^+ = S^{-1}$. Сопряженная матрица S^+ определяется условием $S^+ = \tilde{S}^*$, где знак (*) означает комплексное сопряжение элементов матрицы, а (\sim) — транспонирование матрицы.

Поскольку матрица S произвольна, то выберем ее таким образом, чтобы матрица SDS^{-1} была диагональной

$$(SDS^{-1})_{\gamma\beta} = \omega_{\gamma}^2 \delta_{\gamma\beta}. \quad (58.30)$$

Размерность $(SDS^{-1})_{\gamma\beta}$ есть размерность квадрата частоты, так что ω_{γ} по размерности есть частота

Далее покажем, что подобная матрица S существует, а пока предположим, что наложение условия (58.30) на матрицу S возможно. В таком случае уравнение движения (58.29) упрощается:

$$\ddot{q}_{\gamma} + \omega_{\gamma}^2 q_{\gamma} = 0. \quad (58.31)$$

Мы видим, что переменные q_{γ} являются независимыми и изменяются по гармоническому закону

$$q_{\gamma}(t) = q_{\gamma}^0 e^{-i\omega_{\gamma} t}. \quad (58.32)$$

Вектор q_{γ}^0 содержит в себе амплитуду и начальную фазу. Величины ω_{γ} носят название собственных частот. Число собственных частот кристалла равно sN , поэтому кристалл описывается $3sN$ различными колебаниями в соответствии с полным числом степеней свободы совокупности sN атомов.

Независимые переменные, подчиняющиеся уравнению (58.31), называют нормальными, или главными, координатами кристалла. Выразим энергию кристалла в нормальных координатах. Потенциальная энергия

$$\begin{aligned} U &= U_0 + \frac{1}{2} \sum_{nj, ml} D_{nj, ml} w_{nj} w_{ml} = \\ &= U_0 + \frac{1}{2} \sum_{nj, ml} D_{nj, ml} \sum_{\alpha} T_{nj, \alpha} q_{\alpha} \sum_{\beta} T_{ml, \beta} q_{\beta} = \\ &= U_0 + \frac{1}{2} \sum_{\alpha, \beta} \left(\sum_{nj, ml} S_{\alpha, nj}^* D_{nj, ml} T_{ml, \beta} \right) q_{\alpha} q_{\beta}. \end{aligned} \quad (58.33)$$

Переменные u_{nj} , w_{nj} и матрицы A и D вещественны. Если считать нормальные координаты вещественными величинами, т. е. нормальными координатами считать не величину $q_{\alpha}^0 e^{-i\omega_{\alpha} t}$, а только ее вещественную часть $\text{Re}(q_{\alpha}^0 e^{-i\omega_{\alpha} t})$, то матрица S должна быть вещественной:

$$S_{\alpha, nj}^* = S_{\alpha, nj}. \quad (58.34)$$

Учитывая условие вещественности (58.34) матрицы S и уравнение (58.30), запишем уравнение (58.33) в очень простом виде:

$$U = U_0 + \frac{1}{2} \sum_{\alpha} \omega_{\alpha}^2 q_{\alpha}^2. \quad (58.35)$$

Найдем кинетическую энергию T :

$$\begin{aligned} T &= \sum_{nj} \frac{1}{2} M_j \dot{u}_{nj}^2 = \frac{1}{2} \sum_{nj} \dot{w}_{nj}^2 = \frac{1}{2} \sum_{nj} \left(\sum_{\alpha} T_{nj, \alpha} \dot{q}_{\alpha} \right) \cdot \left(\sum_{\beta} T_{nj, \beta} \dot{q}_{\beta} \right) = \\ &= \frac{1}{2} \sum_{\alpha\beta} \left(\sum_{nj} T_{nj, \alpha} T_{nj, \beta} \right) \dot{q}_{\alpha} \dot{q}_{\beta} = \frac{1}{2} \sum_{\alpha} \dot{q}_{\alpha}^2. \end{aligned} \quad (58.36)$$

Для полной энергии $H = T + U$ соответственно получим

$$H = T + U = \sum_{\alpha} \left(\frac{1}{2} \dot{q}_{\alpha}^2 + \frac{1}{2} \omega_{\alpha}^2 q_{\alpha}^2 \right) + U_0 = \sum_{\alpha} H_{\alpha} + U_0, \quad (58.37)$$

где

$$H_{\alpha} = \frac{\dot{q}_{\alpha}^2}{2} + \frac{\omega_{\alpha}^2 q_{\alpha}^2}{2} \quad (58.38)$$

есть функция Гамильтона, соответствующая α -нормальной координате. Полная энергия колебаний решетки складывается из энергий нормальных колебаний.

Выражение (58.38) совпадает с выражением полной энергии гармонического осциллятора единичной массы ($M_{\alpha} = 1$ и $p_{\alpha} = \dot{q}_{\alpha}$):

$$H_{\alpha} = \frac{p_{\alpha}^2}{2} + \frac{\omega_{\alpha}^2 q_{\alpha}^2}{2}. \quad (58.39)$$

Согласно классической физике энергия гармонического осциллятора может быть произвольной, она пропорциональна квадрату амплитуды колебаний $q_{\alpha}^0 = \text{Re } q_{\alpha}^0$:

$$q_{\alpha}(t) = q_{\alpha}^0 \sin(\omega_{\alpha} t + \varphi_{\alpha}); \quad \dot{q}_{\alpha} = \omega_{\alpha} q_{\alpha}^0 \cos(\omega_{\alpha} t + \varphi_{\alpha}) \quad (58.40)$$

и

$$H_{\alpha} = \frac{\omega_{\alpha}^2 q_{\alpha}^{0^2}}{2} \cos^2(\omega_{\alpha} t + \varphi_{\alpha}) + \frac{\omega_{\alpha}^2 q_{\alpha}^{0^2}}{2} \sin^2(\omega_{\alpha} t + \varphi_{\alpha}) = \frac{\omega_{\alpha}^2 q_{\alpha}^{0^2}}{2}. \quad (58.41)$$

Согласно квантовой механике оператор Гамильтона гармонического осциллятора имеет вид

$$\hat{H}_{\alpha} = \frac{\left(-i\hbar \frac{d}{dq_{\alpha}} \right)^2}{2} + \frac{\omega_{\alpha}^2 q_{\alpha}^2}{2}, \quad (58.42)$$

а уравнение Шредингера

$$\hat{H}_{\alpha} \psi_{\alpha} = E_{\alpha} \psi_{\alpha}, \quad (58.43)$$

где

$$E_{\alpha} = \hbar \omega_{\alpha} \left(v_{\alpha} + \frac{1}{2} \right); \quad v_{\alpha} = 0, 1, 2, \dots \quad (58.44)$$

есть энергия осциллятора, а ψ_{α} — его собственная функция. Поскольку гамильтониан решетки равен сумме независимых \hat{H}_{α} :

$$\hat{H} = \sum_{\alpha} \hat{H}_{\alpha} + U_0, \quad (58.45)$$

то энергия решетки может быть представлена в виде суммы энергии гармонических осцилляторов:

$$E = U_0 + \sum_{\alpha} E_{\alpha} = U_0 + \sum_{\alpha} \hbar \omega_{\alpha} \left(v_{\alpha} + \frac{1}{2} \right). \quad (58.46)$$

Квантовый осциллятор с энергией $E_{\alpha} = \hbar \omega_{\alpha} \left(v_{\alpha} + \frac{1}{2} \right)$ может изменять энергию на величину $\Delta E_{\alpha} = \hbar \omega_{\alpha} \Delta v_{\alpha}$. Как известно из квантовой механики, правило отбора для квантового числа осциллятора имеет вид

$$\Delta v_{\alpha} = \pm 1. \quad (58.47)$$

Если $\Delta v_{\alpha} = -1$, то решетка переходит в одно из более низких энергетических состояний, а энергия $\hbar \omega_{\alpha}$ отдается носителем заряда или окружающим телам. Квант энергии $\hbar \omega_{\alpha}$ получил название кванта энергии колебаний решетки, или фонона. Таким образом, переход $\Delta v_{\alpha} = -1$ можно назвать процессом излучения фонона решеткой, а переход $\Delta v_{\alpha} = +1$ — процессом поглощения фонона решеткой.

Можно подойти к этому вопросу иначе. Будем считать, что энергия колебаний E_{α} локализована в объеме кристалла в виде свободных квазичастиц — фононов — в количестве v_{α} каждого сорта. Фононы образуют фононный газ. В этом случае $\Delta v_{\alpha} = +1$ будет означать рождение фонона, а $\Delta v_{\alpha} = -1$ — уничтожение фонона.

Волновая функция решетки представляется в виде произведения волновых функций осцилляторов

$$\Psi = \prod_{\alpha} \Psi_{\alpha}(\mathbf{q}_{\alpha}), \quad (58.48)$$

где $\Psi_{\alpha}(\mathbf{q}_{\alpha}) = \Psi_{\alpha 1} \Psi_{\alpha 2} \Psi_{\alpha 3}$, и, например,

$$\Psi_{\alpha 1} = e^{-\frac{q_{\alpha 1}^2 \omega_{\alpha}}{2\hbar}} H_{v_{\alpha 1}} \left(\frac{q_{\alpha 1}}{q_{\alpha 0}} \right), \quad (58.49)$$

где $q_{\alpha 0} = \sqrt{\frac{\hbar}{\omega_{\alpha}}}$; $H_{v_{\alpha 1}}$ — полиномы Чебышева — Эрмита.

В заключение этого параграфа укажем, что движение ядер, описываемое в адиабатическом приближении, в нормальных координатах имеет вид гармонических колебаний.

Резюме § 58

1. Хаотическое движение атомов (ионов) решетки описывается вектором смещения \mathbf{u}_{nj} . Поскольку смещение атомов решетки меняет их взаимодействие, то потенциальная энергия решетки зависит от смещения атомов:

$$U = U_0 + \frac{1}{2} \sum_{nj, n'j'} A_{nj, n'j'} \mathbf{u}_{nj} \mathbf{u}_{n'j'} + \dots \quad (58.1p)$$

Сила, действующая на mi -й атом, равна

$$F_{mi} = -\frac{\partial U}{\partial u_{mi}} = -\sum_{nj} A_{mi, nj} u_{nj}, \quad (58.2p)$$

и уравнение движения mi -го атома имеет вид

$$M_i \ddot{u}_{mi} = -\sum_{nj} A_{mi, nj} u_{nj}. \quad (58.3p)$$

2. Если ввести приведенные смещения $w_{mi} = \sqrt{M_i} u_{mi}$ и динамическую матрицу $D_{nj, mi} = \frac{A_{nj, mi}}{\sqrt{M_j M_i}}$, то уравнение (58.3p) переходит в уравнение

$$\ddot{w}_{mi} = -\sum_{nj} D_{mi, nj} w_{nj}. \quad (58.4p)$$

3. Если предположить, что существует некоторая унитарная матрица S , которая переводит динамическую матрицу в диагональную

$$(SDS^{-1})_{\nu\beta} = \omega_\nu^2 \delta_{\nu\beta}, \quad (58.5p)$$

то матрица S позволяет ввести новые переменные q_α :

$$q_\alpha = \sum_{mi} S_{\alpha, mi} w_{mi}, \quad (58.6p)$$

которые удовлетворяют уравнению

$$\ddot{q}_\alpha - \omega_\alpha^2 q_\alpha = 0. \quad (58.7p)$$

Независимые переменные q_α , изменяющиеся во времени по гармоническому закону

$$q_\alpha(t) = q_\alpha^0 e^{-i\omega_\alpha t}, \quad (58.8p)$$

называют нормальными, или главными, координатами решетки кристалла.

4. Полная энергия колебаний решетки, выраженная через нормальные координаты, имеет вид энергии совокупности гармонических осцилляторов

$$H = \sum_{\alpha} H_{\alpha} = \sum_{\alpha} (\dot{q}_{\alpha}^2 + \omega_{\alpha}^2 q_{\alpha}^2) \quad (\text{при } U_0 = 0). \quad (58.9p)$$

В таком случае можно ввести понятие фононов. Фононы — это квазичастицы, энергия которых равна $\hbar\omega_{\alpha}$, а их число равно ν_{α} , так что энергия решетки может быть записана в виде

$$E = \sum_{\alpha} \hbar\omega_{\alpha} \left(\nu_{\alpha} + \frac{1}{2} \right), \quad \text{где } \nu_{\alpha} = 0, 1, \dots \quad (58.10p)$$

Совокупность фононов образует фононный газ, свойства которого адекватно описывают свойства колебаний решетки.

§ 59. АКУСТИЧЕСКИЕ И ОПТИЧЕСКИЕ КОЛЕБАНИЯ РЕШЕТКИ

Вернемся к уравнению движения (58.31). Мы записали его в простой форме в предположении, что существует такая матрица S , с помощью которой можно привести динамическую матрицу к диагональному виду:

$$(SDS^{-1})_{\alpha\beta} = \omega_{\alpha}^2 \delta_{\alpha\beta}, \quad (59.1)$$

или

$$\sum_{nj, ml} S_{\alpha, nj} D_{nj, ml} T_{ml, \beta} = \omega_{\alpha}^2 \delta_{\alpha\beta}. \quad (59.2)$$

Умножим уравнение (59.2) справа на $S_{\beta, kl}$ и просуммируем по β :

$$\begin{aligned} \sum_{\beta} \sum_{nj, ml} S_{\alpha, nj} D_{nj, ml} T_{ml, \beta} S_{\beta, kl} &= \sum_{nj, ml} S_{\alpha, nj} D_{nj, ml} \delta_{ml, kl} =, \\ &= \sum_{nj} S_{\alpha, nj} D_{nj, kl} = \sum_{\beta} \omega_{\alpha}^2 \delta_{\alpha\beta} S_{\beta, kl} = \omega_{\alpha}^2 S_{\alpha, kl}, \end{aligned} \quad (59.3)$$

т. е.

$$\sum_{nj} S_{\alpha, nj} (D_{nj, kl} - \omega_{\alpha}^2 \delta_{nj, kl}) = 0. \quad (59.4)$$

Систему уравнений (59.4) можно рассматривать как систему уравнений относительно элементов неизвестной матрицы S . Эта система уравнений однородна; для того чтобы существовало нетривиальное решение, необходимо, чтобы детерминант системы был равен нулю:

$$|D_{nj, kl} - \omega_{\alpha}^2 \delta_{nj, kl}| = 0. \quad (59.5)$$

Мы видим, что корни этого уравнения определяют набор собственных частот колебаний решетки, они зависят только от упругих свойств решетки и масс ядер.

Рассмотрим в качестве примера простейший случай:

$$D_{nj, ml} = \frac{A_{nj, ml}}{\sqrt{M_j M_l}} \delta_{nj, ml} = \frac{A_{nj, nj}}{M_j} \delta_{nj, ml}. \quad (59.6)$$

Так как в этом случае в определителе все недиагональные элементы равны нулю, то определитель равен произведению диагональных членов:

$$\prod_{nj} \left(\frac{A_{nj, nj}}{M_j} - \omega_{\alpha}^2 \right) = 0, \quad (59.6')$$

откуда

$$\omega_{\alpha}^2 = \frac{A_{nj, nj}}{M_j} \quad (59.7)$$

— обычное выражение для частоты собственных колебаний частицы с массой M_j , на которую действует упругая сила $F_{nj} = -A_{nj, nj} u_{nj}$. Поскольку кристалл обладает периодической трансляционной сим-

метрией, то $A_{nj, nj}$ не зависит от n : $A_{nj, nj} = A_{0j}$ и, следовательно, частоты колебаний всех атомов j одинаковы, но собственные частоты различных атомов j вообще говоря различны. В этом простейшем случае кристалл обладает набором из s различных частот. Можно предполагать на этом основании, что и в общем случае имеется s различных наборов собственных частот.

К уравнению (59.5) можно прийти несколько иначе. Вернемся вновь к уравнениям движения (58.10) и (58.17). Будем искать частное решение этих уравнений в виде гармонического колебания всех атомов с одной и той же частотой:

$$u_{nj}(t) = u_{nj}^0 e^{-i\omega t}, \quad (59.8)$$

$$w_{nj}(t) = w_{nj}^0 e^{-i\omega t}. \quad (59.9)$$

Подставим (59.8) и (59.9) в указанные уравнения, после сокращения на $e^{-i\omega t}$ получим

$$-\omega M_{mi} u_{mi}^0 = - \sum_{nj} A_{mi, nj} u_{nj}^0, \quad (59.10)$$

$$-\omega^2 w_{mi}^0 = - \sum_{nj} D_{mi, nj} w_{nj}^0. \quad (59.11)$$

Мы видим, что уравнение (59.11) в точности совпадает с уравнением (59.4), т. е. амплитуды колебаний w_{nj}^0 удовлетворяют тем же условиям, которым удовлетворяет матрица преобразований S . Но это означает, что матрица S может быть построена из частных решений в виде гармонических колебаний. Раскрывая определитель системы (59.11), найдем все возможные частоты колебаний ω_α , являющиеся собственными частотами кристалла. Из частных решений вида

$$w_{nj, \alpha}(t) = w_{nj, \alpha}^0 e^{-i\omega_\alpha t} \quad (59.12)$$

можно получить общее решение

$$w_{nj}(t) = \sum_{\alpha} w_{nj, \alpha}^0 e^{-i\omega_\alpha t}. \quad (59.13)$$

Но уравнение (59.13) подобно уравнению (58.25). Другими словами, преобразование (58.25) можно рассматривать или как линейное преобразование от «старых» координат к «новым», или представление общего решения уравнения движения как совокупности частных решений в виде гармонических колебаний. Благодаря этому становится более отчетливой аналогия между уравнениями (59.11) и (59.4) для $w_{nj, \alpha}^0$ и $S_{\alpha, nj}$.

Рассмотрим некоторые свойства матрицы S , или частных решений. Построим из $w_{nj, \alpha}^0$ матрицу w^0 , столбцы которой представляют амплитуды и фазы колебаний, соответствующие различным частотам ω_α , для одного и того же атома. Строки же матрицы представляют амплитуды и фазы колебаний различных атомов, соответствующие

ющих одной и той же частоте. Заменяем в уравнении (59.11) индексы m и n на $m+1$ и $n+1$, получим

$$\omega_{\alpha}^2 w_{m+1, i; \alpha}^0 = \sum_{nj} D_{m+1, i; n+1, j} w_{n+1, j; \alpha}^0. \quad (59.14)$$

Но $D_{m+1, i; n+1, j} = D_{m1, nj}$, поэтому уравнение (59.14) можно представить в виде

$$\omega_{\alpha}^2 w_{m+1, i; \alpha}^0 = \sum_{nj} D_{m1, nj} w_{n+1, j; \alpha}^0. \quad (59.15)$$

Отсюда следует вывод, что если элементы некоторого столбца α матрицы w^0 являются решением уравнения ((59.15), то решением того же уравнения является набор величин $[w_{n+1, i; \alpha}^0]_{\alpha}$, который получается из $[w_{nj}^0, \alpha]_{\alpha}$ перестановкой элементов на величину 1 (трансляция на вектор 1). Но поскольку все частные решения найдены, $[w_{n+1, j; \alpha}^0]_{\alpha}$ совпадает с одним из частных решений с точностью до постоянного множителя:

$$[w_{n+1, j; \alpha}^0]_{\alpha} = C_{\alpha}(1) [w_{nj}^0, \alpha]_{\alpha}. \quad (59.16)$$

Величина $C_{\alpha}(1)$ обладает свойствами экспоненты:

$$C_{\alpha}(21) = C_{\alpha}^2(1), \quad (59.17)$$

что вытекает из (59.16). Положим

$$C_{\alpha}(1) = e^{t(K_{\alpha}^1)}. \quad (59.18)$$

Так как 1 может быть любой величиной, то из ограниченности амплитуд колебания следует, что K_{α} вещественный вектор, т. е. $C_{\alpha}(1)$ по модулю равен единице. Это означает, что «трансляция» элементов матрицы на вектор 1 не меняет амплитуд колебаний различных атомов, а меняет только фазу их колебания. Вектор K_{α} обладает некоторыми очевидными свойствами: если добавить к нему вектор $2\pi b$, то условие (59.18) не нарушается, следовательно, вектор K_{α} определяется с точностью до вектора обратной решетки, поэтому все различные векторы K_{α} лежат в некотором ограниченном объеме, совпадающем с зоной Бриллюэна для волнового вектора к электрону. Таким образом, значения K_{α} можно рассматривать в том же пространстве, что и к. Далее очевидно, что условие (59.13) не изменится, если вместо K_{α} взять $(-K_{\alpha})$. Учитывая указанные свойства K_{α} и уравнения (59.16), можно выразить w_{nj}^0 через w_{0j}^0 :

$$w_{nj, \alpha}^0 = w_{0j, \alpha}^0 e^{t(K_{\alpha}^n)}. \quad (59.19)$$

Величина $w_{0j, \alpha}^0$ содержит амплитуду и начальную фазу колебаний всех атомов j с частотой ω_{α} :

$$w_{nj, \alpha}(t) = w_{0j, \alpha} e^{-t[\omega_{\alpha} t - (K_{\alpha}^n)]}. \quad (59.20)$$

Если вместо n подставить r , то (59.20) примет вид

$$w_{rj, \alpha}(t) = w_{j, \alpha}(r, t) = w_{0j, \alpha} e^{-t[\omega_{\alpha} t - (K_{\alpha}^r)]}. \quad (59.21)$$

Выражение (59.21) представляет собой уравнение *плоской гармонической волны частоты ω_α , идущей в направлении волнового вектора K_α* . Фазовая скорость волны $v_\phi = \omega_\alpha / K_\alpha$, групповая скорость $v_{гр} = d\omega_\alpha / dK_\alpha$. Плоской волне с частотой ω_α можно поставить в соответствие квазичастицу с энергией $\hbar\omega_\alpha$ и импульсом $\hbar K_\alpha$. Эта частица представляет собой введенный ранее фонон. Таким образом, совокупности гармонических волн $(\omega_\alpha, K_\alpha)$ соответствует фоновый газ квазичастиц с энергией $\hbar\omega_\alpha$ и квазиимпульсом $\hbar K_\alpha$. Между частотой и волновым вектором должна быть определенная функциональная связь

$$\omega_\alpha = \omega_\alpha(K_\alpha); \quad (59.22)$$

Перепишем уравнение (59.21) с учетом уравнения (59.19):

$$\sum_{nj} D_{m1, nj} w_{nj, \alpha}^0 = \sum_{nj} D_{m1, nj} w_{0j, \alpha}^0 e^{i(K_\alpha n)} = \omega_\alpha^2 w_{01, \alpha}^0 e^{i(K_\alpha m)}. \quad (59.23)$$

Перепишем (59.23) следующим образом:

$$\sum_j \left[\sum_n D_{m1, nj} e^{i(K_\alpha n)} \right] w_{0j, \alpha}^0 - \omega_\alpha^2 w_{01, \alpha}^0 e^{i(K_\alpha m)} = 0. \quad (59.24)$$

Система уравнений (59.24) содержит всего лишь s неизвестных векторных величин $w_{0j, \alpha}^0$. Умножим (59.24) на $e^{-i(K_\alpha m)}$. Введем обозначение $n - m = l$ и перепишем сумму по n в (59.24):

$$\sum_n D_{m1, nj} e^{i(K_\alpha n - m)} = \sum_l D_{m1, m+l, j} e^{i(K_\alpha l)} = \sum_l D_{lj}(l) e^{i(K_\alpha l)} = G_{lj}(K_\alpha), \quad (59.25)$$

после чего уравнение (59.24) примет вид

$$\sum_j G_{lj}(K_\alpha) w_{0j, \alpha}^0 - \omega_\alpha^2 w_{01, \alpha}^0 = 0. \quad (59.26)$$

Однородная система уравнений имеет решение при равном нулю определителе:

$$|G_{lj}(K_\alpha) - \omega_\alpha^2 \delta_{lj}| = 0. \quad (59.27)$$

Определитель (59.27) является уравнением степени $3s$ относительно ω_α^2 . Это означает, что *существует $3s$, вообще говоря, различных зависимостей $\omega_\alpha^2(K_\alpha)$, или $3s$ функций $\omega^2(K)$: $\omega_1^2(K)$; $\omega_2^2(K)$, ... $\omega_{3s}^2(K)$* . Если уравнение (59.5) имеет степень $3Ns$ относительно ω^2 , откуда находятся все возможные частоты, то уравнение (59.27) имеет степень $3s$ относительно ω^2 , но из него находится $3s$ функциональных соотношений $\omega^2(K)$, откуда при известных K находятся все собственные частоты. Найдя все корни уравнения (59.27), можем решить систему уравнений (59.26), подставляя в нее последовательно $\omega_1^2(K)$, ... $\omega_{3s}^2(K)$. Для каждого из этих значений существует решение $w_{01, \alpha}^0(l)$, ... $w_{0j, \alpha}^0(s)$. Мы видим, что существует s различных величин $w_{0j, \alpha}^0 = w_{0j}^0$ для каждого j -атома. Разложим w_{0j}^0 по трем ортам,

один из которых (e_3) коллинеарен вектору K , а два других (e_1 и e_2) лежат в плоскости, перпендикулярной к вектору K . Колебания вдоль e_1 и e_2 являются *поперечными*, а вдоль e_3 — *продольными*. Таким образом, в кристалле имеют место $2s$ типа поперечных волн и s типов продольных волн, всего $3s$ типов волн.

Рассмотрим простейший случай: $s=1$ — в элементарной ячейке содержится один атом с массой M . Величина $G_{1j}(K_\alpha)$ примет вид

$$G(K_\alpha) = D(l) e^{i(K_\alpha l)} = \sum_l \frac{A(l)}{M} e^{i(K_\alpha l)} = \sum_l \frac{A_{n, n+1}}{M} e^{i(K_\alpha l)}. \quad (59.28)$$

Из (59.27) следует, что

$$\omega_\alpha^2 = G(K_\alpha) = \sum_l \frac{A_{n, n+1}}{M} e^{i(K_\alpha l)}. \quad (59.29)$$

Если

$$A_{n, n+1} = \begin{cases} A_{00} & \text{при } l = 0, \\ 0 & \text{при } l \neq 0, \end{cases}$$

то

$$\omega_\alpha^2 = \frac{A_{00}}{M}; \quad \omega_\alpha = \sqrt{\frac{A_{00}}{M}}, \quad (59.30)$$

т. е. частота ω_α не зависит от K_α , следовательно, $\frac{d\omega_\alpha}{dK_\alpha} = 0$ и график $\omega_\alpha(K_\alpha)$ имеет вид прямых, параллельных осям координат. Фазовая скорость при этом будет тем меньше, чем больше K_α , или чем меньше длина волны $\lambda_\alpha = \frac{2\pi}{K_\alpha}$. При $K_\alpha \rightarrow 0$ ($\lambda \rightarrow \infty$) фазовая скорость стремится к ∞ , поэтому все атомы решетки смещаются как одно целое (при малых частотах).

Пусть теперь, кроме A_{00} , отличными от нуля будут члены вида A_{01} . Рассмотрим для простоты линейную цепочку атомов (опустим значок α):

$$\omega^2 = \frac{A_{00}}{M} + \frac{A_{01}}{M} e^{iKa} + \frac{A_{01}}{M} e^{-iKa} = \omega_0^2 \left[1 + \frac{2A_{01}}{A_{00}} \cos Ka \right] = \omega_0^2 [1 + \gamma \cos Ka], \quad (59.31)$$

где $\omega_0^2 = \frac{A_{00}}{M}$; $\gamma = \frac{2A_{01}}{A_{00}}$.

Между различными A_{ij} должна существовать вполне определенная связь. Действительно, пусть, например, все атомы цепочки смещаются как одно целое, так что все $u_{nj} = u_0$, в таком случае

$$F_{mi} = - \left(\sum_{nj} A_{mi, nj} u_{nj} \right) = - \left(\sum_{nj} A_{mi, nj} \right) u_0. \quad (59.32)$$

Но поскольку цепочка атомов смещается как единое целое, то $F_{mi} = 0$, поэтому $\sum_{nj} A_{mi, nj} = 0$, и, следовательно, в данном случае

$$A_{00} = - (A_{01} + A_{0, -1}). \quad (59.33)$$

Из соображений симметрии или из уравнения

$$\begin{aligned} u_m = 0; \quad u_{m-1} = u_0; \quad u_{m+1} = -u_0, \\ F_m = (A_{01} - A_{0,-1}) u_0 = 0 \end{aligned} \quad (59.34)$$

следует, что $A_{01} = A_{0,-1}$ и $A_{00} = 2A_{01}$, поэтому

$$\omega = \omega_0 (1 - \cos Ka)^{\frac{1}{2}} = \omega_0 \sqrt{2} \sin \frac{Ka}{2}. \quad (59.35)$$

При малых K фазовая скорость

$$v_\phi = \frac{\omega}{K} = \sqrt{2} \omega_0 \frac{\sin \frac{Ka}{2}}{K} \approx \frac{\sqrt{2}}{2} \omega_0 a \quad (59.36)$$

и групповая скорость

$$v_{гр} = \frac{d\omega}{dK} = \frac{\sqrt{2}}{2} \omega_0 a = v_\phi, \quad (59.37)$$

т. е. групповая скорость совпадает с фазовой скоростью. Выражению для $v_\phi = v_{гр}$ можно придать некоторый наглядный смысл. Обозначим через $T_n = \frac{A_{00}a}{A}$ — среднее «натяжение» линейной цепочки и через $\rho_M = \frac{M}{a}$ — среднюю плотность вещества; в таком случае для скорости волн можем записать выражение

$$v_{гр} = v_\phi = \sqrt{\frac{A_{00}a^2}{2M}} = \sqrt{\frac{T_n}{\rho_M}}. \quad (59.38)$$

Оно совпадает с выражением для скорости упругих волн, или скорости звука.

Таким образом, при $K \rightarrow 0$ имеем равенство $v_{гр} = v_\phi = c$, где c — скорость упругих волн, или скорость звука.

При $K \approx \pm \frac{\pi}{a}$ групповая скорость обращается в нуль. На рис. 82 представлены зависимости $\omega(K)$, $\frac{\omega(K)}{K}$ и $\frac{d\omega(K)}{dK}$ от волнового числа для рассмотренного случая.

Перейдем к рассмотрению сложных ячеек. Пусть ячейка содержит два атома с массами M_1 и M_2 , с координатами $j_1 = 0$ и $j_2 = 1/2(a, a, a)$. Обозначим коэффициенты упругости A_{ij} в виде $A_{11}(l)$; $A_{12}(l) = A_{21}(l)$ и $A_{22}(l)$ для произвольного l и соответственно обозначим элементы динамической матрицы: D_{11} , $D_{12} = D_{21}$ и D_{22} . Мы можем теперь вычислить G_{11} , $G_{21} = G_{12}^*$ и G_{22} . Запишем, например, выражение (59.25) для G_{12} :

$$G_{12}(K) = \sum_l D_{12}(l) e^{i(Kl)}. \quad (59.39)$$

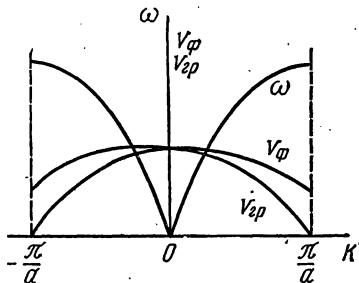


Рис. 82. Зависимость частоты ω , фазовой и групповой скорости волн от волнового числа в линейной цепочке

Составим детерминант (59.27):

$$\begin{vmatrix} G_{11} - \omega^2 & G_{12} \\ G_{21} & G_{22} - \omega^2 \end{vmatrix} = 0. \quad (59.40)$$

Из (59.40) найдем ω^2 :

$$(G_{11} - \omega^2)(G_{22} - \omega^2) - G_{12}G_{21} = 0, \quad (59.41)$$

или

$$\omega^4 - \omega^2(G_{11} + G_{22}) + G_{11}G_{22} - G_{12}G_{21} = 0. \quad (59.42)$$

Решая уравнение (59.42), получим

$$\begin{aligned} \omega^2 &= \frac{G_{11} + G_{22}}{2} \pm \sqrt{\frac{(G_{11} + G_{22})^2}{4} - G_{11}G_{22} + G_{12}G_{21}} = \\ &= \frac{G_{11} + G_{22}}{2} \pm \sqrt{\frac{(G_{11} - G_{22})^2}{4} + G_{12}G_{21}}, \end{aligned} \quad (59.43)$$

т. е. ω^2 принимает два различных значения для каждого K .

Пусть, например, $G_{11} = G_{22}$, тогда

$$\omega_1^2 = G_{11} + |G_{12}|, \quad (59.44)$$

$$\omega_2^2 = G_{11} - |G_{12}|. \quad (59.45)$$

Перепишем систему уравнений (59.26) в следующем виде, опустив нижние значки 0 и α у w_{0j} , α :

$$(G_{11} - \omega^2) w_1^0 + G_{12} w_2^0 = 0, \quad (59.46)$$

$$G_{21} w_1^0 + (G_{22} - \omega^2) w_2^0 = 0.$$

Так как ω^2 имеет два различных значения ω_1^2 и ω_2^2 , то, подставляя их поочередно в (59.46), получим две системы уравнений: при $\omega_1^2 = G_{11} + |G_{12}|$

$$\left. \begin{aligned} -|G_{12}| w_1^0 + G_{12} w_2^0 &= 0, \\ G_{21} w_1^0 - |G_{12}| w_2^0 &= 0; \end{aligned} \right\} \quad (59.47)$$

при $\omega_2^2 = G_{11} - |G_{12}|$

$$\left. \begin{aligned} |G_{12}| w_1^0 + G_{12} w_2^0 &= 0, \\ G_{12} w_1^0 + |G_{12}| w_2^0 &= 0. \end{aligned} \right\} \quad (59.48)$$

Из (59.47) следует, что если $\omega^2 = G_{11} + |G_{12}|$, то

$$w_1^0 = + \frac{G_{12}}{|G_{12}|} w_2^0 \quad (59.49)$$

Из (59.48) аналогично следует, что при $\omega^2 = G_{11} - |G_{12}|$

$$w_1^0 = - \frac{G_{12}}{|G_{12}|} w_2^0 \quad (59.50)$$

Предположим, что G_{12} величина действительная и положительная. В таком случае имеем два типа колебаний

$$\mathbf{w}_1^0 = \mathbf{w}_2^0 \quad (\text{тип } A), \quad (59.51)$$

$$\mathbf{w}_1^0 = -\mathbf{w}_2^0 \quad (\text{тип } O). \quad (59.52)$$

Колебания типа A означают, что смещения обоих атомов одинаковы, ячейка смещается как одно целое. Это приводит к локальным сжатиям и разрежениям в кристалле подобно сжатиям и разрежениям, возникающим при прохождении упругих, или акустических, волн, поэтому колебания, при которых оба атома колеблются в фазе, называются акустическими.

Колебания типа O означают, что атомы смещаются в противоположных направлениях, они колеблются в противофазе, поэтому оставляют центр масс ячейки в покое, но зато в такой ячейке возникает смещение центров тяжести зарядов, и в ячейке возникает дипольный электрический момент. Так как при прохождении электромагнитной волны наблюдается сильное взаимодействие света с колебаниями типа O , то они получили название оптических колебаний.

В общем случае G_{12} может быть комплексной величиной, поэтому представим ее в виде

$$G_{12} = |G_{12}| e^{i\varphi}, \quad (59.53)$$

после чего уравнения (59.49) и (59.50) соответственно примут вид

$$\mathbf{w}_1^0 = \mathbf{w}_2^0 e^{i\varphi} \quad (59.54)$$

и

$$\mathbf{w}_1^0 = -e^{i\varphi} \mathbf{w}_2^0 = e^{i(\varphi+\pi)} \mathbf{w}_2^0. \quad (59.55)$$

Если $\varphi = \pi/2$, то между типами колебаний A и O различие теряется, поскольку и в том и в другом случае сдвиг фаз составляет $\pm \pi/2$.

Чем ближе φ к 0 или π , тем отчетливее проявляется различие между акустическими и оптическими колебаниями решетки, т. е. тем сильнее изменяется распределение заряда или массы в ячейке.

Рассмотрим более общий случай, когда $G_{11} \neq G_{22}$. Подставляя в (59.46) вместо ω^2 значение из (59.43), получим два уравнения, связывающих \mathbf{w}_1^0 и \mathbf{w}_2^0 :

$$\mathbf{w}_2^{0(1)} = \left\{ \frac{G_{11} + G_{22}}{2} + \sqrt{\frac{(G_{11} - G_{22})^2}{4} + |G_{12}|^2} \right\} \frac{1}{G_{12}} \mathbf{w}_1^{0(1)} \quad (59.56)$$

и

$$\mathbf{w}_2^{0(2)} = \left\{ \frac{G_{11} + G_{22}}{2} - \sqrt{\frac{(G_{11} - G_{22})^2}{4} + |G_{12}|^2} \right\} \frac{1}{G_{12}} \mathbf{w}_1^{0(2)}, \quad (59.57)$$

соответствующие плюсу и минусу перед корнем в (59.43). Так же, как и в частном случае $G_{22} = G_{11}$, одно решение соответствует синфазному колебанию атомов ячейки, второе решение — противофазному

колебанию, другими словами, акустической и оптической ветвям колебаний. Рассмотрим зависимость $\omega(K)$ для оптической и акустической ветвей в области малых K (длинных волн).

Чтобы получить зависимость $\omega(K)$, необходимо при вычислении G_{ij} взять, кроме $l=0$, члены с $l \neq 0$. Рассмотрим для простоты одномерную цепочку атомов. Ограничимся простейшим случаем — учтём взаимодействие только соседних атомов, т. е. будем считать, что отличными от нуля являются $A_{11}(0)$; $A_{22}(0)$; $A_{12}(0)$; $A_{21}(0)$; $A_{12}(1)$; $A_{21}(1)$. Для простоты предположим, что $A_{11}(0) = A_{22}(0) = A$; $A_{12}(0) = A_{21}(0) = A_{12}(1) = A_{21}(1) = B$. Между A и B имеется определенная связь; так как

$$\sum_{i,j} A_{ij}(1) = A_{11}(0) + A_{12}(0) + A_{12}(1) = A + 2B = 0, \quad (59.58)$$

то

$$A = -2B. \quad (59.59)$$

Вычислим G_{ij} :

$$\left. \begin{aligned} G_{11} &= \sum_l D_{11} e^{i(Kl)} = \frac{A_{11}(0)}{M_1} = \frac{A}{M_1}; & G_{22} &= \frac{A}{M_2}, \\ G_{12} &= \frac{A_{12}(0)}{\sqrt{M_1 M_2}} + \frac{A_{12}(1)}{\sqrt{M_1 M_2}} e^{i(Ka)} = \frac{B}{\sqrt{M_1 M_2}} (1 + e^{i(Ka)}), \\ G_{21} &= \frac{B}{\sqrt{M_1 M_2}} (1 + e^{-i(Ka)}) = G_{12}^*. \end{aligned} \right\} \quad (59.60)$$

Последнее соотношение в (59.60) записано с учетом того, что вектор l , направленный от первого атома ко второму, противоположен по знаку вектору, направленному от второго атома к первому.

Подставив (59.60) в (59.43), получим

$$\begin{aligned} \omega^2 &= \frac{1}{2} \left(\frac{A}{M_1} + \frac{A}{M_2} \right) \pm \sqrt{\left(\frac{A}{M_1} - \frac{A}{M_2} \right)^2 \frac{1}{4} + \frac{B^2}{M_1 M_2} |1 + e^{i(Ka)}|^2} = \\ &= \frac{A}{2} \left(\frac{M_1 + M_2}{M_1 M_2} \right) \left[1 \pm \sqrt{1 - \frac{4M_1 M_2}{(M_1 + M_2)^2} \sin^2 \frac{Ka}{2}} \right] = \\ &= \omega_0^2 \left[1 \pm \sqrt{1 - \gamma^2 \sin^2 \frac{Ka}{2}} \right], \end{aligned} \quad (59.61)$$

где $\omega_0^2 = \frac{A}{2} \left(\frac{M_1 + M_2}{M_1 M_2} \right)$; $\gamma^2 = \frac{4M_1 M_2}{(M_1 + M_2)^2} \leq 1$; $\omega_0^2 \gamma^2 = \frac{2A}{M_1 + M_2}$.

Рассмотрим зависимость $\omega(K)$ в окрестности $K=0$. Полагая $\sin Ka/2 \approx Ka/2$, получим

$$\omega^2 = \omega_0^2 \left[1 \pm \left(1 - \gamma^2 \frac{K^2 a^2}{4} \right) \right]. \quad (59.62)$$

Для ветви $\omega(K)$, соответствующей знаку минус в (59.62), имеем

$$\omega(K) = \frac{\omega_0 \gamma K a}{2} = \sqrt{\frac{A}{2(M_1 + M_2)}} K a. \quad (59.63)$$

Фазовая v_ϕ и групповая $v_{гр}$ скорости равны и представляют собой скорость звука

$$v_\phi = v_{гр} = \frac{\omega}{K} = \frac{d\omega}{dK} = \frac{\omega_0 \gamma a}{2} = \sqrt{\frac{Aa^2}{2(M_1 + M_2)}} = c. \quad (59.64)$$

Подставляя $\omega = 0$ при $K = 0$ в (59.46), получим

$$\frac{w_2^0}{w_1^0} = \sqrt{\frac{M_2}{M_1}}; \quad \frac{u_1^0}{u_2^0} = 1, \quad (59.65)$$

т. е. атомы колеблются синфазно и с одной и той же амплитудой. Другими словами, ветвь, соответствующая знаку минус перед корнем (59.61), описывает акустические колебания. Плюс в (59.61) описывает оптические колебания:

$$\omega(K) = \sqrt{2} \omega_0 \left(1 - \frac{\gamma^2 a^2 K^2}{16}\right). \quad (59.66)$$

При $K = 0$ частота $\omega(0) = \sqrt{2} \omega_0 = \sqrt{A \left(\frac{M_1 + M_2}{M_1 M_2}\right)}$; с ростом K частота оптических колебаний ω уменьшается по квадратичному закону. Фазовая скорость оптических колебаний стремится к бесконечности при $K \rightarrow 0$, групповая скорость равна нулю при $K = 0$. При $aK/2 = \pi/2$ находится минимальная частота оптических колебаний:

$$\omega^0\left(\frac{\pi}{a}\right) = \omega_0 \sqrt{1 + \sqrt{1 - \gamma^2}} = \omega_{\min}^0 \quad (59.67)$$

и максимальная частота акустической ветви

$$\omega^A\left(\frac{\pi}{a}\right) = \omega_0 \sqrt{1 - \sqrt{1 - \gamma^2}} = \omega_{\max}^A. \quad (59.68)$$

Как видно из общего выражения (59.61), $\omega^O(K) > \omega^A(K)$. Частота акустических колебаний меняется от нуля до ω_{\max}^A , частота оптических колебаний меняется в интервале от $\sqrt{2} \omega_0$ до $\omega_0 \sqrt{1 + \sqrt{1 - \gamma^2}}$. Чем ближе к единице величина γ , тем больше спектральный интервал оптических и акустических колебаний. Групповая скорость $\frac{d\omega}{dK}$ в точке $K = \pm \pi/a$ равна нулю. При $\gamma = 1$ спектры оптических и акустических колебаний сливаются. При $\gamma < 1$ между ними имеется некоторый частотный интервал.

При многих расчетах можно считать, что оптические колебания совершаются примерно с одной и той же частотой $\approx \omega_0$, которая для большинства кристаллов находится в инфракрасной области спектра излучения.

На рис. 83 приведен график зависимости $\omega(K)$ для оптических и акустических ветвей.

Найденные нами свойства оптических и акустических ветвей не зависят от конкретного вида решетки, структура которой лишь влияет на вид функциональной зависимости $\omega(K)$. Наибольшие

максимальные частоты оптических колебаний соответствуют большим длинам волн, поскольку фазовая скорость для оптической ветви обращается в бесконечность при $K \rightarrow 0$, в то время как групповая скорость при этом обращается в нуль. Если максимальная длина волны оптических и акустических ветвей не существует: $\lambda_{\max} = \infty$, то минимальная длина волны существует:

$$\lambda_{\min} = \frac{2\pi}{K_{\max}} = \frac{2\pi}{\pi/a} = 2a, \quad (59.69)$$

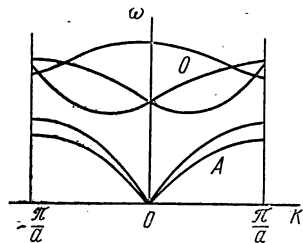


Рис. 83. Оптические и акустические ветви колебаний решетки

что имеет вполне наглядный смысл. Очевидно, что минимальное расстояние в пространстве между точками, состояния колебаний в которых имеют разность фаз 2π , для дискретно распределенных колеблющихся центров не может быть меньше, чем $2a$. Кроме того, дискретность K , вытекающая из периодических граничных условий, имеет тот же смысл — возможны лишь такие колебания, длины волн которых охватывают целое число периодов решетки.

Частоты колебаний решетки лежат в интервале $(0 \div 10^{13})$ Гц. Длинноволновые оптические колебания решетки (колебания с наибольшими частотами) проявляются в спектрах поглощения и отражения кристаллов в области длин волн в несколько десятков микрон.

Резюме § 59

1. Решая уравнение движения (59.8) в предположении, что частое решение имеет вид гармонического колебания с частотой ω , получим для амплитуд уравнение, совпадающее с уравнением матрицы преобразования координат S . Из свойств матриц S и D следует, что колебания соседних атомов происходят с постоянной разностью фаз. Другими словами, гармонические колебания образуют бегущие гармонические волны, идущие в направлении вектора K .

2. Вводя матрицу G , элементы которой представляют собой сумму элементов динамической матрицы, умноженных на некоторый фазовый множитель, сведем систему $3Ns$ уравнений к системе $3s$ уравнений, из решения которой получают $3s$ различных ветвей зависимости $\omega(K)$. Три ветви называются акустическими, остальные $(3s - 3)$ — оптическими.

3. Акустические ветви соответствуют синфазным колебаниям различных атомов элементарной ячейки, когда все атомы совершают колебания с одной и той же амплитудой и фазой. Для окрестности центра зоны Бриллюэна связь между частотой и волновым числом является линейной

$$\omega^A = cK, \quad (59.1p)$$

где c — скорость звука.

4. Оптическим колебаниям соответствуют колебания атомов с противоположными фазами, амплитуды колебаний обратно пропорциональны массам атомов, поэтому центр массы ячейки остается в покое, в то время как смещения атомов приводят к разделению зарядов, вследствие чего в ячейке возникает электрический дипольный момент.

5. Различие между акустическими колебаниями отчетливо проявляется только в окрестности $K=0$, вдали от этой точки различие между оптическими и акустическими колебаниями теряется.

§ 60. ТЕПЛОЕМКОСТЬ РЕШЕТКИ. СТАТИСТИКА ФОНОНОВ

В статистической физике доказывается, что *вероятность нахождения системы в состоянии с энергией E_s равна*

$$\omega_s = \frac{1}{Z} e^{-\frac{E_s}{kT}}, \quad (60.1)$$

где *статистическая сумма Z* определяется условием нормировки

$$\sum_s \omega_s = 1, \quad (60.2)$$

из которого для Z следует выражение

$$Z = \sum_s e^{-\frac{E_s}{kT}}. \quad (60.3)$$

Среднее, или равновесное, значение энергии

$$\langle E \rangle = \sum_s E_s \omega_s. \quad (60.4)$$

Используем выражение (60.4) для нахождения равновесного значения энергии гармонического осциллятора:

$$\begin{aligned} \langle E_\alpha \rangle &= \frac{\sum_{n=0}^{\infty} \hbar \omega_\alpha \left(n + \frac{1}{2} \right) e^{-\frac{\hbar \omega_\alpha \left(n + \frac{1}{2} \right)}{kT}}}{\sum_{n=0}^{\infty} e^{-\frac{\hbar \omega_\alpha \left(n + \frac{1}{2} \right)}{kT}}} = \\ &= \frac{\hbar \omega_\alpha}{2} + \frac{\sum_{n=0}^{\infty} n \hbar \omega_\alpha e^{-\frac{n \hbar \omega_\alpha}{kT}}}{\sum_{n=0}^{\infty} e^{-\frac{n \hbar \omega_\alpha}{kT}}}. \end{aligned} \quad (60.4')$$

Знаменатель (60.4') вычисляется как сумма геометрической прогрессии:

$$\sum_{n=0}^{\infty} \left[e^{-\frac{\hbar\omega_{\alpha}}{kT}} \right]^n = \frac{1}{1 - e^{-\frac{\hbar\omega_{\alpha}}{kT}}}. \quad (60.5)$$

Но так как

$$\sum_{n=0}^{\infty} nq^n = q \sum_{n=0}^{\infty} nq^{n-1} = q \sum_{n=0}^{\infty} \frac{d}{dq} q^n = q \frac{d}{dq} \frac{1}{(1-q)} = \frac{q}{(1-q)^2}, \quad (60.6)$$

то

$$\sum_{n=1}^{\infty} n\hbar\omega_{\alpha} e^{-\frac{n\hbar\omega_{\alpha}}{kT}} = \hbar\omega_{\alpha} \frac{e^{-\frac{\hbar\omega_{\alpha}}{kT}}}{\left(1 - e^{-\frac{\hbar\omega_{\alpha}}{kT}}\right)^2}. \quad (60.7)$$

Учитывая (60.4 – 60.7), запишем выражение для равновесного значения энергии осциллятора:

$$\langle E_{\alpha} \rangle = \frac{\hbar\omega_{\alpha}}{2} + \frac{\hbar\omega_{\alpha} e^{-\frac{\hbar\omega_{\alpha}}{kT}}}{1 - e^{-\frac{\hbar\omega_{\alpha}}{kT}}} = \frac{\hbar\omega_{\alpha}}{2} + \frac{\hbar\omega_{\alpha}}{e^{\frac{\hbar\omega_{\alpha}}{kT}} - 1}, \quad (60.8)$$

что находится в соответствии с функцией распределения Бозе – Эйнштейна. Из (60.8) следует

$$\langle E_{\alpha} \rangle = \hbar\omega_{\alpha} \left(\frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega_{\alpha}}{kT}} - 1} \right). \quad (60.9)$$

Обозначим отношение $\langle E_{\alpha} \rangle$ к $\hbar\omega_{\alpha}$ через $\langle n_{\alpha} \rangle$:

$$\langle n_{\alpha} \rangle = \frac{\langle E_{\alpha} \rangle}{\hbar\omega_{\alpha}} = \frac{1}{2} + \frac{1}{e^{\frac{\hbar\omega_{\alpha}}{kT}} - 1}, \quad (60.10)$$

где $\langle n_{\alpha} \rangle$ есть «среднее» значение колебательного квантового числа осциллятора.

При $T \rightarrow \infty$, когда $\hbar\omega_0 \ll kT$, осциллятор находится в высоком энергетическом состоянии:

$$\langle n_{\alpha} \rangle = \frac{1}{2} + \frac{1}{\frac{\hbar\omega_{\alpha}}{kT}} \approx \frac{kT}{\hbar\omega_{\alpha}}, \quad (60.11)$$

а его средняя энергия равна kT :

$$\langle E_{\alpha} \rangle = \langle n_{\alpha} \rangle \hbar\omega_{\alpha} = kT \quad (60.12)$$

в соответствии с классической статистикой.

При малых температурах, когда $\hbar\omega_\alpha \gg kT$,

$$\langle n_\alpha \rangle = \frac{1}{2} + e^{-\frac{\hbar\omega_\alpha}{kT}} \approx \frac{1}{2}, \quad (60.13)$$

т. е. осциллятор находится в наимизшем колебательном состоянии: $n \approx 0$.

Так как нулевая энергия осциллятора не принимает участия в балансе энергии при взаимодействии осциллятора с окружающей средой, то в дальнейшем ее опустим и будем считать, что *средняя энергия осциллятора равна*

$$\frac{\hbar\omega_\alpha}{e^{\frac{\hbar\omega_\alpha}{kT}} - 1}. \quad (60.14)$$

Воспользовавшись полученными результатами для гармонического осциллятора, найдем среднюю энергию колебаний решетки. С этой целью можем исходить из общего выражения (60.4), учитывая, что

$$E_s = E(v_1, v_2, \dots) = \sum_\alpha \hbar\omega_\alpha \left(v_\alpha + \frac{1}{2} \right) \quad (v_\alpha = 0, 1, 2, \dots) \quad (60.15)$$

Если подставить (60.15) в (60.4), то получим общее выражение для $\langle E \rangle$, которое в силу независимости осцилляторов можно преобразовать в сумму средних значений энергии каждого осциллятора (60.10) или (60.14), так что

$$\langle E \rangle = \sum_\alpha \langle E_\alpha \rangle = \sum_\alpha \frac{\hbar\omega_\alpha}{e^{\frac{\hbar\omega_\alpha}{kT}} - 1} = \sum_{j=1}^{3s} \sum_{K_\alpha} \frac{\hbar\omega_j(K_\alpha)}{e^{\frac{\hbar\omega_j(K_\alpha)}{kT}} - 1}. \quad (60.16)$$

Сумму (60.16) удобно разделить на две суммы — сумму по оптическим ($j > 3$) и сумму по акустическим ($j \leq 3$) ветвям колебаний:

$$\langle E \rangle = \sum_{j=1}^3 \sum_{K_\alpha} \frac{\hbar\omega_j(K_\alpha)}{e^{\frac{\hbar\omega_j(K_\alpha)}{kT}} - 1} + \sum_{j=4}^{3s} \sum_{K_\alpha} \frac{\hbar\omega_j(K_\alpha)}{e^{\frac{\hbar\omega_j(K_\alpha)}{kT}} - 1}. \quad (60.17)$$

Частоты оптических колебаний имеют узкий интервал значений, поэтому $\omega_j(K_\alpha) \approx \omega_j^0$ и

$$\sum_{j=4}^{3s} \sum_{K_\alpha} \frac{\hbar\omega_j^0}{e^{\frac{\hbar\omega_j^0}{kT}} - 1} \approx \sum_{j=4}^{3s} \frac{N\hbar\omega_j^0}{e^{\frac{\hbar\omega_j^0}{kT}} - 1}, \quad (60.18)$$

так как K_α принимает N различных значений.

Введем некоторый характеристический параметр — *температуру Дебая* θ_j^0 — соотношением

$$\hbar\omega_j^0 = k\theta_j^0. \quad (60.19)$$

Если положим $\omega_j^0 = 1,4 \cdot 10^{13} \text{ с}^{-1}$, то для температуры Дебая получим: $\theta_j^0 = 100^\circ \text{ К}$. Для большинства веществ значения θ_j^0 лежат в пределах $100\text{—}400^\circ \text{ К}$. Используя величины θ_j^0 , энергию оптических колебаний можно представить в виде

$$\langle E^0 \rangle = N \sum_{j=4}^{3s} \frac{k\theta_j^0}{e^{\theta_j^0/T} - 1}. \quad (60.20)$$

Если $T \gg \theta_j^0$, то $e^{\theta_j^0/T} = 1 + \theta_j^0/T$ и

$$\langle E^0 \rangle = N \sum_{j=4}^{3s} kT = (3s - 3) NkT, \quad (60.21)$$

т. е. все оптические колебания возбуждены до высоких квантовых состояний. Величину $\langle \bar{n}_\alpha^0 \rangle = \frac{\langle E_\alpha^0 \rangle}{\hbar\omega_\alpha^0}$ можно рассматривать как *число оптических фононов*. При $T \gg \theta_j^0$ в кристалле имеется большое *число оптических фононов*, при $T \ll \theta_j^0$ оптические колебания практически не возбуждены — $\langle n_\alpha^0 \rangle \approx 0$ — и вклад в энергию колебаний решетки они не дают.

В акустической ветви частота меняется от нуля до ω_0^A квази-непрерывно. В этом случае сумму по $\omega_j (K_\alpha)$ можно заменить на интеграл по частотам. Однако при этом необходимо знать плотность состояний по частотам, поскольку в единичном интервале частот ($d\omega = 1$) может содержаться различное число колебаний. Пусть число колебаний в единичном интервале частот равно $g(\omega)$, в таком случае в интервале $d\omega$ их будет

$$dN = g(\omega) d\omega. \quad (60.22)$$

Чтобы найти функцию $g(\omega)$, необходимо знать дисперсионное соотношение $\omega(K)$. Если V — объем кристалла, то в объеме $V \Delta\tau_K \hbar^3$ фазового пространства содержится $\frac{\hbar^3 \Delta\tau_K V}{\hbar^3} = \frac{V \Delta\tau_K}{8\pi^3}$ элементарных ячеек, при этом каждой ячейке соответствует одно возможное колебательное состояние кристалла. Построим в K -пространстве шаровой слой толщиной dK и радиусом K ; объем слоя $4\pi K^2 dK$, ему соответствует $\frac{V 4\pi K^2 dK}{8\pi^3}$ ячеек, на единичный объем кристалла приходится $\frac{4\pi K^2 dK}{8\pi^3}$ ячеек.

Выразим эту величину через частоту: так как $v_{\text{гp}} = d\omega/dK$, то $dK = d\omega/v_{\text{гp}}$. Выражая K через ω , так как $K = K(\omega)$, получим

$$dN = \frac{K^2(\omega) d\omega}{2\pi^2 v_{\text{гp}}} = g(\omega) d\omega. \quad (60.23)$$

Итак, *плотность состояний колебаний решетки* имеет вид

$$g(\omega) = \frac{K^2(\omega)}{2\pi^2 v_{\text{гp}}}. \quad (60.24)$$

В окрестности центра зоны Бриллюэна фазовая и групповая скорости волн акустической ветви равны скорости звука

$$\frac{\omega}{K} = c; \quad \frac{d\omega}{dK} = c; \quad K = \frac{\omega}{c}, \quad (60.25)$$

поэтому

$$g(\omega) = \frac{\omega^2}{2\pi^2 c^3}. \quad (60.26)$$

Таким образом, число колебаний, соответствующих единичному интервалу частот и единичному объему кристалла, определяется соотношением (60.26) для состояний вблизи центра зон Бриллюэна, т. е. для длинных волн. Согласно Дебаю предположим, что соотношение (60.26) справедливо для всего спектра акустических колебаний. Кроме того, учтем, что возможны два типа поперечных и один тип продольных колебаний со скоростями c_t и c_l . В таком случае плотность состояний $g(\omega)$ равна

$$g(\omega) = \frac{\omega^2}{2\pi^2} \left(\frac{2}{c_t^3} + \frac{1}{c_l^3} \right) = \frac{3\omega^2}{2\pi^2 c_0^3}, \quad (60.27)$$

где c_0 , определяемая условием

$$\frac{1}{c_0^3} = \frac{1}{3} \left(\frac{2}{c_t^3} + \frac{1}{c_l^3} \right), \quad (60.28)$$

есть усредненная по направлениям и типам колебаний скорость звука. Запишем теперь энергию акустических колебаний:

$$\langle E^A \rangle = \int_0^{\omega_{\max}^A} \frac{\hbar \omega g(\omega) V d\omega}{e^{\hbar \omega / kT} - 1} = \frac{3V\hbar}{2\pi^2 c_0^3} \int_0^{\omega_{\max}^A} \frac{\omega^3 d\omega}{e^{\hbar \omega / kT} - 1}. \quad (60.29)$$

Положим $\hbar \omega / kT = x$ и перепишем (60.29) в следующем виде:

$$\langle E^A \rangle = \frac{3V (kT)^4}{2\pi^2 c_0^3 \hbar^3} \int_0^{x_{\max}} \frac{x^3 dx}{e^x - 1}. \quad (60.30)$$

Величина x_{\max} , или соответствующая ей предельная частота ω_{\max} , находятся из условия нормировки $g(\omega)$ к полному числу акустических колебаний, равному $3N$:

$$\int_0^{\omega_{\max}^A} g(\omega) d\omega = \frac{3N}{V}. \quad (60.31)$$

Подставляя выражение (60.27) в (60.31), получим

$$\int_0^{\omega_{\max}^A} \frac{3\omega^2}{2\pi^2 c_0^3} d\omega = \frac{\omega_{\max}^3}{2\pi^2 c_0^3} = \frac{3N}{V} = \frac{3}{V_a}; \quad \omega_{\max} = \omega_{\max}^A, \quad (60.32)$$

или

$$\omega_{\max} = \frac{(6\pi^2)^{1/3} c_0}{V_a} = \frac{(6\pi^2)^{1/3} c_0}{a}, \quad (60.33)$$

где a — постоянная решетки, $V_a = a^3 = V/N$ — объем элементарной ячейки.

Оценим максимальное значение волнового вектора:

$$K_{\max} = \frac{\omega_{\max}}{c_0} = \frac{(6\pi^2)^{1/3}}{a} = \frac{7,67}{a}; \quad K_{\max} \sim 10^8 \text{ см}^{-1}. \quad (60.34)$$

Если оценить K_{\max} по минимальной длине волны $\lambda_{\min} = 2a$, то для K_{\max} получим значение $K_{\max} = \pi/a$. Таким образом, интерполяция Дебая приводит к несколько завышенному максимальному значению волнового вектора решетки, поскольку она заменяет зону Бриллюэна сферой радиуса K_{\max} .

Оценим максимальную частоту ω_{\max} :

$$\omega_{\max} = c_0 K_{\max}; \quad \omega_{\max} \approx 10^5 \cdot 10^8 = 10^{13} \text{ с}^{-1}. \quad (60.35)$$

Вместо максимальной частоты ω_{\max} спектр акустических колебаний можно определять *характеристической температурой твердого тела*, или *температурой Дебая* θ :

$$k\theta = h\omega_{\max} = h\omega_{\max}^A; \quad \theta = \frac{\hbar}{k} \left(\frac{6\pi^2}{a^3} \right)^{1/3} c_0. \quad (60.36)$$

При $\omega_{\max} \approx 10^{13} \text{ с}^{-1}$ температура Дебая $\theta = 100^\circ \text{ К}$. Для различных тел температура Дебая различна и составляет величину в несколько сот градусов. Очевидно, что характеристическая температура Дебая не превышает температур Дебая для оптических ветвей: $\theta < \theta_j^0$. Запишем энергию колебаний решетки как функцию температуры:

$$\langle E \rangle = \langle E^A \rangle + \langle E^0 \rangle = \frac{3V (kT)^4}{2\pi^2 c_0^3 \hbar^3} \int_0^{x_M} \frac{x^3 dx}{e^x - 1} + N \sum_{j=4}^{3s} \frac{k\theta_j^0}{e^{\theta_j^0/T} - 1}. \quad (60.37)$$

Учитывая (60.36), можем записать коэффициент перед интегралом из (60.37):

$$\frac{3V (kT)^4}{2\pi^2 c_0^3 \hbar^3} = 9kTN \left(\frac{T}{\theta} \right)^3, \quad (6)$$

после чего можно представить $\langle E \rangle$ в виде

$$\langle E \rangle = NkT \left\{ \frac{9}{\left(\frac{\theta}{T} \right)^3} \int_0^{x_M} \frac{x^3 dx}{e^x - 1} + \sum_{j=4}^{3s} \frac{\frac{\theta_j^0}{T}}{e^{\frac{\theta_j^0}{T}} - 1} \right\}.$$

Рассмотрим теплоемкость твердого тела

$$C = \frac{\partial \langle E \rangle}{\partial T}. \quad (60.41)$$

Выражение для C в общем случае получается достаточно сложным.

Рассмотрим область высоких и низких температур. Когда $T \gg \theta$ и $T \gg \theta_j^0$, оптические колебания дают в этом случае величину, пропорциональную

$$\sum_{j=4}^{3s} \frac{\frac{\theta_j^0}{T}}{e^{\theta_j^0/T} - 1} \approx \sum_{j=4}^{3s} 1 = (3s - 3), \quad (60.42)$$

и акустические колебания дают вклад, пропорциональный значению функции Дебая $D(\theta/T)$:

$$D\left(\frac{\theta}{T}\right) = \frac{3}{\left(\frac{\theta}{T}\right)^3} \int_0^{\theta/T} \frac{x^3 dx}{e^x - 1}. \quad (60.43)$$

Но при больших T , когда $\theta/T \ll 1$ и $x \ll 1$, $e^x \approx 1 + x$, $D(\theta/T) \approx 1$ и, следовательно,

$$\langle E \rangle \approx NkT [3 + (3s - 3)] = 3sNkT \quad (60.44)$$

— на одну степень свободы приходится энергия kT , а теплоемкость решетки не зависит от температуры:

$$C = 3Nsk. \quad (60.45)$$

Для молярной теплоемкости получим величину 6 кал/моль в соответствии с законом Дюлонга — Пти. Он справедлив, пока для металлов $T < F/k = T_{\text{вырожд}}$, а для полупроводников концентрация электронов остается много меньше числа атомов основного вещества в единице объема.

Рассмотрим другой предельный случай: $T \ll \theta$ и, следовательно, $T \ll \theta_j^0$. Это означает, что верхний предел интеграла Дебая можно заменить на ∞ , но

$$\int_0^{\infty} \frac{x^3 dx}{e^x - 1} = \frac{\pi^4}{15}, \quad (60.46)$$

поэтому

$$D\left(\frac{\theta}{T}\right) = \frac{\pi^4}{5} \left(\frac{T}{\theta}\right)^3 \quad (60.47)$$

и

$$\langle E \rangle = \frac{3\pi^4}{5} \frac{kT^4}{\theta^3}. \quad (60.48)$$

Тогда

$$C = \frac{12\pi^4 k}{5} \left(\frac{T}{\theta}\right)^3. \quad (60.49)$$

Закон $C \sim T^3$ хорошо оправдывается в области температур до $20 - 50^\circ \text{K}$. В области промежуточных температур необходимо учитывать как роль акустических, так и оптических колебаний. В этой области теория Дебая недостаточно хорошо совпадает с экспериментальными результатами. Это связано с тем, что происходит возбуждение коротковолновых колебаний, для которых интерполяция Дебая не имеет места. Действительно, рассматривая линейную цепочку типа NaCl и учитывая элементы динамической матрицы, Келлерман путем численных расчетов получил сложную зависимость $\omega(K)$ (или $K(\omega)$), из которой для $g(\omega)$ он построил кривую, изображенную на рис. 84.

В заключение этого параграфа остановимся на двух вопросах: на тепловом расширении решетки и тепловом сопротивлении. Если

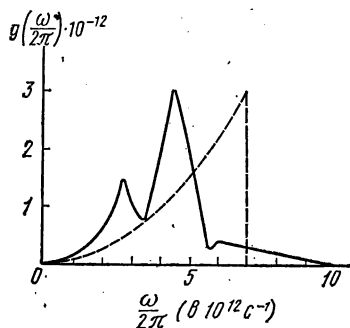


Рис. 84. Плотность колебаний $g(\omega)$ для NaCl по Келлерману (сплошная линия) и Дебаю (пунктирная линия)

рассматривать колебания решетки в гармоническом приближении, то нормальные координаты являются независимыми. Каждой нормальной координате соответствует определенный набор фононов. Они должны быть также независимыми, поэтому в гармоническом приближении фононы не взаимодействуют. Энергия, переносимая фононами, перемещается в решетке с групповой скоростью, поэтому потоку тепла сопротивления нет. Среднее смещение атомов, совершающих гармоническое колебание, равно нулю, и расстояния между атомами не изменяются. Реальные тела испытывают тепловое расширение. Для объяснения этого факта необходимо учесть кубические чле-

ны разложения потенциальной энергии по смещениям ядер. Если учесть тензор $B_{nl, ml, lk}$, то колебания ядер уже не будут гармоническими, кубический член делает их ангармоническими и взаимно зависимыми. Соответствующие нормальным координатам осцилляторы будут также ангармоническими. Ангармонизм осцилляторов превращает фононы во взаимодействующие. Благодаря этому возможно рассеяние фононов на фононах, т. е. происходит нарушение принципа суперпозиции упругих волн. При этом во взаимодействие должны вступать три фонона, что может привести как к рождению нового фонона, так и к уничтожению одного из них (точнее — два фонона превращаются в один). Рассеяние фононов на фононах приводит к появлению теплового сопротивления. Поскольку ангармонизм колебаний проявляется тем меньше, чем меньше амплитуда колебаний, или чем меньше температура, то тепловое сопротивление должно уменьшаться с уменьшением температуры. Ангармонизм колебаний приводит к тому, что среднее смещение атомов отлично от нуля, в результате чего размеры тел изменяются, т. е. имеет место тепловое расширение твердых тел.

Резюме § 60

1. В статистической физике показывается, что вероятность нахождения некоторой системы в состоянии с энергией E_s равна

$$\omega_s = Z^{-1} e^{-\frac{E_s}{kT}}, \quad (60.1p)$$

где статистическая сумма Z определяется из условия нормировки вероятности:

$$\sum_s \omega_s = 1; \quad Z = \sum_s e^{-\frac{E_s}{kT}}. \quad (60.2p)$$

Среднее, или равновесное, значение энергии $\langle E \rangle$ равно

$$\langle E \rangle = \sum_s E_s \omega_s. \quad (60.3p)$$

2. Для гармонического осциллятора равновесное значение энергии согласно (60.3p) равно

$$\langle E_\alpha \rangle = \frac{\hbar\omega_\alpha}{2} + \frac{\hbar\omega_\alpha}{e^{\frac{\hbar\omega_\alpha}{kT}} - 1}. \quad (60.4p)$$

Если считать, что возбуждение осциллятора адекватно рождению фонона, то среднее число фононов определяется уравнением

$$\langle n_\alpha \rangle = \frac{\langle E_\alpha \rangle}{\hbar\omega_\alpha} - \frac{1}{2} = \frac{1}{e^{\frac{\hbar\omega_\alpha}{kT}} - 1}. \quad (60.5p)$$

Из (60.5p) видно, что фононы подчиняются статистике Бозе — Эйнштейна и, следовательно, они имеют целочисленный спин (в единицах \hbar).

3. Итак, фононы — это квазичастицы, соответствующие гармоническим колебаниям решетки кристалла с частотой ω_α и волновым вектором K_α .

Энергия фонона

$$E_\alpha = \hbar\omega_\alpha, \quad (60.6p)$$

квазиимпульс

$$P_\alpha = \hbar K_\alpha. \quad (60.7p)$$

4. Характеристической температурой θ твердого тела, или температурой Дебая, называют величину, определяемую условием

$$\theta = \frac{\hbar\omega_{\max}^A}{k}. \quad (60.8p)$$

5. Температурой Дебая для оптических колебаний решетки называют величину θ_j^o , определяемую условием

$$\theta_j^o = \frac{\hbar\omega_j^o}{k}. \quad (60.9p)$$

6. Число оптических фононов при температуре T равно

$$n_j^o = \frac{1}{\frac{\theta_j^o}{T} - 1}. \quad (60.10p)$$

7. Число акустических фононов зависит от их частоты и температуры в соответствии с (60.5p). С ростом температуры число длинноволновых фононов растет быстрее числа коротковолновых фононов. При больших температурах $T \gg \theta$ энергия, переносимая фононами, не зависит от их частоты и равна kT на каждую степень свободы.

8. Теплоемкость твердого тела при $T \ll \theta$ пропорциональна T^3 , при $T \gg \theta$ она не зависит от температуры.

9. Ангармонизм колебаний решетки приводит к тепловому расширению тел и тепловому сопротивлению, т. е. к рассеянию фононов на фононах.

§ 61. РАССЕЯНИЕ НА ТЕПЛОВЫХ КОЛЕБАНИЯХ РЕШЕТКИ. МЕТОД ПОТЕНЦИАЛА ДЕФОРМАЦИИ

Используя понятие фононов, можно объяснить рассеяние электронов или дырок на колебаниях решетки на языке корпускулярных представлений. Носители заряда, сталкиваясь с фононом, обмениваются с ним энергией и квазиимпульсом. Поскольку число фононов определяется температурой, то и рассеяние носителей заряда должно зависеть от температуры.

Однако, чтобы воспользоваться общей теорией квантовых переходов, изложенной в § 55, необходимо найти возмущение, порождаемое колебаниями решетки, являющееся причиной перехода электронов и дырок из одного состояния в другое. Для нахождения возмущения воспользуемся методом потенциала деформации.

Гармоническая волна (ω , K) приводит к локальным деформациям кристалла. Смещение, создаваемое упругой волной в точке \mathbf{r} в момент t , имеет вид

$$\mathbf{u}(\mathbf{r}, t) = u_0 e^{-i\left[\omega t - (K\mathbf{r}) + \frac{\pi}{2}\right]}, \quad (61.1)$$

где u_0 — амплитуда колебаний, фаза $\pi/2$ введена для упрощения вычислений.

Деформация \mathbf{u} приводит к относительному изменению объема Δ , равному согласно (51.17) $\text{div } \mathbf{u}$:

$$\Delta = \text{div } \mathbf{u} = i(Ku_0) e^{-i\left[\omega t - (K\mathbf{r}) + \frac{\pi}{2}\right]} = (Ku_0) e^{-i[\omega t - (K\mathbf{r})]}. \quad (61.2)$$

Как видно из (61.2), в кубических кристаллах *изменение объема наблюдается только при прохождении продольных волн*. В анизотропных кристаллах происходит изменение объема при прохождении любых волн. Найдем тензор деформации, равный согласно (51.3)

$$u_{mn} = \frac{1}{2} \left(\frac{\partial u_m}{\partial x_n} + \frac{\partial u_n}{\partial x_m} \right). \quad (61.3)$$

Поскольку m -компонент смещения \mathbf{u} определяется m -компонентом u_0 , а дифференцирование (61.1) по r_m приводит к появлению множителя K_m , то получим

$$u_{mn} = \frac{1}{2} e^{-i[\omega t - (\mathbf{K}\mathbf{r})]} (u_{0m}K_n + u_{0n}K_m). \quad (61.4)$$

Направим ось z вдоль вектора \mathbf{K} , вследствие чего $\mathbf{K} = (0, 0, K)$. Для продольной волны $\mathbf{u}_0 = (0, 0, u_0)$; для поперечных волн в зависимости от поляризации: $\mathbf{u}_0 = (u_0, 0, 0)$ или $\mathbf{u}_0 = (0, u_0, 0)$. Для тензора деформации получим

$$\mathbf{u} = u_0 K e^{-i[\omega t - (\mathbf{K}\mathbf{r})]} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (61.5)$$

$$\mathbf{u} = u_0 K e^{-i[\omega t - (\mathbf{K}\mathbf{r})]} \cdot \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (61.6)$$

$$\mathbf{u} = u_0 K e^{-i[\omega t - (\mathbf{K}\mathbf{r})]} \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (61.7)$$

при этом (61.5) относится к *продольной волне*, (61.6) и (61.7) — к *поперечной волне* с колебаниями вдоль осей x и y .

Поскольку *поперечные волны представляют деформацию сдвига*, они не приводят к изменению объема, как это видно из (61.6) и (61.7). Отметим, кстати, что поскольку продольные волны представляют собой деформацию сжатия и растяжения, а поперечные — деформацию сдвига, то они должны иметь различную скорость, так как модуль Юнга и модуль сдвига в общем случае не равны друг другу. *Локальная деформация, создаваемая волной, приводит к смещению дна зоны проводимости и потолка валентной зоны.*

Согласно (52.13) запишем

$$E_c(\mathbf{u}) = E_c(0) + \sum_{ik} \Delta_{ik}^{(c)} u_{ik} = E_c(0) + V_c(\mathbf{r}, t), \quad (61.8)$$

$$E_v(\mathbf{u}) = E_v(0) + \sum_{ik} \Delta_{ik}^{(v)} u_{ik} = E_v(0) + V_v(\mathbf{r}, t), \quad (61.9)$$

где V_c и V_v представляют собой потенциалы деформации для зоны проводимости и валентной зоны. Обозначим элемент $\Delta_{33}^{(c)}$ через Δ_c ,

а элемент $\Delta_{13}^c = \Delta_{31}^c = \Delta_c'$. Из соображений симметрии запишем $\Delta_{23}^c = \Delta_{32}^c = \Delta_c'$. Проведем вычисления для продольной волны и зоны проводимости:

$$V_c(\mathbf{r}, t) = \hat{W}(\mathbf{r}, t) = u_0 K \Delta_c \cdot e^{-i[\omega t - (\mathbf{K}\mathbf{r})]}. \quad (61.10)$$

Вычислим матричный элемент оператора возмущения (61.10) с помощью волн Блоха в приближении эффективной массы и при нормировке в ящике объемом $G = L^3$:

$$\psi_{\mathbf{k}}(\mathbf{r}, t) = \psi_{\mathbf{k}}(\mathbf{r}) e^{-i \frac{E(\mathbf{k})t}{\hbar}} = \frac{1}{\sqrt{G}} e^{-i \left[\frac{Et}{\hbar} - (\mathbf{k}\mathbf{r}) \right]}. \quad (61.11)$$

Запишем

$$\begin{aligned} W_{\mathbf{k}'\mathbf{k}} &= V_{c\mathbf{k}'\mathbf{k}} = \frac{\Delta_c u_0 K}{G} e^{-\frac{i}{\hbar} (E - E' + \hbar\omega)t} \int_{(G)} e^{i(\mathbf{k} + \mathbf{K} - \mathbf{k}', \mathbf{r})} d\tau = \\ &= \Delta_c u_0 K e^{-\frac{i}{\hbar} (E - E' + \hbar\omega)t} \delta_{\mathbf{k} + \mathbf{K}, \mathbf{k}'}. \end{aligned} \quad (61.12)$$

При вычислении интеграла в (61.12) мы учли условие ортонормировки собственных функций $\psi_{\mathbf{k}}(\mathbf{r})$. Для вероятности перехода из состояния \mathbf{k} в состояние \mathbf{k}' согласно (55.23) имеем

$$\omega(\mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar} \Delta_c^2 u_0^2 K^2 \delta(E + \hbar\omega - E') \delta_{\mathbf{k} + \mathbf{K}, \mathbf{k}'}. \quad (61.13)$$

В вероятность перехода входят две δ -величины, обеспечивающие законы сохранения энергии и квазиимпульса. Именно, вероятность отлична от нуля только в том случае, если

$$E' = E + \hbar\omega; \quad \hbar\mathbf{k}' = \hbar\mathbf{k} + \hbar\mathbf{K}, \quad (61.14)$$

которые означают, что энергия и квазиимпульс электрона меняются на энергию и квазиимпульс фонона. Если учесть, что в (61.10) в показателе экспоненты можно взять как минус, так и плюс, то (61.14) можно записать в более общем виде:

$$E' = E \pm \hbar\omega; \quad \hbar\mathbf{k}' = \hbar\mathbf{k} \pm \hbar\mathbf{K}, \quad (61.15)$$

т. е. энергия электрона увеличивается или уменьшается на величину энергии фонона, другими словами электрон или поглощает, или порождает фонон.

Прежде чем переходить к дальнейшим вычислениям, отметим, что соотношения (61.15) могут быть истолкованы на языке волновых представлений:

$$\frac{E'}{\hbar} = \frac{E}{\hbar} \pm \omega; \quad \mathbf{k}' = \mathbf{k} \pm \mathbf{K} \quad (61.16)$$

— частота $\frac{E}{\hbar}$ и волновой вектор \mathbf{k} электронной волны меняются на величину $\pm \omega$ и $\pm \mathbf{K}$ вследствие эффекта Доплера при рассеянии на движущемся объекте — области деформации.

Выражение (61.13) иллюстрируется рис. 85, на котором изображены две изоэнергетические поверхности E и $E + \hbar\omega$. При заданном квазимпульсе фонона $\hbar K$ найдутся такие две точки A и B на поверхности энергии E и E' , которые должны отстоять друг от друга на расстоянии K , так что конец вектора K лежит на поверхности E' , а начало — на поверхности E . Таких точек может быть несколько. Переход $A \rightarrow B$ соответствует рассеянию с поглощением фонона. Точки A' , B' соответствуют рассеянию электронов с излучением фонона. Обозначим угол между k и K через α и выразим соотношение между k и K через него, учитывая (61.15):

$$\begin{aligned}
 E' &= E(k') = \frac{\hbar^2 k'^2}{2m^*} = \frac{\hbar^2}{2m^*} (k \pm K)^2 = \\
 &= \frac{\hbar^2}{2m^*} (k^2 + K^2 \pm 2kK \cos \alpha) = E(k) \pm \hbar\omega = \\
 &= \frac{\hbar^2 k^2}{2m^*} \pm \hbar\omega. \quad (61.17)
 \end{aligned}$$

Сокращая на $\frac{\hbar^2 k^2}{2m^*}$ обе части равенства, подставляя вместо ω величину cK , что справедливо для длинноволновой части акустической ветви, получим:

$$K = \pm \frac{2m^*c}{\hbar} \mp 2k \cos \alpha. \quad (61.18)$$

Так как $c \approx 10^5$ см/с, то при $\frac{2m^*}{\hbar} \sim 1$ первый член в (61.18) равен $\sim 10^5$ см $^{-1}$. Второй член зависит от угла α , при малых углах $\cos \alpha \sim 1$, так что $K \approx \pm 10^5 \mp 2k$. Выразим величину k через температуру из соотношения

$$\frac{\hbar^2 k^2}{2m^*} = \frac{3}{2} kT, \quad (61.19)$$

откуда

$$k = \frac{1}{\hbar} \sqrt{3m^*kT}. \quad (61.20)$$

Полагая $m^* \sim 10^{-27}$ г, получим $k \sim 6 \cdot 10^5 \sqrt{T}$ см $^{-1}$. Для того чтобы электрон имел $k \sim 10^5$ см $^{-1}$, достаточно, чтобы температура тела была $\approx 1^\circ\text{K}$. Другими словами, первый член в (61.18) во всех практически важных случаях можно отбросить, поэтому запишем

$$K \approx \mp 2k \cos \alpha. \quad (61.21)$$

В зависимости от α квазимпульс фонона K может меняться от 0 до $2k$. При $T = 300^\circ\text{K}$ $k \sim 10^7$ см $^{-1}$ и, следовательно, K может принимать значение от 0 до $2 \cdot 10^7$ см $^{-1}$, что соответствует энергии фононов от 0 до $\hbar cK \approx 10^{-27} \cdot 10^5 \cdot 10^7 = 10^{-15}$ эрг = $6 \cdot 10^{-4}$ эВ.

Мы видим, что основную роль в рассеянии играют длинноволновые фононы с энергией, значительно меньшей энергии носителей

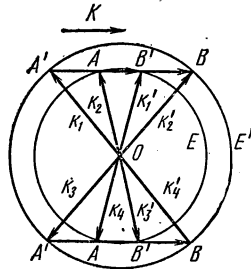


Рис. 85. Переходы частиц при выполнении законов сохранения энергии и квазимпульса

заряда, поэтому можно считать, что рассеяние носителей заряда на фононах упругое. В таком случае можно отбросить энергию фонона в выражении для $\omega(\mathbf{k}, \mathbf{k}')$.

Учитывая (61.21), преобразуем δ -функцию от энергии к виду

$$\begin{aligned} \delta(E \pm \hbar\omega - E') &\approx \delta(E - E') = \delta\left(\frac{\hbar^2\kappa^2}{2m^*} - \frac{\hbar^2\kappa'^2}{2m^*}\right) = \\ &= \frac{2m^*}{\hbar^2} \delta(\kappa'^2 - \kappa^2) = \frac{2m^*}{\hbar^2} \delta(K^2 \pm 2\kappa K \cos \alpha) = \frac{m^*}{\hbar^2 \kappa K} \delta\left(\frac{K}{2\kappa} \pm \cos \alpha\right), \end{aligned} \quad (61.22)$$

после чего для вероятности перехода (61.13) будем иметь

$$\omega(\mathbf{k}, \mathbf{k}') = \frac{2\pi m^* \Delta_c^2 u_0^2 K}{\hbar^3 \kappa} \delta\left(\frac{K}{2\kappa} \pm \cos \alpha\right). \quad (61.23)$$

Чтобы вычислить время релаксации, необходимо учесть рассеяние на фононах всех возможных значений K . Другими словами, в общем случае необходимо представить смещение атомов решетки в виде суперпозиции нормальных колебаний:

$$\mathbf{u}(\mathbf{r}, t) = \sum_{\alpha} \mathbf{u}_{0\alpha} e^{-i\left[\omega_{\alpha} t - (\mathbf{K}_{\alpha} \mathbf{r}) - \frac{\pi}{2}\right]}. \quad (61.24)$$

Для потенциала деформации запишем

$$V_c = \Delta_c \sum_{\alpha} u_{0\alpha} K_{\alpha} e^{-i\left[\omega_{\alpha} t - (\mathbf{K}_{\alpha} \mathbf{r})\right]}. \quad (61.25)$$

Найдя потенциал деформации (61.25), вычислим теперь матричный элемент оператора возмущения, после чего можем записать выражение для вероятности перехода $\omega(\mathbf{k}, \mathbf{k}')$. В общем случае квадрат суммы не равен сумме квадратов, но необходимо учесть, что при возведении в квадрат матричного элемента в общем выражении для вероятности будут члены с произведениями δ -величин, которые при суммировании выпадут, в результате чего под знаком одинарной суммы останутся только квадраты матричных элементов гармонических составляющих смещения (61.24).

Физически этот результат очевиден. Так как гармонические волны, пропорциональные нормальным координатам, независимы, процессы рассеяния на гармонических волнах являются независимыми и для них складываются вероятности, поэтому какие-либо интерференционные эффекты отсутствуют. Интегрируя (61.23) по всем \mathbf{k}' с весом $(1 - \cos \theta)$, найдем $\tau^{-1}(\mathbf{k})$:

$$\frac{1}{\tau(\mathbf{k})} = \frac{1}{4\pi^3} \int_{(V_{\mathbf{k}})} \omega(\mathbf{k}, \mathbf{k}') [1 - \cos \theta] d\tau_{\mathbf{k}'}. \quad (61.26)$$

Для вычисления интеграла (61.26) удобнее перейти к переменной K . Так как $\mathbf{k}' = \mathbf{k} + \mathbf{K}$, то $K = k' - k$ и $d\tau_{\mathbf{k}'} = d\tau_{\mathbf{K}}$. Поскольку мы рассматриваем упругое рассеяние, векторы \mathbf{k} и \mathbf{k}' лежат на одной

поверхности энергии (на одной сфере). Из рис. 86 видно, что для того чтобы вектор \mathbf{K} принял все возможные значения при фиксированной энергии, вектор \mathbf{k} должен принять значения, лежащие на сфере радиуса κ с началом координат в точке $(-\mathbf{k}')$. Обозначим угол между \mathbf{k}' и \mathbf{K} через α , между \mathbf{K}' и \mathbf{k} — через θ . Угол θ меняется от 0 до π , при этом угол α меняется от $\pi/2$ до π , $|\mathbf{K}|$ меняется от нуля до $2|\mathbf{k}|$. Найдем соотношение между α и θ . Так как

$$\cos \alpha = \frac{(\mathbf{k}\mathbf{K})}{\kappa K}, \text{ а } \mathbf{K} = \mathbf{k}' - \mathbf{k}, \text{ то}$$

$$\begin{aligned} \cos \alpha &= \frac{(\mathbf{k}'\mathbf{k}) - \kappa^2}{\kappa K} = \frac{\kappa' \kappa \cos \theta - \kappa^2}{\kappa K} = \\ &= -\frac{\kappa}{K} (1 - \cos \theta), \end{aligned} \quad (61.27)$$

или

$$1 - \cos \theta = -\frac{K}{\kappa} \cos \alpha. \quad (61.28)$$

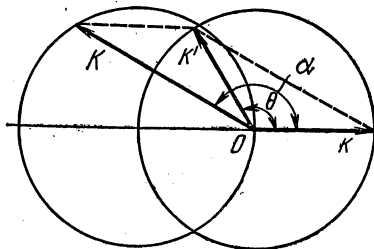


Рис. 86. Связь между углами α и θ

Учитывая (61.28), (61.23), перепишем (61.26) в следующем виде:

$$\begin{aligned} \frac{1}{\tau(\kappa)} &= -\frac{1}{4\pi^3} \frac{2\pi m^* \Delta_c^2}{\hbar^3 \kappa} \int_0^{2\kappa} u_0^2(K) K K^2 dK \int_{\frac{\pi}{2}}^{\pi} \delta\left(\frac{K}{2\kappa} \pm \cos \alpha\right) \times \\ &\times \frac{K}{\kappa} \cos \alpha \sin \alpha d\alpha \int_0^{2\pi} d\varphi. \end{aligned} \quad (61.29)$$

Вычислим интеграл по α :

$$-\int_{\frac{\pi}{2}}^{\pi} \delta\left(\frac{K}{2\kappa} \pm \cos \alpha\right) \cos \alpha d \cos \alpha = \int_0^{-1} \delta\left(\frac{K}{2\kappa} \pm x\right) x dx = \frac{K}{2\kappa} \quad (61.30)$$

и

$$\frac{1}{\tau(\kappa)} = \frac{m^* \Delta_c^2}{4\pi^2 \hbar^3 \kappa^3} \int_0^{2\kappa} u_0^2(K) K^5 dK. \quad (61.31)$$

Чтобы вычислить интеграл (61.31), необходимо знать амплитуды различных колебаний. Для этого рассмотрим *распределение осцилляторов по состояниям колебаний*. При больших температурах все осцилляторы находятся в высоких колебательных состояниях, энергия колебаний совпадает при этом с энергией классического осциллятора:

$$\frac{1}{2} M \omega^2 u_0^2 = kT, \quad (61.32)$$

откуда

$$u_0^2 = \frac{2kT}{M\omega^2} = \frac{2kT}{Mc^2K^2} \quad (61.33)$$

и

$$\frac{1}{\tau(k)} = \frac{m^* \Delta_c^2}{4\pi^2 \hbar^3 \kappa^3} \int_0^{2\kappa} \frac{2kT \cdot K^5}{Mc^2 K^2} dK = \frac{4m^* \Delta_c^2 k T \kappa^4}{\pi^2 M \hbar^3 \kappa^3 c^2} = \frac{4m^* \Delta_c^2 k T}{\pi^2 M \hbar^3 c^2} \kappa, \quad (61.34)$$

т. е.

$$\tau(k) = \frac{\pi^2 M c^2 \hbar^3}{4m^* \Delta_c^2 k T} \frac{1}{\kappa} = \frac{\pi^2 M c^2 \hbar^4}{4\sqrt{2} (m^*)^2 \Delta_c^2} \frac{E^{-\frac{1}{2}}}{kT}. \quad (61.35)$$

Зависимость времени релаксации от энергии имеет вид

$$\tau(k) = \tau(E) = \frac{\tau_0}{kT} E^p \quad \text{при } p = -\frac{1}{2}. \quad (61.36)$$

Длина свободного пробега в атомных кристаллах при рассеянии на акустических колебаниях не зависит от энергии носителей заряда: $l \approx E^0$. Если определить c соотношением $c^2 = \frac{c_{11}}{\rho}$ и положить $M = \frac{\rho}{N}$, то

$$Mc^2 = \frac{\rho}{N} \frac{c_{11}}{\rho} = \frac{c_{11}}{N} \quad \text{и}$$

$$\tau(E) = \frac{\pi^2 \hbar^4 c_{11}}{2 (2m^*)^2 \Delta_c^2 N} \frac{E^{-\frac{1}{2}}}{kT}. \quad (61.37)$$

Таким образом, время релаксации обратно пропорционально концентрации N атомов основного вещества и связано с модулем упругости c_{11} .

Для описания процесса рассеяния носителей заряда на оптических колебаниях используется так называемый *поляризационный потенциал*. Он вводится следующим образом. Смещение атомов в ячейке происходит в противофазе. Разделение зарядов порождает электрическое поле (поляризацию вещества), которое перемещается в пространстве в виде плоской волны. Взаимодействие носителей заряда с этой волной приводит к их рассеянию. Чтобы определить вероятность рассеяния, необходимо найти матричный элемент возмущения. Если вычислить его с помощью плоских электронных волн в приближении эффективной массы, то легко сделать вывод, что при рассеянии должны выполняться законы сохранения

$$\hbar k' = \hbar k \pm \hbar K; \quad E' = E \pm \hbar \omega. \quad (61.38)$$

Поскольку энергия оптических колебаний достаточно велика, необходимо учитывать величину $\hbar\omega$. Кроме того, при достаточно низких температурах число оптических фононов будет невелико, поэтому u_0^2 нельзя заменить на kT , u_0^2 пропорционально числу оптических фононов:

$$\langle n^0 \rangle = \frac{1}{e^{\frac{\hbar\omega^0}{kT}} - 1}, \quad (61.39)$$

поэтому время релаксации при рассеянии на оптических колебаниях должно зависеть от температуры в виде

$$\tau \sim \left[e^{\frac{\hbar\omega^0}{kT}} - 1 \right]. \quad (61.40)$$

При низких температурах это приводит к зависимости

$$\tau \sim e^{\frac{\hbar\omega^0}{kT}}, \quad (61.41)$$

а при больших температурах ($\hbar\omega^0 \ll kT$) число фононов пропорционально kT , так же как и для акустических фононов, но при этом τ от энергии носителей заряда не зависит.

В реальных веществах подвижность определяется как рассеянием на акустических, так и оптических фононах.

В заключение этого параграфа укажем еще на один механизм рассеяния, который имеет место в *многодолинных полупроводниках*. Поскольку волновой вектор решетки K меняется от 0 до $2\pi/\omega$, то в результате поглощения фонона *носитель заряда может перейти из одной долины в другую*. Такие процессы носят название процессов *переброса*.

Резюме § 61

1. Упругая волна, создавая деформацию кристалла, приводит к смещению зон энергии. Величины V_c и V_v , определяющие смещение зон энергии, носят название потенциалов деформации. Они могут быть выражены через тензор деформации u и тензор коэффициентов потенциала деформации, например для зоны проводимости:

$$V_c(\mathbf{r}, t) = \sum_{ik} \Delta_{ik} u_{ik}. \quad (61.1p)$$

Тензор деформации выражается через смещение (61.2) в виде (61.4).

2. Вероятность перехода $\omega(\mathbf{k}, \mathbf{k}')$ носителей заряда, вызываемого потенциалом деформации, имеет вид

$$\omega(\mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar} \Delta_c^2 u_0^2 K^2 \delta(E \pm \hbar\omega - E') \delta_{\mathbf{k}+\mathbf{k}', \mathbf{k}'}, \quad (61.2p)$$

т. е. она отлична от нуля только в том случае, когда энергия электронов и дырок меняется на энергию и квазимпульс фононов. Уч-

тывая одновременное выполнение законов сохранения энергии и квазиимпульса, выражение (61.2) можно преобразовать к виду

$$\omega(\mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar^3} \frac{m^* \Delta_c^2 u_0^2 K^2}{\kappa} \delta\left(\frac{K}{2\kappa} \pm \cos \alpha\right), \quad (61.3p)$$

где α — угол между \mathbf{k} и \mathbf{K} .

3. Учитывая, что процессы рассеяния на нормальных координатах являются независимыми, суммируя (61.3), получим $\tau^{-1}(\mathbf{k})$; переходя к интегрированию по α и K , получим

$$\frac{1}{\tau(\mathbf{k})} = \frac{m^* \Delta_c^2}{4\pi^2 \hbar^3 \kappa^3} \int_0^{2\kappa} u_0^2 K^5 dK. \quad (61.4p)$$

Чтобы вычислить интеграл в (61.4), необходимо знать распределение амплитуд колебаний u_0 по частотам или по K , т. е. необходимо знать число фононов определенной частоты. В общем случае число фононов определяется функцией Бозе — Эйнштейна (60.10). Поскольку при любой конечной температуре некоторая часть осцилляторов возбуждена до высоких квантовых состояний, когда число фононов равно $kT/\hbar\omega$, то, выражая потенциальную энергию колебаний через энергию фононов, получим

$$\frac{1}{2} M \omega_\alpha^2 u_0^2 = \hbar \omega_\alpha \frac{kT}{\hbar \omega_\alpha} = kT. \quad (61.5p)$$

С точки зрения классической статистической физики условие (61.5p) означает, что на каждую степень свободы приходится энергия kT . Для $\tau(\mathbf{k})$ получим

$$\tau(\mathbf{k}) = \tau(E) = \frac{\pi^2 \hbar^4 M c^2}{2 (2m^*)^2 \Delta_c^2} \frac{E^{-1/2}}{kT} = \frac{\pi^2 \hbar^4 c_{II}}{2 (2m^*)^2 \Delta_c^2 N} \frac{E^{-1/2}}{kT}. \quad (61.6p)$$

4. Если выразить транспортное эффективное сечение через время релаксации, то можно сказать, что транспортное эффективное сечение рассеяния на тепловых колебаниях решетки не зависит от энергии частиц. Рассеяние на тепловых колебаниях решетки изотропно.

5. Рассеяние на оптических колебаниях решетки может быть описано введением поляризационного потенциала.

§ 62. ЗАВИСИМОСТЬ ПОДВИЖНОСТИ НОСИТЕЛЕЙ ЗАРЯДА ОТ ТЕМПЕРАТУРЫ

Результаты предыдущих параграфов позволяют рассмотреть температурную зависимость подвижности носителей заряда

$$\mu = \frac{e \langle \tau \rangle}{m^*} = \mu(T). \quad (62.1)$$

При изменении температуры может измениться эффективная масса. Простейшей, но не единственной причиной этого изменения явля-

ется тепловое расширение решетки. Однако мы в дальнейшем не будем учитывать изменение эффективной массы, а ограничимся рассмотрением температурной зависимости усредненного времени релаксации. Согласно (40.4)

$$\langle \tau \rangle = \langle E\tau \rangle = \frac{\int_0^{\infty} x^{\frac{3}{2}} e^{-x\tau} (x) dx}{\int_0^{\infty} x^{\frac{3}{2}} e^{-x} dx}, \quad (62.2)$$

где $x = \frac{E}{kT}$. Вычисления проводятся элементарно, если время релаксации имеет степенную зависимость от энергии, т. е.

$$\tau(E) = \tau_0 E^p = \tau_0 (kT)^p x^p, \quad (62.3)$$

где p — показатель степени, а τ_0 — величина, не зависящая от энергии. Подставляя (62.3) в (62.2), получим

$$\langle \tau \rangle = \tau_0 (kT)^p \frac{\int_0^{\infty} x^{\frac{3}{2} + p} e^{-x} dx}{\int_0^{\infty} x^{\frac{3}{2}} e^{-x} dx} = \tau_0 (kT)^p \frac{\Gamma\left(p + \frac{5}{2}\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (62.4)$$

Отношение Г-функций при некоторых p дано в табл. 8. Подставляя (62.4) в (62.1), запишем

$$\mu = \langle \mu \rangle = \frac{e\tau_0 k^p}{m^*} \frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)} T^p. \quad (62.4')$$

Если τ_0 не зависит от температуры, то температурная зависимость подвижности подобна зависимости времени релаксации от энергии.

Перейдем к рассмотрению конкретных механизмов рассеяния электронов и дырок.

1. Рассеяние на тепловых колебаниях решетки. Учитывая, что при рассеянии на акустических колебаниях решетки $p = -1/2$, согласно (62.4) и (61.37) запишем

$$\mu(T) = \frac{\pi^{\frac{3}{2}} \hbar^4 c_{II} e \Gamma(2)}{2(2m^*)^{\frac{3}{2}} N \Delta_c^2 m^* \Gamma\left(\frac{5}{2}\right)} \frac{1}{(kT)^{\frac{3}{2}}} = \frac{\pi^{\frac{3}{2}} \hbar^4 e c_{II}}{3 \cdot 2^{\frac{3}{2}} N \Delta_c^2 k^{\frac{3}{2}} m^* T^{\frac{3}{2}}}. \quad (62.5)$$

В выражение (62.5) для подвижности входят величины трех видов: 1) универсальные константы e , \hbar , k и числовые множители; 2) характеристики данного вещества c_{II} , N , Δ_c , m^* ; 3) температура. В данном веществе подвижность зависит только от температуры:

$$\mu = \mu_0 T^{-3/2} \quad (62.6)$$

— с ростом температуры подвижность, обусловленная тепловыми колебаниями решетки, уменьшается. Смысл этого выражения достаточно простой: с ростом температуры число фононов возрастает пропорционально температуре, а средняя скорость пропорциональна \sqrt{T} , вследствие чего вероятность столкновения электронов или дырок с фононами растет пропорционально $T^{3/2}$; если температура уменьшается, то подвижность носителей заряда возрастает. Необходимо только помнить, что при $T \rightarrow 0$ выражение (62.5) теряет смысл вследствие неприменимости соотношения (61.32).

Обращает на себя внимание, что величина подвижности обратно пропорциональна эффективной массе в степени $5/2$. Это подтверждается опытом — в веществах с малой эффективной массой подвижность носителей заряда больше, чем в веществах с большей эффективной массой, а подвижность электронов превосходит подвижность дырок.

При рассеянии на оптических колебаниях решетки температурная зависимость подвижности определяется фактором

$$\left[e^{\frac{\hbar\omega_0}{kT}} - 1 \right]. \quad (62.7)$$

2. Рассеяние на ионах примеси. Согласно (56.23) время релаксации τ_I при рассеянии на ионах примеси имеет двойную зависимость от энергии — степенную и логарифмическую:

$$\tau_I = \frac{V\sqrt{2}\epsilon^2 m^{*1/2} E^2}{\pi Z^2 e^4 N_I \ln \left[1 + \left(\frac{\epsilon E}{N_I^{1/3} Z e^2} \right)^2 \right]}. \quad (62.8)$$

При усреднении этого выражения мы вынесем логарифмическую функцию, взяв ее в точке, в которой подынтегральная функция достигает максимума. Для этого необходимо вместо E подставить $3kT$. После вынесения логарифмической функции из-под знака интеграла можем записать

$$\langle \tau_I \rangle = \frac{8V\sqrt{2}k^{3/2}\epsilon^2 m^{*1/2}}{\pi^{3/2} Z^2 e^4 N_I} \frac{T^{3/2}}{\ln \left[1 + \left(\frac{3\epsilon kT}{N_I^{1/3} Z e^2} \right)^2 \right]}. \quad (62.9)$$

Из (62.9) следует выражение для подвижности при рассеянии на ионах примеси:

$$\mu_I = \frac{8V\sqrt{2}k^{3/2}}{\pi^{3/2} e^3} \frac{\epsilon^2}{Z^2 N_I m^{*1/2}} \frac{T^{3/2}}{\ln \left[1 + \left(\frac{3\epsilon kT}{N_I^{1/3} Z e^2} \right)^2 \right]}. \quad (62.10)$$

Подвижность носителей заряда при рассеянии на ионах примеси определяется величинами трех видов: 1) универсальными константами и числовыми множителями; 2) характеристиками вещества N_I , ϵ , Z , m^* ; 3) температурой.

Предполагая, что все величины, кроме температуры, постоянны, исследуем выражение для μ_I . При достаточно больших температурах логарифмической зависимостью μ_I от температуры пренебрежем, поэтому можем сказать, что подвижность с ростом температуры возрастает по закону

$$\mu_I \approx \mu_{0I} T^{3/2}. \quad (62.11)$$

При уменьшении температуры подвижность, обусловленная рассеянием на ионах примеси, уменьшается. Однако необходимо обратить внимание на следующее обстоятельство: при

$$T^2 \ll \frac{N_I^{2/3} Z^2 e^4}{9\epsilon^2 k^2} \quad (62.12)$$

логарифмическая зависимость становится существенной, и мы можем записать:

$$\mu_I \approx \frac{8\sqrt{2}e}{9\pi^{3/2}k^{1/2}} \frac{T^{-1/2}}{N_I^{1/3} m^{*1/2}}, \quad (62.13)$$

т. е. подвижность $\mu_I \sim \frac{1}{\sqrt{T}}$. Этот результат непосредственно вытекает из того факта, что время релаксации в области малых температур согласно (56.24) характеризуется зависимостью $\tau_I \sim E^{-1/2}$. Оценим температуру, исходя из (62.12). Положив $\epsilon \approx 10$; $Z = 1$, получим

$$T^2 \ll 5 \cdot 10^{-9} N_I^{2/3}. \quad (62.14)$$

Как видно из (62.14), температурный интервал δT , в котором справедливо соотношение (62.13), зависит от концентрации ионов примеси N_I ; при $N_I \rightarrow 0$ $\delta T \rightarrow 0$. Однако необходимо иметь в виду, что при $N_I \rightarrow 0$ подвижность $\mu_I \rightarrow \infty$ при *любой* температуре. В заключение необходимо заметить, что N_I зависит от температуры; при $T \rightarrow 0$ $N_I \rightarrow 0$, если в полупроводнике содержится примесь одного вида. Если же в полупроводнике есть и донорная, и акцепторная примеси, то при $T \rightarrow 0$ N_I принимает постоянное значение, равное удвоенной концентрации примеси с меньшим содержанием, т. е. если $N_a < N_d$, то $N_I = 2N_a$ вследствие взаимной компенсации примеси.

3. Рассеяние на нейтральных центрах. Согласно (57.13) при рассеянии частиц на нейтральных центрах время релаксации не зависит от энергии

$$\tau = \frac{m^{*2} e^2}{20\epsilon \hbar^3} \frac{1}{N_n}. \quad (62.15)$$

Отсюда следует, что подвижность, обусловленная рассеянием на нейтральных центрах, имеет вид

$$\mu_n = \frac{e^3 m^*}{20 \epsilon \hbar^3 N_n}. \quad (62.16)$$

В явном виде она не зависит от температуры. С ростом температуры электрически активная примесь ионизуется, вследствие чего при больших температурах μ_n определяется электрически неактивной примесью, т. е. примесью, не поставляющей в зоны свободные электроны и дырки.

При рассеянии на дислокациях $\tau \sim E^{-1/2}$, поэтому

$$\mu_D \approx \frac{1}{\sqrt{T}}. \quad (62.17)$$

4. Общий случай. Рассмотрев действие основных механизмов в «чистом» виде, перейдем к общему случаю, когда действуют *все механизмы рассеяния одновременно*. Считая, что *все механизмы рассеяния независимы*, мы можем сказать, что *полная вероятность рассеяния равна сумме вероятностей рассеяния на каждом типе рассеивающих центров*. Отсюда следует, что полное время релаксации τ имеет вид

$$\tau^{-1} = \frac{1}{4\pi^3} \int_{(V_k)} \sum_i \omega_i(k, k') (1 - \cos \theta) d\tau_k = \sum_i \tau_i^{-1}, \quad (62.18)$$

или

$$\tau = \left(\sum_i \tau_i^{-1} \right)^{-1}. \quad (62.19)$$

Поскольку τ_i зависят от энергии, то и τ является функцией энергии, которую можно найти, подставив явное выражение $\tau_i(E)$. Найдя $\tau(E)$ и усреднив его, получим выражение для подвижности:

$$\mu_d = \langle \mu \rangle = \frac{e \langle \tau \rangle}{m^*} = \frac{e}{m^*} \left\langle \frac{1}{\tau_i(E)} \right\rangle. \quad (62.20)$$

Проанализируем это выражение. Члены, входящие в качестве слагаемых в (62.20) или (62.19), играют различную роль в разных условиях. Предположим, что парциальные времена релаксации одинаковы. В таком случае

$$\tau_1 = \tau_2 = \dots = \tau_r \text{ и } \tau_r = \frac{\tau_i}{r}; \quad \mu = \frac{\mu_i}{r}, \quad (62.21)$$

т. е. если действует r механизмов рассеяния с одинаковыми временами релаксации, то полное время релаксации и полная подвижность меньше парциальных в r раз. В действительности равные времена релаксации маловероятны. Пусть время релаксации l -механизма много меньше других. Поскольку τ_i^{-1} при этом превосходит все остальные слагаемые в (62.19), мы можем сказать, что $\tau \leq \tau_l$. Дру-

гими словами, полное время релаксации τ определяется тем парциальным процессом, который имеет наименьшее время релаксации в данных условиях, вследствие чего и $\langle \tau \rangle \leq \langle \tau_i \rangle$. Поскольку τ_i зависят от энергии, мы должны сравнивать роль того или иного процесса релаксации при данной энергии носителей заряда. Однако при этом медленные и быстрые носители заряда могут играть различную роль в разных процессах. Чтобы получить правильный результат, в этом случае необходимо оценить роль частиц, имеющих среднюю энергию $\frac{3}{2} kT$, поскольку частицы с энергией, близкой к средней, составляют основную часть свободных носителей заряда.

Если вместо E подставить $\langle E \rangle$, то получим приближенно

$$\tau_i(\langle E \rangle) = \tau_0 (kT)^p \left(\frac{3}{2} \right)^p \quad (62.22)$$

вместо истинного

$$\langle \tau_i \rangle = \tau_0 (kT)^p \frac{\Gamma\left(\frac{5}{2} + p\right)}{\Gamma\left(\frac{5}{2}\right)}. \quad (62.23)$$

Таким образом, чтобы оценить роль того или иного процесса рассеяния, необходимо сравнить $\langle \tau_i \rangle$, или, что то же самое, $\langle \mu_i \rangle$. Исходя из этого можем сделать вывод, что при больших температурах основным механизмом рассеяния является рассеяние на тепловых колебаниях решетки: $\tau \sim E^p$ при $p = -1/2$. Подвижность носителей заряда с ростом температуры уменьшается по закону $T^{-3/2}$. При уменьшении температуры роль колебаний решетки уменьшается, подвижность увеличивается, но одновременно с этим увеличивается и роль рассеяния на ионах примеси, вследствие чего при малых температурах $\tau \sim E^p$ при $p = 3/2$ и $\mu \sim T^{3/2}$. Другими словами, если учитывать только рассеяние на ионах примеси и тепловых колебаниях решетки, то можно прийти к выводу, что подвижность μ с ростом температуры растет пропорционально $T^{3/2}$, проходит через максимум и затем уменьшается пропорционально $T^{-3/2}$.

Напомним, что в области температур, когда основную роль играет тепловое колебание решетки, множитель A , входящий в коэффициент Холла, должен быть равен $\frac{3\pi}{8}$, соответствующий $p = -1/2$;

в области низких температур $A = \frac{315\pi}{512}$ ($p = 3/2$). Аналогично должны вычисляться и другие кинетические коэффициенты.

Рассмотрим в заключение этого параграфа некоторые экспериментальные данные. На рис. 87 представлена температурная зависимость подвижности электронов и дырок в арсениде галлия. На рисунке отчетливо видны области роста и уменьшения подвижности

и область максимума. Кривые такого типа наблюдаются у многих веществ. На рис. 88 представлена зависимость подвижности электронов и дырок в кремнии при различной концентрации примеси в интервале температур от -50 до 250°C . Мы видим, что подвижность дырок меньше подвижности электронов. Это положение имеет место во всех полупроводниках.

Как видно из рис. 88, подвижность электронов и дырок с ростом концентрации ионов примеси уменьшается. Очевидно, что в собственном полупроводнике, имеющем совершенную кристаллическую

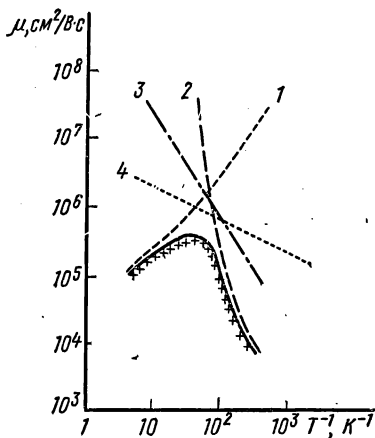


Рис. 87. Температурная зависимость подвижности электронов в арсениде галлия при действии различных механизмов рассеяния: 1—ионы примеси; 2—полярные оптические колебания решетки; 3—деформационный потенциал; 4—пьезоэлектрический потенциал; сплошная линия—суммарный эффект; xxx—эксперимент

структуру, величина подвижности определяется только тепловыми колебаниями решетки. При нарушении совершенства кристаллической структуры подвижность может только уменьшаться: $\mu \leq \mu_t$. На рис. 89 показана зависимость подвижности электронов и дырок от концентрации примеси при комнатной температуре. Из рис. 89 видно, что при концентрации примеси $N_I \leq 10^{14} \text{ см}^{-3}$ подвижность электронов и дырок практически не зависит от концентрации ионов примеси, но начиная с $N_I \approx 10^{15} \text{ см}^{-3}$ их подвижность падает. Другими словами, при $T \approx 300^\circ\text{C}$ рассеяние на ионах примеси в кремнии начинает играть определяющую роль при концентрации $N_I \geq 10^{15} \text{ см}^{-3}$. Существенно отметить, что отношение подвижностей $b = -\mu_n/\mu_p$ с ростом концентрации примеси уменьшается.

Аналогичное происходит с подвижностью электронов и дырок в соединениях A^{III}B^V с ростом концентрации примеси. Обращает на

себя внимание тот факт, что подвижность электронов в дырочном полупроводнике меньше подвижности электронов в электронном полупроводнике. Это различие тем более заметно, чем больше отношение масс m_p^*/m_n^* . Объяснение данного факта состоит в том, что вследствие большого различия в массах m_n^* и m_p^* электроны сильно рассеиваются в кулоновском поле дырок подобно тому, как они рассеиваются в поле ионов.

Из приведенных примеров видно, что подвижность носителей заряда может сильно отличаться от образца к образцу в зависимости от его состава и совершенства кристаллической структуры. Тем не менее мы можем говорить о подвижности электронов и дырок в данном полупроводниковом веществе. Численное значение подвижности, которое обычно связывается с данным полупроводниковым

веществом, относится к наиболее чистому и совершенному монокристаллу. В табл. 19 приведены численные значения подвижности

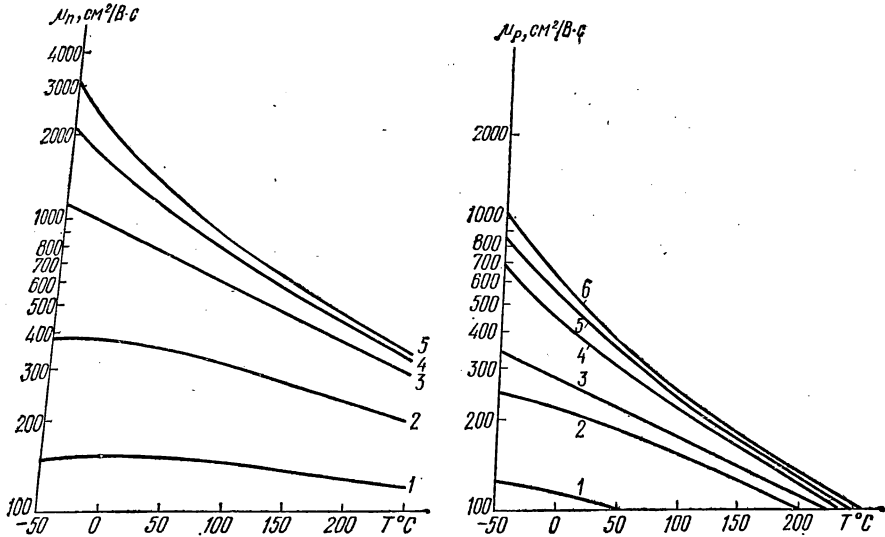


Рис. 88. Зависимость подвижности электронов и дырок от температуры в кремнии, содержащем примесь различной концентрации, $N_I, \text{см}^{-3}$:

- а) 1 — $5 \cdot 10^{18}$
 2 — $1 \cdot 10^{18}$
 3 — $1 \cdot 10^{17}$
 4 — $1 \cdot 10^{16}$
 5 — $1 \cdot 10^{15}$
- б) 1 — $5 \cdot 10^{18}$
 2 — $1 \cdot 10^{18}$
 3 — $5 \cdot 10^{17}$
 4 — $1 \cdot 10^{17}$
 5 — $1 \cdot 10^{16}$
 6 — $1 \cdot 10^{15}$

электронов и дырок в ряде полупроводниковых материалов при двух температурах: 300°K и 77°K и отношение подвижностей b .

Таблица 19

Полупроводниковое вещество	$-\mu_n \sim T^p$		$\mu_p \sim T^p$		$-\mu_n, \text{см}^2 (\text{В} \cdot \text{с})^{-1}$		$\mu_p, \text{см}^2 (\text{В} \cdot \text{с})^{-1}$		$b = \frac{\mu_n}{\mu_p}$	
	p	p								
			300K	77K	300K	77K	300K	77K		
Германий	-1,6	-2,3	3.800	37.100	1.820	43.700	2,1	0,7		
Селенный	-2,6	-2,3	1.300	45.500	500	11.600	2,6	3,9		
Sb	-1,6	-2,1	78.000	1.200.000	750	10.000	100	120		
As	-1,2	-2,3	33.000	82.000	460	690	70	120		
P	-2,0	-2,4	4.600	24.000	150	1.200	30	20		
Bb	-2,0	-0,9	4.000	6.000	1.400	3.600	3	1,7		
As	-1,0	-2,1	8.500	21.000	420	4.200	20	5		
P	-1,5	-1,5	110	500	75	420	1,5	1		
b	—	-1,8	200	—	420	3.700	—	—		

В табл. 19 указан показатель p зависимости подвижности от температуры в виде $\mu \sim T^p$. Обращает на себя внимание, что только в ряде случаев показатель p соответствует теоретическому значению ($-3/2$), обусловленному рассеянием на акустических колебаниях решетки. Достаточно удовлетворительного теоретического объяснения этого расхождения не существует. Однако очевидно, что оно отражает тот факт, что *существует целый ряд различных механизмов рассеяния*, обусловленных колебаниями решетки, приводящих к несколько другой зависимости подвижности от температуры. Действительно, расчеты показывают, что двухфононное рассеяние приводит к зависимости $\mu \sim T^{-3}$. В интерметаллах $A^{III}B^V$ большую роль играет рассеяние на оптических колебаниях решетки (полярные колебания).

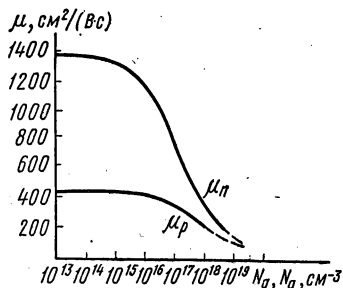


Рис. 89. Зависимость подвижности электронов и дырок от концентрации примеси в кремнии при $T=300\text{K}$

Однако можно сказать, что несмотря на большое число работ теории еще далека от возможности объяснить зависимость подвижности прежде всего от температуры для каждого конкретного вещества.

В заключение этого параграфа остановимся на температурной зависимости удельной электрической проводимости. Она определяется зависимостью подвижности и концентрации от температуры:

$$\sigma(T) = \sum_{\alpha} e_{\alpha} n_{\alpha}(T) \mu_{\alpha}(T). \quad (62.24)$$

В области истощения примеси концентрация основных носителей заряда остается постоянной, и проводимость при изменении температуры меняется только вследствие изменения подвижности. В зависимости от концентрации примеси подвижность может уменьшаться или увеличиваться, что приводит соответственно к уменьшению или увеличению проводимости. В области температур, соответствующих малой ионизации примеси или собственной проводимости, концентрация носителей заряда изменяется по экспоненциальному закону, поэтому $\sigma(T)$ определяется видом $n_{\alpha}(T)$. На рис. 90 в качестве примера приведен график зависимости $\sigma(T^{-1})$ при разной концентрации донорной и акцепторной примеси в кремнии.

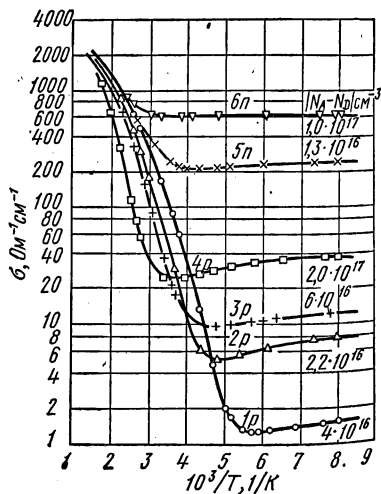


Рис. 90 Зависимость удельной электрической проводимости от обратной температуры в кремнии n - и p -типа

1. В невырожденных полупроводниках при действии одного механизма рассеяния усредненное время релаксации $\langle \tau \rangle$ определяется соотношением

$$\langle \tau \rangle = \tau_0 (kT)^p \frac{\Gamma\left(p + \frac{5}{2}\right)}{\Gamma\left(\frac{5}{2}\right)} \quad (62.1p)$$

при $\tau(E) = \tau_0 E^p$. Если действует несколько механизмов рассеяния, то

$$\tau^{-1} = \sum_i \tau_i^{-1} \quad (62.2p)$$

и

$$\langle \tau \rangle = \left\langle \left(\sum_i \tau_i^{-1} \right)^{-1} \right\rangle. \quad (62.3p)$$

В вырожденных полупроводниках $\langle \tau \rangle = \tau(F)$ и

$$\langle \tau \rangle = \left\langle \left\{ \sum_i \tau_i^{-1}(F) \right\}^{-1} \right\rangle. \quad (62.4p)$$

2. При рассеянии на тепловых колебаниях решетки подвижность пропорциональна $m^{*-5/2}$ и $T^{-3/2}$, если рассеяние происходит на акустических фононах; $\mu_T \sim m^{*-3/2}$ и определяется величиной $(e^{\frac{\theta_T}{T}} - 1)$ при рассеянии на оптических колебаниях.

3. При рассеянии на ионах примеси μ_I пропорциональна $T^{3/2}$ и $m^{*-1/2}$, но при $T \rightarrow 0$ $\mu_I \sim T^{-1/2}$.

4. При больших температурах основную роль играет рассеяние на тепловых колебаниях решетки, поэтому $A = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} = \frac{3\pi}{8}$. В области низких температур основную роль играет рассеяние на ионах примеси, что дает для A величину $\frac{315\pi}{512}$; однако по мере приближения T

к нулю величина $A = \frac{315\pi}{512}$ уменьшается и переходит в величину $3\pi/8$.

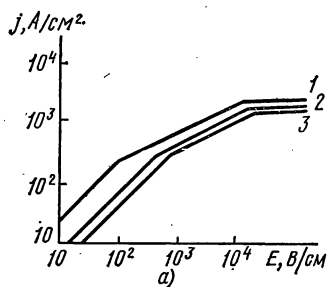
При рассеянии на полярных колебаниях A может меняться от 1 при малых температурах до 1,1–1,3 при $T \sim \theta_T$. При рассеянии на нейтральных примесях μ не зависит от температуры и $A = 1$. При рассеянии электронов на дырках

$$\mu_n \sim T^{-3/2} \text{ и } A_n = \frac{315\pi}{512}.$$

§ 63. ЗАВИСИМОСТЬ ВРЕМЕНИ РЕЛАКСАЦИИ ОТ ВНЕШНИХ ПОЛЕЙ.
НАРУШЕНИЯ ЗАКОНА ОМА

Решая кинетическое уравнение и анализируя связь между временем релаксации и вероятностью переходов, мы указывали, что в простейшем случае время релаксации не должно зависеть от полей. Если же $\tau = \tau(E, V)$, то решение резко усложняется.

На рис. 91 приведена зависимость плотности тока от напряженности электрического поля в электронном германии в двойной логарифмической шкале. На графике отчетливо видны три различные области. В первой области — в полях до 10^2 В/см (при 77° К) график



линеен с углом наклона в 45° , т. е. тангенс угла наклона равен единице, во второй области тангенс угла наклона равен $1/2$ и в третьей — 0. Это значит, что в первой области σ не зависит от поля, во второй области $\sigma \sim E^{-1/2}$ и

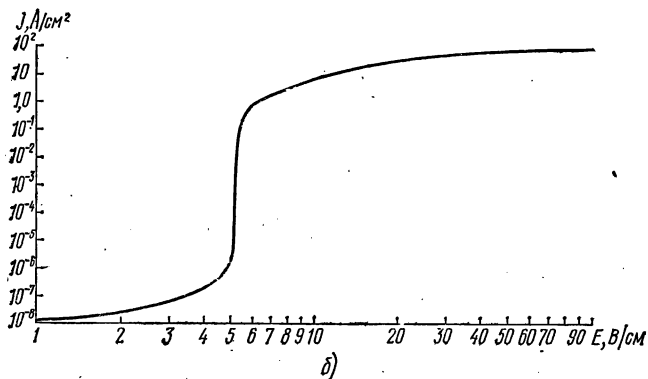


Рис. 91. Влияние сильного электрического поля на проводимость германия при трех различных температурах (а); ударная ионизация в электронном германии при $T = 4,2^\circ$ К (б)

в третьей $\sigma \sim E^{-1}$, так что $j \sim E$; $j \sim E^{1/2}$ и $j \sim E^0$ в указанных трех областях. Но $\sigma = en\mu = \frac{e^2 n \langle \tau \rangle}{m^*}$, поэтому нарушение закона Ома возможно по двум причинам — происходит изменение $\langle \tau \rangle$ или n . Изменение $\langle \tau \rangle$ полем легко понять из следующих простых соображений.

При изотропном рассеянии время релаксации практически соответствует времени свободного пробега, поскольку скорость направленного движения теряется полностью за одно столкновение. Скорость направленного движения зависит только от поля. Энергия, переда-

ваемая носителем заряда за одно столкновение, составляет малую долю полной энергии: $\langle \Delta E \rangle = \frac{2m^*}{M} E$. Но поскольку система частиц находится в стационарном состоянии, электроны или дырки отдают только избыточную энергию. Это означает, что под действием поля частицы за время свободного движения приобретают ту же энергию $\frac{2m^*}{M} E$, следовательно, под действием поля энергия и скорость электрона или дырки меняются незначительно. Если же поле возрастает, то полная скорость и энергия также возрастают. Поскольку длина свободного пробега не зависит от энергии (при рассеянии на тепловых колебаниях решетки), увеличение скорости приводит к уменьшению времени свободного пробега и, следовательно, к уменьшению подвижности.

Найдем изменение тепловой скорости v_T в результате действия поля. Под действием поля частица ускоряется и за единицу времени приобретает энергию, равную работе силы eE :

$$\frac{dE}{dt} = (eE v_d) = eE \mu_d E = \frac{e^2 \tau}{m^*} E^2. \quad (63.1)$$

В результате рассеяния частица теряет энергию $\frac{2m^*}{M} E$ в течение времени τ , так что в среднем за единицу времени теряется энергия $\frac{2m^*}{M} \frac{E}{\tau}$. Поскольку процесс стационарный, в среднем энергия носителей заряда не меняется со временем:

$$\frac{dE}{dt} = \frac{2m^*}{M} \frac{E}{\tau}, \quad (63.2)$$

или

$$\frac{e^2 \tau}{m^*} E^2 = \frac{2m^*}{M} \frac{m^* v^2}{2\tau}. \quad (63.3)$$

Заменяв величину τ на l/v , получим

$$e^2 l^2 E^2 = \frac{m^{*3}}{M} v^4. \quad (63.4)$$

Пусть $l = l_0 v^{2r}$, тогда

$$\frac{e^2 M l_0^2 E^2}{m^{*3}} = v^{(4-4r)} \quad (63.5)$$

и

$$v \sim E^{\frac{1}{2(1-r)}}. \quad (63.6)$$

При рассеянии на тепловых колебаниях решетки длина свободного пробега не зависит от энергии, следовательно, $r = 0$ и

$$v \sim E^{1/2}. \quad (63.7)$$

Увеличение скорости движения носителей заряда согласно (63.6) приводит к уменьшению τ :

$$\tau \approx \frac{l}{v} \sim \frac{1}{E^{1/2}},$$

откуда

$$\mu = \frac{\mu_0}{\sqrt{E}} \quad (63.8)$$

и

$$\sigma = \frac{\sigma_0}{\sqrt{E}}; \quad j = \sigma_0 \sqrt{E} \text{ при } E > E_{кр}. \quad (63.9)$$

Величина критического поля $E_{кр}$ определяется тем условием, что дополнительная скорость, приобретаемая частицей в поле, становится сравнимой с тепловой скоростью. Отсюда ясно, что при уменьшении температуры величина критического поля $E_{кр}$ уменьшается (см. рис. 91). Очевидно также, что чем больше μ_d , тем меньше $E_{кр}$.

Если $r \neq 0$, то зависимость v от поля будет более сложной. При $r = 2$, как это имеет место при рассеянии на ионах примеси, согласно (63.6)

$$v \sim E^{1/2}; \quad \tau \sim E^{5/2}. \quad (63.10)$$

Шокли нашел следующее выражение для зависимости подвижности от напряженности электрического поля при взаимодействии носителей заряда с фононами:

$$\mu(E) = \mu_0 \frac{\sqrt{2}}{\sqrt{1 + \sqrt{1 + \frac{3\pi}{8} \left(\frac{\mu_0 E}{c}\right)^2}}}, \quad (63.10')$$

где μ_0 — подвижность в слабом поле; c — скорость звука. Для кремния скорость звука $\sim 7 \cdot 10^5$ см/с. Поле, при котором электроны получают скорость, равную скорости звука, составляет $\sim 5,5$ кВ/см. При $E \gg c/\mu_0$ из (63.10) следует уже знакомый результат $\mu \approx \mu_0 \times \frac{4}{\sqrt{3\pi}} \left(\frac{c}{\mu_0 E}\right)^{1/2}$, т. е. $\mu \sim E^{-1/2}$ в нарушение закона Ома.

На рис. 92 приведена экспериментальная зависимость дрейфовой скорости от напряженности поля для двух веществ. Ее необходимо иметь в виду при оценке всех величин, зависящих от скорости.

Если в области сильных полей наблюдается насыщение тока, то при дальнейшем увеличении поля ток может резко возрасти, начиная с некоторого поля $E'_{кр}$.

Рост тока можно описать эмпирической зависимостью проводимости от поля:

$$\sigma = \sigma_0 e^{\nu(E - E'_{кр})}. \quad (63.11)$$

Резкое возрастание проводимости в электрическом поле носит название эффекта Пуля. Возрастание проводимости по экспоненциаль-

ному закону, как правило, связан с увеличением концентрации носителей заряда. Рост концентрации обуславливается рядом причин.

Один из механизмов носит название *ударной ионизации*. Электрон (или дырка) под действием поля приобретает энергию за время свободного пробега, достаточную для ионизации примеси или даже основного вещества. При столкновении «горячих» частиц с атомами примеси или основного вещества генерируются дополнительные носители заряда, которые в свою очередь ускоряются и генерируют дополнительные свободные носители заряда. Если учесть, что с ростом концентрации носителей заряда возрастает и их рекомбинация, то легко понять, что *возможно такое стационарное состояние, когда концентрация носителей заряда, а следовательно, и проводи-*

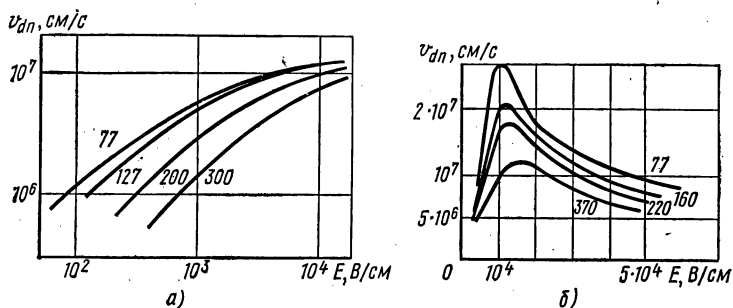


Рис. 92. Зависимость дрейфовой скорости электронов от напряженности электрического поля в кремнии (а) и теллуриде кадмия (б) при разных температурах, °К

мость при действии поля оказываются повышенными. Если же генерация не компенсируется рекомбинацией, то концентрация начинает возрастать лавинообразно, и в полупроводнике возникает необратимое явление пробоя.

На рис. 91, б представлена зависимость плотности тока от электрического поля в электронном германии при $T = 4,2 \text{ K}^\circ$. При $E \sim 5 \text{ В/см}$ происходит резкое скачкообразное возрастание плотности тока, вызванное ударной ионизацией атомов примеси. Однако при дальнейшем повышении поля наступает насыщение тока. Явление ударной ионизации может происходить в результате действия внутренних полей, обусловленных локальными неоднородностями кристалла или полем p - n -перехода. Ударная ионизация проявляется при тем меньших полях, чем меньше температура, меньше энергия активации и больше подвижность.

Вторым механизмом возрастания концентрации является эффект Зинера, который был рассмотрен нами в § 19. Вероятность туннельного перехода зависит от высоты ΔE_0 и ширины a барьера. Ширина барьера зависит от напряженности электрического поля:

$$aeE = \Delta E_0; \quad a = \frac{\Delta E_0}{eE}. \quad (63.12)$$

Как известно из квантовой механики, вероятность перехода для барьера треугольной формы имеет вид

$$D = D_0 e^{-\frac{2\sqrt{m^*}(\Delta E_0)^{3/2}}{2ehE}}. \quad (63.13)$$

Вероятность туннельного перехода одна и та же как для переходов из валентной зоны в зону проводимости, так и из зоны проводимости в валентную зону. Но поскольку концентрация электронов в валентной зоне превосходит концентрацию электронов в зоне проводимости, поток электронов будет направлен из валентной зоны в зону проводимости. Эффект Зинера аналогичен явлению холодной эмиссии. Он наблюдается в полях $E > (10^4 - 10^5) \text{ В/см}$. Таким образом, эффект Зинера приводит к увеличению концентрации свободных носителей заряда.

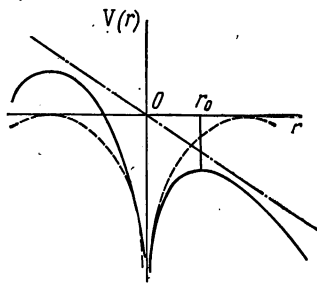


Рис. 93. Понижение потенциального барьера под действием электрического поля

Электростатическая ионизация, проявлением которой служит эффект Зинера, может выступать в виде других эффектов. Одним из них является эффект Штарка, который приводит к увеличению ширины уровней энергии атомов, образующих кристалл. Увеличение ширины уровней атомов кристалла приводит к увеличению ширины зон энергии и уменьшению ширины запрещенных зон. Но уменьшение ширины запрещенной зоны приводит к росту равновесной концентрации электронов и дырок.

Однако объяснение роста концентрации частиц при наложении сильного поля можно построить согласно Френкелю иначе, без учета Штарк-эффекта. На рис. 93 представлена потенциальная энергия электрона в атоме без поля (пунктирная линия) и с полем (сплошная линия). Внешнее электрическое поле понижает потенциальный барьер в направлении, противоположном электрическому полю. Максимум потенциального поля находится в точке r_0 , в которой сила притяжения к ближайшему ядру уравнивается внешней силой:

$$\frac{e^2}{er_0^2} = eE, \quad (63.14)$$

откуда

$$r_0 = \sqrt{\frac{e}{eE}}. \quad (63.15)$$

Можно оценить уменьшение величины потенциального барьера, разделяющего два соседних узла решетки, величиной

$$\delta E_I = 2eEr_0 = 2e\sqrt{\frac{eE}{e}}. \quad (63.16)$$

Но уменьшение потенциального барьера увеличивает вероятность термического возбуждения согласно статистике Больцмана на величину

$$e^{\frac{\delta E_I}{kT}} = e^{\left(\frac{2}{kT} \sqrt{\frac{e^3}{\epsilon}} V E\right)} = e^{\beta' V E}. \quad (63.17)$$

Этот эффект играет роль при $E > 10^5 \div 10^3$ В/см.

В некоторых веществах в сильных электрических полях наблюдаются высокочастотные осцилляции тока. Частота колебаний тока достигает в арсениде галлия величин $(1 \div 6) 10^9$ с⁻¹ при амплитуде колебаний тока более одного ампера. На рис. 94 представлена осциллограмма тока в образце

n-типа GaAs длиной 0,025 мм при подаче на него импульса напряжения величиной 16 В, длительностью 10⁻⁸ с. В верхней части рисунка представлен участок осциллограммы тока в образце *n*-типа GaAs длиной 0,025 мм при подаче на него импульса напряжения величиной 16 В, длительностью 10⁻⁸ с. В верхней части рисунка представлен участок осциллограммы в большем масштабе. Высокочастотные колебания тока при наложении на полупроводник постоянного напряжения могут быть обусловлены различными физическими процессами. Приведенные на рис. 94 осцилляции получены Ганном. Эффект Ганна наблюдается в арсениде галлия, фосфиде галлия и некоторых других веществах.

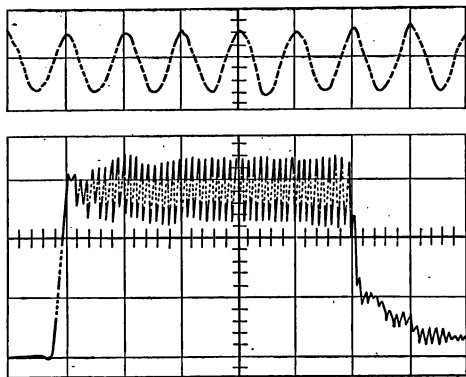


Рис. 94. Осцилляции тока в арсениде галлия при подаче на образец импульса напряжения длительностью 10⁻⁸ с

Одно из объяснений эффекта Ганна может состоять в следующем. При наложении электрического поля электроны переходят в более высокие энергетические состояния, температура электронного газа повышается. Предположим, что зона проводимости имеет минимумы энергии, лежащие выше абсолютного минимума и имеющие значительно большую эффективную массу, чем в нижнем минимуме, подобно тому, как это изображено на рис. 42 для арсенида галлия. Взаимодействуя с фононами, электроны могут быть переброшены в верхние долины. Так как плотность состояний в верхней долине превосходит плотность состояний нижней долины, то электроны будут накапливаться в верхней долине. Но подвижность электронов верхней долины значительно меньше подвижности электронов нижней долины, поэтому дрейфовая скорость электронов уменьшается, их вклад в проводимость уменьшается, и ток падает. Состояния в верхней долине являются неустойчивыми; электроны, взаимодействуя с фононами, переходят в нижнюю долину, что приводит к росту тока. Периодические колебания наблюдаются обычно в тонких образцах, что связано с механизмом возникновения повы-

шенного сопротивления в полупроводниках. Чем больше напряженность поля, тем сильнее происходит процесс переброса электронов в верхние долины, но чем больше электронов перебрасывается в верхнюю долину, тем больше сопротивление в данной области полупроводника и тем большее напряжение падает на ней, что приводит к уменьшению поля в соседних областях полупроводника. Как показывают наблюдения, области повышенного сопротивления зарождаются у «катода» и перемещаются к «аноду». Эффект Ганна наблюдается при таких полях, при которых дрейфовая скорость становится сравнимой с тепловой скоростью, т. е. при $v_d \approx 10^7$ см/с в обычных условиях.

Эффект Ганна может служить для исследований зонной структуры полупроводников и может быть использован для разработки генераторов колебаний большой мощности в области частот порядка 10^9 Гц.

Резюме § 63

1. Внешнее электрическое поле изменяет энергию носителей заряда, что приводит к изменению подвижности. В зависимости от механизма рассеяния подвижность может или возрастать или уменьшаться.

2. Сильное электрическое поле приводит к изменению концентрации свободных носителей заряда в результате ударной ионизации, эффекта Зинера, или внутренней холодной эмиссии, эффекта роста термического возбуждения вследствие понижения потенциального барьера, эффекта Штарка.

3. Переход горячих электронов в более высокие долины, в которых эффективная масса превосходит эффективную массу в нижней долине, приводит к уменьшению дрейфовой скорости электронов и, следовательно, тока. Переход электронов в нижнюю долину обеспечивает рост тока. Это приводит к тому, что при наложении на полупроводник постоянного напряжения возникает осциллирующий высокочастотный ток (эффект Ганна).

Глава VI

РЕКОМБИНАЦИЯ НОСИТЕЛЕЙ ЗАРЯДА

§ 64. УРАВНЕНИЕ НЕПРЕРЫВНОСТИ. ВРЕМЯ ЖИЗНИ

В третьей главе была подробно рассмотрена зависимость концентрации носителей заряда от температуры, концентрации и вида примеси. При этом мы предполагали, что полупроводник находится в состоянии термодинамического равновесия. Концентрацию носителей заряда, соответствующую равновесию, будем называть *равновесной концентрацией*. В четвертой главе, рассматривая кинетические явления, мы исходили из того, что внешние поля нарушают состояние термодинамического равновесия. *В полупроводнике, по которому течет электрический ток, состояние является термодинамически неравновесным.* Однако в однородном полупроводнике, находящемся в изотермических условиях, *концентрация носителей заряда остается равновесной, если поля невелики.* Но поскольку электрическое поле нарушает термодинамическое равновесие, то носители заряда являются неравновесными, ибо их распределение по состояниям описывается неравновесной функцией распределения. Действительно, согласно (29.1р—2р) выражение для концентрации носителей заряда имеет вид

$$n(\mathbf{r}) = \frac{1}{4\pi^3} \int_{(V_k)} f(\mathbf{r}, \mathbf{k}) d\tau_k = \frac{1}{4\pi^3} \int_{(V_k)} [f_0(\mathbf{r}, \mathbf{k}) + f^{(1)}(\mathbf{r}, \mathbf{k})] d\tau_k = n_0(\mathbf{r}) + n_1(\mathbf{r}), \quad (64.1)$$

где через n_0 обозначена *равновесная концентрация электронов или дырок*. Вычислим n_1 :

$$n_1(\mathbf{r}) = \frac{1}{4\pi^3} \int_{(V_k)} f^{(1)}(\mathbf{r}, \mathbf{k}) d\tau_k = -\frac{1}{4\pi^3} \int_{(V_k)} \frac{\partial f_0}{\partial E} (\mathbf{kv}) d\tau_k. \quad (64.2)$$

Легко видеть, что $n_1(\mathbf{r})$ равно нулю. Действительно, \mathbf{k} и $\partial f_0/\partial E$ — четные функции \mathbf{k} , а \mathbf{v} — нечетная, вследствие чего подынтегральная функция в (64.2) является нечетной, а интеграл от нечетной функции в симметричных пределах равен нулю. Таким образом,

$$n(\mathbf{r}) = n_0(\mathbf{r}) \quad (64.3)$$

— *концентрация остается равновесной.* Легко показать, что (64.3) получается из условия (36.6): $df/dt = 0$. Из этого же условия выте-

кает известное из гидродинамики и теории электричества *уравнение непрерывности*

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \operatorname{div} \rho \mathbf{v} = \frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0, \quad (64.4)$$

где ρ — плотность массы или заряда, \mathbf{v} — скорость их перемещения. Уравнение (64.4) означает, например, что *объемная плотность заряда изменяется только в результате расходимости тока*. Оно справедливо в самом общем случае. Однако если его записать в виде уравнения для *одного* типа носителей заряда, то оно существенно видоизменяется, поскольку *число носителей заряда каждого типа может меняться*, как это было видно из § 63, при этом полный заряд остается неизменным. Чтобы описать подобные состояния, необходимо обобщить использованное в четвертой главе уравнение Больцмана. Однако вместо этого обобщим уравнение непрерывности, но прежде всего введем некоторые определения, необходимые для понимания целого ряда явлений. *Если концентрация электронов или дырок $n(\mathbf{r}, t)$ отлична от равновесной концентрации $n_0(\mathbf{r})$, то ее называют неравновесной*. Величину $\delta n(\mathbf{r}, t)$, равную

$$\delta n(\mathbf{r}, t) = n(\mathbf{r}, t) - n_0(\mathbf{r}), \quad (64.5)$$

называют *избыточной концентрацией носителей заряда*.

Возникновение пары свободных носителей заряда называют генерацией пары, обратный процесс, приводящий к исчезновению пары свободных носителей заряда — электрона и дырки, — называют рекомбинацией. Генерация и рекомбинация могут происходить в результате внешних и внутренних причин. Обозначим через g_I и g_E скорости изменения концентрации вследствие генерации в результате внутренних (g_I) и внешних (g_E) причин, а через r_I и r_E — скорости изменения концентрации вследствие рекомбинации в результате действия внутренних (r_I) и внешних (r_E) причин. Учитывая поток частиц, генерацию и рекомбинацию носителей заряда, запишем уравнение непрерывности для электронов и дырок:

$$\frac{\partial n}{\partial t} = -\operatorname{div} \frac{\mathbf{j}_n}{e_n} + g_I^{(n)} + g_E^{(n)} - r_I^{(n)} - r_E^{(n)}, \quad (64.6)$$

$$\frac{\partial p}{\partial t} = -\operatorname{div} \frac{\mathbf{j}_p}{e_p} + g_I^{(p)} + g_E^{(p)} - r_I^{(p)} - r_E^{(p)}. \quad (64.7)$$

Если задать начальные и граничные условия, то при известных зависимостях величин \mathbf{j} , r , g от координат и времени можно в принципе получить решение уравнений (64.6) и (64.7) для $n(\mathbf{r}, t)$ и $p(\mathbf{r}, t)$. Однако решить уравнения непрерывности в общем случае не представляется возможным.

Рассмотрим ряд более простых случаев. В полупроводнике, находящемся в адиабатических условиях $g_E = r_E = 0$. Если в нем нет тока: $\mathbf{j} = 0$, то уравнение непрерывности (64.6) имеет вид

$$\partial n / \partial t = g_I - r_I. \quad (64.8)$$

В стационарных условиях $dn/dt = 0$, поэтому

$$g_I = r_I, \quad (64.9)$$

т. е. рекомбинация и генерация носителей заряда взаимно компенсируют друг друга, обеспечивая стационарность состояния. Поскольку генерация и рекомбинация носителей заряда происходят за счет энергии тепловых колебаний решетки, величины g_I и r_I , которые обозначим в этом случае через g_0 и r_0 , определяют скорости изменения концентрации частиц в результате *тепловой генерации и рекомбинации*. Величины n и p , обозначаемые в этом случае через n_0 и p_0 , представляют собой равновесные концентрации, их называют иногда *концентрациями темновых носителей заряда*. Величины n_0 и p_0 определяются только характеристиками полупроводника, например, для невырожденного полупроводника:

$$n_0 = N_c e^{-\frac{E_c - F}{kT}}; \quad p_0 = N_v e^{-\frac{F - E_v}{kT}}; \quad (64.10)$$

$$n_0 p_0 = N_c N_v e^{-\frac{E_c - E_v}{kT}} = n_i^2.$$

При $r_I \neq g_I$ в адиабатических условиях концентрация частиц меняется. При этом, если $g_I > r_I$, то $dn/dt > 0$ — концентрация частиц растет; если $g_I < r_I$, то $dn/dt < 0$ — концентрация частиц уменьшается. Поскольку полупроводник находится в адиабатических условиях и в нем нет тока, то условие $dn/dt \neq 0$ означает, что один из процессов преобладает. С другой стороны, отсутствие обмена энергией и отсутствие тока являются необходимыми условиями равновесного состояния, поэтому *уравнение*

$$dn/dt = g_I - r_I \neq 0 \quad (64.11)$$

должно описывать релаксационный процесс, т. е. процесс установления равновесного состояния, которое было нарушено током или внешним воздействием. Если внешнее воздействие увеличило концентрацию носителей заряда: $n - n_0 = \delta n > 0$, то в процессе релаксации рекомбинация должна превалировать: $r_I > g_I$. Если $n - n_0 = \delta n < 0$, то в процессе релаксации должна превалировать тепловая генерация: $g_I > r_I$. Обозначим разность между g_I и r_I через R :

$$R = r_I - g_I. \quad (64.12)$$

При $r_I = r_0$ и $g_I = g_0$ скорость изменения концентрации равна нулю — $R = 0$ — вещество находится в стационарном равновесном состоянии. Говоря о рекомбинации, имеют в виду именно случай $R \neq 0$. Другими словами, говоря о рекомбинации, рассматривают *процесс изменения избыточной концентрации носителей заряда*. Для описания релаксационного процесса необходимо знать величину $R = R(r, t)$; в этом случае уравнение непрерывности (64.11)

$$\frac{dn}{dt} = -R(r, t) \quad (64.13)$$

позволяет найти вид $n(r, t)$.

Рассмотрим решение уравнения (64.13) при некоторых предположениях о величине $R(r, t)$. Пусть существует некоторая постоянная величина $1/\tau_f$, равная вероятности рекомбинации одного свободного носителя заряда за единицу времени в единице объема. В таком случае за единицу времени прорекомбинирует $\frac{n-n_0}{\tau_f} = R$ частиц:

$$\frac{\partial n}{\partial t} = -R = -\frac{n-n_0}{\tau_f} = -\frac{\delta n}{\tau_f}. \quad (64.14)$$

Уравнение (64.14) легко интегрируется, в результате чего получаем

$$\delta n(t) = n(t) - n_0 = [n(0) - n_0] e^{-\frac{t}{\tau_f}} = \delta n(0) e^{-\frac{t}{\tau_f}}. \quad (64.15)$$

Таким образом, *неравновесное состояние после прекращения внешнего воздействия релаксирует с некоторой характерной величиной τ_f , которая носит название релаксационного времени жизни неравновесных носителей заряда, или просто времени жизни*. Величина τ_f численно равна времени, по истечении которого избыточная концентрация уменьшается в e раз. Легко видеть, что τ_f представляет собой *среднее время существования избыточной концентрации*.

Действительно, в интервале времени $t, t+dt$ прорекомбинирует $-dn = \delta n(t) \frac{dt}{\tau_f}$ частиц, «проживших» в течение времени t . Если сложим все времена $(-dn)t$ прорекомбинировавших частиц и разделим на начальное число избыточных носителей заряда, то получим среднее время $\langle t \rangle$ их жизни:

$$\langle t \rangle = \frac{1}{\delta n(0)} \int_0^{\infty} t (-dn) = - \int_0^{\infty} t e^{-\frac{t}{\tau_f}} \frac{dt}{\tau_f} = \tau_f. \quad (64.16)$$

Таким образом, величина τ_f , определяющая релаксационный процесс и равная обратной величине вероятности рекомбинации одного носителя заряда в единицу времени и в единице объема, представляет собой величину, равную среднему времени жизни неравновесных носителей заряда, или среднему времени существования избыточной концентрации неравновесных частиц.

Так как можно нарушить равновесное состояние и для электронов и для дырок, то уравнения (64.16) и (65.15) имеют место одновременно для электронов и дырок, при этом времена жизни τ_f^n и τ_f^p могут не совпадать. *Если скорость изменения избыточной концентрации неравновесных носителей заряда пропорциональна их концентрации, то рекомбинация носит название линейной*. При линейной рекомбинации переход носителей заряда из свободного состояния в связанное происходит независимо от наличия избыточных носителей заряда другого знака. Это означает, что не происходит прямого соединения электрона и дырки.

Если происходит прямое соединение неравновесных электрона и дырки, то в одном акте рекомбинации исчезает как электрон, так и дырка. Поскольку мы рассматриваем прямую рекомбинацию электрона и дырки, скорости изменения концентраций электронов и дырок равны; кроме того, они должны быть пропорциональны их произведению, или

$$R = \gamma (np - n_0 p_0) \quad (64.17)$$

и

$$\frac{\partial n}{\partial t} = -\gamma (np - n_0 p_0) = \frac{\partial p}{\partial t}. \quad (64.18)$$

Подставляя вместо n и p величины $n_0 + \delta n$ и $p_0 + \delta p$, получим

$$\begin{aligned} \frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} &= -\gamma [n_0 p_0 + n_0 \delta p + p_0 \delta n + \delta n \delta p - n_0 p_0] = \\ &= -\gamma [n_0 \delta p + p_0 \delta n + \delta n \delta p]. \end{aligned} \quad (64.19)$$

Пусть отклонение концентрации от равновесной невелико:

$$\delta n \delta p \ll n_0 \delta p + p_0 \delta n. \quad (64.20)$$

Предположим, что полупроводник электронный, в таком случае $p_0 \ll n_0$ и

$$\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = -\gamma n_0 \delta p; \quad (64.21)$$

решение уравнения (64.21) имеет вид

$$\delta p(t) = [p(0) - p_0] e^{-\gamma n_0 t} = \delta p(0) e^{-\frac{t}{\tau_f}}, \quad (64.22)$$

где

$$\tau_f = \frac{1}{\gamma n_0}. \quad (64.23)$$

Из (64.22) видно, что избыточная концентрация неосновных носителей заряда уменьшается (увеличивается при $\delta p(0) < 0$) по экспоненте подобно тому, как было в случае линейной рекомбинации. Это естественно, поскольку уравнение (64.21) совпадает с уравнением (64.14) при $\gamma = (n_0 \tau_f)^{-1}$.

Учитывая, что

$$\frac{\partial n}{\partial t} = -\gamma n_0 \delta p(0) e^{-\frac{t}{\tau_f}}, \quad (64.24)$$

и интегрируя (64.24), получим

$$\delta n(t) = \gamma n_0 \delta p(0) \tau_f e^{-\frac{t}{\tau_f}} + C. \quad (64.25)$$

Константу интегрирования C определим из начальных условий

$$\delta n(0) = \gamma n_0 \delta p(0) \tau_f + C = \delta p(0), \quad (64.26)$$

или

$$C = \delta n(0) - \gamma n_0 \delta p(0) \tau_f = \delta n(0) - \delta p(0). \quad (64.27)$$

Подставляя (64.27) в (64.25), получим

$$\delta n(t) = \delta p(0) e^{-\frac{t}{\tau_f}} + \delta n(0) - \delta p(0). \quad (64.28)$$

Но так как $\delta n \rightarrow 0$ при $t \rightarrow \infty$, то $\delta n(0) = \delta p(0)$ и $\delta n(t) = \delta p$, т. е. в случае прямой рекомбинации отклонения от равновесного значения концентраций электронов и дырок равны и изменяются по экспоненте с некоторым временем жизни $\tau_f = (\gamma n_0)^{-1}$.

Если избыточная концентрация велика, так что $\delta n \gg (p_0, n_0)$, то

$$\frac{\partial n}{\partial t} = -\gamma \delta n \delta p. \quad (64.29)$$

Предполагая, что $\delta n = \delta p$, получим

$$\frac{\partial \delta n}{\partial t} = \frac{\partial n}{\partial t} = -\gamma (\delta n)^2. \quad (64.30)$$

Разделяя переменные, запишем

$$\frac{\partial \delta n}{(\delta n)^2} = -\gamma dt, \quad (64.31)$$

откуда

$$\frac{1}{\delta n(t)} = \gamma t + C; \quad C = \frac{1}{\delta n(0)}, \quad (64.32)$$

или

$$\delta n(t) = \frac{\delta n(0)}{1 + \gamma \delta n(0) t}. \quad (64.33)$$

Таким образом, при квадратичной рекомбинации избыточная концентрация уменьшается по гиперболическому закону. Если определить $\langle t \rangle$, то получим $\langle t \rangle = \infty$. Однако это не означает, что в случае квадратичной рекомбинации неравновесное состояние существует сколь угодно долгое время. Гиперболический закон изменения δn переходит в экспоненциальный через некоторое время, по истечении которого условие $\delta n \gg (p_0, n_0)$ нарушается.

Если считать возможным вводить время жизни в любом случае, то для квадратичной рекомбинации время жизни зависит от δp :

$$\frac{\partial \delta p}{\partial t} = -\gamma (\delta p)^2 = -\frac{\delta p}{\tau_f(t)}, \quad (64.34)$$

где

$$\tau_f(t) = \frac{1}{\gamma \delta p(t)}. \quad (64.35)$$

Величину $\tau_f(t)$ можно назвать *мгновенным временем жизни*, оно является функцией избыточной концентрации носителей заряда.

В общем случае можно считать, что изменение концентрации частиц определяется уравнением

$$\frac{\partial n}{\partial t} = -\frac{n - n_0}{\tau_f}, \quad (64.36)$$

где релаксационное время жизни

$$\tau_f = \tau_f(t) = -\frac{n-n_0}{\partial n/\partial t} = -\frac{\delta n}{\partial \delta n/\partial t}. \quad (64.37)$$

Для случая линейной рекомбинации это дает $\tau_f(t) = \tau_f$.

Если известна зависимость $\delta n(t)$, то отношение ординаты кривой $\delta n(t)$ к тангенсу угла наклона касательной в той же точке равно мгновенному времени жизни. Для линейной рекомбинации τ_f равно тангенсу угла наклона касательной к кривой, построенной в координатах $\{\ln \delta n(t), t\}$.

Рассмотрим уравнение непрерывности в однородном полупроводнике в случае, когда в нем существует генерация под действием внешнего возбуждения $G = g_E - r_E$ в отсутствие тока:

$$\frac{\partial n}{\partial t} = G - R. \quad (64.38)$$

Будем считать, что G не зависит от времени. В стационарном состоянии

$$G = R. \quad (64.39)$$

Для линейной рекомбинации $R = \frac{n-n_0}{\tau_f}$ и

$$G = \frac{n-n_0}{\tau_f}, \quad (64.40)$$

или

$$\delta n_{\text{стац}} = G\tau_f. \quad (64.41)$$

Величина

$$\tau_f = \frac{\delta n_{\text{стац}}}{G} = \tau_f^{\text{стац}} \quad (64.42)$$

носит название стационарного времени жизни неравновесных носителей заряда, оно равно отношению стационарного значения избыточной концентрации $\delta n_{\text{стац}}$ к скорости генерации G .

Для квадратичной рекомбинации

$$G = \gamma (\delta n_{\text{стац}})^2; \quad \delta n_{\text{стац}} = \sqrt{\frac{G}{\gamma}} \quad (64.43)$$

и

$$\tau_{\text{стац}} = \frac{\delta n_{\text{стац}}}{G} = \sqrt{\frac{1}{\gamma G}}. \quad (64.44)$$

Если процесс не стационарный, то

$$\frac{\partial n}{\partial t} = G - R \quad (64.45)$$

и

$$R = G - \frac{\partial n}{\partial t}. \quad (64.46)$$

Записав R в виде $R = \frac{n-n_0}{\tau_f}$, можем определить в общем случае мгновенное время жизни соотношением-

$$\tau_f = \tau_f(t) = \frac{n-n_0}{G - \frac{\partial n}{\partial t}}. \quad (64.47)$$

При $\partial n/\partial t = 0$ из (64.47) следует определение стационарного времени жизни; при $G=0$ — релаксационного времени жизни.

Если τ_f есть постоянная величина, то уравнение (64.45) имеет решение; запишем

$$\frac{d(n-n_0)}{dt} = G - \frac{n-n_0}{\tau_f} = \frac{G\tau_f - (n-n_0)}{\tau_f}, \quad (64.48)$$

или

$$\frac{d(n-n_0)}{(n-n_0) - G\tau_f} = -\frac{dt}{\tau_f}.$$

Интегрируя (64.48) с учетом начального условия $\delta n(0) = 0$, получим

$$(n-n_0) - G\tau_f = Ce^{-t/\tau_f}; \quad C = -G\tau_f; \\ \delta n(t) = G\tau_f(1 - e^{-t/\tau_f}). \quad (64.49)$$

Если при $t \gg \tau_f$ выключить возбуждение, то

$$\delta n(t) = \delta n(0) e^{-t/\tau_f} = G\tau_f e^{-t/\tau_f}. \quad (64.50)$$

На рис. 95 приведена зависимость $\delta n(t)$ при периодическом возбуждении полупроводника импульсами света G в предположении, что длительность освещения и затемнения $T \gg \tau_f$.

В случае квадратичной рекомбинации уравнение непрерывности имеет вид

$$\frac{\partial n}{\partial t} = \frac{\partial \delta n}{\partial t} = G - \gamma(\delta n)^2. \quad (64.51)$$

Разделим переменные

$$\frac{d(n-n_0)}{G - \gamma(n-n_0)^2} = dt, \quad (64.52)$$

откуда

$$t = \int \frac{d(n-n_0)}{G - \gamma(n-n_0)^2} + C. \quad (64.53)$$

Интеграл вычисляется следующим образом: определив корни знаменателя

$$G - \gamma(\delta n)^2 = 0; \quad \delta n_{1,2} = \pm \sqrt{\frac{G}{\gamma}} = \pm \delta n_1, \quad (64.54)$$

разложим подынтегральное выражение на множители:

$$\frac{1}{G-\gamma(\delta n)^2} = \frac{1}{(\delta n_1 - \delta n)(\delta n_1 + \delta n)\gamma} = \frac{1}{\gamma} \left(\frac{A}{\delta n_1 - \delta n} + \frac{B}{\delta n_1 + \delta n} \right) =$$

$$= \frac{1}{\gamma} \frac{\delta n_1(A+B) + \delta n(A-B)}{(\delta n_1 - \delta n)(\delta n_1 + \delta n)}, \quad (64.55)$$

откуда

$$A=B, \quad A = \frac{1}{2\delta n_1} = \frac{1}{2} \sqrt{\gamma/G}. \quad (64.56)$$

Учитывая (64.56) и (64.55), получим для (64.53)

$$t = \frac{1}{2\sqrt{G\gamma}} \ln \frac{\delta n_1 + \delta n}{\delta n_1 - \delta n} + C. \quad (64.57)$$

Из условия $\delta n(0) = 0$ следует $C = 0$, или

$$\frac{\delta n_1 + \delta n}{\delta n_1 - \delta n} = e^{2\sqrt{G\gamma}t}, \quad (64.58)$$

откуда

$$\delta n(t) = \delta n_1 \frac{e^{2\sqrt{G\gamma}t} - 1}{e^{2\sqrt{G\gamma}t} + 1} = \sqrt{G/\gamma} \operatorname{th} \sqrt{G\gamma} t. \quad (64.59)$$

Таким образом, при *квадратичной рекомбинации* нарастание избыточной концентрации определяется *гиперболической тангенсоидой*.

При малых временах $t \ll \frac{1}{\sqrt{G\gamma}}$ имеем

$$\delta n(t) \approx \sqrt{G/\gamma} \sqrt{G\gamma} t = Gt. \quad (64.60)$$

При больших временах $t \gg \frac{1}{\sqrt{G\gamma}}$ имеем

$$\delta n(t) \approx \sqrt{G/\gamma} = \delta n_{\text{стан}}. \quad (64.61)$$

Если возбуждение выключается в некоторый момент времени в стационарном состоянии, то спад происходит по гиперболическому закону

$$\delta n(t) = \frac{\delta n(0)}{1 + \sqrt{G\gamma}t} = \frac{\sqrt{G/\gamma}}{1 + \sqrt{G\gamma}t}. \quad (64.62)$$

Чем меньше γ , тем быстрее происходит рост избыточной концентрации при включении возбуждения и тем медленнее происходит спад при его выключении.

Резюме § 64

1. Если состояние отлично от термодинамически равновесного, то оно называется неравновесным, а носители заряда — неравновесными. Неравновесное состояние может быть вызвано перераспределением частиц в зоне Бриллюэна при сохранении концентрации

частиц. Оно наблюдается, например, при прохождении электрического тока через однородный полупроводник. Второй тип неравновесных состояний наблюдается в том случае, когда концентрация частиц изменяется в результате действия внешних или внутренних факторов. Концентрация частиц (электронов или дырок) называется в этом случае неравновесной, а разность

$$n(\mathbf{r}, t) - n_0(\mathbf{r}) = \delta n(\mathbf{r}, t) \quad (64.1p)$$

— избыточной концентрацией. Говоря о неравновесных состояниях, в физике полупроводников и полупроводниковых приборов обычно имеют в виду случай, когда $\delta n(\mathbf{r}, t) \neq 0$. Переход к равновесному состоянию в первом случае происходит с некоторым характеристическим временем τ , называемым временем релаксации, которое было использовано при описании кинетических эффектов. Во втором случае эта характеристическая величина называется временем жизни и обозначается через τ_f (часто используется обозначение времени жизни в виде τ , что, как правило, не приводит к недоразумениям). Смысл понятия времени жизни зависит от условий, в которых находится полупроводник.

2. Для описания неравновесного состояния можно быть использовано уравнение непрерывности в виде

$$\frac{\partial n}{\partial t} = -\operatorname{div} \frac{\mathbf{j}}{e} + g_I + g_E - r_I - r_E, \quad (64.2p)$$

где $n(\mathbf{r}, t)$ — концентрация частиц, \mathbf{j}/e — плотность их потока, g_I , g_E — скорости генерации, а r_I , r_E — скорости рекомбинации частиц в результате действия внутренних (g_I , r_I) и внешних (g_E , r_E) факторов. Вместо четырех величин r_I , g_I , r_E , g_E можно использовать две — R и G :

$$R = r_I - g_I; \quad G = g_E - r_E, \quad (64.3p)$$

при этом для R используется название скорости рекомбинации, а для G — скорости генерации. Уравнение непрерывности может быть записано в виде

$$\frac{\partial n}{\partial t} = -\operatorname{div} \frac{\mathbf{j}}{e} + G - R. \quad (64.4p)$$

3. Мгновенным временем жизни τ_f называют величину, определяемую отношением избыточной концентрации к скорости рекомбинации в отсутствие тока:

$$\frac{\delta n}{\tau_f} = R = G - \frac{\partial n}{\partial t}, \quad (64.5p)$$

или

$$\tau_f = \frac{n - n_0}{R} = \frac{n - n_0}{G - \frac{\partial n}{\partial t}}. \quad (64.6p)$$

4. В стационарном состоянии $\partial n/\partial t = 0$; $n - n_0 = \delta n_{\text{стац}}$ и (64.6p) определяет стационарное время жизни:

$$\tau_f^{\text{стац}} = \frac{\delta n_{\text{стац}}}{G}; \quad \delta n_{\text{стац}} = G\tau_f^{\text{стац}}. \quad (64.7p)$$

5. Рекомбинация называется линейной, если скорость рекомбинации R пропорциональна избыточной концентрации носителей заряда того же знака: $R = \delta n/\tau_f$. При $\tau_f = \text{const}$ это дает

$$\frac{\partial n}{\partial t} = -\frac{\delta n}{\tau_f} + G. \quad (64.8p)$$

При $G = 0$ — генерация отсутствует —

$$\delta n(t) = \delta n(0) e^{-t/\tau_f}. \quad (64.9p)$$

При $G \neq 0$ и $\delta n(0) = 0$ избыточная концентрация стремится к насыщению

$$\delta n(t) = G\tau_f (1 - e^{-t/\tau_f}). \quad (64.10p)$$

6. При квадратичной рекомбинации $R = \gamma(np - n_0p_0)$, и спад избыточной концентрации происходит по гиперболическому закону, а нарастание — по гиперболической тангенсоиде (64.62) и (64.59).

§ 65. МЕХАНИЗМЫ РЕКОМБИНАЦИИ. ЛИНЕЙНАЯ РЕКОМБИНАЦИЯ

Неравновесное состояние свободных носителей заряда может быть описано некоторой функцией распределения $f(\mathbf{r}, \mathbf{k}, t)$, вид которой был найден в четвертой главе для случая, когда отклонение от равновесного состояния вызывается полями E и B , градиентами химического потенциала F (энергии Ферми) и температуры. Неравновесная часть функции распределения $f^{(1)}$ описывает перераспределение частиц по состояниям в зоне Бриллюэна при сохранении числа частиц. При втором типе неравновесных состояний, когда концентрация частиц меняется, характер изменения функции распределения должен быть другим. Предположим, что для описания неравновесных состояний можно воспользоваться функцией Ферми — Дирака с другими значениями входящих в нее параметров.

Системы, в которых вероятность нахождения частиц в верхних состояниях больше, чем в нижних, называют системами с инверсной населенностью. Для их описания необходимо считать температуру величиной отрицательной. Системы с отрицательными температурами используются для создания квантовых генераторов и усилителей.

Для неравновесных систем с нормальным распределением частиц по состояниям используется функция Ферми — Дирака (или функция Больцмана), в которой энергия Ферми F заменяется некоторой величиной F^ , называемой квазиуровнем Ферми. Квазиуровень Ферми F^**

можно ввести как величину, определяющую нормальное распределение частиц по состояниям в системах с неравновесной концентрацией носителей заряда. Величину F^* можно определить из условия нормировки аналогично тому, как это было сделано для уровня Ферми:

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{1/2}(\xi^*), \quad \xi^* = \frac{F^* - E_c}{kT}; \quad (65.1)$$

$$p = \frac{2N_v}{\sqrt{\pi}} \Phi_{1/2}(\eta^*), \quad \eta^* = \frac{E_v - F^*}{kT}. \quad (65.2)$$

Для невырожденных полупроводников можем записать

$$n = N_c e^{-\frac{E_c - F_n^*}{kT}}; \quad p = N_v e^{-\frac{F_p^* - E_v}{kT}} \quad (65.3)$$

и

$$np = N_c N_v e^{-\frac{\Delta E_0}{kT}} e^{-\frac{F_n^* - F_p^*}{kT}} = n_i^2 e^{-\frac{F_n^* - F_p^*}{kT}}. \quad (65.4)$$

Как видно из (65.4), квазиуровни Ферми электронов и дырок в неравновесных системах различны. Чем сильнее различаются квазиуровни Ферми электронов и дырок, тем сильнее отличается произведение неравновесных концентраций от произведения равновесных концентраций.

Прежде чем использовать понятие квазиуровня Ферми для описания процесса рекомбинации в неравновесных системах, рассмотрим некоторые наглядные представления о механизме рекомбинации.

По виду передачи энергии рекомбинирующих частиц различают три основных типа рекомбинации.

1. Рекомбинация называется излучательной, или фотонной, если энергия рекомбинирующих частиц выделяется в виде энергии фотона.

2. Если энергия частицы передается решетке (фононам), то рекомбинация называется безызлучательной, или фононной.

3. Одним из видов безызлучательной рекомбинации является ударная ионизация (процессы Оже), когда энергия рекомбинирующих частиц передается третьей частице, которая благодаря этому становится «горячей». «Горячая» частица в результате ряда столкновений передает свою энергию фононам.

Помимо этих трех основных механизмов, энергия рекомбинирующих частиц может передаваться электронному газу (плазменная рекомбинация). Если электрон и дырка образуют в качестве промежуточного состояния экситон, то такая рекомбинация носит название экситонной.

Фотонная, фононная и рекомбинация Оже могут протекать по-разному в зависимости от механизма перехода электрона из зоны проводимости в валентную зону.

Если частицы рекомбинируют в результате непосредственной встречи электрона и дырки, то такая рекомбинация называется

прямой, или межзонной. Прямая рекомбинация играет основную роль в веществах с малой шириной запрещенной зоны порядка 0,2—0,3 эВ и меньше.

Если ширина зоны больше 0,5 эВ, то рекомбинация происходит через локализованные состояния, лежащие в запрещенной зоне. Эти состояния обычно называют рекомбинационными ловушками. Предположим, что в полупроводнике имеются дефекты, уровни энергии которых лежат в запрещенной зоне, пусть их концентрация равна N_t , а уровень энергии E_t не занят электроном (занят дыркой). Возможен целый ряд процессов, схематически изображенных на рис. 96:

а — нейтральный дефект захватывает свободный электрон;

б — отрицательно заряженный дефект отдает электрон в зону проводимости. Таким образом, электрон, пробыв некоторое время на уровне дефекта, вновь становится свободным. Если дефект с уровнем энергии E_t осуществляет захват свободных электронов с последующим их освобождением, то он называется ловушкой захвата электрона;

в — нейтральный дефект захватывает свободную дырку (отдает электрон валентной зоне);

г — положительно заряженный дефект захватывает электрон из валентной зоны; такой дефект является ловушкой захвата дырки;

д — захватив электрон из зоны проводимости, отрицательно заряженный дефект захватывает свободную дырку — отдает захваченный электрон в валентную зону. Происходит процесс рекомбинации пары электрон — дырка;

е — захватив свободную дырку, положительно заряженный дефект захватывает свободный электрон, превращаясь в нейтральный дефект. Происходит процесс рекомбинации свободной пары электрон — дырка.

Захват носителей заряда не влияет на стационарное время жизни, но оказывает влияние на мгновенное время жизни. Освобождение захваченного носителя заряда может быть вызвано тепловым переобросом. В некоторых случаях это происходит в результате «подсветки».

Если произошел захват носителя заряда нейтральным дефектом, то последующий захват носителя заряда должен быть более вероятным в силу кулоновского притяжения между заряженной рекомбинационной ловушкой и носителем заряда другого знака.

Таким образом, дефекты с энергией E_t могут быть ловушками захвата электронов и дырок. Роль ловушек захвата могут играть атомы доноров и акцепторов. Характерной особенностью ловушек захвата является то, что они в основном взаимодействуют только с одной зоной — зоной проводимости или валентной зоной. Для того

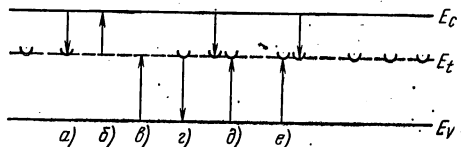


Рис. 96. Схема переходов электронов и дырок при взаимодействии ловушек захвата и рекомбинационных ловушек с зонами энергии

чтобы дефект привел к рекомбинации пары свободных носителей заряда, он должен взаимодействовать с обеими зонами.

Перейдем к описанию процесса рекомбинации, проходящей через рекомбинационные ловушки с энергией E_t и концентрацией N_t . Поскольку полупроводник находится в неравновесном состоянии, то распределение электронов и дырок по состояниям в зонах энергии и локальным уровням E_t характеризуется квазиуровнями Ферми F_n^* , F_p^* и F_t^* . Функцию распределения электронов и дырок по уровням E_t дефекта обозначим через f_t и $f_{tp} = 1 - f_t$.

Предположим, что существуют вероятности захвата в единицу времени одним нейтральным дефектом одного свободного электрона и отрицательно заряженным дефектом одной свободной дырки; обозначим их через $c_n(E)$ и $c_p(E)$.

В интервале энергии dE имеется $2N(E)dE$ состояний, в которых содержится $dn_c = 2N(E)f dE$ электронов. В единицу времени будет захвачено рекомбинационными ловушками

$$dr_n = c_n(E) 2N(E) f dE n_{tp} \quad (65.5)$$

электронов. За то же время в зону проводимости перейдет

$$dg_n = l_n(E) 2N(E) f_p dE n_{tp} \quad (65.6)$$

электронов. Через $l_n(E)$ обозначена вероятность перехода электрона с уровня ловушки на свободный уровень E в зоне проводимости. Между $c_n(E)$ и $l_n(E)$ должна существовать определенная связь, которую можно найти из условия термодинамического равновесия. При термодинамическом равновесии

$$dr_n = dg_n. \quad (65.7)$$

Подставляя (65.6) и (65.5) в (65.7) и сокращая одинаковые члены, получим

$$c_n(E) f n_{tp} = l_n(E) f_p n_{tp}, \quad (65.8)$$

откуда можно записать

$$\frac{l_n(E)}{c_n(E)} = \frac{f}{f_p} \frac{n_{tp}}{n_t}. \quad (65.9)$$

Но так как

$$f_{tp} = 1 - f_t = 1 - \frac{1}{e^{\frac{E_t - F_t^*}{kT}} - 1} = f_t e^{\frac{E_t - F_t^*}{kT}} \quad (65.10)$$

и

$$f_p = f e^{\frac{E - F_n^*}{kT}}, \quad (65.11)$$

то

$$\frac{l_n(E)}{c_n(E)} = e^{-\frac{E - F_n^*}{kT}} \cdot e^{\frac{E_t - F_t^*}{kT}} = e^{-\frac{E - E_t}{kT}}, \quad (65.12)$$

поскольку при равновесии

$$F_i^* = F_n^* = F_p^* = F. \quad (65.13)$$

В отсутствие равновесия $dg_n \neq dr_n$, поэтому

$$\begin{aligned} dR_n &= dr_n - dg_n = c_n(E) 2N(E) f dE N_{iftp} - l_n(E) 2N(E) f_p dE N_{ift} = \\ &= c_n(E) 2N(E) dE f N_{iftp} \left[1 - \frac{l_n(E)}{c_n(E)} \frac{f_p}{f} \frac{f_t}{f_{tp}} \right] = \\ &= c_n(E) 2N(E) dE f N_{iftp} \left[1 - e^{-\frac{F_i^* - F_n^*}{kT}} \right]. \end{aligned} \quad (65.14)$$

Интегрируя величину dR_n по энергии, получим

$$\begin{aligned} R_n &= N_{iftp} \left[1 - e^{-\frac{F_i^* - F_n^*}{kT}} \right] 2 \int_{E_c}^{\infty} N(E) f c_n(E) dE = \\ &= \left[1 - e^{-\frac{F_i^* - F_n^*}{kT}} \right] N_{iftp} n \langle c_n \rangle. \end{aligned} \quad (65.15)$$

Преобразуем выражение для R_n следующим образом. Обозначим

$$N_t \langle c_n \rangle = C_n = \frac{1}{\tau_{on}}, \quad (65.16)$$

перепишем второе слагаемое из (65.15) (без C_n):

$$\begin{aligned} n e^{\frac{F_i^* - F_n^*}{kT}} f_{tp} &= N_c e^{\frac{-E_c + F_n^* + F_i^* - F_n^* + E_t - F_i^*}{kT}} f_t = N_c e^{-\frac{E_c - E_t}{kT}} f_t = \\ &= n_1 f_t, \end{aligned} \quad (65.17)$$

где n_1 численно равно концентрации электронов, когда квазиуровень Ферми F_n совпадает с уровнем ловушки E_t . Учитывая (65.17), запишем скорость рекомбинации электронов R_n в виде

$$R_n = C_n n f_{tp} - C_n n_1 f_t. \quad (65.18)$$

Найдем скорость изменения концентрации дырок. Предположим, что существует вероятность захвата дырки ловушкой $c_p(E)$ и вероятность эмиссии ловушкой дырки $l_p(E)$. Проведя все аналогичные рассуждения относительно дырок, получим

$$R_p = C_p p f_t - C_p p_1 f_{tp}; \quad \frac{1}{C_p} = \tau_{p0}, \quad (65.19)$$

где

$$p_1 = N_v e^{-\frac{E_t - E_v}{kT}}. \quad (65.20)$$

Полные скорости захвата дырок и электронов центрами рекомбинации равны

$$R_p = R_n. \quad (65.21)$$

Отсюда можно найти функцию f_t , т. е. положение уровня F_t^* :

$$C_p p f_t - C_p p_1 f_{tp} = C_n n f_{tp} - C_n n_1 f_t \quad (65.22)$$

или

$$(C_p p + C_n n_1) f_t = (C_n n + C_p p_1) (1 - f_t), \quad (65.23)$$

откуда

$$f_t = \frac{C_n n + C_p p_1}{C_n (n + n_1) + C_p (p + p_1)} \quad (65.24)$$

и

$$f_{tp} = 1 - f_t = \frac{C_n n_1 + C_p p}{C_n (n + n_1) + C_p (p + p_1)}. \quad (65.25)$$

Подставляя f и f_{tp} в (65.18), получим

$$\begin{aligned} R_n &= \frac{C_n n (C_n n_1 + C_p p) - C_n n_1 (C_n n + C_p p_1)}{C_n (n + n_1) + C_p (p + p_1)} = \\ &= \frac{C_n C_p (n p - n_1 p_1)}{C_n (n + n_1) + C_p (p + p_1)}. \end{aligned} \quad (65.26)$$

Но согласно (65.4)

$$n p = n_i^2 e^{\frac{F_n^* - F_p^*}{kT}}, \quad (65.27)$$

а из (65.17) и (65.20) следует

$$n_1 p_1 = N_c e^{-\frac{E_c - E_f}{kT}} N_v e^{-\frac{E_f - E_v}{kT}} = n_i^2, \quad (65.28)$$

поэтому

$$R_n = \frac{C_n C_p [n p - n_i^2]}{C_n (n + n_1) + C_p (p + p_1)} = \frac{C_n C_p n_i^2 \left[e^{\frac{F_n^* - F_p^*}{kT}} - 1 \right]}{C_n (n + n_1) + C_p (p + p_1)}. \quad (65.29)$$

Определим время жизни электронов τ_n и дырок τ_p соотношением (64.6р):

$$\tau_n = \tau_f^n = \frac{n - n_0}{R_n} = \frac{\delta n}{R_n} = \frac{\delta n [C_n (n + n_1) + C_p (p + p_1)]}{C_n C_p [n p - n_i^2]}, \quad (65.30)$$

$$\tau_p = \tau_f^p = \frac{\delta p}{R_p}. \quad (65.31)$$

Так как $R_n = R_p$, то $\delta n = \delta p$ и $\tau_n = \tau_p = \tau_f$, где τ_f — время жизни пары электрон — дырка.

Полагая $n = n_0 + \delta n$, $p = p_0 + \delta p$, получим

$$n p - n_0 p_0 = n_0 p_0 + \delta n p_0 + \delta p n_0 + \delta n \delta p - n_0 p_0 = \delta n (n_0 + p_0 + \delta n) \quad (65.32)$$

и

$$\tau_n = \tau_{p0} \frac{n_0 + n_1 + \delta n}{n_0 + p_0 + \delta n} + \tau_{n0} \frac{p_0 + p_1 + \delta p}{n_0 + p_0 + \delta p}. \quad (65.33)$$

Проанализируем выражение (65.33) для времени жизни электронно-дырочной пары. Время жизни неравновесных носителей заряда определяется двумя членами, которые связаны с временами захвата дефектом свободного носителя заряда τ_{p0} и τ_{n0} соответственно. Помимо вероятности захвата, которая зависит от природы рекомбинационных ловушек и их концентрации N_t , время жизни зависит от избыточной концентрации носителей заряда δn и δp . Положение уровня E_t проявляется через величины n_1 и p_1 . Концентрации донорной и акцепторной примесей входят в выражение для времени жизни через равновесную концентрацию носителей заряда n_0 и p_0 . Таким образом, на время жизни неравновесных носителей заряда влияют:

- 1) тип и концентрация рекомбинационных ловушек (через времена τ_{p0} и τ_{n0});
- 2) концентрация легирующей примеси (через концентрацию n_0, p_0);
- 3) положение уровня дефекта E_t (через n_1, p_1);
- 4) избыточная концентрация неравновесных носителей заряда ($\delta n, \delta p$);
- 5) температура (определяет значения n_0, p_0, n_1, p_1 и, возможно, τ_{n0}, τ_{p0}).

На рис. 97 указано положение уровня энергии E_t дефекта в запрещенной зоне. Для определенности можно считать, что E_t лежит между серединой запрещенной зоны E_i и дном зоны проводимости E_c ; в таком случае $n_1 > n_i > p_1$. Выделим четыре области, на которые можно условно разбить запрещенную зону.

Рассмотрим вначале случай *малого отклонения концентрации от равновесной*:

$$\delta n = \delta p \ll (p_0, n_0). \quad (65.34)$$

Выражение для времени жизни пары имеет вид

$$\tau = \tau_{p0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n0} \frac{p_0 + p_1}{n_0 + p_0}. \quad (65.35)$$

Рассмотрим зависимость τ от равновесной концентрации n_0, p_0 , которая определяется положением уровня Ферми F .

1. Уровень Ферми лежит в I области:

$$E_v < F < 2E_i - E_t = E_v + E_c - E_t, \quad (65.36)$$

полупроводник дырочный: $p_0 \gg n_0$; кроме того, $p_0 \gg p_1$; $p_0 \gg n_1$.

Уравнение (65.35) для τ можно упростить:

$$\tau = \tau_{p0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n0} \frac{p_0 + p_1}{n_0 + p_0} \approx \tau_{n0}. \quad (65.37)$$

Время жизни пары неравновесных электрона и дырки определяется только временем захвата электронов — τ_{n0} — неосновных носителей

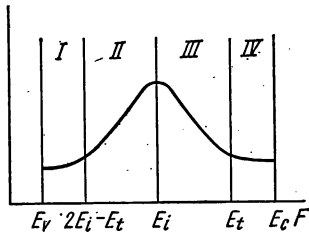


Рис. 97. Зависимость $\ln \tau$ от положения уровня Ферми в запрещенной зоне

заряда в дырочном полупроводнике. Как только неосновной носитель заряда захвачен рекомбинационной ловушкой, в тот же момент захватывается один из основных носителей заряда, концентрация которых велика.

2. Уровень Ферми лежит в IV области:

$$E_i < F < E_c; \quad n_0 \gg n_1 \gg p_1 \gg p_0. \quad (65.38)$$

Упрощая выражение (65.35), получим аналогично первому случаю:

$$\tau \approx \tau_{p0}. \quad (65.39)$$

Таким образом, время жизни неравновесных носителей заряда определяется только природой центров рекомбинации, оно не зависит от концентрации основных и неосновных носителей заряда.

3. Уровень Ферми лежит во II области: $E_c - E_i < F < E_i$, полупроводник p -типа слабо легирован; чем ближе F к E_i , тем ближе концентрация p_0 и n_0 к собственной. Учитывая соотношения между величинами, входящими в (65.33): $n_1 \gg p_0 \gg n_i \gg n_0 \gg p_1$, получим

$$\tau \approx \tau_{n0} + \tau_{p0} \frac{n_1}{p_0}. \quad (65.40)$$

При $F = E_i$ в собственном полупроводнике $\tau \approx \tau_{p0} \frac{n_1}{n_i}$, при $F \approx E_c - E_i$ имеем $p_0 \approx n_1$ и $\tau \approx \tau_{p0} + \tau_{n0}$. Таким образом, при перемещении F от левой границы второй области к правой границе области время жизни τ возрастает от $\tau_{n0} + \tau_{p0}$ до $\tau_{p0} \frac{n_1}{n_i} \gg \tau_{n0} + \tau_{p0}$.

4. Уровень Ферми находится в третьей области: $E_i < F < E_i$; $n_1 \gg n_0 \gg n_i \gg p_0 \gg p_1$, поэтому из (65.35) следует

$$\tau \approx \tau_{n0} + \tau_{p0} \frac{n_1}{n_0}. \quad (65.41)$$

При смещении F от E_i до E_i τ меняется от $\tau_{p0} \frac{n_1}{n_i}$ до $\tau \approx \tau_{p0}$ при переходе F в четвертую область.

На рис. 97 представлена графическая зависимость $\ln \tau$ от $\ln \frac{n}{n_i}$. *Время жизни τ_f максимально в собственном полупроводнике и равно времени захвата неосновного носителя заряда в сильно легированном полупроводнике.*

Если E_i лежит ниже E_i , то характер зависимости времени жизни от n_0 и p_0 не меняется, только во II и III областях время жизни будет определяться величиной $\tau_{n0} \frac{p_1}{n_0}$ или $\tau_{n0} \frac{p_1}{p_0}$. Чем ближе E_i к E_i , тем меньше область концентраций с большим временем жизни. Если в полупроводнике концентрация примеси распределена неравномерно, то время жизни будет меньше в области с большей концентрацией примеси (при равномерном распределении центров рекомбинации) наибольшим время жизни будет в p - n -переходе.

Рассмотрим теперь сильное отклонение концентрации носителей заряда от равновесной (так называемый *большой уровень инжекции*): $n \gg n_0$; $p \gg p_0$.

Преобразуем выражение для τ таким образом, чтобы выделить из него величину $\tau = \tau_0$ для малого уровня инжекции:

$$\begin{aligned} \tau &= \tau_{p0} \frac{n_0 + n_1 + \delta n}{n_0 + p_0 + \delta n} + \tau_{n0} \frac{p_0 + p_1 + \delta p}{n_0 + p_0 + \delta p} = \\ &= \frac{\tau_{p0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n0} \frac{p_0 + p_1}{n_0 + p_0} + (\tau_{p0} + \tau_{n0}) \frac{\delta n}{n_0 + n_0}}{1 + \frac{\delta n}{n_0 + p_0}} = \tau_0 \frac{1 + a \delta n}{1 + b \delta n}, \end{aligned} \quad (65.42)$$

где

$$\begin{aligned} \tau_0 &= \tau_{p0} \frac{n_0 + n_1}{n_0 + p_0} + \tau_{n0} \frac{p_0 + p_1}{n_0 + p_0}; \\ a &= \frac{\tau_{p0} + \tau_{n0}}{\tau_{p0} (n_0 + n_1) + \tau_{n0} (p_0 + p_1)}; \quad b = \frac{1}{n_0 + p_0}. \end{aligned} \quad (65.43)$$

Величины a и b положительны. Обозначим время жизни при $\delta n \rightarrow \infty$ через τ_∞ , из (65.42) следует

$$\tau_\infty = \tau_0 \frac{a}{b}. \quad (65.44)$$

Рассмотрим отношение

$$\frac{a}{b} = \frac{(\tau_{p0} + \tau_{n0}) (n_0 + p_0)}{\tau_{p0} (n_0 + n_1) + \tau_{n0} (p_0 + p_1)}, \quad (65.45)$$

или

$$\frac{b}{a} = \frac{\tau_{p0}}{\tau_{p0} + \tau_{n0}} \frac{n_0 + n_1}{n_0 + p_0} + \frac{\tau_{n0}}{\tau_{p0} + \tau_{n0}} \frac{p_0 + p_1}{p_0 + n_0} = \frac{\tau_0}{\tau_{p0} + \tau_{n0}} \quad (65.46)$$

и

$$\tau_\infty = \tau_{p0} + \tau_{n0}. \quad (65.47)$$

Поскольку τ_0 может быть или меньше, или больше, чем $(\tau_{p0} + \tau_{n0})$, то a/b может быть как больше, так и меньше единицы, поэтому и τ может или превосходить τ_0 , или быть меньше τ_0 , в зависимости от положения уровня Ферми относительно E_t . При $a > b$ время жизни τ возрастает с ростом уровня инжекции неравновесных носителей заряда, при $a < b$ оно уменьшается.

Разобранная в этом параграфе теория рекомбинации Холла — Шокли — Рида является теорией простейшей модели, справедливой при малых концентрациях N_t . В реальных полупроводниках рекомбинационные ловушки могут иметь различную природу, поэтому изменение времени жизни может носить значительно более сложный характер. На время жизни в германии и кремнии существенно влияют атомы примеси никеля, марганца, золота; важную роль играют дислокации.

Время жизни в германии и кремнии в зависимости от их состава может меняться от миллисекунд до долей микросекунд. В других полупроводниках время жизни может иметь значение от секунд до 10^{-16} с и меньше.

Время жизни пары в примесном полупроводнике представляет собой по существу время захвата неосновного носителя заряда, например, в электронном полупроводнике время жизни τ_f определяется временем захвата дырок τ_{p0} . Но время захвата в соответствии с (65.16) можно выразить через концентрацию рекомбинационных ловушек N_t :

$$\tau_{p0} = \frac{1}{C_p} = \frac{1}{\langle c_p \rangle N_t}. \quad (65.48)$$

Если вместо вероятности захвата одного носителя заряда одной ловушкой, отнесенной к единице объема и единице времени, ввести вероятность захвата носителя заряда из единичного потока частиц, то можно прийти к определению эффективного сечения захвата. Меняя вид ловушек и их концентрацию, можно менять время захвата неосновных неравновесных носителей заряда, а тем самым и время жизни пары. В табл. 20 приведены характеристики некоторых примесей в германии и кремнии, играющих роль рекомбинационных ловушек.

Таблица 20

Примесь	Германий			Кремний		
	E_t , эВ	$\langle c_p \rangle \cdot 10^9$, см ³ ·с	$\langle c_n \rangle \cdot 10^9$, см ³ ·с	E_t , эВ	$\langle c_p \rangle \cdot 10^9$, см ³ ·с	$\langle c_n \rangle \cdot 10^9$, см ³ ·с
Медь	+0,31	—	0,25	+0,24	—	—
Серебро	+0,13 —0,28	— 100	1,5 2,5	—	—	—
Золото	+0,16 —0,20	1000 200	2,5 5	+0,35 —0,54	1,6 16	67 9,5
Галлий	+0,0108	—	$2 \cdot 10^{-2}$	+0,065	—	10^{-3}
Железо	+0,35 —0,27	60 200	25 25	+0,40 —0,55	— 2	— —
Никель	+0,23 —0,30	— 200	2,5 7,5	—	—	—
Вакансии	+0,36	7	—	—	—	—
Междоузельные атомы	+0,42	0,14	—	—	—	—

Положение уровня энергии E_t указано с плюсом, если E_t отсчитывается от E_v , и с минусом, если E_t отсчитывается от E_c . Для сравнения приведена вероятность захвата электронов атомом галлия.

Резюме § 65

1. Дефект, способный захватывать свободные носители заряда с последующим их освобождением, называется ловушкой захвата.

2. Дефект, способный захватывать последовательно электрон и дырку, что приводит к рекомбинации пары свободных носителей заряда, называется рекомбинационной ловушкой.

3. Теория рекомбинации электронов и дырок через рекомбинационные ловушки приводит к следующему выражению для времени жизни пары $\tau = \tau_f$:

$$\tau_f = \tau_{p0} \frac{n_0 + n_1 + \delta n}{n_0 + p_0 + \delta p} + \tau_{n0} \frac{p_0 + p_1 + \delta p}{n_0 + p_0 + \delta p}, \quad (65.1p)$$

которое описывает зависимость времени жизни пары:

а) от величины избыточной концентрации неравновесных носителей заряда $\delta n = \delta p$ (от уровня инжекции); б) от равновесной концентрации носителей заряда n_0, p_0 ; в) от положения уровня энергии E_t рекомбинационной ловушки (посредством n_1 и p_1); г) от концентрации дефектов N_t и вероятности захвата $\langle c_n \rangle$ и $\langle c_p \rangle$, определяющих время захвата τ_{n0} и τ_{p0} .

4. При малом уровне инжекции в области примесной проводимости время жизни пары определяется временем захвата неосновных носителей заряда. По мере перемещения уровня Ферми к середине запрещенной зоны время жизни пары возрастает.

5. Для описания неравновесных состояний может быть использована функция Ферми — Дирака, параметры которой должны быть изменены. Если верхние уровни заселены больше, чем нижние, то температуру системы неравновесных частиц необходимо считать величиной отрицательной: $T < 0$. Такие системы называются системами с инверсной населенностью уровней, или системами с отрицательной температурой; они используются для создания квантовых генераторов и усилителей.

Если распределение неравновесных частиц по уровням энергии нормальное, то такая система описывается квазиуровнем Ферми F^* , определяемым из условия нормировки неравновесной функции распределения на неравновесную концентрацию n :

$$n = \frac{2N_c}{\sqrt{\pi}} \Phi_{\frac{1}{2}}(\xi^*). \quad (65.2p)$$

Для невырожденных полупроводников положение квазиуровня Ферми можно определить из уравнений для электронов

$$n = N_c e^{-\frac{E_c - F_n^*}{kT}}; \quad F_n^* = E_c + kT \ln \frac{n}{N_c}, \quad (65.3p)$$

для дырок

$$p = N_v e^{-\frac{F_p^* - E_v}{kT}}; \quad F_p^* = E_v - kT \ln \frac{p}{N_v}. \quad (65.4p)$$

Расстояние между квазиуровнями электронов и дырок определяется уровнем инжекции:

$$F_n^* - F_p^* = kT \ln \frac{np}{n_0 p_0} = kT \ln \frac{np}{n_i^2}. \quad (65.5p)$$

Для вырожденных полупроводников связь между неравновесной концентрацией и квазиуровнем Ферми можно получить, учитывая, что

$$\Phi_{1/2}(\xi^*) = \frac{2}{3} \xi^{*3/2}.$$

§ 66. ДИФФУЗИЯ И ДРЕЙФ НЕРАВНОВЕСНЫХ НОСИТЕЛЕЙ ЗАРЯДА

Если в полупроводниках создать неравновесные носители заряда, то в них наблюдается целый ряд новых физических явлений. Они могут быть описаны на основе обобщенного кинетического уравнения Больцмана или с помощью вытекающего из него уравнения непрерывности:

$$\frac{\partial n}{\partial t} = -\operatorname{div} \frac{j}{e} + G - R. \quad (66.1)$$

Рассмотрим стационарное состояние в отсутствие генерации носителей заряда за счет внешней энергии, т. е. положим $\partial n / \partial t = 0$, $G = 0$, после чего уравнение (66.1) запишем в виде

$$\frac{1}{e} \operatorname{div} j + R = 0. \quad (66.2)$$

Из уравнения (66.2) следует, что при $R \neq 0$ неравновесная концентрация носителей заряда поддерживается током j . Другими словами, *уменьшение избыточной концентрации носителей заряда вследствие рекомбинации компенсируется расходимостью тока.*

Для величины плотности тока j мы можем воспользоваться общим выражением (38.1p), считая, что кинетические коэффициенты выражаются через неравновесную концентрацию n и усредненное время релаксации для неравновесных носителей заряда. Будем считать, что магнитное поле отсутствует ($\mathbf{B} = 0$), полупроводник находится в изотермических условиях ($\nabla T = 0$). В таком случае из (38.1p) запишем

$$j = e^2 K_{11} \mathbf{E} - e K_{11} \nabla F^*. \quad (66.3)$$

Необходимость сохранения члена $-e K_{11} \nabla F^*$ очевидна, если учесть, что он описывает ток, обусловленный неоднородностью полупроводника. В данном случае неоднородность вызывается возможной

зависимостью неравновесной концентрации носителей заряда от координаты. Перепишем уравнение (66.3), учитывая выражения для K_{11} :

$$\mathbf{j} = e^2 n \frac{\langle \tau \rangle}{m^*} \mathbf{E} - en \frac{\langle \tau \rangle}{m^*} \nabla F^* = \sigma \mathbf{E} - n \mu_d \nabla F^*. \quad (66.4)$$

Из (65.3p) можем записать для невырожденного полупроводника

$$\nabla F^* = kT \frac{\nabla n}{n}, \quad (66.5)$$

откуда

$$\mathbf{j} = \sigma \mathbf{E} - kT \mu_d \nabla n = \mathbf{j}_E + \mathbf{j}_D, \quad (66.6)$$

где $\mathbf{j}_E = \sigma \mathbf{E}$ представляет собой *омический, или дрейфовый, ток*, а

$$\mathbf{j}_D = -kT \mu_d \nabla n \quad (66.7)$$

носит название *диффузионного тока*. Последнее название связано с тем, что ток \mathbf{j}_D возникает вследствие диффузии носителей заряда при их неравномерном распределении в кристалле. Поток частиц должен быть пропорционален градиенту концентрации и направлен в сторону уменьшения концентрации частиц. Умножая поток частиц на их заряд и вводя коэффициент пропорциональности D , можем записать

$$\mathbf{j}_D = -eD \nabla n. \quad (66.8)$$

Величину D называют *коэффициентом диффузии*, он имеет размерность $[L^2 T^{-1}]$, как это видно из его определения соотношением (66.8). Сравнивая (66.8) с (66.7), получим связь между коэффициентом диффузии и подвижностью носителей заряда

$$D = \frac{kT}{e} \mu_d \quad (66.9)$$

— выражение, справедливое как для дырок, так и для электронов. Выражение (66.9) называется *соотношением Эйнштейна для подвижности и коэффициента диффузии*. Подобное же соотношение можно получить и для вырожденного полупроводника. Действительно, учитывая связь между n и F^* , получим

$$\frac{\nabla n}{n} = \frac{3}{2} \frac{\nabla F^*}{F^*}; \quad \nabla F^* = \frac{2}{3} F^* \frac{\nabla n}{n}, \quad (66.10)$$

откуда

$$D = \frac{2}{3} \frac{F^*}{e} \mu_d. \quad (66.11)$$

Подставим уравнение (66.6) для \mathbf{j} в (66.2), получим

$$\frac{1}{e} \operatorname{div} (\sigma \mathbf{E} - eD \nabla n) + \frac{n - n_0}{\tau} = 0, \quad (66.12)$$

или

$$n \mu_d \operatorname{div} \mathbf{E} + \mu_d (E \nabla n) - D \nabla^2 n + \frac{n - n_0}{\tau} = 0. \quad (66.13)$$

Выражение (66.13) может быть записано как для дырок, так и для электронов, при этом величина электрического поля, входящего в уравнение (66.13) для электронов и дырок, не зависит от знака заряда частиц. Поскольку при прохождении тока *электро-нейтральность не нарушается*, то на основании уравнения Пуассона можем записать

$$\operatorname{div} \mathbf{E} = 0, \quad (66.14)$$

после чего уравнение (66.13) примет вид

$$\nabla^2 n - \frac{\mu_d}{D} (\mathbf{E} \nabla r) - \frac{n - n_0}{D\tau} = 0. \quad (66.15)$$

Так как неравновесная концентрация поддерживается током, то ∇n должен быть коллинеарен \mathbf{E} . Рассмотрим одномерный случай, положив $\mathbf{E} = (E, 0, 0)$, и запишем уравнение (66.15) для избыточной концентрации $n - n_0$:

$$\frac{d^2 (n - n_0)}{dx^2} - \frac{\mu_d E}{D} \frac{d (n - n_0)}{dx} - \frac{n - n_0}{D\tau} = 0. \quad (66.16)$$

Введем обозначения

$$D\tau = L^2, \quad \frac{\mu_d E}{D} = \frac{\mu_d E\tau}{D\tau} = \frac{l_E}{L^2}. \quad (66.17)$$

В новых обозначениях уравнение (66.16) будет иметь вид

$$\frac{d^2 (n - n_0)}{dx^2} - \frac{l_E}{L^2} \frac{d (n - n_0)}{dx} - \frac{n - n_0}{L^2} = 0. \quad (66.18)$$

Ищем решение линейного однородного уравнения в виде

$$n - n_0 = A e^{ax}. \quad (66.19)$$

Подставив (66.20) в (66.19), после сокращения на $A e^{ax}$ получим характеристическое уравнение

$$a^2 - \frac{l_E a}{L^2} - \frac{1}{L^2} = 0, \quad (66.20)$$

решение которого имеет вид

$$a = \frac{l_E}{2L^2} \pm \sqrt{\frac{l_E^2}{4L^4} + \frac{1}{L^2}} = \frac{1}{L} \left(\frac{l_E}{2L} \pm \sqrt{1 + \frac{l_E^2}{4L^2}} \right). \quad (66.21)$$

Введем величину $l = -1/a$, с помощью которой запишем решение (66.19):

$$n - n_0 = A e^{-x/l}. \quad (66.22)$$

Чтобы решение было конечным при $x \rightarrow \infty$, необходимо выбрать $a < 0$ или $l > 0$, A имеет смысл избыточной концентрации в точке

$x=0$: $[n - n_0]_{x=0} = A$. Из условия $a < 0$ следует, что в (66.21) необходимо оставить знак минус перед корнем, поэтому

$$\frac{1}{\tau} = \frac{1}{L} \left(\sqrt{1 + \frac{l_E^2}{4L^2}} - \frac{l_E}{2L} \right). \quad (66.23)$$

Мы пришли к выводу, что избыточная концентрация с ростом x уменьшается по экспоненте с некоторой характеристической величиной l , которую называют длиной затягивания; она равна расстоянию, на котором избыточная концентрация уменьшается в e раз.

Рассмотрим случай, когда электрического поля нет: $E=0$. В таком случае $l_E=0$ и

$$l = L = \sqrt{D\tau}. \quad (66.24)$$

В веществе нет электрического поля, в нем течет лишь диффузионный ток

$$\mathbf{j} = \mathbf{j}_D = -eD\nabla n. \quad (66.25)$$

Так как избыточная концентрация изменяется только в результате диффузии неравновесных носителей заряда, то $l=L$ называется диффузионной длиной. Для одномерного случая

$$n(x) - n_0 = \delta n(x) = \delta n(0) e^{-\frac{x}{L}} \quad (66.26)$$

и

$$\mathbf{j} = -eD\nabla_x n = \frac{eD\delta n(x)}{L} = e \frac{D\tau}{L\tau} \delta n = e \frac{L}{\tau} \delta n. \quad (66.27)$$

Обозначим отношение L/τ через v_D :

$$v_D = \frac{L}{\tau} = \frac{L^2}{L\tau} = \frac{D}{L}. \quad (66.28)$$

Величину v_D называют диффузионной скоростью, она равна численно скорости, с которой неравновесные носители заряда проходят путь, равный диффузионной длине, за время их жизни. Плотность диффузионного тока можно выразить через диффузионную скорость:

$$j_{Dx} = ev_D (n - n_0) = j_{D0} e^{-x/L}. \quad (66.29)$$

Выражение (66.29) показывает, что плотность диффузионного тока уменьшается по тому же закону, по которому изменяется избыточная концентрация, уменьшение которой обусловлено рекомбинацией неравновесных носителей заряда. Выражения (66.24–29) будут справедливы не только при $E=0$, но и при $E \neq 0$, если $l_E \ll 2L$.

Рассмотрим противоположный случай — случай больших полей: $l_E \gg 4L^2$. Это условие имеет наглядный смысл:

$$\frac{l_E^2}{4L^2} = \frac{(\mu_d E \tau)^2}{4L^2} = \frac{(v_d^2 \tau^2)}{4L^2} = \frac{v_d^2}{4v_D^2} \gg 1. \quad (66.30)$$

Величина l_E численно равна пути, который проходят неравновесные частицы за время жизни со скоростью дрейфа, поэтому ее называют длиной дрейфа неравновесных носителей заряда. Таким образом, поле E будем считать большим, если длина дрейфа намного превосходит диффузионную длину. Однако при нахождении длины затягивания l надо различать два случая $l_E > 0$ и $l_E < 0$. Рассмотрим случай $l_E > 0$, из (66.23) можем записать

$$\frac{1}{l} = \frac{1}{L} \left(\frac{l_E}{2L} \sqrt{1 + \frac{4L^2}{l_E^2}} - \frac{l_E}{2L} \right) = \frac{l_E}{2L^2} \left(1 + \frac{2L^2}{l_E^2} + \dots - 1 \right) \approx \frac{1}{l_E}, \quad (66.31)$$

т. е. $l = l_E$ — длина затягивания равна длине дрейфа:

$$l = l_E = \mu_d E \tau = v_d \tau \quad (l_E > 0). \quad (66.32)$$

Учитывая (66.22) и (66.32), запишем для избыточной концентрации $\delta n(x)$:

$$\delta n(x) = \delta n(0) e^{-\frac{x}{l_E}} = \delta n(0) e^{-\frac{x}{\mu_d E \tau}}. \quad (66.33)$$

Смысл этого выражения состоит в следующем. Если в начале координат создается избыточная концентрация $\delta n(0)$, то в результате прохождения тока, вызываемого сильным электрическим полем, избыточная концентрация «затягивается» полем на расстояния $(2 \div 3)l_E$ в объем полупроводника. Если при этом $\delta n(0) > 0$, то объем обогащается неравновесными носителями заряда; это явление называется инжекцией неравновесных носителей заряда. Инжекция дырок в точки $x > 0$ из точки $x = 0$ наблюдается при $E > 0$, инжекция электронов — при $E < 0$. Если избыточная концентрация отрицательна — $\delta n(0) < 0$ —, то объем обедняется носителями заряда, это явление называется экстракцией.

Рассмотрим случай $l_E < 0$, ($|l_E| > 0$). Для длины затягивания l из (66.23) получим

$$\frac{1}{l} = \frac{1}{L} \frac{|l_E|^2}{2L} = \frac{|l_E|}{L^2}; \quad l = \frac{L^2}{|l_E|}, \quad (66.34)$$

т. е. с ростом поля (по модулю) длина затягивания стремится к нулю. При $\delta n(0) > 0$ происходит уменьшение, а при

$\delta n(0) < 0$ — увеличение концентрации в объеме полупроводника по сравнению со случаем $E = 0$; первый случай называют эксклюзией, второй — аккумуляцией (неравновесных) носителей заряда.

На рис. 98 приведено распределение $p - p_0$ в зависимости от направления поля и величины $\delta p(0)$. (Необходимо заметить, что в некоторых статьях эксклюзией часто называют явление, названное выше экстракцией, и наоборот.)

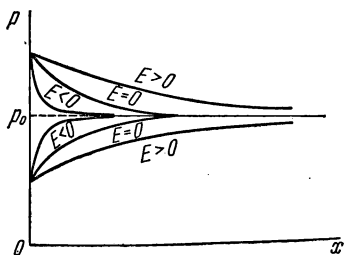


Рис. 98. Распределение избыточной концентрации дырок в случае инжекции, эксклюзии, экстракции и аккумуляции

Говоря об изменении избыточной концентрации носителей заряда, мы не связывали его с видом частиц. Более того, условием $\operatorname{div} \mathbf{E} = 0$ мы по существу предположили, что избыточные концентрации электронов и дырок меняются одинаковым образом. Чтобы доказать это, рассмотрим полупроводник, в котором каким-либо образом создан объемный заряд с плотностью $\rho(\mathbf{r}, 0)$. Введенный заряд порождает электрическое поле $\mathbf{E}(\mathbf{r}, t)$, которое связано с $\rho(\mathbf{r}, t)$ уравнением Пуассона:

$$\operatorname{div} \mathbf{E} = \frac{\rho}{\epsilon \epsilon_0}. \quad (66.35)$$

Электрическое поле \mathbf{E} вызывает ток проводимости $\mathbf{j}_E = \sigma \mathbf{E}$. Из уравнения непрерывности

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} \mathbf{j} \quad (66.36)$$

следует

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} \sigma \mathbf{E} = -\sigma \operatorname{div} \mathbf{E} = -\frac{\sigma \rho}{\epsilon \epsilon_0} = -\frac{\rho}{\tau^M}, \quad (66.37)$$

где через τ^M обозначена величина, имеющая размерность времени:

$$\tau^M = \frac{\epsilon \epsilon_0}{\sigma}. \quad (66.38)$$

Уравнение (66.37) решается элементарно:

$$\rho(\mathbf{r}, t) = \rho(\mathbf{r}, 0) e^{-\frac{t}{\tau^M}}. \quad (66.39)$$

Оценим величину τ^M , положив $\epsilon = 10$ и $\sigma = 10 \text{ Сим} \cdot \text{м}^{-1}$ и учитывая, что $\epsilon_0 = 10^7/4\pi c^2 = 8,85416 \cdot 10^{-12} \text{ Ф/м}$, получим $\tau^M \approx 10^{-11} \text{ с}$. Величину τ^M называют временем релаксации Максвелла, она определяется удельной электрической проводимостью вещества. Таким образом, в результате проводимости объемный заряд, введенный в полупроводник, существует в среднем в течение времени τ^M . Но отсюда вытекает важнейший для понимания многих вопросов физики полупроводников и полупроводниковых приборов, связанных с неравновесными носителями заряда, вывод: если создать в полупроводнике избыточную концентрацию основных носителей заряда, то объемный заряд и избыточная проводимость исчезнут в среднем через τ^M ; но если создать избыточную концентрацию неосновных носителей заряда, то благодаря релаксации Максвелла их объемный заряд будет скомпенсирован основными носителями заряда через τ^M , а избыточная концентрация неосновных и основных носителей заряда будет существовать в течение времени жизни τ неравновесных носителей заряда. Таким образом, неосновные и основные неравновесные носители заряда играют принципиально различную роль в изменении физических свойств полупроводника, поэтому обычно говорят об инжекции или экстракции только неосновных носителей заряда.

Неизменность первоначального распределения объемного заряда с течением времени, вытекающая из уравнения (66.39), возникает потому, что мы учли только ток проводимости. В действительности же будет иметь место диффузия заряда, поэтому объемный заряд вследствие диффузии меняет свое распределение в кристалле.

Резюме § 66 -

1. Рассасывание (или компенсация) объемного заряда вследствие проводимости называется релаксацией Максвелла. Время релаксации Максвелла определяется проводимостью вещества

$$\begin{array}{cc} \text{в СИ} & \text{в системе Гаусса} \\ \tau^M = \frac{\varepsilon\varepsilon_0}{\sigma}; & \tau^M = \frac{\varepsilon}{4\pi\sigma}. \end{array} \quad (66.1p)$$

2. Избыточная концентрация основных носителей заряда рассасывается через τ^M ; избыточная концентрация неосновных носителей заряда исчезает вместе с вызываемой ею избыточной концентрацией основных носителей заряда через время жизни τ_f .

3. Изменение концентрации неравновесных неосновных носителей заряда в стационарных условиях в отсутствие генерации носителей заряда за счет энергии, попадающей извне, описывается уравнением

$$\frac{1}{e} \operatorname{div} \mathbf{j} + \frac{n-n_0}{\tau_f} = 0. \quad (66.2p)$$

4. Ток в изотермических условиях и при $\mathbf{V}=0$ имеет вид

$$\mathbf{j} = e^2 K_{11} \mathbf{E} - e K_{11} \nabla F^* = \mathbf{j}_E + \mathbf{j}_D, \quad (66.3p)$$

где

$$\mathbf{j}_E = e^2 K_{11} \mathbf{E} = en\mu_d \mathbf{E} \quad (66.4p)$$

есть ток проводимости, а

$$\mathbf{j}_D = -n\mu_d \nabla F^* = -e K_{11} \nabla F^* = -e D \nabla n \quad (66.5p)$$

есть диффузионный ток.

5. Коэффициент диффузии D и подвижность носителей заряда связаны соотношением Эйнштейна:

$$D_n = \frac{kT}{e_n} \mu_{dn}; \quad D_p = \frac{kT}{e_p} \mu_{dp}; \quad (66.6p)$$

$$D_n = \frac{2F_n^*}{3e_n} \mu_{dn}; \quad D_p = \frac{2F_p^*}{3e_p} \mu_{dp} \quad (66.7p)$$

соответственно для невырожденного и вырожденного полупроводников.

6. Избыточная концентрация неосновных носителей заряда в одномерном случае определяется уравнением

$$\delta n(x) = \delta n(0) e^{-\frac{x}{l}}, \quad (66.8p)$$

где l называется длиной затягивания; она связана с диффузионной длиной $L = \sqrt{D\tau_f}$ и длиной дрейфа $l_E = \mu_a E \tau_f$ соотношением (66.23). В слабом электрическом поле распределение избыточной концентрации неосновных носителей заряда определяется диффузией, в сильном — дрейфом.

§ 67. ПОВЕРХНОСТНАЯ РЕКОМБИНАЦИЯ

На поверхности полупроводника содержится ряд дефектов, играющих роль рекомбинационных ловушек, что приводит к поверхностной рекомбинации, уменьшающей концентрацию неравновесных носителей заряда в приповерхностном слое полупроводника. Если поверхность одного и того же полупроводника обработать различным образом, например, шлифовкой, полировкой, травлением, нанесением поверхностного покрытия, то время жизни, измеренное экспериментально, будет различным, оно зависит от способа обработки поверхности. Наличие поверхностной рекомбинации можно наблюдать визуально или с помощью электронно-оптического преобразователя. На рис. 99 приведена фотография диода из арсенида галлия, в котором инжекция создается большими импульсами тока. Из рис. 99 видно, что приповерхностные слои образца светятся, в то время как объем остается темным (за исключением p - n -перехода, где возникает стимулированная рекомбинация). Поверхностная рекомбинация обычно характеризуется так называемой скоростью поверхностной рекомбинации s , которую можно ввести следующим образом.

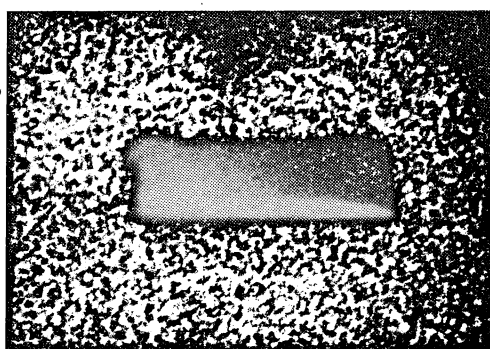


Рис. 99. Свечение на поверхности диода из арсенида галлия вследствие поверхностной рекомбинации

Запишем уравнение непрерывности в виде

$$\frac{\partial n}{\partial t} = -\frac{1}{e} \operatorname{div} \mathbf{j} - \frac{n - n_0}{\tau_f} + G, \quad (67.1)$$

которое мы решали выше в одномерном случае в предположении неограниченности полупроводника вдоль осей y и z и $G = 0$. Учтем

наличие границ у полупроводника. Внешнее электрическое поле, направленное вдоль оси образца (ось x), обеспечивает инжекцию неосновных носителей заряда. Чтобы упростить уравнение (67.1), будем считать, что электрическое поле отсутствует: $E=0$, а избыточная концентрация создается некоторым источником, дающим независимую от координаты скорость генерации носителей заряда G . В этом случае электрическое поле всюду будет равно нулю. Если рассмотреть неограниченную вдоль оси y (и x) пластину, то уравнение непрерывности снова будет одномерным. Так как на поверхности образца существует рекомбинация, то концентрация избыточных носителей заряда у поверхности должна быть меньше, чем в объеме, вдали от поверхности. Наличие градиента концентрации неравновесных носителей заряда должно создать диффузионный поток, направленный из объема полупроводника к его поверхности. Величина тока будет тем больше, чем больше происходит актов рекомбинации и чем ближе к поверхности находится элемент объема, в котором мы рассматриваем ток. Ток должен обращаться в нуль на поверхности. Пусть величина рекомбинационного тока на поверхности равна j_2^s . Уравнение непрерывности (67.1) в стационарных условиях может быть записано в виде

$$0 = -\frac{1}{e} \operatorname{div} \mathbf{j} - \frac{n-n_0}{\tau_f} + G. \quad (67.2)$$

Подставляя вместо \mathbf{j} величину $-eD\nabla_z n$, получим

$$D \frac{d^2 n}{dz^2} - \frac{n-n_0}{\tau_f} + G = 0, \quad (67.3)$$

или

$$\frac{d^2 (n-n_0)}{dz^2} - \frac{n-n_0}{L^2} + \frac{G}{D} = 0. \quad (67.4)$$

Общее решение однородного уравнения имеет вид

$$\delta n(z) = n(z) - n_0 = A e^{-\frac{z}{L}} + B e^{\frac{z}{L}}. \quad (67.5)$$

Частное решение неоднородного уравнения (67.3) можно выбрать следующим образом:

$$\delta n = G\tau_f. \quad (67.6)$$

Общее решение неоднородного уравнения равно их сумме:

$$\delta n(z) = G\tau_f \left[1 - A' e^{-\frac{z}{L}} - B' e^{\frac{z}{L}} \right], \quad (67.7)$$

где

$$A' = -\frac{A}{G\tau_f}; \quad B' = -\frac{B}{G\tau_f}. \quad (67.8)$$

Величины A' и B' находятся из граничных условий

$$\delta n(0) = G\tau_f [1 - A' - B'],$$

$$\delta n(d) = G\tau_f \left[1 - A'e^{-\frac{d}{L}} - B'e^{\frac{d}{L}} \right], \quad (67.9)$$

где d — толщина пластины. Выразим граничные условия через ток $j_z^s(0)$ и $j_z^s(d)$. Поскольку

$$j_z^s = -eD \frac{dn}{dz} = -eDG\tau_f \left[\frac{A'}{L} e^{-\frac{z}{L}} - \frac{B'}{L} e^{\frac{z}{L}} \right], \quad (67.10)$$

то
$$j_z^s(0) = -eGL(A' - B'), \quad (67.11)$$

$$j_z^s(d) = -eGL \left(A'e^{-\frac{d}{L}} - B'e^{\frac{d}{L}} \right). \quad (67.12)$$

Из (67.11) и (67.12) можно получить выражения для A' и B' :

$$A' = \frac{-j_z^s(0) e^{\frac{d}{L}} + j_z^s(d)}{2eLG \operatorname{sh} \frac{d}{L}}, \quad (67.13)$$

$$B' = \frac{-j_z^s(0) e^{-\frac{d}{L}} + j_z^s(d)}{2eLG \operatorname{sh} \frac{d}{L}}. \quad (67.14)$$

Рекомбинационный поверхностный ток можно выразить через избыточную концентрацию носителей заряда на поверхности δn^s :

$$j^s = e\delta n^s s, \quad (67.15)$$

где s — коэффициент пропорциональности, связывающий ток j^s с избыточной концентрацией. Очевидно, что размерность s есть размерность скорости: $[s] = LT^{-1}$. Величина s получила название скорости поверхностной рекомбинации. Можно выразить коэффициенты A' и B' через скорости рекомбинации $s_1(z=0)$ и $s_2(z=d)$. Для этого необходимо выразить δn из (67.9–11) и (67.15):

$$\delta n(0) = G\tau_f [1 - A' - B'] = \frac{j^s(0)}{es_1} = -\frac{GL}{s_1} [A' - B'], \quad (67.16)$$

$$\begin{aligned} \delta n(d) &= G\tau_f \left[1 - A'e^{-\frac{d}{L}} - B'e^{\frac{d}{L}} \right] = \frac{j^s(d)}{es_2} = \\ &= -\frac{GL}{s_2} \left[A'e^{-\frac{d}{L}} - B'e^{\frac{d}{L}} \right]. \end{aligned} \quad (67.17)$$

Решая систему уравнений (67.16, 17), получим выражение для A' и B' :

$$A' = \frac{s_1(v_D + s_2)e^{\frac{d}{L}} - s_2(v_D + s_1)}{2(s_1s_2 - v_D^2) \operatorname{sh} \frac{d}{L} + 2v_D(s_1 - s_2) \operatorname{ch} \frac{d}{L}}, \quad (67.18)$$

$$B' = \frac{s_1(v_D - s_2)e^{-\frac{d}{L}} - s_2(v_D - s_1)}{2(s_1s_2 - v_D^2) \operatorname{sh} \frac{d}{L} + 2v_D(s_1 - s_2) \operatorname{ch} \frac{d}{L}}. \quad (67.19)$$

В выражении (67.18, 19) $v_D = L/\tau_f$ — диффузионная скорость. Величины s_1 и s_2 имеют разный знак; поскольку направление тока к плоскости $z=d$ положительно, а к плоскости $z=0$ — отрицательно, будем считать $s_1 < 0$; $s_2 > 0$.

Рассмотрим распределение $\delta n(z)$ при разных значениях s_1 и s_2 .

1. Поверхностная рекомбинация отсутствует: $s_1 = s_2 = 0$. Из (67.18, 19), видно, что $A' = B'$ и $\delta n(z) = G\tau_f$, т. е. δn не зависит от координаты.

2. Бесконечная скорость поверхностной рекомбинации: $|s_1| = s_2 = \infty$. Физически это означает, что избыточная концентрация носителей заряда, приходящих на поверхность, обращается в нуль вследствие рекомбинации. Из (67.18, 19) следует

$$A' = \frac{e^{\frac{d}{L}} - 1}{2 \operatorname{sh} \frac{d}{L}}; \quad B' = \frac{1 - e^{-\frac{d}{L}}}{2 \operatorname{sh} \frac{d}{L}} \quad (67.20)$$

и

$$\delta n(z) = G\tau_f \left[1 - \frac{\operatorname{sh} \frac{z}{L} + \operatorname{sh} \frac{d-z}{L}}{\operatorname{sh} \frac{d}{L}} \right]. \quad (67.21)$$

Из (67.21) следует, что $\delta n(0) = \delta n(d) = 0$.

3. Аналогично можно исследовать и другие случаи, например, при $s_1 = 0$ и $s_2 = \infty$:

$$A' = B' = \frac{1}{2 \operatorname{ch} \frac{d}{L}} \quad (67.22)$$

и

$$\delta n(z) = G\tau_f \left[1 - \frac{\operatorname{ch} \frac{z}{L}}{\operatorname{ch} \frac{d}{L}} \right]. \quad (67.23)$$

При $z=d$ избыточная концентрация обращается в нуль — $\delta n(d)=0$, — поскольку $s_2 = \infty$. В точке $z=0$

$$\delta n(0) = G\tau_f \left[1 - \frac{1}{\operatorname{ch} \frac{d}{L}} \right].$$

При малой диффузионной длине и большой толщине $\delta n(0) = G\tau_f$. Но если d сравнимо с L , то $\delta n < G\tau_f$, поскольку большая скорость рекомбинации на плоскости $z=d$ приводит к сильному обеднению объема полупроводника. В общем случае *обеднение должно наблюдаться на расстоянии порядка диффузионной длины от поверхности*. На рис. 100 приведено распределение $\delta n(z)$ для случая $d=4L$ при разных значениях s_1 и при $s_2 = \infty$.

В германии наблюдались скорости поверхностной рекомбинации от 50 до 10^5 см/с и более.

Необходимо заметить, что ток, вызванный поверхностной рекомбинацией, не приводит к появлению электрических полей, поскольку при этом электронейтральность не нарушается.

Иногда для характеристики роли поверхностной рекомбинации вводят понятие эффективного времени жизни τ_f^{eff} , которое определяется условием

$$\frac{1}{\tau_f^{eff}} = \frac{1}{\tau_f^v} + \frac{1}{\tau_f^s}, \quad (67.24)$$

где τ_f^v — время жизни в объеме полупроводника, τ_f^s — на поверхности полупроводника.

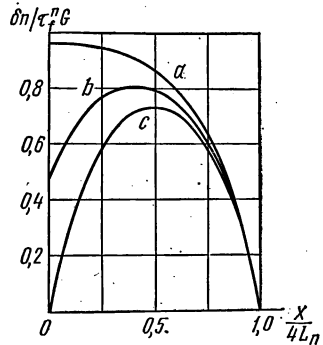


Рис. 100. Влияние скорости поверхностной рекомбинации на избыточную концентрацию неосновных носителей заряда: а) $s_1 = 0$; б) $|s_1| = v_D$; в) $|s_1| = \infty$; во всех случаях $s_2 = \infty$, толщина образца $4L_n$

Глава VII

КОНТАКТНЫЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

§ 68. ДЕБАЕВА ДЛИНА ЭКРАНИРОВАНИЯ

При приведении в контакт двух веществ между ними возникает обмен носителями заряда, что приводит к изменению свойств полупроводника не только на контактной поверхности, но и в объеме.

Практическое использование полупроводников и исследование многих физических явлений в них неизбежно связано с включением полупроводника в определенную цепь, состоящую из различных веществ. Для понимания многих явлений, обусловленных наличием контакта, необходимо рассмотреть изменения свойств полупроводника, внесенного в электрическое поле.

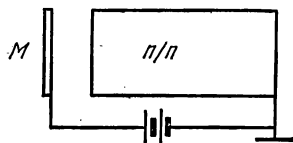


Рис. 101. Полупроводник в однородном электрическом поле

Рассмотрим полупроводник, внесенный для простоты в однородное электрическое поле конденсатора (рис.101). Оно приведет к перераспределению носителей заряда

в полупроводнике, в результате чего в нем возникнет объемный заряд $\rho(r)$ и электрическое поле E , связанное с объемным зарядом уравнением Пуассона

$$\operatorname{div} \varepsilon_0 E(r) = \rho(r). \quad (68.1)$$

Если выразить поле E через потенциал, то уравнение Пуассона (68.1) будет иметь вид

$$\nabla^2 \varphi(r) = -\frac{\rho(r)}{\varepsilon_0}. \quad (68.2)$$

Потенциальная энергия электрона $V(r)$ связана с $\varphi(r)$: $V(r) = e\varphi(r)$. Но, как было показано в § 19, потенциальное электрическое поле $V(r)$ искривляет зоны энергии таким образом, что $E(r) = E_c + V(r)$, или

$$E_c(r) = E_c + V(r). \quad (68.3)$$

Смещение испытывают все уровни энергии, в том числе и дискретные, лежащие в запрещенной зоне. Но если полупроводник находится в состоянии термодинамического равновесия, то уровень Ферми в нем постоянен, следовательно, расстояние между ним и зонами энергии меняется:

без поля

$$E_c - F; F - E_v; \quad (68.4)$$

с полем

$$E_c + V(\mathbf{r}) - F; F - [E_v + V(\mathbf{r})]. \quad (68.5)$$

Сравнивая (68.5) с (68.4), видим, что если расстояние между F и E_c увеличивается на $V(\mathbf{r})$, то между F и E_v уменьшается на ту же величину, и наоборот. Поскольку меняется расстояние между F и уровнями энергии, то меняется и распределение частиц по уровням энергии.

Согласно (30.5), уравнение электронейтральности имеет вид

$$n + n_d - p - p_a = N_d - N_a. \quad (68.6)$$

Поле $V(\mathbf{r})$ меняет входящие в (68.6) величины по-разному, что и приводит к появлению объемного заряда:

$$\rho(\mathbf{r}) = e(\delta n + \delta n_d - \delta p - \delta p_a) \quad (e = e_n). \quad (68.7)$$

Если обозначить равновесные значения концентрации частиц с индексом 0, а при наложении поля — без индекса 0, то

$$\begin{aligned} \delta n &= n(\mathbf{r}) - n_0; \quad \delta p = p(\mathbf{r}) - p_0; \quad \delta n_d = n_d(\mathbf{r}) - n_{d0}; \\ \delta p_a &= p_a(\mathbf{r}) - p_{a0}. \end{aligned} \quad (68.8)$$

Для невырожденного полупроводника имеем

$$n(\mathbf{r}) = N_c e^{-\frac{E_c(\mathbf{r}) - F}{kT}} = N_c e^{-\frac{E_c - F}{kT}} e^{-\frac{V(\mathbf{r})}{kT}} = n_0 e^{-\frac{V(\mathbf{r})}{kT}}, \quad (68.9)$$

$$p(\mathbf{r}) = p_0 e^{\frac{V(\mathbf{r})}{kT}}, \quad (68.10)$$

$$n_d(\mathbf{r}) = \frac{N_d}{\frac{1}{2} e^{\frac{E_d + V(\mathbf{r}) - F}{kT}} + 1}; \quad p_a(\mathbf{r}) = \frac{N_a}{\frac{1}{2} e^{\frac{F - E_a - V(\mathbf{r})}{kT}} + 1}. \quad (68.11)$$

Умножая (68.2) на заряд электрона $e = e_n$ и учитывая (68.7 - 11), получим уравнение для $V(\mathbf{r})$:

$$\begin{aligned} \nabla^2 V(\mathbf{r}) &= -\frac{e\rho(\mathbf{r})}{\epsilon\epsilon_0} = -\frac{e^2}{\epsilon\epsilon_0} (\delta n + \delta n_d - \delta p - \delta p_a) = \\ &= -\frac{e^2}{\epsilon\epsilon_0} \left\{ \left[n_0 \left(e^{-\frac{V(\mathbf{r})}{kT}} - 1 \right) - p_0 \left(e^{\frac{V(\mathbf{r})}{kT}} - 1 \right) \right] + (\delta n_d - \delta p_a) \right\}. \end{aligned} \quad (68.12)$$

Выражения для δn_d и δp_a имеют сложный вид, поэтому мы сохранили краткую форму записи.

Первый интеграл уравнения (68.12) находится стандартным способом: умножим (68.12) на $\nabla V(\mathbf{r})$, после чего левая часть уравнения примет вид

$$[\nabla V(\mathbf{r})] \nabla^2 V(\mathbf{r}) = \frac{1}{2} \nabla [\nabla V(\mathbf{r})]^2. \quad (68.13)$$

Выражение, стоящее в квадратных скобках (68.12), можно преобразовать следующим образом:

$$-\frac{e^2}{\varepsilon\varepsilon_0} \left[n_0 \left(e^{-\frac{V(\mathbf{r})}{kT}} - 1 \right) - p_0 \left(e^{\frac{V(\mathbf{r})}{kT}} - 1 \right) \right] \nabla V(\mathbf{r}) = \\ = \nabla \left\{ -\frac{e^2}{\varepsilon\varepsilon_0} \left[n_0 \left(-e^{-\frac{V(\mathbf{r})}{kT}} kT - V(\mathbf{r}) \right) - p_0 \left(e^{\frac{V(\mathbf{r})}{kT}} kT - V(\mathbf{r}) \right) \right] \right\}. \quad (68.14)$$

Предположим, что аналогично можно подобрать некоторую функцию $\Omega(\mathbf{r})$ таким образом, что

$$(\delta n_x - \delta p_x) \nabla V(\mathbf{r}) = \nabla \Omega(\mathbf{r}). \quad (68.15)$$

В таком случае первый интеграл уравнения (68.12) можно записать в виде

$$\pm |\nabla V(\mathbf{r})| = \left\{ \frac{2e^2 kT}{\varepsilon\varepsilon_0} \left[n_0 \left(e^{-\frac{V(\mathbf{r})}{kT}} + \frac{V(\mathbf{r})}{kT} \right) + p_0 \left(e^{\frac{V(\mathbf{r})}{kT}} - \frac{V(\mathbf{r})}{kT} \right) \right] - \right. \\ \left. - \frac{2e^2}{\varepsilon\varepsilon_0} \Omega(\mathbf{r}) + C \right\}^{\frac{1}{2}}. \quad (68.16)$$

Константа интегрирования C определится из граничных условий для V и ∇V . Дальнейшее решение уравнения (68.16) в общем виде практически невозможно, в связи с чем перейдем к рассмотрению некоторых частных случаев.

1. Собственный полупроводник. В собственном полупроводнике $N_x = N_a = 0$, поэтому $\delta n_x = \delta p_x = 0$, кроме того, $n_0 = p_0 = n_i$, и уравнение (68.16) принимает вид

$$\pm |\nabla V(\mathbf{r})| = \sqrt{\frac{2e^2 n_i kT}{\varepsilon\varepsilon_0} \left[e^{\frac{V(\mathbf{r})}{kT}} + e^{-\frac{V(\mathbf{r})}{kT}} + C \right]^{\frac{1}{2}}}. \quad (68.17)$$

Если за начало отсчета V принять потенциальную энергию в точках, в которых поле равно нулю, то из условия $\nabla V = \dot{V} = 0$ следует

$$0 = [2 + C]^{1/2}; \quad C = -2 \quad (68.18)$$

и

$$\pm |\nabla V(\mathbf{r})| = 2 \sqrt{\frac{e^2 n_i kT}{\varepsilon\varepsilon_0} \left(\operatorname{ch} \frac{V(\mathbf{r})}{kT} - 1 \right)^{\frac{1}{2}}} = 2 \sqrt{\frac{2e^2 n_i kT}{\varepsilon\varepsilon_0}} \operatorname{sh} \frac{V(\mathbf{r})}{2kT}. \quad (68.19)$$

Ограничимся одномерным случаем

$$\frac{dV(x)}{dx} = \pm 2 \sqrt{\frac{2e^2 n_i kT}{\varepsilon\varepsilon_0}} \operatorname{sh} \frac{V(x)}{2kT}. \quad (68.20)$$

Разделив в (68.20) переменные, получим

$$\pm x + C_1 = \int \frac{1}{2} \sqrt{\frac{\varepsilon\varepsilon_0}{2e^2 n_i kT}} \frac{dV}{\operatorname{sh} \frac{V}{kT}}. \quad (68.21)$$

Интеграл в (68.21) вычисляется следующим образом:

$$\int \frac{d\xi}{\operatorname{sh} \xi} = \int \frac{d\xi}{2 \operatorname{sh} \frac{\xi}{2} \operatorname{ch} \frac{\xi}{2}} = \int \frac{d \frac{\xi}{2}}{\frac{\operatorname{sh} \frac{\xi}{2}}{\operatorname{ch} \frac{\xi}{2}} \operatorname{ch}^2 \frac{\xi}{2}} = \int \frac{d \operatorname{th} \frac{\xi}{2}}{\operatorname{th} \frac{\xi}{2}} = \ln \operatorname{th} \frac{\xi}{2}, \quad (68.22)$$

поэтому для (68.21) имеем

$$\pm x + C_1 = \sqrt{\frac{\varepsilon \varepsilon_0 k T}{2e^2 n_i}} \ln \operatorname{th} \frac{V(x)}{4kT}. \quad (68.23)$$

Введем обозначение

$$\sqrt{\frac{\varepsilon \varepsilon_0 k T}{2e^2 n_i}} = L^D, \quad (68.24)$$

L^D называют дебаевой длиной экранирования. Так как гиперболический тангенс не превосходит единицы, а $x > 0$, то при x необходимо сохранить только знак минус:

$$-x + C_1 = L^D \ln \operatorname{th} \frac{V(x)}{4kT}. \quad (68.25)$$

Величина C_1 определится из условия на поверхности: $V(0) = V_s$ при $x=0$, следовательно,

$$C_1 = L^D \ln \operatorname{th} \frac{V_s}{4kT} \quad (C_1 < 0). \quad (68.26)$$

Выразим $\operatorname{th} \frac{V(x)}{4kT}$ через x , учитывая (68.25):

$$\operatorname{th} \frac{V(x)}{4kT} = e^{\frac{C_1 - x}{L^D}}. \quad (68.27)$$

Решая уравнение (68.27) относительно $V(x)$, получим

$$V(x) = 2kT \ln \left(\frac{1 + e^{\frac{C_1 - x}{L^D}}}{1 - e^{\frac{C_1 - x}{L^D}}} \right) = 2kT \ln \left(\frac{1 + e^{-\frac{x}{L^D}} \operatorname{th} \frac{V_s}{4kT}}{1 - e^{-\frac{x}{L^D}} \operatorname{th} \frac{V_s}{4kT}} \right). \quad (68.28)$$

Уравнение (68.28) описывает смещение зон энергии как функцию координаты x и V_s . Мы видим, что величина смещения монотонно меняется от V_s при $x=0$ до нуля при $x \rightarrow \infty$. Учитывая, что гиперболический тангенс не превосходит единицы, упростим выражение для $V(x)$ в случае малых и больших x . При $x \sim (2 \div 3) L^D$ экспонента будет много меньше единицы, в результате чего можем записать, пользуясь соотношением $\ln(1 + \xi) \approx \xi$:

$$V(x) \approx 2kT \cdot 2 \operatorname{th} \frac{V_s}{4kT} e^{-\frac{x}{L^D}} = 4kT \operatorname{th} \frac{V_s}{4kT} e^{-\frac{x}{L^D}}. \quad (68.29)$$

Если $\frac{V_s}{4kT} \ll 1$, то выражение (68.29) будет справедливо при любом x :

$$V(x) = V_s e^{-\frac{x}{L^D}}. \quad (68.30)$$

Рассмотрим случай малых x . Чтобы разложить выражение (68.28) в ряд по x в точке $x=0$, найдем dV/dx и d^2V/dx^2 . Опуская простые, но громоздкие вычисления, запишем

$$\frac{dV(x)}{dx} = -\frac{4kT}{L^D} \operatorname{th} \frac{V_s}{4kT} \frac{e^{-\frac{x}{L^D}}}{1 - \operatorname{th}^2 \frac{V_s}{4kT} e^{-\frac{2x}{L^D}}}; \quad (68.31)$$

$$\frac{d^2V(x)}{dx^2} = \frac{4kT \operatorname{th} \frac{V_s}{4kT}}{L^{2D}} \frac{\left(1 + \operatorname{th}^2 \frac{V_s}{4kT} e^{-\frac{2x}{L^D}}\right) e^{-\frac{x}{L^D}}}{\left(1 - \operatorname{th}^2 \frac{V_s}{4kT} e^{-\frac{2x}{L^D}}\right)^2}. \quad (68.32)$$

Полагая $x=0$, получим

$$\left. \frac{dV(x)}{dx} \right|_{x=0} = -eE_s = -\frac{4kT}{L^D} \operatorname{th} \frac{V_s}{4kT} \frac{1}{1 - \operatorname{th}^2 \frac{V_s}{4kT}} = -\frac{2kT}{L^D} \operatorname{sh} \frac{V_s}{2kT}. \quad (68.33)$$

Ограничиваясь случаем малых x , запишем

$$V(x) = V_s - \frac{2kT}{L^D} \operatorname{sh} \frac{V_s}{2kT} x = V_s \left(1 - \frac{2kT}{V_s} \operatorname{sh} \frac{V_s}{2kT} \cdot \frac{x}{L^D}\right). \quad (68.34)$$

Таким образом, при $x \ll L^D$ потенциальная энергия $V(x)$ изменяется по линейному закону, который при $x > L^D$ переходит в экспоненциальный. Изменение $V(x)$ при больших V_s ($V_s \gg kT$) происходит в области толщины порядка L^D .

Напряженность электрического поля $E(x)$ определяется из соотношения (68.31):

$$E(x) = -\frac{1}{e} \frac{dV(x)}{dx} = \frac{4kT}{eL^D} \operatorname{th} \frac{V_s}{4kT} \frac{e^{-\frac{x}{L^D}}}{1 - \operatorname{th}^2 \frac{V_s}{4kT} e^{-\frac{2x}{L^D}}}. \quad (68.35)$$

На поверхности $x=0$ в соответствии с (68.33)

$$E(0) = E_s = \frac{2kT}{eL^D} \operatorname{sh} \frac{V_s}{2kT}.$$

Плотность объемного заряда $\rho(x)$ определим из (68.2) и (68.32)

$$\rho(x) = -\frac{\epsilon\epsilon_0}{e} \frac{d^2V(x)}{dx^2} = -\frac{4\epsilon\epsilon_0 kT \operatorname{th} \frac{V_s}{4kT} \left[1 + \operatorname{th}^2 \frac{V_s}{4kT} \cdot e^{-\frac{2x}{LD}} \right] e^{-\frac{x}{LD}}}{eL^D \left[1 - \operatorname{th}^2 \frac{V_s}{4kT} e^{-\frac{2x}{LD}} \right]^2}. \quad (68.37)$$

Вычислим объемную плотность заряда ρ_s на поверхности полупроводника:

$$\rho_s = \rho(0) = -\frac{4\epsilon\epsilon_0 kT}{eL^D} \frac{\operatorname{th} \frac{V_s}{4kT} \left(1 + \operatorname{th}^2 \frac{V_s}{4kT} \right)}{\left(1 - \operatorname{th}^2 \frac{V_s}{4kT} \right)^2}. \quad (68.38)$$

Учитывая, что

$$\frac{\operatorname{th} \xi (1 + \operatorname{th}^2 \xi)}{(1 - \operatorname{th}^2 \xi)^2} = \frac{1}{4} \operatorname{sh} 4\xi, \quad (68.39)$$

и подставляя вместо L^D его выражение (68.24), получим

$$\rho_s = -2en_i \operatorname{sh} \frac{V_s}{kT} = en_i \left(e^{-\frac{V_s}{kT}} - e^{\frac{V_s}{kT}} \right). \quad (68.40)$$

Выражение (68.40) можно получить непосредственно из (68.7) с учетом (68.9—10). Мы видим, что знак объемной плотности заряда на поверхности полупроводника ρ_s зависит от знака V_s . При $V_s > 0$ (зоны энергии смещаются вверх) $\rho_s > 0$, при $V_s < 0$ и $\rho_s < 0$. Величина $|\rho_s|$ от знака V_s не зависит.

Полученный результат означает, что у поверхности полупроводника образуется объемный заряд в результате того, что концентрация электронов ($V_s < 0$) или дырок ($V_s > 0$) возрастает. Толщина области объемного заряда невелика, поскольку при малых x функция $V(x)$ пропорциональна x , поэтому $\rho(x)$ уже при малых x уменьшается по экспоненте, при больших x плотность заряда $\rho(x)$ уменьшается еще быстрее с ростом x . В связи с тем, что концентрация носителей заряда у поверхности может значительно превосходить концентрацию частиц в объеме

$$n_s \approx n_i e^{\frac{V_s}{kT}}, \quad (68.41)$$

то проводимость поверхностных слоев может резко возрасти. Это явление называется *эффектом поля*. Рост концентрации носителей заряда в приповерхностных слоях может быть обусловлен не только внешним полем. Как мы видели в § 22, на поверхности полупроводника существует значительное число поверхностных состояний, уровни энергии которых могут находиться в запрещенной зоне.

Поверхностные состояния могут захватить свободные носители заряда — электроны или дырки, в результате чего на поверхности полупроводника создается *поверхностный заряд*, вызывающий электрическое поле, которое окажет такое же действие на свойства

полупроводника в объеме, как и внешнее электрическое поле. На рис. 102 представлена зависимость V , E и ρ от x .

2. Примесный полупроводник. Рассмотрим полупроводник, содержащий, например, донорную примесь, при температурах, когда примесь остается полностью ионизованной:

$$n_0 = N_D^+ = N_D, \quad p_0 = \frac{n_i^2}{N_D}. \quad (68.42)$$

Пренебрегая изменением степени ионизации примеси, получим согласно (68.16)

$$\frac{dV(x)}{dx} = \left\{ \frac{2e^2 kT n_0}{\epsilon \epsilon_0} \left[e^{-\frac{V}{kT}} + \frac{V}{kT} + \frac{n_i^2}{n_0^2} \left(e^{\frac{V}{kT}} - \frac{V}{kT} \right) + C \right] \right\}^{1/2}. \quad (68.43)$$

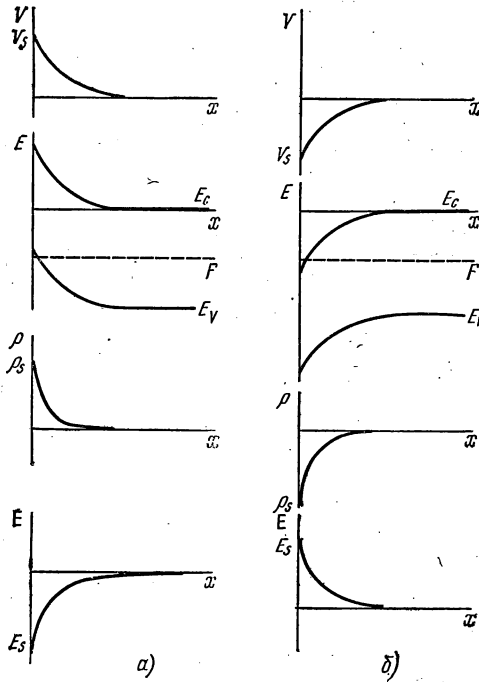


Рис. 102. Искривление зон энергии во внешнем поле; объемный заряд и поле в полупроводнике

Определив постоянную интегрирования тем же условием: $V = 0$ при $\nabla_x V = 0$, получим

$$1 + \frac{n_i^2}{n_0^2} + C = 0; \quad C = -\left(1 + \frac{n_i^2}{n_0^2}\right) \quad (68.44)$$

и

$$\frac{dV(x)}{dx} = \pm \left\{ \frac{2e^2 n_0 kT}{\epsilon \epsilon_0} \left[e^{-\frac{V}{kT}} + \frac{V}{kT} \left(1 - \frac{n_i^2}{N_D^2}\right) + \frac{n_i^2}{N_D^2} e^{\frac{V}{kT}} - \left(1 + \frac{n_i^2}{N_D^2}\right) \right] \right\}^{1/2}. \quad (68.45)$$

Если полупроводник легирован достаточно сильно, то $n_i^2 \ll N_D^2$. Например, для германия, содержащего примесь $N_D = 10^{16}$ при $T \approx 300 \text{ K}$ $\frac{n_i^2}{N_D^2} \approx 10^{-6}$. Пренебрегая величиной $\frac{n_i^2}{N_D^2}$ по сравнению с еди-

ницей, получим для (68.45):

$$\frac{dV(x)}{dx} = \pm \left\{ \frac{2e^2 kT n_0}{\varepsilon \varepsilon_0} \left[e^{-\frac{V(x)}{kT}} + \frac{V}{kT} - 1 + \frac{n_i^2}{N_D^2} e^{\frac{V}{kT}} \right] \right\}^{1/2}. \quad (68.46)$$

При $V \ll kT$ уравнение (68.46) еще упрощается

$$\frac{dV(x)}{dx} = \pm \left\{ \frac{2e^2 kT n_0}{\varepsilon \varepsilon_0} \frac{n_i^2}{N_D^2} \right\}^{1/2} e^{\frac{V}{2kT}}, \quad \left(\frac{n_i^2}{N_D^2} \gg \frac{V}{kT} \right) \quad (68.47)$$

и

$$\frac{dV(x)}{dx} = \pm \left\{ \frac{e^2 n_0}{\varepsilon \varepsilon_0 kT} \right\}^{1/2} V, \quad \left(\frac{n_i^2}{N_D^2} \ll \frac{V}{kT} \right). \quad (68.48)$$

Обозначая в последнем случае величину

$$\sqrt{\frac{\varepsilon \varepsilon_0 kT}{e^2 n_0}} = L^D, \quad (68.49)$$

запишем

$$\frac{dV(x)}{dx} = \pm \frac{V}{L^D}, \quad (68.50)$$

откуда конечное при $x \rightarrow \infty$ решение имеет вид

$$V(x) = V_s e^{-\frac{x}{L^D}}. \quad (68.51)$$

Так же, как и в собственном полупроводнике, происходит *искривление зон энергии и изменение концентраций электронов и дырок*. При $V_s < 0$ увеличивается концентрация электронов, при $V_s > 0$ — дырок, поэтому *приповерхностный слой обедняется основными носителями заряда и обогащается неосновными*. Аналогично ведет себя полупроводник и в случае, описываемом уравнением (68.47).

При $|V| \gg kT$ уравнение (68.46) можно упростить следующим образом. Если $V < 0$, то, сохраняя максимальный по величине член $e^{\frac{V}{kT}}$, будем иметь

$$\frac{dV(x)}{dx} = \pm \left(\frac{2e^2 kT n_0}{\varepsilon \varepsilon_0} \right)^{1/2} e^{-\frac{V}{2kT}}. \quad (68.52)$$

Решая уравнение (68.52), находим

$$e^{\frac{V}{2kT}} = \pm \left(\frac{e^2 n_0}{2\varepsilon \varepsilon_0 kT} \right)^{1/2} x + C. \quad (68.53)$$

Полагая $x = 0$; $V(0) = V_s$, получим

$$C = e^{\frac{V_s}{2kT}} \quad (C \ll 1). \quad (68.54)$$

Из (68.53) имеем асимптотическое решение

$$V(x) = 2kT \ln \left(C + \frac{x}{L D'} \right); \quad L D' = \sqrt{\frac{2\epsilon\epsilon_0 kT}{e^2 n_0}}, \quad (68.55)$$

оно справедливо при таких x , при которых $-\frac{V(x)}{kT} \gg 1$. Найдем изменение концентраций электронов и дырок:

$$n(x) = n_0 e^{-\frac{V}{kT}} = \frac{n_0}{\left(C + \frac{x}{L D'} \right)^2}, \quad (68.56)$$

т. е. концентрация электрона уменьшается по квадратичной гиперболе, приповерхностный слой обогащается электронами:

$$n_s = n(0) = n_0 e^{-\frac{V_s}{kT}} \quad (V_s < 0). \quad (68.57)$$

При $V > 0$ в уравнении (68.46) оставим член $\frac{V}{kT}$:

$$\frac{dV(x)}{dx} = \pm \left(\frac{2e^2 n_0}{\epsilon\epsilon_0} \right)^{1/2} V^{1/2} \quad (68.58)$$

и

$$V^{1/2}(x) = \pm 2 \left(\frac{2e^2 n_0}{\epsilon\epsilon_0} \right)^{1/2} x + C. \quad (68.59)$$

Постоянная интегрирования находится из граничных условий $V(0) = V_s = C^2$, т. е. $C = V_s^{1/2}$, поэтому запишем

$$V(x) = \left[\sqrt{V_s} \pm 2 \left(\frac{2e^2 n_0}{\epsilon\epsilon_0} \right)^{1/2} x \right]^2 = V_s \left[1 \pm 2 \left(\frac{2e^2 n_0}{\epsilon\epsilon_0 V_s} \right)^{1/2} x \right]^2. \quad (68.60)$$

Полагая $\frac{1}{2} \left(\frac{\epsilon\epsilon_0 V_s}{2e^2 n_0} \right)^{1/2} = t$, получим

$$V(x) = V_s \left(1 \pm \frac{x}{t} \right)^2. \quad (68.61)$$

Поскольку $V(x) > 0$, то, сохраняя знак минус, получим

$$V(x) = V_s \left(1 - \frac{x}{t} \right)^2. \quad (68.62)$$

Из этого уравнения следует, что напряженность поля линейно зависит от координаты, а плотность заряда постоянна. У поверхности возникает обедненный основными носителями заряда слой; концентрация неосновных носителей заряда в нем возрастает.

В заключение рассмотрим случай столь больших полей, что при $V > 0$ и $\frac{V}{kT} \gg 1$

$$\frac{n_i^2}{N_D^2} e^{\frac{V}{kT}} \gg \frac{V}{kT} \gg 1. \quad (68.63)$$

Согласно (68.46) для этого случая имеем

$$\frac{dV(x)}{dx} = \pm \left(\frac{2e^2 k T n_0}{\epsilon \epsilon_0} \right)^{1/2} \frac{n_i}{N_d} e^{\frac{V}{2kT}} \quad (68.64)$$

и

$$V(x) = -2kT \ln \left[e^{-\frac{V_s}{kT}} + \frac{n_i}{n_0} \frac{x}{\sqrt{2} L_D} \right]. \quad (68.65)$$

Найдем концентрацию дырок при $x=0$

$$p_s = p_0 e^{\frac{V_s}{kT}} = \frac{n_i^2}{N_d^2} N_d e^{\frac{V_s}{kT}} \gg N_d = n_0, \quad (68.66)$$

следовательно, в этом случае в *приповерхностном слое происходит изменение типа основных носителей заряда — возникает инверсный, или обращенный, слой*. Поскольку у поверхности тип проводимости меняется, при удалении от поверхности можно найти слой, в котором $p \approx n$, т. е. *слой с собственной проводимостью, или i -слой*, а область полупроводника, в которой меняется тип проводимости, носит название *физического p - n -перехода*. Он исчезает при снятии внешнего поля.

Результаты аналитического решения уравнения Пуассона качественно интерпретируются следующим образом. Внешнее электрическое поле перемещает свободные носители заряда в полупроводнике до тех пор, пока они своим зарядом не скомпенсируют внешнее поле. Это значит, что при подаче на металл отрицательного потенциала относительно полупроводника поле направлено от полупроводника к металлу, оно перемещает электроны в глубь полупроводника, а дырки к поверхности, и приповерхностный слой заряжаются положительно. Положительный потенциал металла приводит к появлению отрицательного заряда в приповерхностном слое. В зависимости от типа электропроводности полупроводника это приводит либо к обогащению, либо к обеднению слоя носителями заряда с последующей инверсией типа электропроводности. Если приложение поля происходит мгновенно, то заряд образуется в течение времени, равном времени релаксации Максвелла. Указанную на рис. 103 структуру, состоящую из металлического электрода, полупроводника и разделяющего их слоя диэлектрика — воздуха, — можно рассматривать как плоский конденсатор, одной из обкладок которого и является полупроводник. Это позволяет оценивать величину объемного заряда по толщине слоя диэлектрика и разности потенциалов между полупроводником и металлом. Изменяя потенциал металла, можно изменять объемный заряд. Если система электродов M позволяет получить электрическое поле определенной конфигурации, то и заряд в приповерхностном слое будет отражать конфигурацию внешнего поля. Меняя конфигурацию внешнего поля, мы тем самым изменим конфигурацию объемного заряда и проводимость приповерхностных слоев. Перемещение заряда вдоль поверхности возможно при созда-

нии поля вдоль поверхности либо с помощью дополнительных электродов, либо с помощью системы поверхностных электродов. Рассматриваемая структура лежит в основе одного из важнейших направлений микроэлектроники — микроэлектроники МДП-структур, в которых в качестве диэлектрических слоев используют тонкие слои изоляционных материалов, наносимых на полупроводник тем или другим способом. Если используется слой окисла, то такие структуры называют МОП-структурами, или в английской транскрипции

MOS-структурами. Приборы на основе МДП-структур обычно называют полевыми. Широкое распространение МДП-структур обеспечивается развитием планарной технологии.

Возникновение заряда под действием поля происходит в условиях термодинамического равновесия. Однако он может быть образован и неравновесными носителями заряда. Действительно, если под действием света создать

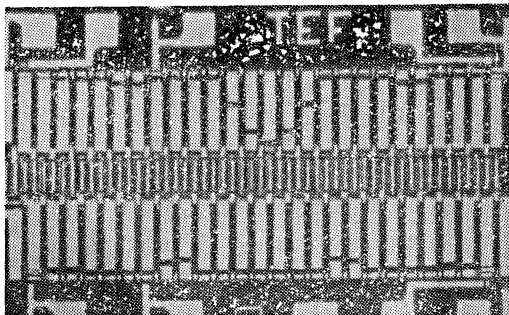


Рис. 103. Использование эффекта поля в кремнии для формирования изображений

локально объемный заряд, то он должен существовать в течение времени, равном в среднем времени жизни, после чего исчезнет вследствие рекомбинации. Это значит, что избыточный заряд способен хранить информацию о создавшем его свете по крайней мере в течение времени жизни. Меняя конфигурацию поля, можно заставить этот локальный заряд перемещаться вдоль поверхностного слоя, что обуславливает простоту сканирования изображения. На рис. 103 приведена схема линейного устройства, используемого для передачи изображения. Эти приборы обладают существенными преимуществами по сравнению с полупроводниковыми приборами на основе фотопроводимости — видиконами.

Резюме § 68

1. При помещении полупроводника во внешнее однородное электрическое поле в результате перемещения свободных носителей заряда, вызванного внешним полем, в объеме полупроводника вдали от поверхности электрическое поле отсутствует. Поле существует только в приповерхностном слое, в котором существует и объемный заряд. Он экранирует внешнее поле, вследствие чего оно и не проникает в объем полупроводника.

2. В области объемного заряда зоны энергии искривляются. Если они изгибаются кверху, то приповерхностный слой обогаща-

ется дырками, при изгибании их вниз происходит обогащение электронами. В собственном полупроводнике проводимость приповерхностных слоев увеличивается независимо от направления поля. В примесном полупроводнике проводимость приповерхностного слоя увеличивается, если внешнее поле вытягивает из объема основные носители заряда, и уменьшается, если приповерхностный слой обогащается неосновными носителями заряда. Изменение проводимости приповерхностного слоя называют эффектом поля.

3. Толщина слоя, в котором существует объемный заряд и электрическое поле и искривляются заметным образом зоны энергии, характеризуется некоторой величиной L^D :

$$L^D = \sqrt{\frac{\epsilon\epsilon_0 kT}{2e^2 n_i}}; \quad L^D = \sqrt{\frac{\epsilon\epsilon_0 kT}{e^2 n_0}} \quad (68.1p)$$

соответственно в собственном и примесном полупроводнике, L^D называется дебаевой длиной экранирования.

4. При достаточно большой напряженности поля в приповерхностном слое может измениться тип проводимости. Слой с измененным типом проводимости называют инверсным слоем; слой, в котором $n \approx p \approx n_i$, называют i -слоем. Область, в которой происходит изменение типа проводимости, называют физическим p - n -переходом.

5. Если поверхностные состояния заполняются свободными носителями заряда, то между поверхностью и объемом возникает электрическое поле, локализованное в приповерхностном слое, свойства которого меняются таким же образом, как и в результате действия внешнего электрического поля.

6. В металле концентрация свободных электронов велика, поэтому дебаева длина экранирования значительно меньше межатомного расстояния.

§ 69. РАБОТА ВЫХОДА

Хорошо известно, что для перевода электрона из твердого тела в вакуум необходимо затратить определенную энергию. Примем за начало отсчета энергию электрона, находящегося в вакууме вдали от данного тела и покоящегося относительно него. В таком случае полная энергия электронов, покоящихся в твердом теле, отрицательна. Электрон, находящийся на дне зоны проводимости, имеет нулевую кинетическую энергию, его полная энергия, с точки зрения классической физики, является потенциальной энергией. Пусть положение уровня E_c в вышеуказанной «абсолютной» шкале энергии характеризуется некоторой величиной $(-W)$, ($W > 0$). Величина W равна работе, необходимой для перевода покоящегося в твердом теле электрона в вакуум без сообщения ему кинетической энергии, W называется истинной работой выхода, она определяет глубину потенциальной ямы, которая соответствует электронам проводимости металла. В теории квазисвободного электрона величина $(-W)$ обозначалась как $\langle U \rangle$. Уровни энергии, лежащие выше E_c , практически

полностью заняты вплоть до уровня Ферми. Электроны, имеющие полную отрицательную энергию, не могут выйти из металла. Но среди электронов есть и такие, которые имеют положительную полную энергию; они способны выйти из металла. Вычислим поток электронов из металла в вакуум за счет их тепловой энергии.

Пусть металл заполняет полупространство $x < 0$. Подсчитаем число электронов, способных преодолеть прямоугольный потенциальный барьер высотой W и выйти из металла. Для этого кинетическая энергия электрона, обусловленная скоростью v_x , должна быть не меньше глубины потенциальной ямы

$$\frac{m^*v_x^2}{2} \geq W; \quad \frac{m^*v_x^2}{2} + E_c = E \geq 0. \quad (69.1)$$

Будем считать, что все вышедшие из металла электроны захватываются некоторым внешним полем и не возвращаются в металл. При этом предположении легко записать выражение для плотности тока, текущего из вакуума в металл:

$$j_x = e \int_{v_{x\min}}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_x 2 \frac{m^{*3}}{h^3} e^{-\frac{E-F}{kT}} dv_x dv_y dv_z = e \int_{v_{x\min}}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v_x e^{-\frac{E-F}{kT}} \frac{d\tau_p}{h^3}. \quad (69.2)$$

Соотношение (69.2) записано с учетом того, что в единичном объеме кристалла и объеме $d\tau_p$ содержится $2 \frac{1 \cdot d\tau_p}{h^3}$ состояний, занятых электронами с вероятностью

$$f_0(E, T) = \frac{1}{e^{\frac{E-F}{kT}} - 1} \approx e^{-\frac{E-F}{kT}}, \quad (69.3)$$

поскольку для электронов, способных вылетать из металла, $E - F \gg kT$.

Учитывая, что энергию E можно выразить через скорость:

$$E = E_c + \frac{m^*v^2}{2} = E_c + \frac{m^*}{2} (v_x^2 + v_y^2 + v_z^2), \quad (69.4)$$

получим для j_x :

$$j_x = \frac{2ee \frac{F-E_c}{kT} m^{*3}}{h^3} \int_{-\infty}^{\infty} e^{-\frac{m^*v_y^2}{2kT}} dv_y \int_{-\infty}^{\infty} e^{-\frac{m^*v_z^2}{2kT}} dv_z \times \\ \times \int_{v_{x\min}}^{\infty} v_x e^{-\frac{m^*v_x^2}{2kT}} dv_x. \quad (69.5)$$

Интегралы по v_y и v_z сводятся к интегралу Пуассона

$$\int_{-\infty}^{\infty} e^{-\xi^2} d\xi = \sqrt{\pi}. \quad (69.6)$$

с помощью замены переменных, например,

$$\frac{m^* v_y^2}{2kT} = \xi^2, \quad dv_y = \sqrt{\frac{2kT}{m^*}} d\xi, \quad (69.7)$$

поэтому

$$\int_{-\infty}^{\infty} e^{-\frac{m^* v_y^2}{2kT}} dv_y = \int_{-\infty}^{\infty} e^{-\frac{m^* v_z^2}{2kT}} dv_z = \sqrt{\frac{2\pi kT}{m^*}}. \quad (69.8)$$

Интегрирование по v_x выполняется непосредственно:

$$\begin{aligned} \int_{v_{x\min}}^{\infty} e^{-\frac{m^* v_x^2}{2kT}} v_x dv_x &= \frac{kT}{m^*} \int_{v_{x\min}}^{\infty} e^{-\frac{m^* v_x^2}{2kT}} d\left(\frac{m^* v_x^2}{2kT}\right) = \\ &= \frac{kT}{m^*} e^{-\frac{m^* v_{x\min}^2}{2kT}} = \frac{kT}{m^*} e^{-\frac{W}{kT}} = \frac{kT}{m^*} e^{-\frac{E_c}{kT}}. \end{aligned} \quad (69.9)$$

Учитывая (69.8) и (69.9), запишем выражение для j_x :

$$j_x = \frac{4\pi e m^* k^2 T^2}{h^3} e^{\frac{F - E_c + E_c}{kT}} = AT^2 e^{-\frac{F}{kT}}; \quad (69.10)$$

$$\left[A = \frac{4\pi e m^* k^2}{h^3} = 120 \left(\frac{m^*}{m} \right) \frac{\text{А}}{\text{см}^2 \text{град}^2} \right].$$

Обозначим расстояние от начала отсчета энергии до уровня Ферми F через Φ , т. е. положим $F = -\Phi$, ($\Phi > 0$), после чего для j_x будем иметь

$$j_x = AT^2 e^{-\frac{\Phi}{kT}} = j_T. \quad (69.11)$$

Величина Φ носит название термодинамической работы выхода электронов из металла. Численно она равна работе, необходимой для удаления из металла электрона, находящегося на уровне Ферми. Выражение (69.11) называется формулой Ричардсона, а j_T — плотностью тока термоэлектронной эмиссии. В реальных условиях термоэлектронный ток не может достигать величины j_T , если не созданы для этого специальные условия. Действительно, металл, потеряв электроны, заряжается положительно. Возникшее электрическое поле удерживает термоэлектроны, возвращая их в металл. Устанавливается динамическое равновесие между двумя встречными потоками электронов, что обеспечивает обращение j_x в нуль. Для того чтобы поддерживать j_T постоянным, необходимо внешним электрическим полем

рассасывать термоэлектронное облако и в то же время компенсировать потерю заряда металлом. Обе эти функции может выполнять одно и то же поле подобно тому, как это имеет место в двухэлектродной лампе. Величина термоэлектронной эмиссии резко зависит от температуры. Действительно, при $\Phi = 2,5$ эВ и $T \approx 300^\circ \text{K}$ плотность тока $j_T \sim 10^{-36}$ А/см², но при повышении температуры в пять раз ($T \approx 1500^\circ \text{K}$) $j_T \approx 0,8$ А/см², т. е. возрастает в 10^{36} раз.

Найдем величину плотности тока термоэлектронной эмиссии j_T электронов из полупроводника. Для этого рассмотрим энергетическую диаграмму полупроводника. Примем, как и в предыдущем случае, за начало отсчета энергию электрона, покоящегося в вакууме.

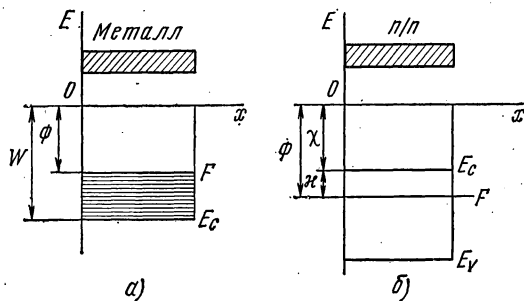


Рис. 104. Схема уровней энергии металла (а) и полупроводника (б). За начало отсчета энергии принята энергия электрона, покоящегося в вакууме

Представим энергию Ферми в виде суммы двух членов:

$$E_c = -\chi \quad (\chi > 0) \\ \kappa = E_c - F, \quad (69.12)$$

или из (69.12) и (69.11)

$$F = E_c - \kappa = -(\chi + \kappa). \quad (69.13)$$

Представление величины F в виде суммы двух членов удобно, поскольку величина χ зависит только от вида полупроводникового вещества, в то время как κ зависит от температуры, вида и количества легирующей примеси, как это было найдено в третьей главе. На рис. 104 представлена схема расположения уровней энергии в металле и полупроводнике при условии, что за начало отсчета энергии принята энергия электрона, покоящегося в вакууме.

Повторяя выкладки, проделанные выше для металла, получим для плотности тока термоэлектронной эмиссии из полупроводника:

$$j_T = AT^2 e^{\frac{F}{kT}} = AT^2 e^{-\frac{\chi + \kappa}{kT}} = AT^2 e^{-\frac{\Phi}{kT}}, \quad (69.14)$$

где Φ — термодинамическая работа выхода электронов из полупроводника. Как и в металле, она определяется положением уровня Ферми в «абсолютной» шкале энергии:

$$\Phi = -F = \chi + \kappa, \quad (69.15)$$

несмотря на то, что на уровне Ферми в невырожденном полупроводнике нет электронов. Используя результаты третьей главы, запишем выражения для работы выхода электронов в различных случаях.

Для собственного полупроводника работа выхода Φ равна

$$\Phi(T) = \chi + \frac{\Delta E_0}{2} + \frac{kT}{2} \ln \frac{N_c}{N_v}; \quad \Phi(0) = \chi + \frac{\Delta E_0}{2}. \quad (69.16)$$

Для донорного полупроводника

$$\Phi(T) = \chi + \frac{\Delta E_d}{2} + \frac{kT}{2} \ln \frac{N_c}{N_d} \quad (69.17)$$

и

$$\Phi(T) = \chi + kT \ln \frac{N_c}{N_d} \quad (69.18)$$

соответственно в области слабой и сильной ионизации примеси.

Для акцепторного полупроводника аналогично предыдущему случаю имеем

$$\Phi(T) = \chi + \Delta E_0 - \frac{\Delta E_a}{2} - \frac{kT}{2} \ln \frac{N_v}{N_a}, \quad (69.19)$$

$$\Phi(T) = \chi + \Delta E_0 - kT \ln \frac{N_v}{N_a}. \quad (69.20)$$

Как видно из (69.17–20), работа выхода электронов из дырочного полупроводника значительно больше, чем из электронного.

Резюме § 69

1. Энергию, необходимую для перевода электрона, имеющего нулевую скорость, из металла в вакуум без изменения скорости, называют истинной работой выхода W .

2. Термодинамической работой выхода электрона из твердого тела называют величину энергии, необходимую для перевода электрона с уровня Ферми на уровень $E=0$ при его переходе из кристалла в вакуум:

$$\Phi = -F. \quad (69.1p)$$

3. Термодинамическая работа выхода электрона из полупроводника зависит от температуры, вида и концентрации примеси. Работа выхода электронов из дырочного полупроводника превосходит работу выхода из электронного полупроводника почти на ширину запрещенной зоны.

4. Работа выхода Φ определяет плотность термоэлектронного тока

$$j_T = AT^2 e^{-\frac{\Phi}{kT}}. \quad (69.2p)$$

§ 70. КОНТАКТНАЯ РАЗНОСТЬ ПОТЕНЦИАЛОВ. КОНТАКТ МЕТАЛЛ—МЕТАЛЛ

На рис. 105, *a* изображена энергетическая диаграмма двух различных изолированных металлов M_1 и M_2 . Величины, характеризующие металлы, имеют индекс «1» или «2». Индекс «0» соответствует тому, что металлы не находятся во взаимодействии друг с другом.

Если привести металлы M_1 и M_2 в контакт, то они начнут обмениваться электронами. Как известно, частицы переходят из той под-

системы, в которой уровень Ферми находится выше, в ту подсистему, в которой уровень Ферми расположен ниже. Это можно показать следующим образом.

Рассмотрим две подсистемы, которые обмениваются частицами. Изменение термодинамического потенциала Φ системы равно

$$d\Phi = \sum_i F_i dN_i = F_1 dN_1 + F_2 dN_2. \quad (70.1)$$

Но так как $dN_1 = -dN_2$, то

$$d\Phi = dN_1 (F_1 - F_2). \quad (70.2)$$

Если $F_1 = F_2$, то $d\Phi = 0$, и рассматриваемое состояние является равновесным.

Если же $F_1 \neq F_2$, то $d\Phi \neq 0$, и, следовательно, система не находится в равновесном состоянии. Переход к равновесному состоянию

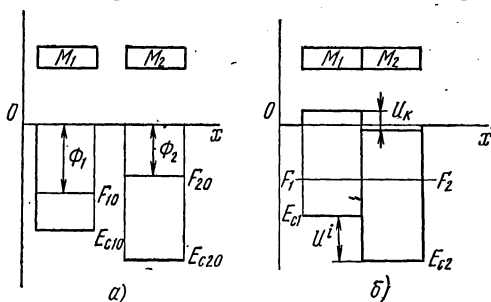


Рис. 105. Контакт двух металлов

означает, что $d\Phi < 0$. Если $F_1 < F_2$, то из $d\Phi < 0$ следует $dN_1 > 0$, т. е. число частиц первой подсистемы увеличивается, а второй уменьшается; если же $F_1 > F_2$, то $dN_1 < 0$. Таким образом, число частиц подсистемы с большим химическим потенциалом должно уменьшаться или, другими словами, переход системы к равновесному состоянию означает наличие

направленного потока частиц из подсистемы с большим химическим потенциалом к подсистеме с меньшим химическим потенциалом. Этот направленный поток частиц будет длиться до тех пор, пока химические потенциалы не выровняются благодаря каким-либо физическим процессам, связанным с направленным потоком.

Вывод о характере направленного потока частиц в нашем случае можно получить из простых соображений: электроны должны занимать все нижележащие свободные уровни; уровень Ферми отделяет уровни, в основном занятые от почти полностью свободных, поэтому если F_{20} лежит выше F_{10} , то это означает, что электроны с верхних уровней M_2 должны переходить на ниже расположенные свободные уровни в M_1 , в полном соответствии с выводом, вытекающим из условия $d\Phi \leq 0$.

Так как положение уровня Ферми в «абсолютной» (т. е. в единой для всех тел) системе отсчета энергии определяется только термодинамической работой выхода Φ , то можно сформулировать следующий вывод: при контакте двух тел возникнет направленный поток частиц из тела с меньшей работой выхода в тело с большей работой выхода. Направленный поток частиц прекратится, когда произойдет выравнивание населенности уровней энергии в обеих подсистемах.

Однако для этого количество перешедших электронов может быть очень незначительным. Действительно, уход электронов приводит к тому, что тело заряжается положительно, и электроны этого тела получают дополнительную отрицательную энергию, поэтому все уровни энергии опускаются на некоторую величину. Уровни энергии второго металла, зарядившегося отрицательно, поднимаются на некоторую, вообще говоря другую, величину. Направленный поток электронов из тела с меньшей работой выхода в тело с большей работой выхода прекратится в тот момент, когда уровни Ферми первого и второго тела выровняются: $F_1 = F_2$. При этом необходимо иметь в виду, что $F_1 = F_2$ отличны как от F_{10} , так и от F_{20} . На рис. 105, б показано расположение уровней энергии двух контактирующих металлов в равновесном состоянии.

Так как тела зарядились, то между ними возникает разность потенциалов и электрическое поле. Как было показано в § 68, в металле электрическое поле локализуется в тонком слое, значительно меньшем периода решетки.

Заряд, сообщенный металлу, распределяется в общем случае по его поверхности. Следовательно, и в данном случае это электрическое поле будет локализовано в тонком поверхностном слое металла. Между любыми двумя точками наружной поверхности металлов M_1 и M_2 должна существовать разность потенциалов Φ^k , называемая *внешней контактной разностью потенциалов*. Ее величину можно определить из следующих соображений: подъем и опускание уровней энергии в M_1 и M_2 приводит к тому, что они смещаются на величину, равную разности работ выхода, следовательно,

$$e\Phi_{21}^k = U_{21}^k = \Phi_1 - \Phi_2, \quad (70.3)$$

U^k представляет собой потенциальный барьер, обеспечивающий равенство термодинамических работ выхода относительно перехода электронов из M_1 и M_2 , находящихся в контакте, что вытекает из закона сохранения энергии. Действительно, возьмем электрон в металле с некоторой полной энергией E и совершим обход по замкнутому контуру в двух противоположных направлениях (рис. 106). Полная энергия электрона в металлах M_1 и M_2 остается постоянной и равной E . Совершим обход против часовой стрелки. Вынесем электрон из M_2 в вакуум, энергия его станет равной $E + \Phi_2$; внесем его в M_1 , это уменьшит энергию электрона на величину Φ_1 : $E^{(1)} = E + \Phi_2 - \Phi_1$. Совершим теперь обход по часовой стрелке. По возвращении в M_1 заряд будет иметь энергию $E^{(2)} = E + \Phi_1 - \Phi_2$. Изменение энергии в первом случае равно $\Delta E^{(1)} = E^{(1)} - E = \Phi_2 - \Phi_1$, во втором случае: $\Delta E^{(2)} = E^{(2)} - E = \Phi_1 - \Phi_2$. Поскольку $\Phi_2 \neq \Phi_1$, то при одном направлении обхода энергия увеличивается, при другом — уменьшается, т. е. имеет место противоречие с законом сохра-

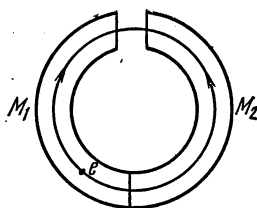


Рис. 106. Направление обхода по замкнутому контуру

нения энергии. Это произошло потому, что мы не учли энергию U^k , обусловленную внешней контактной разностью потенциалов Φ^k , которую необходимо добавить к общей энергии или отнять от нее в зависимости от направления обхода. Учитывая U^k , получим

$$\left. \begin{aligned} \Delta E^{(1)} &= \Phi_2 + U^k - \Phi_1 = 0, \\ \Delta E^2 &= \Phi_1 - (\Phi_2 + U^k) = 0. \end{aligned} \right\} \quad (70.4)$$

Из (70.4) следует, что потенциальный барьер U^k , разделяющий две произвольные внешние точки двух металлов M_1 и M_2 , имеет величину, равную разности термодинамических работ выхода. Физически это означает, что потенциал металла с меньшей работой выхода повышается относительно металла с большей работой выхода на величину Φ^k , благодаря чему работа выхода в любой точке пары металлов определяется наибольшей из (Φ_1, Φ_2) , а внешнее контактное поле U^k локализовано в непосредственной близости к металлу с меньшей работой выхода.

В табл. 21 приведены в качестве примера значения работы выхода электронов Φ из некоторых веществ.

Внешняя контактная разность потенциалов, определяющаяся разностью работ выхода электронов из металлов, обеспечивает равенство плотностей токов термоэлектронной эмиссии.

Таблица 21

Вещества	Li	Be	C	Na	K	Sn	W	Pt	Cs
F , эВ	4,71	9,07	11,05	3,07	2,04	4,07	5,81	6,00	1,53
Φ , эВ	2,39	3,92	4,62	2,35	2,22	4,3	4,52	5,36	1,93

Возникновение внешней контактной разности потенциалов между металлами возможно не только при их соприкосновении, но и в том случае, когда непосредственного контакта между ними не возникает, но металлы могут обмениваться электронами вследствие термоэлектронной эмиссии.

Теперь рассмотрим, какие условия существуют на самом контакте. Если уровни Ферми металлов M_1 и M_2 при равновесии совпадают, то это означает, что на контакте существует потенциальный барьер и обусловленное им электрическое поле. Это поле E^k называют *контактным*, а соответствующую разность потенциалов Φ^i — *внутренней контактной разностью потенциалов*. Ее можно найти из условия

$$\Phi_{21}^i = \frac{U_{12}^i}{e}, \quad (70.5)$$

где U_{12}^i — величина потенциального барьера между M_1 и M_2 . Непосредственно из рис. 105, б видно, что для перевода электрона с ну-

левой кинетической энергией из первого металла во второй без изменения скорости необходимо затратить энергию, равную разности в положении E_{c1} и E_{c2} ; следовательно, величина потенциального барьера

$$U_{12}^i = E_{c1} - E_{c2}. \quad (70.6)$$

С другой стороны, U_{12}^i можно выразить через изменение кинетической энергии частицы при переходе через контакт:

$$U_{12}^i = T_2 - T_1. \quad (70.7)$$

Условия (70.7) и (70.6) совпадают, поскольку $T_1 = E - E_{c1}$; $T_2 = E - E_{c2}$. Но так как все уровни сместились на величину U^k , то

$$E_{c1} - E_{c2} = E_{c10} - E_{c20} + U_{12}^k = E_{c10} - E_{c20} + \Phi_1 - \Phi_2. \quad (70.8)$$

Учитывая, что $\Phi_1 = -F_{10}$; $\Phi_2 = -F_{20}$, получим

$$U_{12}^i = E_{c1} - E_{c2} = (E_{c10} - F_{10}) - (E_{c20} - F_{20}) = (F_{20} - E_{c20}) - (F_{10} - E_{c10}). \quad (70.9)$$

Таким образом, величина потенциального барьера U_{12}^i , определяющего внутреннюю контактную разность потенциалов $\Phi^i = U^i/e$, равна разности энергий Ферми изолированных металлов M_1 и M_2 , отсчитываемых от дна зоны проводимости каждого металла. Величина внутреннего потенциального барьера определяется концентрацией электронов в металлах; так как

$$F_0 - E_{c0} = \frac{h^2}{2m^*} \left(\frac{3n}{8\pi} \right)^{2/3}, \quad (70.10)$$

то

$$U_{12}^i = \frac{h^2}{2} \left(\frac{3}{8\pi} \right)^{2/3} \left(\frac{n_2^{2/3}}{m_1^*} - \frac{n_1^{2/3}}{m_2^*} \right). \quad (70.11)$$

Наличие контактного поля U^i обеспечивает равенство потоков электронов из одного металла в другой. Действительно, запишем выражения для плотности потоков электронов из первого металла во второй (j_{21}/e) и из второго в первый (j_{12}/e):

$$\frac{j_{21}}{e} = \frac{2m_1^{*3}}{h^3} \int_{v_{x1}}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{v_x dv_x dv_y dv_z}{e^{\frac{E_1 - F_1}{kT}} + 1} \quad (70.12)$$

и

$$\frac{j_{12}}{e} = \frac{2m_2^{*3}}{h^3} \int_{v_{x2}}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{v_x dv_x dv_y dv_z}{e^{\frac{E_2 - F_2}{kT}} + 1}. \quad (70.13)$$

Из равенств потоков (70.12) и (70.13) можно получить (70.9) и, наоборот, из (70.9) или (70.7) следует равенство потоков $j_{21}/e = j_{12}/e$.

В табл. 21 приведены значения энергий Ферми некоторых металлов, по которым можно вычислить U^i .

§ 71. КОНТАКТ МЕТАЛЛ — ПОЛУПРОВОДНИК

Перейдем теперь к рассмотрению контакта металла и полупроводника. На рис. 107, а изображена энергетическая диаграмма металла и полупроводника в первый момент после их соприкосновения. Это состояние является неравновесным, поэтому в процессе обмена электронами преимущественный переход электронов будет происходить из вещества с большей энергией Ферми в вещество с меньшей энергией Ферми. Если уровень Ферми металла F_M лежит ниже уровня Ферми F_{nn} полупроводника ($F_M < F_{nn}$), то электроны будут переходить из полупроводника в металл. Металл заряжается отрицательно, а полупроводник — положительно. Направленный поток электронов будет иметь место до тех пор, пока уровни Ферми не выровняются, после чего установится динамическое равновесие. При $F_M > F_{nn}$ электроны перейдут из металла в полупроводник. Обмен

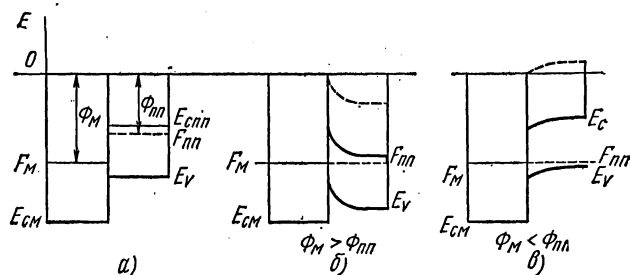


Рис. 107. Контакт металл — полупроводник

электронами возможен не только при непосредственном контакте металла и полупроводника, но и благодаря термоэлектронной эмиссии, которая обуславливает внешнюю контактную разность потенциалов $e\Phi_{Mnn}^k = U_{nnM}^k = \Phi_{nn} - \Phi_M$. При контакте металлов Φ^k распределяется между металлами примерно одинаково, поскольку концентрации электронов в них близки, при этом потенциал внутри каждого металла постоянен. Иначе обстоит дело в полупроводнике. Потенциал полупроводника может быть различным в разных точках, поскольку внешнее электрическое поле может проникать на большую глубину в полупроводник, создавая в нем объемный заряд и искривляя зоны энергии. Точно такое же положение будет иметь место и в данном случае. *Контактная разность потенциалов практически полностью будет падать в приконтактном слое полупроводника, искривляя в нем зоны энергии* (рис. 107, б, в). Зоны энергии будут искривлены в приконтактной области кверху, если $\Phi_{nn} < \Phi_M$, и искривлены книзу, если $\Phi_{nn} > \Phi_M$. Но искривление зон энергии в приконтактном слое меняет концентрацию электронов и дырок: изгиб зон энергии кверху увеличивает концентрацию дырок, а их изгиб книзу увеличивает концентрацию электронов. Таким образом, *если работа выхода электронов из полупроводника меньше, чем из металла, то приконтактный слой обогащается дырками; если же*

$\Phi_{nn} > \Phi_M$, то происходит обогащение приконтактного слоя электронами.

Если искривление зон происходит в собственном полупроводнике, то проводимость приконтактного слоя увеличивается. Иначе обстоит дело в случае примесной проводимости. Проводимость повышается, когда происходит обогащение приконтактного слоя основными носителями заряда и понижается при обогащении слоя неосновными носителями заряда. Слой с повышенной проводимостью (обогащенный основными носителями заряда) называют антизапирающим. Слой с пониженной проводимостью (обогащенный неосновными носителями заряда) называют запирающим. Слой называют инверсным, если в нем в результате сильного искривления зон меняется тип проводимости.

В табл. 22 указаны условия возникновения запирающих, антизапирающих и инверсных слоев.

Таблица 22

Тип полупроводника	Антизапирающий слой	Запирающий слой	Инверсия знака проводимости
Электронный полупроводник	$F_M > F_{nn};$ $\Phi_M < \Phi_{nn}$	$F_M < F_{np};$ $\Phi_M > \Phi_{np}$	$F_M < F_{np};$ $\Phi_M \geq \Phi_{np} + \Delta E_0$
Дырочный полупроводник	$F_M < F_{np};$ $\Phi_M > \Phi_{np}$	$F_M > F_{nn};$ $\Phi_M < \Phi_{nn}$	$F_M > F_{nn};$ $\Phi_M \leq \Phi_{nn} - \Delta E_0$

Обогащение приконтактного слоя полупроводника неосновными носителями заряда имеет важное значение при любых физических измерениях в полупроводниках, связанных с прохождением тока. Действительно, при приложении внешнего напряжения к цепи, состоящей из металла и полупроводника, в ней возникает ток, переносимый электронами и дырками:

$$\mathbf{j} = \mathbf{j}_n + \mathbf{j}_p. \quad (71.1)$$

В связи с тем что концентрация электронов и дырок различна в разных точках полупроводника, состав тока меняется, а величина его остается неизменной. Предположим для определенности, что полупроводник электронный. Обозначим через γ долю тока, переносимого неосновными носителями заряда, т. е. в данном случае дырками:

$$\gamma(x) = \frac{j_p}{j} = \frac{j_p(x)}{j_n(x) + j_p(x)} = \frac{p(x)}{p(x) + bn(x)}; \quad \left(b = \frac{|\mu_n|}{\mu_p} \right). \quad (71.2)$$

Обозначим через γ_0 величину, определяемую равновесной концентрацией неосновных носителей заряда:

$$\gamma_0 = \frac{p_0}{p_0 + bn_0}. \quad (71.3)$$

Через $\gamma(x)$ обозначено отношение тока, переносимого неосновными носителями заряда к полному току с учетом нарушения равновесной концентрации носителей заряда. При $\frac{\gamma(x)}{\gamma_0(x)} < 1$ концентрация неосновных носителей заряда в точке x уменьшается, при $\frac{\gamma(x)}{\gamma_0(x)} > 1$ — повышается при прохождении тока. Другими словами, ток, проходя через контактный слой, меняет концентрацию основных и неосновных носителей заряда в объеме полупроводника, что приводит к изменению свойств полупроводника. Если $\frac{\gamma(x)}{\gamma_0(x)} > 1$, то объем полупроводника обогащается неосновными (и неравновесными) носителями заряда. Это будет возможным как для запирающего, так и для антизапирающего слоев. В первом случае мы имеем инжекцию, во втором — аккумуляцию. Если $\frac{\gamma(x)}{\gamma_0(x)} < 1$, то объем обедняется неосновными носителями заряда. При этом если обеднение вызвано наличием запирающего слоя, то явление называется эксклюзией, в случае антизапирающего — экстракцией. Эти явления соответствуют тем явлениям, которые разбирались в § 66. В этом параграфе мы нашли причину, порождающую неравновесную концентрацию неосновных носителей заряда — искривление зон энергии контактным полем, а внешнее электрическое поле, вызывая дрейф носителей заряда, приводит к явлениям инжекции, экстракции, аккумуляции и эксклюзии.

Если контакт не меняет концентрацию неосновных носителей заряда: $\gamma = \gamma_0$, то его называют омическим. Создание омических контактов при исследовании физических явлений в полупроводниках представляет очень важную, однако крайне трудную задачу. Поскольку создание омического контакта не всегда возможно, то для того, чтобы исключить роль контактных явлений на результаты измерений, в физике полупроводников широко применяют компенсационные методы исследования, благодаря чему ток через контакт не проходит и, следовательно, контактные явления не сказываются на результатах измерений.

В табл. 23 указаны условия наблюдения контактных явлений в зависимости от типа контакта, типа проводимости полупроводника и направления внешнего поля в предположении, что контакт ($x = 0$) находится на левой границе полупроводника.

Однако исследования точечных контактов на кремнии и германии, проведенные еще в пятидесятых годах, показали, что в ряде случаев выбор материала металла не влияет на свойства контакта. Другими словами, нарисованная картина искривления зон, приводящая к обеднению или обогащению слоев полупроводника неосновными носителями заряда в приконтактной области, как будто в действительности не имеет места. Объяснение этих фактов было найдено при учете действия поверхностных состояний на свойства полупроводника в объеме, о чем было сказано в § 68. Мы видели, что наличие поверхностных состояний приводит к искривлению зон энергии

в приповерхностном слое. И если это искривление зон велико (V_s велико), то дополнительное искривление зон, обусловленное контактом полупроводника с металлом, может мало сказаться на свойствах этих контактов. Другими словами, *свойства запирающих и антизапирающих слоев будут мало зависеть от природы металла в том случае, когда в действительности они создаются поверхностными состояниями*. Наличие поверхностных состояний проявляется также в уменьшении времени жизни носителей заряда, за счет поверхностной рекомбинации, о чем было сказано в § 67. Различают «медленные» и «быстрые» поверхностные состояния в зависимости от времени обмена носителями заряда между ними и объемом. Для «медленных» состояний времена установления равновесия имеют значение от десятых долей секунды до многих часов. Для «быстрых» состояний эти времена имеют значение примерно от 10^{-4} до 10^{-6} с. «Медленные» состояния связаны с атомами или ионами, адсорбированными на слой окиси или какой-либо другой промежуточный слой полупроводника. «Быстрые» состояния локализованы на границе между объемом полупроводника и окисной пленкой.

Таблица 23

Контакт	Выпрямляющий (запирающий слой)	Невыпрямляющий (антизапирающий слой)	Омический
	$\frac{\gamma(0)}{\gamma_0} \neq 1$	$\frac{\gamma(0)}{\gamma_0} \neq 1$	$\frac{\gamma(0)}{\gamma_0} = 1$
Инжекция	$\gamma(x) > \gamma_0(x)$ $E > 0$, (n-тип) $E < 0$, (p-тип)	—	—
Эксклюзия	$\gamma(x) < \gamma_0(x)$ $E < 0$, (n-тип) $E > 0$, (p-тип)	—	—
Экстракция	—	$\gamma(x) < \gamma_0(x)$ $E > 0$, (n-тип) $E < 0$, (p-тип)	—
Аккумуляция	—	$\gamma(x) > \gamma_0(x)$ $E < 0$, (n-тип) $E > 0$, (p-тип)	—

Резюме § 70—71

1. При контакте металлов между любыми двумя внешними точками разных металлов существует разность потенциалов φ^k , называемая внешней контактной разностью потенциалов. Наличие скачка потенциальной энергии $U^k = e\varphi^k$ обеспечивает равенство потоков

электронов термоэмиссии из одного металла в другой. Из закона сохранения энергии или из равенства потоков электронов термоэмиссии следует

$$U_{12}^k = e\varphi_{21}^k = \Phi_1 - \Phi_2. \quad (70.1p)$$

Из равенства потоков электронов через контакт следует, что на контакте существует разность потенциалов φ^i , называемая внутренней контактной разностью потенциалов. Она определяет скачок потенциальной энергии U^i .

$$U_{12}^i = e\varphi_{21}^i = T_2 - T_1 = (F_{20} - E_{c20}) - (F_{10} - E_{c10}). \quad (70.2p)$$

2. В связи с тем, что в объеме металла электрическое поле в отсутствие тока существовать не может, вся контактная разность потенциалов φ^k падает в приповерхностном слое полупроводника, приводя к искривлению зон энергии и возникновению объемного заряда, подобно тому, как это имеет место при наложении на полупроводник внешнего электрического поля. Искривление зон энергии, возникающее в результате обмена электронами термоэлектронной эмиссии, не изменится и после приведения металла и полупроводника в контакт. Поскольку для перехода электронов из полупроводника в металл существует потенциальный барьер высотой

$$U_{Mnn}^k = e\varphi_{nnM}^k = \Phi_M - \Phi_{nn}, \quad (70.3p)$$

то можно сказать, что при контакте металла и полупроводника возникает внутренняя контактная разность потенциалов, равная внешней контактной разности потенциалов, т. е. равная разности работ выхода электронов из металла и полупроводника.

3. Если приконтактный слой полупроводника обогащается основными носителями заряда, то его называют антизапирающим. Если происходит обогащение неосновными носителями заряда приконтактного слоя, то слой называют запирающим. Если обогащение неосновными носителями заряда происходит настолько сильно, что меняется тип проводимости, то слой называют инверсным, а под ним лежит физический *p-n*-переход. При сильном искривлении зон энергии с обогащением приконтактного слоя основными носителями заряда возможно возникновение вырождения полупроводника.

4. При прохождении тока через контакт металла и полупроводника неосновные носители заряда будут затягиваться полем в области, в которых они могут стать неравновесными. Возникают явления инжекции, экстракции, эксклюзии и аккумуляции, приводящие к изменению свойств полупроводника в объеме. По этой причине при электрических измерениях в физике полупроводников используют компенсационные методы.

5. Контакт называют выпрямляющим, если он инжектирует или эксклюзирует неосновные носители заряда.

6. Контакт называют невыпрямляющим, если он экстрагирует или аккумулирует неосновные носители заряда.

7. Контакт называется омическим, если состав тока одинаков как на самом контакте, так и в объеме.

8. При контакте двух тел электроны переходят из тела с большей энергией Ферми (с меньшей работой выхода) в тело с меньшей энергией Ферми (с большей работой выхода). Направленный поток прекращается, когда уровни Ферми выравниваются.

§ 72. НЕОДНОРОДНЫЙ ПОЛУПРОВОДНИК. *p-n*-ПЕРЕХОД

Многие полупроводниковые приборы основаны на использовании свойств *p-n*-переходов, которые можно рассматривать как контакт образцов одного и того же полупроводника с электронной и дырочной проводимостью. Однако возможен другой подход, основанный на рассмотрении неоднородного полупроводника. Предположим, что мы имеем полупроводник с некоторым произвольным распределением акцепторной и донорной примеси $N_a(\mathbf{r})$ и $N_d(\mathbf{r})$. Это приводит к тому, что концентрация электронов и дырок зависит от координаты. Положение уровня Ферми относительно зон энергии E_c и E_v определяется концентрацией электронов и дырок, поэтому $F = F(\mathbf{r})$. Но так как $F = F(\mathbf{r})$, то состояние является неравновесным, возникает поток носителей заряда, стремящийся выровнять концентрацию электронов и дырок. *Возникновение диффузионного тока j_D приводит к разделению зарядов, вследствие чего возникает объемный заряд и порождаемое им электрическое поле, оно искривляет зоны энергии.* В состоянии термодинамического равновесия уровень Ферми не зависит от координаты: $F = F_0$. Диффузионный ток j_D компенсируется дрейфовым током j_E , поэтому

$$\mathbf{j} = \mathbf{j}_D + \mathbf{j}_E = 0. \quad (72.1)$$

Из (72.1) можно оценить напряженность внутреннего электрического поля E^i :

$$\mathbf{E}^i = -\frac{j_D}{\sigma} = -\frac{j_{Dn} + j_{Dp}}{\sigma_n + \sigma_p}, \quad (72.2)$$

$$\mathbf{E}^i(\mathbf{r}) = \frac{D_p \nabla p - D_n \nabla n}{\mu_p p - \mu_n n} = \frac{D_p (\nabla p - b \nabla n)}{\mu_p (p + bn)}. \quad (72.3)$$

Мы видим, что напряженность поля E^i определяется градиентами концентрации электронов и дырок и проводимостью. Для того чтобы поле E^i было максимальным, необходимо, чтобы проводимость была минимальной, а градиенты концентрации электронов и дырок были противоположны по направлению. Можно выразить поле $E^i(\mathbf{r})$ через градиент носителей заряда одного знака. Ограничимся случаем невырожденного полупроводника, для него $np = n_i^2$ и

$$p \nabla n + n \nabla p = 0. \quad (72.4)$$

Подставляя ∇n из (72.4) в (72.3), получим

$$\mathbf{E}^i(\mathbf{r}) = \frac{D_p \rho + b n \nabla \rho}{\mu_p \rho + b n \rho} = \frac{kT \nabla \rho}{e_p \rho} = \frac{kT \nabla n}{e_n n}. \quad (72.5)$$

Соотношение (72.5) можно записать в следующем виде:

$$\frac{\nabla n}{n} = \nabla \ln n = -\frac{1}{kT} \nabla (E_c - F) \quad (72.6)$$

и

$$\mathbf{E}^i(\mathbf{r}) = -\frac{1}{e_n} \nabla (E_c - F) = -\frac{\nabla E_c}{e_n} = \frac{\nabla E_v}{e_p}. \quad (72.7)$$

Рассмотрим полупроводник, в котором содержится примесь одного вида, например донорная $N_d = N_d(\mathbf{r})$.

Если примесь ионизована полностью $N_d^+ = N_d$ и температура соответствует области примесной проводимости, то

$$\mathbf{E}^i = \frac{kT \nabla N_d(\mathbf{r})}{e_n N_d(\mathbf{r})} = \frac{kT}{e_n} \nabla \ln N_d(\mathbf{r}). \quad (72.8)$$

Пусть концентрация примеси $N_d(\mathbf{r})$ меняется по экспоненциальному закону

$$N_d(\mathbf{r}) = N_d(\mathbf{r}_0) e^{-\kappa(\mathbf{r} - \mathbf{r}_0)}, \quad (72.9)$$

напряженность поля \mathbf{E}^i определяется величиной κ :

$$\mathbf{E}^i(\mathbf{r}) = -\frac{kT}{e_n} \kappa. \quad (72.10)$$

Величина κ^{-1} численно равна расстоянию, на котором концентрация примеси меняется в e раз. Если $\kappa = 1 \text{ см}^{-1}$, то $\mathbf{E}^i = 0,026 \text{ В/см}$; при $\kappa = 10^7 \text{ см}^{-1}$ $\mathbf{E}^i = 0,026 \cdot 10^7 = 2,6 \cdot 10^5 \text{ В/см}$ ($T \approx 300^\circ \text{ K}$).

Эти результаты будут справедливы и в том случае, когда в полупроводнике содержится два типа примеси при условии, что одна примесь распределена равномерно, а другая — неравномерно. Например, если в дырочный полупроводник с концентрацией дырок $p_0 = N_a^+ = N_a$ ввести донорную примесь, концентрация которой меняется произвольным образом, то в точках, где проводимость остается дырочной, напряженность поля \mathbf{E}^i можно найти на основе соотношения (72.8). Это же соотношение будет справедливо и в области, где $N_d > N_a$. В области компенсации необходимо пользоваться уравнением Пуассона типа (68.12).

Рассмотрим один частный случай, когда примесь одного вида распределена равномерно при $x < 0$ и $x > 0$, но при этом в точке $x = 0$ меняется тип примеси:

$$N_d(x) = \begin{cases} N_d & \text{при } x < 0; \\ 0 & \text{при } x > 0; \end{cases} \quad (72.11)$$

$$N_a(x) = \begin{cases} 0 & \text{при } x < 0; \\ N_a & \text{при } x > 0. \end{cases}$$

Слева от границы раздела $\delta p_a = 0$; $n_0 = N_d$, справа $\delta n_d = 0$; $p_0 = N_a$. Кроме того, вдали от начала координат $\delta n = \delta p = 0$, т. е. $V(x) = 0$. Другими словами, вдали от плоскости $x = 0$ *объемный заряд и поле отсутствуют*. Рассмотрим малые расстояния от плоскости $x = 0$. Запишем уравнение (68.12) для

$$x < 0$$

$$\frac{d^2 V}{dx^2} = -\frac{e^2}{\epsilon \epsilon_0} \left\{ N_d \left(e^{-\frac{V(x)}{kT}} - 1 \right) - \frac{n_i^2}{N_d} \left(e^{\frac{V(x)}{kT}} - 1 \right) + \delta n_d \right\}, \quad (72.12)$$

$$x > 0$$

$$\frac{d^2 V}{dx^2} = \frac{e^2}{\epsilon \epsilon_0} \left\{ N_a \left(e^{\frac{V(x)}{kT}} - 1 \right) - \frac{n_i^2}{N_a} \left(e^{-\frac{V(x)}{kT}} - 1 \right) + \delta p_a \right\}. \quad (72.13)$$

Чтобы решить уравнения (72.12) и (72.13), необходимо выразить δp_a и δn_d через $V(x)$. Однако эти уравнения можно существенно упростить. Прежде всего определим знак $V(x)$. Для этого учтем, что электроны из n -области уйдут в p -область, а дырки из p -области перейдут в n -область, в результате чего n -область заряжается положительно, а p -область отрицательно.

В области перехода типа примеси возникнет электрическое поле, направленное от n -области к p -области. Так как градиент потенциальной энергии электрона совпадает по направлению с электрическим полем, то мы можем сказать, что в окрестности $x = 0$ зоны энергии изгибаются вверх по отношению к их положению в объеме, т. е. $V(x) > 0$ при $x < 0$. В p -области соответственно зоны изгибаются вниз по отношению к их положению вдали от $x = 0$, т. е. $V(x) < 0$ при $x > 0$. Так как уровни энергии поднимаются кверху при $x < 0$, то число электронов на донорном уровне может измениться. Однако рассмотрим случай, когда этим изменением можно пренебречь, т. е. положим $\delta n_d = 0 = \delta p_a$.

Для области, где $|V(x)| > kT$, можно записать

$$\frac{d^2 V(x)}{dx^2} = \frac{e^2}{\epsilon \epsilon_0} \left(N_d + \frac{n_i^2}{N_d} e^{\frac{V(x)}{kT}} \right) \quad (x < 0); \quad (72.14)$$

$$\frac{d^2 V(x)}{dx^2} = -\frac{e^2}{\epsilon \epsilon_0} \left(N_a + \frac{n_i^2}{N_a} e^{-\frac{V(x)}{kT}} \right) \quad (x > 0). \quad (72.15)$$

Если полупроводник достаточно сильно легирован, так что

$$\frac{n_i^2}{N_d^2} e^{\frac{V(x)}{kT}} \ll 1; \quad \frac{n_i^2}{N_a^2} e^{-\frac{V(x)}{kT}} \ll 1, \quad (72.16)$$

то

$$\rho^+(x) = e^+ N_d^+ = e^+ N_d \quad \text{при } x < 0 \quad \text{и}$$

$$\rho^-(x) = e^- N_a^- = -e^+ N_a \quad \text{при } x > 0,$$

т. е. в области $x = 0$ возникает *объемный заряд постоянной плотности*, равной плотности заряда ионов примеси. Предположим, что

область объемного заряда определяется точками $(-t_n, 0)$ и $(0, t_p)$ соответственно. Напряженность поля найдем из (72.14–15) с учетом (72.16):

$$\frac{dV(x)}{dx} = \frac{e^2}{\epsilon\epsilon_0} N_d x + C_1 \quad (x < 0), \quad (72.17)$$

$$\frac{dV(x)}{dx} = -\frac{e^2}{\epsilon\epsilon_0} N_a x + C_2 \quad (x > 0). \quad (72.18)$$

C_1 и C_2 определим из условия обращения в нуль поля в точках $x_1 = -t_n$ и $x_2 = t_p$, получим

$$C_1 = \frac{e^2}{\epsilon\epsilon_0} N_d t_n; \quad \frac{dV(x)}{dx} = \frac{e^2 N_d}{\epsilon\epsilon_0} (t_n + x) \quad (-t_n < x < 0); \quad (72.19)$$

$$C_2 = \frac{e^2}{\epsilon\epsilon_0} N_a t_p; \quad \frac{dV(x)}{dx} = \frac{e^2 N_a}{\epsilon\epsilon_0} (t_p - x) \quad (0 < x < t_p). \quad (72.20)$$

Из условия непрерывности поля в точке $x=0$ имеем

$$N_d t_n = N_a t_p, \quad (72.21)$$

т. е. толщина каждой области объемного заряда обратно пропорциональна концентрации легирующей примеси. Если обозначить полную толщину области объемного заряда через $t = t_n + t_p$, то можем записать

$$t_p = \frac{N_d}{N_a + N_d} t; \quad t_n = \frac{N_a}{N_a + N_d} t. \quad (72.22)$$

Таким образом, учитывая (72.19) и (72.20), можем сказать, что напряженность поля в области объемного заряда линейно зависит от координаты. Интегрируя (72.19, 20) по x , получим зависимость потенциальной энергии от координаты в виде квадратичной параболы:

$$V(x) = \frac{e^2 N_d}{2\epsilon\epsilon_0} (t_n + x)^2 + C'_1 \quad (72.23)$$

и

$$V(x) = -\frac{e^2 N_a}{2\epsilon\epsilon_0} (t_p - x)^2 + C'_2. \quad (72.24)$$

Найдем значения C'_1 и C'_2 из следующих соображений. Положим $V(x) = 0$ при $x < -t_n$ и $V(x) = U^\kappa$ при $x > t_p$, т. е. будем отсчитывать смещение зон энергии от их положения в объеме n -области; величина U^κ представляет собой величину потенциального барьера для электронов, возникающего на границе n - и p -областей:

$$V(x) = \frac{e^2 N_d}{2\epsilon\epsilon_0} (t_n + x)^2 = V_n(x) \quad (x < 0); \quad (72.25)$$

$$V(x) = -\frac{e^2 N_a}{2\epsilon\epsilon_0} (t_p - x)^2 + U^\kappa = V_p(x) \quad (x > 0). \quad (72.26)$$

Полагая $x=0$ и учитывая непрерывность $V(x)$, получим

$$V_n(0^+) = V_p(0); \quad \frac{e^2 N_d}{2\epsilon\epsilon_0} t_n^2 = -\frac{e^2 N_a}{2\epsilon\epsilon_0} t_p^2 + U^k, \quad (72.27)$$

или, с учетом (72.22),

$$U^k = \frac{e^2}{2\epsilon\epsilon_0} (N_d t_n^2 + N_a t_p^2) = \frac{e^2 N_d t_n}{2\epsilon\epsilon_0} (t_n + t_p) = \frac{e^2 N_d^2 t_n^2}{2\epsilon\epsilon_0} \left(\frac{1}{N_d} + \frac{1}{N_a} \right). \quad (72.28)$$

Высоту потенциального барьера U^k можно выразить в виде разности расстояний между F и E_c в n - и p -областях:

$$U^k = E_c(x > t_p) - E_c(x < -t_n) = \kappa_p - \kappa_n = (\chi + \kappa_p) - (\chi + \kappa_n) = \Phi_p - \Phi_n. \quad (72.29)$$

Таким образом, высота потенциального барьера для электронов, разделяющего n - и p -области, равна разности работ выхода электронов из n - и p -областей, что совпадает с внешней контактной разностью потенциалов, если рассматривать возникновение потенциального барьера на контакте n - и p -полупроводника. Область объемного заряда, разделяющая n - и p -области полупроводника, называется p - n - или n - p -переходом. Между шириной и высотой потенциального барьера существует связь. Из (72.28) можно записать

$$N_d t_n = \sqrt{\frac{2\epsilon\epsilon_0 N_a N_d U^k}{e^2 (N_a + N_d)}} = N_a t_p, \quad (72.30)$$

или

$$t = \sqrt{\frac{2\epsilon\epsilon_0 U^k N_a + N_d}{e^2 N_a N_d}}. \quad (72.31)$$

Потенциальный барьер U^k одинаковым образом действует и на электроны, и на дырки — поле p - n -перехода способствует переходу неосновных и препятствует переходу основных носителей заряда из каждой области. Мы получили параболическую зависимость (72.23—24), отбросив целый ряд членов в (72.12, 13). Если их учесть, то можно получить другую зависимость $V(x)$; однако основной результат, связанный с возникновением объемного заряда и объемного поля в области изменения типа примеси, остается в силе.

По аналогии с p - n -переходом используется название n - n^+ -или p - p^+ -переход для области, в которой возникает объемный заряд и электрическое поле, обусловленные достаточно резким изменением концентрации примеси одного и того же вида.

Объемные заряды и поля возникают в веществах с переменной шириной зоны. Например, тройное соединение $Cd_x Hg_{1-x} Te$ может быть получено при любом x : $0 < x < 1$, при этом в зависимости от x ширина запрещенной зоны соединения может быть разной, изменясь от 1,5 эВ при $x=1$ до 0 при $x=0$. Если монокристалл имеет переменный состав по длине образца, то в нем существует градиент концентрации носителей заряда, который приводит к возникновению электрического поля.

Наличие объемных полей нарушает закон Ома для участка цепи, другими словами, *сопротивление участка цепи зависит от величины тока, протекающего по данному полупроводнику; это приводит к нелинейной зависимости между напряжением и током.* Прежде всего запишем выражение для плотности тока:

$$\mathbf{j} = \sigma \mathbf{E} = \frac{\mathbf{E}}{\rho}. \quad (72.32)$$

Найдем разность потенциалов между двумя произвольными точками $M_1(\mathbf{r}_1)$ и $M_2(\mathbf{r}_2)$:

$$d\varphi = -(\mathbf{E} d\mathbf{r}); \quad \varphi(M_2) - \varphi(M_1) = - \int_{M_1}^{M_2} (\mathbf{E} d\mathbf{r}) = -j \int_{M_1}^{M_2} \rho(\mathbf{r}) \left(\frac{j}{j} d\mathbf{r} \right), \quad (72.33)$$

или

$$j = \frac{\varphi(M_1) - \varphi(M_2)}{\int_{M_1}^{M_2} \rho(\mathbf{r}) dl} = \frac{\varphi_{M_2 M_1}}{R_{M_2 M_1}}. \quad (72.34)$$

Если сечение образца остается неизменным вдоль линий тока, то плотность тока является величиной постоянной. *Внешнее напряжение $\varphi_{M_2 M_1} = jR_{M_2 M_1}$ падает в основном на участках с большим удельным сопротивлением. Однако величина сопротивления на участке $M_1 M_2$ цепи вдоль линий тока сама зависит от напряжения, поэтому можно записать:*

$$j = \frac{\int_{M_1}^{M_2} d\varphi(M)}{\int_{M_1}^{M_2} \frac{dl}{e_n \mu_n n_0(M) e^{-\frac{e_n \varphi(M)}{kT}} + e_p \mu_p p_0(M) e^{-\frac{e_p \varphi(M)}{kT}}}}. \quad (72.35)$$

Смысл уравнения (72.35) состоит в следующем. *Удельная электрическая проводимость $\sigma(M)$ определяется не исходной концентрацией $n_0(M)$, $p_0(M)$ электронов и дырок в каждой точке M , а той концентрацией, которая создается в результате протекания тока.* Внешнее поле искривляет зоны энергии, смещая их на величину $V(M) \equiv e_n \varphi(M)$, что приводит к изменению концентрации электронов и дырок. Можно объяснить это явление несколько иначе. В результате явления затягивания, разобранный в § 66, концентрация неосновных и основных носителей заряда меняется, становится неравновесной. Следует обратить внимание, что если при одном направлении поля концентрация неосновных носителей заряда увеличивается, то при изменении направления поля на противоположное концентрация неосновных носителей заряда уменьшается. Как мы видели в § 66, одновременно с этим изменяется и концентрация основных носителей заряда и, следовательно, сопротивление. Таким образом, *сопротив-*

ление участка цепи зависит не только от величины тока, но и от его направления. Отсюда следует, что сопротивление полупроводника, имеющего объемные поля и объемные заряды, будет различным. Наиболее отчетливо это проявляется при прохождении тока через p - n -переход. Если к полупроводнику с p - n -переходом приложить внешнее напряжение $\varphi_{вн}$, то оно согласно (72.33) практически полностью будет падать на области объемного заряда, т. е. на p - n -переходе. Предположим, что внешнее поле направлено против контактного поля E_k , для этого оно должно быть направлено от p -области к n -области. Внешнее электрическое поле уменьшает высоту потенциального барьера U^k , в результате чего поток электронов из n -области (и дырок из p -области) возрастает почти по экспоненциальному закону в соответствии с функцией распределения Больцмана. При противоположной полярности напряжения внешнее электрическое поле увеличивает высоту потенциального барьера, основные носители заряда оттягиваются из области перехода, что приводит к увеличению толщины области объемного заряда. Ток через переход создается только в результате движения неосновных носителей заряда, концентрация которых мала, поэтому и ток, называемый обратным, мал.

Резюме § 72

1. В неоднородном полупроводнике существуют объемные поля и объемные заряды. Сопротивление неоднородного полупроводника зависит от величины и направления проходящего по нему тока, поскольку в результате затягивания неосновных носителей заряда будут иметь место явления инжекции, экстракции, эксклюзии и аккумуляции в области всех неоднородностей.

2. Область, в которой происходит резкое изменение типа примеси, называется p - n -переходом. В области p - n -перехода изменяется тип проводимости, возникает сильное объемное поле и объемный заряд. Зоны искривляются так, как это показано на рис. 108. Величина потенциального барьера равна U^k .

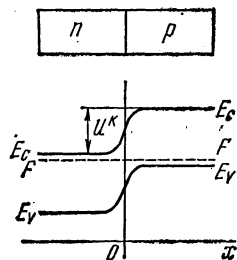


Рис. 108. Возникновение потенциального барьера для основных носителей заряда в области p - n -перехода

Глава VIII

ОПТИЧЕСКИЕ И ФОТОЭЛЕКТРИЧЕСКИЕ ЯВЛЕНИЯ В ПОЛУПРОВОДНИКАХ

§ 73. СПЕКТР ПОГЛОЩЕНИЯ СВЕТА

Свет, попадая в твердое тело, вступает с ним во взаимодействие, связанное с обменом энергией. Обозначим через J *интенсивность света*, т. е. количество световой энергии, проходящей в единицу времени через нормальное единичное сечение. Коэффициент отражения R определяет долю энергии, которая отражается на границе тела:

$$R = \frac{J_R}{J_0}. \quad (73.1)$$

Зависимость коэффициента отражения от частоты $R(\omega)$ или длины волны $R(\lambda)$ называют спектром отражения. Свет, попавший в твердое тело, поглощается им в соответствии с законом Бугера — Ламберта:

$$J(x) = J_0(1 - R)e^{-\alpha x}, \quad (73.2)$$

где x — расстояние от границы тела до данной точки вдоль луча, α называют коэффициентом поглощения, обратная к нему величина α^{-1} численно равна толщине слоя, по прохождении через который интенсивность света уменьшается в e раз. *Зависимость коэффициента поглощения от частоты $\alpha(\omega)$ или длины волны $\alpha(\lambda)$ называют спектром поглощения тела.* Вместо коэффициента поглощения иногда рассматривают *показатель поглощения μ_k* , связанный с коэффициентом поглощения α соотношением

$$\alpha = \frac{4\pi\mu_k}{\lambda} = 4\pi\tilde{\nu}\mu_k \quad (\tilde{\nu} = \lambda^{-1}). \quad (73.3)$$

Электромагнитная теория света показывает, что коэффициент отражения при нормальном падении выражается через показатели преломления и поглощения:

$$R = \frac{(n-1)^2 + n^2\kappa^2}{(n+1)^2 + n^2\kappa^2} = \frac{(n-1)^2 + \frac{\lambda^2}{16\pi^2}\alpha^2}{(n+1)^2 + \frac{\lambda^2}{16\pi^2}\alpha^2}. \quad (73.4)$$

Важной особенностью выражения (73.5) является зависимость коэффициента отражения R от коэффициента поглощения: *с ростом поглощения растет и отражение; $R \approx 1$ при $\kappa \gg 1$, т. е. происходит*

почти полное отражение падающего света. Этим объясняется сильное отражение света металлами (металлический блеск). Таким образом, *если в некоторой области спектра вещество сильно поглощает, то оно и сильно отражает свет в той же области.* Но из (73.4) видно, что отражение имеет место и в отсутствие поглощения; $R \neq 0$ при $\alpha = 0$:

$$R = \frac{(n-1)^2}{(n+1)^2} \text{ (нормальное падение).} \quad (73.5)$$

Для таких прозрачных диэлектриков, как стекло, $n = 1,5$ и $R = 0,04 = 4\%$. Большинство полупроводников имеет большее значение n , например, для германия $n \approx 4$, поэтому чисто диэлектрическое отражение для германия равно $R = (3/5)^2 = 36\%$.

В табл. 24 приведены значения коэффициента преломления n некоторых полупроводниковых веществ, на основе которых можно оценить коэффициент отражения в области слабого поглощения.

Таблица 24

Вещество	n	Вещество	n
C (алмаз)	2,417	InSb	3,988
Si	3,446	GaP	2,97
Ge	4,006	GaAs	3,348
InP	3,37	GaSb	3,748
InAs	3,428	AlSb	3,188

Закон Бугера — Ламберта может быть получен на основе самых общих физических принципов. Рассмотрим слой вещества толщиной $(x, x + dx)$, через который проходит свет. Количество поглощенной энергии в слое толщиной dx должно быть пропорционально толщине слоя dx и количеству падающей на слой световой энергии $J(x)$. Если обозначить коэффициент пропорциональности между поглощенной энергией и падающей через α , то можно записать

$$-dJ(x) = \alpha J(x) dx. \quad (73.6)$$

Поскольку поглощенная энергия приводит к ослаблению интенсивности света, то перед dJ стоит знак минус ($dJ < 0$). Таким образом, α есть *количество поглощенной энергии из пучка единичной интенсивности в слое единичной толщины.* Уравнение (73.6) легко интегрируется:

$$J(x) = J(0) e^{-\alpha x}. \quad (73.7)$$

Величине α можно придать более наглядный смысл в соответствии с принципом, который был использован ранее неоднократно. С этой целью выразим интенсивность J через число фотонов в световом пучке. Если обозначить число фотонов в единице объема пучка через q_1 , то за единицу времени через единичную площадку пройдет $q_1 c$ фотонов, которые несут энергию $q_1 c \hbar \omega = q \hbar \omega$, где $q = q_1 c$ есть поток фотонов, следовательно,

$$J(x) = \hbar \omega q(x). \quad (73.8)$$

Ослабление интенсивности J с точки зрения потока фотонов означает, что число фотонов в пучке уменьшается. Ослабление пучка света может быть связано с рассеянием фотонов или их поглощением. Обозначим вероятность поглощения однофотонного потока на одном поглощающем центре через σ , число поглощающих центров в единице объема через N . В слое толщиной dx содержится $N dx$ поглощающих центров. Число поглощенных фотонов в единицу времени будет равно

$$-dq = \sigma q(x) N dx. \quad (73.9)$$

Интегрируя уравнение (73.10), получим

$$q(x) = q(0) e^{-\sigma N x}. \quad (73.10)$$

Умножим равенство (73.10) на энергию фотона $\hbar\omega$:

$$\hbar\omega q(x) = J(x) = \hbar\omega q(0) e^{-\sigma N x} = J(0) e^{-\sigma N x}. \quad (73.11)$$

Мы видим, что соотношение (73.11) представляет собой закон Бугера — Ламберта, причем коэффициент поглощения связан с концентрацией поглощающих центров и эффективным сечением поглощения одного фотона в единицу времени:

$$\alpha = \sigma N. \quad (73.12)$$

Соотношения типа (73.12) уже неоднократно использовались нами. Если взять величину $(\sigma N)^{-1}$, то ее можно назвать *средней длиной свободного пробега $l_{\text{фот}}$ фотона в поглощающей среде*:

$$l_{\text{фот}} = (\sigma N)^{-1} = \alpha^{-1}. \quad (73.13)$$

Величина α — коэффициент поглощения — есть *вероятность поглощения фотона на расстоянии в единицу длины*. Эффективное сечение σ зависит от энергии фотона и природы поглощающих центров. Если в полупроводнике имеются поглощающие центры N_i различной природы, характеризующиеся своим эффективным сечением $\sigma_i(\omega)$, то

$$a_i(\omega) = \sigma_i(\omega) N_i. \quad (73.14)$$

Полный коэффициент поглощения α есть сумма парциальных коэффициентов поглощения (вероятности независимых процессов складываются):

$$\alpha = \sum_i \alpha_i(\omega) = \sum_i \sigma_i(\omega) N_i = \alpha(\omega). \quad (73.15)$$

Таким образом, полный спектр поглощения складывается из спектров поглощения различных поглощающих центров.

Чтобы вычислить $\alpha(\omega)$, необходимо учесть законы сохранения энергии и импульса. Однако, не вычисляя указанных величин, можно оценить α . Величина σ_i может быть грубо оценена из следующих простых соображений: при поглощении фотона атомом основного вещества и дефектами величину эффективного сечения поглощения можно считать равной площади геометрического сечения

атома или дефекта для тех частот, которые обеспечивают выполнение закона сохранения энергии. Положим $\sigma \sim (10^{-16} - 10^{-17}) \text{ см}^2$.

Если поглощают атомы основного вещества, то $N \approx 10^{22} \text{ см}^{-3}$ и $\alpha \approx (10^{-17} \div 10^{-16}) \cdot 10^{22} = (10^5 \div 10^6) \text{ см}^{-1}$. Приняв эту оценку, для длины свободного пробега фотонов, способных оторвать электрон от атомов основного вещества, т. е. имеющих энергию не меньше ширины запрещенной зоны: $\hbar\omega \geq \Delta E_0$, получим величину $l_{\text{фот}} \sim (10^{-5} \div 10^{-6}) \text{ см} = (0,10 \div 0,01) \text{ мкм}$. Несмотря на грубость оценки для σ , получаем правильный порядок коэффициента поглощения атомами основного вещества. Такое поглощение называют собственным, или фундаментальным.

Если происходит поглощение света дефектами (вакансией, примесным атомом), то их поглощение должно составлять такую же долю от собственного поглощения, какую составляет концентрация дефектов $N_{\text{деф}}$ от концентрации основного вещества. Если $N_{\text{деф}} \approx 10^{16} \text{ см}^{-3}$, то при $\sigma = 10^{-16} \text{ см}^2$ $\alpha = 1 \text{ см}^{-1}$, а при $N_{\text{деф}} \approx 10^{18} \text{ см}^{-3}$ $\alpha \approx 10^2 \text{ см}^{-1}$.

В заключение этого параграфа рассмотрим основные виды поглощения в полупроводниках.

1. *Собственное, или фундаментальное, поглощение света* приводит к переходу электрона из связанного состояния в свободное, т. е. из валентной зоны в зону проводимости. Собственное поглощение возможно при условии, что $\hbar\omega \geq \Delta E_0$. Оно наблюдается в видимой и ближней инфракрасной областях в зависимости от ширины запрещенной зоны.

2. *Примесное поглощение* вызвано ионизацией атомов примеси, т. е. или переходом электронов от атома примеси в зону проводимости, или из валентной зоны на уровни примеси.

3. *Поглощение свободными носителями заряда* обусловлено их движением под действием электрических полей световой волны. На ускорение свободных носителей заряда волна отдает часть своей энергии, что приводит к ослаблению волны.

4. Световая волна вступает во взаимодействие с колебаниями решетки, изменяя число оптических фононов. Это поглощение носит название *поглощения колебаниями решетки, или решеточным поглощением*.

5. Если происходит образование связанной пары электрон — дырка, то такое поглощение называют *экситонным*.

6. Внутризонное поглощение наблюдается в веществах, имеющих сложную структуру зон, подобно валентной зоне германия и кремния.

7. Поглощение света совокупностью свободных электронов и дырок называют *плазменным поглощением*.

Из перечисленных видов поглощения следует, что спектр поглощения должен зависеть от всех внешних воздействий, способных изменить состояние атомов основного вещества, дефектов, колебаний решетки. Поэтому можно ожидать, что на спектр поглощения должны влиять температура, степень легирования, давление, магнитные и электрические поля, облучение частицами.

1. Ослабление света при прохождении пути x в веществе определяется законом Бугера — Ламберта

$$J(x) = J(0) e^{-\alpha x}. \quad (73.1p)$$

Коэффициент поглощения α определяется эффективным сечением $\sigma(\omega)$ поглощения фотонов и концентрацией поглощающих центров N :

$$\alpha(\omega) = \sigma(\omega) N; \quad l_{\text{фот}} = (\sigma N)^{-1} = \alpha^{-1}. \quad (73.2p)$$

2. Зависимость коэффициента поглощения от частоты ω или длины волны λ называют спектром поглощения. Зависимость коэффициента отражения от ω или λ называют спектром отражения.

3. В полупроводниках наблюдается несколько механизмов поглощения: собственное, или фундаментальное, поглощение; поглощение свободными носителями заряда; внутризонное поглощение; поглощение примесью; поглощение колебаниями решетки; экситонное поглощение; плазменное поглощение.

§ 74. ПОГЛОЩЕНИЕ СВЕТА СВОБОДНЫМИ НОСИТЕЛЯМИ ЗАРЯДА

Рассмотрим в качестве первого механизма поглощение света свободными носителями заряда, так как его можно описать в рамках классической электродинамики. Световая волна, попадая в проводящую среду, оказывает воздействие на электроны зоны проводимости и дырки валентной зоны. Электроны, ускоряясь, увеличивают свою энергию за счет энергии волны. Сталкиваясь с решеткой, они отдают свою энергию решетке. В конечном счете энергия световой волны превращается в тепловую энергию решетки.

Запишем уравнения Максвелла (в системе Гаусса) для случая, когда отсутствуют сторонние поля и объемные заряды:

$$\text{rot } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c} \mathbf{j}; \quad (74.1)$$

$$\text{rot } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}; \quad (74.2)$$

$$\text{div } \mathbf{D} = 0; \quad (74.3)$$

$$\text{div } \mathbf{B} = 0; \quad (74.4)$$

$$\mathbf{D} = \epsilon \mathbf{E}; \quad (74.5)$$

$$\mathbf{B} = \mu \mathbf{H}; \quad (74.6)$$

$$\mathbf{j} = \sigma \mathbf{E}. \quad (74.7)$$

Перейдем к волновому уравнению, для чего найдем ротор уравнения (74.1) или (74.2):

$$\begin{aligned} \operatorname{rot} \operatorname{rot} \mathbf{H} &= \operatorname{grad} \operatorname{div} \mathbf{H} - \nabla^2 \mathbf{H} = -\nabla^2 \mathbf{H} = \frac{1}{c} \operatorname{rot} \frac{\varepsilon \partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \operatorname{rot} \sigma \mathbf{E} = \\ &= \frac{1}{c} \left(\varepsilon \frac{\partial}{\partial t} + 4\pi\sigma \right) \operatorname{rot} \mathbf{E} = -\frac{1}{c^2} \left(\varepsilon \frac{\partial}{\partial t} + 4\pi\sigma \right) \mu \frac{\partial \mathbf{H}}{\partial t} \end{aligned} \quad (74.8)$$

или

$$\nabla^2 \mathbf{H} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} - \frac{4\pi \mu \sigma}{c^2} \frac{\partial \mathbf{H}}{\partial t} = 0. \quad (74.9)$$

При выводе уравнения (74.9) мы использовали соотношения (74.1—7). Если $\sigma = 0$, то уравнение (74.9) переходит в волновое уравнение

$$\Delta \mathbf{H} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0, \quad (74.10)$$

решение которого имеет вид плоской волны:

$$\mathbf{H} = \mathbf{H}_0 \left(t - \frac{(\mathbf{k}^0 \mathbf{r})}{v} \right), \quad (74.11)$$

идущей со скоростью $v = \frac{c}{\sqrt{\varepsilon \mu}}$ в направлении единичного вектора \mathbf{k}^0 .

Беря ротор от (74.2), получим уравнение для \mathbf{E} :

$$\Delta \mathbf{E} - \frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \frac{4\pi \mu \sigma}{c^2} \frac{\partial \mathbf{E}}{\partial t} = 0. \quad (74.12)$$

Решение уравнений (74.12) и (74.9) удобно искать методом монохроматических плоских волн, для этого представим $\mathbf{E} = \mathbf{E}(\mathbf{r}, t)$ в виде интеграла Фурье по всем возможным частотам:

$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, \omega) e^{i\omega t} d\omega, \quad (74.13)$$

где фурье-компонент $\mathbf{E}(\mathbf{r}, \omega)$ поля равен

$$\mathbf{E}(\mathbf{r}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, t) e^{-i\omega t} dt. \quad (74.14)$$

Найдем уравнения для фурье-компонента поля \mathbf{E} :

$$\nabla^2 \mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \nabla^2 \mathbf{E}(\mathbf{r}, \omega) e^{i\omega t} d\omega, \quad (74.15)$$

$$\frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} = \int_{-\infty}^{\infty} i\omega \mathbf{E}(\mathbf{r}, \omega) e^{i\omega t} d\omega, \quad (74.16)$$

$$\frac{\partial^2 \mathbf{E}(\mathbf{r}, t)}{\partial t^2} = - \int_{-\infty}^{\infty} \omega^2 \mathbf{E}(\mathbf{r}, \omega) e^{i\omega t} d\omega. \quad (74.17)$$

Подставляя (74.15 — 17) в (74.12), получим

$$\int_{-\infty}^{\infty} \left\{ \nabla^2 \mathbf{E}(\mathbf{r}, \omega) + \left[\frac{\omega^2 \varepsilon \mu}{c^2} - i \frac{4\pi \mu \sigma \omega}{c^2} \right] \mathbf{E}(\mathbf{r}, \omega) \right\} e^{i\omega t} d\omega = 0. \quad (74.18)$$

Из (74.18) следует, что амплитуда $\mathbf{E}(\mathbf{r}, \omega)$ монохроматической волны удовлетворяет уравнению

$$\Delta \mathbf{E}(\mathbf{r}, \omega) + \left[\frac{\omega^2 \varepsilon \mu}{c^2} - i \frac{4\pi \mu \sigma \omega}{c^2} \right] \mathbf{E}(\mathbf{r}, \omega) = 0. \quad (74.19)$$

Запишем коэффициент при $\mathbf{E}(\mathbf{r}, \omega)$ в виде

$$\frac{\omega^2 \varepsilon \mu}{c^2} - i \frac{4\pi \mu \sigma \omega}{c^2} = \frac{\omega^2}{c^2} n^2 (1 - i\kappa)^2, \quad (74.20)$$

где n и κ — две неизвестные величины.

Возводя в квадрат правую часть (74.19) и приравнявая соответственно действительные и мнимые величины, получим

$$\begin{aligned} \frac{\omega^2 \varepsilon \mu}{c^2} &= \frac{\omega^2}{c^2} n^2 (1 - \kappa^2); & n^2 (1 - \kappa^2) &= \varepsilon \mu; \\ i \frac{4\pi \mu \sigma \omega}{c^2} &= i \frac{2\kappa n^2 \omega^2}{c^2}; & n^2 \kappa &= \frac{2\pi \mu \sigma}{\omega}. \end{aligned} \quad (74.21)$$

Из (74.21) можно выразить n и κ :

$$n = \sqrt{\frac{\varepsilon \mu}{2} \left(\sqrt{1 + \frac{16\pi^2 \sigma^2}{\omega^2 \varepsilon^2}} + 1 \right)}, \quad (74.22)$$

$$n\kappa = \sqrt{\frac{\varepsilon \mu}{2} \left(\sqrt{1 + \frac{16\pi^2 \sigma^2}{\omega^2 \varepsilon^2}} - 1 \right)}; \quad (74.23)$$

n носит название *показателя преломления*, $n\kappa$ — *показателя поглощения*. Если $\sigma = 0$, то $n = \sqrt{\varepsilon \mu}$ и $n\kappa = 0$. Используя (74.19), запишем уравнение (74.18) в таком виде:

$$\Delta \mathbf{E}(\mathbf{r}; \omega) + \frac{\omega^2 n^2 (1 - i\kappa)^2}{c^2} \mathbf{E}(\mathbf{r}; \omega) = 0. \quad (74.24)$$

Решение этого уравнения легко находится:

$$\mathbf{E}(\mathbf{r}; \omega) = \mathbf{E}_{0\omega} e^{\pm i \frac{\omega n}{c} (1 - i\kappa) (\mathbf{k}^0 \mathbf{r})} = \mathbf{E}_{0\omega} e^{\mp \frac{\omega n \kappa}{c} (\mathbf{k}^0 \mathbf{r})} e^{\pm i \frac{\omega n}{c} (\mathbf{k}^0 \mathbf{r})} = \mathbf{E}_{\omega}. \quad (74.25)$$

Вводя волновой вектор $\mathbf{k} = \frac{\omega}{c} \mathbf{k}^0$, запишем уравнение (74.25) для монохроматической волны, в котором сохраним только знак минус:

$$\mathbf{E}_{\omega}(\mathbf{r}, t) = \mathbf{E}_{0\omega} e^{-n\kappa (\mathbf{k} \mathbf{r})} e^{-i[\omega t - n(\mathbf{k} \mathbf{r})]}. \quad (74.26)$$

Выражение (74.26) показывает, что у плоской монохроматической волны по мере прохождения в среде, обладающей конечной проводимостью σ , уменьшается амплитуда (если $\kappa < 0$, то среда усили-

вает излучение). Поскольку интенсивность света пропорциональна квадрату амплитуды, то, отсчитывая расстояние вдоль луча (kx) = kx , получим

$$J(x) = J_{0\omega} e^{-2\pi kx} = J_{0\omega} e^{-\alpha x}, \quad (74.27)$$

$$\alpha = 2\pi kx = \frac{2\pi k\omega}{c} = \frac{2\pi \cdot 2\pi}{\lambda} = \frac{4\pi}{\lambda} \pi k. \quad (74.28)$$

Таким образом, коэффициент поглощения света α определяют согласно теории Максвелла проводимостью вещества:

$$\alpha = \frac{4\pi}{\lambda} \pi k = \frac{4\pi}{\lambda} \sqrt{\frac{\epsilon\mu}{2} \left[\sqrt{1 + \frac{16\pi^2\sigma^2}{\omega^2\epsilon^2}} - 1 \right]}. \quad (74.29)$$

Оценим α при некоторых значениях величин ϵ , μ , σ , ω . Положим $\omega \approx 10^{15} \text{ с}^{-1}$, $\sigma = 1 \text{ Ом}^{-1} \text{ см}^{-1} = 9 \cdot 10^{11} \text{ СГСЭ}\sigma$; $\epsilon \approx 10$, $\mu \approx 1$. В таком случае

$$\frac{16\pi^2\sigma^2}{\omega^2\epsilon^2} \approx \frac{16 \cdot 10 \cdot 8 \cdot 10^{23}}{10^{30} \cdot 10^2} = 1,28 \cdot 10^{-6} \ll 1,$$

причем неравенство сохраняется вплоть до $\sigma \sim 10^2 \text{ Ом}^{-1} \text{ см}^{-1}$, которая наблюдается в сильно легированных полупроводниках. Разложив корень в (74.28) в ряд, запишем

$$\alpha = \frac{4\pi}{\lambda} \sqrt{\frac{\epsilon\mu}{2} \cdot \frac{1}{2} \cdot \frac{16\pi^2\sigma^2}{\omega^2\epsilon^2}} = \frac{8\pi^2 \sqrt{\epsilon\mu} \sigma}{\epsilon\lambda\omega} = \frac{4\pi\sigma \sqrt{\epsilon\mu}}{\epsilon c}. \quad (74.30)$$

Если учесть оценку подкоренного выражения в α или πk , то выражение (74.22) для n примет вид

$$n = \sqrt{\epsilon\mu}. \quad (74.31)$$

Величина μ для частот видимого света практически равна единице, поэтому можно записать $n = \sqrt{\epsilon}$ и

$$\alpha = \frac{4\pi\sigma}{c n}. \quad (74.32)$$

Если измерять σ в $\text{Ом}^{-1} \text{ см}^{-1}$ и α в см^{-1} , то

$$\alpha (\text{см}^{-1}) = \frac{120\pi}{n} \sigma (\text{Ом}^{-1} \text{ см}^{-1}). \quad (74.33)$$

При $\sigma = 1 \text{ Ом}^{-1} \text{ см}^{-1}$ и $n = 3,14$ $\alpha = 120 \text{ см}^{-1}$.

Так как проводимость σ пропорциональна концентрации носителей заряда, то и коэффициент поглощения определяется концентрацией свободных носителей заряда. Обозначим концентрацию свободных носителей заряда через p , запишем

$$\alpha = \frac{4\pi\sigma}{c n} = \frac{4\pi e^2 \langle \tau \rangle}{c n m^*} p. \quad (74.34)$$

Таким образом, коэффициент поглощения света свободными носителями заряда с концентрацией p зависит от их эффективной

массы, усредненного времени релаксации и показателя преломления среды. Но показатель преломления является в действительности функцией длины волны, следовательно, и коэффициент поглощения свободными носителями заряда должен зависеть от длины волны. Например, Коши нашел теоретически задолго до теории Максвелла, что показатель преломления можно представить в виде

$$n = a + \frac{b}{\lambda^2} + \frac{c'}{\lambda^3} + \dots, \quad (74.35)$$

где a , b , c' — константы, которые необходимо определять экспериментально для каждого вещества.

Электронная теория позволяет найти зависимость n от λ . Рассмотрим с этой целью поляризацию вещества, которая возникает под действием электрического поля:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \epsilon\mathbf{E}. \quad (74.36)$$

Если $\mathbf{P} = \chi\mathbf{E}$, то $\epsilon = 1 + 4\pi\chi$. Чтобы найти ϵ , подсчитаем смещение зарядов. Пусть в атомах содержатся несколько сортов зарядов e_i числом f_i с концентрацией атомов N_i . Смещение зарядов равно \mathbf{r}_i , полный электрический момент единицы объема (поляризация) будет равен

$$\mathbf{P} = \sum_i N_i e_i f_i \mathbf{r}_i = \chi\mathbf{E}. \quad (74.37)$$

Чтобы найти смещение заряда e , запишем уравнение его движения. На него должна действовать квазиупругая, возвращающая сила $\mathbf{f}_{\text{возвр}} = -\gamma\mathbf{r}$, сила «трения», зависящая от его скорости и приводящая к расходу энергии $\mathbf{f}_{\text{трени}} = -g\dot{\mathbf{r}}$, и вынуждающая сила, обусловленная полем \mathbf{E} :

$$m^*\ddot{\mathbf{r}} = -\gamma\mathbf{r} - g\dot{\mathbf{r}} + e\mathbf{E}. \quad (74.38)$$

Обозначая $\gamma/m^* = \omega_0^2$, $g/m^* = b$, получим

$$\ddot{\mathbf{r}} + b\dot{\mathbf{r}} + \omega_0^2\mathbf{r} = \frac{e}{m^*}\mathbf{E} = \frac{e}{m^*}\mathbf{E}_0 e^{-i\omega t}, \quad (74.39)$$

где \mathbf{E}_0 — амплитуда колебаний волны в точке нахождения заряда. Уравнение (74.39) представляет собой уравнение вынужденных колебаний. В стационарном состоянии колебания происходят с частотой вынуждающей силы, но со сдвигом фазы, поэтому положим

$$\mathbf{r}(t) = \mathbf{r}_0 e^{-i\omega t - i\varphi}. \quad (74.40)$$

Подставим (74.40) в (74.39), получим после сокращения на $e^{-i\omega t}$

$$-\omega^2\mathbf{r}_0 e^{-i\varphi} - i\omega b\mathbf{r}_0 e^{-i\varphi} + \omega_0^2\mathbf{r}_0 e^{-i\varphi} = \frac{e}{m^*}\mathbf{E}_0, \quad (74.41)$$

откуда

$$\mathbf{r}_0 = \frac{\frac{e}{m^*}\mathbf{E}_0 e^{i\varphi}}{\omega_0^2 - \omega^2 - i\omega b}. \quad (74.42)$$

Подставляя величину стационарного смещения зарядов в выражение для поляризации и вынося поле E_0 , получим

$$\epsilon = 1 + 4\pi\chi = 1 + 4\pi \sum_l N_l e_l f_l \frac{\frac{e_l}{m_l^*} e^{i\varphi_l}}{[\omega_{0l}^2 - \omega^2] - i\omega b_l} = n^2 (1 - i\kappa)^2. \quad (74.43)$$

Приравняем действительные и мнимые части двух последних выражений в (74.43)

$$n^2 (1 - \kappa^2) = 1 + 4\pi \sum_l N_l \frac{e_l^2}{m_l^*} f_l \frac{[\omega_{0l}^2 - \omega^2] \cos \varphi_l - b_l \omega \sin \varphi_l}{[\omega_{0l}^2 - \omega^2]^2 + b_l^2 \omega^2}; \quad (74.44)$$

$$2n^2 \kappa = 1 + 4\pi \sum_l N_l \frac{e_l^2}{m_l^*} f_l \frac{[\omega_{0l}^2 - \omega^2] \sin \varphi_l + b_l \omega \cos \varphi_l}{[\omega_{0l}^2 - \omega^2]^2 + b_l^2 \omega^2}. \quad (74.45)$$

Решая систему (74.44–45), получим выражение для n и $n\kappa$. Мы видим, что они действительно зависят от частоты. Если провести расчет поляризуемости атомов согласно квантовой теории, то вид частотной зависимости n и $n\kappa$, найденный в классической электронной теории, сохраняется. Меняется только смысл величины f_l : в классической электронной теории f_l — число электронов в одном атоме в данном состоянии, в квантовой теории вместо f_l появляется величина

$$f_{ln} = \frac{2m |x_{ln}|^2 \omega_{ln}}{\hbar} = \frac{2m\omega_{ln}}{e^2 \hbar} |d_{ln}|^2, \quad (74.46)$$

которая называется силой осциллятора для перехода $E_l \rightarrow E_n$; $\omega_{ln} = \frac{E_l - E_n}{\hbar}$; d_{ln} является матричным элементом дипольного момента $\mathbf{d} = e\mathbf{r}$ атома, который индуцируется полем. Опуская расчеты, запишем выражение для ϵ :

$$\epsilon \approx n^2(\omega) = 1 + \frac{4\pi N}{m^*} \sum_{ln} \omega_l \frac{f_{ln}}{\omega_{ln}^2 - \omega^2}. \quad (74.47)$$

Здесь ω_l — вероятность того, что атом находится в состоянии E_l ; $\omega_l \sim e^{-\frac{E_l}{kT}}$. Выражение (74.47) получено без учета затухания колебаний. Как видим из (74.47) и (74.44), частотные зависимости n^2 по классической и квантовой теории совпадают.

Если применять выражение для $n(\omega)$ с учетом того, что электроны проводимости являются квазисвободными, то необходимо считать $\gamma = 0$ и, следовательно, $\omega_0 = 0$. Положим для простоты $\varphi = 0$ и будем считать все электроны проводимости одинаковыми; с учетом

этого выражения (74.44) и (74.45) можно переписать в виде

$$n^2(1 - \kappa^2) = 1 - \frac{4\pi}{m^*} \rho e^2 \frac{1}{\omega^2 + b^2} = 1 - A = C, \\ \left(A = \frac{4\pi\sigma}{\langle \tau \rangle} \frac{1}{\omega^2 + b^2} \right); \quad (74.48)$$

$$2n^2\kappa = \frac{4\pi\rho e^2}{m^*} \frac{l\omega}{\omega^4 + b^2\omega^2} = \frac{1}{2} \frac{b}{\omega} A = B. \quad (74.49)$$

Решая систему уравнений (74.48 – 49), получим

$$n = \frac{B}{\sqrt{\frac{C}{2} \left[\sqrt{1 + \frac{4B^2}{C^2}} - 1 \right]}}, \quad (74.50)$$

$$n\kappa = \sqrt{\frac{C}{2} \left[\sqrt{1 + \frac{4B^2}{C^2}} - 1 \right]}. \quad (74.51)$$

Учитывая, что для больших частот $A \ll 1$ и $B \ll 1$, получим

$$n\kappa = \sqrt{\frac{C}{2} \left[1 + \frac{2B^2}{C^2} + \dots - 1 \right]} = \frac{B}{\sqrt{C}} \approx B = \frac{2\pi\sigma b}{\langle \tau \rangle \omega} \frac{1}{\omega^2 + b^2}; \quad (74.52)$$

$$n = \frac{B}{\sqrt{\frac{C}{2} \left[1 + \frac{2B^2}{C^2} + \dots - 1 \right]}} = \\ = \sqrt{C} = \sqrt{1 - A} = 1 - \frac{A}{2} = 1 - \frac{2\pi\sigma}{\langle \tau \rangle} \frac{1}{\omega^2 + b^2}. \quad (74.53)$$

Найдем теперь коэффициент поглощения α :

$$\alpha = \frac{4\pi}{\lambda} n\kappa = \frac{8\pi^2\sigma b}{\langle \tau \rangle \lambda \omega} \frac{1}{\omega^2 + b^2} = \frac{4\pi\sigma b}{c \langle \tau \rangle} \frac{1}{\omega^2 + b^2}. \quad (74.54)$$

Если сравнить полученный результат (74.54) с (74.32), то легко увидеть различие и сходство между ними: по феноменологической теории Максвелла

$$\alpha = \alpha_{\text{фен}} = \frac{4\pi\sigma}{c} \frac{1}{\sqrt{\epsilon}}, \quad (74.55)$$

по электронной теории

$$\alpha = \alpha_{\text{эл}} = \frac{4\pi\sigma}{c} \frac{b}{\langle \tau \rangle} \frac{1}{\omega^2 + b^2}. \quad (74.56)$$

Из (74.56) и (74.55) имеем

$$\frac{\alpha_{\text{эл}}}{\alpha_{\text{фен}}} = \frac{b\sqrt{\epsilon}}{\langle \tau \rangle} \frac{1}{\omega^2 + b^2}. \quad (74.57)$$

При $\omega^2 \ll b^2$ коэффициент поглощения не зависит от частоты и равен

$$\alpha_{\text{эл}} = \frac{4\pi\sigma}{c} \frac{1}{b \langle \tau \rangle}. \quad (74.58)$$

Поскольку электронная теория Лоренца при больших длинах волн должна переходить в феноменологическую теорию Максвелла, или, другими словами, $\varepsilon(\omega)$, $\mu(\omega)$, $\sigma(\omega)$ можно заменить их статическими значениями, то отсюда можем оценить величину b :

$$b \langle \tau \rangle = \sqrt{\varepsilon}; \quad b = \frac{\sqrt{\varepsilon}}{\langle \tau \rangle}. \quad (74.59)$$

Подставляя выражение для b из (74.59) в (74.56), запишем

$$\alpha = \frac{4\pi\sigma}{c} \frac{\sqrt{\varepsilon}}{\langle \tau \rangle^2} \frac{1}{\omega^2 + \frac{\varepsilon}{\langle \tau \rangle^2}} = \frac{4\pi\sigma}{c} \frac{\sqrt{\varepsilon}}{\varepsilon + \omega^2 \langle \tau \rangle^2}. \quad (74.60)$$

Таким образом, если $\omega^2 \ll b^2 = \frac{\varepsilon}{\langle \tau \rangle^2}$, то коэффициент поглощения света свободными носителями заряда не зависит от частоты (или длины волны) света, он остается постоянным и равным

$$\alpha = \alpha_0 = \frac{4\pi\sigma}{c\sqrt{\varepsilon}} = \frac{4\pi e^2 \langle \tau \rangle}{c\sqrt{\varepsilon} m^*} p. \quad (74.61)$$

Оценим величину эффективного сечения поглощения света электроном; положим $\langle \tau \rangle \approx 10^{-13}$ с; $\varepsilon = 10$; $m^* = 10^{-27}$ г; получим $\alpha/p = 3 \cdot 10^{-15}$ см².

Если концентрация электронов или дырок $p = 10^{16}$ см⁻³, то $\alpha = 30$ см⁻¹, но при $p = 10^{19}$ см⁻³ $\alpha = 30\,000$ см⁻¹.

Если $\omega^2 \gg b^2$, то согласно (74.56) коэффициент поглощения зависит от длины волны:

$$\alpha = \frac{4\pi\sigma}{c} \frac{b}{\langle \tau \rangle} \frac{1}{\omega^2} = \frac{4\pi\sigma}{c\sqrt{\varepsilon}} \frac{\varepsilon}{\langle \tau \rangle^2} \frac{1}{\omega^2} = \alpha_0 \frac{\varepsilon}{\langle \tau \rangle^2} \frac{\lambda^2}{4\pi^2}. \quad (74.62)$$

Таким образом, в области больших частот $\alpha \sim \lambda^2$, при этом $\alpha < \alpha_0$, поскольку $\omega^2 \gg \varepsilon/\langle \tau \rangle^2$.

Оценим граничную частоту, при которой происходит изменение зависимости $\alpha \sim \lambda^2$ на $\alpha \sim \lambda^0$. Выразим для этого α через α_0 :

$$\alpha = \frac{4\pi\sigma}{c} \frac{b}{\langle \tau \rangle} \frac{1}{\omega^2 + b^2} = \frac{4\pi\sigma}{c} \frac{1}{b \langle \tau \rangle} \frac{b^2}{\omega^2 + b^2} = \alpha_0 \frac{b^2}{\omega^2 + b^2}. \quad (74.63)$$

Определяя границу перехода условием $\alpha = \alpha_0/2$, получим $\omega_{\text{гр}}^2 = b^2$, или

$$\omega_{\text{гр}} = b = \frac{\sqrt{\varepsilon}}{\langle \tau \rangle}. \quad (74.64)$$

При $\langle \tau \rangle = 10^{-13}$ с, $\varepsilon = 10$, $\omega_{\text{гр}} \approx 3 \cdot 10^{13}$ с⁻¹, что дает для $\lambda_{\text{гр}} = \frac{2\pi c}{\omega_{\text{гр}}} \approx \frac{3 \cdot 10^{10}}{3 \cdot 10^{13}} = 6,3 \cdot 10^{-3}$ см = 63 мкм. Выразим $\alpha(\lambda)$ через $\lambda_{\text{гр}}$, учитывая (74.64) и (74.63):

$$\alpha = \alpha_0 \frac{\omega_{\text{гр}}^2}{\omega_{\text{гр}}^2 + \omega^2} = \alpha_0 \frac{\lambda^2}{\lambda_{\text{гр}}^2 + \lambda^2}. \quad (74.65)$$

Вместо оценки величины α и $\langle \tau \rangle$ удобно использовать экспериментальное значение подвижности μ_d , поскольку α_0 можно выразить через μ_d :

$$\alpha_0 = \frac{4\pi e}{c\sqrt{\epsilon}} \mu_d \rho. \quad (74.66)$$

Электронная теория в отличие от феноменологической теории Максвелла приводит к квадратичной зависимости коэффициента поглощения от длины волны, при этом он пропорционален подвижности, или усредненному времени релаксации. Но при поглощении свободными носителями заряда энергии световой волны происходит

переход частиц с одного уровня энергии на другой. Следует ожидать поэтому различной зависимости коэффициента поглощения от длины волны при доминировании различных механизмов рассеяния.

Рассматривая взаимодействие свободных носителей заряда с фотонами, мы должны иметь в виду, что поглощение свободной частицей фотона невозможно — не могут быть выполнены одновременно законы сохранения импульса и энергии. Поглощение же свободными носителями заряда происходит с участием поля решетки, поэтому для этого процесса важное значение имеет рассеяние носителей заряда на дефектах кристаллического поля. Учет механизма рассеяния приводит к различной спектральной зависимости коэффициента поглощения, для одного механизма рассеяния можно записать в случае невырожденного газа:

$$\alpha \sim \lambda^p, \quad (74.67)$$

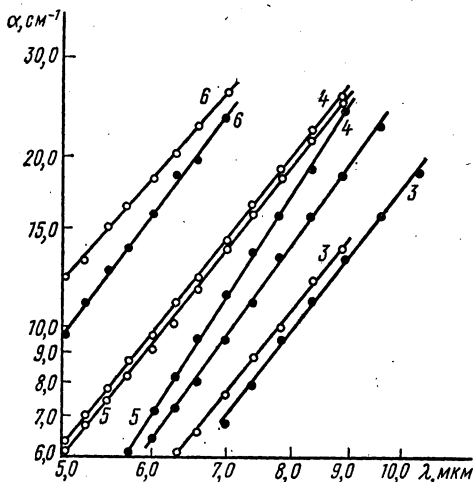


Рис. 109. Спектр поглощения свободными носителями заряда в теллуриде кадмия

Параметры исследованных образцов

Образец	$N, \text{см}^{-3}$
1	$6 \cdot 10^{15}$
2	$3,6 \cdot 10^{16}$
3	$1,7 \cdot 10^{17}$
4	$2,0 \cdot 10^{17}$
5	$3,1 \cdot 10^{17}$
6	$5,8 \cdot 10^{17}$

где степень p принимает значения $3/2$; $5/2$ и $7/2$ при рассеянии на акустических колебаниях, на оптических колебаниях и ионах примеси. Значения показателя степени указаны для параболических зон и невырожденного газа. Вырождение газа приводит к некоторому уменьшению показателя степени, а отклонение от параболическости, т. е. от квадратичного закона дисперсии, — к увеличению.

Например, расчеты для фотонов с достаточно большой энергией $\hbar\omega > \Delta E_0$ показывают, что ρ равно 1,66; 2,83 и 4,21 соответственно для акустических и оптических колебаний решетки и ионов примеси. Если в рассеянии принимают участие два или более механизмов рассеяния, то спектр поглощения свободными носителями заряда становится более сложным. На рис. 109 приведен спектр поглощения электронами в теллуриде кадмия. Как видно из рис. 109, наклон прямых с ростом концентрации электронов возрастает, что свидетельствует о возрастающей роли поглощения при рассеянии на ионах примеси. Аналогичный эффект имеет место при понижении температуры.

§ 75. ЦИКЛОТРОННЫЙ РЕЗОНАНС

Чтобы найти зависимость $\alpha(\lambda)$, мы использовали результаты теории дисперсии для атомов, положив $\omega_0 = 0$. Можно получить эти же результаты другим способом, учитывая, что проводимость должна зависеть от частоты электрического поля.

Пусть v — скорость направленного движения носителей заряда, изменение ее происходит в результате действия внешней силы и соударений. Если τ есть среднее время, в течение которого полностью теряется скорость направленного движения частиц, то уравнение движения носителей заряда можно записать в виде

$$\frac{dv}{dt} = \frac{eE}{m^*} - \frac{v}{\tau}. \quad (75.1)$$

Первый член справа в (75.1) определяет приращение скорости за единицу времени под действием поля E , второй член определяет потерю скорости за единицу времени вследствие столкновений. Этим уравнением мы пользовались в сущности неоднократно. Действительно, в стационарном состоянии при постоянном поле имеем

$$\frac{eE}{m^*} - \frac{v}{\tau} = 0, \quad (75.2)$$

или

$$v = \frac{e\tau}{m^*} E = \mu E. \quad (75.3)$$

Предположим теперь, что внешнее поле является переменным. Так как любое переменное поле можно представить в виде суперпозиции гармонических колебаний, то мы предположим, что

$$E = E_0 e^{i\omega t} \quad (75.4)$$

и

$$\frac{dv}{dt} = \frac{eE_0}{m^*} e^{i\omega t} - \frac{v}{\tau}. \quad (75.5)$$

Общее решение однородного уравнения описывает процесс установления стационарного состояния. Считая, что $t \gg \tau$, найдем установившееся движение, определяемое частным решением неоднородного уравнения. Установившееся движение представляет собой вынужден-

ные колебания с частотой вынуждающей силы, поэтому положим

$$\mathbf{v}(t) = \mathbf{v}_0 e^{i\omega t}. \quad (75.6)$$

Подставляя (75.6) в (75.5) и сокращая на $e^{i\omega t}$, получим

$$i\omega \mathbf{v}_0 = \frac{e}{m^*} \mathbf{E}_0 - \frac{\mathbf{v}_0}{\tau}, \quad (75.7)$$

или

$$\mathbf{v}_0 = \frac{e\mathbf{E}_0}{m^* \left(i\omega + \frac{1}{\tau} \right)} = \mu \frac{1 - i\omega\tau}{1 + \omega^2\tau^2} \mathbf{E}_0, \quad (75.8)$$

откуда для $\mathbf{v}(t)$ имеем

$$\mathbf{v}(t) = \frac{\mu}{1 + \omega^2\tau^2} [1 - i\omega\tau] \mathbf{E}_0 e^{i\omega t} = \frac{\mu}{1 + \omega^2\tau^2} \left[\mathbf{E} - \tau \frac{d\mathbf{E}}{dt} \right]. \quad (75.9)$$

Умножим величину $\mathbf{v}(t)$ на заряд частиц и их концентрацию ρ , получим плотность тока

$$\mathbf{j} = \frac{e^2\tau\rho}{m^* [1 + \omega^2\tau^2]} \left[\mathbf{E} - \tau \frac{\partial \mathbf{E}}{\partial t} \right]. \quad (75.10)$$

Ток, определяемый выражением (75.10), состоит из двух частей: член, пропорциональный \mathbf{E} (или действительной части \mathbf{v}_0), представляет собой омический ток

$$\mathbf{j}_E = \frac{e^2\tau\rho}{m^* [1 + \omega^2\tau^2]} \mathbf{E}(t) = \sigma(\omega) \mathbf{E}(t), \quad (75.11)$$

где

$$\sigma(\omega) = \frac{e^2\tau\rho}{m^*} \frac{1}{1 + \omega^2\tau^2} = \frac{\sigma_0}{1 + \omega^2\tau^2}. \quad (75.12)$$

Вторая составляющая тока, пропорциональная $\partial\mathbf{E}/\partial t$, есть ток поляризации

$$\mathbf{j}^{(n)} = -\sigma\tau \frac{\partial \mathbf{E}}{\partial t} = \frac{\partial \chi \mathbf{E}}{\partial t}, \quad (75.13)$$

где поляризуемость χ равна

$$\chi = -\sigma\tau = -\frac{e^2\tau^2\rho}{m^* [1 + \omega^2\tau^2]}. \quad (75.14)$$

Из (75.14) можем записать для диэлектрической проницаемости

$$\epsilon(\omega) = 1 + 4\pi\chi = 1 - \frac{4\pi e^2\tau^2\rho}{m^* [1 + \omega^2\tau^2]} = 1 - \frac{4\pi\sigma_0\tau}{1 + \omega^2\tau^2}. \quad (75.15)$$

Положив $\epsilon = 0$, найдем некоторую критическую частоту $\omega_{кр}$:

$$\omega_{кр} = \frac{1}{\tau} \sqrt{4\pi\sigma_0\tau - 1}. \quad (75.16)$$

Если $\omega > \omega_{кр}$, то $\epsilon > 0$ и n — действительная величина, коэффициент поглощения света согласно (74.32) имеет вид

$$\alpha = \frac{4\pi\sigma}{c n} = \frac{4\pi}{c \sqrt{1 - \frac{4\pi\sigma_0\tau}{\omega^2\tau^2}}} \frac{\sigma_0}{1 + \omega^2\tau^2} \approx \frac{4\pi\sigma_0}{c\tau^2} \frac{1}{\omega^2}, \quad (75.17)$$

т. е. в области больших частот $\alpha \sim \lambda^2$. Если $\omega \ll \omega_{кр}$, то $\epsilon < 0$ и p — мнимая величина, поглощение велико:

$$\alpha \approx \frac{4\pi}{\lambda} \sqrt{\frac{4\pi\sigma_0\tau}{1+\omega^2\tau^2} - 1}. \quad (75.18)$$

Таким образом, мы пришли к новому и несколько неожиданному результату. В области «малых» частот коэффициент поглощения зависит от λ в виде $\alpha \sim \lambda^{-1}$, при этом поглощение велико. В области больших частот $\alpha \sim \lambda^2$ и поглощение сравнительно мало. Оценим критическую частоту согласно (75.16):

$$\omega_{кр} = \frac{1}{\tau} \sqrt{4\pi\sigma_0\tau - 1} = \sqrt{\frac{4\pi e^2 p}{m^*} - \frac{1}{\tau^2}}. \quad (75.19)$$

Критическая частота $\omega_{кр}$ существует, если концентрация носителей заряда велика:

$$p > \frac{m^*}{\tau^2 4\pi e^2}. \quad (75.20)$$

Полагая $\tau \sim 10^{-13}$ с, $m^* \sim 10^{-27}$ г, получим

$$p > \frac{10^{-27}}{10^{-26} \cdot 1,26 \cdot 2,3 \cdot 10^{-18}} \approx 3,5 \cdot 10^{16} \text{ см}^{-3}.$$

При меньших значениях концентрации частиц критическая частота отсутствует. Для металлов $\omega_{кр} \sim \sqrt{\frac{4\pi e^2 p}{m^*}} \approx 5 \cdot 10^{15} \text{ с}^{-1}$, т. е. критическая частота $\omega_{кр}$ лежит в ультрафиолетовой области. Если $\omega > \omega_{кр}$, то α мало. Многие металлы действительно обладают малым поглощением в ультрафиолетовой области спектра. При $\omega < \omega_{кр}$ металлы обладают очень большим поглощением. Для полупроводников $\omega_{кр}$ может не существовать и область сильного «металлического» поглощения в них может отсутствовать. Нормальное же поглощение определяется обычным соотношением (75.17) и оно пропорционально λ^2 .

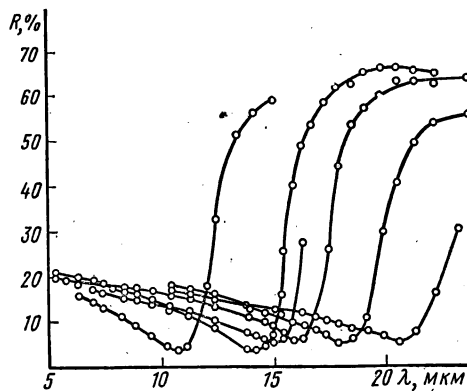


Рис. 110. Плазменное отражение в селе- ниде ртути

Поскольку коэффициент отражения определяется значением коэффициентов преломления и поглощения согласно (73.4), изменение этих величин в области критических частот приводит к резкому изменению коэффициента отражения. По мере приближения к критической частоте со стороны больших значений происходит уменьшение коэффициента отражения практически до нуля, после чего наблюдается его резкий рост (рис. 110). Это явление носит назва-

ние плазменного отражения, оно позволяет определять эффективную массу свободных носителей заряда. Для наблюдения плазменного отражения необходима большая концентрация свободных носителей заряда.

Перейдем к рассмотрению поглощения энергии излучения свободными носителями заряда в полупроводнике, на который наложено постоянное магнитное поле с индукцией \mathbf{B} . Уравнение движения носителей заряда с учетом действия соударений, электрического и магнитного полей имеет вид

$$\frac{d\mathbf{v}}{dt} = \frac{e\mathbf{E}_0}{m^*} e^{i\omega t} + \frac{e}{m^*} [\mathbf{v}\mathbf{B}] - \frac{\mathbf{v}}{\tau}. \quad (75.21)$$

В стационарном состоянии решение имеет вид (75.6), однако уравнение для определения амплитуды колебаний отличается от (75.7):

$$i\omega \mathbf{v}_0 = \frac{e\mathbf{E}_0}{m^*} + \frac{e}{m^*} [\mathbf{v}_0\mathbf{B}] - \frac{\mathbf{v}_0}{\tau}. \quad (75.22)$$

Перепишем уравнение (75.22) следующим образом:

$$\mathbf{v}_0 = \frac{\mu\mathbf{E}_0}{1+i\omega\tau} + \frac{\mu}{1+i\omega\tau} [\mathbf{v}_0\mathbf{B}]. \quad (75.23)$$

Вводя для упрощения обозначения

$$\frac{\mu\mathbf{E}_0}{1+i\omega\tau} = \mathbf{A}; \quad \frac{\mu\mathbf{B}}{1+i\omega\tau} = \Phi, \quad (75.24)$$

запишем согласно (37.27) и (37.32) решение векторного уравнения (75.23):

$$\mathbf{v}_0 = \frac{\mathbf{A} + [\mathbf{A}\Phi] + \Phi[\mathbf{A}\Phi]}{1+\Phi^2}. \quad (75.25)$$

Ограничимся случаем поперечных полей, поскольку в случае продольных полей в веществах со скалярной эффективной массой явления протекают так же, как и в отсутствие магнитного поля; запишем

$$\mathbf{v}_0 = \frac{\mu\mathbf{E}_0}{(1+i\omega\tau) \left[1 + \frac{\mu^2 B^2}{(1+i\omega\tau)^2} \right]} + \frac{\mu^2 [\mathbf{E}_0\mathbf{B}]}{(1+i\omega\tau)^2 \left[1 + \frac{\mu^2 B^2}{(1+i\omega\tau)^2} \right]}. \quad (75.26)$$

Как видно из (75.26), скорость частиц имеет продольную v_0^{\parallel} и поперечную v_0^{\perp} составляющие относительно электрического поля. Поглощение энергии вызвано тем, что электрическое поле совершает работу по перемещению заряженных частиц. Из (75.26) видно, что поперечный компонент скорости не приводит к потере энергии, поэтому ограничимся рассмотрением продольного компонента скорости:

$$\begin{aligned} v_0 &= \mu E_0 \frac{1+i\omega\tau}{\left[(1+i\omega\tau)^2 + \frac{e^2\tau^2}{m^{*2}} B^2 \right]} = \\ &= \mu E_0 \frac{[1 + (\omega_c^2 + \omega^2)\tau^2] + i\omega\tau [(\omega_c^2 - \omega^2)\tau^2 - 1]}{[1 + (\omega_c^2 - \omega^2)\tau^2]^2 + 4\omega^2\tau^2}, \end{aligned} \quad (75.27)$$

$$\omega_c = \frac{eB}{m^*}$$

Так же, как и в случае (75.10), действительная часть (75.27) связана с током проводимости, мнимая часть связана с током поляризации. Из (75.27) имеем для проводимости

$$\sigma = e\mu\rho \frac{1 + (\omega_c^2 + \omega^2) \tau^2}{[1 + (\omega_c^2 - \omega^2) \tau^2]^2 + 4\omega^2\tau^2}. \quad (75.28)$$

Если $B = 0$, то $\omega_c = 0$ и выражение (75.28) переходит в (75.12), которое согласно (74.34) дает зависимость $\alpha(\omega)$ вида (74.65). Не останавливаясь на этом случае, рассмотрим зависимость $\alpha(\omega)$, вытекающую из (74.34) и (75.28), при $B \neq 0$. Обозначив

$$\alpha_0 = \frac{4\pi\sigma_0}{cn}, \quad (75.29)$$

получим

$$\alpha(\omega) = \alpha_0 \frac{1 + (\omega_c^2 + \omega^2) \tau^2}{[1 + (\omega_c^2 - \omega^2) \tau^2]^2 + 4\omega^2\tau^2}. \quad (75.30)$$

Исследуем выражение (75.30) при некоторых значениях входящих в него параметров.

1) $\omega_c^2\tau^2 \gg 1$. Выражение (75.30) можно упростить следующим образом:

для низких частот ($\omega \ll \omega_c$)

$$\alpha(\omega) \approx \alpha_0 \frac{\omega_c^2\tau^2}{\omega_c^4\tau^4} = \frac{\alpha_0}{\omega_c^2\tau^2} = \frac{\alpha_0}{\mu^2 B^2}; \quad (75.31)$$

для высоких частот ($\omega \gg \omega_c$)

$$\alpha(\omega) \approx \alpha_0 \frac{\omega^2\tau^2}{\omega^4\tau^4} = \frac{\alpha_0}{\omega^2\tau^2}. \quad (75.32)$$

Для частот $\omega \approx \omega_c$ коэффициент поглощения достигает максимума:

$$\alpha(\omega) \approx \frac{\alpha_0}{2}. \quad (75.33)$$

Мы видим, что поглощение «света» имеет резонансный характер. При этом резонанс проявляется тем резче, чем лучше выполняется условие $\omega_c^2\tau^2 \gg 1$. Смысл этого неравенства очень простой: за время свободного пробега частица совершает несколько оборотов, причем чем больше оборотов делает частица, тем резче, острее проявляется резонанс. Явление резонансного поглощения излучения при $\omega \approx \omega_c = eB/m^*c$ называют циклотронным резонансом, ω_c называют циклотронной частотой.

2) $\omega_c^2\tau^2 \ll 1$ — за время свободного пробега частица совершает лишь часть оборота.

В области высоких частот ($\omega^2\tau^2 \gg 1$)

$$\alpha(\omega) = \alpha_0 \frac{\omega^2\tau^2}{\omega^4\tau^4} = \frac{\alpha_0}{\omega^2\tau^2}; \quad (75.34)$$

в области низких частот ($\omega^2\tau^2 \ll 1$)

$$\alpha(\omega) \approx \alpha_0. \quad (75.35)$$

Таким образом, в случае $\omega_c^2 \tau^2 \ll 1$ резонанс не наблюдается. На рис. 111 представлена зависимость $\alpha(\omega)$ при трех различных значениях $\omega_c \tau = \mu B$. При экспериментальном изучении циклотронного

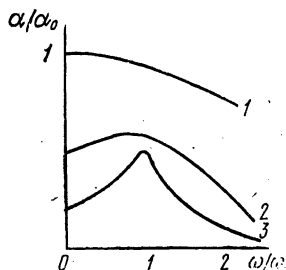


Рис. 111. Условия наблюдения циклотронного резонанса: 1 — $\omega_c \tau = 0,2$; 2 — $\omega_c \tau = 1$; 3 — $\omega_c \tau = 2$

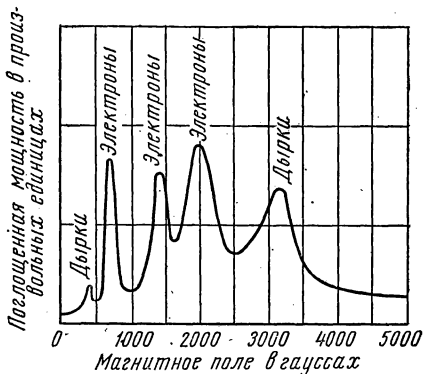


Рис. 112. Пики поглощения при циклотронном резонансе в германии

резонанса выбирается определенная частота излучения, и резонанс достигается изменением магнитного поля, т. е. ω_c . Определив ω_c , можно вычислить эффективную массу m^* . Циклотронный резонанс

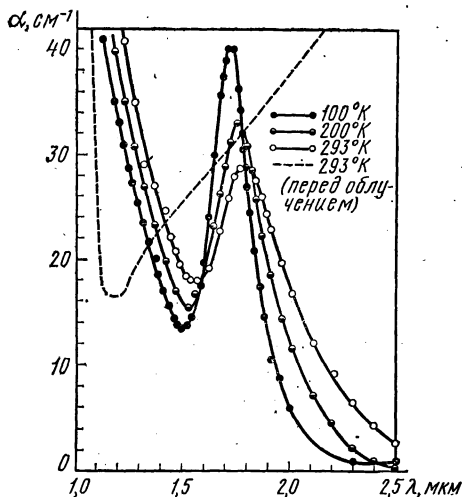


Рис. 113. Влияние облучения нейтронами р-кремния на поглощение света свободными носителями заряда

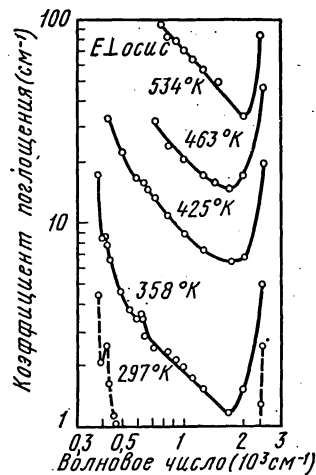


Рис. 114. Влияние температуры на поглощение света свободными носителями заряда в монокристаллах теллура

позволяет получать сведения об эффективной массе и в том случае, когда она является тензорной величиной. На рис. 112 приведены пики поглощения в относительных единицах в германии.

Циклотронный резонанс может быть объяснен квантованием энергии электронов и дырок в магнитном поле, при котором возникают уровни Ландау. Расстояние между уровнями равно

$$\hbar\omega_c = \frac{\hbar e B}{m^*}. \quad (75.36)$$

Если энергия фотона $\hbar\omega$ равна расстоянию между уровнями

$$\hbar\omega = \hbar\omega_c, \quad (75.37)$$

то возможно поглощение фотонов. Но уравнение (75.37) и есть условие циклотронного резонанса.

Рассмотрим некоторые экспериментальные данные в качестве иллюстрации расчетов, изложенных в § 74, 75. На рис. 113 представлена зависимость коэффициента поглощения кремния p -типа в области от 1 до 2,5 мкм. Резкое уменьшение α в области $\lambda \approx 1,1$ мкм связано с уменьшением собственного поглощения, которое переходит в поглощение свободными носителями заряда (правая часть пунктирной кривой). Исследованный образец затем был подвергнут облучению нейтронами, проводимость его упала, и мы видим, что характер кривой поглощения сильно изменился.

На рис. 114 приведены кривые поглощения монокристалла теллура в двойной логарифмической шкале. По оси абсцисс отложен логарифм волнового числа $\tilde{\nu} = \lambda^{-1}$ (в 10^3 см^{-1}). Отчетливо видно, что логарифм коэффициента поглощения в области от 300 см^{-1} до 2000 см^{-1} линейно зависит от логарифма волнового числа. Кроме того, коэффициент поглощения растет с ростом температуры, которая обуславливает возращание концентрации носителей заряда.

На рис. 115 представлено поглощение, обусловленное дырками, инжектированными в германий n -типа. Поглощение свободными носителями заряда может быть использовано для изучения явлений, связанных с неравновесными процессами на контактах.

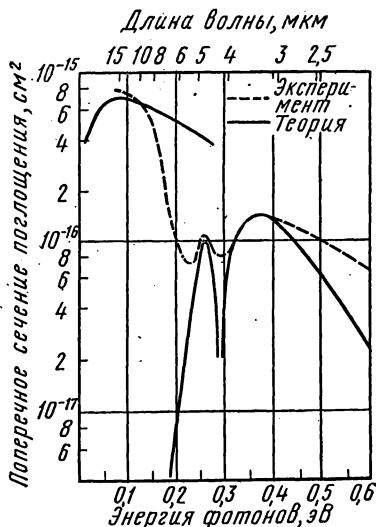


Рис. 115. Поглощение света дырками, инжектированными в германий n -типа

Резюме § 74, 75

1. Поглощение энергии излучения свободными носителями заряда обусловлено тем, что электрическое поле волны вызывает омический ток, на поддержание которого расходуется энергия волны, переходя

дующая в ленц-джоулево тепло. Коэффициент поглощения пропорционален проводимости

$$\alpha = \frac{4\pi}{cn} \sigma. \quad (75.1p)$$

2. Зависимость коэффициента поглощения от частоты может быть найдена или на основе теории дисперсии, или с учетом зависимости проводимости от частоты. И в том и в другом случае зависимость α от частоты имеет вид

$$\alpha = \frac{\alpha_0 \omega_{\text{ГР}}^3}{\omega_{\text{ГР}}^2 + \omega^2}, \quad (75.2p)$$

т. е. в области $\omega^2 \ll \omega_{\text{ГР}}^2$ $\alpha = \alpha_0$, а при $\omega^2 \gg \omega_{\text{ГР}}^2$ α уменьшается по закону: $\alpha \sim \omega^{-2}$. Физический смысл $\omega_{\text{ГР}}$ в разных теориях разный, однако во всех случаях $\omega_{\text{ГР}}$ зависит от времени релаксации.

3. При наложении магнитного поля на полупроводник его сопротивление возрастает, поэтому поглощение свободными носителями заряда уменьшается. Выражение для $\alpha(\omega)$ имеет вид

$$\alpha(\omega) = \frac{4\pi e \mu p}{cn} \frac{1 + (\omega_c^2 + \omega^2) \tau^2}{[1 + (\omega_c^2 - \omega^2) \tau^2]^2 + 4\omega^2 \tau^2}, \quad (75.3p)$$

где p — концентрация носителей заряда, n — коэффициент преломления.

4. При совпадении частоты излучения ω с циклотронной частотой ω_c наблюдается возрастание коэффициента поглощения по сравнению с частотами $\omega \neq \omega_c$. Это явление называется циклотронным, или диамагнитным, резонансом. Он наблюдается достаточно отчетливо в том случае, когда частица за время релаксации совершает не менее одного оборота. Чем лучше выполняется неравенство $\omega_c^2 \tau^2 \gg 1$, тем отчетливее наблюдается резонанс. С квантовой точки зрения циклотронный резонанс есть результат квантования энергии частиц в магнитном поле. Фотон поглощается в том случае, когда расстояние между уровнями Ландау равно энергии фотона:

$$\hbar\omega = \hbar\omega_c; \quad \omega = \omega_c. \quad (75.4p)$$

Циклотронный резонанс позволяет определить m^* ; для его наблюдения необходимы низкие, обычно гелиевые, температуры.

5. При поглощении света свободными носителями заряда происходит обмен импульсом между носителями заряда и решеткой, вследствие чего коэффициент поглощения зависит от механизма рассеяния. Для рассеяния на акустических и оптических колебаниях решетки и ионах примеси коэффициент поглощения пропорционален $\lambda^{3/2}$, $\lambda^{5/2}$ и $\lambda^{7/2}$ в полупроводниках со стандартными зонами энергии и невырожденном газе носителей заряда. Вырождение или отклонение закона дисперсии от квадратичного приводит к увеличению показателя степени.

§ 76. СОБСТВЕННОЕ ПОГЛОЩЕНИЕ СВЕТА

Для описания собственного поглощения света воспользуемся теорией возмущений. В качестве возмущения мы возьмем энергию электрона в поле световой волны, которую можно характеризовать напряженностью \mathbf{E} электрического поля и индукцией \mathbf{B} магнитного поля. Вместо двух величин удобно ввести векторный потенциал $\mathbf{A}(\mathbf{r}, t)$, который позволяет выразить через него \mathbf{E} и \mathbf{B} (в системе Гаусса):

$$\mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}, \quad (76.1)$$

$$\mathbf{B} = \text{rot } \mathbf{A}. \quad (76.2)$$

Гамильтониан электрона в кристалле, находящемся в поле излучения, в приближении эффективной массы согласно (23.29) и (23.11) имеет вид

$$\hat{H} = \frac{(\mathbf{p} - \frac{e}{c} \mathbf{A})^2}{2m^*} = -\frac{\hbar^2}{2m^*} \Delta + \frac{i\hbar e}{m^* c} (\mathbf{A} \nabla) + \frac{i\hbar e}{2m^* c} \text{div } \mathbf{A} + \frac{e^2 \mathbf{A}^2}{2m^* c^2}. \quad (76.3)$$

Если ограничиться слабыми световыми потоками, которые можно получить от обычных источников света, то последний член, пропорциональный \mathbf{A}^2 , по сравнению с линейным можно отбросить¹.

Учитывая, что вектор-потенциал должен удовлетворять условию Лоренца

$$\text{div } \mathbf{A} = 0, \quad (76.4)$$

выделим в (76.3) оператор возмущения

$$\hat{W} = \frac{i\hbar e}{m^* c} (\mathbf{A} \nabla) = -\frac{e}{c} \left(\mathbf{A}, -\frac{i\hbar \nabla}{m^*} \right) = -\frac{1}{c} (\mathbf{A} \mathbf{j}), \quad (76.5)$$

где \mathbf{j} — оператор плотности тока:

$$\hat{\mathbf{j}} = e \hat{\mathbf{v}} = -\frac{ie\hbar}{m^*} \nabla. \quad (76.6)$$

Таким образом, оператор Гамильтона для электронов полупроводника, находящегося в световом поле, имеет вид

$$\hat{H} = \hat{H}^0 + \hat{W} = -\frac{\hbar^2}{2m^*} \Delta + \hat{W}. \quad (76.7)$$

Как было показано в § 55, возмущение индуцирует переходы из одного состояния в другое. Рассмотрим некоторое состояние

¹ Учет \mathbf{A}^2 необходим при освещении полупроводников мощными световыми потоками, создаваемыми квантовыми генераторами, что приводит к описанию двухфотонных переходов.

в валентной зоне с энергией $E_1(\kappa_1)$, ему соответствует волновая функция $\psi_{1\kappa_1}(\mathbf{r}, t)$:

$$\psi_{1\kappa_1}(\mathbf{r}, t) = \frac{1}{L^3} e^{i(\kappa_1 \mathbf{r})} e^{-i \frac{E_1(\kappa_1)t}{\hbar}}. \quad (76.8)$$

Возьмем некоторое состояние в зоне проводимости с энергией $E_2(\kappa_2)$ и волновой функцией $\psi_{2\kappa_2}(\mathbf{r}, t)$:

$$\psi_{2\kappa_2}(\mathbf{r}, t) = \frac{1}{L^3} e^{i(\kappa_2 \mathbf{r})} e^{-i \frac{E_2(\kappa_2)t}{\hbar}}. \quad (76.9)$$

Чтобы вычислить вероятность перехода, необходимо задать возмущение $\hat{W} = \hat{W}(\mathbf{r}, t)$. Вектор-потенциал возьмем в виде плоской волны:

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 e^{i[\omega t - (\mathbf{g}\mathbf{r})]}. \quad (76.10)$$

Для оператора возмущения \mathbf{W} согласно (76.5) и (76.10) имеем

$$\hat{W} = \frac{ie\hbar}{cm^*} (\mathbf{A}\nabla) = \frac{ie\hbar}{cm^*} e^{i[\omega t - (\mathbf{g}\mathbf{r})]} (\mathbf{A}_0\nabla). \quad (76.11)$$

Вычислим матричный элемент \hat{W} с помощью волновых функций $\psi_{1\kappa_1}(\mathbf{r}, t)$ и $\psi_{2\kappa_2}(\mathbf{r}, t)$:

$$\begin{aligned} \int_{(L^3)} \psi_{1\kappa_1}^*(\mathbf{r}, t) \hat{W}(\mathbf{r}, t) \psi_{2\kappa_2}(\mathbf{r}, t) d\tau &= \frac{1}{L^3} e^{i[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega]t} \times \\ &\times \frac{ie\hbar}{cm^*} \int_{(L^3)} e^{-i(\kappa_1 \mathbf{r})} e^{-i(\mathbf{g}\mathbf{r})} (\mathbf{A}_0\nabla) e^{i(\kappa_2 \mathbf{r})} d\tau = \\ &= e^{i[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega]t} (-1) \frac{e\hbar}{cm^*} (\mathbf{A}\kappa_2) \frac{1}{L^3} \int e^{-i(\kappa_1 + \mathbf{g} - \kappa_2, \mathbf{r})} d\tau = \\ &= -\frac{e\hbar}{cm^*} (\mathbf{A}_0\kappa_2) e^{i[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega]t} \delta_{\kappa_1 + \mathbf{g}, \kappa_2} \end{aligned} \quad (76.12)$$

Выражение (76.12) показывает, что матричный элемент отличен от нуля только в том случае, когда

$$\kappa_1 + \mathbf{g} = \kappa_2, \quad (76.13)$$

или

$$\mathbf{P}_2 = \mathbf{P}_1 + \hbar\mathbf{g}, \quad (76.14)$$

т. е. при поглощении света должен выполняться закон сохранения квазиимпульса: *квазиимпульс конечного состояния равен векторной сумме квазиимпульса начального состояния и импульса фотона.*

Если $\kappa_1 = 0$, то $\kappa_2 = \mathbf{g}$. Но такие переходы невозможны, поскольку в этом случае $(\mathbf{A}_0\kappa_2) = (\mathbf{A}_0\mathbf{g}) = 0$ (условие поперечности световых волн).

Квазиимпульс электронов с тепловой энергией имеет величину $P = \sqrt{2m^*kT}$. При $T = 300$ К и $m^* = 10^{-27}$ г имеем $P \approx 10^{-20}$ г·см/с и $\kappa \approx 10^7$ см⁻¹. Для света с длиной волны $\lambda \approx 1$ мкм $g = \frac{2\pi}{\lambda} \approx 6 \cdot 10^4$ см⁻¹, что значительно меньше κ для тепловых электронов. В таком случае, пренебрегая величиной $\hbar g$ по сравнению с P_1 , получим

$$P_2 = P_1; \quad \kappa_2 = \kappa_1. \quad (76.15)$$

Переходы из валентной зоны в зону проводимости в соответствии с правилами отбора (76.15) — с сохранением волнового вектора электрона — носят название прямых, или вертикальных. Электрон, поглотив фотон, переходит из некоторой точки зоны Бриллюэна валентной зоны в эквивалентную точку зоны Бриллюэна зоны проводимости.

Как видно из (76.12) в матричный элемент оператора возмущения входит член $e^{\frac{i}{\hbar}[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega]t}$. При вычислении вероятности перехода в единицу времени этот множитель дает δ -функцию (см. § 55):

$$\delta[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega],$$

которая отражает выполнение закона сохранения энергии при поглощении света:

$$E_2(\kappa_2) = E_1(\kappa_1) + \hbar\omega. \quad (76.16)$$

В соответствии с (55.23) вероятность перехода электрона из единичного объема κ_1 -пространства в единичный объем κ_2 -пространства в единицу времени равна

$$\omega(\kappa_1, \kappa_2) = \frac{2\pi}{\hbar} \frac{e^2 \hbar^2}{c^2 m^{*2}} (A_0 \kappa_2)^2 \delta[E_1(\kappa_1) - E_2(\kappa_2) + \hbar\omega] \delta_{\kappa_1 + g, \kappa_2}. \quad (76.17)$$

Предполагая в дальнейшем, что законы сохранения энергии и квазиимпульса выполняются, мы можем опустить δ -функции. Выразим вероятность перехода электрона (76.17) через число фотонов, проходящих через полупроводник. Для этого учтем, что средняя плотность световой энергии равна $\frac{\epsilon E_0^2}{8\pi}$ (мгновенная плотность равна $\frac{\epsilon E^2}{8\pi} + \frac{\mu H^2}{8\pi} = \frac{\epsilon E^2}{4\pi}$), поток энергии равен $c \frac{\epsilon E_0^2}{8\pi}$, где n — коэффициент преломления вещества и c/n — скорость света в нем. Поток фотонов q найдем, если поток энергии разделим на энергию одного кванта:

$$q = \frac{c\epsilon E_0^2}{8\pi \hbar \omega}. \quad (76.18)$$

Выразим вероятность перехода электрона через поток фотонов, для этого выразим A_0^2 через q . Найдем связь между E_0 и A_0 :

$$E = -\frac{1}{c} \frac{\partial A}{\partial t} = -\frac{i\omega}{c} A_0 e^{i[\omega t - (gr)]} = \frac{\omega A_0}{c} e^{i\left[\omega t - (gr) - \frac{\pi}{2}\right]} \quad (76.19)$$

и

$$E_0 = \frac{\omega}{c} A_0 = g A_0. \quad (76.20)$$

Из (76.18) и (76.20) следует

$$q = \frac{ceE_0^2}{8\pi\hbar\omega} = \frac{eA_0^2\omega}{8\pi\hbar c}, \quad (76.21)$$

или

$$A_0^2 = \frac{8\pi\hbar c}{e\omega} q. \quad (76.22)$$

Для вероятности перехода (76.17) с учетом (76.22) имеем

$$\omega(\kappa_1, \kappa_2) = \frac{16\pi^2\hbar^2 e^2 \pi \kappa_3^2 \cos^2 \theta}{cem^* \omega} q \quad (\theta = \widehat{A_0, \kappa_2}). \quad (76.23)$$

Переход электрона из одного состояния в другое возможен только в результате поглощения фотонов, поэтому $\omega(\kappa_1, \kappa_2)$ есть вероятность поглощения фотонов. Так как она пропорциональна потоку фотонов q , то эффективное сечение поглощения однофотонного потока на одном электроне получим, разделив $\omega(\kappa_1, \kappa_2)$ на поток фотонов q :

$$\sigma_q = \frac{\omega}{q} = \frac{16\pi^2\hbar^2 e^2 \pi \kappa_3^2 \cos^2 \theta}{cem^* \omega}. \quad (76.24)$$

Оценим величину эффективного сечения поглощения однофотонного потока электроном, предположив, что $m^* \approx 10^{-27}$ г, $\varepsilon \approx 16$, $n \approx 4$, $\kappa_2 \approx 10^8$ см⁻¹, при $\omega \approx 10^{14}$ с⁻¹ и $\cos^2 \theta = 1/3$, получим $\sigma_q = 10^{-26}$ см². Это значение соответствует площади поперечного сечения электрона в классической физике. При вычислении σ_q положили $\cos^2 \theta = 1/3$, что соответствует среднему значению $\langle \cos^2 \theta \rangle$:

$$\langle \cos^2 \theta \rangle = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin \theta d\theta d\varphi = \frac{1}{3}. \quad (76.25)$$

Если рассматривать вероятность (76.23), отнесенную к одному фотону, который образует поток c/n , то

$$\omega_\Phi = \sigma_q \frac{c}{n} = \frac{16\pi^2\hbar^2 e^2 \pi \kappa_3^2 \cos^2 \theta}{em^* \omega}. \quad (76.26)$$

Если рассматривать поглощение одного фотона одним электроном, то (76.26) необходимо разделить на поток $\frac{\hbar\kappa_2}{m^*}$, образуемый электро-

ном; таким образом, вероятность поглощения одного фотона одним электроном

$$\omega_{сф} = \frac{\omega_{\phi}}{\hbar \kappa_2} = \frac{16\pi^2 \hbar c^2 \cos^2 \theta \kappa_2}{em^* \omega}. \quad (76.27)$$

Перейдем к нахождению коэффициента поглощения α . В элементе объема $d\tau_{\kappa_2}$ имеется $d\tau_{\kappa_2} (4\pi^3)^{-1} f_{p_2}(\kappa_2)$ свободных и $d\tau_{\kappa_2} (4\pi^3)^{-1} f_2(\kappa_2)$ занятых состояний. В элементе объема $d\tau_{\kappa_1}$ имеется $d\tau_{\kappa_1} (4\pi^3)^{-1} f_1(\kappa_1)$ занятых и $d\tau_{\kappa_1} (4\pi^3)^{-1} f_{p_1}(\kappa_1)$ свободных состояний. Поскольку вероятности прямых и обратных переходов равны, при вычислении коэффициента поглощения света необходимо учитывать как прямые, так и обратные переходы, вызванные светом. Спонтанными переходами (рекомбинационными переходами) будем при этом пренебрегать. За единицу времени будет поглощено фотонов

$$\delta q = \iint [f_1(\kappa_1) f_{p_2}(\kappa_2) \omega(\kappa_1, \kappa_2) - f_2(\kappa_2) f_{p_1}(\kappa_1) \omega(\kappa_2, \kappa_1)] \frac{d\tau_{\kappa_1}}{4\pi^3} \frac{d\tau_{\kappa_2}}{4\pi^3}. \quad (76.28)$$

Первый член в (76.28) определяет число поглощенных фотонов, второй — излученных. В обычных условиях уровни энергии заняты электронами в соответствии с равновесным распределением их по состояниям:

$$\left. \begin{aligned} f_1(\kappa_1) &= \frac{1}{e^{\frac{E_1 - F}{kT}} + 1} \approx 1; & f_2(E_2) &\approx 0; \\ f_{p_1}(\kappa_1) &= \frac{1}{e^{\frac{F - E_1}{kT}} + 1} \approx 0; & f_{p_2}(E_2) &\approx 1, \end{aligned} \right\} \quad (76.29)$$

т. е. валентная зона практически заполнена, а зона проводимости практически свободна, поэтому обратными переходами можно пренебречь. (Но если в полупроводнике создать инверсную заселенность уровней энергии, то (76.28) станет отрицательной величиной, и такой полупроводник будет усиливать излучение, а не поглощать.) Умножим (76.28) на $\hbar\omega$, получим количество энергии, поглощаемой в единице объема полупроводника в единицу времени.

Но согласно (73.6) коэффициент поглощения α представляет собой величину, определяющую количество энергии, поглощаемой в единичном объеме в единицу времени:

$$\alpha = \frac{\delta q \hbar \omega}{q \hbar \omega}. \quad (76.30)$$

Интегрируя (76.28) с учетом (76.23) и (76.17), получим

$$\alpha = \frac{\hbar^2 e^2 \pi}{\pi^4 \omega c e m^{*2}} \int_{V_{\kappa_2}} \cos^2 \theta \kappa_2^3 \delta(E_1(\kappa_2) - E_2(\kappa_2) + \hbar\omega) d\tau_{\kappa_2}. \quad (76.31)$$

Исключим зависимость α от E_1 , E_2 и κ_2^2 , учитывая, что $E_2 = E_1 + \hbar\omega$. Кроме того, необходимо иметь в виду, что величина эффективной массы m^* , входящая во все выражения (76.1—31), есть эффективная масса электрона в валентной зоне, или масса дырки. Выразим κ_2^2 через ω :

$$\frac{\hbar^2 \kappa_2^2}{2m_n^*} = E_2 - E_c = E_1 + \hbar\omega - E_c = E_v - \frac{\hbar^2 \kappa_1^2}{2m_p^*} + \hbar\omega - E_c. \quad (76.32)$$

Из (76.32) при $\kappa_2 = \kappa_1$ имеем

$$\begin{aligned} & \delta(E_1(\kappa_2) - E_2(\kappa_2) + \hbar\omega) = \\ & = \delta\left(E_v - \frac{\hbar^2 \kappa_2^2}{2m_p^*} - E_c - \frac{\hbar^2 \kappa_2^2}{2m_n^*} + \hbar\omega\right) = \delta\left(\hbar\omega - \Delta E_0 - \frac{\hbar^2 \kappa_2^2}{2m_{np}^*}\right), \end{aligned} \quad (76.33)$$

где

$$m_{np}^* = \frac{m_n^* m_p^*}{m_n^* + m_p^*} \quad (76.34)$$

— приведенная эффективная масса электрона и дырки. Для коэффициента поглощения α с учетом (76.33) можем записать

$$\alpha = \frac{\hbar^2 e^2 \eta}{\pi^4 c \omega m_p^{*2}} \frac{4\pi}{3} \int_0^\infty \kappa_2^4 \delta\left(\hbar\omega - \Delta E_0 - \frac{\hbar^2 \kappa_2^2}{2m_{np}^*}\right) d\kappa_2. \quad (76.35)$$

Множитель $4\pi/3$ возникает вследствие интегрирования по углам, при этом полярный угол отсчитывается от направления вектор — потенциала. Для вычисления интеграла в (76.35) положим

$$\frac{\hbar^2 \kappa_2^2}{2m_{np}^*} = x; \quad \kappa_2 = \left(\frac{2m_{np}^*}{\hbar^2}\right)^{1/2} x^{1/2}, \quad (76.36)$$

откуда

$$\kappa_2^4 d\kappa_2 = \frac{1}{2} \left(\frac{2m_{np}^*}{\hbar^2}\right)^{5/2} x^{3/2} dx, \quad (76.37)$$

после чего интеграл легко вычисляется

$$\begin{aligned} & \int_0^\infty \frac{1}{2} \left(\frac{2m_{np}^*}{\hbar^2}\right)^{5/2} x^{3/2} \delta(\hbar\omega - \Delta E_0 - x) dx = \\ & = \frac{1}{2} \left(\frac{2m_{np}^*}{\hbar^2}\right)^{5/2} (\hbar\omega - \Delta E_0)^{3/2}. \end{aligned} \quad (76.38)$$

Подставляя (76.38) в (76.35), получим

$$\begin{aligned} \alpha & = \frac{\hbar^2 e^2 \eta}{\pi^4 c \omega m_p^{*2}} \frac{4\pi}{3} \cdot \frac{1}{2} \left(\frac{2m_{np}^*}{\hbar^2}\right)^{5/2} (\hbar\omega - \Delta E_0)^{3/2} = \\ & = \frac{2}{3\pi^3} \frac{e^2}{c \hbar^3 \omega} \frac{(2m_{np}^*)^{5/2}}{m_p^{*2} \eta} (\hbar\omega - \Delta E_0)^{3/2} = A (\hbar\omega - \Delta E_0)^{3/2}; \\ A & = \frac{2}{3\pi^3} \frac{e^2}{c \hbar^3 \omega} \frac{(2m_{np}^*)^{5/2}}{m_p^{*2} \eta}. \end{aligned} \quad (76.39)$$

Выражение (76.39) показывает, что при вертикальных переходах в области малых $(\hbar\omega - \Delta E_0)$ коэффициент поглощения пропорционален $(\hbar\omega - \Delta E_0)^{3/2}$. Если $\hbar\omega < \Delta E_0$, то $\alpha = 0$ — собственное поглощение имеет резкую границу со стороны малых частот. Граница собственного поглощения определяется (оптической) шириной запрещенной зоны для вертикальных переходов:

$$\omega_{\text{гp}} = \frac{\Delta E_0}{\hbar}; \quad \lambda_{\text{гp}} = \frac{\hbar c}{\Delta E_0}. \quad (76.40)$$

Если выразить ширину запрещенной зоны в эВ, то границу собственного поглощения можно вычислить из соотношения:

$$\lambda_{\text{гp}}, \text{ мкм} = \frac{1,24}{\Delta E_0 (\text{эВ})}. \quad (76.41)$$

Как мы видим из (76.39), γ края собственного поглощения α определяется разностью между энергией фотона и шириной запрещенной зоны в степени $3/2$.

Выражение (76.39) справедливо только при переходе в полупроводниках со сферическими поверхностями энергии и экстремумами, лежащими в одной точке, например в центре зоны Бриллюэна.

Рассмотрим теперь периодическое поле кристалла. С учетом условия Лоренца (76.4) гамильтониан электрона может быть записан в виде:

$$\hat{H} = -\frac{\hbar^2}{2m_0} \Delta + U(\mathbf{r}) + \frac{i\hbar e}{m_0 c} (\mathbf{A}\nabla) = \hat{H}_0 + \hat{W}, \quad (76.42)$$

где для оператора возмущения имеем:

$$\hat{W} = \frac{i\hbar e}{m_0 c} (\mathbf{A}\nabla) = -\frac{eA_0 e^{i[\omega t - (\mathbf{g}\mathbf{r})]}}{m_0 c} (\mathbf{a}^0 \hat{\mathbf{p}}). \quad (76.43)$$

В этом выражении для возмущения используется масса свободного электрона m_0 вместо эффективной массы. Кроме того, для удобства дальнейших вычислений введен единичный вектор \mathbf{a}^0 , определяющий поляризацию волны. Вычисления матричных элементов возмущения проводятся с использованием функций Блоха $\psi_{n\mathbf{k}} = e^{i(\mathbf{k}\mathbf{r})} \varphi_{n\mathbf{k}}(\mathbf{r})$, у которых указаны два индекса: положение точки в зоне Бриллюэна \mathbf{k} и номер зоны n . Временная зависимость волны Блоха аналогична временной зависимости волны де-Бройля, поэтому при вычислении вероятности перехода снова получим δ -функцию, из которой вытекает закон сохранения энергии при переходе электрона из валентной зоны в зону проводимости. Опуская временную зависимость, получим для матричного элемента оператора возмущения

$$\begin{aligned} W_{n\mathbf{k}_1, n'\mathbf{k}_2} &= \int e^{-i(\mathbf{k}_1\mathbf{r})} \varphi_{n\mathbf{k}_1}^* \frac{ie\hbar A_0}{m_0 c} e^{-i(\mathbf{g}\mathbf{r})} (\mathbf{a}^0 \nabla) e^{i(\mathbf{k}_2\mathbf{r})} \varphi_{n'\mathbf{k}_2} d\tau = \\ &= -\frac{eA_0}{m_0 c} \int e^{-i(\mathbf{k}_1\mathbf{r})} \varphi_{n\mathbf{k}_1}^* e^{-i(\mathbf{g}\mathbf{r})} (\mathbf{a}^0 \hat{\mathbf{p}}) e^{i(\mathbf{k}_2\mathbf{r})} \varphi_{n'\mathbf{k}_2} d\tau. \end{aligned} \quad (76.44)$$

Учитывая, что $\nabla\psi_{n'\kappa_2} = i\kappa_2\psi_{n'\kappa_2} + e^{i(\kappa_2 r)}\nabla\varphi_{n\kappa_2}$, получим

$$W_{n\kappa_1, n'\kappa_2} = \frac{i\hbar e A_0}{m_0 c} \int e^{-i(\kappa_1 + \mathbf{g} - \kappa_2, \mathbf{r})} \varphi_{n\kappa_1}^* (\mathbf{a}^0, i\kappa_2) \varphi_{n'\kappa_2} d\tau + \\ + \frac{i\hbar e A_0}{m_0 c} \int e^{-i(\kappa_1 + \mathbf{g} - \kappa_2, \mathbf{r})} \varphi_{n\kappa_1}^* (\mathbf{a}^0 \nabla) \varphi_{n'\kappa_2} d\tau. \quad (76.45)$$

Поскольку $|\mathbf{g}| \ll (\kappa_1, \kappa_2)$, первый член дает

$$\kappa_1 = \kappa_2 \text{ или } \frac{e\hbar (A_0 \kappa_2)}{m_1 c}, \quad (76.46)$$

т. е. то же, что мы имели при использовании волн де-Бройля. Условие $\kappa_1 = \kappa_2$ отражает закон сохранения квазиимпульса, поэтому будем считать, что оно справедливо и для второго члена в (76.45). В таком случае, объединив оба члена в (76.45), введем обозначение

$$P_{nn'}(\kappa_2) = \int \varphi_{n\kappa_2}^* (\mathbf{a}^0, -i\hbar\nabla + \hbar\kappa_2) \varphi_{n'\kappa_2} d\tau, \quad (76.47)$$

что позволяет записать для матричного элемента оператора возмущения

$$W_{n\kappa_1, n'\kappa_2} = W_{nn'}(\kappa_2) = -\frac{eA_0}{m_0 c} P_{nn'}(\kappa_2) \quad (76.48)$$

и для вероятности перехода

$$\omega(\kappa_1, \kappa_2) = \frac{2\pi}{\hbar} \frac{e^2 A_0^2}{m_0^2 c^2} |P_{nn'}(\kappa_2)|^2 \delta_{\kappa_1 \kappa_2} \delta[E(\kappa_1) - E(\kappa_2) + \hbar\omega]. \quad (76.49)$$

Предполагая, что экстремумы энергии находятся в точке $\mathbf{k} = 0$, разложим $P_{nn'}(\kappa_2)$ в ряд

$$P_{nn'}(\kappa_2) = P_{nn'}(0) + \frac{dP_{nn'}(0)}{d\kappa_2} \kappa_2 + \dots \quad (76.50)$$

В приближении эффективной массы первый член равен нулю, а второму члену согласно (76.17) * соответствует $(\mathbf{a}^0, \hbar\kappa_2) \frac{m_0}{m_p^*}$.

Учитывая (76.30), (76.28) и (76.22), получим

$$\alpha = \frac{\delta q}{q} = \frac{1}{q} \iint \omega(\kappa_1, \kappa_2) \frac{d\tau_{\kappa_1}}{4\pi^3} \frac{d\tau_{\kappa_2}}{4\pi^3} = \\ = \frac{8\pi\hbar n c}{\varepsilon \omega A_0^2} \frac{2\pi}{\hbar} \frac{e^2 A_0^2}{c^2 m_0^2} \iint |P_{nn'}(\kappa_2)|^2 \delta_{\kappa_1 \kappa_2} \delta[E(\kappa_1) - E(\kappa_2) + \hbar\omega] \frac{d\tau_{\kappa_1}}{4\pi^3} \frac{d\tau_{\kappa_2}}{4\pi^3} = \\ = \frac{e^2}{c m_0^2 \pi^4 n \omega} \int |P_{nn'}(\kappa_2)|^2 \delta[E_1(\kappa_2) - E_2(\kappa_2) + \hbar\omega] d\tau_{\kappa_2}. \quad (76.51)$$

Переходы называются разрешенными, если $P_{nn'}(0) \neq 0$. Для них имеем:

$$\alpha = \frac{e^2}{\pi^4 c m_0^2} \frac{|P_{nn'}(0)|^2}{n \omega} \int \delta[E_1(\kappa_2) - E_2(\kappa_2) + \hbar\omega] d\tau_{\kappa_2}. \quad (76.52)$$

Используя (76.33) и (76.36), вычислим интеграл в (76.52)

$$\int \delta [E_1(\mathbf{k}_2) - E_2(\mathbf{k}_2) + \hbar\omega] d\tau_{\mathbf{k}_2} = \int_0^{\infty} \delta \left[(\hbar\omega - \Delta E_0) - \frac{\hbar^2 \mathbf{k}_2^2}{2m_{\text{пр}}^*} \right] 4\pi k_2^2 dk_2 =$$

$$= 2\pi \left(\frac{2m_{\text{пр}}^*}{\hbar^2} \right)^{3/2} (\hbar\omega - \Delta E_0)^{1/2}, \quad (76.53)$$

что дает для α :

$$\alpha = \frac{e^2}{\pi^3 c m_0^3} \frac{|P_{nn'}(0)|^2}{\rho\omega} \frac{2\pi (2m_{\text{пр}}^*)^{3/2}}{\hbar^3} (\hbar\omega - \Delta E_0)^{1/2} =$$

$$= \frac{2e^2}{\pi^3 c m_0^3 \hbar^3} \frac{|P_{nn'}(0)|^2 (2m_{\text{пр}}^*)^{3/2}}{\rho\omega} (\hbar\omega - \Delta E_0)^{1/2} = B (\hbar\omega - \Delta E_0)^{1/2}, \quad (76.54)$$

где

$$B = \frac{2e^2}{\pi^3 c m_0^3 \hbar^3} \frac{|P_{nn'}(0)|^2 (2m_{\text{пр}}^*)^{3/2}}{\rho\omega}. \quad (76.55)$$

Если $P_{nn'}(0) = 0$, то переходы называют запрещенными. Вычисляя α на основании (76.51) с учетом линейного члена в (76.50), получим выражение типа (76.39) с несколько другим значением константы A . Учет последующих членов разложения (76.50) приведет к зависимости $\alpha \sim (\hbar\omega - \Delta E_0)^{m+1/2}$, где m — порядок производной матричного элемента по волновому вектору. В общем случае могут иметь место переходы различной «мультипольности», однако их вклад оказывается различным. Для проверки этого оценим отношение коэффициентов поглощения, соответствующих разрешенным (76.54) и запрещенным (76.39) переходам:

$$\frac{\alpha_s}{\alpha_p} = \frac{A}{B} (\hbar\omega - \Delta E_0) = \frac{(2m_{\text{пр}}^*)^3 m_0^3 (\hbar\omega - \Delta E_0)}{3m_p^{*2} |P_{nn'}(0)|^2}. \quad (76.56)$$

Мы видим, что при любом конечном значении $|P_{nn'}(0)|^2$ форма края собственного поглощения определяется разрешенными переходами. Однако с ростом энергии $(\hbar\omega - \Delta E_0)$ вклад запрещенных переходов возрастает.

Таким образом, прямые переходы должны давать зависимость α от $(\hbar\omega - \Delta E_0)$ в виде $\alpha = \alpha_0 (\hbar\omega - \Delta E_0)^r$, где r может принимать значения $1/2$ и $3/2$ при прямых разрешенных и запрещенных переходах. Граница поглощения $\omega_{\text{гр}}$ определяет оптическую ширину запрещенной зоны $\Delta E_0^g = \hbar\omega_{\text{гр}}$, которая превосходит минимальное расстояние между валентной зоной и зоной проводимости, определяющее термическое возбуждение электронов.

Поскольку существуют состояния, разделенные энергетическим промежутком меньше, чем ΔE_0^g , возникает вопрос, возможно ли поглощение фотонов с энергией меньше, чем $\hbar\omega_{\text{гр}}$. Очевидно, что при этом правила отбора (76.13, 14) должны нарушаться. Однако нарушение правил отбора $\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{g}$ не может означать нарушение закона сохранения квазиимпульса (импульса). Переход электрона

из состояния $\kappa_1 \approx 0$ в состояние $\kappa_2 \approx \kappa_0$ возможен, если изменение импульса электрона будет компенсироваться изменением импульса фонона.

Рассмотрим следующие процессы. Пусть электрон находится в состоянии $\kappa_1 = 0$. Переведем его в состояние $\kappa_2 = 0$, для этого ему необходимо сообщить энергию ΔE_0^0 . Но такой электрон оказывается «горячим» и в результате столкновения с решеткой электрон может отдать свою энергию решетке, породив фонон с энергией $E_2(0) - E_c$ и квазиимпульсом $\hbar K = -\hbar \kappa_0$, при этом электрон окажется в состоянии $E_2(\kappa_0)$. Его энергия изменилась на $E_2(\kappa_0) - E_1(0) = \Delta E_0^T < \Delta E_0^0$. Такой переход можно представить следующим образом: электрон переходит из состояния $E_1(0)$ в состояние $E_2(\kappa_0)$ в результате поглощения фотона с энергией ΔE_0^T и длинноволнового фонона с энергией $[\Delta E_0^0 - \Delta E_0^T]$, при этом электрон оказывается в состоянии $E_2(0)$. Испустив фонон с энергией $[\Delta E_0^0 - \Delta E_0^T]$ и волновым вектором $-\kappa_0$, электрон оказывается в состоянии $E_2(\kappa_0)$. Таким образом, электрон переходит из $E_1(0)$ в $E_2(\kappa_0)$, поглотив фотон с энергией $\hbar \omega = \Delta E_0^T$. Необходимая для переброса электрона энергия $\Delta E_0^0 - \Delta E_0^T$ сообщается электрону решеткой и решетке же она передается электроном. Переход электрона происходит через промежуточное состояние, в котором происходит преобразование длинноволнового фонона в коротковолновый. Другими словами, *переход электрона из зоны проводимости в валентную зону происходит за счет энергии фотона, изменение же импульса электрона компенсируется решеткой (фононом)*.

Разобранная выше схема не является единственно возможной. Действительно, электрон в состоянии $E_1(0)$ может поглотить фонон с энергией $(\Delta E_0^0 - \Delta E_0^T)$ и квазиимпульсом $\hbar \kappa_0$, в результате чего он окажется в некотором виртуальном состоянии; испустив длинноволновый фонон с энергией $\Delta E_0^0 - \Delta E_0^T$, электрон окажется в состоянии $E_2(\kappa_0)$. Электрон в состоянии $E_1(0)$ может испустить фонон с энергией $\Delta E_0^0 - \Delta E_0^T$ и с импульсом $(-\hbar \kappa_0)$ и быть при этом в некотором виртуальном состоянии, поглотив затем фонон с энергией ΔE_0^T , электрон окажется в состоянии $E_2(\kappa_0)$. Таким образом, переход электрона из состояния $E_1(\kappa_1)$ в состояние $E_2(\kappa_2)$ при $\kappa_1 \approx 0$ и $\kappa_2 \approx \kappa_0$ происходит через ряд виртуальных состояний. Чтобы получить зависимость α от частоты, необходимо учесть законы сохранения энергии и импульса

$$E_2(\kappa_2) = E_1(\kappa_1) + \hbar \omega \pm \hbar \omega_{\text{фон}}, \quad (76.57)$$

$$\kappa_2 = \kappa_1 \pm K_{\text{фон}}, \quad (76.58)$$

где $\omega_{\text{фон}}$, $K_{\text{фон}}$ — частота и волновой вектор поглощаемого (плюс) или испускаемого (минус) фонона. Граница поглощения должна определяться условием

$$\hbar \omega_{\text{гр}} = E_2(\kappa_0) - E_1(0) \pm \hbar \omega_{\text{фон}} = \Delta E_0^T \pm \hbar \omega_{\text{фон}}. \quad (76.59)$$

Таким образом, существует две границы собственного поглощения, оптическая ширина запрещенной зоны (минимальная) при не-

прямых неперпендикулярных переходах должна быть меньше термической ширины запрещенной зоны ΔE_0^T на энергию фонона $\hbar\omega_{\text{фон}}$.

Если предположить, что возмущение содержит теперь некоторую характеристику фононов, то вероятность перехода электрона будет определяться как матричным элементом возмущения со стороны электромагнитного поля, так матричным элементом возмущения со стороны решетки.

Предполагая, что матричный элемент возмущения со стороны решетки не зависит от частоты фонона, мы получим зависимость коэффициента поглощения от частоты в виде

$$\alpha(\omega) \sim (\hbar\omega - \Delta E_0^T \pm \hbar\omega_{\text{фон}})^2. \quad (76.60)$$

Если учесть, что число фононов зависит от их энергии и температуры, то выражение для α можно записать в виде

$$\alpha(\omega) \sim \left\{ \frac{(\hbar\omega - \Delta E_0^T + \hbar\omega_{\text{фон}})^2}{\frac{\hbar\omega_{\text{фон}}}{e \frac{kT}{-1}}} + \frac{(\hbar\omega - \Delta E_0^T - \hbar\omega_{\text{фон}})^2}{1 - e \frac{\hbar\omega_{\text{фон}}}{kT}} \right\}. \quad (76.61)$$

Первый член описывает процесс поглощения света с поглощением фононов, число которых пропорционально величине $\frac{1}{e \frac{\hbar\omega_{\text{фон}}}{kT} - 1}$,

второй член описывает процесс поглощения фотона с испусканием фонона, вероятность испускания фонона равна вероятности того, что данное колебательное состояние не возбуждено, т. е.

$$1 - \frac{1}{e \frac{\hbar\omega_{\text{фон}}}{kT} - 1} = \frac{1}{1 - e \frac{\hbar\omega_{\text{фон}}}{kT}}. \quad (76.62)$$

Для запрещенных непрямых переходов, показатель степени должен быть на единицу больше по сравнению с непрямими разрешенными переходами, т. е. вместо (76.60) следует записать

$$\alpha(\omega) \sim (\hbar\omega - \Delta E_0^T \pm \hbar\omega_{\text{фон}})^3. \quad (76.63)$$

В заключение этого параграфа сравним коэффициенты поглощения света в результате прямых и непрямых переходов. Прямой переход определяется вероятностью встречи двух частиц — электрона и фотона. При непрямых переходах должны встретиться три частицы — электрон, фотон и фонон. Но это означает, что непрямой переход является процессом менее вероятным, чем прямой, поэтому коэффициент поглощения света при прямых переходах должен достигать больших величин, чем для непрямых переходов.

На рис. 116 представлена схема прямых (а) и непрямых (б) переходов. Точки 1 и 2 представляют собой виртуальные состояния. Примерами прямых переходов могут служить переходы электронов при поглощении света в InSb, непрямые переходы имеют место, например, в германии и кремнии.

На рис. 117 представлены спектры поглощения германия и кремния. Граница собственного поглощения, определяющая оптическую

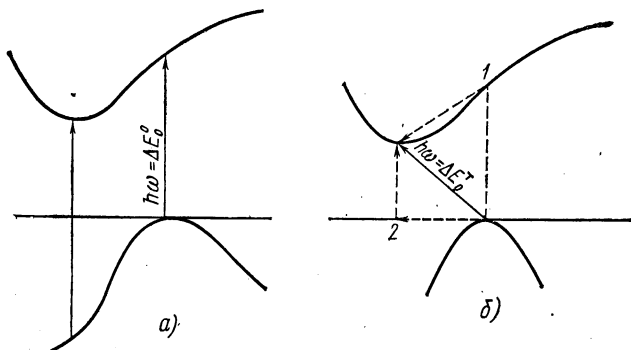


Рис. 116. Прямые (а) и непрямые (б) переходы при собственном поглощении света

ширину запрещенной зоны, находится при 0,66 и 1,09 эВ. Край поглощения обусловлен непрямими переходами. В области 0,8 эВ в германии наблюдается резкое возрастание коэффициента поглощения, оно объясняется началом прямых переходов, которые, как было

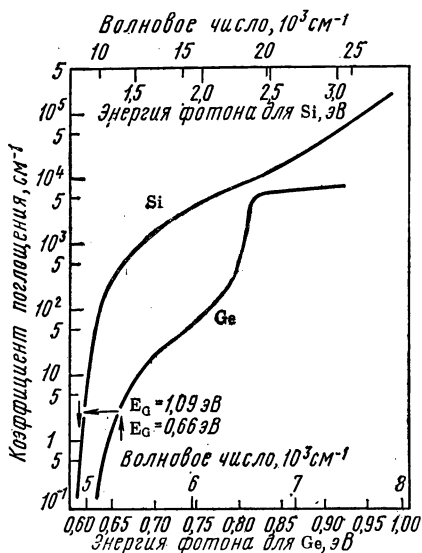


Рис. 117. Спектр поглощения германия и кремния

сказано выше, являются более вероятными по сравнению с непрямими переходами. В кремнии прямые переходы наблюдаются при $h\omega \geq 2,5$ эВ.

На рис. 118 представлена графическая зависимость собственного поглощения в $A^{III}BV$, а на рис. 110 — в некоторых соединениях $A^{IV}BV^{VI}$. Резкое возрастание коэффициента поглощения у края основной полосы связано с быстрым ростом числа частиц, способных поглощать фотоны с ростом их энергии. На рис. 119 видно, что за краем собственного поглощения α определяется поглощением свободными носителями заряда.

Исследуя форму кривой в области края собственного поглощения, можно в ряде случаев сделать заключение о структуре зон энергии. Собственное поглощение приводит к быстрому уменьшению интенсивности света, поскольку коэффициент собственного поглощения имеет величину $10^5 - 10^6 \text{ см}^{-1}$.

При построении теории поглощения света, приводящего к межзонным переходам электронов, мы не учитывали кулоновского взаимодействия между электронами и дырками, возникающими вследствие поглощения фотона. Благодаря кулоновскому притяжению образуется связанная система «электрон — дырка» — экситон, имеющая водородоподобную систему дискретных уровней энергии $E_n^3 = E_c - E_g/n^2$, расположенных ниже дна зоны проводимости, о чем сказано в § 27. Основное состояние экситона согласно (27.9) лежит ниже E_c на величину

$$E_1^3, \text{ эВ} = \frac{13,5}{\epsilon^2} \left(\frac{m_{np}^*}{m_0} \right). \quad (76.64)$$

При прямых межзонных переходах из состояния $k = k_n$ образуется дырка с волновым вектором $k_p = -k$. Поскольку экситон движется как единое целое, отсюда следует, что движение электрона и дырки коррелировано, их относительная

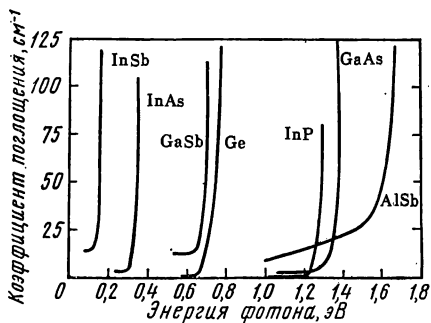


Рис. 118. Край собственного поглощения в соединениях $A^{III}B^V$ и германия

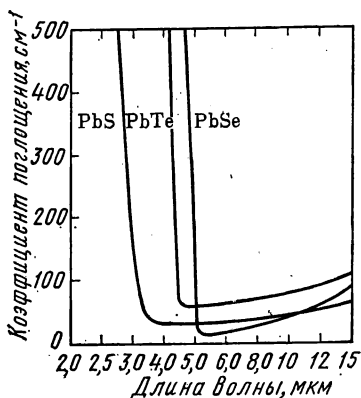


Рис. 119. Край собственного поглощения и поглощение свободными носителями заряда в соединениях $A^{IV}B^{VI}$

скорость равна нулю. Это возможно в том случае, если экситон образуется при переходах $k_n = k_p = 0$, т. е. переходы происходят в центре зоны Бриллюэна, или в точках экстремумов энергии в общем случае. Малость интервала возможных состояний, из которых разрешены переходы электронов с образованием экситонов, приводит к возникновению узких спектральных полос поглощения, примыкающих к фундаментальной полосе с длинноволновой стороны. Пример наблюдения экситонных полос поглощения приведен на рис. 120 для теллурида и селенида кадмия. Для теллурида кадмия наблюдается одна полоса, соответствующая переходу в основное состояние экситона. Для селенида наблюдается три полосы, соответствующие переходам из трех валентных зон, одна из которых опущена вниз вследствие спин-орбитального взаимодействия на величину порядка 0,4 эВ (полоса С). Две другие возникают при снятии вырождения полем

кристаллической решетки селенида кадмия, имеющего структуру вюрцита. Для теллурида кадмия двукратное вырождение не снижается, поскольку он имеет кубическую решетку типа сфалерита, а спин-орбитальное взаимодействие для теллурида кадмия составляет примерно 0,9 эВ, и полоса С попадает в область сильного межзонного поглощения и плохо разрешается.

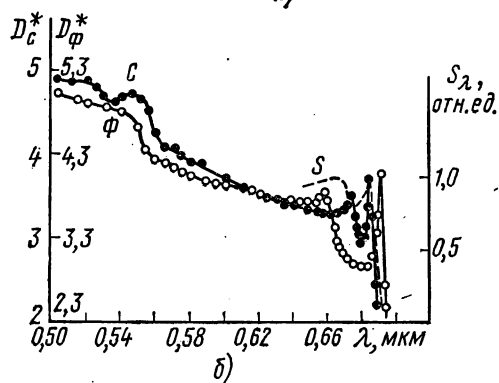
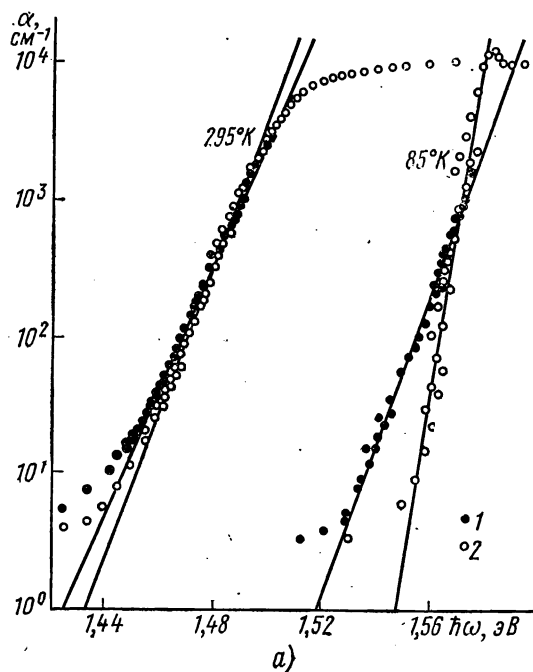


Рис. 120. Край собственного поглощения в теллуриде (а) и селениде (б) кадмия. Спектры поглощения селенида кадмия на подложках из сапфира (кривая С) и флюорита (кривая Φ) при 80 К и спектр фоточувствительности толстого слоя (~ 20 мкм) на флюорите (кривая S)

Возникновение узких дискретных полос поглощения — не единственное изменение, вносимое экситонными состояниями в спектр собственного поглощения. Расчет показывает, что существование экситонных состояний, т. е. кулоновского взаимодействия электрона и дырки, приводит к изменению формы полосы собственного поглощения.

Как известно, любая атомная система имеет бесконечный набор дискретных уровней энергии, соответствующих финитному движению электрона. При нормировке потенциальной энергии взаимодействия на ноль в бесконечность полная энергия является величиной отрицательной. Для положительных значений энергии движение электрона инфинитно, происходит ионизация атома, а спектр энергии свободного движения электрона непрерывен. Этот факт позволяет характеризовать собственную полосу поглощения как результат перехода электрона в состоянии, соответствующие непрерывному спектру энергии экситона. Наложение переходов

в дискретные и непрерывные области спектра энергии приводит к тому, что коэффициент поглощения при $\hbar\omega = \Delta E_0$ не равен нулю. Расчеты Эллиота, Декстера, Мак-Лина и др. дают следующую форму спектра при прямых разрешенных межзонных переходах:

$$\alpha = \frac{2\pi e^2}{m_0^2 c \omega \pi} \left(\frac{2m_{np}^*}{\hbar^2} \right)^{3/2} |P_{nn'}(0)|^2 (E_I^2)^{1/2} \frac{e^z}{\text{sh } z}, \quad (76.65)$$

где

$$z = \pi \sqrt{\frac{E_I^2}{\hbar\omega - \Delta E_0}}. \quad (76.66)$$

При $\hbar\omega \rightarrow \Delta E_0$ получим

$$\alpha(\Delta E_0) = \frac{4\pi e^2}{m_0^2 c \omega \pi} \left(\frac{2m_{np}^*}{\hbar^2} \right)^{3/2} |P_{nn'}(0)|^2 (E_I^2)^{1/2}, \quad (76.67)$$

т. е., чем больше энергия ионизации экситона E_I^2 , тем больше α при $\hbar\omega = \Delta E_0$. При $E_I^2 \rightarrow 0$ и $\alpha(\Delta E_0) \rightarrow 0$, а уравнение (76.65) дает зависимость вида $\alpha \sim (\hbar\omega - \Delta E_0)^{1/2}$.

Аналогичные поправки появляются и для прямых запрещенных переходов. Развита теория и непрямых экситонных переходов.

Как следует из модели экситонных переходов, должна существовать резкая длинноволновая граница фундаментальной полосы, тогда как в действительности длинноволновый край фундаментальной полосы поглощения более или менее затянут в сторону $\hbar\omega < \Delta E_0$.

На рис. 121 приведена зависимость вида $\alpha^{1/2}(\hbar\omega)$ для теллурида кадмия, построенная по данным рис. 120, а. На кривой зависимости $\alpha^{1/2}(\hbar\omega)$ отчетливо видны несколько прямоугольных участков, экстраполяция которых к $\alpha^{1/2} = 0$ дает энергии с расстоянием между ними 0,021 эВ,

что хорошо совпадает с энергией продольного оптического фонона. Это означает, что край поглощения формируется с участием фононов, взаимодействующих с экситонами. При непрямых межзонных переходах должны наблюдаться только два прямолинейных участка на зависимости $\alpha^{1/2}(\hbar\omega)$, разделенных при экстраполяции их к $\alpha^{1/2} = 0$ интервалом энергии, равным удвоенной энергии фонона. Теория экситон-фононного поглощения разработана Б. Сегаллом.

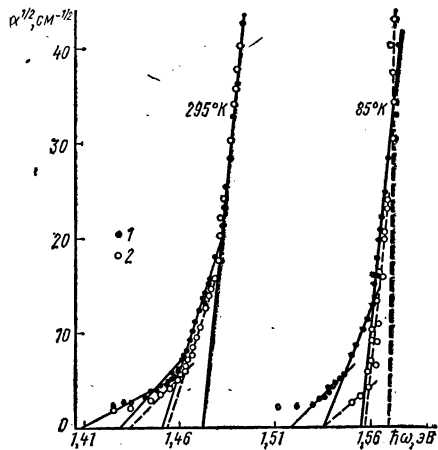


Рис. 121. Зависимость $\alpha^{1/2}$ от энергии фотона для теллурида кадмия (по данным рис. 120, а);

1—исходный образец; 2—отожженный при 500 °С

При экситон-фоонных межзонных переходах форма края собственного поглощения многих веществ может быть описана в широком интервале значений α уравнением Урбаха

$$\alpha(\hbar\omega) = \alpha_0 e^{-\frac{\sigma(\hbar\omega - E_0)}{kT}} \quad (76.68)$$

Параметр E_0 закона Урбаха может быть связан с энергией максимума экситонной полосы поглощения, при этом параметр α_0 совпадает со значением коэффициента поглощения в максимуме экситонной полосы. Параметр σ принимает значения в интервале $1 \div 3$. Если считать его зависящим от температуры, то можно описать и температурное изменение фундаментальной полосы.

Резюме § 76

1. Переходы электрона при поглощении света называются прямыми, или вертикальными, если выполняются правила отбора

$$k_2 = k_1 + g, \text{ или } k_2 \approx k_1. \quad (76.1p)$$

Переходы называют непрямыми, или невертикальными, если

$$k_2 \approx k_1 + K_{\text{фон}}. \quad (76.2p)$$

2. Край собственного поглощения определяется при прямых и не прямых переходах соответственно соотношениями:

$$\hbar\omega_{\text{rp}} = \Delta E_0^0, \quad (76.3p)$$

$$\hbar\omega_{\text{rp}} = (\Delta E_0^T \pm \hbar\omega_{\text{фон}}). \quad (76.4p)$$

3. Коэффициент поглощения α у края собственного поглощения пропорционален разности $(\hbar\omega - \Delta E_0)$ в степени r :

$$\alpha \sim (\hbar\omega - \Delta E_0)^r = (\hbar\omega - \hbar\omega_{\text{rp}})^r, \quad (76.5p)$$

где r принимает значения:

1/2 — для прямых разрешенных; 3/2 — для прямых запрещенных; 2 — для не прямых разрешенных; 3 — для не прямых запрещенных переходов.

4. Собственное поглощение приводит к генерации пары свободных носителей заряда — электрона и дырки.

Кулоновское взаимодействие генерированных светом электрона и дырки приводит к изменению формы края собственного поглощения при межзонных переходах, благодаря чему коэффициент поглощения не обращается в нуль на границе полосы.

5. Собственное поглощение приводит к быстрому поглощению света; длина свободного пробега фотона $l_{\text{фт}}$ при $\hbar\omega > \hbar\omega_{\text{rp}}$ имеет величину порядка $10^{-5} - 10^{-6}$ см.

6. Оператор возмущения имеет вид

$$\hat{W} = -\frac{1}{c} (A_j) = \frac{ie\hbar}{m^*c} (A_{\nabla}). \quad (76.6p)$$

§ 77. ПОГЛОЩЕНИЕ СВЕТА РЕШЕТКОЙ

Поглощение света решеткой происходит в результате взаимодействия электромагнитного поля световой волны с движущимися зарядами узлов решетки. В ионных кристаллах локализованный заряд иона, имеющего координату \mathbf{R}_{nj} и скорость $\dot{\mathbf{R}}_{nj}$, взаимодействует со световой волной, которую можно описать вектор-потенциалом $\mathbf{A}(\mathbf{r}, t)$. В гомополярных кристаллах атомы будут иметь дипольный момент вследствие движения ядер или вследствие индукции полем световой волны. Для определенности будем рассматривать движение ионов решетки. Энергию взаимодействия света с решеткой можно записать согласно (76.5) в виде

$$\dot{\mathbf{W}} = \sum_{nj} \left\{ -\frac{1}{c} e_j (\mathbf{A}[\mathbf{R}_{nj}] \dot{\mathbf{R}}_{nj}) + \frac{e_j^2}{2M_j c^2} [\mathbf{A}(\mathbf{R}_{nj})]^2 \right\}. \quad (77.1)$$

Если длина волны света велика по сравнению с периодом решетки и тем более с величиной смещения \mathbf{u}_{nj} , то можно взять значение вектор-потенциала при $\mathbf{u}_{nj} = 0$, т. е. в положении равновесия ядер \mathbf{R}_{nj}^0 . Считая опять-таки интенсивность света не слишком большой, можно отбросить второй член в выражении (77.1). Введем вместо величин $\dot{\mathbf{R}}_{nj}$ нормальные координаты

$$\dot{\mathbf{R}}_{nj} = \dot{\mathbf{u}}_{nj} = \sqrt{M_j} \dot{\mathbf{w}}_{nj}. \quad (77.2)$$

Величины приведенных смещений можно записать в виде

$$\mathbf{w}_{nj, \alpha(s)} = \mathbf{w}_{0j, \alpha(s)} e^{-i[\omega_{\alpha(s)} t - (\mathbf{K}_{\alpha(s)} \cdot \mathbf{n})]}. \quad (77.3)$$

Из (77.3) и (77.2) запишем

$$\dot{\mathbf{R}}_{nj} = -i \sum_{\alpha(s)} \mathbf{u}_{0j, \alpha(s)}^0 e^{-i[\omega_{\alpha(s)} t - (\mathbf{K}_{\alpha(s)} \cdot \mathbf{n})] \omega_{\alpha(s)}}, \quad (77.4)$$

поэтому для оператора возмущения (77.1) из (77.4) имеем

$$\dot{\mathbf{W}} = -\frac{i}{c} \sum_{nj} \sum_{\alpha(s)} e_j \omega_{\alpha(s)} (\mathbf{A}[\mathbf{n} + \mathbf{j}, t] \mathbf{u}_{0j, \alpha(s)}^0) e^{-i[\omega_{\alpha(s)} t - (\mathbf{K}_{\alpha(s)} \cdot \mathbf{n})]}. \quad (77.5)$$

Запишем вектор-потенциал

$$\mathbf{A}(\mathbf{n} + \mathbf{j}, t) = \mathbf{A}_0 e^{i[\omega t - (\mathbf{gR}_{nj}^0)]}. \quad (77.6)$$

Чтобы искать вероятность перехода, необходимо указать, между какими состояниями происходит переход. Нас будет интересовать система фононов; она характеризуется энергией E и волновой функцией Ψ :

$$E = \sum_{\alpha(s)} \hbar \omega_{\alpha(s)} \left(v_{\alpha(s)} + \frac{1}{2} \right), \quad (77.7)$$

$$\Psi = \prod_{\alpha(s)} \Psi_{v_{\alpha(s)}}(q_{\alpha(s)}). \quad (77.8)$$

Для нахождения вероятности перехода из одного состояния (Ψ_1, E_1) в другое (Ψ_2, E_2) необходимо найти матричный элемент оператора возмущения:

$$W_{12} = \int \Psi_1^\dagger \hat{W} \Psi_2 d\Gamma', \quad (77.9)$$

где $d\Gamma'$ — элемент объема в конфигурационном пространстве. Если подставить в (77.9) вместо \hat{W} его выражение (77.5), то можно вычислить W_{12} .

При таком способе вычисления мы не учитываем изменений, происходящих в системе фотонов. Чтобы учесть изменение в системе фотонов, нужно записать волновые функции фотонов с использованием методов вторичного квантования, вводя операторы рождения и уничтожения. Мы ограничимся качественными соображениями.

Вычисляя матричный элемент W_{12} , необходимо иметь в виду, что интеграл берется по всем нормальным координатам; кроме того, примем во внимание, что величина $u_{0j, \alpha(s)}$ есть нормальная координата, так что $u_{0j, \alpha(s)} = q_{\alpha(s)}$, поэтому при интегрировании все величины, кроме $u_{0j, \alpha(s)}$, можно вынести из-под знака интеграла:

$$W_{12} = -\frac{i}{c} \sum_{nj} \sum_{\alpha(s)} e_j \omega_{\alpha(s)} A_0 \cos \theta_j e^{-i[\omega_{\alpha(s)} t - (K_{\alpha(s)}, n)]} \times \\ \times \int \Psi_{\alpha(s)}^* (q_{\alpha(s)}) q_{\alpha(s)} \Psi_{\alpha(s)} (q_{\alpha(s)}) dq_{\alpha(s)}, \quad (77.10)$$

остальные функции фотонов дают единицы при интегрировании.

Матричный элемент координаты гармонического осциллятора равен

$$\int \Psi_{\alpha(s)}^* (q_{\alpha(s)}) q_{\alpha(s)} \Psi_{\alpha(s)} (q_{\alpha(s)}) dq_{\alpha(s)} = \\ = \sqrt{\frac{\hbar}{2M_j \omega_{\alpha(s)}}} \begin{cases} \sqrt{v_{\alpha(s)}''}; & v_{\alpha(s)}' = v_{\alpha(s)}'' - 1; \\ \sqrt{v_{\alpha(s)}'' + 1}; & v_{\alpha(s)}' = v_{\alpha(s)}'' + 1. \end{cases} \quad (77.11)$$

Таким образом, матричный элемент оператора возмущения W_{12} отличен от нуля только в том случае, когда колебательные квантовые числа меняются на единицу: $v'' = v' \pm 1$ для любых частот. Если $v'' = v' - 1$, то число фотонов уменьшается на единицу — происходит уничтожение фотона, если $v'' = v' + 1$ — происходит рождение фотона за счет фотона.

Подставляя (77.11) в (77.10), получим

$$W_{12} = -\frac{i}{c} \sum_{nj} \sum_{\alpha(s)} e_j \sqrt{\frac{\hbar}{2M_j}} \omega_{\alpha(s)}^{\frac{1}{2}} \sqrt{v_{\alpha(s)}'' + 1} A_0 \cos \theta_j \times \\ \times e^{i(\omega - \omega_{\alpha(s)})t} e^{i(K_{\alpha(s)} - g, n)} e^{-t(g)}, \quad (77.12)$$

Для вероятности перехода $\omega(1, 2)$ получим

$$\omega(1, 2) = \frac{2\pi}{\hbar^2 c^2} A_0^2 \left| \sum_j \sum_{\alpha(s)} e_j e^{-i(g_j)} \cos \theta_j \sqrt{\frac{\hbar \omega_{\alpha(s)} (v'_{\alpha(s)} + 1)}{2M_j}} \right. \times \\ \left. \times \sum_{\mathbf{n}} e^{i(K_{\alpha(s)} - \mathbf{g}, \mathbf{n})} \right|^2 \times \delta(\omega - \omega_{\alpha(s)}). \quad (77.13)$$

Мы видим, что поглощение света колебаниями решетки носит резонансный характер: $\omega(1, 2) = 0$, если $\omega \neq \omega(s)$. Из сплошного спектра излучения колебаниями решетки поглощаются только собственные частоты $\omega = \omega_{\alpha(s)}$. Значения $\omega_{\alpha(s)}$ определяются совокупностью оптических и акустических ветвей. Учитывая, что $\omega_{\alpha(s)} = \omega$, можем существенно упростить выражение (77.13). Однако прежде всего учтем, что последняя сумма в (77.13) легко вычисляется:

$$\sum_{\mathbf{n}} e^{i(K_{\alpha(s)} - \mathbf{g}, \mathbf{n})} = N \delta_{K_{\alpha(s)}, \mathbf{g} + 2\pi\mathbf{b}}, \quad (77.14)$$

где N — число ячеек в кристалле, \mathbf{b} — вектор обратной решетки. Выражение (77.14) имеет наглядный смысл: при поглощении света должен выполняться закон сохранения квазиимпульса:

$$K_{\alpha(s)} = \mathbf{g} + 2\pi\mathbf{b}. \quad (77.15)$$

Если мы, как всегда, ограничимся основной зоной Бриллюэна для колебания решетки ($\mathbf{b} = 0$), то

$$K_{\alpha(s)} = \mathbf{g}, \quad (77.16)$$

т. е. поглощаются только такие фотоны, импульс которых равен импульсу фононов. Объединяя соотношения (77.13) и (77.15), можно записать

$$\hbar \omega_{\alpha(s)} = \hbar \omega, \\ \hbar K_{\alpha(s)} = \hbar \mathbf{g} \quad (77.17)$$

— законы сохранения энергии и импульса. Они позволяют провести геометрическое построение для нахождения спектра поглощения колебаниями решетки. На рис. 122 приведены акустическая и оптическая ветви колебаний решетки. Тангенс угла наклона акустических ветвей равен скорости звука. Нанесем на том же графике дисперсионную зависимость $\omega(\mathbf{g})$ для света:

$$\omega = c\mathbf{g}, \quad (77.18)$$

прямолинейная зависимость частоты фотона от его волнового вектора изображена на рис. 122 пунктиром. Поскольку угол наклона должен быть равен c — скорости света, то дисперсионная кривая для света должна быть практически вертикальной линией. Поглощаются

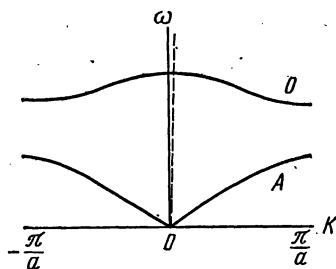


Рис. 122. Определение спектра поглощения света колебаниями решетки

будут только те частоты, которые находятся в точках пересечения $\omega_{\alpha(s)}(K)$ и $\omega = cg$, поскольку именно для этих точек выполняются одновременно законы сохранения энергии и импульса.

Отсюда следует, что поглощение света акустической ветвью действительно не происходит, поскольку она не может пересечься с прямой $\omega = cK$. Из каждой оптической ветви в поглощении проявляются только длинноволновые колебания, практически можно считать, что спектр поглощения колебаниями решетки сводится к $(3s - 3)$ частотам $\omega_{(s)}^0(0)$.

Запишем теперь выражение для вероятности перехода, учитывая правила отбора (77.17) и (77.11):

$$\omega(1, 2) = \frac{\pi A_0^2 N^2 \hbar \omega}{\hbar^2 c^2} \left| \left[\sum_j e_j e^{-i(g, l)} \cos \theta_j \sqrt{\frac{v'_0 + 1}{M_j}} \right] \right|^2. \quad (77.18')$$

Если проквантовать A_0^2 тем же способом, который был использован нами в предыдущем случае

$$A_0^2 = \frac{8\pi \hbar c n}{\epsilon \omega} q, \quad (77.19)$$

то получим

$$\omega(1, 2) = \frac{8\pi^2 N^2 n}{\epsilon \epsilon} q \left| \sum_j e_j e^{-i(g, l)} \cos \theta_j \sqrt{\frac{v'_0 + 1}{M_j}} \right|^2. \quad (77.20)$$

Эффективное сечение поглощения одного фотона пропорционально числу фононов $v'_0 + 1$ данной частоты.

Величина j имеет порядок постоянной решетки, в то время как длина волны света в области поглощения много больше постоянной решетки, поэтому $e^{-i(g, l)} \approx 1$.

Рассмотрим теперь зависимость $\omega(1, 2)$ от поляризации света и фононов. Угол θ_j есть угол между векторами поляризации фотона и фонона, и поскольку $K = g$, то для продольных оптических колебаний $\theta_j = 90^\circ$, и поглощения не происходит. В поглощении будут принимать участие только поперечные оптические фононы.

При освещении кристалла естественным светом поляризацию фотонов и фононов можно разложить по двум взаимно ортогональным направлениям, благодаря чему вместо $\cos \theta_j$ можно поставить единицу с тем или другим знаком. Знак $\cos \theta_j$ должен совпадать со знаком e_j (или быть противоположного с ним знака). Это связано с тем, что смещения ионов ячейки происходят в противофазе. Учитывая последнее замечание, запишем

$$\omega(1, 2) = \frac{8\pi^2 N^2 e^2 n (v'_0 + 1)}{\epsilon \epsilon M_{np}} q. \quad (77.21)$$

Мы заменили M_j на некоторую приведенную массу M_{np} . Для решетки с двумя ионами $M_{np} = \frac{M_1 M_2}{M_1 + M_2}$. Полагая $N(M_1 + M_2) = \rho V$, запишем

$$\omega(1, 2) = \frac{8\pi^2 N V \rho e^2 (v'_0 + 1) n}{\epsilon \epsilon M_1 M_2} q. \quad (77.22)$$

Подставим вместо ν_0' их равновесное значение:

$$\langle \nu_0' \rangle = \frac{1}{\frac{\hbar\omega}{e^{kT}} - 1}. \quad (77.23)$$

Выражение (77.23) определяет частотно-температурную зависимость поглощения света колебаниями решетки. С ростом температуры, когда $\frac{\hbar\omega}{kT} \gg 1$, $\omega(1, 2) \sim T$. Оценим необходимую для этого

температуру: Положим, что длина волны равна 20 мкм; тогда $\hbar\omega \approx 0,06$ эВ, что соответствует температуре 400° С.

Поскольку вероятности прямых и обратных переходов равны, то сильному поглощению должно соответство-

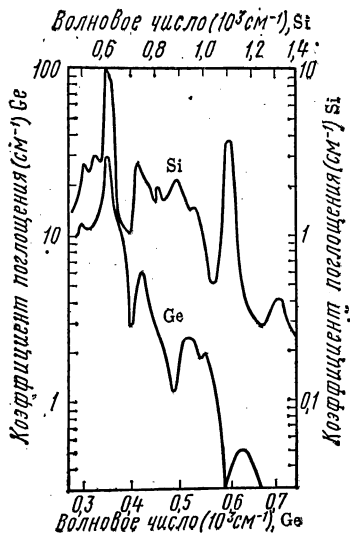


Рис. 123. Спектр поглощения колебаниями решетки в германии и кремнии

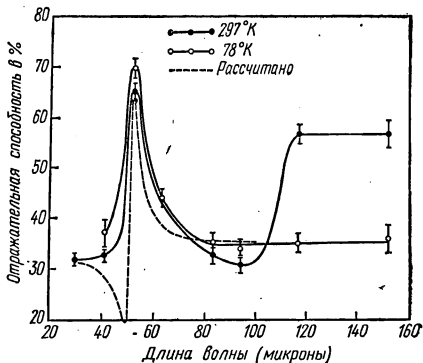


Рис. 124. Спектр отражения антимионида индия

вать и сильное отражение. Спектр поглощения решеткой наблюдается в виде так называемых *остаточных лучей*. Если направить на кристалл пучок широкого спектрального интервала, то в отраженном луче останутся только частоты собственных колебаний решетки, удовлетворяющих правилам отбора (77.17). В действительности же спектр поглощения колебаниями решетки имеет полосы определенной ширины. На рис. 123 приведен спектр колебаний решетки германия и кремния. На рис. 124 приведен спектр отражения антимионида индия.

Как видно из рис. 123, спектр поглощения света колебаниями решетки в германии и кремнии имеет ряд сравнительно широких полос: в области от 300 до 700 см^{-1} (33 — 14 мкм) в германии и в области от 500 до 1400 см^{-1} (20 — 7 мкм) в кремнии. Коэффициент поглощения измеряется единицами и десятками см^{-1} . Коэффициент поглощения в германии больше, чем в кремнии, поскольку плотность вещества в германии больше, чем в кремнии.

Спектр отражения антимолибдита индия содержит хорошо выраженную полосу в области $\lambda \approx 50$ мкм, обусловленную колебаниями решетки. Обращают на себя внимание два момента на рис. 124. При малой температуре, когда концентрация носителей заряда мала, отражение имеет чисто диэлектрический характер, обусловленный коэффициентом преломления. При комнатной температуре проявляется отражение, связанное с поглощением свободными носителями заряда. Коэффициент отражения возрастает с ростом длины волны, пока $\lambda < \lambda_{гр}$; при $\lambda > \lambda_{гр}$ он остается постоянным.

Резюме § 77

1. Колебания решетки приводят к поглощению света в соответствии с правилами отбора:

$$\hbar\omega_{\alpha(s)} = \hbar\omega, \quad (77.1p)$$

$$\hbar K = \hbar g. \quad (77.2p)$$

2. Из правил отбора (77.1—2p) следует, что свет поглощается только оптическими колебаниями решетки, при этом

$$\omega \approx \omega_{\alpha(s)}^0(0). \quad (77.3p)$$

Свет поглощается поперечными оптическими колебаниями решетки. Колебания решетки проявляются в спектре отражения в виде «остаточных» лучей.

§ 78. ПОГЛОЩЕНИЕ СВЕТА ЭЛЕКТРОНАМИ В ЛОКАЛИЗОВАННЫХ СОСТОЯНИЯХ

Электрон или дырка, находясь в локализованном состоянии, может поглотить фотон, перейдя при этом в другое локализованное состояние или в свободное состояние. Полосы поглощения света при переходах носителей заряда с дискретных уровней лежат за границей собственного поглощения в сторону меньших частот. Положение полос поглощения можно определить соотношением

$$\lambda_{лок} = \frac{1,24}{|E_{лок} - E_{extr}| (\text{эВ})} \text{ (мкм)}. \quad (78.1)$$

Рассмотрим элементарную теорию поглощения света электроном в локализованном состоянии. Обозначим исходное состояние значком «1», конечное состояние — «2». Волновую функцию локализованного состояния возьмем в виде

$$\psi_1 = C e^{-\rho r} e^{-\frac{i}{\hbar} E_{лок} t}, \quad (78.2)$$

где $\rho = |\rho| = |g - R_{лок}|$, $E_{лок}$ — энергия локального уровня. Величина $\frac{x^{-1}}{2}$ равна расстоянию, на котором вероятность нахождения элек-

трона уменьшается в e раз. Функция (78.2) подобна волновой функции s -состояния одноэлектронной атомной системы. Из условия нормировки волновой функции можно найти коэффициент C :

$$1 = |C|^2 \int e^{-2\kappa\rho} d\tau = |C|^2 4\pi \int_0^\infty e^{-2\kappa\rho} \rho^2 d\rho = \frac{\pi |C|^2}{\kappa^3}, \quad (78.3)$$

откуда

$$C = C^* = \frac{\kappa^{3/2}}{\sqrt{\pi}}. \quad (78.4)$$

Для конечного состояния возьмем волновую функцию электрона в зоне проводимости

$$\psi_2 = \frac{1}{L^{3/2}} e^{\left[-\frac{iE_2 t}{\hbar} + i(\kappa_2 r)\right]}. \quad (78.5)$$

Энергию взаимодействия выберем в обычном виде

$$\hat{W} = -\frac{ie\hbar}{cm^*} (\mathbf{A}\nabla) = -\frac{ie\hbar}{cm^*} e^{i[\omega t - (\mathbf{g}, \mathbf{r})]} (\mathbf{A}_0\nabla). \quad (78.6)$$

Найдем матричный элемент W_{12} :

$$W_{12} = \int \psi_1^* \hat{W} \psi_2 d\tau = -\frac{ie\hbar}{cm^*} \frac{\kappa^2}{\sqrt{\pi} L^{3/2}} e^{\frac{i}{\hbar} [E_{\text{лок}} + \hbar\omega - E_2] t} \times \\ \times \int e^{-\kappa\rho} e^{-i(\mathbf{g}, \mathbf{r})} (\mathbf{A}_0\nabla) e^{-i(\kappa_2 r)} d\tau. \quad (78.7)$$

Поместим начало отсчета в точке $\mathbf{R}_{\text{лок}}$, тогда $|\rho| = |\mathbf{r}|$. Учитывая, что

$$(\mathbf{A}_0\nabla) e^{i(\kappa_2 r)} = i(\mathbf{A}_0\kappa_2) e^{i(\kappa_2 r)}, \quad (78.8)$$

запишем

$$W_{12} = \frac{e\hbar\kappa^{3/2} (\mathbf{A}_0\kappa_2) e^{\frac{i}{\hbar} [E_{\text{лок}} + \hbar\omega - E_2] t}}{cm^* \sqrt{\pi} L^{3/2}} \int e^{-\kappa r + i(\kappa_2 - \mathbf{g}, \mathbf{r})} d\tau. \quad (78.9)$$

Вычислим интеграл в (78.9); положив

$$\kappa_2 - \mathbf{g} = \mathbf{Q}, \quad (78.10)$$

запишем

$$\int e^{-\kappa r} e^{i(\mathbf{Q}, \mathbf{r})} d\tau = \int_0^\infty e^{-\kappa r} r^2 dr \int_0^\pi e^{iQr \cos\theta} \sin\theta d\theta \int_0^{2\pi} d\varphi = \\ = \frac{2\pi}{iQ} \int_0^\infty [e^{-(\kappa+iQ)r} - e^{-(\kappa-iQ)r}] r dr = \frac{2\pi}{iQ} \left[\frac{1}{(\kappa-iQ)^2} - \frac{1}{(\kappa+iQ)^2} \right] = \frac{8\pi\kappa}{(\kappa^2 + Q^2)^2}. \quad (78.11)$$

Из (78.11) и (78.9) для матричного элемента возмущения имеем

$$W_{12} = \frac{8\sqrt{\pi} e \hbar \kappa^{\frac{5}{2}}}{3 c m^* L^2} \frac{(A_0 \kappa_2)}{(\kappa^2 + Q^2)^2} e^{\frac{i}{\hbar} [E_{\text{лок}} + \hbar \omega - E_2] t}. \quad (78.12)$$

Вероятность перехода в единицу времени из состояния «1» в состояние «2» равна

$$\omega(1, 2) = \frac{128\pi^2 \hbar e^2}{c^2 m^{*2} L^3} \frac{\kappa^5}{(\kappa^2 + Q^2)^4} A_0^2 \kappa_2^2 \cos(A_0 \hat{\kappa}_2) \delta[E_{\text{лок}} + \hbar \omega - E_2]. \quad (78.13)$$

Из закона сохранения энергии

$$E_2 = E_{\text{лок}} + \hbar \omega \quad (78.14)$$

следует, что поглощение света дискретным центром может происходить при любой энергии $\hbar \omega \geq E_c - E_{\text{лок}} = E_I$. Подставляя вместо A_0^2 его выражение (76.22) через поток фотонов, перепишем

$$\omega(1, 2) = \frac{2^{11} \pi^3 \hbar^2 e^2 n}{3 c m^{*2} L^3 \epsilon \omega} \frac{\kappa^5}{(\kappa^2 + Q^2)^4} \kappa_2^2 q = \frac{2^{11} \pi^3 e^2 n}{3 c m^* L^3 \hbar \epsilon} \frac{\kappa^5}{(\kappa^2 + Q^2)^4} \frac{(\hbar \omega - E_I)}{\omega^2} q. \quad (78.15)$$

Умножая (78.15) на концентрацию локальных центров поглощения $N_{\text{лок}}$, получим коэффициент поглощения:

$$\alpha_{\text{лок}} = \frac{2^{11} \pi^3 e^2 n N_{\text{лок}}}{3 c m^* \hbar L^3 \epsilon} \frac{\kappa^5}{(\kappa^2 + Q^2)^4} \frac{(\hbar \omega - E_I)}{\omega^2}. \quad (78.16)$$

Поскольку энергия E_2 электрона согласно (78.14) может быть любой, то и квазиимпульс электрона не ограничивается никакими правилами отбора. Чтобы найти правила отбора, рассмотрим с этой целью величину β :

$$\beta = \frac{\kappa^5 \kappa_2^2}{(\kappa^2 + Q^2)^4} = \frac{\kappa^5 \kappa_2^2}{[\kappa^2 + (\kappa_2 - g)^2]^4}. \quad (78.17)$$

Поскольку для поглощения света электроном на дискретном уровне $g \sim 10^3 \text{ см}^{-1}$, а для электронов проводимости с энергией порядка kT волновой вектор достигает значений 10^7 см^{-1} , то величиной g по сравнению с κ_2 можно пренебречь, поэтому

$$\beta = \frac{\kappa^5 \kappa_2^2}{(\kappa^2 + \kappa_2^2)^4}. \quad (78.18)$$

Если рассматривать α как функцию κ_2^2 , то из (78.18) видно, что α проходит через максимум, положение которого найдем из условия

$$\frac{d}{d\kappa_2} \frac{\kappa^5 \kappa_2^2}{(\kappa^2 + \kappa_2^2)^4} = 0 = (\kappa^2 + \kappa_2^2)^4 2\kappa_2 - 4\kappa_2^3 (\kappa^2 + \kappa_2^2)^3 2\kappa_2 = 0, \quad (78.19)$$

или

$$\kappa_2 = \frac{\kappa}{\sqrt{3}}. \quad (78.20)$$

Вследствие резкого уменьшения α при больших κ_2 ($\alpha \sim \kappa_2^{-6}$ при $\kappa_2 \gg \kappa$) мы можем считать, что волновой вектор электрона меняется

в интервале $0 < \kappa_2 < \kappa$, другими словами, при поглощении света электронами в локализованных состояниях поглощаются только такие фотоны, которые переводят электрон в состояние с энергией $E_2(\kappa_2) = E_{\text{лок}} + \hbar\omega$, где

$$\kappa_2 \approx \kappa. \quad (78.21)$$

Уравнение (78.21) представляет собой своеобразное правило отбора при поглощении света электроном в локализованном состоянии. Легко показать, что оно является следствием соотношений неопределенности. Действительно, если электрон находится в потенциальной яме, размеры которой порядка $2\kappa^{-1}$, то его средний импульс равен нулю, но неопределенность в импульсе

$$\Delta p \approx \frac{\hbar}{2\kappa^{-1}} = \frac{\hbar\kappa}{2}. \quad (78.22)$$

При переходе электрона в свободное состояние он может иметь импульс $\hbar\kappa_2 \approx \Delta p$, т. е.

$$\kappa_2 \approx \frac{\kappa}{2} \quad (78.23)$$

в полном соответствии с (78.21). Итак, приходим к выводу, что при поглощении света электроном в локализованном состоянии спектр поглощения должен иметь вид полосы сравнительно большой, но ограниченной ширины. Максимум полосы находится при частоте, определяемой условием

$$\hbar\omega = E_2(\kappa_2) - E_{\text{лок}} = E_I + \frac{\hbar^2\kappa^2}{8m^*}. \quad (78.24)$$

Если электрон находится на примесном уровне, то энергия фотоионизации, определяемая по максимуму примесного поглощения, превосходит энергию термической ионизации примеси $E_I = E_c - E_{\text{лок}}$ на величину порядка $\frac{\hbar^2\kappa^2}{8m^*}$.

Форму кривой коэффициента поглощения можно оценить на основании соотношения (78.16), если вместо κ_2^2 подставить (78.14):

$$\alpha \sim \frac{\hbar\omega - E_I}{\hbar^2\omega^2 [\hbar^2\kappa^2 + 2m^*(\hbar\omega - E_I)]^4}. \quad (78.25)$$

Как видно из (78.25), α линейно растет относительно частоты при $\hbar\omega < 2E_I$, проходит через максимум, уменьшается сначала медленно $\sim \omega^{-1}$, затем значительно быстрее.

Если локальные состояния имеют водородоподобный спектр, то к ним можно применить теорию излучения (поглощения) атома водорода. В этом случае выражение для вероятности перехода позволяет записать выражение для коэффициента поглощения в виде

$$\alpha = N_{\text{лок}} \frac{2^{10}\pi^2\hbar e^2}{3\pi m^* E_I} \left(\frac{E_I}{\hbar\omega}\right)^4 \frac{\exp\left[4\left(1 - \frac{E_I}{\hbar\omega}\right) \arctg\left(1 - \frac{E_I}{\hbar\omega}\right)\right]}{1 - \exp\left[2\pi\left(\frac{E_I}{\hbar\omega} - 1\right)\right]}, \quad (78.26)$$

где n — показатель преломления, E_I — энергия ионизации примеси. При $m^* \sim 10^{-27}$ г $n=4$, $E_I=0,05$ эВ

$$\sigma = \frac{\alpha}{N_{\text{лок}}} \approx 4 \cdot 10^{-16} \text{ см}^2.$$

Как видим, разные методы расчета приводят к различным выражениям для спектра поглощения при переходах между зоной и локализованным состоянием. Это различие не только методическое, но и отражает различия в механизме поглощения. Проще всего это показывает переход из валентной зоны на дискретный уровень локализованного состояния, если поменять места функции ψ_1 и ψ_2 при вычислении матричного элемента оператора возмущения в (78.7).

Рис. 125. Спектр поглощения атомов бора в кремнии

Переходы, описываемые матричным элементом (78.9), соответствуют запрещенным переходам. Для разрешенных переходов следует ввести матричный элемент оператора возмущения, независящий от волнового вектора. Вследствие простоты вычисления его по приведенной схеме ограничимся сделанным замечанием.

Локализованные состояния могут иметь различную физическую природу. Простейший способ их создания состоит во введении примеси в полупроводник. Донорные и акцепторные примеси в обычных условиях ионизованы, поэтому они не могут поглощать свет. Но при низких температурах, когда примесь ионизована не полностью, наблюдается поглощение примесью.

На рис. 125 видна широкая полоса фотоионизации атомов бора в кремнии. Максимум коэффициента поглощения находится при $\hbar\omega \approx 0,055$ эВ, после чего коэффициент поглощения медленно уменьшается до нуля в интервале $\delta\hbar\omega \approx 2E_I$. На область линейного роста коэффициента поглощения накладывается ряд узких полос поглощения. Их можно объяснить наличием возбужденных уровней энергии примесного атома бора, лежащих в запрещенной области, как

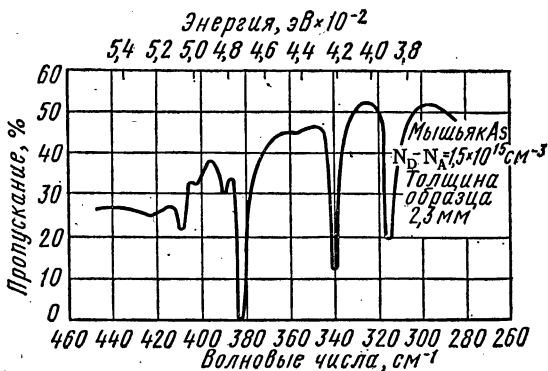


Рис. 126. Спектр примесного поглощения кремния, легированного мышьяком

это изображено на рис. 30. Переходу атома примеси из основного в возбужденное состояние соответствует узкая полоса поглощения. На рис. 126 приведены полосы поглощения атомов мышьяка в кремнии, соответствующих фотоионизации и переходу атомов мышьяка в возбужденное состояние.

Таблица 25

Примесь	Энергия термической ионизации, эВ	Энергия оптической ионизации, эВ
B	0,045	0,046
Al	0,057	0,067
Ga	0,065	0,071
In	0,16	0,154
P	0,044	0,0503
As	0,049	0,0533
Sb	0,039	0,0426

В табл. 25 приведены значения энергий ионизаций атомов примеси при фото- и термоионизации. Обращает на себя внимание тот факт, что энергия фотоионизации несколько больше энергии термической ионизации. Если считать, что это различие обусловлено соотношением неопределенности и вытекающим из него правилом отбора, то при разности энергий фото- и термоионизации $\delta E_I \approx 10^{-2}$ эВ = $1,6 \cdot 10^{-14}$ эрг на основании (78.24) при $m^* = m$ получим $\lambda \approx 10^7$ см⁻¹.

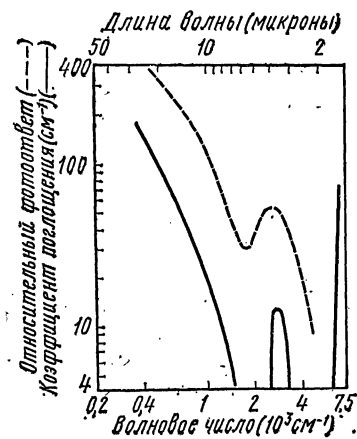


Рис. 127. Коэффициент поглощения германия *p*-типа (сплошная кривая) и относительный фотоответ германия *p*-типа, легированного индием, при температуре 5 К

На рис. 127 приведена полоса поглощения германия *p*-типа, легированного индием, при температуре 5К.

В ряде случаев узкие полосы поглощения наблюдаются за краем собственного поглощения в сторону больших длин волн. Эти линии могут быть вызваны глубокими уровнями или мелкими уровнями, но при переходе на них электронов из валентной зоны. Для некоторых веществ подобные линии могут быть интерпретированы как экситонные линии. Экситонные линии наблюдаются в закиси меди, сульфиде кадмия, селениде кадмия, германии и некоторых других веществах.

На рис. 113 приведен спектр поглощения *p*-кремния, который был подвергнут облучению нейтронами. Нейтроны, сталкиваясь с ядрами кремния, создают в нем большое число дефектов. Сопротивление кремния и германия после облучения обычно сильно возрастает. На рис. 113 это видно по резкому уменьшению поглощения свободными носителями заряда. За границей собственного поглощения появляется пик поглощения, обусловленный дефектами. Площадь кривой поглощения дефектами пропорциональна дозе облучения, которая определяет концентрацию дефектов.

В ионных кристаллах, прозрачных в видимой области, дефекты, захватывая электрон, окрашивают кристалл. Такие дефекты получили название *центров окраски*. К ним принадлежит, например, F-центр — это вакантный узел, захвативший электрон.

В некоторых случаях примесь является электронейтральной, т. е. термически она не ионизируется и, следовательно, не является поставщиком свободных носителей заряда. Но при действии света такая примесь может ионизоваться, приводя к появлению полос поглощения в инфракрасной области. Примером такой примеси может служить кислород в кремнии, который не влияет на его электрические свойства, но дает полосу поглощения в области 9 мкм.

Резюме § 78

1. Правила отбора при поглощении света носителями заряда в локализованном состоянии имеют вид

$$\hbar\omega = E_2 - E_{\text{лок}}, \quad (78.1p)$$

$$\kappa_2 \approx \kappa, \quad (78.2p)$$

где $2\kappa^{-1}$ — размеры области локализации электрона или дырки.

2. Примесное поглощение имеет вид узких полос, если происходит переход электрона между дискретными уровнями энергии, соответствующими возбужденному состоянию примеси, и сравнительно широким полос при фотоионизации. Энергия фотоионизации несколько больше энергии термической ионизации.

3. Примесное поглощение приводит к генерации носителей заряда одного типа.

§ 79. ВЛИЯНИЕ ВНЕШНИХ УСЛОВИЙ НА СПЕКТР ПОГЛОЩЕНИЯ

Спектр поглощения полупроводника меняется при изменении внешних условий, к числу которых относятся температура, давление, внешние поля.

Температура влияет на спектр поглощения следующим образом.

При повышении температуры ширина запрещенной зоны для большинства полупроводников уменьшается, следовательно, граница собственного поглощения смещается в сторону длинных волн. На рис. 128 приведен спектр поглощения пленок селена при комнатной (1, 2, 3) и азотной (4, 5) температурах. Отчетливо видно, что при переходе от азотной температуры к комнатной край собственного поглощения смещается в сторону больших длин волн.

Если примесь электрически активна, то ее спектр поглощения может наблюдаться только при очень низких температурах, когда свободные носители заряда «вымораживаются» на примесные уровни. Другими словами, с ростом температуры по мере ионизации атомов

примеси спектр примесного поглощения исчезает. Поскольку концентрация свободных носителей заряда изменяется при изменении температуры, то изменяется и коэффициент поглощения в том же направлении, как это было видно на рис. 114.

Температура может влиять на положение границы собственного поглощения косвенным образом. Если зоны энергии имеют малую плотность состояний, то они быстро заполняются носителями заряда, вследствие чего переход на занятые уровни невозможен, и граница собственного поглощения смещается в сторону коротких длин волн. Обычно этот эффект рассматривается в качестве концентрационного — положение границы собственного поглощения определяется концентрацией примеси. Это обусловлено тем, что при большой концентрации примеси энергия ионизации примеси обращается в нуль. Наиболее отчетливо этот эффект наблюдается в антимониде индия, который имеет малую плотность состояния в зоне проводимости. На рис. 129

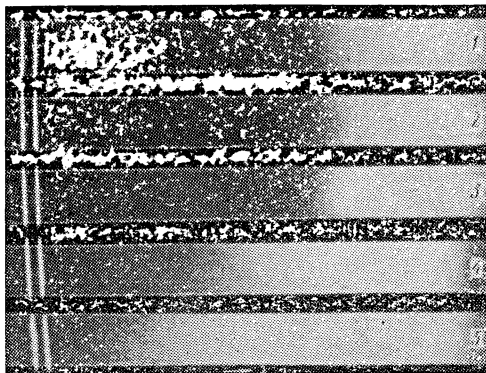


Рис. 128. Положение границы собственного поглощения пленок селена при комнатной (1—3) и азотной (4—5) температурах

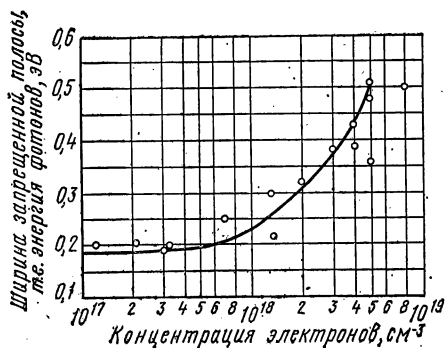


Рис. 129. Положение границы собственного поглощения InSb в зависимости от концентрации электронов

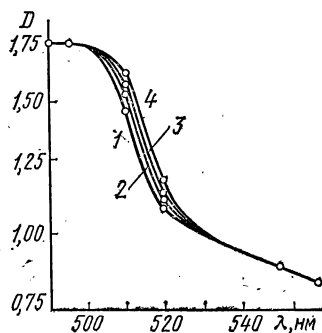


Рис. 130. Смещение края собственного поглощения пленки CdS при наложении на нее электрического поля. Напряжение в кВ: 1) 0; 2) 1,2; 3) 1,6; 4) 1,8

представлена зависимость положения границы собственного поглощения от концентрации электронов в InSb. Как видно из рис. 129, при изменении концентрации электронов от 10^{17} до $5 \cdot 10^{18}$ см⁻³ зона проводимости заполняется электронами на 0,3 эВ выше E_c . Смеще-

ние края поглощения с ростом степени вырождения называется эффектом Бурштейна.

Сжатие или растяжение кристалла меняет ширину запрещенной зоны, что приводит к смещению границы собственного поглощения.

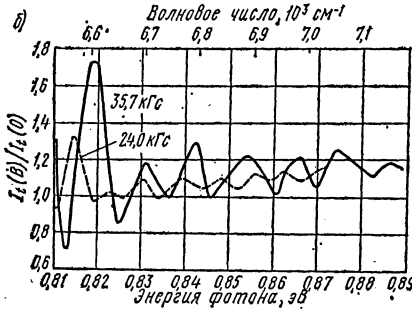
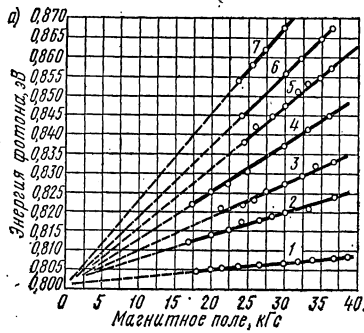


Рис. 131. Осцилляторный магнитно-абсорбционный эффект в германии

На положение края собственного поглощения оказывает влияние электрическое поле. На рис. 130 представлено положение края собственного поглощения в пленках CdS при трех различных напряжениях, поданных на пленку. Край собственного поглощения смещается в длинноволновую область спектра, что равносильно уменьшению ширины запрещенной зоны в сильном электрическом поле. Этот эффект называют *эффектом Келдыша — Франца*.

На положение границы собственного поглощения оказывает влияние и магнитное поле. Как было показано во второй и третьей главах, дно зоны проводимости смещается вверх при наложении магнитного поля на величину $\frac{\hbar\omega_c}{2} = \frac{\hbar eB}{2m_n^*}$. Потолок валентной зоны смещается вниз на величину $\frac{\hbar eB}{2m_p^*}$, вследствие чего шири-

на запрещенной зоны увеличивается и граница собственного поглощения смещается в коротковолновую сторону:

$$\begin{aligned} \hbar\omega_{\text{гп}} &= E_c(B) - E_v(B) = E_{c0} + \frac{\hbar eB}{2m_n^*} - \left(E_{v0} - \frac{\hbar eB}{2m_p^*} \right) = \\ &= \Delta E_0 + \frac{\hbar eB}{2} \left(\frac{1}{m_n^*} + \frac{1}{m_p^*} \right) = \Delta E_0 + \frac{\hbar eB}{2m_{\text{np}}^*}. \end{aligned} \quad (79.1)$$

Измерив смещение границы собственного поглощения как функцию магнитного поля, можно определить приведенную эффективную массу электрона и дырки. Однако это не единственный эффект, наблюдаемый при поглощении света полупроводником, находящимся в магнитном поле. Согласно (76.36) коэффициент собственного поглощения пропорционален плотности состояний в зонах энергии. Но в § 35 было показано, что плотность состояний при наложении магнитного поля изменяется — она обращается в бесконечность на уровнях Ландау и практически равна нулю между минимумами парабол Ландау (рис. 58). Но отсюда следует, что коэффициент собственного погло-

щения отличен от нуля только при переходах электронов между уровнями Ландау, вследствие чего спектр собственного поглощения должен представлять собой ряд узких полос. Этот эффект при низких температурах и больших магнитных полях наблюдается в ряде веществ, он называется *осцилляторным магнитноабсорбционным эффектом, или магнетопоглощением*. На рис. 131, а представлено отношение интенсивностей света, прошедшего через образец германия в магнитном поле $J(B)$ и без магнитного поля $J(0)$. Мы видим, что пропускание имеет целый ряд пиков, положение которых зависит от величины поля. На рис. 131, б представлена зависимость положения максимумов магнетопоглощения от величины поля. При $B \rightarrow 0$ положение всех максимумов сходится к одной точке, представляющей собой оптическую ширину запрещенной зоны германия при вертикальных переходах. Этот эффект позволяет определить ширину запрещенной зоны с наибольшей точностью. Осцилляторный магнитноабсорбционный эффект подобен циклотронному резонансу, поскольку и в том и другом эффектах происходит переход электронов между уровнями Ландау, но при циклотронном резонансе происходит переход электронов между уровнями Ландау одной и той же зоны, а при магнетопоглощении — между уровнями Ландау различных зон. Энергии фотонов, необходимых для таких переходов, отличаются на величину порядка ширины запрещенной зоны.

§ 80. ФОТОРЕЗИСТИВНЫЙ ЭФФЕКТ

Изменение электрического сопротивления полупроводника, обусловленное непосредственным действием излучения, называют фоторезистивным эффектом, или внутренним фотоэлектрическим эффектом. Изменение сопротивления, или проводимости, вызывают изменением концентрации носителей заряда. Фоторезистивный эффект можно описать величиной световой проводимости $\sigma_{св}$:

$$\sigma_{св} = e_n \mu_n \delta n + e_p \mu_p \delta p, \quad (80.1)$$

где δn , δp — избыточная концентрация электронов и дырок, создаваемая светом. Если $\sigma_{св} > 0$, то фоторезистивный эффект называют положительным; если $\sigma_{св} < 0$ — отрицательным. Проводимость $\sigma_0 = \sigma_t$, обусловленную равновесными носителями заряда, называют темновой. Полная проводимость может быть представлена в виде суммы темновой и световой проводимостей:

$$\sigma = e_n \mu_n (n_0 + \delta n) + e_p \mu_p (p_0 + \delta p) = \sigma_t + \sigma_{св}. \quad (80.2)$$

Собственное поглощение света приводит к генерации электронов и дырок в равных количествах: $\delta n = \delta p$. Экситонное поглощение приводит к созданию связанной пары электрон — дырка, являющейся электрически нейтральным образованием, поэтому экситонное поглощение непосредственно не увеличивает концентрации носителей заряда. Но если экситон при движении в решетке диссоциирует в результате поглощения дополнительной энергии, то каждый экситон

приводит к появлению двух свободных носителей заряда: электрона и дырки. Если же экситон рекомбинирует, то экситонное поглощение не приводит к увеличению проводимости. Кстати говоря, само понятие экситона было введено Френкелем для объяснения отсутствия фоторезистивного эффекта при наличии сильного поглощения света.

При фотоионизации локализованных состояний типа примесных атомов или F -центров возрастает концентрация носителей заряда только одного типа. При этом возможно два случая: неравновесные носители заряда являются основными или неосновными. Если неравновесные носители заряда являются неосновными, а их концентрация превосходит темновую концентрацию основных носителей заряда, то меняется тип проводимости. При освещении полупроводника все кинетические явления могут протекать иначе, чем без освещения, например, может изменяться знак коэффициента Холла.

Поглощение света колебаниями решетки может привести к возрастанию концентрации носителей заряда лишь в результате вторичного эффекта — поглощение света увеличивает концентрацию фононов, которые отдают свою энергию на возбуждение носителей заряда.

Поглощение света свободными носителями заряда не приводит к изменению их концентрации, однако при этом нарушается равновесное распределение носителей заряда по состояниям, они становятся более «горячими», вследствие чего их подвижность может изменяться. Это в свою очередь приводит к изменению проводимости.

До сих пор, говоря о фоторезистивном эффекте, имели в виду увеличение проводимости. Однако при освещении полупроводника возможно увеличение сопротивления. Объяснение этого явления состоит в следующем: нарушение равновесия между тепловой генерацией и рекомбинацией вблизи поверхности полупроводника, вызванное светом, приводит к возрастанию скорости рекомбинации в объеме, куда диффундируют неосновные носители заряда, что приводит к уменьшению концентрации основных носителей заряда и, следовательно, к росту сопротивления.

Чтобы описать фоторезистивный эффект, необходимо знать световую концентрацию носителей заряда. Она определяется из уравнений непрерывности (64.6) и (64.7):

$$\frac{\partial n}{\partial t} = -\operatorname{div} \frac{j_n}{e_n} + G_n - \frac{n-n_0}{\tau_f^n}, \quad (80.3)$$

$$\frac{\partial p}{\partial t} = -\operatorname{div} \frac{j_p}{e_p} + G_p - \frac{p-p_0}{\tau_f^p}. \quad (80.4)$$

Рассмотрим полупроводник, в котором тока нет: $j_n = j_p = 0$. В таком случае уравнения (80.3) и (80.4) упрощаются:

$$\frac{\partial n}{\partial t} = G_n - \frac{n-n_0}{\tau_f^n}, \quad (80.5)$$

$$\frac{\partial p}{\partial t} = G_p - \frac{p-p_0}{\tau_f^p}. \quad (80.6)$$

Если известны скорости генерации носителей заряда G_n, G_p , то из уравнений (80.5), (80.6) можно найти концентрацию носителей заряда и тем самым световую проводимость $\sigma_{св}$. В стационарном состоянии $\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = 0$, и уравнения (80.5), (80.6) имеют простое очевидное решение:

$$n - n_0 = \delta n = G_n \tau_f^n, \quad (80.7)$$

$$p - p_0 = \delta p = G_p \tau_f^p. \quad (80.8)$$

Выражения (80.7) и (80.8) носят название *первого характеристического соотношения* для фоторезистивного эффекта. Они определяют зависимость световой проводимости от интенсивности света и области спектра. Запишем величину световой проводимости

$$\sigma_{св} = e_p \mu_p (G_p \tau_f^p + b G_n \tau_f^n) = \Delta \sigma. \quad (80.9)$$

Величина $\sigma_{св}$ зависит от длины волны падающего света и от его интенсивности посредством зависимости от λ и J величин G и τ_f . Скорость генерации G зависит от λ и J непосредственно, а τ_f — посредством зависимости времени жизни от избыточной концентрации, которая зависит в свою очередь от скорости генерации.

Скорость генерации G определяется интенсивностью света J и коэффициентом поглощения α . Действительно, в объеме $1 dx$ в единицу времени поглощается энергия $-dJ$, равная

$$-dJ = \alpha J dx. \quad (80.10)$$

В единице объема поглощается энергия

$$-\frac{dJ}{dx} = \alpha J. \quad (80.11)$$

Если выразить интенсивность света J через поток фотонов $q = \frac{J}{\hbar\omega}$, то число поглощенных фотонов в единице объема в единицу времени равно

$$-\frac{1}{\hbar\omega} \frac{dJ}{dx} = (-)q_1 = \frac{\alpha J}{\hbar\omega} = \alpha q. \quad (80.12)$$

Таким образом, число поглощаемых в единицу времени в единице объема фотонов q_1 равно произведению коэффициента α и потока фотонов q . Пусть каждый поглощенный фотон с вероятностью η порождает свободный носитель заряда (или пару частиц). Величина ηq_1 представляет собой скорость генерации носителей заряда:

$$G_n = \eta_n q_1 = \eta_n \alpha q, \quad (80.14)$$

$$G_p = \eta_p q_1 = \eta_p \alpha q. \quad (80.15)$$

Величины η_n и η_p называют *квантовым выходом фотоионизации*. Если световая проводимость обусловлена поглощением локализован-

ными состояниями, то одна из величин — η_n или η_p — равна нулю. Из физических соображений ясно, что величина η может принимать одно из двух возможных значений: 0 или 1. Других значений с точки зрения протекания элементарных процессов быть не должно. Однако любая физическая величина имеет смысл только в том случае, если указан метод ее измерения. Очевидно, что величину η можно измерить только по величине δn или δp . Экспериментально измеряемая величина η имеет различные значения — как меньше единицы, так и больше единицы. Однако это не означает, что один фотон непосредственно может генерировать несколько свободных носителей заряда. Это означает только, что в полупроводнике происходят какие-то вторичные эффекты, в результате которых на каждый поглощенный квант приходится эффективное число свободных носителей η_n и η_p , превосходящее единицу. Учитывая выражение для скорости генерации носителей заряда, запишем выражение для плотности тока, обусловленного световой проводимостью, — плотность фототока j_Φ :

$$j_\Phi = \sigma_{св} E = e_p \mu_p (\eta_p \tau_f^p + b \eta_n \tau_f^n) \alpha q E. \quad (80.16)$$

Если через l обозначить размеры полупроводника в направлении поля и через V — напряжение на нем, то $E = V/l$, $\mu_p E = v_{dp}$, $\mu_n E = v_{dn}$. Время дрейфа $t_n = l/v_{dn}$ и $t_p = l/v_{dp}$. Выразим напряженность поля через время, в течение которого носители заряда проходят через полупроводник:

$$E = \frac{v_{dp}}{\mu_p} = \frac{l}{\mu_p t_p} = - \frac{l}{\mu_n t_n}, \quad (80.17)$$

и подставив (80.17) и (80.16), получим

$$j_\Phi = \left(e_p \mu_p \eta_p \tau_f^p \frac{l}{\mu_p t_p} - e_n \mu_n \eta_n \tau_f^n \frac{l}{\mu_n t_n} \right) \alpha q = e_p \left(\eta_p \frac{\tau_f^p}{t_p} + \eta_n \frac{\tau_f^n}{t_n} \right) \alpha q l. \quad (80.18)$$

Умножим выражение (80.18) на сечение S полупроводника; Sl есть объем полупроводника, $j_\Phi S = I$ — фототок. Если образец достаточно тонкий, а поглощение света достаточно слабое, то αq во всех точках полупроводника одинаково. Обозначим через q_v полное число поглощенных фотонов во всем объеме полупроводника и через A' обозначим величину

$$A' = \eta_n \frac{\tau_f^n}{t_n} + \eta_p \frac{\tau_f^p}{t_p}. \quad (80.19)$$

В таком случае фототок равен

$$I_\Phi = e_p q_v A'. \quad (80.20)$$

Если $\eta_n = \eta_p = \eta$ (или один из них равен нулю), то величина

$$q_v \eta = G \quad (80.21)$$

представляет собой полную генерацию носителей заряда в единицу времени, а величина

$$A' = \frac{\tau_f^n}{t_n} + \frac{\tau_f^p}{t_p} \quad (80.22)$$

носит название *коэффициента усиления*. Выражение для фототока может быть записано в виде

$$I_\Phi = e_p G A'. \quad (80.23)$$

Оно называется *вторым характеристическим соотношением для фоторезистивного эффекта*.

Если поле достаточно велико, то время пролета $t = l/v_d$ может стать меньше времени жизни. Выражение (80.23) справедливо только в том случае, если контакты омические, поскольку для них вместо каждого ушедшего из объема через контакт носителя заряда входит такая же частица через другой контакт. Соотношение (80.22) представляет собой коэффициент использования световых носителей заряда.

Если, кроме η , все величины, входящие в (80.18), известны, то, измеряя I_Φ , можно определить квантовый выход. На рис. 132, а приведена спектральная характеристика квантового выхода для германия, из которой видно, что вплоть до 2,7 эВ квантовый выход равен единице. При дальнейшем росте энергии фотона квантовый выход увеличивается. Этот факт можно объяснить следующим образом.

При поглощении фотона в результате прямого перехода квазинимпульсы электрона и дырки равны, и их кинетические энергии при этом обратно пропорциональны эффективным массам. Когда кинетическая энергия одной из частиц достигает величины, равной ширине запрещенной зоны, то «горячий» носитель заряда может отдать свою энергию на образование дополнительной пары свободных электрона и дырки. Если $m_n^* \approx m_p^*$, то граница роста квантового выхода должна лежать в области $\hbar\omega \approx 3\Delta E_0$. Если $m_p^* \gg m_n^*$, то эта граница смещается в область $\hbar\omega \approx 2\Delta E_0$. При повышении температуры ширина запрещенной зоны уменьшается и граница роста η смещается в сторону меньших энергий, как это и видно в случае

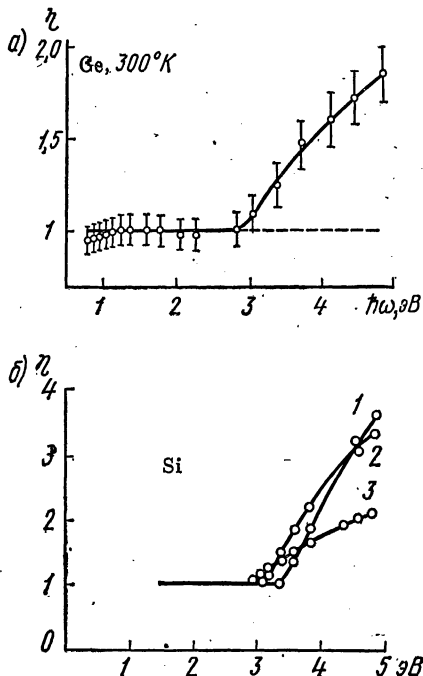


Рис. 132. Спектральная зависимость квантового выхода в германии (а) и влияние температуры на спектральную зависимость квантового выхода кремния (б): 1 — 100 К; 2 — 300 К; 3 — 400 К

кремния (рис. 132, б). Таким образом, *превышение квантового выхода над единицей является вторичным эффектом, оно вызвано ударной ионизацией, а не непосредственным рождением двух пар носителей заряда одним фотоном.*

Величина световой проводимости зависит от интенсивности света по-разному для разных случаев рекомбинации. Если рекомбинация линейная, то избыточная концентрация пропорциональна интенсивности света и

$$\sigma_{св} \sim J; I_{\phi} \sim J, \quad (80.24)$$

т. е. фототок I_{ϕ} пропорционален интенсивности света J . При квадратичной рекомбинации

$$\sigma_{св} \sim \sqrt{J}, I_{\phi} \sim \sqrt{J}. \quad (80.25)$$

В общем случае можно считать, что

$$\sigma_{св} \sim J^{\gamma}. \quad (80.26)$$

При $\gamma = 1$ фоторезистивный эффект называют линейным, при $\gamma < 1$ — нелинейным и при $\gamma > 1$ — сверхлинейным.

Однако и при линейной рекомбинации зависимость $\sigma_{св}$ или $\delta n_{стac}$ от J имеет более сложный вид. Как было показано в § 65, время жизни τ_{∞} может быть больше или меньше τ_0 , где τ_0 — время жизни пары носителей заряда при малом уровне, а τ_{∞} — при большом уровне генерации. Поэтому в области малых и больших интенсивностей $\delta n \sim J$, но с разным наклоном прямых, в области «средних» интенсивностей возможен нелинейный или сверхлинейный ход δn с ростом J .

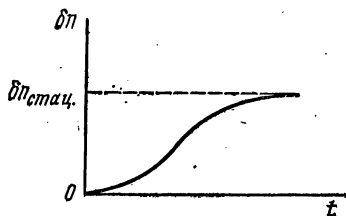


Рис. 133. S-образное нарастание стационарной концентрации неравновесных носителей заряда

Рассматривая процессы нарастания и спада стационарной концентрации неравновесных носителей заряда, мы показали, что она определяется экспоненциальной или тангенсоидальной (и гиперболической) зависимостью от времени.

Однако в ряде случаев кинетика фоторезистивного эффекта отличается от разобранных случаев. Кривые нарастания имеют S-образный вид (рис. 133), при этом чем меньше интенсивность импульса света, тем сильнее выражен S-образный характер кривой нарастания избыточной концентрации, другими словами, тем медленнее происходит переход к стационарной избыточной концентрации. Объяснение подобных фактов удастся получить с помощью понятий *ловушек захвата*.

Предположим, что имеются ловушки захвата с концентрацией M , они могут захватывать и удерживать частицы одного типа в течение времени θ . При этом возможно: $\theta \ll \tau_f$; $\theta \gg \tau_f$. В первом случае ловушка может захватывать частицу неоднократно (α -тип ловушек), во втором случае — не более одного раза (β -тип ловушек). Установле-

ние стационарного состояния можно описать на основе уравнения непрерывности, в которое необходимо добавить члены, описывающие обмен носителями заряда между ловушками захвата и зонами энергии. Опуская соответствующие выкладки, ограничимся некоторыми качественными соображениями. Если ловушки свободны, то в первый момент они будут захватывать, например, электроны, уменьшая скорость нарастания стационарной концентрации и затягивая время нарастания. Аналогично будет происходить затягивание и спад стационарной концентрации за счет опустошения ловушек. Начальный этап нарастания и спада должен определяться минимальным характерным временем, т. е. временем захвата θ , в случае α -ловушек и временем жизни τ_f — в случае β -ловушек. Однако по мере насыщения ловушек носителями заряда их роль в установлении стационарной концентрации уменьшается, чем и объясняется S-образный вид кривых нарастания стационарной концентрации.

Наличие ловушек позволяет понять, например, влияние «подсветки»: если вещество освещать длинноволновым светом, приводящим к фотоионизации ловушек захвата, то световая проводимость возрастает.

Концепция ловушек очень широко используется в настоящее время в теории фоторезистивного эффекта. Предполагая наличие ловушек различного типа и наделяя их определенными свойствами, можно объяснить практически любые экспериментальные факты. Рассмотрим в качестве иллюстрации один пример. Пусть в полупроводнике имеются ловушки захвата, лежащие ниже уровня Ферми. Предположим, что они обладают некоторыми значениями сечений захвата свободных дырок σ_p^* и свободных электронов σ_n^* . При равновесии, квазиуровни Ферми F_n^* , F_p^* , F_M^* совпадают. Если при этом концентрация доноров велика, то уровень Ферми лежит вблизи E_c , поэтому ловушка захвата заполнена электронами. Если полупроводник облучить светом $\hbar\omega \geq \Delta E_0$, то в результате переходов электронов из валентной зоны в зону проводимости концентрация электронов и дырок возрастает на одну и ту же величину. Нарушение термодинамического равновесия приводит к тому, что квазиуровни смещаются: F_n^* поднимается вверх, F_p^* опускается вниз. Смещение F_M^* будет определяться соотношением между σ_p^* и σ_n^* .

Если $\sigma_p^* > \sigma_n^*$, то ловушки с большей вероятностью захватывают дырки, чем электроны, — F_M^* смещается вниз до тех пор, пока не установится равновесие между зоной и дискретным уровнем. Концентрация электронов возрастает в большей степени, чем дырок, — световая проводимость будет в основном электронной (монополярной).

Если $\sigma_n^* > \sigma_p^*$, то уровень Ферми F_M^* смещается вверх, ловушки заполняются электронами, концентрация дырок возрастает в большей степени, чем электронов. Световая проводимость (а возможно и полная, если $\sigma_{св} > \sigma_n$) становится дырочной. Таким образом объясняется монополярный фоторезистивный эффект при собственном поглощении света,

Аналогичным образом можно объяснить и целый ряд других результатов, получаемых при исследовании кинетики фоторезистивного эффекта.

Спектральная зависимость фотопроводимости в первую очередь определяется спектральной зависимостью скорости генерации.

Для примесного внутреннего фотоэффекта кривые $\sigma_{св}(\lambda)$ достаточно хорошо совпадают с кривыми поглощения $\alpha(\lambda)$ (рис. 134 и 127). В ряде случаев спектр $\sigma_{св}(\lambda)$ имеет вид более широких полос, чем полосы поглощения (рис. 135).

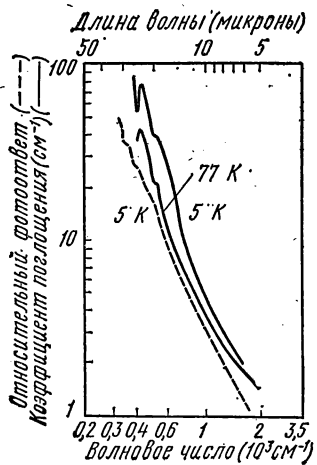


Рис. 134. Коэффициент поглощения и относительный фототок для германия *n*-типа, легированного сурьмой

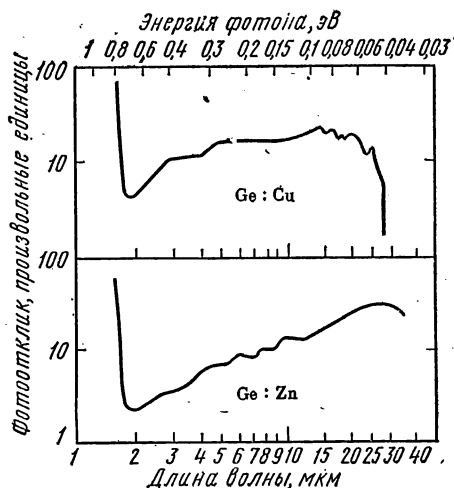


Рис. 135. Спектральная зависимость фоторезистивного эффекта (в относительных единицах) в германии с примесью меди и цинка

Спектральная зависимость фоторезистивного эффекта, соответствующего собственному поглощению, отличается от спектров собственного поглощения. Положение границы собственного фоторезистивного эффекта соответствует границе собственного поглощения, однако с ростом энергии фотонов спектральная кривая внутреннего фотоэффекта проходит через максимум и быстро уменьшается, несмотря на то, что поглощение света в этой области велико (рис. 136):

$$G(x) = \eta \alpha_0 e^{-\alpha x} = G(0) e^{-\alpha x}. \quad (80.27)$$

Чтобы это понять, необходимо учесть, что вследствие сильного поглощения света скорость генерации резко уменьшается по мере проникновения света в глубь вещества.

При $x \gg \alpha^{-1}$ $G(x) \approx 0$, поэтому проводимость не увеличивается. Отсюда следует, что световая проводимость должна зависеть от геометрических размеров образца.

Характеристикой вещества является фоточувствительность S_{ϕ} , которая определяется как отношение световой проводимости $\sigma_{св}$ к интенсивности света J :

$$S_{\phi} = \frac{\sigma_{св}}{J}. \quad (80.28)$$

В СИ S_{ϕ} измеряется в $\text{Ом}^{-1} \text{м}^{-1} / (\text{Вт} \cdot \text{м}^{-2}) = \text{м} / (\text{Ом} \cdot \text{Вт})$. Спектральная характеристика фоторезистивного эффекта должна определяться

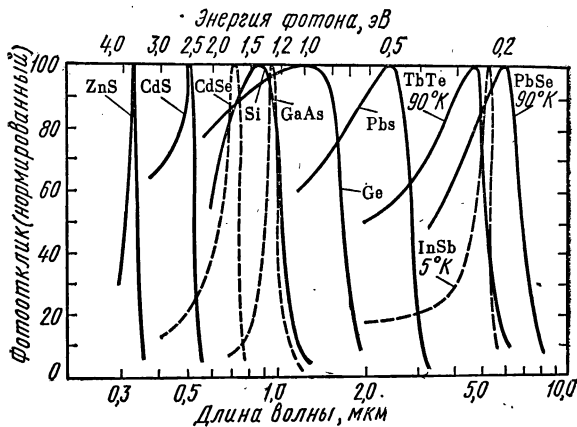


Рис. 136. Спектры фоторезистивного эффекта в области собственного поглощения

именно как зависимость фоточувствительности от частоты или длины волны света. Величину фототока J_{ϕ} называют иногда фотооткликом или фотоответом.

Резюме § 80

1. Фоторезистивный эффект, или внутренний фотоэффект, состоит в изменении сопротивления (или проводимости) полупроводника, вызванного непосредственным действием излучения.

2. Изменение удельной проводимости $\delta\sigma = \sigma_{св}$ называется световой проводимостью, или фотопроводимостью. Отношение световой проводимости к интенсивности падающего света J называется фоточувствительностью

$$S_{\phi} = \frac{\sigma_{св}}{J}. \quad (80.1p)$$

3. Световая проводимость $\sigma_{св}$ определяется избыточной концентрацией основных и неосновных носителей заряда:

$$\sigma_{св} = e_n \mu_n \delta n + e_p \mu_p \delta p. \quad (80.2p)$$

В стационарном состоянии избыточная концентрация носителей заряда определяется первым характеристическим соотношением:

$$\delta n = G_n \tau_f^n = \eta_n \alpha q \tau_f^n, \quad (80.3p)$$

$$\delta p = G_p \tau_f^p = \eta_p \alpha q \tau_f^p, \quad (80.4p)$$

$$\sigma_{св} = (e_n \mu_n \eta_n \tau_n + e_p \mu_p \eta_p \tau_p) \alpha q. \quad (80.5p)$$

4. Фототок I_Φ определяется выражением

$$I_\Phi = S \sigma_{св} E = \left(\frac{\eta_n \tau_f^n}{t_n} + \frac{\eta_p \tau_f^p}{t_p} \right) e_p S l \alpha q, \quad (80.6p)$$

или

$$I_\Phi = e_p A' G, \quad (80.7p)$$

где $G = S l \alpha q$ — полное число фотонов, поглощенных в объеме Sl

полупроводника, $A' = \frac{\eta_n \tau_f^n}{t_n} + \frac{\eta_p \tau_f^p}{t_p}$ — «коэффициент усиления»;

t_n и t_p — время пролета электронов и дырок, определяемые условием: $\mu_n E t_n = \mu_p E t_p = l$

Соотношение (80.7p) называется вторым характеристическим соотношением фоторезистивного эффекта.

5. Зависимость световой проводимости от частоты света определяется спектром поглощения $\alpha(\omega)$ и косвенно — временем жизни.

6. Для объяснения кинетики фоторезистивного эффекта используется понятие ловушек захвата, обладающих различными характеристиками.

7. Фоторезистивный эффект используется для создания фотоэлектрических приборов — фоторезисторов, или фотоспротивлений, позволяющих преобразовывать энергию излучения в электрическую.

§ 81. ЭФФЕКТ ДЕМБЕРА. ФОТОГАЛЬВАНИЧЕСКИЙ ЭФФЕКТ

Рассмотрим однородный полупроводник, на который падает свет. Интенсивность света уменьшается по мере его проникновения в глубь полупроводника в соответствии с законом Бугера — Ламберта:

$$J(x) = J(0) e^{-\alpha x}; \quad q(x) = q(0) e^{-\alpha x}. \quad (81.1)$$

Поглощение света приводит к генерации носителей заряда, скорость которой падает по экспоненте:

$$G(x) = G(0) e^{-\alpha x}; \quad G(0) = \eta \alpha q(0). \quad (81.2)$$

Неравномерная генерация частиц приводит к тому, что они диффундируют в глубь полупроводника. Но так как коэффициенты диффузии электронов и дырок различны, то происходит пространственное разделение подвижных носителей заряда — электроны, имеющие, как правило, большую подвижность, уходят в глубь полупроводника в большей мере, чем дырки. Освещенная поверхность заряжается

положительно, неосвещенная — отрицательно, возникает электрическое поле, направленное вдоль луча света.

Возникновение электрического поля в однородном освещенном полупроводнике называют эффектом Дембера, или кристалл-фотоэффектом. Найдем напряженность поля, возникающего вследствие эффекта Дембера, исходя из кинетического уравнения, которое в использованной нами форме справедливо для стационарных состояний, при условии, что в уравнение необходимо подставлять неравновесные значения кинетических коэффициентов и других величин.

Итак, предположим, что имеем однородный полупроводник, находящийся в изотермических условиях, $\mathbf{V} = 0$. Для плотности тока запишем уравнение

$$\mathbf{j} = \sum_{\alpha} e_{\alpha} K_{11(\alpha)} \mathbf{E} - \sum_{\alpha} e_{\alpha} K_{11(\alpha)} \nabla F_{\alpha} \quad (81.3)$$

Эффект Дембера наблюдается в полупроводнике в отсутствие тока: $\mathbf{j} = 0$ и, следовательно,

$$\mathbf{E} = \frac{\sum_{\alpha} e_{\alpha} K_{11(\alpha)} \nabla F_{\alpha}}{\sum_{\alpha} e_{\alpha}^2 K_{11(\alpha)}}. \quad (81.4)$$

Предположим, что в полупроводнике имеются электроны ($\alpha = 1$) и дырки ($\alpha = 2$):

$$e_1^2 K_{11(1)} = e_n \mu_n n = \sigma_n; \quad e_2^2 K_{11(2)} = e_p \mu_p p = \sigma_p. \quad (81.5)$$

Кроме того, согласно (66.5)

$$\nabla F_{\alpha} = kT \frac{\nabla n_{\alpha}}{n_{\alpha}}, \quad (81.6)$$

поэтому

$$\begin{aligned} \mathbf{E} &= \frac{\sum_{\alpha} e_{\alpha} \frac{n_{\alpha} \langle \tau_{\alpha} \rangle}{m_{\alpha}^*} \frac{\nabla n_{\alpha}}{n_{\alpha}} kT}{\sigma_n + \sigma_p} = \frac{\sum_{\alpha} \mu_{\alpha} \nabla n_{\alpha} kT}{\sigma_n + \sigma_p} = \frac{\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla n_{\alpha}}{\sigma} = \\ &= \frac{-\sum_{\alpha} \mathbf{j}_{D\alpha}}{\sum_{\alpha} \sigma_{\alpha}} = -\frac{\mathbf{j}_{Dn} + \mathbf{j}_{Dp}}{\sigma_n + \sigma_p}. \end{aligned} \quad (81.7)$$

Уравнение (81.7) показывает, что при неоднородном освещении возникают диффузионные потоки электронов и дырок, направленные в одну сторону — в глубь полупроводника, но диффузионные токи \mathbf{j}_{Dn} и \mathbf{j}_{Dp} направлены в противоположные стороны. Возникающее разделение зарядов порождает электрическое поле \mathbf{E} , которое вызывает дрейфовый ток, компенсирующий диффузионный ток. Электрическое поле \mathbf{E} определяется разностью диффузионных потоков электронов и дырок.

Полагая потенциал на неосвещенной стороне полупроводника равным нулю ($\varphi(\infty) = 0$), найдем потенциал на освещенной поверхности $\varphi(0)$:

$$-\int_0^{\infty} (\text{Edl}) = \varphi(\infty) - \varphi(0) = -\varphi(0), \quad (81.8)$$

$$\varphi(0) = \int_0^{\infty} \frac{e_n D_n \frac{dn}{dx} + e_p D_p \frac{dp}{dx}}{\sigma} dx. \quad (81.9)$$

Чтобы вычислить этот интеграл, необходимо знать $\sigma(x)$, поскольку проводимость полупроводника в объеме изменяется вследствие диффузии в объем неравновесных носителей заряда. Рассмотрим слабую генерацию, такую, что можно считать $\sigma \approx \sigma_0$, в этом случае интеграл вычисляется элементарно:

$$\varphi(0) = \frac{e_n D_n}{\sigma_0} \int_0^{\infty} dn + \frac{e_p D_p}{\sigma_0} \int_0^{\infty} dp = \frac{-e_n D_n \delta n(0) - e_p D_p \delta p(0)}{\sigma}. \quad (81.10)$$

Если генерация вызвана собственным поглощением, то $\delta n(0) = \delta p(0) = \alpha q \eta$ и

$$\begin{aligned} \varphi(0) &= \mathcal{E}^D = \frac{e_p \delta p(0)}{\sigma} (D_n - D_p) = -\frac{kT \delta p(0)}{\sigma} (\mu_n + \mu_p) = \\ &= \frac{kT}{\sigma} (|\mu_n| - \mu_p) \alpha q \eta. \end{aligned} \quad (81.11)$$

Э. д. с. эффекта Дембера \mathcal{E}^D определяется разностью модулей подвижностей, $\mathcal{E}^D = 0$ при $|\mu_n| = \mu_p$; она пропорциональна интенсивности света J , пропорциональность сохраняется до тех пор, пока можно считать $\sigma \approx \sigma_0$. Но при больших интенсивностях проводимость σ будет возрастать, а рост \mathcal{E}^D — замедляться.

В связи с эффектом Дембера отметим одну особенность в явлении диффузии неравновесных носителей заряда. Поле Дембера направлено так, что оно препятствует диффузии более подвижных носителей заряда и способствует диффузии менее подвижных носителей заряда. Само явление называется биполярной диффузией. Оно может быть описано на основе уравнения непрерывности подобно тому, как это было проведено при рассмотрении монополярной диффузии в § 66. С этой целью подставим выражение для токов j_n и j_p в уравнения (80.3) и (80.4), для стационарного состояния получим-

$$D_n \nabla^2 n + \mu_n n \operatorname{div} \mathbf{E} + \mu_n (\nabla n \mathbf{E}) + G_n - \frac{n - n_0}{\tau_n} = 0 \quad (81.12)$$

и

$$D_p \nabla^2 p + \mu_p p \operatorname{div} \mathbf{E} + \mu_p (\nabla p \mathbf{E}) + G_p - \frac{p - p_0}{\tau_p} = 0. \quad (81.13)$$

Уравнения (81.12) и (81.13) связаны между собой посредством поля \mathbf{E} :

$$\operatorname{div} \mathbf{E} = -\frac{4\pi}{\varepsilon} \rho = -\frac{4\pi}{\varepsilon} [e_n \delta n + e_p \delta p]. \quad (81.14)$$

Рассмотрим решение уравнений (81.12) и (81.13) для одного частного, но практически важного случая; в объеме обеспечивается электронейтральность: $\rho = 0$; $\delta n = \delta p$, поэтому $\operatorname{div} \mathbf{E} = 0$. Поскольку $\delta n = n - n_0 = \delta p = p - p_0$, то и $\tau_f^n = \tau_f^p = \tau_f$ — время жизни электронов, дырок и пары носителей заряда равны.

Умножив (81.12) на σ_{p0} , а (81.13) на σ_{n0} и сложив их, получим

$$(\sigma_{p0} D_n + \sigma_{n0} D_p) \nabla^2 n + (\mu_n \sigma_{p0} + \mu_p \sigma_{n0}) (\mathbf{E} \nabla n) + G_n \sigma_{p0} + G_p \sigma_{n0} - \frac{n - n_0}{\tau_f} (\sigma_{p0} + \sigma_{n0}) = 0, \quad (81.15)$$

или в одномерном случае

$$\frac{D_n \sigma_{p0} + D_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} \frac{d^2 (n - n_0)}{dx^2} + \frac{(\mu_n \sigma_{p0} + \mu_p \sigma_{n0})}{\sigma_{p0} + \sigma_{n0}} \mathbf{E} \frac{d(n - n_0)}{dx} + G_n \sigma_{p0} + G_p \sigma_{n0} - \frac{n - n_0}{\tau_f} (\sigma_{p0} + \sigma_{n0}) = 0. \quad (81.16)$$

Введем обозначения: коэффициент биполярной диффузии

$$D = \frac{D_n \sigma_{p0} + D_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}}; \quad (81.17)$$

биполярная дрейфовая подвижность

$$\mu_E = \frac{\mu_n \sigma_{p0} + \mu_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}}. \quad (81.18)$$

Как видно из (81.18) и (81.17), μ_E и D не удовлетворяют уравнению Эйнштейна, ему будет удовлетворять величина μ_D :

$$\mu_D = \frac{e_p}{kT} D = \frac{\mu_p \sigma_{n0} - \mu_n \sigma_{p0}}{\sigma_{p0} + \sigma_{n0}} \quad (81.19)$$

— биполярная диффузионная подвижность.

Важной особенностью величин D , μ_E и μ_D является их зависимость от соотношения между концентрациями n_0 и p_0 .

В собственном полупроводнике ($n_0 = p_0$) величины D , μ_E и μ_D равны соответственно

$$D = 2 \frac{D_p D_n}{D_p + D_n}; \quad \frac{1}{D} = \frac{1}{2} \left(\frac{1}{D_p} + \frac{1}{D_n} \right); \quad (81.20)$$

$$\mu_E = 0; \quad (81.21)$$

$$\mu_D = 2 \frac{\mu_p |\mu_n|}{\mu_p + |\mu_n|}; \quad \frac{1}{\mu_D} = \frac{1}{2} \left(\frac{1}{\mu_p} + \frac{1}{|\mu_n|} \right). \quad (81.22)$$

В области примесной проводимости, например при $\sigma_{p0} \ll \sigma_{n0}$,

$$D = D_p; \quad \mu_E = \mu_p; \quad \mu_D = \mu_p \quad (81.23)$$

и при $\sigma_{n0} \ll \sigma_{p0}$,

$$D = D_n; \quad \mu_E = \mu_n; \quad \mu_D = -\mu_n. \quad (81.24)$$

Таким образом, в области примесной проводимости величины D и μ_E совпадают с соответствующими характеристиками неосновных носителей заряда, а μ_D представляет собой подвижность неосновных носителей заряда по модулю.

Учитывая обозначения (81.17) и (81.18), перепишем (81.16):

$$D \frac{d^2(n-n_0)}{dx^2} + \mu_E E \frac{d(n-n_0)}{dx} - \frac{n-n_0}{\tau_f} = G, \quad (81.25)$$

где

$$G = \frac{G_n \sigma_{p0} + G_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}} \quad (81.26)$$

— скорость «биполярной» генерации.

Уравнение (81.25) при $G = 0$ совпадает с уравнением (66.15). Обозначая

$$L^2 = D\tau_f, \quad (81.27)$$

$$\frac{\mu_E E}{D} = \frac{\mu_E E \tau_f}{D\tau_f} = \frac{l_E}{L^2}, \quad (81.28)$$

получим для (81.25)

$$\frac{d^2(n-n_0)}{dx^2} + \frac{l_E}{L^2} \frac{d(n-n_0)}{dx} - \frac{n-n_0}{L^2} = -\frac{G}{D}. \quad (81.29)$$

Общее решение однородного уравнения может быть записано, например, в виде

$$(n-n_0) = Ae^{-\frac{x}{l}} + Be^{\frac{x}{l}}, \quad (81.30)$$

а частное решение неоднородного уравнения может быть найдено при известной зависимости $G(x)$. Если носители заряда генерируются светом, то

$$G(x) = G(0) e^{-\alpha x}. \quad (81.31)$$

В этом случае частное решение можно представить в виде

$$(n-n_0) = -\frac{G(0) e^{-\alpha x}}{D \left[\alpha^2 - \frac{\alpha l_E}{L^2} - \frac{1}{L^2} \right]}, \quad (81.32)$$

что проверяется непосредственной подстановкой (81.31) в (81.29).

Для собственного поглощения в большинстве случаев можно считать $G(x) \approx 0$ ($x \neq 0$). Коэффициенты A и B общего решения могут быть найдены из граничных условий для концентраций и токов, с учетом токов поверхностной рекомбинации.

Опустим эти расчеты, они несложны, но громоздки. Отметим одну особенность в распределении $[n - n_0](x)$. При малых полях E распределение неравновесной концентрации определяется диффузионной длиной L . В области больших полей длина затягивания равна длине дрейфа, если полупроводник примесный. Но чем ближе полупроводник к собственному, тем ближе длина затягивания к диффузионной длине. В собственном полупроводнике электрическое поле, независимо от его величины, не влияет на распределение избыточной концентрации.

Перейдем к качественному рассмотрению фотогальванического эффекта.

Как было показано в § 72, в неоднородном полупроводнике существуют объемные поля E^i и объемные заряды, величина которых может быть найдена из соотношения (72.4), которое мы перепишем в виде

$$E^i = - \frac{J_{Dn0} + J_{Dp0}}{\sigma_{p0} + \sigma_{n0}} = \frac{D_p \nabla p_0 - b \nabla n_0}{\mu_p p_0 + b n_0} \quad (81.33)$$

Выражение (81.33) в точности совпадает с (81.7), однако между ними есть принципиальное различие. Поле E^i является термодинамически равновесным, оно компенсирует диффузионный ток, обеспечивая равенство нулю полного тока в объеме полупроводника. Поэтому во внешней цепи поле E^i не может создать электрического тока, в то время как поле Дембера, будучи неравновесным, вызывает электрический ток во внешней цепи.

Предположим теперь, что свет падает на неоднородный полупроводник. Инжектированные светом электрон и дырка находятся в объемном поле E^i , которое приводит их в дрейфовое движение. Но если диффузия приводит к тому, что они двигаются в одном и том же направлении, то поле E^i перемещает их в противоположных направлениях — электроны перемещаются против поля E^i , дырки — по полю E^i .

Разделение неравновесных электронов и дырок, их перемещение в противоположных направлениях приводят к тому, что они вызывают неравновесное электрическое поле E^* , которое направлено противоположно полю E^i . Таким образом, неравновесное объемное поле E^* препятствует разделению неравновесных носителей заряда. Очевидно, что в пределе при $\delta n = \delta p \rightarrow \infty$ объемные поля E^* и E^i полностью скомпенсируют друг друга, поле и объемный заряд обратятся в нуль, потенциальные барьеры, существовавшие в объеме полупроводника, исчезнут (полупроводник станет однородным и собственным $\delta n = n = \delta p = p$, но неравновесным!). Между точками полупроводника, которые разделялись потенциальными барьерами, возникнет разность потенциалов, равная по модулю отношению высоты барьера к заряду электрона. Таким образом, объемные поля приводят к возникновению разности потенциалов между двумя любыми точками полупроводника, находящимися в области объемного поля, при генерации

в ней неравновесных носителей заряда. Неравновесные носители заряда могут проявить себя и в том случае, если их генерация происходит от области полей E^i на расстояниях, меньших диффузионной длины. В этом случае неравновесные носители заряда, диффундируя, попадут в область объемного поля, которое приведет к их разделению.

Рассмотрим один важный случай — генерация носителей заряда происходит в области p - n -перехода, где существует объемный заряд

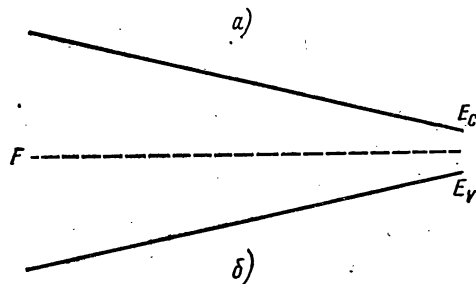
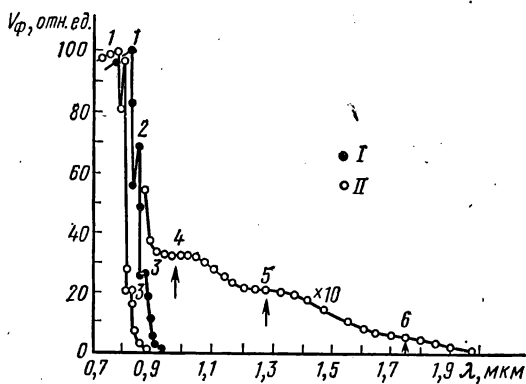


Рис. 137. Спектр фотогальванического эффекта в теллуриде кадмия (а); энергетическая диаграмма варизонной структуры (б)

и объемное поле E_k , направленное от n -области к p -области. Электроны и дырки, генерируемые в n -области и p -области на расстояниях порядка L_p и L_n соответственно от области объемного заряда, будут разделены полем E_k так, что неравновесные электроны будут переведены в n -область, а дырки — в p -область, т. е. контактное поле способствует переходу через область объемного заряда неосновных носителей заряда. Электронная область заряжается отрицательно, дырочная — положительно. Если такой полупроводник включить в замкнутую цепь, то в ней потечет ток, направленный от p -области к n -области, приводящий к уничтожению избыточной концентрации электронов и дырок. Из самого механизма возникновения полей E^* следует, что максимальная ф. э. д. с. на p - n -переходе не может превзойти величину ϕ^* , определяющую высоту потенциального барьера. Это явление лежит в основе работы p - n -фотоэлементов и преобразователей энергии света в электрическую («солнечные батареи»). Фотогальванический эффект может наблюдаться при освещении любых выпрямляющих контактов, он может быть использован для исследования их свойств подобно тому, как неоднородности в полупроводнике могут быть обнаружены с помощью светового зонда.

На рис. 137, а приведен спектр S_λ фотогальванического эффекта в теллуриде кадмия. Длинноволновые полосы фоточувствительности находятся вне области собственного поглощения. Это означает, что

На рис. 137, а приведен спектр S_λ фотогальванического эффекта в теллуриде кадмия. Длинноволновые полосы фоточувствительности находятся вне области собственного поглощения. Это означает, что

фотогальванический эффект может возникать и при монополярной генерации носителей заряда, обеспечивающей возникновение при оптических или термооптических переходах неосновных носителей заряда. На кривой S_{λ} можно наблюдать пик, соответствующий экситонному поглощению. Его возникновение возможно только благодаря диссоциации экситона.

К фотогальваническому эффекту можно отнести явление возникновения электродвижущей силы в твердых растворах, имеющих переменный состав и, тем самым, переменную ширину запрещенной зоны (варизонные структуры). На рис. 137, б приведена энергетическая диаграмма твердого раствора $Cd_xHg_{1-x}Te$, в котором ширина запрещенной зоны может меняться при комнатной температуре от 1,5 эВ в теллуриде кадмия ($x=1$) до 0 в теллуриде ртути ($x=0$). Поскольку экспериментальные значения энергий зависят от координаты, то возникает электрическое поле, действующее на электроны и дырки проводимости:

$$E_n = - \frac{1}{e_n} \frac{dE_c(x)}{dx}, \quad (81.34)$$

$$E_p = - \frac{1}{e_p} \frac{dE_v(x)}{dx}. \quad (81.35)$$

Эти поля различны по величине и противоположны по направлению. Силы, действующие на электроны и дырки, направлены в одну и ту же сторону — от области с большей шириной запрещенной зоны в сторону области с меньшей шириной запрещенной зоны. В условиях термодинамического равновесия эти поля не приводят к возникновению электродвижущей силы во внешней цепи, но при нарушении равновесия, например при поглощении света, происходит дрейф носителей заряда в сторону узкозонной области. Поскольку обычно электроны имеют большую подвижность, то их скорость дрейфа превосходит скорость дрейфа дырок, благодаря чему происходит пространственное разделение зарядов. Следствием этого является возникновение неравновесного электрического поля, вызывающего ток во внешней цепи. Спектр фоточувствительности варизонных структур достаточно широк — от видимой области до 10—20 мкм.

Резюме § 81

1. Диффузия неравновесных носителей заряда противоположных знаков носит название биполярной. Она характеризуется биполярным коэффициентом диффузии

$$D = \frac{D_n \sigma_{p0} + D_p \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}}, \quad (81.1p)$$

где $\sigma_{p0} + \sigma_{n0}$ — равновесная удельная проводимость. Распределение носителей заряда определяется длиной затягивания, которая прак-

тически равна биполярной диффузионной длине L , определяемой уравнением

$$L^2 = D\tau_f = \frac{L_n^2 \sigma_{p0} + L_p^2 \sigma_{n0}}{\sigma_{p0} + \sigma_{n0}}, \quad (81.2p)$$

если поля E слабые, или для сильных полей в почти собственном полупроводнике.

2. Распределение неравновесной концентрации определяется уравнением

$$n(x) - n_0 = p - p_0 = A e^{-\frac{x}{L}} + B e^{\frac{x}{L}} - \frac{G(0) e^{-\alpha x}}{D \left[\alpha^2 - \frac{\alpha E}{L^2} - \frac{1}{L^2} \right]}. \quad (81.3p)$$

Величины A и B определяются из граничных условий.

3. Эффект Дембера, или кристалл-фотоэффект, состоит в возникновении электрического поля в направлении луча сильно поглощаемого света. Напряженность поля, или э. д. с. \mathcal{E}^D , пропорциональна разности коэффициентов диффузии электронов и дырок, поскольку поле Дембера компенсирует разность диффузионных токов электронов и дырок. Если коэффициенты диффузии равны, то токи численно равны и противоположно направлены, и поле Дембера равно нулю.

4. Разделение неравновесных электронов и дырок объемными полями, существующими в неоднородных полупроводниках, приводит к возникновению неравновесных объемных полей, способных создавать ток во внешней цепи. Фотогальванический эффект состоит в возникновении э. д. с. при генерации неравновесных носителей заряда в области p - n -перехода или выпрямляющего контакта металла и полупроводника.

§ 82. ФОТОМАГНИТНОЭЛЕКТРИЧЕСКИЙ ЭФФЕКТ

В полупроводнике, находящемся в магнитном поле, при освещении сильно поглощаемым светом, возникает электрическое поле. Это явление называется эффектом Кикоина—Носкова, или фотомагнитноэлектрическим эффектом (ФМЭЭ).

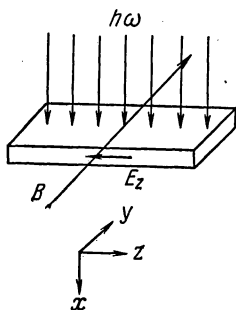


Рис. 138. Схема наблюдения эффекта Кикоина—Носкова

Причиной возникновения электрического поля является отклонение от первоначального направления дрейфовых и диффузионных потоков носителей заряда под действием магнитного поля. При экспериментальном исследовании фотомагнитноэлектрического эффекта свет направляется на одну из граней образца, магнитное поле направлено перпендикулярно световому потоку и другой паре граней, а электрическое поле наблюдается в направлении, перпендикулярном к магнитному полю и световому потоку (рис. 138). Явление Кикоина—Нос-

кова аналогично двум явлениям: эффекту Холла, поскольку в полупроводнике имеется дрейфовый ток, и термомагнитному эффекту Нернста — Эттингсгаузена, поскольку в полупроводнике при неоднородном освещении возникает диффузионный ток.

В стационарных условиях фотомагнитноэлектрический эффект может быть описан на основе кинетического уравнения. Учитывая действие электрического и магнитного полей и градиента химического потенциала, запишем

$$\mathbf{j} = \sum_{\alpha} e_{\alpha}^2 K'_{11(\alpha)} \mathbf{E} - \sum_{\alpha} e_{\alpha} K'_{11(\alpha)} \nabla F_{\alpha} + \left[\left(\sum_{\alpha} \frac{e_{\alpha}^3}{m_{\alpha}^*} K_{12(\alpha)} \mathbf{E} - \sum_{\alpha} \frac{e_{\alpha}^2}{m_{\alpha}^*} K'_{12(\alpha)} \nabla F_{\alpha} \right), \mathbf{B} \right] \quad (82.1)$$

Первый член описывает дрейфовый ток, второй — диффузионный, третий — гальваномагнитный ток Холла, четвертый — магнитнодиффузионный ток. Выражение (82.1) учитывает токи, создаваемые носителями зарядов разных типов. Преобразуем члены, входящие в (82.1), следующим образом:

$$e_{\alpha}^2 K'_{11(\alpha)} = e_{\alpha} \mu'_{\alpha} n_{\alpha} = \sigma'_{\alpha}; \quad (82.2)$$

$$e_{\alpha} K'_{11(\alpha)} \nabla F_{\alpha} = \mu'_{\alpha} n_{\alpha} k T \frac{\nabla n_{\alpha}}{n_{\alpha}} = e_{\alpha} D'_{\alpha} \nabla n_{\alpha}; \quad (82.3)$$

$$\frac{e_{\alpha}^3}{m_{\alpha}^*} K'_{12(\alpha)} = \frac{e_{\alpha}^3 n_{\alpha} \langle \tau'_{\alpha}{}^2 \rangle}{m_{\alpha}^{*2}} = \frac{e_{\alpha} \langle \tau'_{\alpha}{}^2 \rangle}{m_{\alpha}^* \langle \tau'_{\alpha} \rangle} \frac{e_{\alpha} \langle \tau'_{\alpha} \rangle}{m_{\alpha}^*} e_{\alpha} n_{\alpha} = \mu_{\alpha}^H \sigma'_{\alpha}; \quad (82.4)$$

$$\frac{e_{\alpha}^2}{m_{\alpha}^*} K'_{12(\alpha)} \nabla F_{\alpha} = \frac{e_{\alpha}^2}{m_{\alpha}^*} \frac{n_{\alpha} \langle \tau'_{\alpha}{}^2 \rangle}{m_{\alpha}^*} k T \frac{\nabla n_{\alpha}}{n_{\alpha}} = \mu_{\alpha}^H \mu'_{\alpha} k T \nabla n_{\alpha} = e_{\alpha} D'_{\alpha} \nabla n_{\alpha} \mu_{\alpha}^H. \quad (82.5)$$

В выражениях (82.2 — 5) все величины связаны с неравновесными концентрациями, а штрих учитывает величину $(1 + \mu_{\alpha}^2 B^2)$ в знаменателе усредненных времен релаксаций. В области слабых магнитных полей величиной $\mu_{\alpha}^2 B^2$ можно пренебречь, что позволит считать соответствующие величины не зависящими от магнитного поля.

Перепишем (82.1) с учетом (82.2 — 5):

$$\mathbf{j} = \left(\sum_{\alpha} \sigma_{\alpha} \right) \mathbf{E} - \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla n_{\alpha} \right) + \left[\left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \mathbf{E} - \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla n_{\alpha} \mu_{\alpha}^H \right), \mathbf{B} \right] = \sum_{\alpha} \mathbf{j}_{E\alpha} + \sum_{\alpha} \mathbf{j}_{D\alpha} + \left[\sum_{\alpha} \mu_{\alpha}^H \mathbf{j}_{E\alpha}, \mathbf{B} \right] + \left[\sum_{\alpha} \mu_{\alpha}^H \mathbf{j}_{D\alpha}, \mathbf{B} \right]. \quad (82.6)$$

Из (82.6) достаточно отчетливо виден состав тока, включающий ток проводимости, ток Холла, диффузионный и магнитнодиффу-

зионный токи. Учитывая, что $\mathbf{V} = (0, V, 0)$, перепишем (82.6) в проекциях на оси координат:

$$j_x = \sigma E_x - \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) E_z B + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha} \mu_{\alpha}^H \right) B; \quad (82.7)$$

$$j_y = \sigma E_y - \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_y n_{\alpha}; \quad (82.8)$$

$$j_z = \sigma E_z - \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha} + \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) E_x B - \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B. \quad (82.9)$$

Решая систему уравнений (82.7) и (82.9), выразим E_x и E_z . Определитель системы Δ равен:

$$\Delta = \begin{vmatrix} \sigma & \left(- \sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B \\ \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B & \sigma \end{vmatrix} = \sigma^2 + \left(\sum_{\alpha} \sigma_{\alpha} \mu_{\alpha}^H \right)^2 B^2, \quad (82.10)$$

в области слабых магнитных полей ($\mu_{\alpha}^2 B^2 \ll 1$)

$$\Delta = \sigma^2. \quad (82.11)$$

Для E_x и E_z имеем выражения:

$$E_x = \frac{1}{\Delta} \left\{ \sigma \left[j_x + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) - \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B \right] + \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B \left[j_z + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha} \right) + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B \right] \right\}; \quad (82.12)$$

$$E_z = \frac{1}{\Delta} \left\{ \sigma \left[j_z + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha} \right) + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B \right] - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B \left[j_x + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) - \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha} \mu_{\alpha}^H \right) B \right] \right\}. \quad (82.13)$$

При $B=0$ выражения (82.12) и (82.13) примут вид:

$$E_x = \frac{j_x + \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha}}{\sigma}, \quad (82.14)$$

$$E_z = \frac{j_z + \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_z n_{\alpha}}{\sigma}, \quad (82.15)$$

т. е. E_x и E_z представляют собой поле омического падения напряжения и поле, компенсирующее диффузионные потоки. Если $j_x = j_z = 0$, то E_x — поле Дембера, а $E_z = 0$, поскольку $\nabla_z n_{\alpha} = 0$.

Рассмотрим уравнения (82.12) и (82.13) в предположении, что магнитное поле не влияет на распределение носителей заряда

вдоль оси z , т. е. считая, что $\nabla_z n_\alpha = 0$. Пренебрегая членами, содержащими B^2 , запишем

$$E_x = \frac{1}{\sigma^2} \left\{ \sigma \left[j_x + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) \right] + \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) j_z B \right\}; \quad (82.16)$$

$$E_z = \frac{1}{\sigma^2} \left\{ \sigma \left[j_z + \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B \right] - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B \left[j_x + \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right] \right\}. \quad (82.17)$$

В выражении (82.17) содержится описание эффекта Кикоина — Носкова при различных условиях наблюдения. При экспериментальном изучении эффекта Кикоина — Носкова $j_x = 0$. Полагая $j_z = 0$, приходим к уравнению для поля E_z , или соответствующей разности потенциалов:

$$E_z = B \frac{\left(\sum_{\alpha} \sigma_{\alpha} \right) \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right)}{\left(\sum_{\alpha} \sigma_{\alpha} \right)^2}. \quad (82.18)$$

В режиме «короткого замыкания» $E_z = 0$ и из (82.17) находим выражение для плотности тока j_z :

$$-j_z = B \frac{\left(\sum_{\alpha} \sigma_{\alpha} \right) \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right)}{\sigma}. \quad (82.19)$$

Сравнивая (82.18) и (82.19), мы видим, что E_z , измеренная в режиме «э. д. с.», и j_z , измеренный в режиме «короткого замыкания», связаны между собой совершенно очевидным соотношением

$$(-) j_z^{к. з.} = \sigma E_z^{э. д. с.} \quad (82.20)$$

Для общности рассмотрим режим «двойного короткого замыкания», при котором $E_z = E_x = 0$. Выразим j_x из (82.16):

$$-j_x = \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) + \frac{\left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B j_z}{\sigma} \quad (82.21)$$

и подставим в (82.17):

$$0 = \sigma j_z + \sigma \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B - \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) B + \left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) B \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \right) = \sigma j_z + \sigma \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B, \quad (82.22)$$

или

$$j_z = - \left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B. \quad (82.23)$$

В режиме «второго короткого замыкания» ($E_x = 0$) можно найти E_z при $j_z = 0$.

Из (82.16) имеем

$$j_x = - \sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \quad (82.24)$$

и из (82.17) с учетом (82.24) получим

$$E_z = \frac{\left(\sum_{\alpha} e_{\alpha} D_{\alpha} \nabla_x n_{\alpha} \mu_{\alpha}^H \right) B}{\sigma} \quad (82.25)$$

в полном соответствии с (82.23).

Уравнения (82.16) и (82.17) описывают более общий случай чем разобранные выше, поскольку можно пропускать через исследуемый образец ток j_z и одновременно с этим измерять E_z .

Запишем выражение для E_z при условии $j_x = j_z = 0$ с учетом генерации электронов и дырок:

$$\begin{aligned} E_z &= \frac{e_n D_n \nabla_x n \mu_n^H + e_p D_p \nabla_x p \mu_p^H}{\sigma} B - \frac{\left(\sum_{\alpha} \mu_{\alpha}^H \sigma_{\alpha} \right) (e_n D_n \nabla_x n + e_p D_p \nabla_x p)}{\sigma^2} B = \\ &= \frac{e_p \nabla_x p (D_p \mu_p^H - D_n \mu_n^H)}{\sigma} B - \frac{(\sigma_p \mu_p^H + \sigma_n \mu_n^H) e_p \nabla_x p (D_p - D_n)}{\sigma^2} B. \end{aligned} \quad (82.26)$$

В области примесного поглощения возникает монополярная диффузия и, например, при $\nabla_x n = 0$ имеем

$$E_z = \frac{e_p D_p \nabla_x p \mu_p^H}{\sigma_n + \sigma_p} B - \frac{(\sigma_n \mu_n^H + \sigma_p \mu_p^H) e_p D_p \nabla_x p}{(\sigma_n + \sigma_p)^2} B. \quad (82.27)$$

В области примесной проводимости выражение (82.27) упрощается, поскольку можно пренебречь проводимостью, создаваемой неосновными носителями заряда. Предполагая, что образец имеет дырочную проводимость, получим

$$E_z = B \frac{e_p D_p \nabla_x p}{\sigma_p} \left(\mu_p^H - \frac{\sigma_p \mu_p^H}{\sigma_p} \right) = 0. \quad (82.28)$$

Таким образом, в случае примесного поглощения, приводящего к генерации основных носителей заряда, поле Кикоина — Носкова равно нулю. Предположим теперь, что образец имеет электронную проводимость и в нем генерируются неосновные носители заряда:

$$E_z = \frac{B e_p D_p \nabla_x p}{\sigma_n} (\mu_p^H - \mu_n^H) \neq 0. \quad (82.29)$$

Интересно отметить, что в режиме $E_x = 0$ поле Кикоина — Носкова отлично от нуля в любом случае. Из (82.25) имеем

$$E_z = \frac{e_p D_p \nabla_x p \mu_p^H B}{\sigma_n + \sigma_p} \neq 0. \quad (82.30)$$

Как видно из выражений для E_z , поле Кикоина — Носкова в области слабых интенсивностей света растет линейно с ростом J . Действительно, пока $\delta n = \delta p \ll n_0 + p_0$, $E_z \sim J$, но по мере роста J $\delta n = \delta p$ может превзойти $n_0 + p_0$, и E_z достигает насыщения, величина которого линейно зависит от B , однако с ростом B линейность нарушается. Характер изменения E_z при больших полях можно получить из (82.12) и (82.13) при $B \rightarrow \infty$.

Фотомагнитноэлектрический эффект широко используется для исследования различных неравновесных процессов в полупроводниках. При теоретическом анализе экспериментальных результатов необходим тщательный анализ условий эксперимента, поскольку в различных условиях могут быть получены различные результаты.

Ограничимся одним примером практического применения эффекта Кикоина — Носкова.

Распределение избыточной концентрации по толщине образца определяется диффузионной длиной

$$\delta n(x) = \delta n(0) e^{-\frac{x}{L}}, \quad (82.31)$$

что дает для производной

$$\nabla_x n(x) = -\frac{\delta n(x)}{L} = -\frac{\delta n(0) e^{-\frac{x}{L}}}{L}. \quad (82.32)$$

Если в условиях опыта достигнуто насыщение $\delta n = \delta p \gg n_0 + p_0$, то $\sigma_n + \sigma_p = e_p \mu_p (1+b) \delta p$, и на основании (82.26) получаем

$$E_z \sim \frac{2D_n(1-b)}{(1+b)^2} \frac{1}{L} = \frac{2D_n(1-b)}{(1+b)^2 \sqrt{D}} \frac{1}{\tau_f}. \quad (82.33)$$

В связи с тем, что $E_z \sim \frac{1}{\sqrt{\tau_f}}$, фотомагнитноэлектрический эффект используется для определения малых времен жизни: $\tau_f \approx (10^{-9} \div 10^{-10})$ с.

§ 83. ЭФФЕКТ ФАРАДЕЯ

Физические явления, обусловленные взаимодействием излучения с веществом, находящимся в магнитном поле, называются магнитно-оптическими. К ним относятся осцилляторное магнетопоглощение, циклотронный резонанс, плазменное магнетоотражение и вращение плоскости поляризации света, или эффект Фарадея. Остановимся на последнем явлении.

Существование магнитно-оптических явлений означает, что оптические свойства вещества изменяются под действием магнитного поля. Характер внесистемных полей изменений в ряде случаев можно понять на основе теоремы Лармора. Пусть квантовая система (атомы, молекулы и т. п.) имеют механический и магнитный моменты,

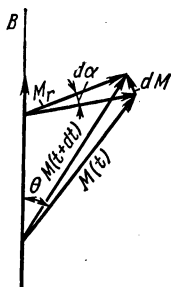
связанные между собой обычным соотношением посредством гиромагнитного отношения G :

$$\mu = GM. \quad (83.1)$$

При наложении магнитного поля B на квантовую систему будет действовать пара сил с моментом N :

$$N = [\mu B], \quad (83.2)$$

стремящаяся повернуть μ вдоль поля. Однако с μ связан механический момент, поэтому изменение направления μ и M означало бы нарушение закона сохранения момента количества движения. Изменение M можно найти, пользуясь основным уравнением вращательного движения:



$$\frac{dM}{dt} = N \quad (83.3)$$

или

$$\frac{dM}{dt} = [\mu B] = G [MB]. \quad (83.4)$$

Считая гиромагнитное отношение скалярной величиной, можем записать

$$dM = G [MB] dt, \quad (83.5)$$

Рис. 139. Возникновение прецессии моментов в магнитном поле

т. е. приращение момента dM перпендикулярно самому моменту M и индукции B магнитного поля, поэтому модуль вектора M не изменяется, а происходит лишь поворот M вокруг B . На рис. 139 показано положение M в два момента времени t и $t + dt$. Угол между M и B не изменяется, расстояние между концом вектора M и осью B сохраняется. Это означает, что вектор M прецессирует вокруг поля B . Мы имеем дело с типичным гироскопическим эффектом.

Найдем угловую скорость прецессии, равную угловой скорости поворота вектора M , вокруг оси B (рис. 139):

$$d\alpha = \omega_L dt = \frac{|dM|}{|M_r|} = \frac{GMB \sin \theta dt}{M \sin \theta} = GB dt, \quad (83.6)$$

или $\omega_L = GB$. Учитывая векторный характер величин и их направление, можем записать в векторной форме:

$$\omega_L = GB. \quad (83.7)$$

Угловая скорость прецессии ω_L называется частотой Лармора.

Если ввести вращающуюся вокруг B со скоростью ω_L систему координат, то в ней M и μ остаются постоянными. Теперь можно

сформулировать следующее: при наложении на физическую систему магнитного поля свойства системы изменяются таким образом, что при переходе к вращающейся системе координат свойства остаются неизменными. Применим это утверждение к эффекту Фарадея.

Экспериментально было найдено, что поворот плоскости поляризации линейно поляризованного света зависит от толщины слоя вещества l и индукции магнитного поля B :

$$\varphi = \theta Bl. \quad (83.8)$$

Величина θ называется коэффициентом Верде. Угол поворота принято считать положительным, если плоскость поляризации поворачивается по часовой стрелке (вправо) при распространении света вдоль магнитного поля. Отсюда следует, что знак вращения меняется при изменении направления поля.

Объяснение эффекта Фарадея основывается на идее Френеля о различии коэффициентов преломления для право- и левовращающейся поляризованных волн. Линейно поляризованную волну при этом можно рассматривать как суперпозицию двух циркулярно поляризованных волн, соответственно по правому (+) и левому (-) кругу. Пройдя слой толщиной l , волны получают разность хода Δ и разность фаз $\Delta\varphi$:

$$\Delta\varphi = \frac{2\pi\Delta}{\lambda} = \frac{2\pi}{\lambda} (ln^+ - ln^-). \quad (83.9)$$

Угол поворота плоскости поляризации равен половине набираемой волнами разности фаз, т. е.

$$\varphi = \frac{1}{2} \Delta\varphi = \frac{\pi}{\lambda} (n^+ - n^-) l = \frac{\omega (n^+ - n^-) l}{2c}. \quad (83.10)$$

Вращение возможно в том случае, когда $n^+ \neq n^-$. Выразим n^+ и n^- через n . Согласно сказанному выше, можно записать

$$n^\pm = n (\omega \mp \omega_L). \quad (83.11)$$

Поскольку $\omega_L \ll \omega$, то можно разложить $n(\omega)$ в ряд, взяв только два члена

$$n(\omega \pm \omega_L) = n(\omega) \pm \frac{dn}{d\omega} \omega_L \quad (83.12)$$

и

$$n^+ - n^- = -2 \frac{dn}{d\omega} \omega_L, \quad (83.13)$$

откуда

$$\varphi = -\frac{\omega l}{2c} \frac{dn}{d\omega} 2\omega_L. \quad (83.14)$$

Учитывая, что $\omega \frac{dn}{d\omega} = -\lambda \frac{dn}{d\lambda}$, перепишем (83.14) в виде:

$$\varphi = \frac{\lambda}{c} \frac{dn}{d\lambda} l \omega_L = \frac{\lambda}{c} \frac{dn}{d\lambda} GlB. \quad (83.15)$$

Это выражение представляет собой одну из форм соотношения Беккереля. Для коэффициента Верде имеем

$$\theta = \frac{\lambda}{c} \frac{dn}{d\lambda} G. \quad (83.16)$$

Для орбитального движения электрона $G = \frac{e}{2mc}$ (в системе Гаусса), поэтому

$$\theta = \frac{\lambda}{2c^2} \frac{e}{m_0} \frac{dn}{d\lambda} \quad (83.17)$$

и

$$\varphi = \frac{\lambda}{2c^2} \frac{e}{m_0} \frac{dn}{d\lambda} lB. \quad (83.18)$$

Поворот плоскости поляризации отличен от нуля только в том случае, когда в данной области спектра имеет место дисперсия света. Но дисперсия тесно связана с поглощением света, а именно, в окрестности линии или полосы поглощения $\frac{dn}{d\lambda} < 0$ — область нормальной дисперсии, в самой полосе поглощения имеет место аномальная дисперсия $-\frac{dn}{d\lambda} > 0$, поэтому при переходе через полосу поглощения следует ожидать изменения знака угла поворота. Инверсия знака должна быть двукратной в соответствии с тем, что при переходе через полосу поглощения $\frac{dn}{d\lambda}$ по крайней мере дважды обращается в нуль.

Поскольку в полупроводниках имеется несколько видов поглощения света, то следует ожидать столько же видов вращения плоскости поляризации, если считать, что каждый тип поглощения обуславливает дисперсию показателя преломления.

Вращение, обусловленное дисперсией вследствие собственного поглощения, называется межзонным вращением. Оно наблюдается обычно вне полосы собственного поглощения при $\frac{\hbar\omega}{\Delta E_0} \sim 0,1 \div 1$. Наблюдение межзонного вращения при $\hbar\omega > \Delta E_0$ осложняется сильным поглощением. Аналогичным образом говорят о вращении на свободных носителях заряда и на экситонах. Вращение на примесных состояниях возможно, хотя практически трудно осуществимо. Наблюдаемый эффект представляет собой суперпозицию вращений, обусловленных различными механизмами.

Рассмотрим вращение плоскости поляризации на свободных носителях заряда. Используем метод тензора комплексной диэлектрической проницаемости. Так как n^+ и n^- отличаются незначительно, то можно записать:

$$n^+ + n^- = 2n \quad (83.19)$$

и

$$\varphi = \frac{\pi l}{2\lambda n} (n^{+2} - n^{-2}). \quad (83.20)$$

Заменяя n^2 на ϵ , получим

$$\varphi = \frac{\pi l}{2\lambda n} (\epsilon^+ - \epsilon^-). \quad (83.21)$$

Как показано в § 74, амплитуда Фурье — компонента поля излучения удовлетворяет уравнению

$$\Delta E(\mathbf{r}, \omega) + \frac{\omega^2 \mu}{c^2} \left[\epsilon - i \frac{4\pi\sigma}{\omega} \right] E(\mathbf{r}, \omega) = 0. \quad (83.22)$$

Обозначим выражение в квадратной скобке через $\tilde{\epsilon}$:

$$\tilde{\epsilon} = \epsilon - i \frac{4\pi\sigma}{\omega}, \quad (83.23)$$

т. е. введем комплексную диэлектрическую проницаемость. При наложении магнитного поля σ становится тензорной величиной (гиротропия), поэтому и $\tilde{\epsilon}$ — величина тензорная (величину ϵ будем считать для простоты скалярной). Если \vec{B} направлено вдоль оси z , то для σ согласно § 42 и 45 можем записать:

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & 0 \\ \sigma_{yx} & \sigma_{yy} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix}, \quad (83.24)$$

где $\sigma_{xx} \approx \sigma_{yy} \approx \sigma_{zz} \approx \sigma(0)$ в случае слабого магнитного поля, σ_{xy} — проводимость Холла, для которой из кинетического уравнения следует

$$\sigma_{xy} = - \frac{e^3}{m^*} K'_{12} B_z \quad (83.25)$$

с учетом ограниченности образца. Из явного выражения для σ_{xy} следует, что $\sigma_{yx} = -\sigma_{xy}$. Действительно, при замене $x \rightarrow y$ $y \rightarrow x$ при сохранении правой системы координат должно быть $z \rightarrow z$, но при этом $B_z \rightarrow -B_z$, что и доказывает антисимметричность тензора. Поэтому можно записать:

$$\tilde{\epsilon} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & 0 \\ \epsilon_{yx} & \epsilon_{yy} & 0 \\ 0 & 0 & \epsilon_{zz} \end{pmatrix}; \quad \begin{aligned} \epsilon_{xy} &= -i \frac{4\pi}{\omega} \sigma_{xy} \\ \epsilon_{yx} &= -\epsilon_{xy}. \end{aligned} \quad (83.26)$$

Вводя скалярные показатели преломления n и поглощения κ соотношением

$$n^2 (1 - i\kappa)^2 E = \epsilon E, \quad (83.27)$$

приравнявая коэффициенты при E_x и E_y слева и справа, получим:

$$\begin{aligned} n^2 (1 - i\kappa)^2 &= \epsilon_{xx} + \epsilon_{yx}, \\ n^2 (1 - i\kappa)^2 &= \epsilon_{yy} + \epsilon_{xy} = \epsilon_{yy} - \epsilon_{yx}. \end{aligned} \quad (83.28)$$

Из двух последних соотношений следует, что существуют два решения уравнений

$$n^2(1 - ik)^2 = \epsilon_{xx} \mp \epsilon_{xy}, \quad (83.29)$$

поскольку для циркулярно поляризованной волны можем записать

$$E_x = \mp iE_y \quad (83.30)$$

для лево-(+) и право(-) поляризованных волн (в соответствии с традиционными в оптике обозначениями поляризации; в квантовой механике и ядерной физике понятия правой и левой поляризации противоположны понятию в оптике). Два различных решения уравнения (83.29) соответствуют двум различным поляризациям:

$$n^{\pm 2} = \epsilon_{xx} \pm \epsilon_{xy}, \quad (83.31)$$

поэтому

$$n^{+2} - n^{-2} = i2\epsilon_{xy} \quad (83.32)$$

и

$$\varphi = i \frac{\pi l}{2\pi l} 2\epsilon_{xy} \quad (83.33)$$

или

$$\varphi = \frac{2\pi l}{\pi c} \sigma_{xy}. \quad (83.34)$$

Для коэффициента Верде можем записать

$$\theta = \frac{2\pi}{\pi c} \frac{\sigma_{xy}}{B_z}, \quad (83.35)$$

или

$$\theta = - \frac{2\pi}{\pi c} \frac{e^3}{m^*} K'_{12}. \quad (83.36)$$

Но согласно (74.32) имеем

$$\alpha = \frac{4\pi}{\pi c} \sigma_{xx} = \frac{4\pi}{\pi c} e^2 K'_{11}, \quad (83.37)$$

поэтому

$$\frac{\theta}{\alpha} = - \frac{2\pi e^3 K'_{12}}{\pi c m^* \frac{4\pi}{\pi c} e^2 K'_{11}} = - \frac{1}{2} \frac{e \langle \tau^2 \rangle}{m^* \langle \tau \rangle} = - \frac{1}{2} \mu^H, \quad (83.38)$$

т. е.

$$\theta = - \frac{\mu^H}{2} \alpha. \quad (83.39)$$

Для электронов $\mu^H < 0$ и коэффициент Верде положителен, для дырок $\theta < 0$.

Подвижность Холла μ^H зависит от механизма рассеяния, от него же зависит и α , поэтому та же зависимость и у θ . Действительно, непосредственно из выражения для θ через K_{12} следует

$$\theta = - \frac{2\pi}{\pi c} \frac{e^3}{m^{*2}} \rho < \tau^2 >. \quad (83.40)$$

Наиболее важной особенностью этого соотношения является линейная зависимость удельного угла поворота от концентрации носителей заряда p и от обратного квадрата эффективной массы m^{*2} .

Последнее соотношение получено без учета дисперсии. Основанное на феноменологических уравнениях Максвелла, оно справедливо в низкочастотной области спектра при $\omega^{-2} \gg \tau^2$. Чтобы получить выражение, применимое в высокочастотной области спектра, необходимо учесть множитель $(1 + \omega^2\tau^2)^{-1}$, входящий в поляризуемость (75.14) и диэлектрическую проницаемость (75.15). Добавляя его в проводимость Холла, в слабом поле можем записать

$$\theta = -\frac{2\pi}{nc} \frac{e^3 p}{m^{*2}} \left\langle \frac{\tau^2}{1 + \omega^2\tau^2} \right\rangle. \quad (83.41)$$

В высокочастотной области, где $\omega^{-2} \gg \tau^2$, имеем $\omega^2\tau^2 \gg 1$, поэтому

$$\theta = -\frac{2\pi e^3 p \omega^{-2}}{ncm^{*2}} = -\frac{e^3 p \lambda^2}{2\pi c^3 n m^{*2}}. \quad (83.42)$$

Это уравнение широко используется для определения эффективной массы свободных носителей заряда в образцах с примесной проводимостью. Условие его применимости проверяется экспериментально: график $\theta(\lambda^2)$ или $\varphi(\lambda^2)$ должен иметь вид прямой линии, в этом случае вращение плоскости поляризации от механизма рассеяния не зависит. Отступление от линейности в зависимости θ , или φ , от λ^2 может быть связано либо с невыполнением условия $\omega^2\tau^2 \gg 1$, либо с наложением других механизмов вращения, например, межзонного вращения. Рассмотрим вновь отношение θ/α , учитывая частотную зависимость проводимости. Согласно (75.12) можем записать с учетом усреднения по энергии:

$$\sigma = \frac{e^2 p}{m^*} \left\langle \frac{\tau}{1 + \omega^2\tau^2} \right\rangle, \quad (83.43)$$

$$\alpha = \frac{4\pi e^2 p}{ncm^*} \left\langle \frac{\tau}{1 + \omega^2\tau^2} \right\rangle \quad (83.44)$$

и

$$\frac{\theta}{\alpha} = -\frac{e}{2m^*} \frac{\left\langle \frac{\tau^2}{1 + \omega^2\tau^2} \right\rangle}{\left\langle \frac{\tau}{1 + \omega^2\tau^2} \right\rangle}. \quad (83.44)$$

При $\omega^2\tau^2 \ll 1$ имеем уже известное нам выражение

$$\frac{\theta}{\alpha} = -\frac{\mu^H}{2}. \quad (83.45)$$

При $\omega^2\tau^2 \gg 1$ получим

$$\frac{\theta}{\alpha} = -\frac{\mu^H}{2} \langle \tau \rangle^{-1} \langle \tau^{-1} \rangle^{-1}. \quad (83.46)$$

В более общем случае α может зависеть не от λ^2 , а пропорционально λ^γ при $\gamma \neq 2$, и отношение θ/α может зависеть от λ .

На рис. 140 приведен пример вращения на свободных электронах в теллуриде кадмия. Из угла наклона графика получено $m^*/m_0 = 0,09$ при 100 К.

Теория межзонного вращения также основывается на методе тензора проводимости. При вычислении тензора проводимости учитываются переходы электронов из валентной зоны в зону проводимости. Возмущение записывается в обычном виде:

$$W = \sum_i e (v_i A'), \quad (83.47)$$

где v_i — скорость электрона в некотором i -состоянии, в простейшем случае она равна отношению импульса или квазиимпульса к массе электрона, A' — вектор-потенциал световой волны. Используя волновые функции электрона в исходном k и конечном k' состояниях в зоне Бриллюэна, вычисляются матричные элементы оператора импульса, для которых предполагаются определенные соотношения, например, можно предполагать:

$$\frac{|P_{kk'}^+|}{\omega_{kk'}^+} = \frac{|P_{kk'}^-|}{\omega_{kk'}^-} = \frac{|P_{kk'}|}{\omega_{kk'}}. \quad (83.48)$$

Теоретические расчеты в соответствии с экспериментом приводят к выводу, что межзонное вращение изменяется как ω^2 при $\omega \rightarrow 0$. Однако вблизи края собственного поглощения различные теоретические модели приводят к различным выражениям для зависимости угла поворота от энергии фотонов. Согласно расчетам Л. Рот $\varphi \sim F_1(x)$, где

$$F_1(x) = \frac{1}{x} \left(\frac{1}{(1-x)^{1/2}} - \frac{1}{(1+x)^{1/2}} \right) - \frac{4}{x^2} [2 - (1-x)^{1/2} - (1+x)^{1/2}]; \quad (83.49)$$

$$x = \frac{\omega}{\omega_g} = \frac{\hbar\omega}{\Delta E_0}.$$

При $x \rightarrow 1$ функция $F_1(x)$ имеет асимптотический вид

$$F_1(x) \approx \frac{1}{(1-x)^{1/2}}. \quad (83.50)$$

У края поглощения $\omega = \omega_g$, или $x = 1$, функция $F_1(x)$ имеет сингулярность, от которой можно избавиться введением времени релаксации. Согласно другой теории (Босварва, Говард, Лидиард) вращение определяется функцией $F_2(x)$:

$$F_2(x) = \frac{1}{x} \left[\frac{1}{(1-x)^{1/2}} - \frac{1}{(1+x)^{1/2}} \right] - 1 \quad (83.51)$$

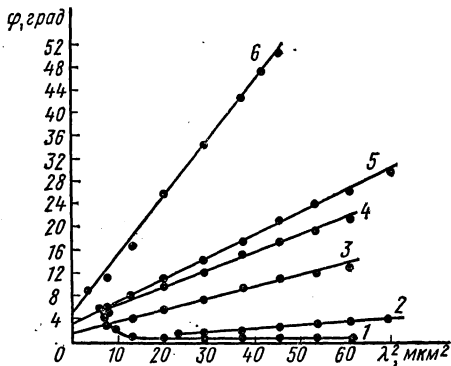


Рис. 140. Эффект Фарадея на свободных электронах в теллуриде кадмия концентрации электронов, см⁻³: 1 — $< 10^{15}$; 2 — $2,6 \cdot 10^{16}$; 3 — $1,2 \cdot 10^{17}$; 4 — $1,5 \cdot 10^{17}$; 5 — $2,0 \cdot 10^{17}$; 6 — $3,5 \cdot 10^{17}$

с тем же асимптотическим представлением. При $x \rightarrow 0$ обе функции ведут себя одинаковым образом, при этом $F_2(x) = 2F_1(x)$. Обе функ-

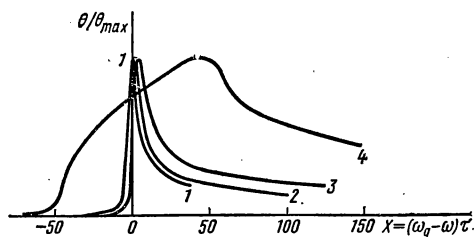


Рис. 141. Дисперсионная кривая при прямых межзонных переходах, Y : 1—0,1; 2—1; 3—5; 4—50

ции широко используются для анализа дисперсионных зависимостей, получаемых экспериментально, обычно в области $x \sim 0,6 \div 0,9$. Эксперимент показывает, что при $x \gtrsim 1$ вращение конечно. Этот результат находит объяснение в теории Б. Лэкса, в которой показано, что при вращении, обусловленном дисперсией вследствие перехода между парой уровней Ландау, φ определяется для прямых переходов выражением:

$$\begin{aligned} \varphi_{\text{Ландау}} = A \left\{ \frac{1}{[(X-Y)^2 + 1]} \left\{ [(X-Y)^2 + 1]^{1/2} - (X+Y) \right\}^{1/2} - \right. \\ \left. - \frac{1}{(X+Y)^2 + 1} \left\{ [(X+Y)^2 + 1]^{1/2} + (X+Y) \right\}^{1/2} \right\}, \quad (83.52) \end{aligned}$$

где A — величина, независящая от частоты, $X = (\omega_n - \omega) \tau$; $Y = \gamma B \tau$; γ — эффективный магнитный момент, выраженный через гиромагнитные отношения g_c и g_v и магнетон Бора μ_B ; $\gamma = \frac{g_c + g_v}{2\hbar} \mu_B$; $\omega_n = \omega_g + (n + 1/2) \omega_c$ — частота перехода между n -ми уровнями Ландау, ω_c — циклотронная частота, $\omega_c = 2\omega_L$.

Как видно из этого уравнения, знак угла поворота зависит от знака члена $X - Y$: при $X - Y < 0$ вращение положительно, при $X - Y > 0$ — отрицательно.

Таким образом, поворот, вызываемый дисперсией при переходах между парой уровней Ландау, испытывает инверсию знака. Величина вращения не симметрична по частоте относительно ω_n . Чтобы учесть вращение, создаваемое всей совокупностью уровней Ландау, необходимо провести суммирование по n , что приводит к выражению вида

$$\begin{aligned} \varphi = \frac{2A}{\omega_c \tau} \left\{ \left\{ [(X+Y)^2 + 1]^{1/2} + (X+Y) \right\}^{1/2} - \right. \\ \left. - \left\{ [(X-Y)^2 + 1]^{1/2} + (X-Y) \right\}^{1/2} \right\}, \quad (83.53) \end{aligned}$$

где $X = (\omega_g - \omega) \tau$, а остальные обозначения те же. Мы видим, что теперь межзонное вращение положительно. Вид дисперсионных кривых межзонного вращении плоскости поляризации при некоторых значениях параметра Y приведен на рис. 141.

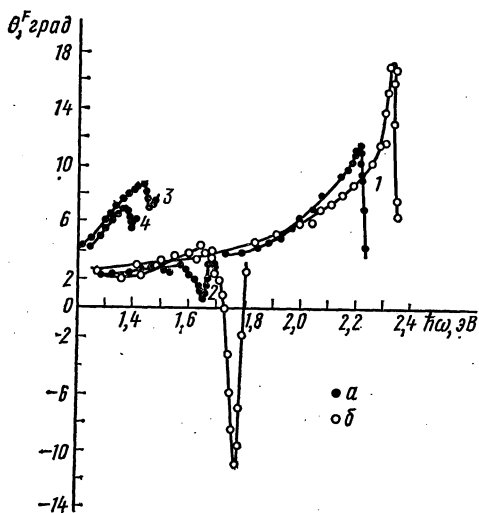


Рис. 142. Межзонное и экситонное вращение плоскости поляризации: a —300 °К; b —77 °К; 1—ZnTe; 2—CdSe; 3—CdTe; 4—CdTe_{0,85}Se_{0,05}

При экспериментальном исследовании межзонного вращении наблюдается инверсия знака вблизи края поглощения, а в ряде случаев инверсия наблюдается многократно на экситонных полосах поглощения (рис. 142).

§ 84. СПИН-ОРБИТАЛЬНОЕ РАСЩЕПЛЕНИЕ ЭНЕРГЕТИЧЕСКИХ ЗОН

Спин-орбитальное расщепление энергетических зон непосредственно следует из его наличия для уровней энергии свободных атомов, образующих энер-

гетические зоны при взаимодействии атомов друг с другом.

Спин-орбитальное расщепление уровней энергии атомов в атомной физике и спектроскопии обычно называют тонкой структурой уровней и спектральных линий.

Тонкая структура уровней энергии свободных атомов и ионов хорошо изучена. Теоретически существование тонкой структуры следует из релятивистского квантовомеханического уравнения Дирака, поэтому можно сказать, что тонкая структура, или спин-орбитальное расщепление, уровней является квантово-релятивистским эффектом. Наиболее удобным способом введения в рассмотрение спин-орбитального взаимодействия является переход от уравнения Дирака к уравнению Паули, т. е. переход от четырехкомпонентной волновой функции Дирака к двухкомпонентной функции Паули. Исключение «малых» компонентов из уравнения Дирака приводит к уравнению вида:

$$\left(\frac{\hat{p}^2}{2m_0} + U(r) - E \right) \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \left\{ \frac{p^4}{8m_0^3c^2} + \frac{e\hbar}{2m_0c} (\sigma B) + \frac{e\hbar}{4m_0^3c^2} \left(\sigma [\hat{E}\hat{p}] - \frac{\hbar^2}{8m_0c^2} \nabla U \right) \right\} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \quad (84.1)$$

Правая часть уравнения (84.1) представляет собой кванто-релятивистскую поправку к нерелятивистскому уравнению, стоящему

в левой части. Первый член $-p^4/8m_0^2c^2$ — учитывает релятивистскую зависимость массы от скорости. Его необходимость можно показать, исходя из релятивистской связи между энергией и импульсом:

$$E = \sqrt{m_0^2c^4 + p^2c^2}. \quad (84.2)$$

Разложим $E(p)$ в ряд по p^2 :

$$E = m_0c^2 + \frac{p^2}{2m_0} - \frac{p^4}{8m_0^3c^2} + \dots \quad (84.3)$$

Второй член дает энергию электрона во внешнем магнитном поле B , с которым электрон взаимодействует вследствие существования у него собственного магнитного момента

$$\mu_e = \frac{e\hbar}{2m_0c} \sigma = \frac{e}{m_0c} \mathbf{S}, \quad (84.4)$$

где $\mathbf{S} = \frac{\hbar}{2} \sigma$ — собственный механический момент, или спин; σ — двухрядные матрицы Паули.

Третий член представляет собой энергию спин-орбитального взаимодействия. Четвертый член носит название контактного взаимодействия, или поправки Дарвина. Обладая преимуществом строгого вывода, такой способ введения спин-орбитального взаимодействия мало нагляден, поэтому рассмотрим полуклассический способ его описания.

С релятивистской точки зрения описание электрического и магнитного полей как существующих независимо лишено смысла. В действительности существует единое электромагнитное поле, описываемое четырехмерным антисимметричным тензором поля $F_{\mu\nu}$:

$$\{F_{\mu\nu}\} = \begin{pmatrix} 0 & B_z & -B_y & -iE_x \\ -B_z & 0 & B_x & -iE_y \\ B_y & -B_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix}. \quad (84.5)$$

Компоненты тензора поля зависят от состояния движения физической системы, или, выражаясь более формально, от системы координат. Если в некоторой системе координат существуют и электрическое E , и магнитное B поля, то можно найти системы координат, в которых существует либо только электрическое, либо только магнитное поле.

Предположим, что тензор $F_{\mu\nu}$ известен нам в некоторой системе координат, которую примем в качестве «неподвижной» системы (x, y, z, t) или (x, y, z, ict) . Систему (x', y', z', t') или (x', y', z', ict') будем считать «движущейся» со скоростью $v = \beta c$ вдоль направления оси $x - x'$. Переход от неподвижной системы координат к «дви-

жущейся» системе координат определяется преобразованиями Лоренца:

$$\begin{pmatrix} x' \\ y' \\ z' \\ ict' \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix} = \frac{1}{\sqrt{1-\beta^2}} \begin{pmatrix} 1 & 0 & 0 & i\beta \\ 0 & \sqrt{1-\beta^2} & 0 & 0 \\ 0 & 0 & \sqrt{1-\beta^2} & 0 \\ -i\beta & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix} \quad (84.6)$$

Тензор поля $F'_{\mu\nu}$ в «движущейся» системе координат находится по обычным правилам преобразования компонентов тензора

$$F'_{\mu\nu} = \sum_{\alpha, \gamma=1}^4 a_{\mu\alpha} a_{\nu\gamma} F_{\alpha\gamma}. \quad (84.7)$$

Предположим, что в «неподвижной» системе координат электрическое поле отсутствует ($\mathbf{E} = 0$), а магнитное поле отлично от нуля ($\mathbf{B} \neq 0$). На покоящуюся частицу магнитное поле не оказывает воздействия. Предположим теперь, что заряженная частица движется со скоростью v . Введем систему координат, в которой частица покоится, и найдем тензор поля в «движущейся» системе координат. Проводя несложные преобразования (84.7) с матрицей (84.6), найдем для пространственно-подобных компонентов ($\mu, \nu \neq 4$):

$$\begin{aligned} F'_{12} &= \frac{F_{12}}{(1-\beta^2)^{1/2}}, & \text{или} & \quad B'_z = \frac{B_z}{(1-\beta^2)^{1/2}}; \\ F'_{13} &= \frac{F_{13}}{(1-\beta^2)^{1/2}}, & \text{или} & \quad B'_y = \frac{B_y}{(1-\beta^2)^{1/2}}; \\ F'_{23} &= \frac{F_{23}}{(1-\beta^2)^{1/2}}, & \text{или} & \quad B'_x = \frac{B_x}{(1-\beta^2)^{1/2}}. \end{aligned} \quad (84.8)$$

В трехмерных обозначениях соотношения (84.8) можно записать в виде:

$$\bar{\mathbf{B}}' = \frac{\mathbf{B}}{(1-\beta^2)^{1/2}}. \quad (84.9)$$

Для временно-подобных компонентов аналогично найдем:

$$\begin{aligned} F'_{41} &= 0, & \text{или} & \quad E'_x = 0, \\ F'_{42} &= \frac{-i\beta}{(1-\beta^2)^{1/2}} F_{12}, & \text{или} & \quad E'_y = \frac{-\beta}{(1-\beta^2)^{1/2}} B_z, \\ F'_{43} &= \frac{-i\beta}{(1-\beta^2)^{1/2}} F_{13}, & \text{или} & \quad E'_z = \frac{\beta}{(1-\beta^2)^{1/2}} B_y. \end{aligned} \quad (84.10)$$

В трехмерных обозначениях (84.10) можно записать

$$\mathbf{E}' = \frac{1}{(1-\beta^2)^{1/2}} \frac{1}{c} [\mathbf{vB}]. \quad (84.11)$$

Следовательно, на покоящуюся в «движущейся» системе координат частицу действует сила $e\mathbf{E}'$:

$$e\mathbf{E}' = \frac{e}{(1-\beta^2)^{1/2}} \frac{1}{c} [\mathbf{vB}]. \quad (84.12)$$

При малых скоростях $\beta^2 \ll 1$ имеем

$$e\mathbf{E}' = \frac{e}{c} [\mathbf{v}\mathbf{B}], \quad (84.13)$$

т. е. неоднократно использовавшееся нами выражение для силы Лоренца.

Теперь рассмотрим случай, когда в «неподвижной» системе координат существует только электрическое поле: $\mathbf{E} \neq 0$, $\mathbf{B} = 0$.

Проводя вычисления, получим:

$$\begin{aligned} F'_{41} &= F_{41}, \text{ или } E'_x = E_x, \\ F'_{42} &= \frac{F_{42}}{(1-\beta^2)^{1/2}}, \text{ или } E'_y = \frac{E_y}{(1-\beta^2)^{1/2}}, \\ F'_{43} &= \frac{F_{43}}{(1-\beta^2)^{1/2}}, \text{ или } E'_z = \frac{E_z}{(1-\beta^2)^{1/2}}. \end{aligned} \quad (84.14)$$

Для пространственно-подобных компонентов найдем:

$$\begin{aligned} F'_{12} &= -\frac{i\beta}{(1-\beta^2)^{1/2}} F_{42}, \text{ или } B'_z = \frac{\beta}{(1-\beta^2)^{1/2}} E_y, \\ F'_{13} &= \frac{i\beta}{(1-\beta^2)^{1/2}} F_{43}, \text{ или } B'_y = -\frac{\beta}{(1-\beta^2)^{1/2}} E_z, \\ E'_{23} &= 0, \text{ или } B'_x = 0. \end{aligned} \quad (84.15)$$

В трехмерных обозначениях можем записать

$$\mathbf{B}' = \frac{1}{(1-\beta^2)^{1/2}} \frac{1}{c} [\mathbf{E}\mathbf{v}]. \quad (84.16)$$

Таким образом, в «движущейся» системе координат существует магнитное поле, хотя в «неподвижной» системе координат оно отсутствует.

Если частица имеет магнитный момент μ , то при движении она будет взаимодействовать с электрическим полем, которое будет выступать как взаимодействие с магнитным полем \mathbf{B} . Энергия взаимодействия

$$W = -(\mu\mathbf{B}') = -\frac{1}{(1-\beta^2)^{1/2}} \frac{1}{c} (\mu [\mathbf{E}\mathbf{v}]). \quad (84.17)$$

Выразим магнитный момент μ частицы через механический момент или спин:

$$\mu = g \frac{e}{2mc} \mathbf{S} = g \frac{e\hbar}{4mc} \boldsymbol{\sigma}, \quad (84.18)$$

где g — гиромагнитное отношение, представляющее фактор Ланде в общем случае. Выражая скорость через импульс, запишем:

$$W = -\frac{ge\hbar}{(1-\beta^2)4m^2c^2} (\boldsymbol{\sigma} [\mathbf{E}\mathbf{p}]) = -\frac{ge\hbar}{4m^2c^2} (\boldsymbol{\sigma} [\mathbf{E}\mathbf{p}]). \quad (84.19)$$

При последнем преобразовании мы учли, что $m^2(1-\beta^2) = m_0^2$.

Для гиромагнитного отношения, связывающего собственный магнитный и механический моменты, необходимо взять значение, равное 2. Однако, положив $g=2$, придем к выражению для W , кото-

рое в два раза превосходит третий член в (84.1). Чтобы исключить различие с точным выражением, следует считать $g=1$ (поправка Томаса — Френкеля). Поправка Томаса — Френкеля свидетельствует о том, что по отношению к трансляционному движению собственный магнитный момент подобен орбитальному, для которого $g=1$, в то время как для собственных моментов должно быть $g=2$. Учитывая поправку Томаса — Френкеля, запишем энергию взаимодействия спинового магнитного момента с электрическим полем E в виде:

$$W = -\frac{e\hbar}{4m_0^2c^2} (\sigma [E\rho]) = \frac{\hbar^2}{4m_0^2c^2} (\sigma [\nabla U\rho]). \quad (84.20)$$

Применим выражение (84.20) к движению электронов в атомах. Напряженность электрического поля ядра можно выразить в виде

$$E = \frac{(Z - \sigma_n) |e|}{r^3} \mathbf{r}, \quad (84.21)$$

где σ_n — константа экранирования; $(Z - \sigma_n) |e|$ — эффективный заряд ядра, экранируемого внутренними электронными оболочками.

Подставляя (84.21) в (84.20), получим

$$\begin{aligned} W &= -\frac{e\hbar}{4m_0^2c^2} \left(\sigma \left[\frac{(Z - \sigma_n) |e|}{r^3} \mathbf{r}, \mathbf{p} \right] \right) = \\ &= \frac{(Z - \sigma_n) e^2 \hbar}{4m_0^2c^2 r^3} (\sigma [\mathbf{r}\mathbf{p}]) = \frac{(Z - \sigma_n) e^2 \hbar^2}{4m_0^2c^2 r^3} (\sigma \mathbf{L}), \end{aligned} \quad (84.22)$$

где \mathbf{L} — орбитальный механический момент, выраженный в \hbar , т. е.

$$[\mathbf{r}\mathbf{p}] = \mathbf{M} = \hbar \mathbf{L}. \quad (84.23)$$

Переходя к квантовомеханическому описанию, следует заменить W на оператор \hat{W} по обычным правилам и найти среднее значение в состоянии с волновой функцией ψ .

Если использовать водородоподобные волновые функции

$$\psi = \psi_{nlm}(\mathbf{r}) = R_{nl}(\mathbf{r}) Y_{lm}(\theta, \varphi), \quad (84.24)$$

то можно получить

$$W = W_{nlj} = \frac{\alpha^2 Ry (Z - \sigma_n)^4}{n^3} \frac{j(j+1) - l(l+1) - s(s+1)}{2l(l+1/2)(l+1)}, \quad (84.25)$$

где $\alpha = \frac{e^2}{\hbar c} = 7,29717 \cdot 10^{-3} \approx 1/137$ — постоянная тонкой структуры; $Ry = 13,6$ эВ — ридберг, равный энергии ионизации атома водорода; n — главное, l — орбитальное, j — полное или внутреннее, квантовые числа.

В электронвольтах W_{nlj} запишется в виде:

$$W_{nlj}, \text{ эВ} = 0,362 \cdot 10^{-3} \frac{(Z - \sigma_n)^4}{n^3} \frac{j(j+1) - l(l+1) - s(s+1)}{l(l+1/2)(l+1)}. \quad (84.26)$$

Спин-орбитальное расщепление должно наблюдаться для всех состояний, кроме s , поскольку в s -состоянии орбитальный момент равен нулю и согласно (84.27) $W = 0$.

Для водородоподобных или, в общем случае, для одноэлектронных атомов $j = l \pm 1/2$, поэтому все уровни, кроме s , становятся

дублетными. Для более сложных атомов следует заменить одноэлектронные квантовые числа j, l, s на квантовые числа всей электронной оболочки J, L, S .

Невозбужденное состояние электронной оболочки атомов IV и VI групп подгруппы В имеет $S=1$ и $L=1$, поэтому полное квантовое число может принимать три различных значения: $J=0, 1, 2$ в соответствии с тремя различными ориентациями моментов L, S . Ориентационный множитель имеет три различных значения: $-4/3; -2/3$ и $2/3$ соответственно для $J=0, 1$ и 2 . Расстояния между уровнями тонкой структуры относятся как $1:2$. На рис. 143 показана схема спин-орбитального расщепления исходного уровня на три подуровня при нормальном (а) и обращенном (б) расположении их. Ориентационный множитель определяет число компонентов тонкой структуры, а величина расщепления зависит от номера оболочки и константы экранирования.

В пределах одного ряда периодической таблицы Менделеева n и σ_n постоянны, а Z возрастает, поэтому величина спин-орбитального расщепления растет. Например, расщепление уровней энергии пар атомов $C-O, Si-S, Ge-Se, Sn-Te$ близко, но у элементов

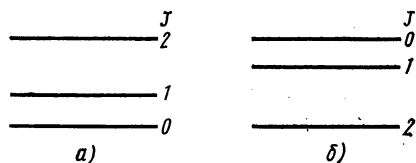


Рис. 143. Схема расположения уровней тонкой структуры в атомах

VI В группы оно больше, чем у элементов IV группы. Для элементов одной группы n растет с ростом номера периода (для валентных электронов), одновременно с этим растут Z и σ_n . Однако σ_n растет медленнее, чем Z , поэтому в пределах одной группы эффективный заряд растет с ростом Z быстрее, чем n . Это приводит к тому, что для атомов одной группы величина спин-орбитального расщепления возрастает с ростом порядкового номера элемента. В табл. 26 приведены примеры величин спин-орбитального расщепления основного состояния атомов некоторых элементов. У большинства атомов из табл. 26 основным является состояние 3P_0 , однако у селена имеет место инверсное расположение подуровней энергии, а у теллура основным является уровень 3P_2 , а верхние уровни имеют «нормальное» расположение. Правило интервалов выполняется приближенно для германия и кремния, а для теллура имеет место практически «дублетное» расщепление, поскольку расстояние $^3P_2 - ^3P_1$ составляет $0,586$ эВ, а $^3P_1 - ^3P_0$ — всего $0,006$ эВ.

Как видно из таблицы, расщепление уровней атомов элементов одной и той же группы действительно возрастает с ростом Z , и у таких сравнительно тяжелых элементах, как теллур, оно может составлять $0,6$ эВ.

При образовании кристаллов спин-орбитальное расщепление зон энергии обуславливается взаимодействием магнитного момента электрона с электрическим полем решетки $E = -\frac{1}{e} \nabla U$. Величина расщепления зоны энергии определяется в основном спин-орбитальным

расщеплением тех уровней энергии, из которых возникает зона. Поскольку к полю ядра данного атома добавляется поле взаимодействия ядер и электронов, то величина расщепления зон отличается от величины расщепления уровней энергии атомов. Опыт и расчет показывают, что, как правило, величина спин-орбитального расщепления зон энергии несколько больше расщепления соответствующих уровней энергии. Однако основные закономерности в изменении величины расщепления зон энергии с ростом заряда ядер атомов кристалла сохраняются. В настоящее время известны величины спин-орбитального расщепления зон энергии многих полупроводников. Величина расщепления зон в точке Γ , т. е. при $k=0$, обычно обозначается через Δ_0 , в точке L , т. е. в точке, лежащей на границе зоны Бриллюэна в направлении типа [111], через Δ_1 , при этом во многих случаях $\Delta_1 = \frac{2}{3} \Delta_0$.

Спин-орбитальное расщепление зон энергии, образованных из p -уровней, может быть только дублетным, «чистые» s -зоны не имеют спин-орбитального расщепления. В полупроводниках IV группы и в соединениях элементов типа $A^{III}B^V$ и $A^{II}B^{VI}$ спин-орбитальное расщепление наблюдается для валентной зоны. При фиксированном элементе B величина расщепления остается практически постоянной. Это находится в хорошем соответствии со сказанным в § 26 о характере образования зон энергии соединений $A^{II}B^{IV}$.

Таблица 26

Спин-орбитальное расщепление основного состояния атомов элементов четвертой и шестой групп. За начало отсчета принята энергия наиболее глубокого уровня

Элементы	Энергия уровня						$^3P_2 - ^3P_1$	
	см ⁻¹ , эВ		см ⁻¹ , эВ		см ⁻¹ , эВ		см ⁻¹	эВ
	P_0		P_1		P_2			
Углерод, С . . .	0,00	0,00	16,4	0,0020	43,5	0,0054	27,1	0,0033
Кремний, Si . .	0,00	0,00	77,15	0,009	223,2	0,0276	146,2	0,018
Германий, Ge .	0,00	0,00	557,1	0,069	1409,9	0,174	852,8	0,101
Кислород, О . .	0,00	0,00	68,9	0,0084	157,5	0,019	88,6	0,011
Сера, S	0,00	0,00	176,8	0,0216	396,8	0,049	220,0	0,027
Селен, Se	2534	0,317	1990	0,244	0,00	0,00	-545	-0,073
Теллур, Te . .	4707	0,580	4751	0,586	0,00	0,00	-44	-0,006

Таблица 27

Величина спин-орбитального расщепления валентной зоны в точке Γ (Γ_{15}) некоторых полупроводников, эВ

Полупроводник	Si	Ge	ZnS	ZnSe	ZnTe	CdTe	InSb
Эксперимент . . .	0,03	0,09	0,07	0,43	0,93	0,92	0,82
Расчет	0,03	0,29	0,08	0,42	0,93	0,90	0,82

Основные условные обозначения

$x, y, z, \mathbf{r}, \hat{\mathbf{r}}$	— координаты, радиус-вектор и оператор координаты.
$p_x, p_y, p_z, \mathbf{p}, \hat{\mathbf{p}}$	— компоненты импульса, импульс и его оператор.
ρ	— концентрация дырок, показатель степени зависимости времени релаксации от энергии.
$P, \hat{P}, P_x, P_y, P_z$	— квазиимпульс, его оператор и компоненты.
\hat{P}	— обобщенный импульс и квазиимпульс.
$P_{\text{фон}}$	— квазиимпульс фононов.
$v, \hat{v}, v_x, v_y, v_z$	— скорость, оператор и компоненты скорости.
E	— полная энергия.
E, E_x, E_y, E_z	— напряженность электрического поля.
H	— напряженность магнитного поля.
\hat{H}	— функция Гамильтона, коэффициент и магнетосопротивления.
\hat{H}	— оператор Гамильтона.
v_d	— дрейфовая скорость.
v_D	— диффузионная скорость.
τ	— время релаксации, время свободного движения, коэффициент Томсона.
τ_f, τ	— время жизни.
l	— длина свободного пробега.
μ_d, μ_n, μ_p	— подвижность дрейфовая, электронов и дырок.
μ^H	— подвижность Холла.
μ	— магнитная проницаемость.
n	— концентрация электронов, концентрация носителей заряда.
η	— коэффициент преломления.
\mathbf{n}	— вектор трансляции.
$\mathbf{j}, j_x, j_y, j_z$	— плотность тока и ее компоненты.
$\sigma, \sigma_n, \sigma_p$	— удельная электрическая проводимость, электронная и дырочная проводимости.
σ^*, σ	— эффективное сечение рассеяния и поглощения.
m	— масса свободного электрона.
m^{*-1}, m^*	— тензор обратной эффективной массы, тензор эффективной массы.
m^*	— скалярная эффективная масса.
m_l, m_t	— продольная и поперечная эффективные массы.
m_d^*	— эффективная масса плотности состояний.
F, F^*	— уровень (энергия) и квазиуровень Ферми.
F	— сила.
T	— температура абсолютная.
T, \hat{T}	— кинетическая энергия, оператор кинетической энергии.
$\hat{T}(\mathbf{n}), T(\mathbf{n})$	— оператор трансляции и его собственные значения.
U, V, \hat{U}, \hat{V}	— потенциальная энергия и оператор потенциальной энергии
u, u_{ij}	— тензор деформации и его элемент.
u	— вектор смещения.
$\mathcal{W}, \hat{\mathcal{W}}$	— возмущение, оператор возмущения.
\mathcal{W}	— истинная работа выхода.
\hat{w}	— приведенные смещения.
\mathcal{W}	— плотность потока энергии.
Φ	— термодинамическая работа выхода.
$\Phi, (\xi)$	— интеграл Ферми порядка j .
ϕ^k, ϕ^l	— контактная разность потенциалов, внешняя и внутренняя.
ϕ^H	— угол Холла.

k	— константа Больцмана.
k	— волновой вектор электрона.
$k_{\text{фон}}$	— волновой вектор фонона.
K'_{rs}	— кинетический коэффициент.
L	— число Лоренца, диффузионная длина.
α	— абсолютная дифференциальная термо-э. д. с., коэффициент поглощения.
R	— коэффициент Холла, скорость рекомбинации.
G, G	— скорость генерации, объем основной области.
LD	— дебаева длина экранирования.
g	— волновой вектор фотона.
D	— индукция электрического поля.
D_{nj}, m_i	— элементы динамической матрицы.
s	— скорость поверхностной рекомбинации.
S^*, S	— интегральное эффективное сечение.
W_{ij}	— матричный элемент оператора возмущения.
w	— вероятность перехода.
e	— заряд частицы.
e_n, e^-	— заряд электрона.
e_p, e^+	— заряд дырки.
b	— отношение модулей подвижностей электрона и дырки.
b	— вектор обратной решетки.
$A(r)$	— вектор-потенциал.
c	— скорость звука, скорость света.
θ	— температура Дебая.
Z	— статистическая сумма, заряд ядра, заряд иона/в $ e $
$h, \hbar = \frac{h}{2\pi}$	— постоянная Планка.
\hat{V}	— оператор Гамильтона (оператор набла).

Предметный указатель

- Аккумуляция 446, 476, 477
Акцентор 35, 36, 121, 122, 124
Алмаза решетка 31
Антисимметричный единичный тензор третьего ранга 320
- Батареи солнечные 552
Беккереля соотношение 562
Блоха волна 52, 53, 354
Бозе—Эйнштейна статистика 388
Больцмана кинетическое уравнение 218, 220, 227
— функция распределения 20, 21, 165, 175, 177, 234
Бора магнетон 210
Бриллюэна зоны 72, 74, 86, 87, 95, 105, 133, 139, 153, 154, 164, 179, 180, 214, 391, 431, 509
Бугера — Ламберта закон 487, 488, 546
- Ванье функции 116
Вариационный принцип 146
Варизонные структуры 552, 553
Вектор волновой 51, 52, 53, 57, 73, 80, 92, 101, 125, 133, 379, 509
Вектор обратной решетки 73, 80, 87, 91, 94
Вектор-потенциал 127, 128, 507, 523
Верде коэффициент 561
Вероятность захвата 434
— перехода 217, 354, 356, 403, 509, 525, 530
— рассеяния 337, 338
Видемана-Франца закон 286
Возмущение 81, 94, 113, 123, 351, 507, 508, 513, 523
Возмущения оператор — см. возмущения
Время жизни 424, 448
— мгновенное 452, 456, 426, 430
— релаксации 10, 11, 221, 226, 346
— релаксационное 427, 430
— свободного движения 9, 10, 11, 13, 15, 139
— стационарное 427, 430
— эффективное 453
Всестороннего сжатия тензор 323
Вульфа-брегга уравнение 87, 94
- Вымораживание 205
Выврождения фактор 106, 169, 182
Выход квантовый 539
- Гамильтона функция 110, 127, 373
Гамильтониан 37, 43, 48, 67, 81, 107, 113, 128, 350
Гамма-функция 175
Ганна эффект 419
Гаусса эффект 252
Гейзенберга соотношения 164
Генерация 32, 213, 290, 422, 423, 537, 546
Германий 118, 120, 135, 153, 185, 334, 440, 518
Гиббса потенциал 18
Главные координаты 372
Граничные условия циклические 78
Гаука закон 324
- Дебаева длина экранирования 454, 465
Дебая температура 389, 392, 395, 396
Де-Бройля волна 82, 87, 128, 354
Дельта-функция 19, 77, 177, 212
Дембера эффект 547, 554
Деформации тензор 322, 403
Диагональная матрица 372
Диаманитный резонанс 506
Диэлектрик 26, 28, 134, 138
Дипольный момент 365
Диффузии коэффициент 303, 443, 448
Диффузия биполярная 549, 553
Длина диффузионная 445, 449
— дрейфа 476, 446, 449
— затягивания 476, 445, 449
— свободного пробега 9, 10, 13, 249, 253, 488
Донор 34, 36, 121, 124
Доплера эффект 398
Дырка 33, 35, 138, 143
Дырка квантовомеханическая 69
Дюлонга-Пти закон 22, 393
- Запирающий слой 475
Зинера эффект 112, 417
Значения собственные 37, 50, 52, 54, 63, 77, 114
Зона валентная 160, 134, 138, 151

Зона запрещенная 98, 100, 107, 110, 114, 137, 138, 156
— примесная 118, 124, 206
— проводимости 157, 134, 138, 150
— свободная 134, 138
— экситонная 162
— энергетическая 90, 93, 94, 101, 106
Зоны Бриллюэна — см. Бриллюэна зоны
Зоны вырожденные 151, 186
Зон энергии искривление 111, 112, 454

Инверсия населенности 431
Инъекции уровень 439, 441
Инъекция 446, 477
Импульс обобщенный 127, 132
Интеграл перекрытия 97
Интеграл соударений 217
Ионизация ударная 417, 542
— электростатическая 418
Истощение 191

Квазимпульс 53, 54, 55, 57
Квазимпульса оператор 54, 55, 57
Квазисвободного электрона теория 81
Квазисвязанного электрона теория 81, 95
Квазичастицы 33, 158, 162
Квантовые системы — см. системы вырожденные
Кейна метод 147
— параметр 149
Келдыша — Франца эффект 536
Кикоина — Носкова эффект 554
Колесания акустические 383, 386
— оптические 405, 383, 385, 387
— поперечные 380
— продольные 380
Концентрация избыточная 422, 423, 430
— критическая 203, 205
— неравновесная 422, 430
— носителей заряда 172
Контакт антизапирающий 475
— выпрямляющий 475
— запирающий 475
— омический 475
Координаты оператор 63, 67
Коэффициент биполярной диффузии 581, 549, 553
— кинетический 232
Конуэлл—Вайскопфа формула 362
Кристалл-фотоэффект — см. Дембера эффект

Лармора частота 560
Лиувилля теорема 215
Ловушки захвата 433, 441

Ловушки рекомбинационные 433, 441
Лоренца сила 222, 247
— условие 507
— число 286

Магнетопоглощение 537
Максвелла время релаксации 447
— уравнения 490
Масса приведенная 160
— эффективная 58
Матрица диагональная 372, 375
— динамическая 369, 375
— унитарная 371, 375
Металл 26, 27, 133, 138
Метод последовательных приближений 47
Методы расчета энергетической структуры 143
Модуль упругости 324

Набор динамических величин полный 132
Напряжения тензор 324
Непрерывности уравнение 422, 430
Нернста—Эттингсгаузена эффект 247, 254, 255, 305, 307
Нернста эффект 247, 254, 255, 305
Нормальные координаты 370, 372, 373, 375 -
Нормировки условие 13, 16, 76, 96, 344, 387
Носители заряда 7, 30, 33
— — неосновные 34, 35, 191
— — неравновесные 430
— — основные 34, 35, 191
— — темновые 423

Область истощения примеси 191
Обменная энергия 100, 107
Обменный эффект 117, 124, 159
Обратной решетки вектор 74
Объемные поля 313
Оже процессы 432
Ома закон 10, 239, 414
Оператор Гамильтона — см. Гамильтона оператор
— Лапласа 37
— квазимпульса — см. квазимпульса оператор
— кинетической энергии 37
— координаты — см. координаты оператора
— потенциальной энергии 37
— скорости 62
— ускорения 66
Ортогонализированных плоских волн метод 144

Паули принцип 132
Пельтье коэффициент 246, 292
— поток 246, 292
Переходы вертикальные 509, 522
— запрещенные 515
— квантовые 350
— не прямые 548, 516, 522
— прямые 509, 519, 522
— разрешенные 514
Плоских волн метод 144
Плотность потока энергии 228, 284
— состояний 164, 206
— электрического тока 10, 11, 17, 228, 239
Поверхность изоэнергетическая 59, 71, 92, 95
Поглощение внутризонное 489
— плазменное 489
— примесное 489, 528
— решеточное 549, 454, 489, 523
— свободными носителями заряда 489, 490
— собственное 489, 507
— фундаментальное — см. поглощение собственное
— экситонное — 489, 520
Показатель поглощения 492
— преломления 492
Подвижность 9, 11, 14, 139, 142, 239, 245, 404, 416, 443
— биполярная диффузионная 549
— — дрейфовая 549
Поле самосогласованное 44, 46, 48
Полупроводники 26, 27, 134, 138
— атомарные 28
— вырожденные 25, 175, 179, 202, 236
— дырочные 33
— ионные 28
— невырожденные 25, 175, 177, 179, 234
— скомпенсированные 35, 197
— собственные 33, 183
— электронные 28
Порядок ближний 158, 163
— дальний 158, 163
Потенциал деформации 330, 396, 403
— ионизации 136, 160
— самосогласованный 44
— химический 18
— экранирующий 123, 357
Потока вероятности плотность 355
Предел квантовый 211
Приближение адиабатическое 39
— Борна—Оппенгеймера 39
— квазисвободного электрона 42, 81
— квазисвязанного электрона 42, 95
— нулевое 81, 94, 95, 98
— одноэлектронное 44
— первое 82, 94, 98

Присоединенных плоских волн метод 145
Проводимость примесная 34
— световая 545
— собственная 33
— темновая 545
— удельная электрическая 10, 11, 239, 503
Псевдопотенциала метод 145
Пуассона интеграл 21, 175, 467
Пуля эффект 416

Работа выхода истинная 83, 466, 469
— термодинамическая 467, 469
Расстояние прищельное 358
Резерфорда формула 358
Резонанс диамагнитный 132
— циклотронный 132, 499
Рекомбинация 32, 290, 425
— безызлучательная 432
— излучательная 432
— квадратичная 426, 431
— линейная 424, 427, 431
— плазменная 432
— поверхностная 449
— прямая 433
— тепловая 423
— фотонная 432

Решетка обратная 72
Риги—Ледюка—Маджи эффект 305
Риги—Ледюка эффект 305
Ричардсона формула 467

Связь вакантная 30, 32
Сдвига тензор 324
Сечение эффективное дифференциальное 344, 349, 357, 363
— — интегральное 344, 349
— — поглощения 488
— — транспортное 426, 343, 349
Сжатие всестороннее 329
— объемное 329
— одностороннее 331
Системы вырожденные 20
— невырожденные 20
Скобки Пуассона квантовые 62
Скорость 66
— биполярной генерации 550
— генерации 422
— групповая 64, 379, 381, 385
— диффузионная 445
— дрейфа 9, 416
— звука 381, 385
— рекомбинации 422
— фазовая 379, 381, 385
Слой антизапирающий 475

Слой запирающий 475
— инверсный 465
Слэтера определитель 47
Смещения приведенные 369
— вектор 322, 368
Соединения полупроводниковые 28,
29
Соударение 9, 10, 216, 220
Сохранения законы 53
Спектр отражения 486
— поглощения 486
Спин-орбитальное расщепление 148,
152, 153, 568
Столкновение — см. соударение
Сумма статистическая 416, 387

Температура вырождения 23, 25,
205
— истощения 193
— отрицательная 431
— характеристическая 395
Тензочувствительность 328
Теплопроводность 284
— фононная 285
— электронная 284
Термозонд 297
Термомагнитные явления 305
Термоэлектрический эффект 291
Термо-э. д. с. абсолютная 295
— относительная 296
Термоэлектронной эмиссии плотность
тока 467
Ток диффузионный 443, 448
— дрейфовый 443, 448
— поляризации 500
Томсона соотношения 301, 304, 305
— эффект 293

Удельное сопротивление продольное
327
Упругости коэффициент 369
Уравнение вековое 85, 113
— движение квантовое 109
— секулярное — см. уравнение веко-
вое
Уровни акцепторные 124
— глубокие 122
— дискретные 124
— донорные 121
— Ландау 132
— мелкие 122
— поверхностные 127
— примесные 124
— Тамма 127
Усиления коэффициент 541
Условие трансляционное 51

Фарадея эффект 559
Ферми — Дирака функция 18, 25, 165,
172, 173, 222
Ферми интеграл 173, 174
— квазиуровень 431, 432, 441
— поверхность 180, 240
— уровень 18, 22, 179, 183, 184, 189,
190, 192, 195, 196, 198, 204, 211,
293, 437, 468, 470, 473, 479
— энергия — см. Ферми уровень
Фононы 374, 375, 395, 396
Фотогальванический эффект 551, 554
Фотомагнитноэлектрический эффект
554
Фотоответ 545
Фотоотклик 545
Фоторезистивный эффект 537, 545
Фоточувствительность 545
Функция распределения 12, 16, 17, 25,
195
Функция собственная 37, 44, 49
Фурье интеграл 491
— компонент 491
— ряд 82

Характеристическое соотношение пер-
вое 539, 546
— — второе 541, 546
Хартри уравнение 47
Хартри — Фока уравнение 47
Холла коэффициент 251, 255, 258, 265,
269, 276
— подвижность 257, 258, 260, 263,
275
— проводимость 280, 563
— угол 250, 257, 258, 280
— фактор 260, 263
— эффект 249, 255, 256, 263
Холла — Шокли — Рида теория 434

Центры окраски 534
Циклоида 248

Частота осциллятора 129
— собственная 398, 129
— циклотронная 129, 132, 206, 503
Число состояний эффективное 173, 174,
179

Шредингера уравнение 37, 44
Штарк-эффект 418

Экситон 160
Эксклюзия 446, 477
Экстракция 446, 477

Энергия активации 26, 184, 198, 201

— ионизации примеси 119, 121

Энергии квантование 127, 131, 132

Эффективной массы метод 107, 112

Эффективности параметр 305

Эффекты адиабатические 254

— изотермические 254

— нечетные 255

— поперечные 224

— продольные 224

Юнга модуль 328

Явления кинетические 214, 219

Явления переноса — см. явления кинетические

Ячейка фазовая 164

Оглавление

	Стр.
Предисловие ко второму изданию	3
Предисловие к первому изданию	4
Глава I. Электронная теория проводимости	7
§ 1. Электронная теория проводимости. Закон Ома	7
§ 2. Функция распределения времен и длин свободного пробега	12
§ 3. Функция распределения электронов по состояниям. Средние значения физических величин	16
§ 4. Полупроводники. Классификация веществ по величине проводимости	25
§ 5. Модельные представления о проводимости полупроводников. Понятие дырки	30
§ 6. Собственная и примесная проводимость	33
Глава II. Основы зонной теории полупроводников	37
§ 7.- Уравнение Шредингера для кристалла	37
§ 8. Адиабатическое приближение	39
§ 9. Одноэлектронное приближение	44
§ 10. Периодическое поле решетки кристалла. Оператор трансляции	48
§ 11. Квазимпульс	53
§ 12. Эффективная масса электрона	58
§ 13. Связь скорости с квазимпульсом	62
§ 14. Оператор ускорения	66
§ 15. Зоны Бриллюэна	72
§ 16. Нормировка в ящике и дискретность квазимпульса	76
§ 17. Теория квазисвободного электрона	81
§ 18. Теория квазисвязанного электрона	95
§ 19. Метод эффективной массы. Влияние внешних полей на спектр энергии кристалла	107
§ 20. Локализованные состояния	113
§ 21. Элементарная теория примесных состояний	118
§ 22. Поверхностные состояния	124
§ 23. Квантование энергии электрона в магнитном поле. Уровни Ландау	127
§ 24. Принцип Паули. Понятие металла, полупроводника и диэлектрика	132
§ 25. Основные характеристики дырки	138
§ 26. Зонная структура некоторых полупроводников	143
§ 27. Понятие о квазичастицах	158
Глава III. Статистика электронов и дырок в полупроводниках	164
§ 28. Плотность состояний	164
§ 29. Концентрация электронов и дырок	172
§ 30. Уравнение электронейтральности	180
§ 31. Собственный полупроводник	183

§ 32. Примесный полупроводник. Примесь одного вида	187
§ 33. Полупроводник, содержащий акцепторную и донорную примесь	197
§ 34. Вырожденный полупроводник	202
§ 35. Плотность состояний в магнитном поле	206
Глава IV. Кинетические явления в полупроводниках	214
§ 36. Кинетическое уравнение Больцмана	214
§ 37. Время релаксации	220
§ 38. Плотность электрического тока и плотность потока энергии	228
§ 39. Кинетические коэффициенты	232
§ 40. Электропроводность полупроводников	239
§ 41. Гальваномагнитные эффекты	247
§ 42. Эффект Холла в области примесной проводимости	256
§ 43. Эффект Холла в веществе с носителями заряда нескольких типов	263
§ 44. Зависимость коэффициента Холла от магнитного поля	269
§ 45. Магниторезистивный эффект	276
§ 46. Теплопроводность полупроводников	284
§ 47. Термоэлектрические явления	291
§ 48. Термомагнитные явления	305
§ 49. Общий анализ кинетических явлений	309
§ 50. О кинетических явлениях в полупроводниках с тензорной эффективной массой	319
§ 51. Тензорезистивный эффект. Тензочувствительность	322
§ 52. Тензорезистивный эффект. Коэффициенты пьезосопротивления	329
Глава V. Теория рассеяния носителей заряда	337
§ 53. Эффективное сечение рассеяния	337
§ 54. Связь времени релаксации с эффективным сечением	346
§ 55. Понятие о теории квантовых переходов	350
§ 56. Рассеяние на ионах примеси	357
§ 57. Рассеяние на нейтральных атомах примеси	364
§ 58. Колебания решетки. Нормальные координаты. Фононы	368
§ 59. Акустические и оптические колебания решетки	376
§ 60. Теплоемкость решетки. Статистика фононов	387
§ 61. Рассеяние на тепловых колебаниях решетки. Метод потенциала деформации	396
§ 62. Зависимость подвижности носителей заряда от температуры	404
§ 63. Зависимость времени релаксации от внешних полей. Нарушения закона Ома	414
Глава VI. Рекомбинация носителей заряда	421
§ 64. Уравнение непрерывности. Время жизни	421
§ 65. Механизмы рекомбинации. Линейная рекомбинация	431
§ 66. Диффузия и дрейф неравновесных носителей заряда	442
§ 67. Поверхностная рекомбинация	449
Глава VII. Контактные явления в полупроводниках	454
§ 68. Дебаева длина экранирования	454
§ 69. Работа выхода	465
§ 70. Контактная разность потенциалов. Контакт металл — металл	469
§ 71. Контакт металл — полупроводник	474
§ 72. Неоднородный полупроводник. p - n -переход	479
	583

Глава VIII. Оптические и фотоэлектрические явления в полупроводниках 486

	§ 73. Спектр поглощения света	486
	§ 74. Поглощение света свободными носителями заряда	490
	§ 75. Циклотронный резонанс	499
	§ 76. Собственное поглощение света	507
	§ 77. Поглощение света решеткой	523
	§ 78. Поглощение света электронами в локализованных состояниях	528
	§ 79. Влияние внешних условий на спектр поглощения	534
	§ 80. Фоторезистивный эффект	537
Пр	§ 81. Эффект Дембера. Фотогальванический эффект	546
	§ 82. Фотомагнитноэлектрический эффект	554
Пр	§ 83. Эффект Фарадея	559
	§ 84. Спин-орбитальное расщепление энергетических зон	568
	Основные условные обозначения	575
Гл	Предметный указатель	577

Г

Петр Семенович Киреев

ФИЗИКА ПОЛУПРОВОДНИКОВ

Редактор *Г. А. Сорокина*, Художественный редактор *Т. М. Скворцова*. Переплет художника *Ю. Д. Федичкина*. Технический редактор *Л. А. Муравьева*.
Корректор *М. М. Малиновская*

Т-03108. Сдано в набор 10/IX 1974 г. Подп. к печати 9/I 1975 г. Формат 60×90¹/₁₆.
Бум. тип. № 3. Объем 36,5 печ. л. Усл. п. л. 36,5. 32,55 уч.-изд. л. Изд. № ЭР-173.
Тираж 30 000 экз. Цена 1 р. 37 к. Зак. 1592. План выпуска литературы издательства «Высшая школа» (вузы и техникумы) на 1975 г. Позиция № 172.

Москва, К-51, Неглинная ул., д. 29/14, Издательство «Высшая школа»

Ордена Трудового Красного Знамени Ленинградское производственно-техническое объединение «Печатный Двор» имени А. М. Горького Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательства, полиграфии и книжной торговли, 197136, Ленинград, П-136, Гатчинская ул., 26,

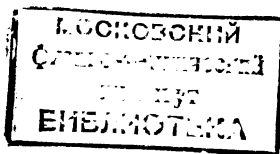
537.8
К 45

П. С. Киреев

ФИЗИКА ПОЛУПРОВОДНИКОВ

ИЗДАНИЕ ВТОРОЕ,
ДОПОЛНЕННОЕ

ДОПУЩЕНО МИНИСТЕРСТВОМ
ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО
ОБРАЗОВАНИЯ СССР
В КАЧЕСТВЕ УЧЕБНОГО ПОСОБИЯ
ДЛЯ СТУДЕНТОВ ВТУЗОВ



379402



МОСКВА «ВЫСШАЯ ШКОЛА» 1975

1р.37к.

ИЗДАТЕЛЬСТВО МОСКВА
1975 ГОД
ВЫСШАЯ ШКОЛА

