


В. ШОКЛИ

ТЕОРИЯ
ЭЛЕКТРОННЫХ
ПОЛУПРОВОДНИКОВ



В. ШОКЛИ

ТЕОРИЯ ЭЛЕКТРОННЫХ ПОЛУПРОВОДНИКОВ

ПРИЛОЖЕНИЯ К ТЕОРИИ ТРАНЗИСТОРОВ

*Перевод с английского
под редакцией и с предисловием*

В. П. ЖУЗЕ

И * Л

ИЗДАТЕЛЬСТВО
ИНОСТРАННОЙ ЛИТЕРАТУРЫ

Москва, 1953

ELECTRONS AND HOLES
IN SEMICONDUCTORS

WITH APPLICATIONS TO TRANSISTOR ELECTRONICS

By
WILLIAM SHOCKLEY

TORONTO—NEW YORK—LONDON

1950

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Предлагаемая вниманию советского читателя монография В. Шокли охватывает ряд важнейших вопросов современной теории электронных полупроводников. Весьма значительная ее часть посвящена описанию кристаллических усилителей — транзисторов — и теории их действия. Как указывает в своем предисловии автор, в основу книги положены лекции, прочитанные им для сотрудников руководимой им научно-исследовательской лаборатории в период выполнения программы работ по кристаллическим детекторам и усилителям.

Книга Шокли состоит из трех частей различной степени сложности. Первые две части рассчитаны на инженеров и физиков, не имеющих специальной подготовки в области теоретической физики, а третья часть предназначена для лиц, обладающих серьезной теоретической подготовкой.

В первой части книги даются общие сведения о гомеоплярных полупроводниках, структуре их энергетического спектра, их электрических свойствах и подробно излагаются основные сведения о кристаллических усилителях и явлениях, лежащих в основе их действия. Значительный интерес представляет гл. 3, посвященная рассмотрению нового явления — введения („инъекции“) носителей тока в полупроводник — и описанию методики определения подвижности по измерению скорости дрейфа таких носителей в электрическом поле. Гл. 4 целиком отведена теории транзистора. Материал излагается в значительной степени качественно, без громоздкого математического аппарата.

Вторая часть книги, наибольшая по объему, написана на гораздо более высоком теоретическом уровне и посвящена описательной теории полупроводников на основе зонной теории твердого тела и квантовой статистики. Подробно рассмотрена теория электропроводности и эффекта Холла и проанализирован вопрос о подвижности носителей тока в кристаллической решетке. Достаточно подробно разбираются также различные механизмы рассеяния электронов при их движении через кристалл. Наглядные графики во многом облегчают понимание материала. В гл. 12 дается теория $p-n$ -переходов и подробно рассматривается важный вопрос о времени жизни носителей тока в полупроводниках.

Как в первой, так и во второй части книги автор почти совершенно не касается технических применений кристаллических усилителей.

В третьей части книги рассматриваются соотношения квантовой механики и квантовой статистики, лежащие в основе современной теории полупроводников. Особый интерес представляют § 6—8 гл. 15, в которых автор анализирует законы движения электронов и дырок в электрических и магнитных полях, основываясь на представлении о волновых пакетах, а также те параграфы гл. 17, в которых речь идет о разработанном автором книги методе потенциалов деформации, применяемом для вычисления матричного элемента вероятности перехода.

Серьезным недостатком книги является замалчивание работ советских ученых по физике полупроводников. В книге приведены всего две ссылки на работы советских физиков, несмотря на то, что весьма многие вопросы и даже целые разделы теории полупроводников полностью разработаны и созданы в СССР.

При редактировании перевода мы стремились по возможности восполнить этот пробел, дав в примечаниях ссылки (конечно, далеко не исчерпывающие) на соответствующие исследования и монографии советских авторов¹⁾.

При переводе книги мы сочли необходимым совсем опустить § 1 гл. 14 „О характере прикладной математической теории“, не имеющий непосредственного отношения к вопросам, составляющим содержание книги, а также ссылки на этот параграф в других частях книги. В этом параграфе автор излагает идеалистические взгляды американского физика Бриджмена, с которыми он полностью солидарен. Мы прибегали также к частичным сокращениям текста (когда это можно было сделать, не затрагивая основного физического содержания книги), или давали соответствующие примечания, которые должны помочь советскому читателю разобраться в некоторых ошибочных высказываниях автора и понять значение советской науки в развитии ряда вопросов, затронутых в книге.

При переводе значительные трудности вызвали допущенные автором стилистические небрежности. В ряде случаев материал изложен недостаточно последовательно. Например, гл. 16 носит название „Статистическая теория в приложении к полупроводникам“. Однако большая часть главы посвящена формулировке некоторых общих положений статистической физики, выводу формулы Планка и закона распределения Ферми — Дирака. Все эти вопросы достаточно подробно разобраны в учебниках по статистической физике и здесь могли бы быть опущены. В остальной же части главы рассмотрены численные (правда, полезные) примеры графического решения всего лишь одной задачи об определении положения уровня Ферми в германии при разных температурах. Таким образом, читатель, надеявшийся найти

¹⁾ См. литературу в конце каждой главы, а также Библиографию, стр. 633.

в этой главе обзор весьма многочисленных приложений статистической физики к полупроводникам, будет обманут в своих ожиданиях.

Библиография, помещенная в конце книги, дополнена работами по кристаллическим усилителям, опубликованными уже после выхода монографии Шокли из печати. Работы советских и зарубежных исследователей, упоминающиеся в примечаниях редактора, внесены в списки литературы в конце каждой главы и отмечены звездочкой.

Монография Шокли вышла в свет в самом конце 1950 г., т. е. всего через три года после изобретения транзистора. В связи с этим целый ряд вопросов, разработанных уже после выхода американского издания книги, не нашел в ней отражения. Поэтому было сочтено целесообразным дополнить книгу переводом двух статей, посвященных $n-r-n$ -транзисторам. В первой статье (Шокли с сотрудниками) изложена теория вопроса, а вторая статья (Мортон), печатаемая в несколько сокращенном виде, дает представление о технических характеристиках современных (середина 1952 г.) транзисторов, устойчивости и воспроизводимости их параметров, а также их практических применениях в области средств связи, в счетных устройствах и т. п.

Несмотря на указанные выше недостатки, монография Шокли, безусловно, будет полезна для широкого круга советских читателей, работающих в области полупроводников и в особенности полупроводниковой электроники.

Сентябрь 1952 г.

В. Жузе.

ИЗ ПРЕДИСЛОВИЯ АВТОРА

В основу этой книги положен ряд лекций, прочитанных автором в связи с развитием программы работ по транзисторам. Содержание ее и структура в значительной степени определились этими лекциями. Ударение в изложении сделано, в соответствии с этим, на явления, наиболее существенные для электроники транзисторов.

Первая часть книги содержит лишь самое элементарное введение в теорию полупроводников и транзисторов, основное внимание уделено рассмотрению результатов экспериментальных исследований. Предполагается, что этот материал будет доступен инженерам-электрикам и студентам-физикам, не знающим квантовой механики. Эта часть должна служить основой для понимания работы транзисторов, а также элементарных расчетов.

Третья часть книги имеет своей целью показать, как основные положения квантовой теории приводят нас к представлениям о свойствах дырок и электронов. Чтобы сделать понятным хотя бы основной смысл, если не детали, этих математических исследований для читателя, не очень искушенного в теоретической физике, эта часть книги содержит некоторые вводные главы. Так, гл. 13 и 14 помогут ответить на вопросы, часто возникающие при рассмотрении волновых функций и при попытке их физической интерпретации. В гл. 14 содержится изложение значительной части основных квантовомеханических принципов, необходимых для последующего рассмотрения поведения дырок и электронов. Она включает также введение в статистическую механику и рассматривает некоторые вопросы, относящиеся к теории электронной проводимости в кристаллах.

Во второй части книги делается попытка перебросить мост между первой и третьей частями книги. Результаты и соображения, приведенные в третьей части книги, представлены здесь в более наглядной и описательной форме.

В заключительной главе этой части вновь рассматриваются некоторые теории в аналитическом виде, полезном при количественном исследовании явлений, протекающих в транзисторах. Эта глава содержит также некоторые новые данные по электронике транзисторов.

В конце многих глав помещены задачи. Некоторые из них представляют собой просто численные примеры, рассчитанные на то, чтобы дать читателю возможность оценить порядок величин, другие

помещены с целью дополнить или расширить математическую сторону изложения; наконец, последняя группа задач посвящена вопросам, которые вполне могли бы обсуждаться и в тексте, но вынесены в задачи для большей краткости изложения.

В основе исследовательской деятельности нашей лаборатории лежит стремление проникнуть в логические следствия основной теории, придать этим следствиям наглядную форму и найти из них экспериментальные выходы. Открытие транзистора последовало в результате выполнения исследовательской программы лаборатории, основанной на этих принципах. Структура и содержание этой книги в какой-то мере отражают этот принцип. Принцип выполнения исследовательской программы, связанной, например, с изобретением транзистора, заключается в необходимости добиваться основательного понимания сущности исследуемого явления и в то же время настойчиво искать возможности его технических применений.

Автор считает, что для глубокого понимания многих проблем физики твердого состояния часто бывает полезно ответить на следующие четыре вопроса:

1) Какие атомы принимают участие в явлении и как они расположены?

2) Как достигается такое расположение?

3) Как это расположение приводит к определенному механизму движения электронов и атомов?

4) Как этот механизм обуславливает наблюдаемые свойства?

Постановка этих вопросов помогла понять как явление электропроводности, так и процесс введения носителей заряда с p — n -переходов. Многие другие явления, рассматриваемые в этой книге, понятны, с точки зрения четвертого и отчасти третьего вопросов; следущей задачей является дальнейшее исследование некоторых из этих явлений.

ОБОЗНАЧЕНИЯ

Индексы

Следующие индексы встречаются часто и во многих случаях не повторяются в основном перечне обозначений:

- a, c, d, v, G, F — акцептор, зона проводимости, донор, зона валентных связей, ширина запрещенной зоны, уровень Ферми;
 - b, c, ϵ, J — основной электрод, коллектор, эмиттер и точка контакта (см. фиг. 37);
 - n, p — применяются для обозначения электронов (отрицательных) и дырок (положительных), а также для обозначения областей с проводимостью n - или p -типа;
 - s — квантовое состояние;
 - P — квази-импульс.
- $i, j, k, \alpha, \beta, \gamma, a, b, c$ — иногда стоят в качестве целых чисел, отличающих отдельные члены ряда сходных величин.

Основной перечень обозначений

- a, a_1, a_2, a_3 — постоянная решетки, единичные векторы или периоды других периодических структур;
- a — постоянная затухания;
- A, A_x, A_y, A_z, A_1 и т. д. — протяженность периодической структуры;
- A — антисимметричная волновая функция;
- A — вектор-потенциал;
- b — обозначение основного электрода транзистора; отношение подвижности электронов к подвижности дырок; постоянная разделения в нитевидном образце;
- B, B_M — плотность магнитного потока;
- $B = P/h$;
- B — полуширина нитевидного образца;
- c — скорость света; скорость звука; постоянная разделения в нитевидном образце;
- c_{ij} — упругие постоянные;
- c_{II} — среднее значение упругой постоянной в продольном направлении;
- C — емкость; полуширина нитевидного образца;
- D, D_p, D_n — коэффициенты диффузии;
- e — абсолютное значение заряда электрона ($e = q$ в гл. 12);
- E, E — напряженность электрического поля;
- $\mathcal{E}, \mathcal{E}(P), \mathcal{E}_a$ и т. д. — энергии квантовых состояний, разности энергий, термодинамическая энергия системы;

- $E_{\text{кин}}$ — кинетическая энергия в классической и квантовой механике;
 f — частота;
 f, f_p — функции распределения Ферми — Дирака для электронов и дырок;
 F — сила;
 \mathcal{E} — свободная энергия;
 g — число пар электрон — дырка, образуемых в единицу объема за единицу времени;
 g, G — обратные векторы;
 $g(x)$ — четная функция;
 G_{ka} — единичный вектор;
 h — постоянная Планка;
 \hbar — постоянная Планка, деленная на 2π ;
 H, H_x, H — магнитное поле;
 \mathcal{H} — оператор, или функция Гамильтона;
 i_b, i_c и т. д. — переменные составляющие тока;
 i_x, i_y, i_z — единичные векторы;
 I, I_p, I_b — плотности тока; составляющие тока; полная величина тока (в случае постоянного тока);
 k — постоянная Больцмана;
 $k = P/\hbar$ — волновой вектор;
 K, K' — постоянные величины;
 l_p, l_n — средняя длина свободного пробега;
 L — длина; индуктивность;
 \mathcal{L} — оператор смещения;
 m — масса электрона (всюду, кроме гл. 17);
 m_e — масса электрона (гл. 17);
 m_n, m_p — эффективные массы;
 n — концентрация электронов в зоне проводимости;
 N — число объектов или их концентрация;
 N_c, N_v — эффективные плотности состояний в зонах;
 N_a, N_d, N_I — концентрация акцепторов и доноров; концентрация ионов;
 N_s — число атомов в кристалле;
 $N_x, N_y, N_z, N_1, N_2, N_3$ — число элементарных ячеек вдоль ребра периодической структуры;
 p — концентрация дырок в зоне валентных связей;
 p, p — импульс электрона;
 P, P, P', P_x — квази-импульс;
 q, q_i — координаты;
 q — координаты и спин электрона;
 Q — число систем;
 r — сопротивления для малых сигналов переменного тока; число пар электрон — дырка, рекомбинирующих в единицу объема за единицу времени;
 r, r — радиус-вектор и радиус;

- R — сопротивление;
 $R_{\text{H}}, R_{\text{L}}, R_{\text{M}}$ — постоянная Холла;
 $R_{\text{p}}, \delta R, R_{\text{T}}$ — векторы, определяющие положения ядер, узлов решетки кристалла и т. д.;
 s — скорость рекомбинации на поверхности; спиновое квантовое число;
 s, S — число квантовых состояний;
 \mathcal{S} — энтропия;
 t — время;
 T — абсолютная температура;
 $u(x)$ — нечетная функция;
 $u_{\text{P}}(r)$ — периодический множитель функции Блоха;
 $U(r)$ — потенциальная энергия электронов;
 v, v — скорость;
 $v_{\text{g}}, v_{\text{p}}$ — групповая и фазовая скорости;
 $v_{\text{b}}, v_{\text{c}}, v_{\text{e}}$ — напряжения в случае переменного тока;
 $V_{\text{b}}, V_{\text{c}}, V_{\text{e}}$ — напряжения постоянного тока;
 V — объем кристалла;
 dV (с различными индексами) — элементы объема;
 V_{P} — объем зоны Бриллюэна;
 ψ — потенциальная энергия ядра;
 w — волновые пакеты;
 W — число способов, которыми может быть осуществлено данное распределение;
 W_{ij}, W_{ij} — вероятности перехода;
 x, X, y, Y, z, Z — координаты;
 Z — заряд ядра;
 α — коэффициент усиления по току транзистора;
 $\alpha_{\text{e}}, \alpha_{\text{f}}$ — коэффициенты усиления эквивалентной схемы;
 α — спиновая волновая функция;
 β — доля вошедшего тока, достигающая коллектора;
 β — спиновая волновая функция;
 γ — доля тока эмиттера, переносимая введенными носителями;
 δ_{rs} — дельта-функция, равная 1 при $r = s$ и равная 0 при $r \neq s$;
 Δ — расширение;
 ∇ — вектор-оператор;
 ε — обозначение эмиттера;
 ε_0 — диэлектрическая проницаемость в системе MKS;
 $\varepsilon(\Lambda)$ — знак перестановки;
 $\theta_{\text{m}}, \theta_{\text{p}}$ — углы Холла;
 $\kappa = \kappa_{\text{e}}$ — диэлектрическая постоянная;
 λ — длина волны;
 Λ — перестановочный оператор;
 μ_0 — магнитная проницаемость вакуума в системе MKS;
 $\mu, \mu_{\text{D}}, \mu_{\text{H}}$ — подвижность; дрейфовая подвижность; подвижность Холла;

- ν — частота; постоянная затухания;
 ρ — удельное сопротивление; плотность заряда;
 ρ_f — плотность квантовых состояний;
 σ — удельная проводимость;
 τ_p, τ_n, τ_f — время жизни электронов и дырок; время пролета;
 $\tau, \tau(v), \tau_e$ — среднее время жизни; время релаксации; среднее время между столкновениями;
 τ_H, τ_μ — среднее время жизни (из эффекта Холла); среднее время жизни (из измерений подвижности);
 $\varphi, \varphi_n, \varphi_p$ — уровни и квази-уровни Ферми;
 ϕ — скалярный потенциал; угол в цилиндрических и сферических координатах; вертикальное смещение;
 $\varphi, \varphi_f, \varphi(n)$ — волновые функции;
 Φ — волновая функция кристалла, содержащего один избыточный электрон;
 χ — волновая функция;
 ψ — электростатический потенциал;
 Ψ_P — функция Блоха;
 $\omega = 2\pi\nu = 2\pi f$ — круговая частота;
 Ω — поверхность постоянной энергии (изоэнергетическая поверхность);
 $\langle \rangle$ — среднее значение (например, $\langle x \rangle$ — среднее значение x);
 \bullet — \dot{E} , т. е. $\partial E/\partial t$;
 $*$ — \mathcal{U}_{ij}^* , т. е. комплексная сопряженная \mathcal{U}_{ij} .

Основные физические постоянные

	В системе CGS		В системе MKS	
	Значение	Логарифм	Значение	Логарифм
Заряд электрона e	$4,80 \cdot 10^{-10}$	— 9,319	$1,60 \cdot 10^{-19}$	—18,796
Масса электрона m	$9,11 \cdot 10^{-28}$	—27,041	$9,11 \cdot 10^{-31}$	—30,041
Масса протона	$1,67 \cdot 10^{-24}$	—23,777	$1,67 \cdot 10^{-27}$	—26,777
Постоянная Планка h	$6,62 \cdot 10^{-27}$	—26,179	$6,62 \cdot 10^{-34}$	—33,179
$\hbar = h/2\pi$	$1,054 \cdot 10^{-27}$	—26,977	$1,054 \cdot 10^{-34}$	—33,977
Радиус бортовой орбиты $a_0 = \hbar/m_e v$	$0,529 \cdot 10^{-8}$	— 8,277	$0,529 \cdot 10^{-10}$	—10,277
Атомная единица энергии me^4/\hbar^2	$4,36 \cdot 10^{-11}$	—10,360	$4,36 \cdot 10^{-18}$	—17,360
Постоянная Больцмана k	$1,38 \cdot 10^{-16}$	—15,860	$1,38 \cdot 10^{-23}$	—22,860
Число Авогадро	$6,02 \cdot 10^{23}$	23,780		
Скорость света c	$2,998 \cdot 10^{10}$	10,477	$2,998 \cdot 10^8$	8,477
Диэлектрическая проницаемость вакуума, ϵ_0 ($\text{ф} \cdot \text{м}^{-1}$)			$8,854 \cdot 10^{-12}$	—11,053
Магнитная проницаемость вакуума, μ_0 ($\text{гн} \cdot \text{м}^{-1}$)			$1,257 \cdot 10^{-6}$	— 5,901

ЧАСТЬ I

ВВЕДЕНИЕ В ЭЛЕКТРОНИКУ
ТРАНЗИСТОРОВ

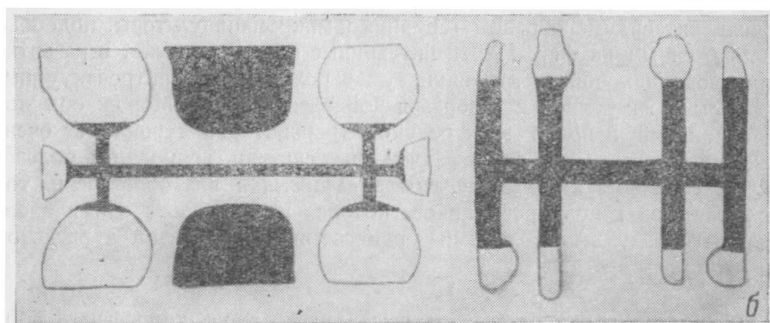
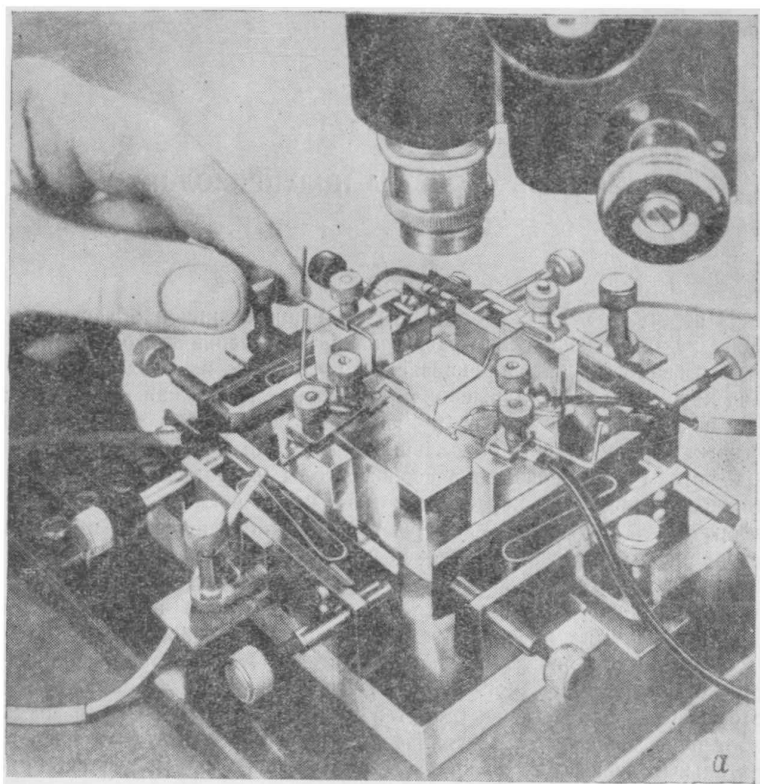
Глава 1

ОБЪЕМНЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ

В период между 1940 и 1950 гг. наши знания в области полупроводников¹⁾ значительно расширились. В течение этого десятилетия были развиты достаточно надежные представления о схемах энергетических уровней и движении электронов в кремнии и германии. Это в значительной степени явилось результатом исследований твердых выпрямителей, проводившихся многими лабораториями в связи с развитием радиолокации в военные годы и продолжающихся в настоящее время. Ряд исследований военного времени был проведен лабораторией излучения Массачузетского технологического института, находившейся в ведении национального комитета оборонных исследований. Работы этой лаборатории в области твердых выпрямителей описаны в книге Торрея и Уитмера [1], в которой имеются ссылки на работы других лабораторий, в частности английских.

С появлением кристаллических усилителей (транзисторов) роль полупроводников в технической электронике значительно возросла. Кроме того, специальная техника, развитая в связи с выполнением программы работ по транзисторам, позволила получить целый ряд новых экспериментальных данных, касающихся основных процессов электронной проводимости в полупроводниках. Центральное место в этих исследованиях заняли нитевидные образцы германия и микроманипуляторы, подобные изображенным на фиг. 1. Эти нитевидные образцы играют в развитии полупроводниковой электроники ту же роль, что и электровакуумные приборы в вакуумной и газоразрядной электронике. Между обычной электроникой и новой электроникой транзисторов существует очень близкая аналогия, в связи с чем мы считаем возможным пользоваться термином *транзисторная электроника* для обозначения той области физики полупроводников, которая имеет дело с этими вновь открытыми фундаментальными процессами, с анализом и расчетом

¹⁾ Как уже было отмечено в нашем предисловии к переводу, автор почти совершенно умалчивает о работах по полупроводникам, выполненных в СССР. Необходимо подчеркнуть, что советские исследователи внесли огромный вклад в развитие физики и техники полупроводников. Ряд основных теоретических и принципиальных экспериментальных работ, в частности, непосредственно касающихся и вопросов, затронутых в настоящей книге, был выполнен советскими учеными [21—32], см. также [33—36]. — *Прим. ред.*



Фиг. 1. *а* — Микроманипулятор, использованный при исследовании движения введенных в германий дырок и электронов. *б* — Нитевидные образцы германия сложной формы с контактами большой площади.

Длина левого образца 7,5 мм, правого — 3 мм.

транзисторов, а также с теорией схем, где применяются эти новые приборы. Все эти вопросы рассматриваются в последующих главах.

Полупроводники, представляющие интерес для транзисторной электроники, являются электронными, а не ионными проводниками. В ионных проводниках перенос значительной части электрического тока сопровождается перемещениями ионов. Поскольку физические свойства вещества и его химический состав определяются взаимным расположением ионов и атомов, ионная проводимость обычно приводит к радикальным изменениям в образце. Напротив, в электронных проводниках при прохождении электрического тока атомы остаются в своих фиксированных положениях. Поэтому в процессе проводимости число электронов в объеме полупроводника может уменьшаться или увеличиваться, но структура и химический состав его остаются при этом неизменными.

§ 1. ИЗБЫТОЧНЫЕ ЭЛЕКТРОНЫ И ДЫРКИ — НОСИТЕЛИ ЭЛЕКТРИЧЕСКОГО ТОКА

Основой теории полупроводников, как известно, является представление о том, что электроны могут переносить электрический ток двумя различными способами, один из которых называют *избыточной проводимостью*, *проводимостью избыточными электронами* или просто *электронной проводимостью*, другой же обычно называется *дефектной проводимостью*, или *дырочной проводимостью*. Возможность одновременного и раздельного существования этих двух процессов в полупроводниках и лежит в основе наших представлений о механизме действия транзисторов.

Квантовомеханическое объяснение этих процессов при помощи представлений об *энергетических зонах* и *зонах Бриллюэна* в настоящее время хорошо разработано и явится как раз основным содержанием последующих глав. В настоящей вводной главе дается лишь элементарный обзор поведения электронов в кремнии и германии¹⁾. Рассматриваемые здесь общие представления находят затем более строго теоретическое объяснение в последующих главах.

Кремний и германий являются так называемыми *ковалентными кристаллами*, атомы которых связаны между собой ковалентными связями (*парно-электронными* связями), образуемыми валентными электронами²⁾. Прежде чем рассматривать электронную структуру кристаллов, рассмотрим сначала ковалентную связь в молекуле водорода, являющаяся примером простейшей парно-электронной связи. На фиг. 2 изображены два атома и одна молекула водорода. Каждый атом

¹⁾ См. обзор [36], а также реферативные сборники [37, 38]. — *Прим. ред.*

²⁾ Подробнее см., например, [39] и [40]. — *Прим. ред.*

состоит из протона и одного электрона. Протон примерно в 2 000 раз тяжелее электрона и является относительно неподвижной частицей, вокруг которой по своей орбите вращается электрон. В квантовой механике движение электрона описывается волновой функцией. (Интерпретация волновых функций будет дана позднее, в гл. 5 и 14.) В изолированном атоме эта волновая функция обладает сферической симметрией и заряд электрона диффузно распределен по сфере с центром около протона. Если сблизить два таких атома, то между электронами возникает взаимодействие, вследствие чего вид электронного облака изменяется; это показано схематически на фиг. 2, б. В результате между двумя протонами возникает как бы добавочный отрицательный заряд, который и связывает их вместе. В соответствии с законами квантовой механики и принципом Паули, эта связь особенно устойчива в том случае, когда в ней участвуют точно два электрона. При удалении одного из электронов связь эта значительно ослабляется, а при прибавлении третьего электрона лишь немного усиливается. Эта особая устойчивость связи пары электронов или так называемой ковалентной связи в настоящее время вполне понятна на основе волновой механики [2]¹⁾.

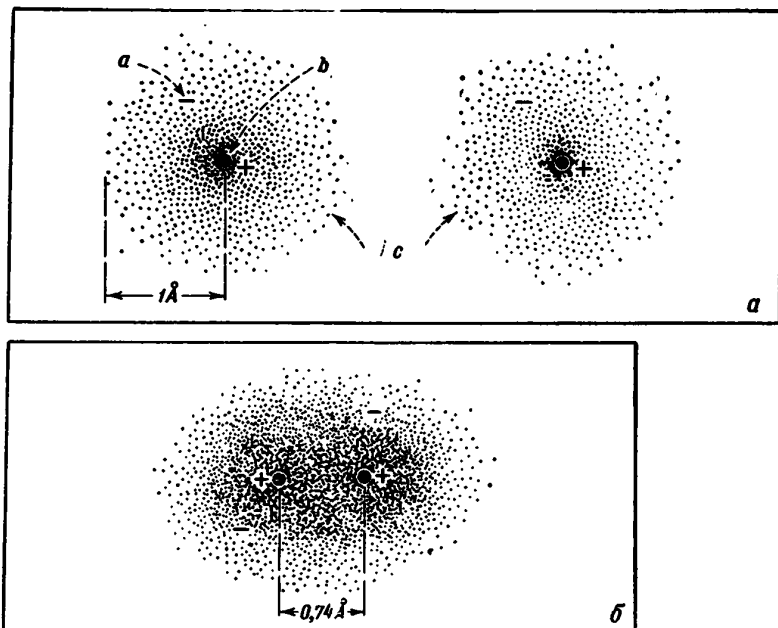
В периодической системе элементов (см. приложение, стр. 46) ковалентная связь играет особенно существенную роль у элементов класса III, которые, согласно классификации Юм-Розери [3], образуют группу так называемых *валентных кристаллов*. Атомные номера (Z) углерода, кремния и германия — элементов IV группы периодической системы — равны соответственно 6, 14 и 32. Эти числа равны величине заряда атомных ядер указанных элементов, измеренных в единицах заряда электрона. Таким образом, нейтральный атом состоит из ядра с зарядом, равным $+Ze$, и из Z окружающих ядро электронов, каждый из которых имеет заряд $-e$.

Валентность элементов углерода, кремния и германия одинакова и равна 4, как это изображено на фиг. 3. Хотя электронные оболочки этих элементов содержат соответственно 6, 14 и 32 электрона, лишь четыре из них участвуют в химических реакциях, остальные тесно связаны с ядрами, образуя стабильный *атомный остаток* с результирующим зарядом, равным $+4$. В электронных процессах, имеющих место при химических реакциях и в полупроводниках, этот атомный остаток совершенно не участвует.

Каждый из этих атомов стремится образовать ковалентные или парно-электронные связи с четырьмя другими атомами. Эта тенденция полностью осуществляется в структуре алмаза, характеризующей все три перечисленные элемента. Эта структура изображена на фиг. 4 и является одним из основных типов структур кубической системы;

¹⁾ Книга Паулинга, на которую ссылается автор, подверглась справедливой критике со стороны советской научной общественности. См., например, [41]. — *Прим. ред.*

она построена как бы из восьми вставленных одна в другую простых кубических решеток, подобных изображенному на рисунке большому кубу с восемью атомами в его вершинах. Теория проводимости в алмазной решетке тесно связана с деталями структуры кристалла однако в настоящей главе для нас представляет интерес лишь одна



Фиг. 2. Парно-электронная связь в молекуле водорода. а — два атома водорода; б — молекула водорода.

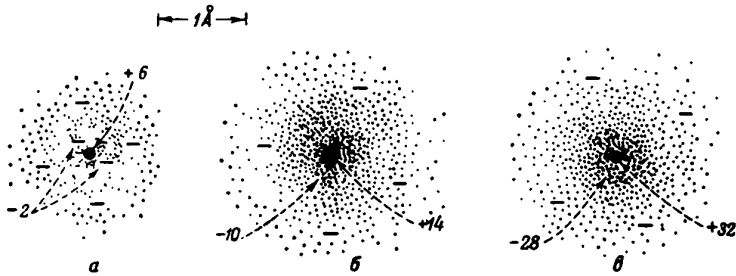
а — электрон; б — протон; точками обозначена плотность заряда электронов; $1 \text{ \AA} = 10^{-8} \text{ см}$.

особенность этой структуры (см. верхний левый угол фиг. 4). Как это видно из рисунка, здесь каждый атом окружен четырьмя соседними атомами, равномерно расположенными вокруг него и связанными с ним четырьмя ковалентными связями. Эти соседние атомы располагаются в вершинах правильного тетраэдра в соответствии с известными химическими свойствами тетраэдрического атома углерода¹⁾.

1) Задолго до того, как с помощью рентгеновских лучей было установлено пространственное расположение атомов в решетке алмаза, органическая химия установила, что углерод характеризуется тетраэдрическим расположением четырех связей в пространстве. Этот правильный вывод был сделан из наблюдений оптических свойств растворов органических соединений. — *Прим. авт.*

Длина ребра куба, изображенного на рисунке, является постоянной решетки и равна a .

Для алмаза, кремния, германия и серого олова постоянная решетки равна соответственно 3,56; 5,42; 5,62 и 6,46 Å, где $1\text{Å} = 10^{-8}\text{ см}$. Расстояние между ближайшими атомами равно $\sqrt{3}/16 a$. Из изображенных 18 атомов в действительности лишь 8 принадлежат элементарной ячейке объема a^3 . Из 8 атомов, находящихся в вершинах куба,



Фиг. 3. Электронная структура атомов углерода, кремния и германия.

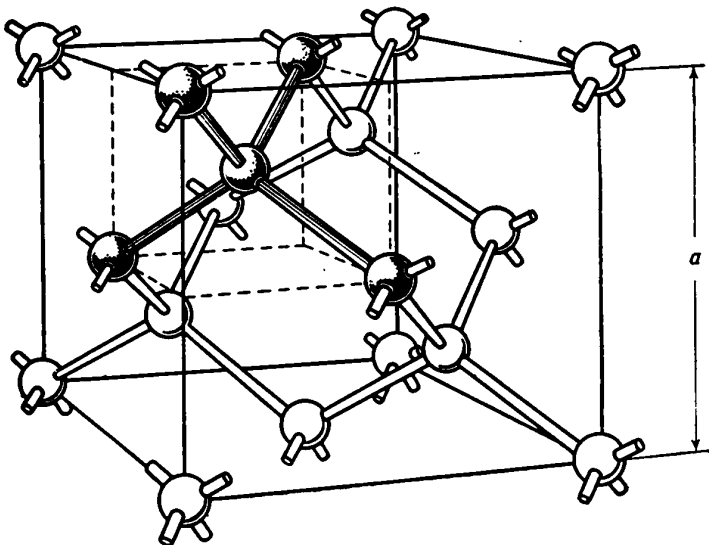
a —углерод ($Z=6$); $б$ —кремний ($Z=14$); $в$ —германий ($Z=32$).

на одну элементарную ячейку приходится всего 1 атом, поскольку каждая такая вершина является общей для 8 кубов; из 6 атомов, находящихся в гранях куба, на одну элементарную ячейку приходится 3 атома, поскольку каждая грань является общей для двух кубов. Кроме того, внутри куба имеется 4 атома, что и дает в результате 8 атомов на элементарную ячейку. Таким образом, число атомов в 1 см^3 , равно $(8/a^3)$, у алмаза равно $17,7 \cdot 10^{22}$, у кремния $5,00 \cdot 10^{22}$, у германия $4,52 \cdot 10^{22}$ и у серого олова $2,82 \cdot 10^{22}$.

Для рассмотрения вопроса проводимости в этих кристаллах будем изображать их структуру не в трех, а в двух измерениях, как это сделано на фиг. 5; здесь каждый атом углерода участвует в ковалентной связи с четырьмя окружающими его соседними атомами. Кристалл в целом, конечно, электрически нейтрален; это следует из рассмотрения в отдельности каждого атомного остатка и приходящейся на него доли заряда из четырех окружающих его ковалентных связей. Каждая такая ячейка, как это показано на фиг. 5, б, электрически нейтральна. В случае наличия в решетке кристалла примесных атомов, о чем речь будет дальше, такие ячейки не всегда нейтральны.

Исходя из описанной выше картины ковалентной связи, можно понять, почему алмаз должен быть изолятором. Как и в металле, в алмазе имеется большое число электронов, однако ковалентная связь обладает иными особенностями, нежели металлическая связь. Металлическая связь часто описывается следующим образом: валент-

ные электроны образуют нечто подобное газу из свободных электронов, создающему однородное облако отрицательного заряда, в котором „плавают“ положительные ионы металла. В алмазе, однако, электроны не ведут себя подобно электронному газу, заполняющему кристалл;



Фиг. 4. Структура алмаза, из которой видно, каким образом каждый атом образует четыре связи со своими ближайшими соседями.

a — постоянная решетки.

наоборот, здесь электроны образуют скорее структурные элементы, связывающие кристалл в одно целое. В идеальном кристалле в каждой валентной связи участвуют два электрона, поэтому каждый электрон прочно связан и не может принимать участие в процессах проводимости. Это положение можно иллюстрировать при помощи грубой аналогии с закрытым со всех сторон гаражом (фиг. 5, 2), в котором один из этажей полностью заставлен автомобилями. Естественно, что в таком гараже ни один из автомобилей двигаться не может¹⁾.

Однако проводимость в алмазе может быть создана самыми различными способами. Все эти способы связаны с нарушениями идеальной

¹⁾ Теория энергетических зон, рассматриваемая в основном в § 6 гл. 5, показывает, каким образом может происходить непрерывный переход от структуры с валентной связью к металлическим структурам. С этой точки зрения металлические свойства олова и свинца не противоречат рассматриваемой здесь теории. — Прим. авт.

Этот вырванный электрон, как показано на фиг. 6, в, образует в каком-то месте кристалла, которое до этого было электрически нейтральным, локализованный отрицательный заряд. Такой электрон, являющийся избыточным по отношению к тем, которые обеспечивают валентные связи соседних атомов, называют *избыточным электроном*. Так как этот электрон не может войти ни в одну из завершенных связей кристалла (см. фиг. 6, в), то под влиянием теплового движения он будет беспорядочно перемещаться по кристаллу. При приложении к кристаллу электрического поля этот электрон будет дрейфовать в направлении приложенной силы и создавать электрический ток.

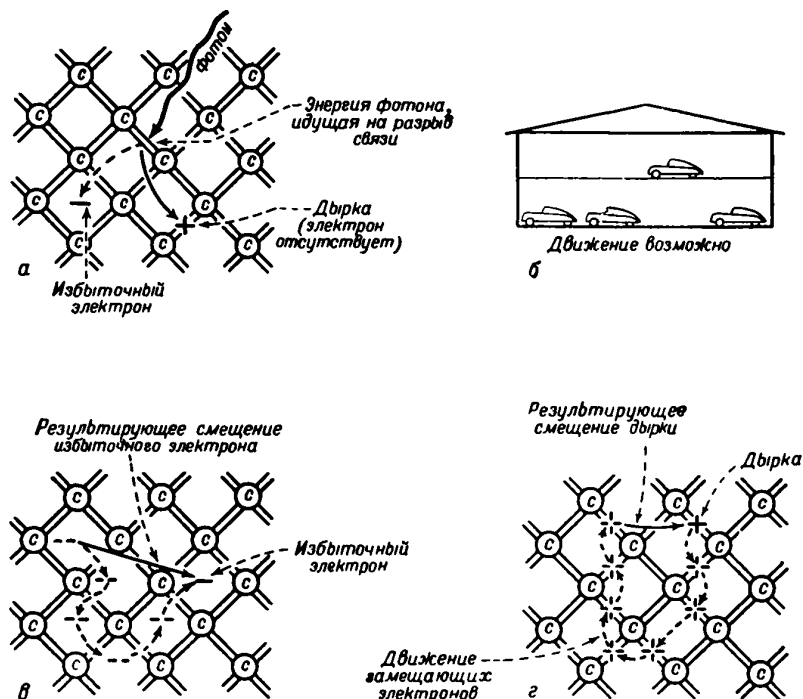
Поведение такого электрона сходно с положением автомобиля, находящегося на втором этаже гаража (см. фиг. 6, б), который может в этом случае свободно перемещаться. Подобно этому автомобилю электрон также должен быть переведен в состояние с более высокой энергией.

Как уже упоминалось, проводимость, обусловленная избыточными электронами, называется просто *электронной проводимостью*. Этот вид проводимости необходимо отличать от так называемой *дырочной* проводимости, к рассмотрению которой мы сейчас и перейдем.

Место, освободившееся на первом этаже гаража (см. фиг. 6, б) после подъема одного из автомобилей на второй этаж, может теперь перемещаться по первому этажу. Используя эту очень грубую аналогию, можно сказать, что нечто подобное происходит и в кристалле при движении дырок, образовавшихся в полосе валентных связей вследствие удаления электронов. Как это схематически представлено на фиг. 6, г, такая дырка образует локальный положительный заряд в местах кристалла, бывших до ее появления электрически нейтральными. Из этого же рисунка следует, что перемещение подобных дырок осуществляется при движении электронов в полосе валентных связей в обратном направлении (точно так же, как перемещение пустого места в гараже можно произвести при помощи последовательного перемещения автомобилей в противоположном направлении). Под влиянием электрического поля на беспорядочное движение дырки накладывается направленный дрейф и она также может принимать участие в переносе электрического тока ¹⁾.

¹⁾ В этой главе рассматривается самая упрощенная картина процесса проводимости. На самом деле, ни электроны, ни дырки нельзя рассматривать столь локализованными, как это показано на фиг. 6 или даже на фиг. 15. Процессы перемещения дырок не так просты. Изображенная здесь картина справедлива лишь в той мере, в какой она дает возможность разграничить два вида проводимости в полупроводниках. В гл. 5 и 8 эта картина будет представлена в более строгом виде. Основы теории полупроводников разработаны Вильсоном [7]. — *Прим. авт.*

Процесс протекания электрического тока в освещенном кристалле алмаза схематически изображен на фиг. 7. Электроны и дырки, парно образуемые фотонами, дрейфуют в электрическом поле в

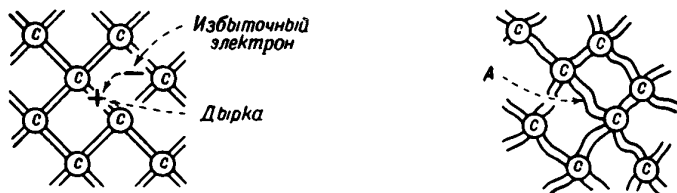
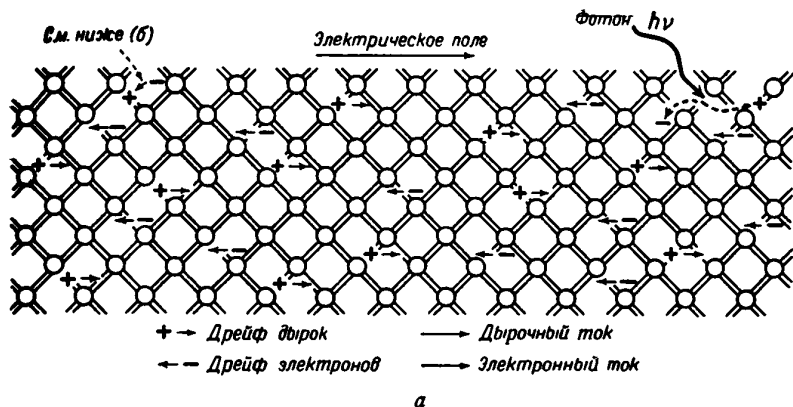


Фиг. 6. Избыточные электроны и дырки в кристалле алмаза.

а — образование фотонно пары электрон—дырка; в — беспорядочное движение избыточного электрона; г — беспорядочное движение дырки.

противоположных направлениях. Электрон со своим отрицательным зарядом перемещается в направлении, противоположном направлению приложенного электрического поля (т. е. налево), а создаваемый им ток совпадает, конечно, с направлением поля. Перемещение дырки осуществляется в результате движения электронов так же, как и для случая электронной проводимости, в направлении, противоположном электрическому полю (в среднем). При этом дырка дрейфует направо (см. фиг. 7, а), и, поскольку она несет положительный заряд, создаваемый ею ток также течет слева направо. [Это легко понять, если представить себе, что мы наклонили гараж, приподняв его правую часть (см. фиг. 6, б). В этом случае машины, находящиеся на верх-

нем этаже, будут перемещаться влево, а пустые места на первом этаже будут смещаться вправо.] Если прекратить освещение кристалла, то фотопроводимость исчезает и кристалл возвращается в свое



Фиг. 7. Фотопроводимость алмаза.

а — фотопроводимость: дрейф дырок и избыточных электронов в электрическом поле; б — рекомбинация дырки и электрона приводит к восстановлению нормальной валентной связи. А — тепловые колебания в решетке, возникающие в результате рекомбинации.

нормальное состояние в результате рекомбинации дырок и электронов¹⁾. Если электрон рекомбинирует с дыркой, то при этом происходит их взаимное уничтожение — валентная связь становится завершённой, а избыточная энергия передается атомам в форме тепловых колебаний, как это показано на схеме фиг. 7, б²⁾.

¹⁾ О рекомбинации электронов и дырок в германии см. [42, 43]. — *Прим. ред.*

²⁾ Процессы рекомбинации на самом деле, повидимому, более сложны и могут сопровождаться промежуточными состояниями, при которых захватывается либо электрон, либо дырка. Более детально об этих вопросах см. работу Мак-Кэя [8], а также § 1 гл. 3 и гл. 12. — *Прим. авт.*

На основе квантовомеханической теории (подробно этот вопрос рассмотрен в гл. 7) установлено, что в поведении электронов и „дырок“ существует весьма высокая степень симметрии. Пример заполненного автомобилями гаража для иллюстрации этой симметрии не пригоден. В гараже автомобилю, очевидно, легче будет перемещаться по второму этажу, нежели пустому месту по первому этажу. Согласно же квантовомеханической теории движения электронов в кристаллах, между упомянутыми выше процессами существует лишь незначительная разница и перемещающиеся через кристалл дырки можно рассматривать как положительно заряженные частицы, обладающие в основном, за исключением знака заряда, теми же свойствами, что и свободные электроны.

Ясно, что между движением электронов, возбужденных из валентной полосы и оставшихся в ней, имеется существенное различие. Движение как тех, так и других описывается при помощи теории зон *Брюллюэна* и *энергетических зон*. Согласно этой теории, рассматриваемой в гл. 5 и 7, электроны валентных связей заполняют ряд энергетических уровней, образующих зону дозволённых энергий.

В идеальном кристалле все энергетические уровни этой зоны заполнены электронами, а поэтому зону называют *заполненной зоной валентной связи*, или просто *валентной зоной*. Эта зона соответствует первому этажу гаража. Электрон не может обладать более высокой энергией, чем та, которая соответствует наивысшему энергетическому состоянию заполненной зоны, до тех пор, пока при достаточно большом увеличении его энергии он не становится свободным или избыточным электроном в рассмотренном выше смысле. В свободном состоянии электрон также может обладать различными значениями энергии и, следовательно, занимать различные уровни в энергетической схеме *пустой зоны*, или *зоны проводимости* (второй этаж).

Заполненная зона также является зоной, в которой возможна проводимость, но ее не принято называть зоной проводимости. Заполненная и свободная зоны разделены областью запрещённых энергий, в которой не существует энергетических уровней для электронов кристалла.

В этой вводной главе мы будем пользоваться результатами зонной теории, не ссылаясь на ее теоретическую основу.

При достаточном повышении температуры тепловые колебания начинают спонтанно разрушать валентные связи, что приводит к образованию свободных электронов и такого же количества дырок. В алмазе это явление могло бы наступить лишь при столь высоких температурах, при которых само существование алмаза становится уже невозможным. Однако в кремнии и германии этот процесс играет существенную роль уже при температурах, легко достижимых в лабораторных условиях.

§ 2. ПРИМЕСНЫЕ ПОЛУПРОВОДНИКИ. ДОНОРЫ И АКЦЕПТОРЫ

Если бы при исследовании полупроводников мы встречались только с теми случаями, когда проводимость их обусловлена электронами и дырками, присутствующими в равных количествах, то интерпретация полученных данных представляла бы большие трудности. К счастью, кремний и германий являются такими полупроводниками, проводимость которых при не очень высоких температурах обуславливается либо одними избыточными электронами, либо одними дырками. Сначала рассмотрим некоторые специальные примеры для кремния¹⁾, а затем укажем, как полученные результаты относятся к германию.

Полупроводник, проводимость которого в основном обусловлена наличием избыточных электронов в почти свободной зоне, называется полупроводником *n*-типа, а полупроводник, проводимость которого вызывается наличием дырок в почти заполненной зоне, называется полупроводником *p*-типа²⁾.

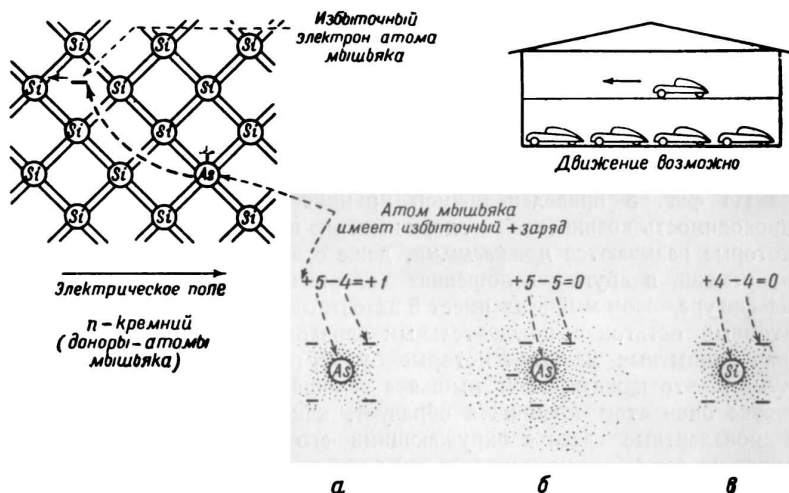
На фиг. 8 приведен пример кремния *n*-типа. В данном случае проводимость возникает благодаря наличию в кремнии атомов мышьяка, которые называются *примесными*, даже если они специально введены в кремний, в других отношениях совершенно чистый. Как показано на фигуре, атом мышьяка имеет 5 валентных электронов, окружающих атомный остаток с положительным зарядом $+5$ единиц. На основании опытных данных, которые будут рассмотрены ниже, можно думать, что каждый атом мышьяка замещает в кристаллической решетке один атом кремния и образует, как это показано на фиг. 8, 4 ковалентные связи с окружающими его атомами кремния, затрачивая на это 4 (из имеющихся пяти) валентных электрона. Оставшийся пятый валентный электрон, не участвующий в создании ковалентной связи, может свободно перемещаться по кристаллу. Как уже указывалось выше, этот избыточный электрон составляет *подвижный* локализованный отрицательный заряд. С другой стороны, атом мышьяка является *неподвижным* локализованным положительным зарядом, поскольку атомный остаток его, обладающий зарядом $+5$, полностью не нейтрализуется зарядом (-4) электронов, обуславливающих валентную связь. Поэтому его результирующий положительный заряд как раз компенсируется оставшимся в кристалле избыточным электроном. Следовательно, атомы мышьяка добавляют в решетку кремния избыточные электроны, не нарушая при этом электрической нейтраль-

1) Систематическое исследование свойств кремния с примесями было впервые осуществлено в работе Скаффа, Тёерера и Шумахера [9], появление которой было вызвано стремлением усовершенствовать кремниевые детекторы для микроволн, нашедшие еще в предвоенные годы практическое применение. — *Прим. авт.*

2) *n* и *p* соответствуют первым буквам слов „negative“ — отрицательный (электронная проводимость) и „positive“ — положительный (дырочная проводимость). — *Прим. ред.*

ности кристалла в целом. Эти электроны притягиваются положительными атомами мышьяка и при низких температурах связываются с ними. Однако уже при комнатных температурах тепловые колебания решетки нарушают эти связи, и электроны (см. фиг. 8) становятся свободными. (Ценные данные об этом процессе можно получить, как это будет показано ниже, из измерений электропроводности в широком интервале температур.)

Полупроводник *p*-типа изображен на фиг. 9. В этом случае в качестве примеси введены атомы бора; они имеют валентность, равную 3,

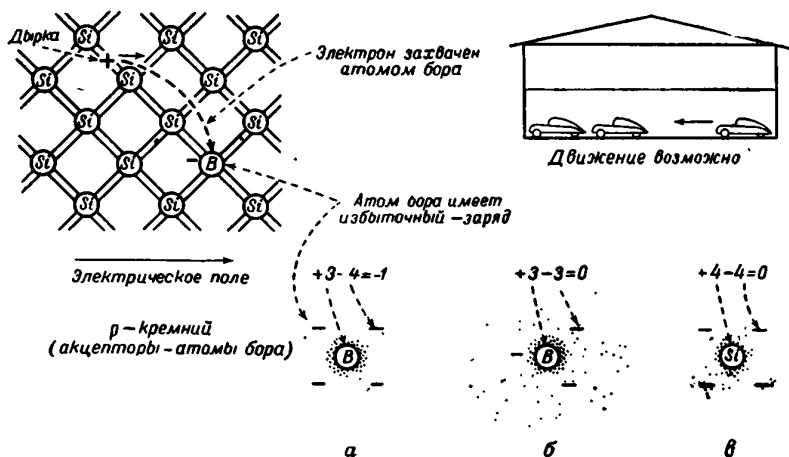


Фиг. 8. Избыточная, или электронная, проводимость кремния с примесью мышьяка.

a — атом мышьяка в кристалле кремния; *б* — свободный атом мышьяка; *в* — нейтральный атом кремния в кристалле.

и поэтому не могут полностью завершить окружающую их структуру валентных связей. Дырка в одной из связей атома бора может быть заполнена электроном из смежной связи, вследствие чего, как это показано на фиг. 6, 2, дырка получает возможность мигрировать по кристаллу; бор, таким образом, становится неподвижным локализованным отрицательным зарядом. Благодаря симметрии свойств дырок и электронов изображенную на фиг. 9 картину можно описать следующим образом: атом бора притягивает к себе положительно заряженную дырку, но вследствие тепловых колебаний эта связь между ними уже при комнатной температуре нарушается, дырка становится свободной и блуждает по кристаллу, обуславливая проводимость.

Примесные атомы с валентностью, равной 5, называются *донорными примесями*, или *донорами*, поскольку они отдают избыточный электрон кристаллу; трехвалентные примесные атомы называются *акцепторными примесями*, или *акцепторами*, так как они захватывают по одному электрону откуда-нибудь из кристалла для заполнения структуры валентных связей, создавая тем самым в кристалле дырку, участвующую в электропроводности¹⁾. Все эти термины вместе с другими характеристиками полупроводников, о которых речь шла уже выше, сведены в табл. 1.



Фиг. 9. Дырочная проводимость кремния с примесью бора.

a — атом бора в кристалле кремния; *б* — свободный атом бора; *в* — нейтральный атом кремния в кристалле.

На фиг. 10 показана зависимость концентрации носителей тока от температуры для гипотетического образца кремния с пятью донорами и двумя акцепторами. При температуре абсолютного нуля все три добавочных электрона, доставляемые этими семью примесными атомами, связаны с донорами. С повышением температуры до -125°C , или 148°K , тепловые колебания становятся достаточными для отрыва электронов. Однако эти электроны так быстро рекомбинируют, что в среднем только один из трех свободен и может участвовать в проводимости. При комнатной температуре уже два электрона из трех могут участвовать в проводимости; при этом, однако, тепловые колебания еще недостаточно велики, чтобы разорвать валентные связи,

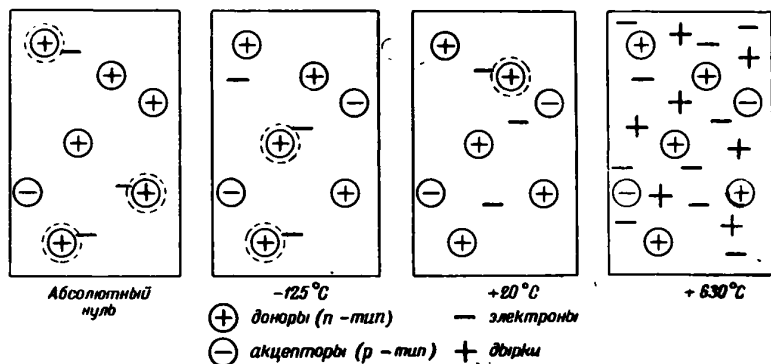
¹⁾ О механизме примесной проводимости в кристаллах кремния или германия см., например, [21, 24, 34]. — *Прим. ред.*

Таблица 1

СВОДКА НЕКОТОРЫХ ДАННЫХ О ПРИМЕСНОЙ ПРОВОДИМОСТИ В КРЕМНИИ И ГЕРМАНИИ

Тип проводимости	Проводимость обуславливается	Энергетическая зона, в которой движутся носители тока	Знак носителей тока	Валентность примесного атома	Наименование примесного атома	Типичные примеси
<i>n</i> -тип, или избыточная (электронная) проводимость	(Избыточными) электронами	Свободная зона, или зона проводимости	Отрицательный	5	Донор	Элементы V группы: фосфор, P; мышьяк, As; сурьма, Sb
<i>p</i> -тип, или дефектная (дырочная) проводимость	Дырками	Заполненная зона, или зона валентных электронов	Положительный	3	Акцептор	Элементы III группы: бор, B; алюминий, Al; галлий, Ga; индий, In

и число дырок составляет всего 10^{-12} числа электронов¹⁾. При комнатной температуре в образцах германия с высоким сопротивлением концентрация дырок составляет около 1% от концентрации электронов²⁾. При более высоких температурах заметная доля валентных



Фиг. 10. Температурная зависимость концентрации дырок и электронов в кремнии *n*-типа.

¹⁾ Эта оценка получена экстраполяцией рассматриваемой ниже кривой фиг. 13 при помощи кривой фиг. 138 (стр. 550). — *Прим. авт.*

²⁾ Вопрос о свойствах германия при комнатной температуре рассматривается в § 4 гл. 10 и иллюстрируется фиг. 93 (стр. 298). — *Прим. авт.*

связей разрушается тепловым движением, причем попарно создаются электроны и дырки. На фиг. 10 показано, что при температуре 630°C количество термически созданных носителей тока уже в 6 раз превышает количество носителей, вносимых примесями, и проводимость практически имеет то же значение, какое она имела бы, если бы в этом образце совсем отсутствовали примеси. В этом случае проводимость образца принято называть *собственной* (intrinsic), поскольку она определяется свойствами самого кремния и не зависит от содержания примесей.

Картина, изображенная на фиг. 10, примерно соответствует положению, которое имело место в образце кремния, исследованном Пирсоном и Бардином (см. фиг. 13).

§ 3. ИНТЕРПРЕТАЦИЯ ДАННЫХ ПО ПРОВОДИМОСТИ И ЭФФЕКТУ ХОЛЛА

Пирсон и Бардин [10] исследовали большое количество образцов кремния с различным содержанием примеси фосфора и бора. Мы приведем здесь некоторые данные, заимствованные из работы этих авторов. Однако прежде чем приступить к рассмотрению результатов этой работы, мы кратко, в самых общих чертах, остановимся на вопросах экспериментальной методики, применявшейся для получения основных данных о концентрации и движении электронов и дырок в полупроводниках.

Проводимость полупроводника, содержащего электроны и дырки, зависит от их числа и от легкости их перемещения при приложении к образцу электрического поля. (Эти вопросы подробно рассматриваются в гл. 8, 10 и 11.) Скорость движения электронов и дырок в кристалле при приложении электрического поля называется *подвижностью* и определяется как скорость дрейфа частицы в см/сек в электрическом поле напряженностью в 1 в/см . Подвижность имеет поэтому размерность $\text{см}^2/\text{в.сек}$. Проводимость (в $\text{ом}^{-1} \cdot \text{см}^{-1}$), обусловленная избыточными электронами, равна произведению суммарного заряда имеющихся в единице объема вещества свободных электронов (в кулон/см^3) на величину их подвижности. В свою очередь, полный заряд единицы объема полупроводника равен, очевидно, произведению концентрации электронов (в см^{-3}) на заряд электрона e ($1,60 \cdot 10^{-19}$ кулон). Если проводимость обусловливается не только движением электронов, но и дырок, то общее выражение для электропроводности полупроводника можно записать в виде

$$\sigma = e(n\mu_n + p\mu_p), \quad (1.1)$$

где n и p (соответственно для отрицательных и положительных зарядов) — концентрации электронов и дырок, а μ_n и μ_p — их подвижности. (Вывод этой формулы сделан при помощи фиг. 79, стр. 247.)

Измерение только одной электропроводности дает поэтому лишь односторонние сведения о четырех неизвестных, входящих в уравнение (1.1). К счастью, другое соотношение между этими же четырьмя неизвестными величинами можно получить из измерений эффекта Холла.

Эффект Холла, более подробно рассмотренный в гл. 8, наблюдается в том случае, когда к образцу, через который проходит электрический ток, под прямым углом к направлению тока приложено магнитное поле. Установлено, что при этом в образце возникает поперечная разность потенциалов, которая и может быть измерена. Измерения эффекта Холла дают второе уравнение, которое, так же как и (1.1), содержит указанные выше 4 неизвестные величины. Но если измерения проводятся на образце с проводимостью n -типа и притом в области температуры, где имеет место лишь одна примесная проводимость, то в этом случае концентрацию дырок p можно считать равной нулю. В этих условиях измерения проводимости и эффекта Холла дают возможность в отдельности определить значения n и μ_n . Аналогично можно определить p и μ_p в образце с проводимостью p -типа. Сопоставляя, таким образом, данные, полученные на различных образцах, можно определить зависимость μ_n и μ_p от температуры. Зная μ_n и μ_p , легко определить значения n и p даже в случае одновременного участия в проводимости электронов и дырок. Значения подвижности носителей тока в алмазе, кремнии и германии приводятся в табл. 2¹⁾. Методы измерения подвижности рассматриваются в § 1 гл. 3, а также в гл. 8, 11 и 12.

Перейдем теперь к рассмотрению данных, полученных Пирсоном и Бардином для ряда образцов кремния, содержавших различные количества добавленного фосфора. Фосфор, так же как и мышьяк, — пятивалентный элемент и создает проводимость n -типа. На фиг. 11 представлена температурная зависимость сопротивления четырех образцов кремния с различным содержанием фосфора²⁾.

На фиг. 12 изображен температурный ход концентрации избыточных электронов для тех же образцов кремния. По оси ординат на

¹⁾ Электрические свойства серого олова (α -Sn) были достаточно подробно изучены лишь в последнее время в работах [44, 45] и [46]. — *Прим. ред.*

²⁾ Указанная для каждого образца концентрация атомов фосфора, рассчитана по количеству фосфора, добавленного в расплавленный кремний. Часть этого фосфора, вероятно, испаряется еще до того, как он успевает смешаться с затвердевающим кремнием; кроме того, следует заметить, что при кристаллизации расплава происходит сегрегация (неравномерное распределение примеси). По мере затвердевания сплава концентрация фосфора в остающейся еще жидкой части сплава возрастает. В результате концентрация фосфора в образцах, определенная на основании измеренного эффекта Холла, оказывается ниже расчетной. При этом предполагается, что каждый донорный атом в решетке кристалла создает один электрон проводимости. Экспериментальное доказательство этого предположения получено на образцах германия с примесью радиоактивной сурьмы. Мы рассмотрим этот вопрос подробнее в конце главы. — *Прим. авт.*

Таблица 2

ПОДВИЖНОСТЬ НОСИТЕЛЕЙ ТОКА, $\text{см}^2/\text{в} \cdot \text{сек}$

	Подвижность электронов μ_n	Подвижность дырок μ_p
Углерод (алмаз)	>400 [5] 900 [11] 156 [12] 300 [10]	>200 [5] — — 100 [10]
Кремний*	300 [10]	100 [10]
Германий**	2 600 [13]	1 700 [13]
(„холловская“ подвижность)	3 600***	1 700***
(„дрейфовая“ подвижность)		

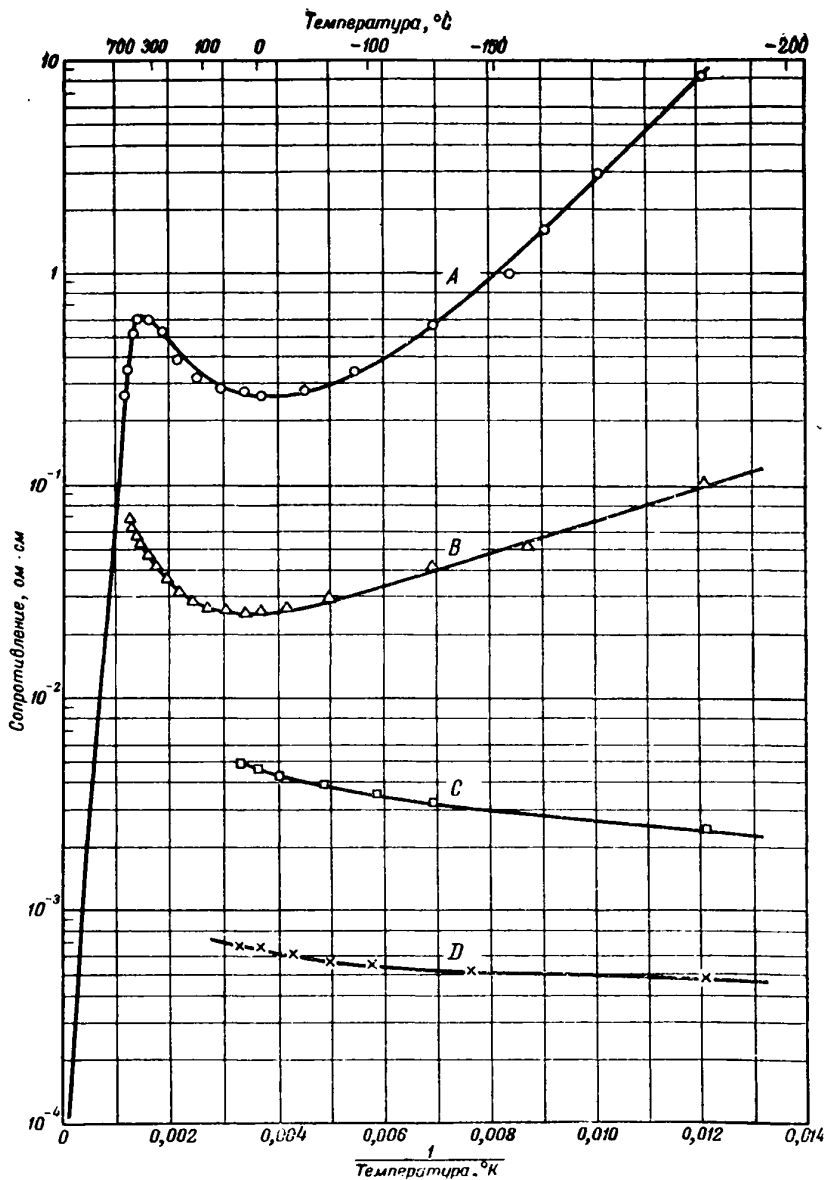
* Подвижность электронов и дырок в монокристаллах *p*-Si и *n*-Si соответственно была недавно измерена методом дрейфа [48]. Среднее значение подвижности электронов при комнатной температуре оказалось равным $1\,210 \text{ см}^2/\text{в} \cdot \text{сек}$, а подвижности дырок — $250 \text{ см}^2/\text{в} \cdot \text{сек}$. Малые значения подвижности, приведенные в табл. 2 и полученные из данных по эффекту Холла и электропроводности, связаны, вероятно, с неоднородностями, вносимыми границами раздела между отдельными зернами поликристаллического образца. — Прим. ред.

** Дюлап [49] определял подвижность электронов и дырок в монокристаллах высокоомного германия ($\rho_{20^\circ} = 18,8 \text{ ом} \cdot \text{см}$). Подвижность электронов оказалась равной $3\,800 \text{ см}^2/\text{в} \cdot \text{сек}$, а подвижность дырок — $2\,730 \text{ см}^2/\text{в} \cdot \text{сек}$. При температуре жидкого воздуха подвижность электронов достигает $18\,000 \text{ см}^2/\text{в} \cdot \text{сек}$. — Прим. ред.

*** Неопубликованные данные, сообщенные Хайнсом на заседании физического общества в Чикаго (ноябрь 1949 г.).

этом графике отложена концентрация электронов проводимости. Так как в кремнии в каждом см^3 содержится $5,00 \cdot 10^{23}$ атомов, то нетрудно убедиться, что в образце *D* один электрон приходится на 250 атомов кремния. В образце *A* при комнатной температуре лишь один электрон проводимости приходится в среднем на 500 000 атомов кремния. На фиг. 13 приведен температурный ход концентрации электронов и дырок в этом образце как в области примесной, так и собственной проводимости. На графике нанесены как экспериментальные точки, так и теоретические кривые (сплошные линии), вычисленные на основе схемы энергетических уровней полупроводника и статистики Ферми — Дирака (см. последующие главы). Из этого графика видно, что теоретическая кривая хорошо совпадает с экспериментальными данными. Физические представления, лежащие в основе теоретической интерпретации температурного хода концентрации носителей, уже рассматривались ранее в связи с фиг. 10. Концентрация акцепторов и доноров в образце *A* составляет соответственно $1,5 \cdot 10^{16}$ и $12 \cdot 10^{16} \text{ см}^{-3}$ (отношение этих концентраций равно 1 : 8, в то время как на фиг. 10 было выбрано отношение 2 : 5).

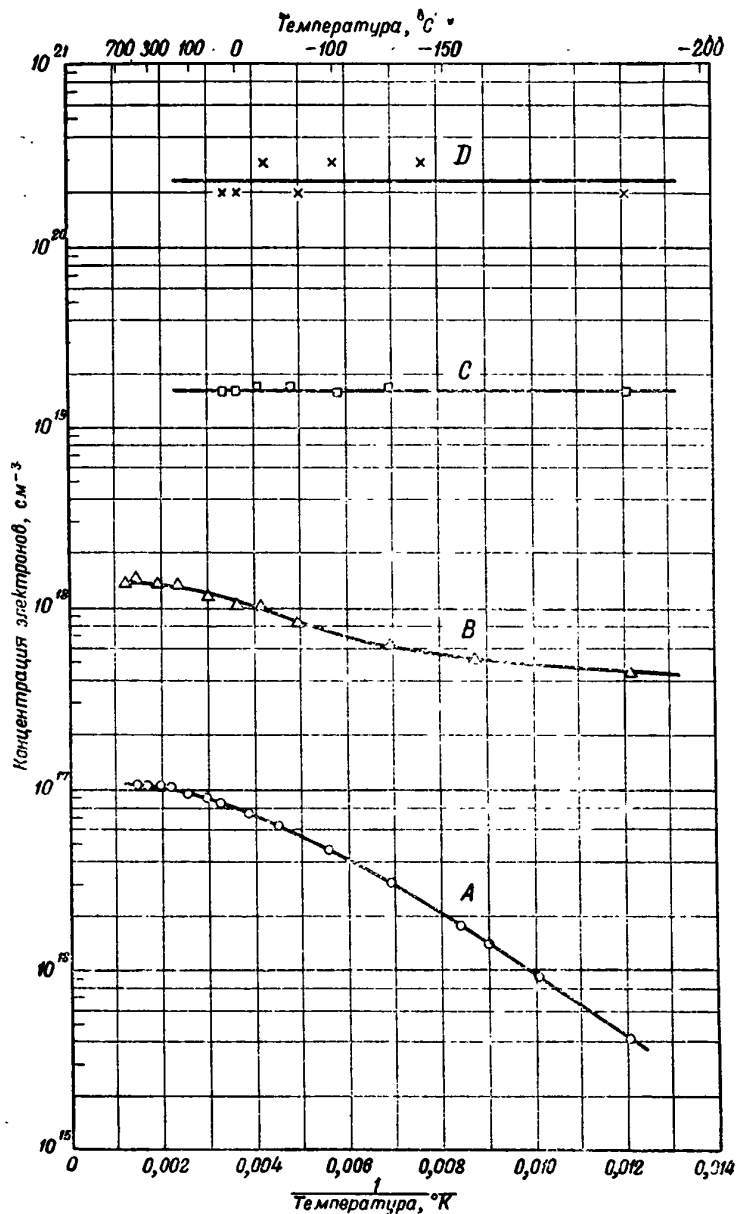
Данные о подвижности электронов в зависимости от температуры для различных образцов кремния с примесью фосфора приведены на фиг. 14. Следует подчеркнуть, что при высоких температурах здесь имеет место *собственная подвижность* (intrinsic mobility), по крайней мере для образцов *A* и *B* (об образцах *C* и *D* речь будет



Ф и г. 11. Зависимость сопротивления сплавов кремния с фосфором от температуры.

Концентрация атомов фосфора в расплаве соответственно равна:

- Образец A 4,7 · 10¹⁷ см⁻³
- " B 2,7 · 10¹⁸ " "
- " C 4,7 · 10¹⁹ " "
- " D 4,7 · 10²⁰ " "



Фиг. 12. Концентрация электронов проводимости в сплавах кремния с фосфором в зависимости от температуры.

Концентрация атомов фосфора в расплаве соответственно равна:

Образец	A	4,7 · 10 ¹⁷	см ⁻³
"	B	2,7 · 10 ¹⁸	"
"	C	4,7 · 10 ¹⁹	"
"	D	4,7 · 10 ²⁰	"

идти ниже), которая обуславливается взаимодействием электронов с тепловыми колебаниями атомов кристалла. При малой концентрации примесных атомов подвижность носителей заряда можно считать не зависящей от содержания примесей. При низких температурах примесные центры уменьшают подвижность; в этом случае процессы, ограничивающие скорость дрейфа электронов в электрическом поле, сводятся к отклонению электронов электростатическим полем заряженных примесных атомов. Эти вопросы рассматриваются в гл. 11.

Читателю, вероятно, известно, что по кривым, подобным тем, которые изображены на фиг. 12 и 13, можно определить энергию активации точно так же, как это делается по ричардсоновским кривым для термоионной эмиссии или из температурной зависимости постоянной химической реакции. Во всех этих случаях мы встречаемся с выражением вида

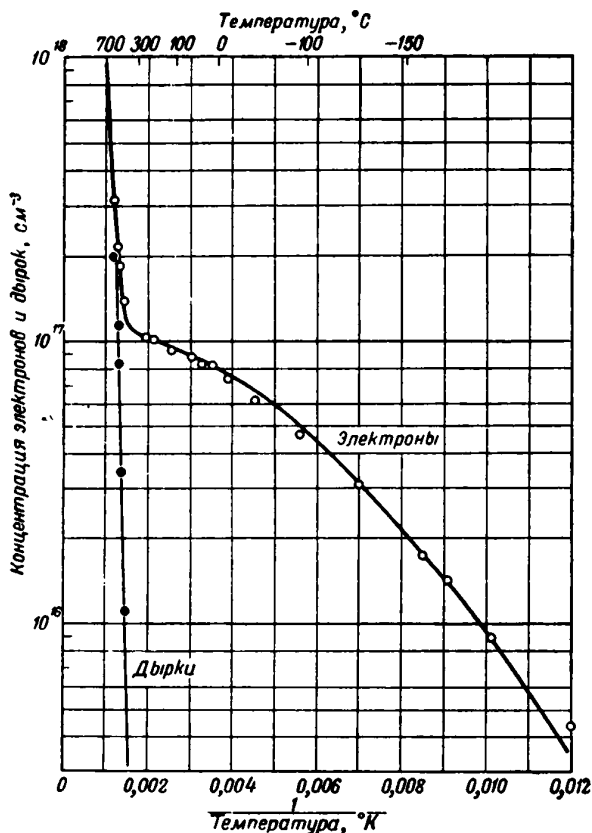
$$y = e^{-\mathcal{E}/kT}, \quad (1.2)$$

где y — исследуемая величина, \mathcal{E} — „энергия активации“ процесса, а kT — тепловая энергия. Если построить график зависимости $\ln y$ от $1/T$, то получим прямую линию, уравнение которой имеет вид

$$\ln y = -\frac{\mathcal{E}}{k} \left(\frac{1}{T} \right). \quad (1.3)$$

Угловой коэффициент прямой равен \mathcal{E}/k . Поскольку k (постоянная Больцмана) известна, всегда можно определить значение \mathcal{E} .

По наклонам прямолинейных участков графиков $\ln n = f_1(1/T)$ и $\ln p = f_2(1/T)$ можно определить 3 энергии активации, представляющие для нас особое значение. [Анализ этих графиков сводится, однако, не только к простому определению угловых коэффициентов (наклонов) прямолинейных участков.] Здесь мы имеем в виду: 1) энергию, необходимую для разрыва валентной связи и образования пары электрон — дырка, 2) энергию, необходимую для отрыва электрона от донора, и 3) энергию отрыва дырки от акцептора. Энергия, необходимая для разрыва валентной связи, легко может быть определена по наклону прямой на фиг. 13 в области собственной проводимости. Теория показывает, что из наклона этого участка температурной зависимости концентрации носителей тока определяется значение $\mathcal{E}_1/2k$, где \mathcal{E}_1 — энергия, необходимая для разрыва валентной связи. Значительно более сложен анализ значений \mathcal{E}_2 в образцах с проводимостью n -типа или \mathcal{E}_3 в образцах с проводимостью p -типа. При очень низких температурах следует ожидать, что наклоны будут равны \mathcal{E}_2/k , а не $\mathcal{E}_2/2k$. (Подробнее об этом см. фиг. 136 и 137, стр. 548 и 549.)



Фиг. 13. Сопоставление теоретических и экспериментальных значений концентраций электронов и дырок при различных температурах.

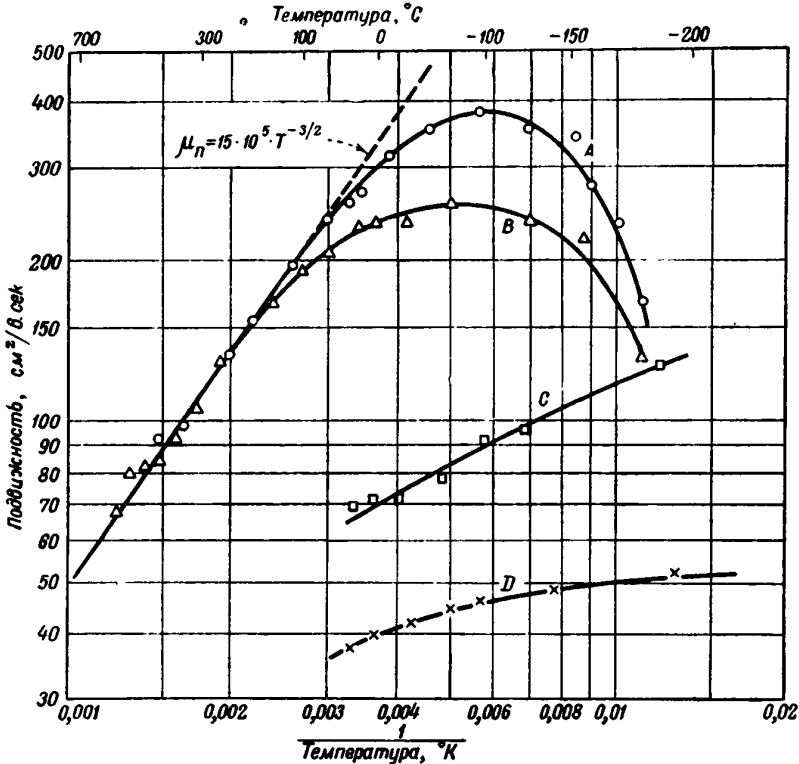
— теоретическая кривая; O ● — экспериментальные точки. Обе кривые относятся к образцу А, для которого концентрация атомов фосфора в расплаве составляет $4,7 \cdot 10^{17} \text{ см}^{-3}$.

Это необходимо иметь в виду при определении \mathcal{E}_2 и \mathcal{E}_3 . Пирсон и Бардин получили для кремния следующие данные:

Энергия разрыва валентной связи	1,11 эв
" отрыва электрона от донора	0,054 "
" " дырки от акцептора	0,08 "

(Для германия столь точные данные не были опубликованы; мы приняли для германия следующие характерные значения энергии активации: 0,72 и 0,04 эв. Для алмаза энергия разрыва валентной связи,

вычисленная из данных по поглощению в ультрафиолетовой области спектра, составляет 6 или 7 эв. Энергия отрыва электронов от



Фиг. 14. Зависимость подвижности электронов в сплавах кремния с фосфором от температуры.

Концентрация атомов фосфора в сплаве соответственно равна:

Образец А	$4,7 \cdot 10^{17}$	см^{-3}
•	$2,7 \cdot 10^{18}$	•
•	$4,7 \cdot 10^{19}$	•
•	$4,7 \cdot 10^{21}$	•

доноров или дырок от акцепторов составляет, повидимому, величину порядка 0,35 эв.)¹⁾

¹⁾ Данные о поглощении алмаза в ультрафиолетовой области спектра приведены в работе Ахирна [14]. Он установил также, что при бомбардировке кристалла германия α -частицами для образования каждого электрона проводимости затрачивается около 10 эв. По данным Мак-Кэя [5], облучение алмаза электронами с энергией 14 кэ также приводит к появлению проводимости, причем на образование одного электрона проводимости расходуется около 10 эв. Диэлектрическая постоянная алмаза, вычисленная по коэффи-

Малые значения энергий связи электронов с донорами и дырок с акцепторами можно объяснить, основываясь на модели, предложенной Бете. Предполагается, что электрон движется в поле донора примерно так же, как и электрон вокруг протона в атоме водорода¹⁾. Однако при этом учитывается одно существенное обстоятельство: поскольку результирующий положительный заряд иона фосфора находится в диэлектрической среде, сила притяжения между электроном и ионом уменьшается обратно пропорционально диэлектрической постоянной (приблизительно в 13 раз для кремния). Поэтому связанный с атомом фосфора электрон движется по очень большой орбите, радиус которой в 13 раз больше радиуса орбиты электрона в атоме водорода, что дает для энергии связи величину порядка 0,08 эв (см. гл. 9). Волновая функция в этом случае перекрывает большое число атомов кремния (фиг. 15). Тот факт, что волновая функция простирается на область кристалла, занятую большим числом атомов, означает, что детальный вид ее вблизи самого примесного атома не имеет существенного значения и что энергия связи электрона с примесным атомом определяется почти исключительно зарядом примесного атома. Этим и объясняется наблюдающийся на опыте факт независимости энергии связи электронов с донорами и дырок с акцепторами от сорта донорных и акцепторных примесных атомов. Различие в значениях энергии связи электронов и дырок Пирсон и Бардин приписывают различию их эффективных масс (вопрос этот рассматривается в гл. 7). Более того, приведенные выше значения энергии активации получены ими лишь для самых чистых образцов; для других образцов с большим содержанием примесей энергия активации еще меньше. Авторы объясняют это тем фактом, что взаимодействие между примесными атомами приводит к уменьшению энергии отрыва электрона от примесного атома²⁾.

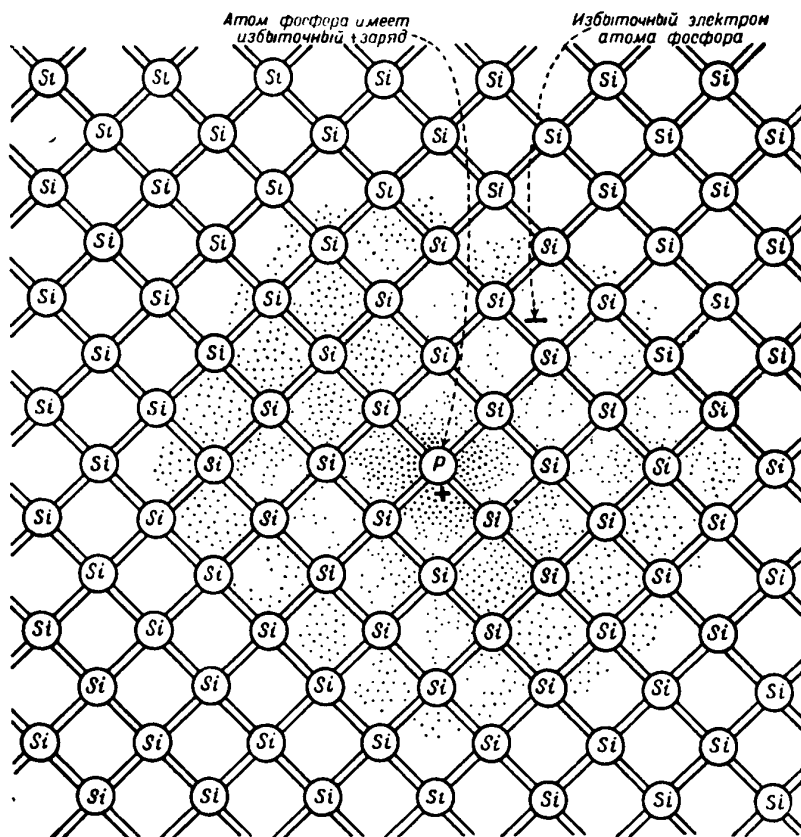
При относительно больших концентрациях примесей возникают новые эффекты. Концентрация электронов в образцах кремния *C* и *D* (см. фиг. 12) почти не зависит от температуры. Объяснение этого факта заключается в том, что концентрация примесей здесь настолько велика, что наступает перекрытие волновых функций примесных центров, подобных изображенному на фиг. 15. В этом случае „избыточные“ электроны и примесные атомы играют почти ту же роль, что электроны и ионы в металле: избыточные электроны образуют вырожденный электронный газ, который движется через систему беспорядочно расположенных в пространстве положительных ионов. Теория электропроводности таких полупроводников весьма близка к теории электропроводности металлических сплавов и будет цо-

цену преломления, составляет величину порядка 6,2. Мак-Кэй установил, что в алмазе электроны захватываются ловушками с энергией связи порядка 0,3 эв. — *Прим. авт.*

¹⁾ Такой вид приближения впервые был введен Ванисером [15]. — *Прим. авт.*

²⁾ См. в связи с этим [49] и [50]. — *Прим. ред.*

дробно изложена в гл. 11. Электропроводность таких полупроводников значительно меньше электропроводности металлических сплавов, так как концентрация носителей тока в них меньше; в остальном же, если не считать различия в величине электропроводности, свойства



Фиг. 15. Волновая функция электрона, связанного с атомом фосфора в кремнии.

Плотность точек приближенно соответствует распределению волновой функции.

образцов *C* и *D* при низких температурах весьма сходны со свойствами металлических сплавов. Однако при приближении к области собственной проводимости число носителей тока и проводимость быстро возрастают с ростом температуры.

Такие же примерно результаты получены и для дырок в кремнии, а также для электронов и дырок в германии. Существенной и характерной особенностью германия является столь малое значение энергии, необходимой для отрыва электронов от доноров и дырок от акцепторов, что уже при комнатной температуре лишь ничтожно малая часть избыточных электронов и дырок связана с примесями центрами. Иными словами, при комнатной температуре доноры и акцепторы в германии полностью ионизированы. Это обстоятельство значительно упрощает создание теории электронных явлений в германиевых транзисторах.

Германий обладает и другой интересной особенностью: при термической обработке при температуре выше 500°C в нем возникают примесные центры, создающие электропроводность p -типа. Длительная термическая обработка при более низких температурах приводит к исчезновению этих примесей. Одна из существующих теорий заключается в том, что этот эффект обуславливается на самом деле совсем не химическими примесями, а связан с явлением разупорядочения решетки германия. Определенное на основании теории общее число разупорядоченных атомов примерно соответствует наблюдаемому на опыте значению концентрации носителей заряда ¹⁾.

Ларк-Горовитц и его сотрудники [17—19] получили непосредственное доказательство того, что разупорядочение решетки германия приводит к возникновению в нем проводимости p -типа. При бомбардировке α -частицами или дейтронами германия n -типа высокого сопротивления последний переходит в германий p -типа, а его проводимость возрастает с увеличением времени облучения ²⁾. Установлено, что этот эффект не связан с превращением германия в другой элемент, а возникает лишь при нарушении правильности кристалла, т. е. вызывается переходом атомов германия в результате соударения с бомбардирующими частицами в междоузлия с образованием при этом вакантных узлов в решетке ³⁾.

Такие „дефекты решетки“ действуют как акцепторы. Однако нет никаких оснований предполагать, что волновые функции, характеризующие поведение электронов вблизи таких дефектов, должны быть

¹⁾ Доля разупорядоченных атомов при температуре T должна быть приблизительно пропорциональна $\exp(-\mathcal{E}/kT)$, где \mathcal{E} — энергия разупорядочения на один атом. Из данных по самодиффузии в металлах [16] для \mathcal{E} получаются значения, лежащие в пределах от $10 kT_m$ до $20 kT_m$, где T_m — температура плавления. Это дает для экспоненты значения порядка 10^{-7} — 10^{-10} . Наблюдаемые на опыте данные соответствуют примерно 10^{15} акцепторным центрам, или 10^{-8} всех атомов. — *Прим. авт.*

²⁾ Обзор работ по облучению полупроводников электронами и нейтронами, а также фотонами различной энергии опубликован Ларк-Горовитцом [51]. — *Прим. ред.*

³⁾ Данные о влиянии бомбардировки поверхности кремния газовыми ионами с энергией до $50\,000$ эв на свойства кремниевых детекторов приведены в работе Оля [52]. Изменение структуры поверхностного слоя кристаллов под действием α -частиц описано в работе [53]. — *Прим. ред.*

сходны с волновыми функциями, изображенными на фиг. 15, и, следовательно, можно ожидать, что подобные образцы будут иметь различные значения энергии активации при низких температурах. Данные, относящиеся к этому вопросу, рассматриваются в гл. 9.

Прежде чем закончить вводную главу, целесообразно привести еще некоторые дополнительные экспериментальные данные, подтверждающие представление о том, что проводимость германия и кремния обуславливается примесными атомами, образующими твердые растворы замещения. Следующие данные свидетельствуют о том, что примесные атомы преимущественно замещают атомы основного вещества в узлах решетки, а не внедряются в междоузлия:

1. Проводимость, вызываемая примесями, оказывается проводимостью именно такого рода, какой и следовало бы ожидать, исходя из того, заполняет или не заполняет примесный атом, замещающий нормальный узел решетки, окружающую его структуру валентных связей (см. табл. 1). С другой стороны, если бы введенные в решетку атомы примеси располагались в междоузлиях, то они вообще не входили бы в схему валентной связи и поэтому не было бы оснований предполагать, что введение в качестве примеси атомов трехвалентных элементов должно приводить к возникновению дырочной проводимости, а атомов пятивалентных элементов — к появлению электронной проводимости.

2. Значение энергии активации, необходимой для отрыва электрона от атома примеси, как это показано в упомянутой работе Пирсона и Бардина, находится в хорошем согласии с теоретическим значением, ожидаемым в соответствии с ходом волновой функции, изображенным на фиг. 15.

3. Существует также непосредственное доказательство, основанное на рентгенографическом исследовании постоянной решетки сплавов кремния с фосфором и бором более высокой концентрации, подтверждающее, что в решетке примеси располагаются в виде атомов замещения. Постоянная решетки таких сплавов определялась в зависимости от концентрации примеси; было показано, что по мере введения атомов фосфора и бора она уменьшается. Такое уменьшение следует ожидать, когда вводимые атомы примеси образуют с кремнием твердые растворы замещения, так как по своим размерам эти атомы меньше атомов кремния. Напротив, если бы атомы фосфора или бора внедрялись в междоузлия, то это приводило бы к увеличению постоянной решетки, что находится в противоречии с опытом.

Установлено, что даже в случае относительно очень чистых образцов германия, в которых содержание примеси столь мало, что не поддается определению при помощи химического анализа, можно контролировать величину проводимости путем введения ничтожных количеств примеси. Эти исследования проводились с радиоактивной сурьмой, которая вводилась в виде присадок в расплав германия. Количество

введенной сурьмы контролировалось при помощи радиоактивных измерений. Опыт дал хорошее согласие между содержанием сурьмы и числом электронов проводимости [20].

ЗАДАЧИ

1. Слиток германия получен сплавлением 100 г германия и $3,22 \cdot 10^{-6}$ г сурьмы. Предполагая, что сурьма распределена в образце равномерно, нужно показать, что концентрация атомов сурьмы при этом равна $8,70 \cdot 10^{14} \text{ см}^{-3}$. [Плотность германия $5,46 \text{ г/см}^3$, а его атомный вес 72,60; атомный вес сурьмы 121,76; 1 граммом, т. е. 72,60 г, германия содержит $6,02 \cdot 10^{23}$ атомов (число Авогадро).]

2. Предполагая, что избыточные электроны, внесенные атомами сурьмы (см. задачу 1), полностью возбуждены при комнатной температуре и что их подвижность равна $3600 \text{ см}^2/\text{в} \cdot \text{сек}$, нужно показать, что удельная проводимость этого слитка равна $0,50 \text{ ом}^{-1} \cdot \text{см}^{-1}$, или $50 \text{ ом}^{-1} \cdot \text{м}^{-1}$ в единицах MKS. Показать, что образец сечением 1 мм^2 и длиной 2 см будет иметь сопротивление 400 ом.

3. Предполагая, что образец вместо сурьмы содержит $0,78 \cdot 10^{-6}$ г галлия с атомным весом 69,72, нужно показать, что концентрация атомов галлия составит $3,68 \cdot 10^{-14} \text{ см}^{-3}$. Принимая $\mu_p = 1700 \text{ см}^2/\text{в} \cdot \text{сек}$, показать, что проводимость при этом равна $0,10 \text{ ом}^{-1} \cdot \text{см}^{-1}$.

4. Предполагая, что в слиток германия весом 100 г введены одновременно галлий и сурьма в указанных выше (см. задачи 1 и 3) количествах, показать, почему проводимость равна $0,29 \text{ ом}^{-1} \cdot \text{см}^{-1}$, а не $0,6 \text{ ом}^{-1} \cdot \text{см}^{-1}$. Будет ли полученный материал обладать проводимостью *n*-типа или *p*-типа?

5. Предполагая, что на образец (см. задачу 2) падает равномерный пучок света, общее число поглощенных в образце фотонов равно $1,74 \cdot 10^{18}$ и каждый фотон образует пару электрон—дырка, показать, что сопротивление образца при освещении понизится до 162 ом.

6. Показать, что если к образцу задачи 2 приложить напряжение в 10 в (между концами), то электрон будет дрейфовать со скоростью 18000 см/сек и пройдет через весь образец за $1,11 \cdot 10^{-4}$ сек.

ЛИТЕРАТУРА

1. Torrey H. C., Whitmer C. A., Crystal Rectifiers, New York, 1948 (см. перевод: Торрей и Уитмер, Кристаллические детекторы, т. I и II, М., 1950).
2. Pauling L., The Nature of the Chemical Bond, Cornell Univ. Press, 1939.
3. Hume-Rothery W., The Structure of Metals and Alloy, London, 1936 (revised 1950).

4. Wooldridge D. E., Ahearn A. J., Burton J. A., Phys. Rev., **71**, 913 (1947); Van Heerden P. J., Диссертация, Utrecht, 1945.
5. McKay K. G., Phys. Rev., **77**, 816 (1950).
6. Goucher F. S., Bull. of the 298th Meeting of the Amer. Phys. Soc.
7. Wilson A. H., Proc. Roy. Soc., **133A**, 458 (1931).
8. McKay K. G., Phys. Rev., **74**, 1606 (1948).
9. Scaff J. H., Theuerer H. C., Schumacher E. E., Journ. of Metals, **185**, 383 (1949).
10. Pearson G. L., Bardeen J., Phys. Rev., **75**, 865 (1949).
11. Klick C. C., Maurer R. J., Phys. Rev., **76**, 179 (1949).
12. Seitz F., Phys. Rev., **73**, 549 (1948).
13. Pearson G. L., Phys. Rev., **76**, 179 (1949).
14. Ahearn A. J., Phys. Rev., **73**, 1113 (1948).
15. Wannier G., Phys. Rev., **52**, 191 (1937).
16. Stelgman J., Shockley W., Nix F. C., Phys. Rev., **56**, 13 (1939).
17. Lark-Horovitz K., Bleuler E., Davis R., Tendam D., Phys. Rev., **78**, 1256A (1948).
18. Davis R. E., Johnson V. A., Lark-Horovitz K., Siegel S., Phys. Rev., **74**, 1255A (1948).
19. Johnson V. A., Lark-Horovitz K., Phys. Rev., **76**, 442 (1949).
20. Pearson G. L., Struthers J. D., Theuerer H. C., Phys. Rev., **75**, 344 (1949); **77**, 809 (1950).
- 21*. Иоффе А. Ф., Электронные полупроводники, М. — Л., 1933.
- 22*. Иоффе А. Ф., Изв. АН СССР, серия физич., **15**, 477 (1951).
- 23*. Давыдов Б. И., Успехи физич. наук, **33**, 157 (1947).
- 24*. Волькенштейн Ф. Ф., Электропроводность полупроводников, М. — Л., 1947.
- 25*. Волькенштейн Ф. Ф., Успехи физич. наук, **43**, 11 (1952); Успехи химии, **18**, 228 (1949).
- 26*. Френкель Я. И., Вестник АН СССР, **10**, 61 (1946).
- 27*. Пекар С. И., Исследования по электронной теории кристаллов, М. — Л., 1951.
- 28*. Труды 1-й конференции по твердым выпрямителям и фотоэлементам, ЖТФ, **1**, № 7 (1931).
- 29*. Совещание по полупроводникам и фотоэлементам 1938 года, Изв. АН СССР, серия физич., № 5 — 6 (1938).
- 30*. 6-я конференция по полупроводникам, 1941, Изв. АН СССР, серия физич., **5**, № 4—5 (1941).
- 31*. 7-я конференция по полупроводникам, 1950, Изв. АН СССР, серия физич., **16**, № 1—2 (1952).
- 32*. Mott N. E., Gurney R. W., Electronic Processes in Ionic Crystals, Oxford, 1948 (см. перевод: Мотт Н., Гурни Р., Электронные процессы в ионных кристаллах, ИЛ, 1949).
- 33*. Tetzner S., Semi-conducteurs électroniques et leur complexes dérivés, Paris, 1950 (см. перевод: Полупроводники, теория и применения, под ред. Е. Я. Пумпера, Изд. Сов. радио, М., 1952).
- 34*. Буш Г., Электронная проводимость неметаллов, Успехи физич. наук, **47**, 258 (1952).
- 35*. Henisch H. K., Metal Rectifiers, Oxford, 1949 (см. перевод: Хениш Г., Полупроводниковые выпрямители, ИЛ, 1952).
- 36*. Болтакс Б. И., ЖТФ, **20**, 3 (1950).
- 37*. „Диэлектрики и полупроводники“, Проблемы современной физики, научно-реферативный сборник, серия 2, вып. 10, ИЛ (1950).
- 38*. „Полупроводники“, Проблемы современной физики, научно-реферативный сборник, серия 3, вып. 19, ИЛ (1951).
- 39*. Волькенштейн М. В., Строение молекул, Изд. АН СССР, 1947, —
- 40*. Эванс Р., Введение в кристаллохимию, М. — Л., 1948.

- 41*. „Состояние теории химического строения в органической химии“, Все-союзное совещание 11—14 июня 1951 г. (стенографич. отчет), АН СССР, Отд. хим. наук, М., 1952.
- 42*. Herring K., Bell Syst. Techn. Journ., 28, 401 (1949); Leverenz H. W., North D. O., Phys. Rev., 85, 930 (1952).
- 43*. Hall R. H., Phys. Rev., 87, No 2, 387 (1952).
- 44*. Блум А. И., Горюнова Н. А., ДАН СССР, 75, 367 (1950).
- 45*. Kendall J. T., Proc. Phys. Soc., 63B, 821 (1950).
- 46*. Busch C., Wieland J., Zoller H., Helv. Phys. Acta, 24, 49 (1951).
- 47*. Haynes I. R., Westphal W. C., Phys. Rev., 85, 680 (1952).
- 48*. Dunlap W. C., Phys. Rev., 79, 286 (1950).
- 49*. Hibi T., Matsumura T., Phys. Rev., 81, 884 (1951).
- 50*. Kubo R., Phys. Rev., 86, 929 (1952).
- 51*. Lark-Horovitz K., Semi-conducting Materials, London, 1951.
- 52*. Ohl R. S., Bell Syst. Techn. Journ., 31, 104 (1952).
- 53*. Stech B., Zs. f. Naturforsch., (B), 7a, 175 (1952).

ПРИЛОЖЕНИЕ К ГЛАВЕ 1

ПЕРИОДИЧЕСКАЯ СИСТЕМА ЭЛЕМЕНТОВ МЕНДЕЛЕЕВА (классификация по Юм-Розери)

										H 1								He 2					
										Li 3	Be 4	B 5	C 6	N 7	O 8	F 9	Ne 10						
										No 11	Mg 12	Al 13	Si 14	P 15	S 16	Cl 17	Ar 18						
K 19	Ca 20	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Sc 21</td> <td style="text-align: center;">Ti 22</td> <td style="text-align: center;">V 23</td> <td style="text-align: center;">Cr 24</td> <td style="text-align: center;">Mn 25</td> <td style="text-align: center;">Fe 26</td> <td style="text-align: center;">Co 27</td> <td style="text-align: center;">Ni 28</td> </tr> </table>						Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36
Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28																
Rb 37	Sr 38	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Y 39</td> <td style="text-align: center;">Zr 40</td> <td style="text-align: center;">Nb 41</td> <td style="text-align: center;">Mo 42</td> <td style="text-align: center;">Tc 43</td> <td style="text-align: center;">Ru 44</td> <td style="text-align: center;">Rh 45</td> <td style="text-align: center;">Pd 46</td> </tr> </table>						Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54
Y 39	Zr 40	Nb 41	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46																
Cs 55	Ba 56	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">[57-71]</td> <td style="text-align: center;">Hf 72</td> <td style="text-align: center;">Ta 73</td> <td style="text-align: center;">W 74</td> <td style="text-align: center;">Re 75</td> <td style="text-align: center;">Os 76</td> <td style="text-align: center;">Ir 77</td> <td style="text-align: center;">Pt 78</td> </tr> </table>						[57-71]	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
[57-71]	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78																
Fr 87	Ra 88	Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96														
I	II	III	IV	V	VI	VII	VIII	VIII	VIII	I	II	III	IV	V	VI	VII	VIII						
Класс I										Класс II				Класс III									
Металлы; большинство металлических элементов кристаллизуется в граноцентрированной кубической решетке, либо в объемноцентрированной кубической, либо по закону гексагональной плотнейшей упаковки										Переходные структуры				Валентные кристаллы									

ТРАНЗИСТОР КАК ЭЛЕМЕНТ СХЕМЫ

§ 1. СОСТОЯНИЕ ВОПРОСА О ПОЛУПРОВОДНИКОВЫХ
УСИЛИТЕЛЬНЫХ УСТРОЙСТВАХ ДО ПОЯВЛЕНИЯ ТРАНЗИСТОРА

Введение. В гл. 1 было рассмотрено поведение электронов и дырок в тех случаях, когда их распределение в пространстве и во времени существенно однородно. Однако для понимания механизма действия транзистора, а также и других относящихся к этому вопросу явлений в полупроводниках необходимо рассмотреть те случаи, когда концентрация дырок и электронов изменяется во времени и в пространстве. Изучение этих явлений в последнее время значительно облегчилось благодаря появлению новой экспериментальной техники и методов, возникших в связи с исследованиями транзисторов.

Теперь имеется возможность изолировать и отдельно анализировать ряд отдельных процессов, происходящих при работе транзистора. Эти вопросы будут рассмотрены в гл. 3, а затем в гл. 4 мы перейдем к изложению теории контактного транзистора, или, как его обычно называют, транзистора типа А, изобретенного Бардином и Браттейном¹⁾.

В этой главе будет дан краткий обзор истории исследований полупроводников, в частности связанных с усилительными устройствами, и прослежен ход этих исследований вплоть до изобретения транзистора. Затем будет дано описание самого транзистора с феноменологической точки зрения и рассмотрена его эквивалентная схема. Последующие параграфы настоящей главы посвящены в основном исследованию физической сущности описанной ранее чисто формальной эквивалентной схемы транзистора.

В части I этой книги дана главным образом качественная картина движения электронов и дырок в полупроводниках и сделана попытка изложить необходимые сведения из теории полупроводников, включая явления теплового возбуждения дырок и электронов, их рекомбинацию, а также движение в электрических и магнитных полях. В частях II и III изложена зонная теория специально в применении к интересующим нас явлениям; однако чтобы дать правильное представление об основных процессах, имеющих место при работе транзистора, нет никакой необходимости прибегать к общей теории полу-

¹⁾ Впервые транзистор типа А был описан Бардином и Браттейном в статье „Физические принципы действия транзистора“ [1], за которой последовал ряд других сообщений [2—6]. — *Прим. авт.*

проводников. В некоторых случаях мы приводим лишь качественные результаты и формулы; однако общую количественную теорию транзистора еще не удалось довести до такой же степени совершенства, как лежащую в ее основе теорию дырок и электронов.

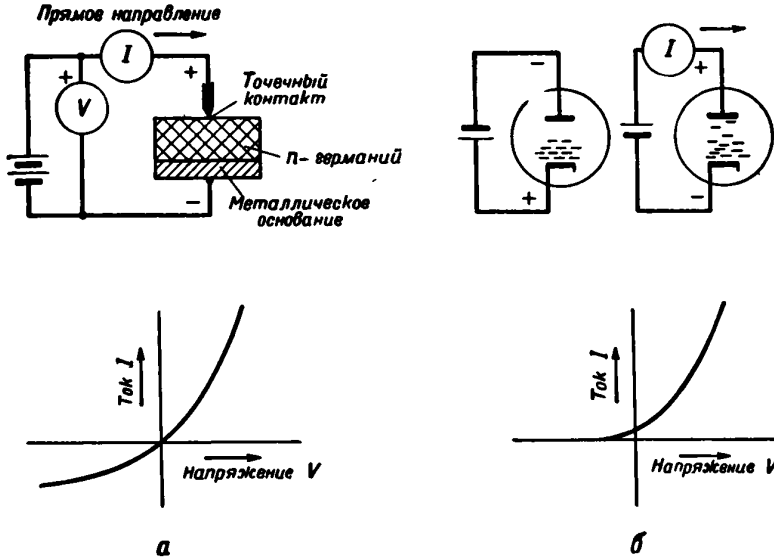
Основная идея о том, что полупроводники могут быть использованы для усиления электрических сигналов, относительно стара. Как это будет показано ниже, усиление электрических сигналов при помощи термистора с отрицательным температурным коэффициентом сопротивления было получено еще задолго до изобретения транзистора. Сходство вольтамперных характеристик вакуумных диодов и кристаллических детекторов (фиг. 16) наталкивало многих исследователей на мысль о том, что на основе кристаллического детектора можно было бы создать усилитель¹⁾, если каким-нибудь образом добавить к нему сетку²⁾. Это, собственно говоря, и осуществлено в настоящее время в транзисторе. Однако, как будет показано ниже, процесс усиления совсем не требует наличия сетки. Другое существенное различие заключается в том, что в случае вакуумного диода сетка служит для регулировки прямого тока. В транзисторе же, напротив, на выходной контакт подается смещение в запирающем направлении, а усиление тока и его регулировка достигаются при помощи входного контакта.

Значительно раньше появления транзисторов было получено и изучено усиление, основанное на использовании отрицательной температурной характеристики сопротивления термистора. Это отрицательное сопротивление термистора обусловлено отрицательным температурным коэффициентом сопротивления полупроводников, например в области их собственной проводимости (см. фиг. 11). По мере возрастания проходящего через термистор тока напряжение на нем вначале растет линейно, в соответствии с законом Ома. Однако при дальнейшем увеличении тока термистор начинает заметно нагреваться, причем его сопротивление быстро падает и вольтамперная характеристика становится нелинейной. При пропускании достаточно больших токов температура термистора может повыситься настолько, что уменьшение его сопротивления приводит к быстрому падению на нем напряжения с увеличением силы тока и, следовательно, к возникновению отрицательного дифференциального сопротивления. Такая характеристика и может быть использована для устройства генератора или четырехполосника, при условии, что частота изменения силы тока достаточно низка и температура термистора успевает за ней

¹⁾ Первые практические результаты по применению точечных контактов металл — полупроводник и полупроводник — полупроводник для усиления и генерации переменных сигналов были получены в период с 1922 по 1927 г. советским физиком О. В. Лосевым [15—20]. — *Прим. ред.*

²⁾ Ван-Гель [21] описал селеновый выпрямитель с управляющей сеткой, введенной в искусственный запирающий слой. О практическом применении такого устройства в качестве генератора и усилителя никаких указаний в литературе не имеется. — *Прим. ред.*

следовать. Это требование определяет верхнюю частотную границу таких устройств, которая зависит от постоянных, определяющих изменение температурного режима во времени при заданных условиях опыта. С термисторами, обладающими весьма малыми размерами и большой теплопроводностью, удавалось получать устойчивые колебания с частотой вплоть до 100 000 гц¹⁾. Однако дополнительным фактором, лимитирующим частоту независимо от размеров термистора,



Фиг. 16. Выпрямители и их вольтамперные характеристики.

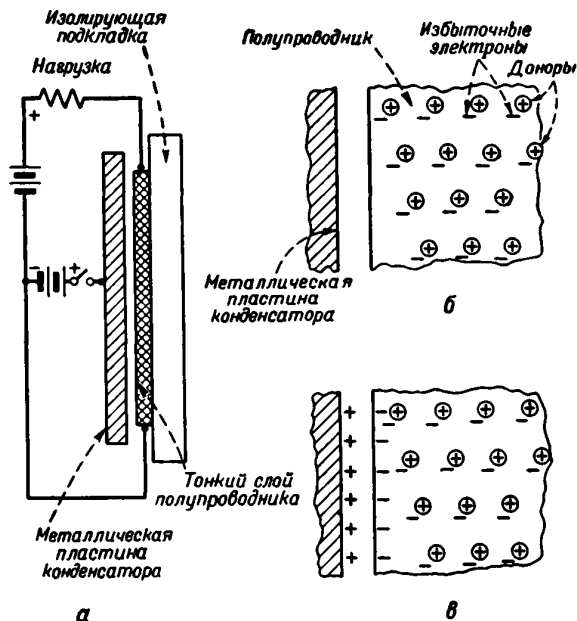
а — кристаллический детектор; б — ламповый диод.

является время, необходимое для того, чтобы электроны и дырки могли прийти в равновесие после изменения температуры. Об этом факторе для большинства полупроводников известно очень мало. Ниже, в § 1 гл. 3, будут рассмотрены некоторые данные для случая германия при комнатной температуре.

Модуляция проводимости поверхностными зарядами. Получение чисто электронного, а не теплового полупроводникового усилителя стало значительно более вероятным с тех пор, как экспериментально было доказано, что создание такого усилителя возможно или, по крайней мере, не находится в противоречии с современной теорией полупроводников и статистической механикой. На фиг. 17

1) Частное сообщение Беккера. — Прим. авт.

схематически показано устройство прибора, при помощи которого проведены соответствующие опыты. Его основу составляет очень тонкий слой полупроводника, нанесенный на изолирующую подкладку. Этот слой полупроводника образует одну из обкладок плоского конденсатора. Второй обкладкой служит расположенная в непосредственной близости к полупроводниковому слою металлическая пластинка.



Фиг. 17. Модуляция проводимости³ поверхностными зарядами.

а — физическая схема; б — незаряженный конденсатор; в — заряженный конденсатор. (Дополнительные электроны, индуцированные на поверхности в случае «в», переносят дополнительный ток.)

Если зарядить конденсатор так, чтобы на металлической пластинке появился положительный заряд (см. фиг. 17), то на полупроводнике образуется дополнительный отрицательный заряд, обусловленный увеличением числа электронов. В германии при комнатной температуре лишь незначительная часть этих электронов связана с донорами (подробнее см. § 4 гл. 10). Следовательно, появившиеся в полупроводнике добавочные электроны могут свободно перемещаться и участвовать в проводимости полупроводника. Меняя напряжение, приложенное к пластине конденсатора, можно модулировать проводимость полупроводника. Так как в случае идеального диэлектрика этот входной сигнал не требует затраты мощности, то в результате получается

усиление по мощности. Такое устройство должно давать весьма значительный эффект¹⁾. Для иллюстрации, а также чтобы показать, как в таких случаях могут быть рассчитаны концентрации электронов и дырок, ниже приведен численный пример.

Предположим, что полупроводниковый слой изготовлен из германия n -типа с сопротивлением $2,4 \text{ ом} \cdot \text{см}^2$). Концентрация электронов может быть вычислена при помощи следующей формулы для проводимости в $\text{ом}^{-1} \cdot \text{см}^{-1}$ (или $\text{кулон/в} \cdot \text{сек} \cdot \text{см}$):

$$\sigma = \frac{1}{\rho} = n e \mu = n \text{ см}^{-3} \cdot 1,6 \cdot 10^{-19} \text{ кулон} \cdot 2600 \text{ см}^2/\text{в} \cdot \text{сек} = \\ = n \cdot 4,2 \cdot 10^{-16} \text{ кулон/в} \cdot \text{сек} \cdot \text{см}. \quad (2.1)$$

Это дает

$$n = \frac{1}{2,4} \cdot 4,2 \cdot 10^{-16} = 1,0 \cdot 10^{15} \text{ электронов/см}^3. \quad (2.2)$$

В слое толщиной $T = 1000 \text{ \AA} = 10^{-5} \text{ см}$ будет, таким образом, находиться $nT = 1,0 \cdot 10^{10} \text{ электронов/см}^2$. Вычислим теперь, сколько электронов индуцируется полем, если к диэлектрику приложено 30000 в/см , или 100 CGSE ед. потенциала. Такое поле создает смещение в $\chi E/4\pi \text{ CGSE}$ ед. заряда на 1 см^2 . Полагая, что для диэлектрика между обкладками конденсатора $\chi = 2$, получим заряд, равный $2 \cdot 100/4\pi = 16 \text{ CGSE}$ ед. заряда на 1 см^2 , что соответствует $16/4,8 \cdot 10^{-10} = 3,3 \cdot 10^{10} \text{ электронов/см}^2$. Это в 3 раза больше числа электронов, имеющих там в нормальном состоянии. Такое поле должно, таким образом, увеличить проводимость слоя германия в 4 раза.

Опыты проводились со слоями из различных полупроводников. При этом степень модуляции была несколько меньше вычисленной выше. Анализ экспериментальных данных показал, что в слоях германия толщиной около 5000 \AA , в которых подвижность составляла всего около $40 \text{ см}^2/\text{в} \cdot \text{сек}$, участвовало фактически в модуляции проводимости около 10% индуцированных зарядов (в данном случае речь идет о дырках, так как слой испаренного германия обладал проводимостью p -типа [7]). Эту пониженную эффективность модуляции

¹⁾ Кристаллический усилитель с высоким входным сопротивлением и большим коэффициентом усиления по току (до 100 000), основанный на принципе модуляции проводимости слоя полупроводника сильным электростатическим полем, подробно описан в статье Штётцера [22]. — *Прим. ред.*

²⁾ В системе MKS этот слой имеет толщину $T = 10^{-7} \text{ м}$, с проводимостью $\sigma = 10^2/2,4 \text{ ом}^{-1} \cdot \text{см}^{-1} = 42 \text{ ом}^{-1} \cdot \text{см}^{-1}$, создаваемой электронами, подвижность которых равна $0,26 \text{ м}^2/\text{в} \cdot \text{сек}$. Число электронов в единице объема определяется по формуле $\sigma = n e \mu$, так что $n = 42/0,26 \cdot 1,6 \cdot 10^{-19} \text{ м}^{-3}$; это составляет $1,0 \cdot 10^{14}$ электронов на 1 м^2 поверхности. Поле в 30000 в/см , или $3 \cdot 10^6 \text{ в/м}$, в среде диэлектрической постоянной, равной 2, создает поверхностный заряд, равный $2 \cdot \epsilon_0 \cdot 3 \cdot 10^6 = 5,3 \cdot 10^{-6} \text{ кулон} \cdot \text{м}^{-2}$, или $3,3 \cdot 10^{14} \text{ электронов} \cdot \text{м}^{-2}$. — *Прим. авт.*

можно объяснить на основе теории поведения носителей тока на поверхности полупроводника. -

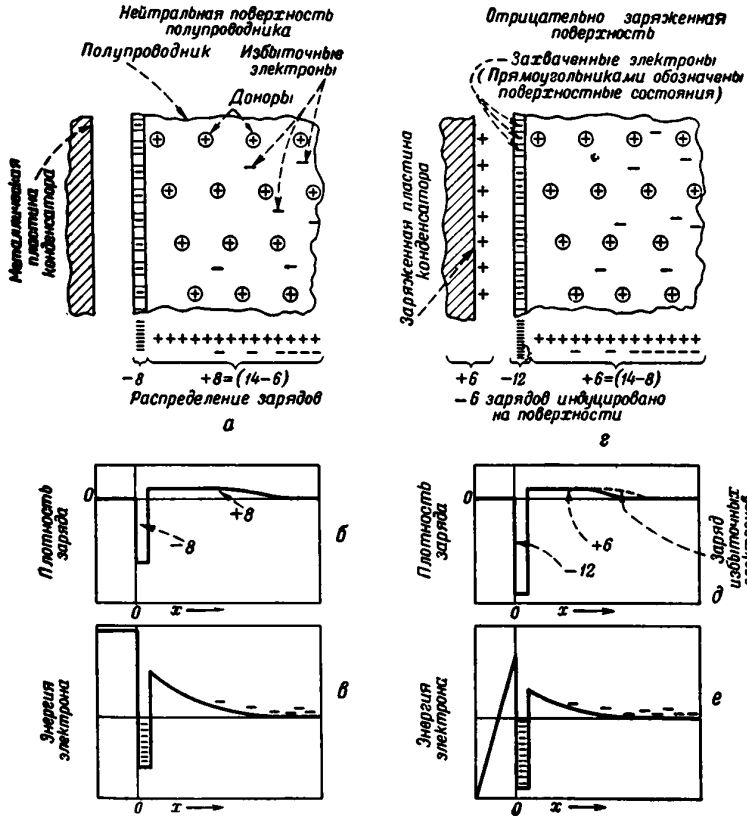
Теория поверхностных состояний по Бардину. Бардин высказал предположение, что не участвующая в модуляции проводимости часть индуцированных зарядов находится в состояниях, локализованных на поверхности полупроводника. Для исследования этого вопроса он воспользовался теорией *поверхностных состояний*, примененной им ранее столь успешно для объяснения ряда других явлений, имеющих место на поверхности полупроводника¹⁾.

Подробное рассмотрение теории поверхностных состояний по Бардину потребовало бы применения математического аппарата зонной теории и статистики Ферми — Дирака, что, как уже указывалось, выходит за рамки части I этой книги. Поэтому мы ограничимся тем, что приведем здесь лишь результаты, полученные при применении этой теории к вопросу о модуляции проводимости поверхностными зарядами, не вдаваясь в их подробное обоснование. Аналогичным образом мы поступим при обсуждении вопроса о механизме работы транзистора с точечными металлическими контактами.

Вывод, к которому пришел Бардин, заключается в том, что электроны могут оказаться прочно связанными в *поверхностных состояниях* на поверхности полупроводника и, таким образом, стать неподвижными. На фиг. 18 приведена схема поверхностных состояний для свободной поверхности полупроводника *n*-типа. На фиг. 18, *a* показано, что в этих поверхностных состояниях находится 8 электронов. Они отталкивают от себя другие электроны, находящиеся в полосе проводимости полупроводника, в результате чего вблизи поверхности образуется слой полупроводника с пониженной проводимостью. В этом слое имеется 8 ионизированных доноров, так что поверхность полупроводника в целом нейтральна. Для наглядности и облегчения счета донорные ионы на рисунке расположены правильно; в образце германия, о котором здесь идет речь, доноры распределены в действительности беспорядочно и находятся друг от друга на расстояниях, в среднем превышающих в 500 раз нормальные расстояния между атомами в решетке. На фиг. 18, *б* представлено распределение плотности заряда в полупроводнике в зависимости от расстояния до его поверхности. При таком распределении плотности заряда изменение электростатического потенциала от точки к точке соответствует ходу кривой энергии электрона, приведенной на фиг. 18, *в*. Влияние взаимного отталкивания электронов, находящихся в поверхностных состояниях, и притяжения их ионизированными донорами характеризуется приблизительно

¹⁾ Представление о поверхностных электронных состояниях теоретически обосновал И. Е. Тамм [23]. Вопрос о работе выхода и проводимости полупроводника при наличии поверхностного заряда исследовал В. Е. Лашкарев [24]. Важную роль поверхностных электронных состояний в явлениях абсорбции и катализа обосновал Ф. Ф. Волькенштейн [33]. — *Прим. ред.*

параболическим ходом кривой энергии в зависимости от расстояния. Поверхностные состояния прочно связывают электроны, при этом энергия электронов даже ниже тех ее значений, которыми электроны обладали бы в полосе проводимости, если бы они находились внутри



Фиг. 18. Диаграмма поверхностных состояний, иллюстрирующая теорию Бардина.

кристалла. Полупроводники, так же как и металлы, характеризуются определенными значениями работы выхода электронов. Энергия, необходимая для удаления электрона из полупроводника в некоторую точку пространства вне его, определяется по высоте кривой энергии, приведенной на фиг. 18, в, и отсчитывается от оси абсцисс.

С этой точки зрения здесь уместно сделать несколько замечаний о соотношении между энергией электрона (см. фиг. 18) и электростатическим потенциалом. Прежде всего, благодаря отрицатель-

ному заряду электрона кривая, характеризующая энергию, и кривая для электростатического потенциала имеют противоположный ход; это соответствует тому, что на диаграмме электростатического потенциала электроны как бы стремятся занимать более высокое положение. Более существенным, особенно в случае движения электрона вблизи поляризующейся поверхности, например поверхности металла, является, однако, то обстоятельство, что электрон создает индуцированный заряд, вследствие чего изменяется электростатический потенциал [8]. На диаграмме энергии это взаимодействие учтено так, что диаграмма характеризует изменение энергии всей системы в зависимости от положения электрона. Однако различие энергий разных состояний определяется не только различием электростатического взаимодействия. Электрон, находящийся в одном из глубоких поверхностных состояний, обладает не только иной электростатической, но и иной кинетической энергией, нежели электрон, находящийся в полосе проводимости. Напомним, например, что в атоме водорода (см. фиг. 2) электрон имеет отрицательную потенциальную энергию, численное значение которой в 2 раза больше его кинетической энергии. В общем, различие энергий разных состояний электронов обусловлено различием их энергий, кинетической и потенциальной, которые практически часто бесполезно пытаться разграничить.

Когда металлическая обкладка конденсатора заряжается положительно, то в цепи возникает ток компенсирующих отрицательных зарядов, при этом в случае, рассмотренном на фиг. 18, *г*, на поверхности полупроводника появляется 6 избыточных электронов. Из этих шести электронов 4 занимают поверхностные состояния, а 2 остаются в качестве избыточных электронов в полосе проводимости. Окончательный вид распределения объемного заряда (см. фиг. 18, *д*) и изменения энергии электрона (см. фиг. 18, *е*) непосредственно следуют из фиг. 18, *г*. Электрическое поле между пластинами конденсатора характеризуется наклоном прямолинейного участка энергетической кривой вне полупроводника (см. фиг. 18, *е*). Рассмотренный на фиг. 18 пример, конечно, чисто качественный, и естественно предположить, что в действительности до 90% зарядов занимают поверхностные состояния в согласии с результатами измерений Пирсона [7].

Исследования, предпринятые в последнее время, позволили выявить различные свойства поверхностных состояний¹⁾.

§ 2. ТРАНЗИСТОР ТИПА А

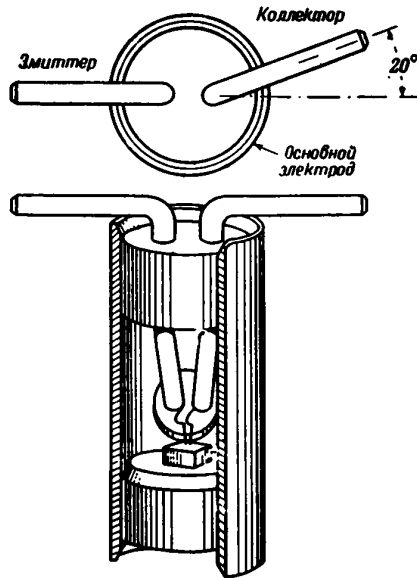
Открытие транзистора. Его устройство. Экспериментальное и теоретическое изучение поверхностных состояний кристаллов особенно усилилось в связи с исследовательской программой, выполнявшейся в

¹⁾ Многие из этих опытов описаны в оригинальной работе Бардина [9]. Данные о контактных потенциалах имеются в работе Браттейна и Шокли [10], а о фотоэлектронной эмиссии — в работе Апкера, Тафта и Дикей [11]. Там же приведена большая библиография. — *Прим. авт.*

лаборатории, руководимой автором. Эта лаборатория, основываясь на теории Бардина, поставила ряд опытов для изучения поверхностных состояний. Во многих случаях искомые эффекты были ниже предела чувствительности применявшихся методов исследования. При изучении этих вопросов Бардин и Браттейн столкнулись с рядом новых явлений, исследование которых и привело к открытию новой области теории и техники полупроводников и к созданию транзистора¹⁾.

Сначала рассмотрим транзистор с феноменологической точки зрения, а затем приведем некоторые теоретические положения, объясняющие принцип его действия.

На фиг. 19 показан общий вид транзистора (в разрезе), а на фиг. 20 — схема включения батарей смещения, их полярность и направления токов, текущих в системе. Каждый точечный контакт транзистора представляет собой детектирующий контакт обычного типа (см. фиг. 16). Входной контакт, или эмиттер, включается в положительном направлении (в прямом направлении, или направлении с малым сопротивлением), а выходной точечный контакт, или коллектор, — в отрицательном направлении (в обратном, или запорном, направлении). Важным открытием Бардина и Браттейна явилось установление того факта, что при прохождении тока через входной точечный контакт (эмиттер) в прямом направлении вокруг него создается некоторая область взаимодействия; если коллектор (выходной точечный контакт) поместить внутри этой области, то ток на выходе будет регулироваться входным током так, что в результате получится усиление по мощности. Эта область взаимодействия не имеет резких границ, и потому усиление по мощности для такого



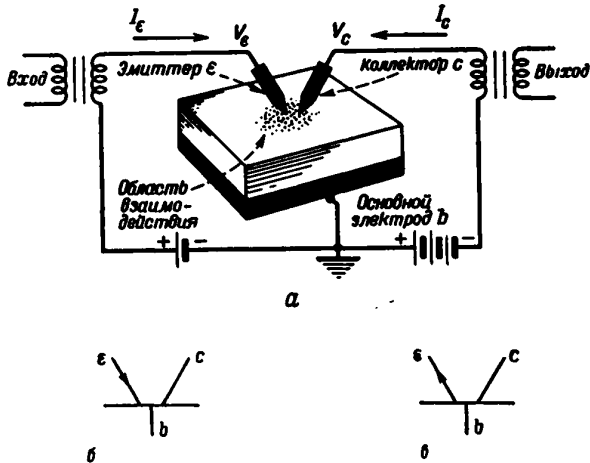
Фиг. 19. Разрез транзистора типа А.

Две проволоочки из фосфористой бронзы диаметром 127 μ заточены, согнуты и припаяны к никелевым электродам, заделанным в пробку из пластмассы, вставленную в цилиндрическую оправу. В эту же оправу вставлена и латунная пробка, к которой припаяна германиевая облатка толщиной 0,5 мм и сечением 0,3 мм². После электрической формовки все это устройство заполняется в вакууме воском через отверстие в цилиндре (о формовке см. § 5 гл. 4).

¹⁾ Дополнительные сведения об истории открытия транзистора, а также обсуждение многих теоретических и экспериментальных вопросов, касающихся транзистора, см. в работе Бардина и Браттейна [1]. — *Прим. авт.*

устройства колеблется в пределах от 100 (20 дб), при расстоянии коллектора от эмиттера в 0,005 см, и до 1 (0 дб), при расстоянии между ними порядка 0,025 см.

Излагаемая в этой главе теория транзистора основана на представлении о том, что эмиттер создает дырки в материале с проводимостью n -типа, который обычно содержит только электроны. Небольшое количество дырок имеется в материале n -типа, но их участие в проводимости ничтожно мало (см. гл. 10).



Фиг. 20. Схема транзистора.

a — транзистор типа А; b — обозначение транзистора из германия n -типа; ϵ — обозначение транзистора из германия p -типа. Показана область взаимодействия, окружающая контакт эмиттера, а также указаны напряжения и направления токов. (Типичные значения для транзистора типа А: $I_{\epsilon} = 0,6$ ма, $V_{\epsilon} = 0,7$ в, $I_c = -2,0$ ма, $V_c = -40$ в.)

Эти дырки двигаются к находящемуся под отрицательным потенциалом коллектору и увеличивают текущий через него ток. В § 1 гл. 3 будут описаны некоторые опыты, подтверждающие эти представления, и приведены подробные данные о концентрации и движении таких дырок. В этой главе мы познакомимся с основными характеристиками транзистора и покажем, каким образом возникает усиление по мощности. Подробная теория транзистора типа А будет дана после разбора более простых случаев.

Статические характеристики ¹⁾. Транзистор представляет собой трехклеммное устройство, работа которого может быть описана

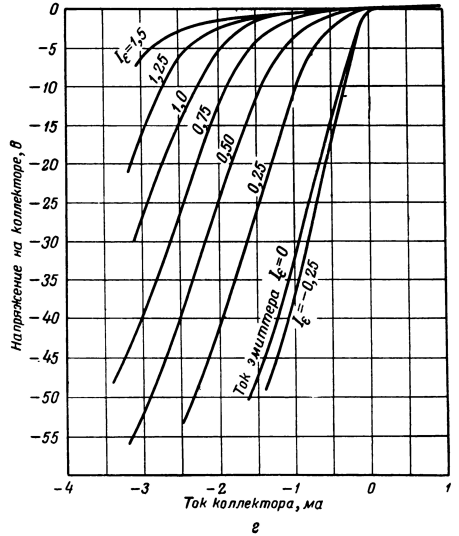
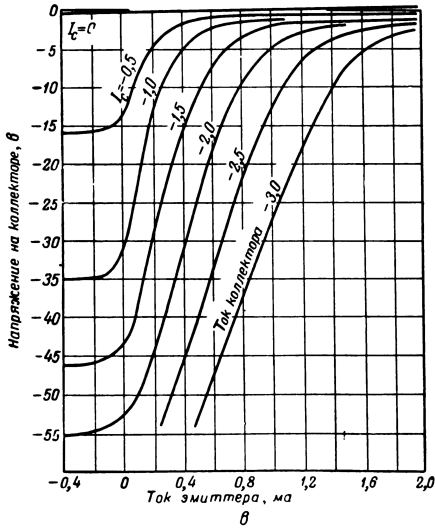
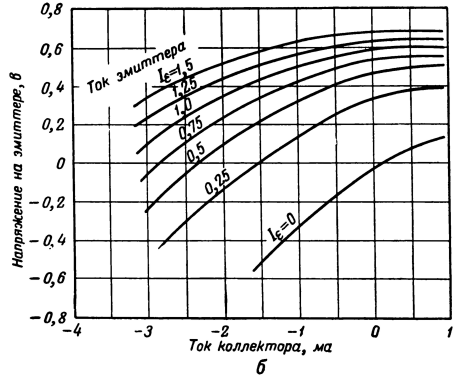
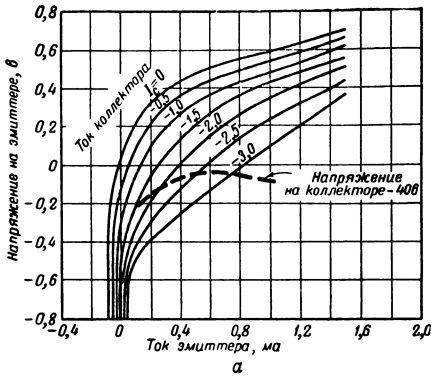
¹⁾ Метод графического анализа характеристик транзистора приведен в статье Хантера [25]. — Прим. ред.

заданием двух приложенных к нему напряжений и соответствующих им двух токов. Для этой цели можно воспользоваться соотношениями, применяющимися обычно при рассмотрении простого трехклемного четырехполюсника. Как указано на фиг. 20, напряжение считается положительным, если оно положительно по отношению к основному электроду, а положительным направлением тока считается направление от металлического точечного контакта к полупроводнику (т. е. когда электроны переходят из полупроводника в металл). Здесь мы имеем дело с более сложными условиями, чем в вакуумном триоде, так как в последнем сеточным током обычно пренебрегают и для получения его статических характеристик достаточно определить анодный ток как функцию анодного и сеточного напряжений. В транзисторе же существенную роль играют как анодные, так и сеточные токи и напряжения. Есть и другое важное различие между транзистором и вакуумным триодом: в случае транзистора данному ряду напряжений могут соответствовать два или три возможных ряда значений токов. Иными словами, токи здесь являются многозначными функциями напряжений. (Для электронной лампы с сеточной эмиссией справедливо обратное: данному значению тока сетки могут соответствовать два значения сеточного напряжения.) Однако в нормальном рабочем режиме транзистора каждому напряжению отвечают лишь 2 значения тока. Поэтому при рассмотрении характеристик транзистора удобнее выражать напряжения как функции токов.

Для эмиттера, основного электрода и коллектора приняты соответственно следующие обозначения: e , b , c .

На фиг. 21 приведены системы статических вольтамперных характеристик транзистора. На фиг. 21, a и b приведены зависимости V_e от I_e и I_c , а на фиг. 21, $в$ и $г$ — зависимости V_c от I_e и I_c . Такой способ графического изображения характеристик транзистора выбран вследствие удобства сопоставления их наклонов с другими параметрами при рассмотрении теории работы транзистора на малых сигналах.

На графике фиг. 21, a пунктирная кривая соответствует напряжению на коллекторе, равному -40 в. Эта кривая продлена за область рабочих характеристик транзистора. Она показывает, что при заданном смещении на коллекторе напряжение на эмиттере с ростом эмиттерного тока сначала возрастает, а затем, при выходе за пределы рабочей области транзистора, убывает. Другими словами, существуют два ряда значений токов эмиттера и коллектора, соответствующих напряжениям $V_c = -40$ в и $V_e = -0,1$ в. Эта многозначность представляет значительные неудобства при построении вольтамперных характеристик, но при переходе к анализу работы самого транзистора дело еще больше усложняется. Работа транзистора на этих двух напряжениях, т. е. от внешних источников с малыми полными сопротивлениями, была бы относительно устойчива в первом режиме и неустойчива во втором. Если переходный импульс вывел бы транзистор



Фиг. 21. Статические характеристики транзистора типа А.
 а — входная характеристика; наклон равен r_{11} ; б — характеристика обратной связи; наклон равен r_{12} ; в — характеристика в прямом направлении; наклон равен r_{21} ; г — характеристика в обратном направлении; наклон равен r_{22} .

а — r_{11} ; б — r_{12} ; в — r_{21} ; г — r_{22}

из устойчивого режима, то токи, протекающие через транзистор, возросли бы выше их предельных значений, в результате чего произошло бы выгорание точечных контактов. Во избежание этого транзисторные схемы рассчитываются так, чтобы исключить неустойчивость усилителей и ограничить токи в опрокидывающих схемах, исключая или контролируя тем самым эту неустойчивость. В § 1 гл. 4 будет показано, что отрицательное сопротивление входного точечного контакта обусловлено усилением тока на пути между входной и выходной цепями транзистора и многозначность характеристик определяется уменьшением сопротивления выпрямляющего эмиттерного контакта по мере возрастания эмиттерного тока.

Статические характеристики имеют важное значение при определении нелинейности транзисторов и для проектирования счетных и переключающих схем с применением кристаллических усилителей. Некоторые задачи, иллюстрирующие приведенные в этом параграфе положения, помещены в конце главы.

Эквивалентная схема ¹⁾. Если записать соотношения, изображенные графически на фиг. 21, в общем виде:

$$V_s = f_1(I_s, I_c) \quad (2.3a)$$

и

$$V_c = f_2(I_s, I_c), \quad (2.3b)$$

то малые переменные напряжения v_s и v_c , обусловленные прохождением малых переменных токов i_s и i_c , равны соответственно ²⁾

$$v_s = \frac{\partial f_1}{\partial I_s} i_s + \frac{\partial f_1}{\partial I_c} i_c = r_{11} i_s + r_{12} i_c \quad (2.4a)$$

и

$$v_c = \frac{\partial f_2}{\partial I_s} i_s + \frac{\partial f_2}{\partial I_c} i_c = r_{21} i_s + r_{22} i_c. \quad (2.4b)$$

Здесь коэффициенты r , представляющие собой сокращенное обозначение частных производных, характеризуют наклон кривых на фиг. 21. Если $r_{12} = r_{21}$ и $r_{11} > r_{12} < r_{22}$, то систему, описываемую приведенными выше уравнениями, можно представить в виде пассивного четырехполюсника сопротивлений. Можно поступить и иначе, введя в рассмотрение отрицательные сопротивления либо генераторы тока или напряжения. На фиг. 22 изображены 3 эквивалентные схемы

¹⁾ Общий обзор теории пассивных четырехполюсников можно найти в книге Гюйсмена [12], а теория активных четырехполюсников изложена в статье Петерсона [13]. — *Прим. авт.* (См. также книгу Р. А. Воронова [26], специально посвященную теории четырехполюсников. — *Прим. ред.*)

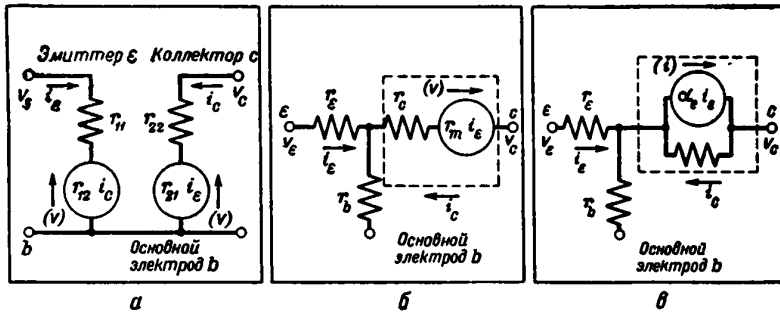
²⁾ В этом параграфе мы будем предполагать, что частота столь низка, что все элементы контура обладают лишь активным сопротивлением. — *Прим. авт.*

транзистора и здесь же указаны соотношения между параметрами этих схем ¹⁾.

Величина

$$\alpha = - \left[\frac{\partial I_c}{\partial I_e} \right]_{V_c = \text{const}} = \frac{r_{21}}{r_{22}} \quad (2.5)$$

есть мера влияния дырочного тока эмиттера на ток коллектора. Физический смысл величины α будет рассмотрен в следующих параграфах. Сопротивление r_{12} , характеризующее влияние коллектора на



$r_e = r_{11} - r_{12}$	$r_{11} = r_e + r_b$	$\alpha = r_{21} / r_{22}$
$r_b = r_{12}$	$r_{12} = r_b$	$\alpha_e = r_m / r_c$
$r_c = r_{22} - r_{21}$	$r_{21} = r_m + r_b$	$\alpha_e = \alpha + (\alpha - 1) \frac{r_b}{r_c}$
$r_m = r_{21} - r_{12}$	$r_{22} = r_b + r_c$	

Фиг. 22. Эквивалентные схемы транзистора и соотношения между параметрами схем.

Генераторы, заключенные внутри пунктирных прямоугольников, эквивалентны, если $r_m = \alpha_e r_c$.

эмиттер, называется *сопротивлением основного электрода*; оно играет роль положительной обратной связи в схеме усилителя с заземленным основным электродом и, как таковое, повышает усиление схемы, но понижает ее устойчивость.

На фиг. 21, б сопротивление r_m в генераторе напряжения $\alpha_e i_e$ есть общее активное сопротивление. Как будет показано в конце

¹⁾ Эквивалентная схема транзистора типа А подробно рассматривается в работах Райдера и Кирхера [27, 28]. — *Прим. ред.*

этого параграфа, усиление по мощности примерно пропорционально r_m^2 .

Значения сопротивлений r_{11} и др. для любой рабочей точки можно определить из статических характеристик, изображенных на фиг. 21. В табл. 3 приведены различные характеристики среднего образца транзистора типа А.

Таблица 3

НЕКОТОРЫЕ ДАННЫЕ ДЛЯ ТРАНЗИСТОРА ТИПА А

Типичные рабочие условия	Средние значения параметров эквивалентной схемы, ом	
Ток эмиттера 0,6 ма	$r_{11} = 530$	Сопротивление эмиттера $r_a = 240$
Напряжение на эмиттере 0,7 в	$r_{12} = 290$	Сопротивление основного электрода $r_b = 290$
Ток коллектора . . . — 2 ма	$r_{21} = 34\ 000$	Сопротивление коллектора $r_c = 19\ 000$
Напряжение на коллекторе — 40 в	$r_{22} = 19\ 000$	Общее сопротивление $r_m = 34\ 000$
	$\alpha_e \approx \alpha \equiv r_{21}/r_{22} = 1,8$	

Максимальные характеристики: при непрерывной работе не достигаются.

Напряжения указаны относительно основного электрода:

Напряжение на коллекторе — 70 в

Рассеиваемая мощность на коллекторе 0,2 вт

Работа транзистора в схеме с заземленным основным электродом: транзистор типа А работает от генератора с внутренним сопротивлением 500 ом на нагрузку в 20 000 ом.

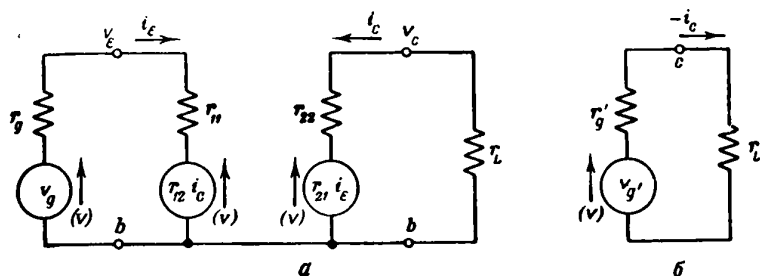
Усиление по мощности в рабочем режиме — около 17 дБ

Выходная мощность — около 5 мвт

Из приведенных в табл. 3 данных следует, что полное выходное сопротивление транзистора составляет величину порядка сотен ом. Значения α показывают, что ток на выходе, например в эквивалентной схеме фиг. 21, в, даже больше входного тока; полное сопротивление нагрузки составляет 20 000 ом. На этом, очевидно, и основана возможность усиления по мощности.

Усиление по мощности связано с наличием в эквивалентных схемах фиг. 22, б и в генераторов тока или напряжения. (Усиление по мощности может иметь место даже в том случае, если α_e будет меньше единицы.) В последующих параграфах мы рассмотрим вопрос о внутреннем механизме работы транзистора и, сопоставив физическую картину его работы с эквивалентной схемой, дадим физическую интерпретацию механизма усиления по мощности. В настоящем параграфе мы разберем лишь формальную эквивалентную схему, не касаясь физической стороны явления. Такой способ изложения вопроса представляет, тем не менее, известный интерес, и отчасти поэтому мы не пользуемся специальной техникой теории цепей. Хотя при

рассмотрении затронутых здесь вопросов было бы удобно использовать результаты теории усилителей с обратной связью, мы, однако, воспользовались этой возможностью лишь в незначительной степени. Это вызвано тем, что, с одной стороны, для читателя, знакомого с этими представлениями, не составит большого труда перевести полученные нами результаты на язык общей теории электрических цепей, а, с другой стороны, было бы нецелесообразно излагать теорию цепей слишком подробно для тех читателей, которые впервые встречаются с основными представлениями теории полупроводников.



Ф и г. 23. Транзистор как усилитель с заземленным основным электродом.

а—работает от генератора сигнала v_g, r_g на нагрузку r_L ; б—эквивалентная схема выхода.

Для дальнейшего изложения схемные характеристики транзистора представляют интерес главным образом как проявление основных физических процессов.

В качестве примера применения эквивалентной схемы, а также с целью иллюстрации особенностей транзистора типа А здесь приведен расчет транзистора как усилителя с заземленным основным электродом. Это, конечно, лишь одна из многочисленных возможных схем с применением транзистора. На фиг. 23 изображена эквивалентная схема усилителя с заземленным основанием. Эмиттер присоединен к входному генератору v_g, r_g , представленному в виде генератора напряжения v_g , включенного последовательно с сопротивлением r_g . На выходе включено нагрузочное сопротивление r_L . Источники, питающие цепи смещения, на схеме не показаны. Генератор имеет некоторую мощность, определяемую как его *полезная мощность*. Это максимальная мощность, развиваемая, как известно из общей теории электрических цепей, генератором при работе на согласованную нагрузку. В этом случае она равна

$$\frac{v_g^2}{4r_g} = \text{полезная мощность входного генератора.} \quad (2.6)$$

В дальнейшем усиление по мощности всегда будет определяться как отношение выходной мощности к мощности входного генератора.

Усиление по мощности транзистора с заземленным основным электродом определяется непосредственно из анализа схемы фиг. 23. Чтобы выразить выходную мощность транзистора через параметры эквивалентного генератора v'_g, r'_g (см. фиг. 23, б), запишем выражения для напряжений на электродах e и c . Для электрода e получаем

$$v_e = v_g - r_g i_e \quad (\text{от генератора}), \quad (2.7a)$$

$$v_e = r_{11} i_e + r_{12} i_c \quad (\text{от транзистора}). \quad (2.7б)$$

Аналогично для электрода c имеем

$$v_c = -r_L i_c \quad (\text{от нагрузки}), \quad (2.8a)$$

$$v_c = r_{21} i_e + r_{22} i_c \quad (\text{от транзистора}). \quad (2.8б)$$

Решим теперь последнее уравнение, выражая i_e через v_g . Прделав это, найдем, что уравнение для v_c станет очень похожим на уравнение для генератора, работающего в цепи эмиттера, так что транзистор можно рассматривать как устройство, преобразующее генератор v_g, r_g в новый генератор v'_g, r'_g . Решение уравнения для v_e относительно i_e дает

$$i_e = \frac{v_g - r_{12} i_c}{r_{11} + r_g}. \quad (2.9)$$

Подставляя (2.9) в (2.8б), получим

$$v_c = \left(r_{22} - \frac{r_{12} r_{21}}{r_{11} + r_g} \right) i_c + \frac{r_{21}}{r_{11} + r_g} v_g = v'_g + r'_g i_c. \quad (2.10)$$

Последнее выражение соответствует замене транзистора генератором v'_g, r'_g ; впереди $r'_g i_c$ стоит знак, обратный знаку перед $r_g i_e$ в уравнении (2.7а), что связано с изменением направления тока (i_e — ток, выходящий из генератора v_g, r_g ; i_c — ток, входящий в r'_g, i'_c). Отсюда мы делаем заключение, что уравнения

$$v'_g = \frac{r_{21} v_g}{r_{11} + r_g}, \quad (2.11a)$$

$$r'_g = \frac{r_{22} - r_{12} r_{21}}{r_{11} + r_g} \quad (2.11б)$$

описывают работу транзистора на нагрузку.

Мощность, выделяемая генератором в нагрузке, определяется выходным током $i_c = v'_g / (r'_g + r_L)$, проходящим через r_L :

$$r_L i_c^2 = r_L \left(\frac{v'_g}{r'_g + r_L} \right)^2. \quad (2.12)$$

Отношение этой мощности к полезной мощности генератора v_g, r_g при работе на согласованную нагрузку и есть *усиление в рабочем режиме*.

Усиление в рабочем режиме =

$$\begin{aligned} &= G(r_g, r_L) = \frac{r_L v_g'^2}{(r_g' + r_L)^2} \frac{4r_g}{v_g^2} = \\ &= \frac{4r_L r_g}{(r_g' + r_L)^2} \left[\frac{r_{21}}{r_{11} + r_g} \right]^2 = 4r_g r_L \left[\frac{r_{21}}{\Delta} \right]^2, \end{aligned} \quad (2.13)$$

где

$$\Delta = (r_{11} + r_g)(r_{22} + r_L) - r_{12}r_{21} \quad (2.14)$$

есть *детерминант цепи*. Уже само обозначение $G(r_g, r_L)$ указывает на зависимость усиления по мощности в рабочем режиме от характеристик внешних входных и выходных цепей, а также от параметров самого транзистора. Задача нахождения максимального значения $G(r_g, r_L)$ посредством соответствующего подбора r_g и r_L будет нами рассмотрена ниже. Детерминант цепи есть показатель ее устойчивости. При $\Delta < 0$ схема будет самопроизвольно генерировать, даже если $v_g = 0$. При $\Delta > 0$ работа схемы может быть стабильной, однако величина Δ , определяемая уравнением (2.14), есть низкочастотный детерминант цепи, а поэтому стабильность цепи при высоких частотах будет определяться его свойствами при высоких частотах. К этому вопросу мы еще вернемся.

Рассмотрим теперь параметры выходного генератора, исходя из данных табл. 3. Отношение напряжений генераторов, при условии, что $r_g = 500$ ом, приближенно равно

$$\frac{v_g'}{v_g} = \frac{r_{21}}{r_{11} + r_g} = \frac{34\,000}{530 + r_g} \approx 30. \quad (2.15)$$

Отношение v_g'/v_g есть коэффициент усиления по напряжению при работе на нагрузку с бесконечно большим полным сопротивлением и соответствует крутизне μ электронной лампы. Значение r_g' зависит от величины обратной связи, определяемой сопротивлением r_{12} , и равно

$$\begin{aligned} r_g' &= r_{22} \left[1 - \frac{\alpha r_{12}}{r_{11} + r_g} \right] = r_{22} \left[1 - \frac{1,8 \cdot 290}{530 + r_g} \right] \approx \\ &\approx r_{22} [1 - 0,45] = 10\,500 \text{ ом}. \end{aligned} \quad (2.16a)$$

Последние два члена соответствуют $r_g = 500$ ом. Очевидно, что положительная обратная связь благодаря появлению члена, содержащего αr_{12} , приведет к уменьшению полного сопротивления выходного генератора, а следовательно, к возрастанию полезной мощности генератора v_g' , r_g' . В принципе полное сопротивление генератора может быть уменьшено до нуля или сделано отрицательным, если в цепь основного электрода включить внешнее сопротивление обратной связи.

Такое уменьшение величины полного сопротивления эквивалентно увеличению всех значений сопротивлений r эквивалентной схемы на величину этого добавочного сопротивления, что особенно сказывается на значениях r_{11} и r_{12} . Когда r'_g проходит через нуль и становится отрицательным, полное сопротивление между электродами c и b , равное $r'_g r_L / (r'_g + r_L)$, также становится отрицательным и при этом могут возникнуть колебания, связанные с паразитными емкостями и индуктивностями. Это есть пример неустойчивости, возникающей еще до того, как величина Δ (для низкой частоты) обращается в нуль; Δ не обращается в нуль до тех пор, пока $r'_g + r_L = 0$. Задача нахождения оптимального значения сопротивления обратной связи сходна с другими задачами расчета усилителей с обратной связью.

Воспользовавшись вычисленным выше [см. (2.16)] значением для r'_g , равным 10 500 ом , а также приведенными в табл. 3 данными, характеризующими работу транзистора в схеме с заземленным основным электродом, мы на основании (2.13) получаем

$$\text{Усиление в рабочем режиме} = 47, \text{ или } 16,7 \text{ дб.} \quad (2.16\text{б})$$

Часто пользуются еще двумя другими определениями: *полезное усиление* и *максимальное полезное усиление*. Величину полезного усиления можно определить, выбирая значение r_L так, чтобы $G(r_g, r_L)$ было максимальным. Очевидно, что это имеет место при $r_L = r'_g$ и дает

$$G(r_g, r'_g) = \frac{r_g}{r'_g} \left[\frac{r_{21}}{r_{11} + r_g} \right]^2 = \text{полезное усиление.} \quad (2.17)$$

Для вышеприведенных параметров схемы

$$G(r_g, r'_g) = \frac{500}{10\,500} \left[\frac{34\,000}{1\,030} \right]^2 = 52, \text{ или } 17,2 \text{ дб.} \quad (2.18)$$

Для указанного в табл. 3 значения нагрузочного сопротивления 20 000 ом (которое обеспечивает значительно большую устойчивость работы схемы, нежели нагрузка в 10 500 ом , при которой усиление в рабочем режиме имеет максимальное значение) рабочее усиление уменьшается в число раз, равное отношению выражений (2.13) и (2.18):

$$\frac{4r_L r'_g}{(r'_g + r_L)^2} = \frac{4 \cdot 20\,000 \cdot 10\,500}{(30\,500)^2} = 0,9. \quad (2.19)$$

Уменьшение коэффициента усиления, как видно, составляет всего 0,5 дб .

Полезное усиление является еще и функцией параметров внешней цепи из-за входящей в формулу (2.17) величины r_g . Максимальное полезное усиление получается при согласовании транзистора и нагрузки. При этом согласованные с транзистором сопротивления r_g

и r_L имеют следующие значения:

$$r_g = \left[\frac{(r_{11}r_{22} - r_{12}r_{21})r_{11}}{r_{22}} \right]^{1/2}, \quad (2.20a)$$

$$r_L = \left[\frac{(r_{11}r_{22} - r_{12}r_{21})r_{22}}{r_{11}} \right]^{1/2}. \quad (2.20б)$$

Если положительная обратная связь слишком велика, то приведенные выше соотношения становятся мнимыми. Критическое условие есть

$$r_{11}r_{22} - r_{12}r_{21} = r_{22}(r_{11} - \alpha r_{12}) = 0. \quad (2.21)$$

Выражая максимальное полезное усиление через r_g и r_L [см. (2.20a) и (2.20б)] — параметры, соответствующие согласованию генератора и нагрузки, имеем

$$\text{Максимальное полезное усиление} = \frac{r_{21}^2}{(r_{11} + r_g)(r_{22} + r_L)}. \quad (2.22)$$

Подставляя сюда численные значения из табл. 3

$$r_g = 76 \text{ ом}, \quad r_L = 2750 \text{ ом}, \quad (2.23)$$

получаем, что максимальное полезное усиление равно 88, или 19 дб. Получившееся по сравнению с коэффициентом усиления, вычисленным для случая, когда $r_g = 500 \text{ ом}$ и $r_L = 20000 \text{ ом}$, дополнительное усиление в 2 дб связано со столь сильным уменьшением внешних сопротивлений, что запас стабильности схемы снижается до опасного уровня. Лучше поэтому в известной мере пожертвовать коэффициентом усиления, но иметь большую стабильность работы схемы.

Для упрощения вычисления максимального полезного усиления в следующих параграфах пренебрежем членом обратной связи r_{12} . В этом случае величины r_g и r_L для максимального полезного усиления соответственно равны r_{11} и r_{22} , а формула для коэффициента усиления принимает следующий вид:

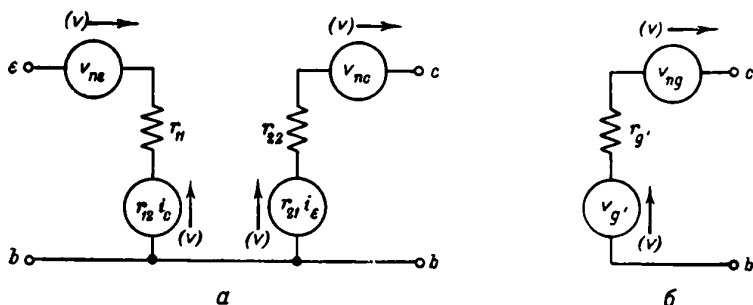
$$\text{Максимальное полезное усиление} = \frac{r_{21}^2}{4r_{11}r_{22}} = \frac{\alpha^2 r_{22}}{4r_{11}}, \quad (2.24)$$

так как $r_{21} = \alpha r_{22}$. Из этой формулы видно, что как большое усиление по току, выражаемое коэффициентом α , так и большое значение отношения полных сопротивлений коллектора и эмиттера способствуют увеличению коэффициента усиления транзистора.

В заключение этих замечаний по поводу эквивалентной схемы транзистора следует указать, что рассматривался лишь транзистор, работающий в схеме с „заземленным основным электродом“. В этом случае режим работы транзистора в общем аналогичен режиму работы электронной лампы с заземленной сеткой; при $\alpha = 1$ теории этих электронных устройств становятся, в сущности, весьма сходными, по

крайней мере для низких частот¹⁾. Помимо этой схемы включения транзистора, он может работать еще в схеме с заземленным эмиттером, что соответствует заземленному катоду, в схеме с заземленным коллектором, что соответствует заземленному аноду, и, наконец, в схеме катодного повторителя. Эквивалентную схему транзистора можно, конечно, с равным правом применить к любому из этих случаев. Более подробный разбор эквивалентной схемы транзистора читатель может найти в работе Райдера [6].

Шумы. Частотный спектр шумов, возникающих в транзисторе типа А, соответствует спектру контактных, или токовых, шумов.



Фиг. 24. Шумы в транзисторах.

a — генераторы шумов в эквивалентной схеме; *б* — эквивалентная схема выхода.

Мощность шумов в полосе шириной 1 гц меняется обратно пропорционально частоте, так что мощность шумов для полосы частот, соответствующей октаве, для всех октав остается одной и той же, по крайней мере в той области частот, для которой параметры эквивалентной схемы не зависят от частоты.

Подробный анализ теории шумов выходит за рамки этой книги²⁾. Здесь укажем только, как надо видоизменить эквивалентную схему транзистора, чтобы учесть этот эффект. Для этого в эквивалентную схему (фиг. 24, *a*) необходимо ввести два генератора э. д. с. шума с напряжениями v_{ne} и v_{nc} соответственно в цепь эмиттера и цепь коллектора. Действующие значения э. д. с. шума этих генераторов в полосе шириной 1 гц и при частоте 1000 гц равны:

$$v_{ne} = 1 \text{ мкв}, \quad (2.25a)$$

$$v_{nc} = 100 \text{ мкв}. \quad (2.25б)$$

¹⁾ Новые методы расчета цепей, содержащих транзисторы, основанные на приближительной аналогии с ламповыми схемами, описаны в работах Уоллеса [29], Уоллеса и Райзбека [30] и Райзбека [31]. — *Прим. ред.*

²⁾ Теория шумов в кристаллических детекторах подробно рассмотрена в книге Торрея и Уитмера [14]. — *Прим. авт.*

Так как мощность шумов меняется пропорционально $1/f$, то для любой другой частоты f значения v_{ns} и v_{nc} будут в $(1000/f)^{1/2}$ раз больше или меньше (если $f > 1000$) их значений при 1000 гц.

В схеме усиления с заземленным основным электродом, изображенной на фиг. 23, v_{ns} будет действовать так же, как и v_g , а v_{nc} просто аддитивно складывается с v'_g . В этом случае эквивалентная схема транзистора будет иметь вид, изображенный на фиг. 24, б, где значение v_{ng} определяется уравнением

$$v_{ng} = \frac{r_{21}v_{ns}}{r_{11} + r_g} + v_{nc} = 33v_{ns} + v_{nc}. \quad (2.26)$$

Здесь мы воспользовались параметрами схемы, указанными в табл. 3, а r_g приняли равным 500 ом. Имеются данные, что генераторы шумов v_{ns} и v_{nc} в некоторой степени взаимно связаны друг с другом («коррелированы»); это говорит о том, что в обоих генераторах шумы возникают отчасти из общего источника. В нашем примере доли v_{nc} и v_{ns} в v_{ng} составляют соответственно 100 и 33 мкв. Если бы они были в действительности идеально конструктивно коррелированы, то они дали бы в сумме 133 мкв вместо 67 мкв, соответствующих случаю идеальной отрицательной корреляции. Для v_{ng} это составило бы разницу в 6 дб. Коэффициент корреляции C двух источников э. д. с. шумов будет в нашем примере определяться уравнением

$$v_{ng}^2 = (33)^2 v_{ns}^2 + v_{nc}^2 + 2 \cdot 33C \sqrt{v_{ns}^2 v_{nc}^2}. \quad (2.27)$$

Наблюдающиеся значения коэффициентов корреляции для транзистора типа А лежат в пределах от $-0,8$ до $+0,4$.

Если пренебречь корреляционными эффектами, то средний квадрат э. д. с. шума в интервале частот от f_1 до f_2 будет равен¹⁾:

$$v_{ng}^2(f_1, f_2) = (100v_{ns}^2 + v_{nc}^2) 1000 \ln \left(\frac{f_2}{f_1} \right). \quad (2.28)$$

В интервале частот от 1000 до 10000 гц эта формула дает

$$v_{ng}^2(10) = (0,25 + 2,3) \cdot 10^{-5} \delta^2 = 2,5 \cdot 10^{-5} \delta^2, \quad (2.29)$$

что соответствует эффективному значению э. д. с. шума в 5 мв. Цифра 10 в выражении $v_{ng}^2(10)$ указывает на то, что в данном случае полоса частот имеет ширину в одну декаду. Это соотношение

¹⁾ Входящий в выражение для v_{ng}^2 член v_{nc} может быть вычислен путем суммирования значений v_{nc} для каждого элемента полосы частот df , что дает

$$\int_{f_1}^{f_2} v_{nc}^2 \left(\frac{1000}{f} \right) df = v_{nc}^2 1000 \ln \left(\frac{f_2}{f_1} \right).$$

— Прим. авт.

является характеристикой частотного спектра контактных шумов и указывает на то, что уровень шумов зависит в этом случае только от отношения f_2/f_1 , а не от значения каждой частоты в отдельности.

Сравним эту величину со значением v'_g для того случая, когда неискаженная мощность в 5 мвт выделяется в нагрузочном сопротивлении $r_L = 20\,000$ ом:

$$5 \cdot 10^{-3} = \frac{v'_g{}^2 r_L}{(r_L + r'_g)^2} = \frac{v'_g{}^2 20\,000}{(30\,500)^2},$$

откуда

$$v'_g{}^2 = 230 \text{ в}^2, \quad (2.30)$$

что соответствует эффективному напряжению в 15 в, или около $1/8$ напряжения смещения на коллекторе. Отношение $(v'_g)^2$ к квадрату э. д. с. шума равно $9,2 \cdot 10^6$, или 70 дб. Эта величина является наиболее удобной характеристикой добротности усилителей; мы назовем ее *динамическим коэффициентом усиления для полосы частот шириной в одну декаду*. Ее можно определить так же как *отношение сигнала к шуму при максимальной выходной мощности для полосы частот шириной в одну декаду*. В нашем случае сигналы на выходе с действующим значением напряжения, превышающим 15 в, будут уже искажаться, а выходные сигналы с действующими значениями меньше 5 в будут сравнимы с уровнем шумов.

Интенсивность полезных выходных сигналов будет поэтому лежать в области шириной порядка 70 дб; точное значение ширины этой области определяется допустимыми пределами величины отношения сигнал/шум и искажений.

Полезное усиление (динамический коэффициент усиления) в 70 дб соответствует полосе частот шириной в одну декаду. Для более широкой полосы частот, естественно, и мощность шумов будет больше, а динамический коэффициент усиления соответственно уменьшается.

Например, полоса частот в пределах от 100 до 10 000 гц охватывает две декады, чему соответствует мощность шумов на 3 дб больше, чем в рассмотренном нами выше примере, так что динамический коэффициент усиления равен теперь уже не 70, а 67 дб.

Динамический коэффициент усиления для любой полосы частот может быть вычислен, если из динамического коэффициента усиления для полосы частот шириной в одну декаду вычесть величину, равную десятикратному значению логарифма при основании 10 от числа декад в полосе.

В качестве еще одного примера рассмотрим тот случай, когда полоса частот простирается от 5 до 6 мгц. Ширина такой полосы равна $\lg 6/5 = \lg 1,2 = 0,080$ декады. Динамический коэффициент усиления в этом случае будет $70 - 10 \lg 0,080 = 70 - 10(-1,1) = 81$ дб.

Динамический коэффициент усиления, очевидно, тесно связан с допустимым значением затухания между каскадами усиления. Если затухание достигает предельного значения динамического коэффициента усиления, то либо будет перегружен выход одного из усилителей, либо первый каскад последующего усилителя будет на выходе иметь сигнал, меньший чем уровень шума; при этом предполагается, что задачи согласования полных сопротивлений решены. Для проектирования усилительного устройства необходимо, кроме того, произвести согласование полных сопротивлений, рассчитать величину допустимых искажений, отношения сигнал/шум, а также число последовательно включенных усилителей.

При рассмотрении шумов часто пользуются и другой количественной характеристикой их свойств, которая называется *фактором шумов*. Можно определить эту величину, воспользовавшись для этой цели эквивалентной усилительной схемой, подобной, например, изображенной на фиг. 23.

Если единственным источником шумов являются тепловые шумы (или, как их иногда называют, шумы Джонсона), генерируемые во внешнем сопротивлении генератора r_g , v_{nT} (индекс T указывает, что здесь идет речь о тепловых шумах), то генерируемый тепловой шум может быть представлен и рассчитан введением на выходе кажущегося шумового генератора v'_{nT} . Отношение среднего квадрата э. д. с. шума $(v_{ng})^2$, действительно наблюдающегося на выходе, к среднему квадрату э. д. с. шумового генератора $(v'_{nT})^2$ и называется *фактором шумов*. Величина фактора шумов зависит от параметров отдельных элементов генератора; минимальное значение фактора шумов достигается лишь при соответствующем выборе этих параметров. Здесь мы не будем вдаваться в детали теории, позволяющей произвести такой выбор, а рассмотрим лишь общее выражение для фактора шумов и вычислим с его помощью фактор шумов для ранее рассмотренной эквивалентной схемы транзистора типа А, в которой $r_g = 500$ ом. Формула для э. д. с. теплового шума при комнатной температуре ($kT/e \approx 0,025$ в) имеет вид

$$\begin{aligned} v_{nT}^2 &= 4r_g kT \Delta f = r_g \left(\frac{4kT}{e} \right) (e \Delta f) = r_g (0,1 \text{ в}) (1,6 \cdot 10^{-19} \Delta f \text{ а}) = \\ &= 1,6 \cdot 10^{-20} r_g \text{ ом } \Delta f \text{ сек}^{-1} \text{ в}^2. \end{aligned} \quad (2.31)$$

Из (2.31), в сочетании с выражениями для усиления по напряжению и э. д. с. шумов выходного генератора (вследствие наличия v_{ne} и v_{nc}), получаем выражения для полной э. д. с. выходного шумового генератора

$$\begin{aligned} v_{ng}^2(f_1, f_2) &= \left[\frac{r_{21}}{r_{11} + r_g} \right]^2 4r_g kT (f_2 - f_1) + \\ &+ \left\{ \left[\frac{r_{21}}{r_{11} + r_g} \right]^2 v_{ne}^2 \oplus v_{nc}^2 \right\} 1000 \ln \left(\frac{f_2}{f_1} \right) \end{aligned} \quad (2.32)$$

и фактора шумов

$$F = 1 + \frac{1}{4r_g kT(f_2 - f_1)} \left\{ v_{ns}^2 \oplus \left[\frac{r_{11} + r_g}{r_{21}} \right]^2 v_{nc}^2 \right\} 1000 \ln \left(\frac{f_2}{f_1} \right). \quad (2.33)$$

Знак \oplus указывает здесь, что при комбинации отдельных членов принята во внимание корреляция этих членов. Прежде чем прямо подставить численные значения в эту формулу, рассмотрим ранее разобранный случай, когда $f_1 = 1000$ и $f_2 = 10000$ гц , воспользовавшись значением $r_g = 500$ ом .

Для этой полосы частот $v_{nT}^2 = 1,6 \cdot 10^{-20} \cdot 500 \cdot 9000 \text{ в}^2$ и, следовательно, в 33^2 раза меньше величины v_{ng}^2 , равной $8 \cdot 10^{-11} \text{ в}^2$. Даваемое транзистором значение э. д. с. шума $2,5 \cdot 10^{-5} \text{ в}^2$ больше этого числа в $3 \cdot 10^{-5}$ раза, или на 55 дб .

Фактор шумов транзистора, равный, как мы только что видели, 55 дб для полосы частот от 1000 до 10000 гц , во много раз превышает значение фактора шума хорошей электронной лампы, для которой эта величина может быть доведена почти до 0 дб . Если принять во внимание частотную зависимость, благодаря которой фактор шумов транзистора при полосе шириной 1 мгц снижается до 30 дб , то это сравнение с фактором шумов электронной лампы при видео-частотах становится более благоприятным, в особенности если внести в устройство транзистора некоторые улучшения.

Фактор шумов зависит как от вида примененной транзисторной схемы в целом, так и от всех ее параметров. Более подробные данные по этому вопросу читатель может найти в специальной литературе¹⁾.

Фактор шумов служит, очевидно, показателем того, в какой степени транзистор не соответствует требованиям, предъявляемым к бесшумовым усилителям. Очевидно, что понижение мощности шумов (добавляемых транзистором) больше чем на 3 дб , по сравнению с мощностью шумов генератора, не даст заметной пользы. Так как шумовой генератор в эквивалентной схеме обладает значительно большей мощностью, чем тепловой шум соответствующих активных составляющих эквивалентной схемы (а также гораздо большей мощностью, чем дробовой шум), то, повидимому, нет достаточно веских физических оснований, которые указывали бы на невозможность сильного понижения мощности шумов в транзисторе.

ЗАДАЧИ

1. Рассмотрев схему транзистора с заземленным эмиттером, когда входной сигнал подается на основной электрод от входного генератора r_g , v_g , а на выходе включено нагрузочное сопротивление r_L , показать, что усилитель будет иметь следующие характеристики:

¹⁾ См. статью Райдера и Кирхера [6], в которой приведены формулы для ряда случаев. — *Прим. авт.*

Детерминант цепи

$$\Delta = (r_g + r_b + r_e)(r_L + r_e + r_c - r_m) + r_e(r_m - r_e)$$

($\Delta > 0$ для устойчивости работы).

Полное входное сопротивление

$$r_{вх.} = r_b + r_e + \frac{r_e(r_m - r_e)}{r_L + r_e + r_c - r_m}.$$

Полное выходное сопротивление

$$r_{вых.} = r_c + r_e - r_m + \frac{r_e(r_m - r_e)}{r_g + r_b + r_e}.$$

Усиление в рабочем режиме (в прямом направлении)

$$G_F = 4r_g r_L \left[\frac{(r_m - r_e)}{\Delta} \right]^2.$$

Усиление в рабочем режиме (в обратном направлении)

$$G_R = 4r_g r_L \left(\frac{r_e}{\Delta} \right)^2.$$

Показать, что для транзистора, параметры которого приведены в табл. 3 ($r_g = 500$ ом и $r_L = 20\,000$ ом), $r_{вх.} = 2\,100$ ом, $r_{вых.} = -6\,900$ ом, $G_F = 24$ дб, $G_R = -19$ дб.

2. Показать, что при заземленном коллекторе в усилителе (см. задачу 1) уравнения, определяющие его свойства, имеют следующий вид:

Детерминант цепи

$$\Delta = (r_g + r_b + r_c)(r_L + r_e + r_c - r_m) + r_c(r_m - r_c)$$

($\Delta > 0$ для устойчивости работы).

Полное входное сопротивление

$$r_{вх.} = r_b + r_c + \frac{r_c(r_m - r_c)}{r_L + r_e + r_c - r_m}.$$

Полное выходное сопротивление

$$r_{вых.} = r_e + r_c - r_m + \frac{r_c(r_m - r_c)}{r_g + r_b + r_c}.$$

Усиление в рабочем режиме (в прямом направлении)

$$G_F = 4r_g r_L \left(\frac{r_e}{\Delta} \right)^2.$$

Усиление в рабочем режиме (в обратном направлении)

$$G_R = 4r_g r_L \left[\frac{(r_m - r_c)}{\Delta} \right]^2 = (1 - \alpha)^2 G_F.$$

Показать, что для транзистора, параметры которого указаны в табл. 3, при $r_g = 20\,000$ ом и $r_L = 10\,000$ ом, $r_{вх.} = -41\,000$ ом, $r_{вых.} = -7\,500$ ом, $G_F = 15$ дб, $G_R = 13$ дб.

Примечание. Помещенные ниже задачи должны дать читателю некоторое представление о значениях токов и напряжений в схемах с транзисторами, а также познакомить его с нелинейными свойствами транзистора, типичные характеристики которого приведены на фиг. 21.

3. Построить график зависимости напряжения эмиттера от эмиттерного тока для $V_c = -10, -20, -30$ в в схеме с заземленным основным электродом. (Указание. Определить по графикам фиг. 21 для заданных значений V_c величины I_e и I_c и отложить их на дополнительном графике в зависимости от r_{11} .) Показать, что при $I_e \approx +1,0$ ма входное сопротивление равно примерно 80 ом.

4. Построить график зависимости V_e от I_e , полагая, что на V_c подано смещение (-60 в) от батареи через сопротивление в 20 000 ом. (Указание. Провести нагрузочную характеристику на дополнительном графике для r_{22} и затем проделать то же, что и в задаче 3.) Показать, что для I_e в пределах от 0,5 до 1,5 ма полное входное сопротивление равно 300 ом. (Эта задача показывает, что увеличение сопротивления в цепи коллектора приводит к увеличению полного входного сопротивления эмиттера.)

5. Цель задачи показать, как включение сопротивления между основным электродом и землей может привести к изменению знака полного входного сопротивления с положительного на отрицательный. Для этого нужно рассмотреть кривую $V_c = -20$ в задачи 3. Показать, что, когда I_e возрастает от 0,5 до 1,0 ма, I_c меняет свою величину от $-1,8$ до $-2,7$ ма, так что значение $I_b = -(I_e + I_c)$ изменяется от $+1,3$ до $+1,7$ ма. Предположим, что между основным электродом и землей включено сопротивление и V'_b есть разность потенциалов между основным электродом и землей, а $V'_e = V_e + V'_b$ и $V'_c = V_c + V'_b$ равны соответственно потенциалам эмиттера и коллектора относительно земли.

Доказать, что если включенное сопротивление равно 500 ом, то увеличение I_e от 0,5 до 1,0 ма создаст отрицательное изменение в V'_b , определяемое величиной $500 \cdot (1,7 - 1,3) \cdot 10^{-3} = 0,2$ в, и, следовательно, V'_e уменьшится так, что полное входное сопротивление станет отрицательным и равным -300 ом. (В этом примере пренебрегаем малой поправкой, обусловленной влиянием V'_b на V_c .)

6. Рассмотреть график зависимости V_e от I_e для схемы, в которой на коллектор подано смещение -40 в, а основной электрод соединен с землей через сопротивление в 10 000 ом.

Показать, что когда $I_e = 0$, то $I_c \approx -1,0$ ма, $V'_b \approx -10$ в и $V'_e = V'_b + V_e \approx -10 + (-0,3) = -10,3$ в. Отрицательные эмиттерные

токи соответствуют большим отрицательным напряжениям на эмиттере. Другие рабочие точки можно найти графическим путем по кривой r_{21} . Для каждого значения I_e должно выполняться соотношение

$$-40 \text{ в} = 10\,000 (I_e + I_c) + V_c (I_e, I_c).$$

Из этого следует, что напряжение батареи должно быть равно сумме падений напряжения на сопротивлении и между основным электродом и коллектором. Уравнение можно переписать в виде

$$V_c (I_e, I_c) = [-40 \text{ в} - 10\,000 I_e] - 10\,000 I_c.$$

Для заданного значения I_c левая часть представляет собой просто одну из параметрических кривых графика r_{21} . Для этого же значения I_c правая часть есть просто прямая линия с наклоном $-10\,000 \text{ ом}$ на графике r_{21} . Таким образом, для каждого значения I_c решение дается точкой пересечения этой прямой с параметрической кривой.

Показать, что при этом получаются следующие точки:

I_c	I_e	$I_b \text{ (ма)}$	V_c	V'_b	$V'_e \text{ (в)}$
-1	0	1	-30	-10	-10,3
-2	0,5	1,5	-25	-15	-14,9
-3	+1,2	1,8	-22	-18	-17,8

Небольшая разница между V'_b и V'_e определяется величиной V_c . Показать, что отрицательное сопротивление, полученное из графика зависимости V'_e от I_e , будет продолжаться с наклоном в $-6\,000 \text{ ом}$ до значения $V'_b \approx -35 \text{ в}$, соответствующего группе кривых вблизи верхнего края графика r_{21} , экстраполированной к значению $I_e = 3,5 \text{ ма}$ и $I_c \approx -7 \text{ ма}$. При больших значениях I_e значение I_c будет лишь медленно возрастать, V_c изменится мало и на выходе возникнет положительное сопротивление. Это показывает, что кривая зависимости V'_e от I_e имеет N-образную форму, так что каждому заданному пределу смещений эмиттера будут соответствовать два стабильных режима и один нестабильный. Этой особенностью схемы можно воспользоваться для создания так называемых опрокидывающих схем, подавая на эмиттер отрицательное смещение через соответствующее сопротивление¹⁾.

7. а) Транзисторный усилитель с заземленным основным электродом подключен через трансформатор к генератору v_g с внутренним сопротивлением r_g . Показать, что если пренебречь шумом эмиттера, то наибольшее значение отношения сигнал/шум в нагрузке будет в том случае, когда отношение полных сопротивлений транзистора равно $r_g : r_{11}$.

¹⁾ Этот вывод принадлежит Фарли. — *Прим. авт.*

б) Предположим, что генератором служит обычный микрофон с подвижной катушкой, сопротивление которой $r_g = 22 \text{ ом}$, и что в нем при приеме разговорной речи на расстояние 1 м возникает э. д. с. $\mathcal{E} = 5 \cdot 10^{-6} \text{ в}$. Показать, что при надлежащем согласовании отношение сигнал/шум в нагрузке равно примерно $0,6 \text{ дб}$. Ширина полосы частот от 50 до $10\,000 \text{ гц}$.

в) Показать, что конечное отношение сигнал/шум не будет значительно отличаться от приведенного в предыдущем примере значения, если усилитель состоит из ряда каскадов, каждый из которых дает усиление в 20 дб .

8. На основании результатов предыдущей задачи показать, какое усиление можно получить с помощью транзисторного усилителя, рассчитанного на полосу частот от 50 до $10\,000 \text{ гц}$, если необходимо, чтобы уровень шумов на выходе был на 40 дб ниже максимального значения выходного сигнала. Мощность максимального выходного сигнала принимаем равной 10 мвт .

Ответ: примерно 46 дб .

9. Рассмотрим транзистор, работающий как усилитель с заземленным основным электродом на нагрузку в $10\,000 \text{ ом}$, последовательно включенную с батареей в 40 в .

а) Воспользовавшись кривыми фиг. 21, определить максимальное значение двойной амплитуды выходного тока при условии сравнительно малого его искажения.

б) Полагая, что входной ток (эмиттера) синусоидален, определить величину постоянного смещения эмиттера.

в) Определить величину мощности, рассеиваемой на коллекторе, если устранена переменная составляющая на входе. Какая мощность отбирается от батарей?

г) Пусть переменная составляющая входа искажена, но таким образом, что *выходной* ток имеет *синусоидальную* форму при максимальном значении двойной амплитуды. Определить значение мощности переменного тока, выделяемой в нагрузке.

д) Какая мощность будет рассеиваться коллектором и какая мощность будет отбираться от батареи при условиях, указанных в предыдущем примере?

Ответ: а) $2,2 \text{ ма}$; б) $0,75 \text{ ма}$; в) $39,6 \text{ мвт}$, 88 мвт ;
г) $6,1 \text{ мвт}$; д) $33,9 \text{ мвт}$, 80 мвт .

10. Показать таким же образом, как и в задаче 7, что отношение сигнал/шум при условии оптимального согласования не будет изменяться при параллельном включении ряда транзисторов; другими словами, фактор шумов согласованного транзистора не изменяться при параллельном включении таких же транзисторов.

ЛИТЕРАТУРА

1. Bardeen J., Brattain W. H., Phys. Rev., **75**, 1208 (1949).
2. Bardeen J., Brattain W. H., Phys. Rev., **74**, 230 (1948).
3. Ryder R. M., Bell Lab. Rec., **27**, 89 (1949).
4. Shive J. N., Phys. Rev., **75**, 689 (1949).
5. Becker J. A., Shive J. N., Electr. Eng., **68**, 222 (1949).
6. Ryder R. M., Kircher R. J., Bell Syst. Techn. Journ., **28**, 367 (1949).
7. Shockley W., Pearson G. L., Phys. Rev., **74**, 232 (1948).
8. Dow W. G., Fundamentals of Electronic Engineering, New York, 1937.
9. Bardeen J., Phys. Rev., **71**, 717 (1947).
10. Brattain W. H., Shockley W., Phys. Rev., **72**, 345 (1947).
11. Apker L., Taft E., Dickey J., Phys. Rev., **74**, 1462 (1948).
12. Guillemin E. A., Communication Networks, v. II, New York, 1935.
13. Peterson L. C., Bell Syst. Techn. Journ., **27**, 593 (1948).
14. Torrey H. C., Whitmer C. A., Crystal Rectifiers, New York, 1948 (см. перевод: Торрей и Уитмер, Кристаллические детекторы, М., 1950).
- 15*. Лосев О. В. Описание детекторного радиоприемника гетеродина, патент № 467 от 18 декабря 1923 г.
- 16*. Лосев О. В., Описание способа генерирования незатухающих колебаний, патент № 996 от 21 февраля 1922 г.
- 17*. Лосев О. В., Телегр. и телеф. без проводов, № 14, 374, № 15, 564 (1922).
- 18*. Лосев О. В., там же, № 18, 45; № 21, 349; № 22, 506 (1923).
- 19*. Лосев О. В., там же, № 26, 404 (1924).
- 20*. Лосев О. В., там же, № 5 (44), 485 (1927).
- 21*. Van Geel W. C., Патент США 2428400, 7 октября 1947.
- 22*. Stuetzer O. M., Proc. of the Inst. Rad. Eng., **38**, 865 (1950).
- 23*. Тамм И. Е., Sow. Phys., **1**, 733 (1932); ЖЭТФ, **3**, 34 (1933).
- 24*. Лашкарев В. Е., Изв. АН СССР, серия физич., **16**, 203 (1952).
- 25*. Hunter L. P., Proc. of the Inst. Rad. Eng., **38**, 1387 (1950).
- 26*. Воронов Р. А., Общая теория четырехполюсников и многополюсников, М.—Л., 1951.
- 27*. Ryder R. M., Kircher R. J., Bell Syst. Techn. Journ., **28**, 367 (1949).
- 28*. Ryder R. M., Kircher R. J., Proc. of the Inst. Rad., Eng., **40**, 360 (1952).
- 29*. Wallace R. L., Proc. of the Inst. Rad. Eng., **39**, 702 (1951).
- 30*. Wallace R. L., Raisbeck G., Bell Syst. Techn. Journ., **30**, 381 (1951).
- 31*. Raisbeck G., Electronics, **24**, 128 (1951).
- 32*. Ляшенко В. И., Павленко А. М., ЖТФ, **17**, 1097 (1947).
- 33*. Волькенштейн Ф. Ф., ЖФХ, **22**, 311 (1948).

КОЛИЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ ПРОЦЕССА ВВЕДЕНИЯ ДЫРОК И ЭЛЕКТРОНОВ В ГЕРМАНИИ

§ 1. ВВЕДЕНИЕ НОСИТЕЛЕЙ ТОКА В ГЕРМАНИИ¹⁾

Введение. Изобретение транзистора [2—4] послужило прямым стимулом к исследованию взаимодействия дырок и электронов в полупроводниках. Описанная в этом параграфе экспериментальная техника для изучения поведения дырок в германии n -типа частично уже применялась и ранее для исследования эмиттерного тока в транзисторах. В первых опубликованных работах Бардина и Браттейна [2,3] было высказано предположение, что дырочный ток протекает от эмиттера к коллектору лишь в поверхностном слое полупроводника. Предположение о том, что дырки могут проходить через всю толщу германия n -типа возникло в связи с появлением $p-n-p$ -транзистора [5]. Независимо от этого возможность прохождения дырок через сравнительно толстый слой германия n -типа была доказана Шайвом [6], изготовившим транзистор с точечными контактами, расположенными на противоположных сторонах образца германия. Это же было подтверждено и на примере коаксиального транзистора, исследованного Коком и Уоллесом [7]. Подробные данные о введении дырок в германий n -типа в условиях сильного поля были получены Райдером [8].

Все эти факты позволили сделать заключение [4], что когда оба точечных контакта расположены близко друг к другу на поверхности германия, как, например, в транзисторе типа А²⁾, дырки могут проходить от эмиттера как в поверхностном слое, так и через толщу образца германия. Для того чтобы поток дырок в поверхностном слое германия был значительным, необходимо, повидимому, провести специальную обработку поверхности. В описываемых здесь опытах такая обработка поверхности не производилась и полученные результаты вполне согласуются с предположением о том, что дырочный ток от эмиттера течет через толщу образца.

Описанные в этом параграфе опыты имеют большую практическую ценность. Кроме того, они позволили на количественной основе объяснить протекающие у контакта эмиттера явления и открыли новое направление исследования процессов проводимости в полупроводниках. В связи с этим здесь будет уместно сопоставить некоторые новые

¹⁾ Этот параграф несколько отличается по содержанию от ранее опубликованной статьи Шокли, Пирсона и Хайнса [1]. — *Прим. авт.*

²⁾ Подробнее относительно транзистора типа А см. [4, 9—11]. — *Прим. авт.*

представления об объемных свойствах полупроводников, вытекающие из этих опытов, с прежными.

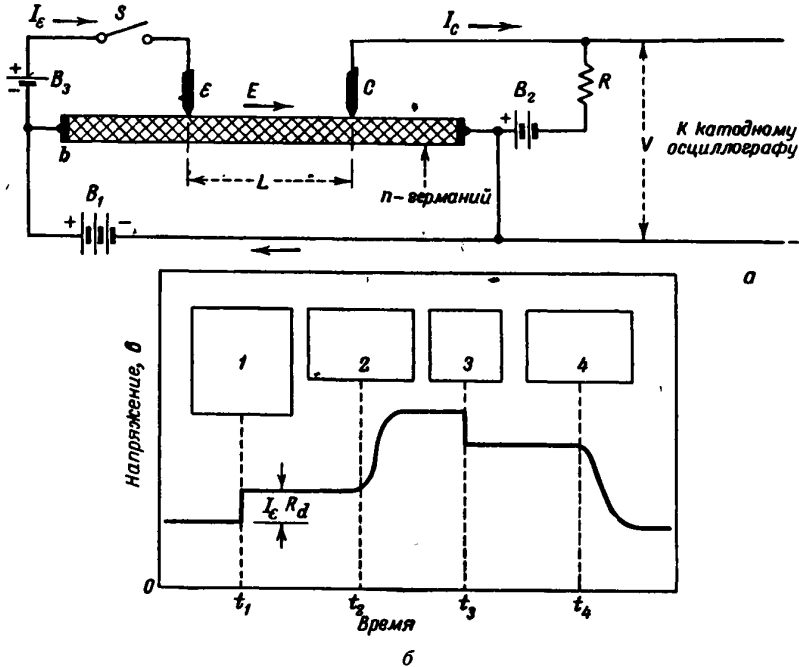
До появления транзистора все выводы, касающиеся поведения электронов и дырок в полупроводниках, делались в основном на базе измерения проводимости и эффекта Холла. В опытах по исследованию этих эффектов в стационарных условиях измерялись размеры образца, токи, разности потенциалов и магнитные поля. Временные характеристики при этом не изучались. Тем не менее на основании этих данных составлялось определенное представление о движении электронов и дырок в полупроводниках и, в частности, о скорости их дрейфа в электрическом поле и поперечном отклонении в магнитном поле. Новые опыты дали ряд прямых доказательств существования дрейфа электронов и дырок в полупроводниках, находящихся в полном согласии с ожидаемой величиной подвижности и направлением движения носителей тока. Кроме того, опыты, описываемые в § 2, показали, что в магнитном поле действительно наблюдается поперечное отклонение носителей тока, проявляющееся в изменении концентрации дырок и электронов вблизи одной из сторон тонкого нитевидного образца германия.

Рассмотрим здесь некоторые экспериментальные доказательства того факта, что дырки на самом деле „входят“ в германий n -типа при пропускании через эмиттер тока прямого направления, и покажем также, как на основании этих данных можно оценить концентрацию и время жизни дырок. В дальнейшем этот важнейший процесс мы просто будем называть *введением дырок*. Вопрос о введении дырок в германий из эмиттера для случая контакта металла с полупроводником подробно исследовали Бардин и Браттейн, а для случая p — n -перехода — Шокли. Относящиеся к этому вопросу теории рассматриваются нами в последующих параграфах. Существуют также другие возможности создания полупроводниковых усилителей и без использования явления введения дырок в германий n -типа или электронов в германий p -типа [2, 12]. В этой главе, однако, мы ограничимся рассмотрением только таких полупроводников, которые в состоянии теплового равновесия содержат в заметных количествах лишь один вид носителей тока; в подобных случаях теория значительно упрощается, но все же, видимо, находится в хорошем согласии с опытными данными.

Измерение концентрации и тока введенных носителей. Принципиальная постановка такого опыта в его полуквантитативной форме относительно проста и схематически иллюстрируется (фиг. 25¹). С помощью батареи V_1 вдоль тонкого стержня из германия n -типа создается продольное электрическое поле E . Точечные контакты коллектора и эмиттера прижимаются к образцу с помощью микромини-

¹ Такого рода опыты впервые были описаны Хайнсом и Шокли [13]. — *Прим. авт.*

пулятора. На коллектор, как и в случае транзистора типа А, с помощью батареи B_2 подается некоторое смещение, а снимаемый с нагрузочного сопротивления R сигнал подается на вход осциллографа. В момент времени t_1 цепь эмиттера замыкается и через эмиттер проходит создаваемый батареей B_3 ток в прямом направлении.



Фиг. 25. Опыт по исследованию движения дырок, введенных в германий n -типа.

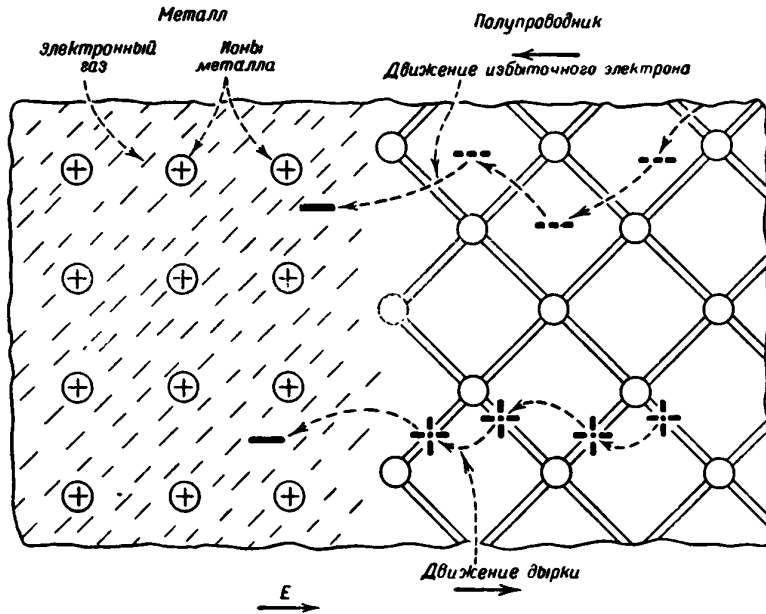
a — схема опыта; b — форма сигнала на осциллографе.

1 — замыкается контакт S (впуск дырок через b); 2 — дырки достигают коллектора C ; 3 — контакт S размыкается; 4 — дырки уходят за коллектор.

В момент времени t_1 цепь замыкается. Наблюдаемый при этом на осциллографе ход изменения величины напряжения на коллекторе изображен на фиг. 25, b .

Эту кривую можно интерпретировать следующим образом: при замыкании эмиттерной цепи электроны в проволочке эмиттера начинают двигаться в направлении от германия (т. е. в германий течет положительный ток). Эти электроны доставляются током, текущим в объеме германия по направлению к точке контакта. Ток в германии может быть обусловлен как имеющимися в германии избыточными электронами, так и процессом перемещения дырок.

На фиг. 26 показаны эти две возможности. На первый взгляд кажется, что различие между этими двумя процессами не существенно, поскольку в обоих случаях в конечном счете получается один и тот же результат — движение электронов из германия к контакту эмиттерного электрода. Однако на самом деле между этими двумя процессами имеется существенное различие, что и позволило осуществить



Фиг. 26. Сравнение электронного и дырочного токов в области точечного контакта.

некоторые виды транзисторов. В случае дырочного механизма проводимости электрон переходит в металл из основной (валентной) полосы полупроводника, после чего дырка перемещается глубже в германий. В результате, благодаря присутствию введенных в германий дырок, электронная структура германия вблизи контакта эмиттера видоизменяется. Под влиянием электрического поля E вошедшие в германий дырки дрейфуют по направлению к коллектору со скоростью $\mu_p E$, где μ_p — подвижность дырок. Эти дырки проходят расстояние до коллектора L за время $L/\mu_p E$. Достигая коллектора, они увеличивают его обратный ток и создают сигнал, регистрируемый в момент времени t_2 .

Между сигналами, отмечаемыми в моменты времени t_1 и t_2 , имеется два существенных различия. Сигнал в момент времени t_1 соответствует

возникновению нормального тока через образец при замыкании цепи эмиттера; этот сигнал, очевидно, будет наблюдаться и в тех случаях, когда никакого введения дырок в германий происходить не будет. Разъясним это на примере простого омического сопротивления. Для этого воспользуемся обычной теорией электрических цепей; однако чтобы сопоставить этот чисто омический случай со случаем, когда происходит введение дырок в германий, изложим здесь теорию передачи электрических сигналов с точки зрения движения носителей заряда. В соответствии с обычной теорией электрических цепей, добавочный ток I_e создаст просто дополнительный скачок потенциала IR , вызванный прохождением тока через часть образца, находящуюся справа от коллектора. Этот скачок потенциала на фиг. 25, б обозначен как $I_e R_d$, где R_d — сопротивление разветвленной цепи, по которой протекает ток I_e . Этот сигнал будет передаваться от эмиттера к коллектору практически со скоростью света, и к нему применима обычная теория распространения электрических сигналов вдоль проводников. Столь большая скорость распространения сигнала отнюдь не означает, что с такой же большой скоростью будут двигаться и носители тока. На самом деле скорость передачи сигнала не имеет ничего общего со скоростью движения носителей тока, что легко понять на основании следующих соображений: если вещество является чисто омическим электронным проводником, то уход из него некоторого числа электронов при прохождении эмиттерного тока создает в нем локальный положительный заряд. В свою очередь этот заряд создает электрическое поле, распространяющееся со скоростью света. Под действием поля близлежащие электроны устремляются в область объема заряда для его нейтрализации. В результате практически мгновенно начинают перемещаться электроны во всем объеме образца. Электроны, покидающие образец через эмиттерный электрод, замещаются электронами, входящими через концевые электроды, так что нигде внутри образца концентрация электронов заметным образом не меняется¹⁾.

Различие между только что описанным процессом и тем, который происходит, когда в германий вводятся дырки, имеет важнейшее значение для понимания многих явлений, связанных с действием транзисторов. Это положение до некоторой степени можно резюмировать следующим образом: в образце, в котором ток переносится носителями заряда лишь одного вида, например электронами, невозможно

¹⁾ Это есть не что иное, как словесное описание результата, получаемого обычно при помощи понятия о времени диэлектрической релаксации, которое вводится следующим образом:

$$\nabla I = -\dot{\rho}, I = \sigma E = -\sigma \nabla \psi, \nabla^2 \psi = -4\pi \rho / \kappa = \dot{\rho} / \sigma,$$

так что $\rho = \rho_0 \exp[-(4\pi\sigma/\kappa)t]$, где I — плотность тока, ρ — плотность зарядов, σ — проводимость, E — напряженность электрического поля, ψ — электростатический потенциал, κ — диэлектрическая постоянная. Величина $\kappa/4\pi\sigma$ и есть время диэлектрической релаксации. — *Прим. авт.*

изменить их концентрацию, пытаясь ввести или вывести из него носителя заряда того же знака. Как уже указывалось выше (и отмечалось в примечании), это получается потому, что в этом случае всякие изменения концентрации носителей сопровождались бы появлением в образце некомпенсированного объемного заряда, который сразу бы и уничтожался¹⁾.

Введение дырок в германий с проводимостью n -типа также сопровождается образованием объемного заряда, который немедленно нейтрализуется электронным потоком. В этом случае, однако, нейтральное состояние не является нормальным состоянием теплового равновесия. Общее количество носителей заряда увеличивается за счет введенных в германий дырок и такого же числа электронов, привлеченных для нейтрализации заряда этих дырок. Общее число электронов в образце возрастает, таким образом, за счет избыточных электронов, входящих в образец из металлических электродов, образующих общую электрическую цепь с эмиттером. Присутствующие вблизи эмиттера избыточные дырки и нейтрализующие их объемный заряд электроны изменяют проводимость вещества в этом месте. Это изменение проводимости, как будет указано ниже, может быть столь большим, что им можно воспользоваться для определения концентрации введенных в образец дырок, а также для получения усиления по мощности в транзисторах различного типа. Все изложенное можно резюмировать следующим образом: *в полупроводнике, содержащем в основном лишь один тип носителей тока, невозможно увеличить общую концентрацию носителей введением носителей тока этого же типа; такое увеличение концентрации может быть достигнуто лишь при введении носителей другого знака, поскольку объемный заряд последних может быть нейтрализован возрастающей концентрацией нормальных носителей тока в этом образце.*

Таким образом, мы приходим к выводу, что существование двух процессов электронной проводимости в полупроводниках, соответствующих наличию положительных и отрицательных подвижных носителей тока, является необходимым условием для работы транзистора.

В вакуумном триоде имеются также два способа переноса тока электронами: 1) обычная металлическая проводимость в электродах и 2) ток в вакууме, переносимый электронами, возникающими при термоионной эмиссии с катода. В § 2 гл. 4 при сравнении транзистора с $p-n-p$ -переходами с вакуумным триодом будет показано,

¹⁾ В случае модуляции проводимости поверхностными зарядами, о чем речь шла в § 1 гл. 2, результирующий заряд создается полем пластины конденсатора. Измененная плотность заряда наблюдается и на некоторой глубине внутри образца, но это явление не следует смешивать с истинным объемным процессом, имеющим место при введении дырок. Такие поверхностные слои зарядов кратко рассмотрены в § 3 гл. 4. — *Прим. авт.*

что в обоих этих случаях существует близкая аналогия в том, как взаимодействие этих двух видов тока приводит к усилению.

В свете изложенного нетрудно объяснить суть опыта, схема которого изображена на фиг. 25. Мгновенный скачок потенциала в момент времени t_1 является просто чисто омическим скачком, обусловленным изменением общего тока в правой части образца германия при возникновении эмиттерного тока. После этого проходит некоторый промежуток времени, в течение которого введенные в германий дырки перемещаются вдоль образца и достигают коллектора. Затем, когда в момент времени t_2 ток в цепи эмиттера выключается, наблюдается подобная же последовательность явлений.

Зная промежуток времени $t_1 - t_2$, величину напряженности электрического поля E и расстояние L , можно определить подвижность дырок. Тот факт, что здесь мы имеем дело с дырками, а не с электронами, следует из полярности этого эффекта; возмущение, создаваемое эмиттером, перемещается в направлении E так, как если бы оно создавалось положительными зарядами; если изменить направление электрического поля, то сигнал, наблюдавшийся в момент времени t_2 , совершенно исчезает. Полученные этим путем значения для подвижности носителей заряда в германии находятся в хорошем согласии с теми значениями, которые получаются из проводимости и эффекта Холла. Подвижность Холла, полученная на монокристаллах германия n - и p -типа, имеет следующие значения [14]:

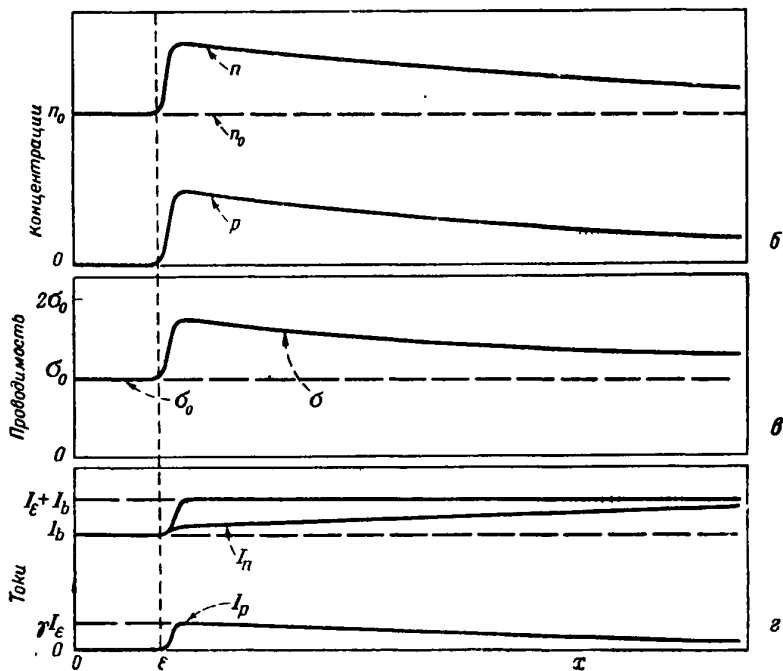
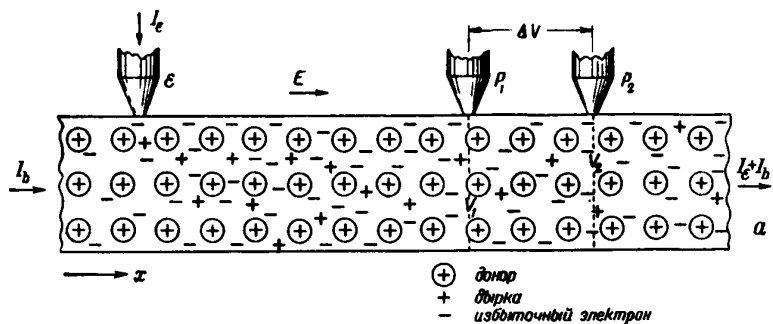
$$\mu_p = 1\,700 \text{ см}^2/\text{в} \cdot \text{сек}, \quad (3.1)$$

$$\mu_n = 2\,600 \text{ см}^2/\text{в} \cdot \text{сек}. \quad (3.2)$$

Совпадение между значениями подвижности Холла и дрейфа, как уже указывалось в начале этого параграфа, является удовлетворительным подтверждением общей теоретической картины движения дырок в направлении электрического поля¹⁾.

Рассмотрим теперь несколько подробнее количественную сторону описанного эксперимента. На фиг. 27 приведена схема постановки такого опыта. Для получения более четко выраженных эффектов в этом случае существенно, чтобы поперечное сечение нитевидного образца германия было малым. Тонкая пластинка германия приклеивается специальным цементом к стеклянной пластинке и шлифуется до желаемой толщины. Затем на образец накладывается металлический защитный шаблон подходящей формы и ненужная часть образца удаляется обработкой пескоструйным аппаратом. После изготовления образца поверхность германия подвергается травлению. Описанный способ позволяет получать образцы с поперечным сечением, меньшим $0,01 \cdot 0,01 \text{ см}$. Концы нитевидного образца обычно делают более широкими, чтобы упростить задачу изготовления контактов.

¹⁾ Более новые данные по этому вопросу приведены в гл. 12. — *Полн. авт.*



Фиг. 27. Метод измерения концентрации дырок и дырочных токов.

а — распределение дырок, электронов и доноров. Для простоты акцепторы не указаны. Разность концентраций доноров N_d и акцепторов N_a равна n_0 ; б — справа от эмиттера дополнительная концентрация дырок p компенсируется равным увеличением концентрации электронов. Концентрация электронов $n = n_0 + p$; разность концентраций доноров и акцепторов равна $n_0 = N_d - N_a$; в — проводимость равна сумме дырочной и электронной проводимостей; $\sigma_0 = e b \mu_p n_0$; $\sigma = e \mu_p p + e b \mu_p n = e \mu_p [p + b(n_0 + p)] = \sigma_0 [1 + (1 + b)(p/n_0)]$; з — полный ток $I_b + I_a$ справа от эмиттера определяется токами I_p и I_n , взятыми в отношении, равном отношению дырочной и электронной проводимостей; I_n — электронный ток; I_p — дырочный ток, $I_p = I_a + I_b \{(\sigma - \sigma_0)/(1 + b)\sigma\}$; x — расстояние вдоль полупроводника.

По условиям опыта указанная на фиг. 25 батарея B_1 создает в образце «развертывающее» поле такого направления, что дырки, введенные в образец с током эмиттера, передвигаются вдоль образца слева направо. В применяемых для этих опытов маленьких нитевидных образцах германия результирующая концентрация введенных дырок настолько велика, что справа от эмиттера наступает сильное изменение сопротивления. Как будет показано ниже, это изменение сопротивления можно измерить и по нему определить дырочный ток эмиттера. Для количественного рассмотрения этого вопроса введем величину γ и дадим ей следующее определение:

$$\gamma = \text{доле эмиттерного тока, переносимой дырками.} \quad (3.3)$$

Следовательно, справа от эмиттера протекает дырочный ток γI_e и создает концентрацию дырок, равную p , которая нейтрализуется равной концентрацией дополнительных электронов. Часть электронного тока, равная $(1 - \gamma) I_e$, протекает влево по образцу, однако эти электроны не приводят к увеличению концентрации электронов слева от эмиттера, так как они имеют тот же знак, что и нормальные носители тока в германии n -типа. Наличие дырок в правой части образца приводит к увеличению проводимости σ (см. фиг. 27, в) как благодаря их собственному присутствию, так и вследствие появления дополнительных электронов, втянутых в эту область образца для нейтрализации положительного объемного заряда, созданного дырками. Подвижность электронов больше подвижности дырок; в германии отношение подвижностей равно ¹⁾

$$b = \frac{\mu_n}{\mu_p} = 1,5. \quad (3.4)$$

Концентрация электронов в этом случае также всегда больше концентрации дырок ²⁾

$$n = n_0 + p. \quad (3.5)^3)$$

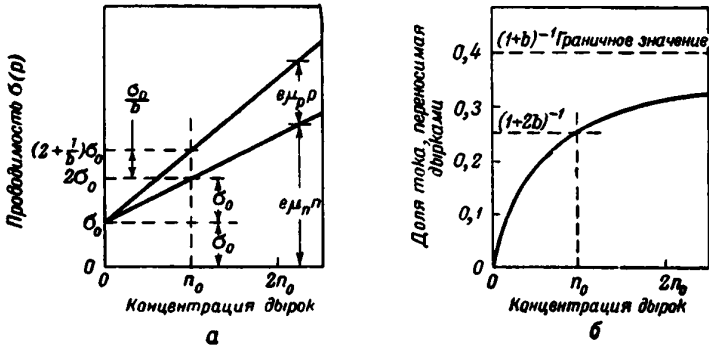
Здесь n_0 — концентрация электронов, необходимая для нейтрализации доноров, при условии, что $p = 0$; отсюда следует, что электронный ток больше дырочного. Уменьшение концентрации дырок справа от эмиттера получается в результате рекомбинации их с электронами при перемещении вдоль образца.

¹⁾ См. [14]. Более новые данные приведены в гл. 12. — *Прим. авт.*

²⁾ Здесь приняты следующие обозначения: n, p, n_0 — соответственно концентрации электронов, дырок и начальная концентрация электронов, имеющаяся в образце до того, как в него были введены дырки; N_d и N_n — концентрации доноров и акцепторов. Если предполагать, что все они ионизированы, то $n_0 = N_d - N_n$. Значения I_e, I_1, I_c указаны на фиг. 27. (На фиг. 25 и 33 через I_c обозначен ток коллектора; он в уравнениях не входит.) — *Прим. авт.*

³⁾ См. задачи в конце этой главы. — *Прим. авт.*

Этот опыт позволяет определить значение коэффициента γ и время жизни дырок в образце. Измерения производятся с помощью двух точечных зондов P_1 и P_2 , расположенных на двух эквипотенциальных



Фиг. 28. Зависимость проводимости $\sigma(p)$ и доли тока, переносимого дырками, от концентрации дырок p (для $b = 1,5$).

поверхностях V_1 и V_2 и измеряющих падение напряжения между ними ¹⁾. Проводимость нитевидного образца германия между этими зондами определяется как частное от деления измеренной разности потенциалов ΔV на ток $I_b + I_a$. Приведенные в подписи к фиг. 27 формулы, по которым вычисляется концентрация дырок и сила дырочного тока, выводятся следующим образом:

Нормальная проводимость

$$\sigma_0 = e\mu_n n_i. \tag{3.6}$$

Проводимость при наличии дырок

$$\sigma = e\mu_n n + e\mu_p p = e\mu_n (n_0 + p) + e\mu_p p = \sigma_0 \left[1 + (1 + b^{-1}) \frac{p}{n_0} \right]. \tag{3.7}$$

На фиг. 28, а изображен график этой зависимости как функции от p/n_0 .

Активная проводимость образца на участке между зондами P_1 и P_2

$$G = \frac{I_a + I_b}{\Delta V} \tag{3.8}$$

пропорциональна локальной проводимости и, следовательно, пропорциональна

$$1 + (1 + b^{-1}) \frac{p}{n_0}. \tag{3.9}$$

¹⁾ Когда концентрации электронов и дырок в полупроводнике не равны их равновесным значениям, измерения потенциала точечным контактом будут зависеть от природы контакта. Подробный разбор этого вопроса см. в работе [5]. — Прим. авт.

Таким образом, измерения проводимости позволяют легко определить значение отношения p/n_0 . Обозначив через G и G_0 значения активной проводимости участка образца между зондами соответственно при наличии и в отсутствие введения дырок из эмиттера, получаем

$$\frac{G}{G_0} = \frac{\sigma}{\sigma_0} = 1 + (1 + b^{-1}) \frac{p}{n_0}, \quad (3.10)$$

или

$$\frac{p}{n_0} = \frac{\sigma - \sigma_0}{\sigma_0(1 + b^{-1})} = \frac{(G/G_0) - 1}{1 + b^{-1}}. \quad (3.11)$$

Отношение величины дырочного тока к общему току, естественно, такое же, как отношение дырочной проводимости к общей проводимости. Это отношение $I_p/(I_n + I_p)$ графически изображено на фиг. 28, б. Величину I_p можно определить графически следующим образом: сначала по измеренному значению G/G_0 определяется p/n_0 (см. фиг. 28, а), а затем, пользуясь значением p/n_0 , по фиг. 27, б находится отношение $I_p/(I_n + I_p)$. Так как величина $(I_n + I_p)$ известна из опыта, то нетрудно вычислить I_p . Этот метод может быть выражен соотношением

$$\begin{aligned} \frac{I_p}{I_n + I_p} &= \frac{e\mu_p p}{e\mu_n n + e\mu_p p} = \frac{p}{bn_0 + (1 + b)p} = \\ &= \frac{p/n_0}{b[1 + (1 + b^{-1})(p/n_0)]} = \frac{1 - (G_0/G)}{1 + b}. \end{aligned} \quad (3.12)$$

Таким образом, по измеренным значениям G можно определить переносимую дырками долю общего тока, протекающего через образец. Дырочный ток, протекающий через образец в его части, находящейся за зондами, может быть вычислен умножением (3.12) на $I_n + I_p = I_a + I_b$. Это приводит к формуле

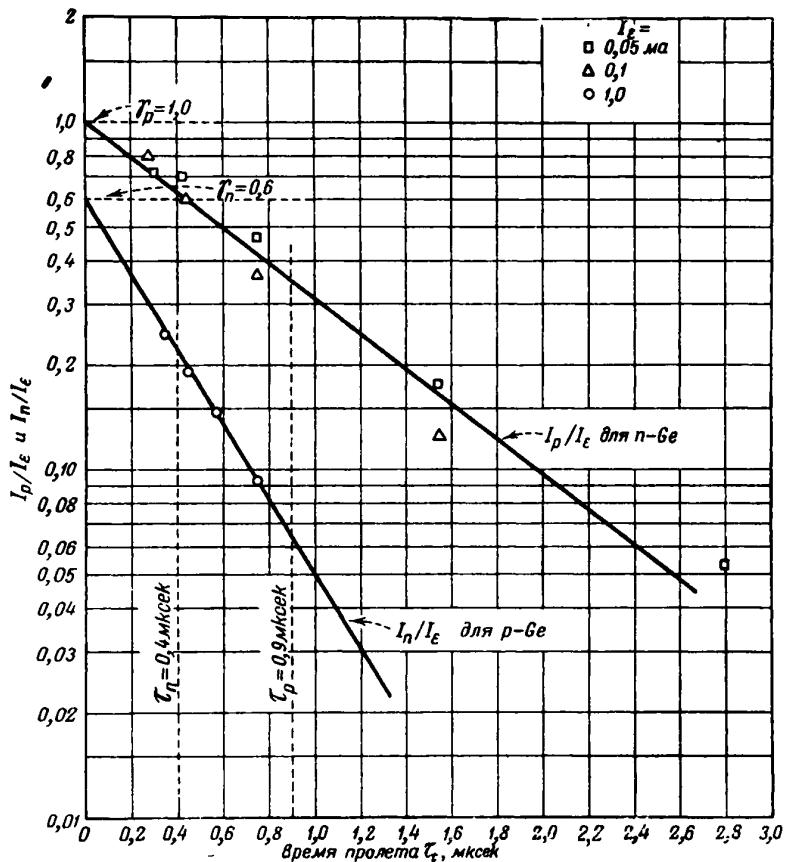
$$I_p = \frac{(I_a + I_b)[1 - (G_0/G)]}{1 + b} = \frac{I_a + I_b - G_0\Delta V}{1 + b}, \quad (3.13)$$

где ΔV определяется из (3.8). Это уравнение можно интерпретировать следующим образом: $G_0\Delta V$ — ток, соответствующий нормальной концентрации электронов n_0 ; числитель дроби представляет собой избыточный ток, переносимый введенными дырками и добавочными электронами. Часть этого избыточного тока, $1/(1 + b)$, переносится дырками, и формула (3.13) определяет величину дырочного тока¹⁾.

Если затухание импульса тока в образце, обусловленное комбинацией носителей, отсутствует, то ток, протекающий в части образца, лежащей за зондами, можно представить как γI_e , а поскольку значение I_e известно, легко определить и γ . В действительности же

¹⁾ При этом расчете мы воспользовались формулой $n = p + n_0$, соответствующей условию полной ионизации акцепторов и доноров. В случае германия это предположение является достаточно хорошим приближением. Для кремния, однако, требуются уже иные предположения. — *Прим. авт.*

спадание тока может быть довольно значительным. Однако при увеличении тока I_b дырки начинают быстрее двигаться от эмиттера к зондам, что приводит к уменьшению скорости спадания тока.



Фиг. 29. Экстраполяция экспериментальных значений дырочного и электронного токов к нулевому значению времени пролета для определения γ .

Увеличивая постепенно I_b , можно свести до минимума влияние рекомбинации, а величину дырочного тока экстраполировать к тому значению, которое она имела бы в отсутствие спадания. Это значение, конечно, и есть γ_{I_s} .

Некоторые графики такого рода показаны на фиг. 29; по оси ординат в логарифмическом масштабе отложены значения отношения I_p/I_s , которое стремится к γ по мере увеличения I_b ; по оси абсцисс

отложено «время пролета»; отсутствие затухания, или нулевое значение времени пролета, соответствует началу координат¹⁾. Вывод, который можно сделать из анализа этого графика, заключается в том, что в образцах с проводимостью n -типа $\gamma = 1$ и весь ток эмиттера является дырочным током. Напротив, в случае вещества с проводимостью p -типа²⁾ через эмиттер вводятся в образец электроны. Величина отношения I_n/I_e , полученная путем экстраполяции для $\tau_e = 0$, равна 0,6, откуда следует, что в данном случае 60% тока переносится электронами, а 40% — дырками. Установлено, что время жизни носителей тока в этих образцах равно 0,9 и 0,41 мксек. соответственно. Имеется ряд данных, часть которых будет рассмотрена ниже, свидетельствующих о том, что дырки рекомбинируют с электронами главным образом на поверхности нитевидного образца германия.

Влияние концентрации дырок на характеристики точечных контактов. Наличие дырок вблизи контакта коллектора приводит к увеличению его обратного тока. Усиление в транзисторе типа А на самом деле обуславливается модуляцией коллекторного тока, производимой дырочной частью тока эмиттера. Влияние концентрации дырок на ток коллектора изучалось в связи с опытами, подобными рассмотренному на фиг. 27.³⁾ После того как измерялись дырочный ток и концентрация дырок, к образцу прикладывалось обратное смещение в 20 или 40 в. При этом было установлено, что обратный ток является линейной функцией концентрации дырок. Эта зависимость иллюстрируется фиг. 30. Как видно из графика, разные коллекторные зонды имеют совершенно различные сопротивления⁴⁾. Но если для какого-нибудь точечного зонда построить кривую, подобную тем, которые изображены на фиг. 30, то по величине тока можно определить и концентрацию дырок. Этот экспериментальный способ определения концентрации дырок проще способа с двумя зондами и лучше применим к изучению переходных явлений. При пользовании этим методом необходимо следить за тем, чтобы ток коллектора был несколько меньше $I_b + I_e$ (в противном случае вносимые

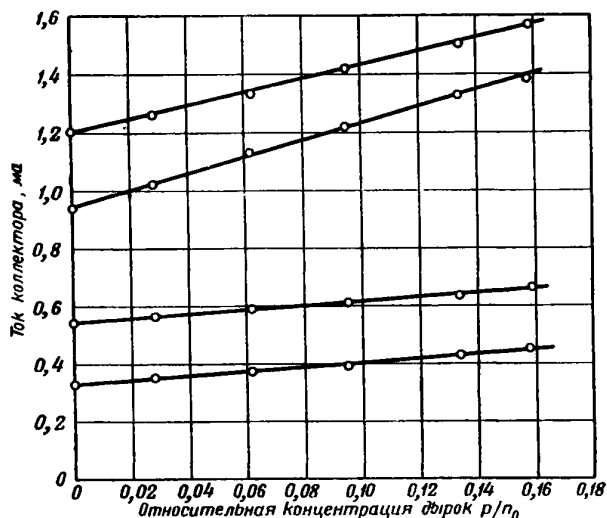
¹⁾ Если время жизни дырки есть τ , то дырочный ток у зондов $I_p = \gamma I_e \exp(-t/\tau)$, где t — «время пролета» дыркой половины расстояния между зондами, скажем, расстояния L от эмиттера. Если электрическое поле $E = \Delta V/\Delta L$, то «время пролета» $t = L\Delta L/\mu_p \Delta V$. Следовательно, если откладывать величину $\ln I_p$, определенную из отношения проводимости как функцию $t = L\Delta L/\mu_p \Delta V$, то при этом получится прямая линия, пересекающая ось ординат в точке $\ln \gamma I_e$ и имеющая наклон $-1/\tau$ (см. гл.12). — *Прим. авт.*

²⁾ Транзисторы, изготовленные из германия p -типа, были описаны Пфаном и Скафом [15]. Введение электронов в германий p -типа наблюдал также Брей [16]. — *Прим. авт.*

³⁾ Влияние состава точечного металлического электрода на свойства выпрямляющих контактов металл — полупроводник исследовалось в работах [32, 33], а также [23]. — *Прим. ред.*

коллекторным током искажения будут очень велики и получаемые данные не будут соответствовать свойствам самого образца). Однако опыты показали, что эти условия легко выполнимы и измерение коллекторного тока может быть с успехом использовано для определения концентрации дырок.

Концентрация дырок изменяет также сопротивление контакта зонда и при малом напряжении. Изучение этого эффекта также производилось в связи с опытом, схематически изображенным на фиг. 27.



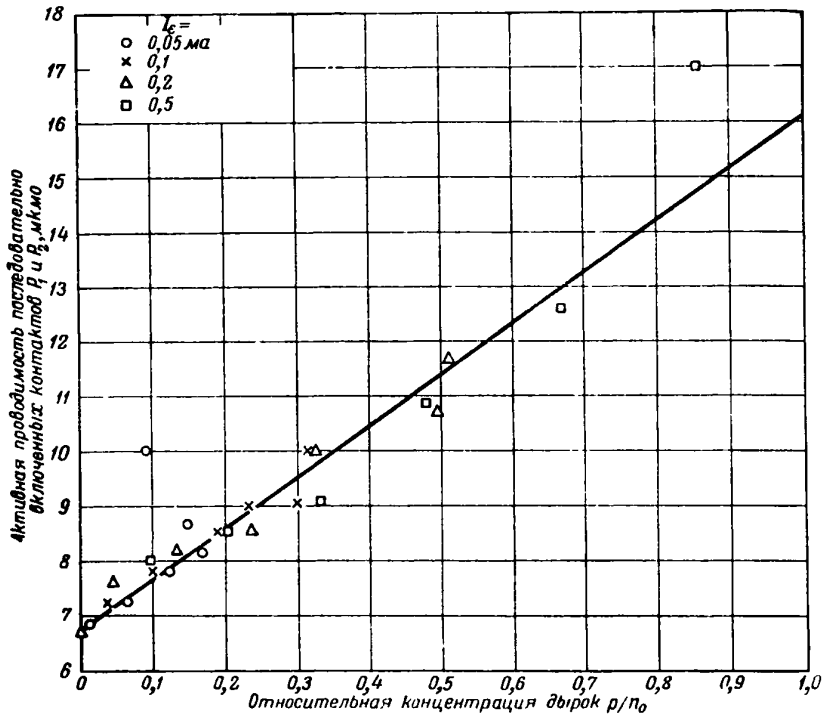
Фиг. 30. Зависимость коллекторного тока I_k от среднего значения концентрации дырок, введенных в коллектор.

На коллектор подано обратное смещение в 20 в.

После того как из ΔV и $I_b + I_c$ была определена концентрация дырок, между зондами P_1 и P_2 прикладывалась небольшая добавочная разность потенциалов (0,015 в) и измерялся ток между ними. На основании этих данных для случая малых токов определялась дифференциальная проводимость участка образца между двумя последовательно включенными зондами P_1 и P_2 .

Как видно из фиг. 31, эта проводимость является линейной функцией концентрации дырок. Проводимость точечного контакта обусловлена частично потоком электронов, а частично — дырок. На основании опытов с магнитными полями [17] установлено, что в равновесных условиях доли участия этих двух процессов в проводимости точечного контакта сравнимы. Следует отметить, что отложенная по оси абсцисс на фиг. 31 концентрация дырок является средней концентрацией дырок по всему поперечному сечению образца; вблизи

самой поверхности (благодаря процессу поверхностной рекомбинации) концентрация дырок может быть значительно меньше. Описанная здесь экспериментальная техника может быть применена для исследования свойств точечных контактов коллекторных зондов. Если поместить коллектор между эмиттером и зондом P_1 (см. фиг. 27), то



Фиг. 31. Активная проводимость последовательно включенных контактов P_1 и P_2 (см. фиг. 27) как функция отношения p/p_0 .

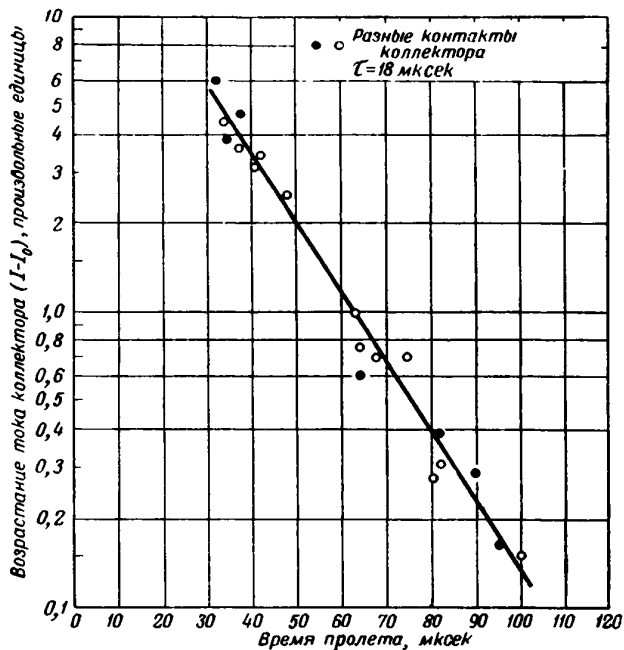
Из рисунка следует, что активная проводимость зависит от концентрации дырок и не зависит от токов в нитевидном образце. При каждом значении I_E концентрация дырок меняется путем изменения значения $I_b + I_e$ в пределах от 0,033 до 0,78 мА.

дырочный ток коллектора может быть определен из значения дырочного тока за электродами P_1 и P_2 . Этим путем можно определить «собственный» коэффициент усиления по току α_i коллектора, который равен отношению изменения коллекторного тока к изменению дырочного тока (фактически достигающего коллектора).

Изучение переходных процессов. Экспериментальная техника, основанная на применении точечного электрода коллектора для измерения

концентрации дырок, впоследствии была использована в ряде опытов, подобных описанному выше (см. фиг. 25). На основании этих опытов получены данные о времени жизни дырок и их подвижности, диффузии и модуляции проводимости.

Один из методов определения времени жизни сводится к измерению увеличения тока коллектора, обусловленного прохождением



Фиг. 32. Спадание тока введенных дырок в образце германия *n*-типа.

фронта импульса дырочного тока в зависимости от «времени пролета» дырками расстояния от эмиттера до коллектора.

На фиг. 32 приведен график такой зависимости, полученный для образца германия размерами $1,0 \times 0,05 \times 0,08$ см. Как это следует из графика, возросший вследствие прихода дырок коллекторный ток экспоненциально затухает с постоянной времени, равной 18 мксек. Эта постоянная времени возрастает с увеличением размеров образца. Для образца германия размерами $2,5 \times 0,35 \times 0,30$ см постоянная времени составляла уже 140 мксек. Так как введенные в объем этого образца дырки могут диффундировать к поверхности и рекомбинировать за время порядка 100 мксек., то процесс может несколько затягиваться вследствие рекомбинации на поверхности. Во всяком

случае, можно утверждать, что время жизни дырок в объеме германия должно быть равным по крайней мере 140 мксек. Используя данные о концентрации электронов, полученные из других измерений, мы приходим к заключению, что поперечное сечение рекомбинации должно быть меньше 10^{-18} см². Это сечение, составляющее меньше 1/400 площади поперечного сечения атома германия, столь мало потому, что для пары электрон — дырка в кристалле существуют затруднения в выполнении условий, аналогичных требованию сохранения энергии и количества движения, препятствующих рекомбинации электронов с положительными ионами в газовом разряде. Необходимо подчеркнуть, что пара электрон — дырка будет находиться в самом низком из возможных энергетических состояний. Электрон и дырка связаны наподобие протона и электрона в водородном атоме [18]. Такую комбинацию электрона и дырки называют *эксцитомом*, а энергию, выделяемую ими при рекомбинации, — *энергией эксцитона*. Чтобы рекомбинировать, эксцитоны должны выделить эту энергию в виде светового кванта (фотона) или в виде кванта тепловых колебаний решетки (фонона). Время рекомбинации эксцитона, сопровождающейся излучением фотона, можно оценить из оптических постоянных германия, воспользовавшись теорией излучения с учетом принципа детального равновесия, из которого следует, что в условиях равновесия скорость образования пар дырка — электрон при поглощении фотона равна скорости их рекомбинации с испусканием фотона; время жизни эксцитона, определенное этим способом, составляет при комнатной температуре величину порядка 1 сек., что указывает на малую роль процессов рекомбинации с испусканием фотонов¹⁾. Лоусон²⁾ указал, что энергия эксцитона во много раз превышает энергию фонона и поэтому выделение энергии эксцитона требует участия многих фононов с соответственно малой вероятностью перехода³⁾.

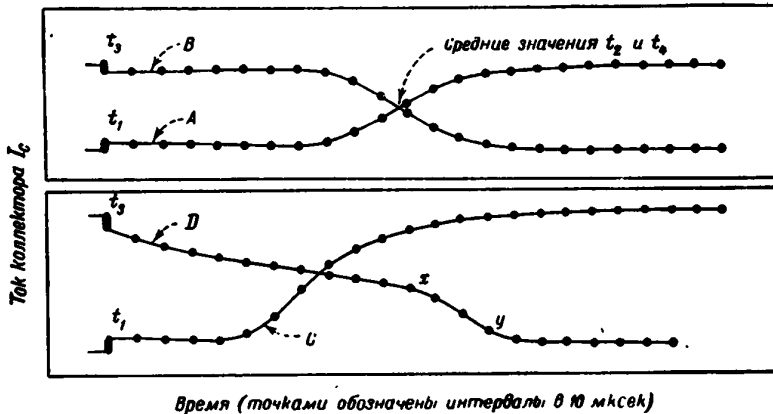
При введении через эмиттер прямоугольных импульсов дырочного тока в опытах, подобных изображенному на фиг. 25, по ряду причин наблюдается искажение переднего фронта и конца импульса, достигающего коллектора. Благодаря тому, что в области контакта эмиттера существует сильное локальное электрическое поле, в первый момент часть дырок начинает двигаться в неправильном направлении — в сторону, противоположную от коллектора; траектории

¹⁾ Оптические постоянные для германия впервые были опубликованы Браттейном и Бриггсом [19] и Бриггсом [20]. Интегрирование по всей области распределения энергии было произведено Ван Роозбруком. — *Прим. авт.*

²⁾ Частное сообщение; аналогичный случай рассмотрен Гудменом, Лоусоном и Шиффом [21]. — *Прим. авт.*

³⁾ Представление об «эксцитонах» было введено в теорию Я. И. Френкелем [24, 25]. В настоящее время имеется ряд убедительных экспериментальных фактов, доказывающих существование эксцитона [26—29]. — *Прим. ред.*

движения их затем искривляются, и они также в конце концов достигают коллектора, но уже за более длительное время, нежели дырки, которые с самого начала стали двигаться по направлению к коллектору. Такого рода разброс значений „времен пролета“, вероятно, в наибольшей степени ответствен за уменьшение коэффициента усиления транзистора на высоких частотах. Однако в описываемых ниже опытах этот эффект пренебрежимо мал по сравнению с двумя другими эффектами, которые мы сейчас рассмотрим.



Фиг. 33. Характеристика тока коллектора для схемы фиг. 25.

A—приход дырок; *B*—уход дырок (ток эмиттера мал, составляет около 4% I_b , так что все дырки движутся в одном и том же поле); *C*—приход дырок (набегающий край импульса при $I_a = \frac{1}{2} I_b$); *D*—уход дырок (задний край дырочного импульса при $I_a = \frac{1}{2} I_b$; участок от *x* до *y* обусловлен запаздыванием дырок).

На направленный дрейф дырок в электрическом поле накладывается беспорядочное перемещение их, вызванное тепловым движением. Это приводит к размытию первоначально резкой формы импульса дырок; за время дрейфа t_d это размытие края импульса достигает величины $\sqrt{Dt_d}$, где D —коэффициент диффузии дырок, равный $kT_p/e \approx 45 \text{ см}^2/\text{сек}$. В результате этого явления импульс доходит до коллектора с размытыми передним фронтом и концом. Это показано на фиг. 33, где кривая *A*—передний набегающий край импульса, а кривая *B*—конец импульса. Отметки времени даны через каждые 10 мксек., а начало отсчета совпадает с моментом включения тока эмиттера. В случае *A* и *B* эмиттерный ток был настолько мал в сравнении с током I_b , что дырки создавали лишь незначительную модуляцию проводимости и каждая дырка двигалась, по существу, в одинаковом электрическом поле. Как видно из графика,

форма импульса почти симметрична во времени относительно средней точки фронта нарастания, а кривые A и B идентичны, за исключением их знака. Именно этого и следовало ожидать, принимая во внимание диффузию. Больше того, анализ показывает, что разброс времен подхода дырок к коллектору находится в хорошем количественном согласии с теоретически ожидаемой крутизной фронта волны с учетом соответствующего коэффициента диффузии для дырок. В этом случае средняя точка кривой нарастания импульса, соответствующая точке пересечения кривых, дает среднее время пролета дырок, которое и используется для точного определения подвижности.

Кривые C и D соответствуют условиям, при которых ток эмиттера сравнительно велик и равен $\frac{2}{3}$ тока основного электрода. В этом случае применялись источники питания с высоким полным сопротивлением, так что ток I_b оставался постоянным, а I_e представлял собой сигнал с плоской верхушкой. При тех токах, которые протекали через образец в этих опытах, дырки заметно модулировали проводимость образца. Это сказывается на форме кривой C , характеризующей последовательность прихода дырок к коллектору.

Как видно из графика, эта кривая не симметрична; она значительно более пологая в части, соответствующей большим временам. Это получается потому, что до коллектора сначала доходят те дырки, которые первыми начинают диффундировать и движутся в веществе с еще низкой проводимостью. Последующие дырки движутся уже в среде с относительно более высокой проводимостью, а следовательно, и в более слабом электрическом поле. (Поскольку во всех точках между эмиттером и коллектором ток одинаков, поле обратно пропорционально проводимости.) В связи с этим время пролета для последующих дырок возрастает и их концентрация для конечной части импульса устанавливается более медленно. Форма сигнала, соответствующего заднему краю импульса эмиттерного тока (кривая D), прямо противоположна форме сигнала от его набегающего края.

Постепенное спадание тока коллектора вначале, вплоть до момента x , обусловлено рекомбинацией дырок и электронов. В момент времени t_3 ток эмиттера становится равным нулю; напряженность электрического поля в образце при этом падает, а время пролета дырок, достигающих коллектора к моменту x , больше, чем для дырок, подходящих к нему в момент времени t_3 , вследствие чего значительная часть дырок рекомбинирует с электронами. Самый задний край импульса, достигающего коллектора (участок кривой D от x до y), значительно круче его набегающего края. Это связано с тем обстоятельством, что дырки, отстающие по отношению к их главной массе, перемещаются уже в области относительно низкой проводимости и сильного электрического поля, нагоняя основную

массу дырок. Таким образом, тот же эффект, который удлинит сигнал C , одновременно укорачивает сигнал D .

Решение соответствующих уравнений, описывающих изложенные здесь явления, получено Херингом [22].

Рассмотренные в связи с фиг. 25 и 33 явления запаздывания сигнала указывают на интересные возможности применения нитевидных образцов германия в качестве элементов задерживающих или накапливающих линий.

§ 2. ВЛИЯНИЕ МАГНИТНОГО ПОЛЯ НА КОНЦЕНТРАЦИЮ ДЫРОК И ЭЛЕКТРОНОВ. ЭФФЕКТ СУЛЯ

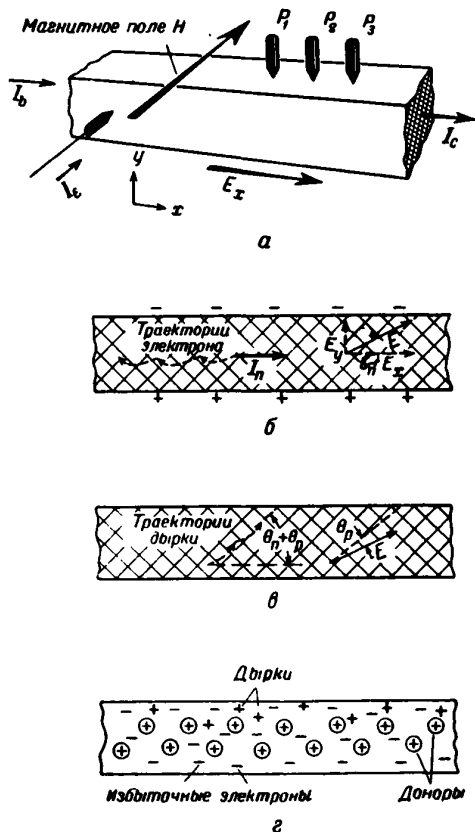
Открытый Сулем и Шокли [17] магнито-концентрационный эффект, который будет рассмотрен здесь, имеет чрезвычайно важное значение для исследования времени жизни дырок и других сходных вопросов. Этот эффект является в некотором роде распространением эффекта Холла за область его обычного применения и дает непосредственное экспериментальное подтверждение того, что электроны и дырки, двигаясь в магнитных полях, подвергаются боковому отклонению.

Опыт производится с тонким нитевидным образцом германия n -типа, помещенным в поперечное магнитное поле (фиг. 34, a). Если электрическое поле E_x создает вдоль образца поток электронов, то поперек образца возникает холловская разность потенциалов, в данном случае такой полярности, что верхняя его грань заряжается отрицательно¹⁾. В соответствии с теорией этот эффект обусловлен поперечной силой, стремящейся отклонять электроны к одной из сторон образца по мере их движения вдоль него. В нашем случае верхняя поверхность образца заряжается отрицательно, и результирующее электрическое поле, составляющие которого обозначим через E_x и E_y , образует с линиями тока угол ψ_n . Если ток I_x вводит в образец небольшое число дырок, то магнитное поле будет отклонять их в ту же сторону, что и электроны. (Как показано в § 7 гл. 8, боковое отклонение пропорционально как скорости, так и заряду частицы, знак которых для дырок противоположен их знаку для электронов, так что магнитное поле отклоняет их в одном и том же направлении.) В дополнение к этому, поперечная компонента электрического поля E_y стремится отвести электроны от верхней грани образца и, наоборот, сконцентрировать там дырки. В результате всех этих воздействий дырки устремляются в направлении к верхней поверхности образца²⁾.

¹⁾ Элементарное описание эффекта Холла дано в § 6 и 8 гл. 8. Существенные детали приведены на фиг. 80 и 82 (стр. 249 и 255). — *Прим. авт.*

²⁾ Вопрос о распределении концентраций носителей тока в образцах полупроводника при эффекте Холла рассматривается теоретически в работах [30, 31]. — *Прим. ред.*

Экспериментальное доказательство описанного здесь движения дырок было получено при помощи точечных зондов. Как следует из фиг. 31, активная проводимость участка образца между точечными

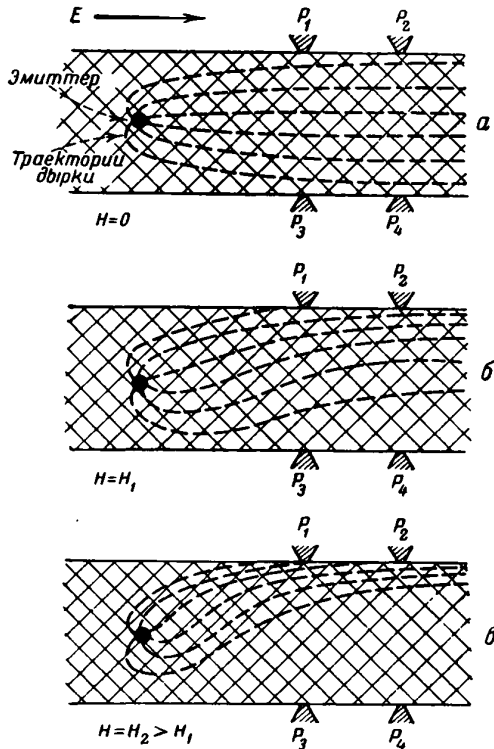


Фиг. 34. Магнитная концентрация дырок и электронов.

а—схема опыта; б—эффект Холла для электронов; в—движение дырки в поле; г—концентрация введенных дырок у поверхности.

контактами является линейной функцией концентрации дырок. Следовательно, по величине активной проводимости контактов можно определить концентрацию дырок. Явление наглядно иллюстрируется фиг. 35, на которой схематически показано, как возникает магнитоконцентрационный эффект. При отсутствии магнитного поля дырочный ток распространяется по сечению образца только благодаря диффузии; предвосхищая излагаемые ниже результаты, укажем, что

некоторая часть линий тока заканчивается на поверхности, где в основном и происходит рекомбинация электронов и дырок. При наличии магнитного поля H_1 дырки отклоняются в направлении к верхней стороне образца. При $H_2 > H_1$ этот эффект еще больше



Ф и г. 35. Траектории дырок, введенных в нитевидный образец германия *n*-типа.

Иллюстрируется эффект концентрации дырок, вызываемый магнитным полем.

выражен; кроме того, появляется второй эффект: по мере движения вправо дырочный ток начинает убывать быстрее. Это происходит потому, что дырки рекомбинируют с электронами на поверхности нитевидного образца, а возросшая концентрация дырок (в случае H_2 по сравнению с H_1) приводит к увеличению скорости рекомбинации и уменьшению времени их жизни.

При соответствующих условиях опыта эти эффекты могут быть значительными. Углы θ_n и θ_p между токами и электрическим полем

равны, как это показано в гл. 8,

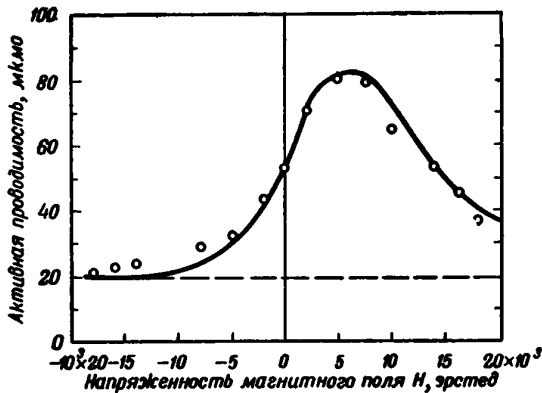
$$\begin{aligned} \theta_n &= 10^{-8} \mu_n H = 10^{-8} \cdot 2600 \cdot 10^4 = 0,26 = \\ &= 15^\circ \text{ для } 10000 \text{ эрстед,} \end{aligned} \quad (3.14)$$

$$\theta_p = 10^{-8} \cdot 1700 \cdot 10^4 = 10^\circ \text{ для } 10000 \text{ эрстед.} \quad (3.15)$$

Для продольного электрического поля E_x с напряженностью 20 в/см результирующая поперечная составляющая приближенно эквивалентна полю

$$E_x \operatorname{tg}(\theta_n + \theta_p) = 8,5 \text{ в/см,} \quad (3.16)$$

что соответствует эффективной разности потенциалов в $0,025 \cdot 8,5 = 0,21$ в для образца шириной 0,025 см. Это составляет около



Фиг. 36. Зависимость активной проводимости точечного контакта от магнитного поля.

8,5 кТ; в этом случае должна наблюдаться значительная концентрация дырок у поверхности образца. Общее расстояние, которое дырка проходит от одной стороны образца к другой, составляет $0,025/\sin 25^\circ = 0,06$ см. Поскольку скорость дрейфа составляет примерно $1700 \cdot 20 = 3,4 \cdot 10^4$ см/сек, время пролета будет меньше $2 \cdot 10^{-6}$ сек. Это значение по крайней мере на один порядок величины меньше времени жизни электронов и дырок в хорошем образце германия, так что концентрация должна была бы выравняться до того, как дырки внутри образца придут в равновесное состояние.

На фиг. 36 приведена полученная на опыте типичная зависимость активной проводимости точечного коллектора, в данном случае P_1 , от напряженности магнитного поля H . При положительных значениях H дырки отклоняются по направлению к P_1 , а при отрицательных H — в обратную сторону. Увеличение H приводит к увеличению

§ 2. Влияние магнитного поля на концентрацию дырок и электронов 101

активной проводимости коллектора, что указывает на возникновение повышенной концентрации дырок вблизи точечного контакта. Дальнейшее увеличение H , как это показано на фиг. 35, приводит соответственно к дальнейшему увеличению концентрации дырок; однако в то же время увеличивается и скорость убывания числа дырок вследствие их рекомбинации с электронами, и поэтому сигнал, приходящий к P_2 , приходит уже уменьшенным. В результате этих двух противоположных тенденций кривая, изображающая зависимость активной проводимости точечного контакта от напряженности магнитного поля, сначала достигает максимума, а затем убывает. При отрицательных значениях H дырки отклоняются в направлении от зонда, время жизни их убывает, как и в случае положительных значений H , и в результате кривая монотонно убывает.

Изображенная на фиг. 36 сплошная кривая вычислена теоретически и получена в предположении, что закон рекомбинации может быть выражен следующим образом:

$$\text{(Скорость рекомбинации дырок на единицу площади поверхности)} = \\ = (\text{Постоянной } s) \times (\text{Концентрацию дырок вблизи самой поверхности}). \quad (3.17)$$

Для постоянной s , имеющей размерность скорости, здесь принято значение 1500 см/сек . Другие методы анализа дали для подобных же образцов сравнимые результаты.

Отступление экспериментальных точек от теоретической кривой в области больших отрицательных значений H имеет свое объяснение. Практически те же значения активной проводимости контакта получаются и при отсутствии какого бы то ни было введения дырок. Скорее всего объяснение этой части кривой связано с дырками, спонтанно возникающими на поверхности образца. При отрицательных значениях H эти дырки отводятся от зонда; однако некоторая часть их все же попадает в приконтактную область зонда и изменяет ее проводимость. Для больших отрицательных значений H , по-видимому, 20 мкмо является предельным значением активной проводимости, соответствующим почти полному отсутствию дырок в образце германия.

При правильном выборе размеров образца и величины полей можно разделить явления поверхностного и объемного образования дырок. Математическое рассмотрение этого вопроса отнесено к гл. 12. Вопрос о практическом применении описанного здесь явления концентрации введенных дырок магнитным полем рассмотрен, в частности, Р. Уоллэсом.

В настоящем параграфе нам, с одной стороны, хотелось привести дополнительные экспериментальные данные относительно поведения электронов и дырок, рассматриваемого теоретически в последующих главах, а с другой стороны, ознакомить читателя с новой и мощной экспериментальной техникой, которая, как полагает автор, будет играть решающую роль при изучении транзисторных материалов.

ЗАДАЧИ¹⁾

1. В этой задаче имеется в виду показать, почему обычно не встречаются отклонения от соотношения

$$n = n_0 + p$$

[см. (3.5)]. Предположим, что в данном образце германия n -типа с диэлектрической постоянной $\epsilon_e = 16$ и проводимостью $\sigma = 0,4 \text{ ом}^{-1} \cdot \text{см}^{-1}$ значение $n - p$ отличается на 1% от n_0 в объеме сферы радиусом 10^{-8} см . Вычислить значение потенциала и величину напряженности электрического поля на поверхности сферы.

Ответ: 0,36 в; 36 в/см.

2. Предположим, что фиг. 27 относится к образцу задачи 2 гл. 1. Пусть ток $I_b = 2,9 \text{ ма}$, $I_c = 1,8 \text{ ма}$, а $\gamma = 0,5$; рекомбинацией пренебрегаем. Полагая, что $\mu_p = 1700$, $\mu_n = 3600 \text{ см}^2/\text{в} \cdot \text{сек}$ и $b = 2,1$ (эти же данные были использованы в задачах гл. 1), показать, что проводимость при этом возрастет от $0,5 \text{ ом}^{-1} \cdot \text{см}^{-1}$ до $1,23 \text{ ом}^{-1} \cdot \text{см}^{-1}$. Показать, что электрическое поле убывает от значения $0,94 \text{ в/см}$, которое оно имеет при $\gamma = 0$, до значения $0,38 \text{ в/см}$, обусловленного наличием введенных дырок, а сопротивление при этом убывает от 400 до 162 ом. Показать, что концентрация электронов возрастает с $8,7 \cdot 10^{14} \text{ см}^{-3}$ до $17,4 \cdot 10^{14} \text{ см}^{-3}$ (ср. эти результаты с задачей 5 гл. 1).

3. Рассмотрим нитевидный образец германия n -типа длиной L , поперечным сечением A и проводимостью $\sigma_0 = en_0\mu_n$. Предположим, что имеется добавочная концентрация дырок $p(x)$, так что в элементе длиной dx имеется $p(x)A dx$ дырок. Показать, что если $p(x)$ много меньше n_0 , то увеличение проводимости образца равно

$$e(\mu_n + \mu_p)L^{-2} \int_0^L Ap(x) dx = e(\mu_n + \mu_p)L^{-2} p_{\text{обм}},$$

где $p_{\text{обм}}$ — общее число добавочных дырок. Этот результат может быть получен, если сопротивление образца записать в виде

$$R = \int_0^L \frac{dx}{Ae} \{ \mu_n [n_0 + p(x)] + \mu_p p(x) \}$$

и разложить это выражение в ряд по степеням отношения $p(x)/n_0$, полагая его малой величиной.

¹⁾ Большая часть задач, основанных на материале этой главы, включает решение уравнения непрерывности. Ряд таких задач приводится в конце гл. 12. — *Прим. авт.*

Эта задача показывает, что для однородного образца и для малой концентрации дырок изменение проводимости зависит только от общего числа добавленных дырок и не зависит от их распределения.

ЛИТЕРАТУРА

1. Shockley W., Pearson G. L., Haynes J. R., Bell Syst. Techn. Journ., **28**, 344 (1949).
2. Bardeen J., Brattain W. H., Phys. Rev., **74**, 230 (1948).
3. Brattain W. H., Bardeen J., Phys. Rev., **74**, 231 (1948).
4. Bardeen J., Brattain W., Phys. Rev., **75**, 1208 (1949).
5. Shockley W., Bell Syst. Techn. Journ., **28**, 453 (1949).
6. Shive J. N., Phys. Rev., **75**, 689 (1949).
7. Kock W. E., Wallace R. L., Electr. Eng., **68**, 222 (1949).
8. Ryder E. J., Bell Lab. Rec., **27**, 89 (1949).
9. Ryder E. J., Shockley W., Phys. Rev., **75**, 310 (1949).
10. Becker J. A., Shive J. N., Electr. Eng., **68**, 215 (1949).
11. Ryder R. M., Bell Syst. Techn. Journ., **28**, 367 (1949).
12. Shockley W., Pearson G. L., Phys. Rev., **74**, 232 (1948).
13. Haynes J. R., Shockley W., Phys. Rev., **75**, 691 (1949).
14. Pearson G. L., Phys. Rev., **76**, 179 (1949).
15. Pfann W. G., Scaff J. H., Phys. Rev., **76**, 459 (1949).
16. Bray R., Phys. Rev., **76**, 152, 458 (1949).
17. Suhl H., Shockley W., Phys. Rev., **75**, 1617 (1949); **76**, 180 (1949).
18. Wannier G. H., Phys. Rev., **52**, 191 (1937).
19. Brattain W. H., Briggs H. B., Phys. Rev., **75**, 1705 (1940).
20. Briggs H. B., Phys. Rev., **77**, 287 (1950).
21. Goodman B., Lawson A. W., Schiff L. I., Phys. Rev., **71**, 191 (1947).
22. Herring C., Bell Syst. Techn. Journ., **28**, 401 (1949).
- 23*. Pfann W. G., Phys. Rev., **81**, 881 (1951).
- 24*. Френкель Я. И., ЖЭТФ, **6**, 647 (1936).
- 25*. Френкель Я. И., Phys. Rev., **37**, 17, 1276 (1931).
- 26*. Жузе В. П., Рывкин С. М., ДАН СССР, **77**, 241 (1951).
- 27*. Жузе В. П., Рывкин С. М., Изв. АН СССР, серия физич., **16**, 93 (1952).
- 28*. Лашкарев В. Е., Сборник памяти С. И. Вавилова, Изд. АН СССР, 1952, стр. 324.
- 29*. Гросс Е. Ф., Каррыев А. Н., ДАН СССР, **84**, 3 (1952).
- 30*. Ансельм А. И., ЖТФ, **22**, 1146 (1952).
- 31*. Welker H., Zs. f. Naturforsch., **6a**, 184 (1951).
- 32*. Ляшенко В. И., Павленко А. М., ЖТФ, **20**, 854 (1950).
- 33*. Пекар С. И., Изв. АН СССР, серия физич., **5**, 422 (1941).

ФИЗИЧЕСКАЯ ТЕОРИЯ ТРАНЗИСТОРОВ

Для объяснения внутреннего механизма работы транзисторов воспользуемся рассмотренными в предыдущих главах теориями введения электронов и дырок в германий. Начнем изложение с рассмотрения нитевидного транзистора. С точки зрения наглядности этот тип транзистора¹⁾ наиболее прост, так как способность усиления в этом случае определяется в основном модуляцией проводимости, рассмотренной в § 2 гл. 3. Затем перейдем к рассмотрению других типов транзисторов с выпрямляющими переходами в их выходных цепях. В § 2 вопрос о выпрямляющих переходах начинается с рассмотрения простейшего случая — выпрямления на $p-n$ -переходе. Там же затем указывается, как, комбинируя такие переходы, можно изготовить транзистор. В § 3 рассматриваются выпрямители с точечными металлическими контактами и сравниваются с выпрямляющими $p-n$ -переходами, а в § 4 теория таких выпрямляющих переходов используется для сопоставления физической картины явления с эквивалентной схемой транзистора типа А. В § 5 обсуждаются вопросы электрической *формовки* контактов транзистора, а также условия получения больших значений коэффициента усиления по току α . В § 6 кратко рассматриваются фототранзисторы и счетчики.

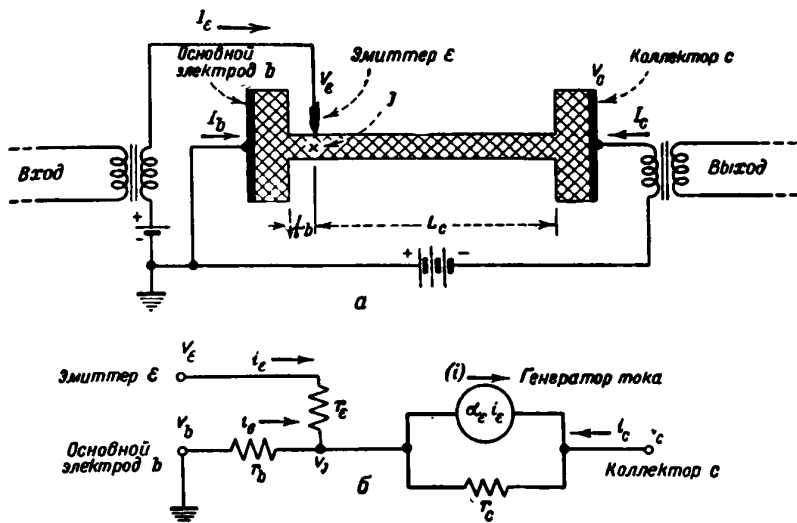
§ 1. ТЕОРИЯ НИТЕВИДНЫХ ТРАНЗИСТОРОВ

Эквивалентная схема. На фиг. 37 изображен нитевидный транзистор²⁾. Модуляция проводимости создается в этом случае тем, что дырки, введенные в германий из эмиттера, устремляются вправо и модулируют сопротивление участка выходной цепи между эмиттером

¹⁾ После открытия транзисторов из германия в зарубежной печати были опубликованы сообщения о кристаллических усилителях из PbS и PbSe с кратким указанием на наличие этого эффекта у кремния. Недавно Тённесен [44] исследовал около 300 различных веществ, главным образом минералов, и показал, что около 40 из них также обладают способностью усиливать сигнал. Наибольший эффект наблюдался у MnO_2 , Ag_2S , SnO_2 , Bi_2S_3 , $CaSi_2$, MoS_2 , ZnO и FeS_2 . — *Прим. ред.*

²⁾ Транзисторы этого типа, основанные на использовании $p-n$ -переходов, так же как и точечных контактов, в качестве эмиттеров были рассмотрены в работе Шокли, Пирсона, Спаркса и Браттейна [1]. Содержание этого параграфа заимствовано из работы Шокли, Пирсона и Хайнса [2]. — *Прим. авт.*

и коллектором. Твердые усилители такого вида можно изготовить при помощи пескоструйного аппарата, о чем уже шла речь в § 1 гл. 3. Утолщенные концы, придающие устройству гантелеобразную форму, упрощают задачу осуществления контактов с таким устройством. Большая поверхность контакта с левой стороны от эмиттера уменьшает эмиссию дырок в германий из металлического электрода и создает благоприятные условия для их быстрой рекомбинации еще до того, как они успевают дойти до узкой части транзистора.



Фиг. 37. Нитевидный транзистор и его эквивалентная схема.
а — нитевидный транзистор; б — эквивалентная схема.

Теория этого транзистора сравнительно проста, и отдельные стороны ее мы будем рассматривать в связи с соответствующими разделами теории транзистора типа А. Рассмотрим тот случай, когда ток, создаваемый введенными носителями, составляет лишь малую долю общего тока в нитевидном образце. При этом условии можно воспользоваться простой линейной теорией. Покажем, что для случая малых переменных сигналов свойства нитевидного транзистора могут быть описаны с помощью эквивалентной схемы, изображенной на фиг. 37 и построенной по тому же принципу, что и эквивалентная схема транзистора типа А. При этом будем исходить из предположения, что $V_b = 0$, что соответствует условиям работы с заземленным основным электродом. Это условие упрощает алгебраические вычисления, необходимые для расчета эквивалентной схемы, так как при этом исключается одна переменная. Сама по себе эквивалентная

схема, разумеется, применима и в тех случаях, когда основной электрод незаземлен.

Точка J на фиг. 37 соответствует некоторой точке в нитевидном образце германия, расположенной вблизи от эмиттера. Ток эмиттера будет определяться разностью между его потенциалом V_e и потенциалом окружающего полупроводника — потенциалом в точке J . Следовательно, можно написать

$$I_e = f_e(V_e - V_J). \quad (4.1)$$

При малых изменениях величин i_e , v_e и v_J это уравнение приводит к соотношению

$$i_e = (v_e - v_J) f'_e, \quad (4.2)$$

где f'_e — производная от f_e по ее аргументу. Полагая, что $f'_e = 1/r_e$, можно (4.2) переписать в виде

$$v_e - v_J = r_e i_e. \quad (4.3)$$

Этому соотношению соответствует ветвь эквивалентной схемы с сопротивлением r_e . Напряжение в точке J , в предположении указанных выше условий работы транзистора и при условии, что I_e положительно и много меньше I_c , будет равно $-I_b r_b$, где r_b — сопротивление участка нитевидного образца от основного электрода до воображаемой эквипотенциальной поверхности, проходящей через точку J . При работе транзистора в схеме с заземленным основным электродом $v_b = 0$ и

$$v_J = -r_b i_b = +r_b i_e + r_b i_c, \quad (4.4)$$

так как $i_b + i_e + i_c = 0$. Это соотношение, очевидно, выполняется для ветви эквивалентной схемы, содержащей сопротивление r_b .

Перейдем теперь к рассмотрению цепи коллектора, которую можно представить в виде сопротивления r_c и параллельно подключенного к нему генератора тока ¹⁾ $\alpha_e i_e$. Значение r_c соответствует здесь тому случаю, когда дырки не вводятся. (Эта схема эквивалентна схеме, в которой параллельно включенный генератор тока заменен генератором напряжения $\alpha_e r_c i_e$, включенным последовательно с сопротивлением r_c .) Следует показать, что эта часть эквивалентной схемы правильно изображает эффект введения дырок в правую часть нитевидного образца германия. Предположим, что рекомбинацией можно пренебречь и дырочный ток, созданный введенными из эмиттера дырками, проходит через весь образец. (Влияние рекомбинации будет рассмотрено в следующем параграфе.) Следовательно, ток I_c , протекающий в коллекторной ветви, содержит компоненту дырочного тока $-\gamma I_e = I_p$. [Знак минус появляется здесь потому, что по

¹⁾ Индекс e при коэффициенте α_e указывает на отношение к эквивалентной схеме; α_e отлично от $\alpha = -(\partial I_c / \partial I_e) v_c$ и связано с ним соотношением $\alpha_e = \alpha + (\alpha - 1)(r_b / r_c)$, эквивалентным уравнению (4.24). — Прим. авт.

условию положительный ток $I_c (= -I_b - I_e)$ течет в направлении справа налево.] Добавочная концентрация электронов и дырок уменьшает сопротивление, и r_c изменяется до значения $r_c + \delta r_c$, где δr_c — отрицательная величина. Соотношение между током и напряжением для этой ветви нитевидного усилителя принимает в таком случае вид

$$V_c - V_J = (r_c + \delta r_c) I_c. \quad (4.5)$$

Задача заключается теперь в том, чтобы преобразовать это соотношение для случая малых изменений всех параметров и показать, что оно сводится при этом к соотношению

$$v_c - v_J = r_c (i_e + \alpha_e i_b), \quad (4.6)$$

соответствующему эквивалентной схеме. При малых значениях тока эмиттера этот анализ удобно провести следующим образом. Отношение дырочного тока к общему току равно $-\gamma I_b / I_c$. Отношение $(r_c + \delta r_c) / r_c$ соответствует отношению G_0 / G , рассмотренному в связи с фиг. 27. Отношение величины дырочного тока к общему току, выраженное формулой (3.12) через величину G_0 / G , может быть переписано в виде

$$-\frac{\gamma I_b}{I_c} = \frac{1 - (G_0 / G)}{1 + b} = \frac{-\delta r_c}{(1 + b) r_c}, \quad (4.7)$$

откуда

$$\delta r_c = r_c (1 + b) \frac{\gamma I_b}{I_c}. \quad (4.8)$$

(Так как I_c отрицательно, а I_b положительно, то из этого уравнения следует, что δr_c отрицательно, так что дырочный ток увеличивает проводимость.) Подстановка этого значения $r_c + \delta r_c$ в уравнение для $V_c - V_J$ дает

$$V_c - V_J = (r_c + \delta r_c) I_c = r_c [I_c + (1 + b) \gamma I_b]. \quad (4.9)$$

В случае малых изменений токов и напряжений это уравнение сводится к уравнению эквивалентной схемы с

$$\alpha_e = (1 + b) \gamma. \quad (4.10)$$

Данные, приведенные в § 1 гл. 3¹⁾, указывают, что для дырок, введенных в германий n -типа, $\gamma = 1$, и при $b = 1,5$ получаем $\alpha_e = 2,5$.

Воспользовавшись значением $v_J = r_b (i_b + i_c)$, исключаем v_J из (4.3) и получаем соотношение между v_c и токами. Точно так же в случае малого сигнала, вид которого описывается уравнением (4.9), получаем уравнение для v_c

$$v_c = (r_e + r_b) i_b + r_b i_c, \quad (4.11a)$$

$$v_c = (r_b + \alpha_e r_e) i_b + (r_c + r_b) i_c. \quad (4.11b)$$

¹⁾ Для более точной оценки см. гл. 12. — *Прим. авт.*

Формально эти уравнения тождественны уравнениям для эквивалентных схем транзистора типа А и потому приводят к эквивалентной схеме, изображенной на фиг. 22. Следует подчеркнуть, что хотя процесс введения дырок в германий *n*-типа играет существенную роль как в транзисторе типа А, так и в нитевидном транзисторе, изображенном на фиг. 37, однако принципы действия этих транзисторов различны.

Одной из существенных особенностей транзистора типа А является высокое значение полного сопротивления выпрямляющего контакта коллектора, что, однако, не препятствует прохождению дырочного тока; другая его особенность заключается в явлении усиления тока, происходящем вблизи точечного контакта коллектора. Этими особенностями не обладает нитевидный транзистор. Высокое полное сопротивление коллекторного электрода возникает здесь вследствие малого сечения нити. Модуляция же выходного тока связана с изменением объемной проводимости германия, обусловленным добавочными дырками, т. е. эффектом, не играющим в транзисторе типа А существенной роли. В нитевидном транзисторе усиление по току создается добавочными электронами, присутствие которых необходимо для нейтрализации объемного заряда дырок. Усиление по току в транзисторе типа А, вероятно, также производится объемным зарядом дырок, однако детали механизма этого явления плохо выяснены. В § 5 этой главы будут рассмотрены теории электрической формовки контактов и причина возникновения высоких значений коэффициента усиления по току α [3].

Положительная обратная связь и нестабильность. Нитевидный транзистор характеризуется такой же обратной связью, возникающей в значительной мере по тем же причинам, что и транзистор типа А. Однако благодаря более простой структуре нитевидного транзистора вопрос об обратной связи легче рассмотреть именно на его примере.

Опишем некоторые явления, связанные с возникновением обратной связи, и попытаемся дать им физическую интерпретацию, а затем приведем ряд уравнений, характеризующих количественную сторону вопроса.

Предположим, что на коллектор подано определенное постоянное смещение V_c , скажем, — 20 в. Предположим затем, что ток эмиттера возрастает постепенно от нуля до большего значения. Как уже отмечалось в связи с анализом статических характеристик транзистора (см. фиг. 21), напряжение на эмиттере возрастает при этом до некоторого максимального значения, а затем падает. Иными словами, эмиттерная цепь обладает нелинейной характеристикой и отрицательным сопротивлением; хотя напряжение на эмиттере однозначно определяется током эмиттера, обратное утверждение несправедливо, и данному значению напряжения на эмиттере соответствует несколько

возможных значений тока эмиттера. Это обстоятельство может привести к появлению нестабильности, если на вход транзистора подано напряжение от источника с малым полным сопротивлением. При этом случайный переходный процесс может вызвать переход от слабotoчного к сильноточному режиму работы транзистора. При соответствующем подборе параметров внешней цепи с целью ограничить величину токов можно добиться, чтобы режим работы в слабotoчных и сильноточных условиях не выходил за допустимые пределы. Подобная же настройка схемы может быть использована в опрокидывающих и счетных схемах. Однако если внешние сопротивления были подобраны для получения максимального усиления (см. § 2 гл. 2), то, несмотря на устойчивость схемы в рабочей точке, эта устойчивость все же условна и большой переходный процесс легко может вызвать переход к сильноточному режиму с возможным последующим разрушением элемента.

Для объяснения этой особой зависимости V_c от I_e , показанной на статических характеристиках фиг. 21, необходимо выяснить, почему в этом случае данная характеристика нелинейна, а сопротивление падает, изменяясь от положительных до отрицательных значений. Таким образом, необходимо объяснить возникновение как отрицательного сопротивления, так и нелинейности. Отрицательное сопротивление обусловлено наличием положительной обратной связи в эквивалентной схеме, чему соответствуют значения α_e , превышающие единицу; нелинейность, в свою очередь, связана, главным образом, с меняющимся сопротивлением выпрямляющего контакта эмиттера. Рассмотрим сначала вопрос об отрицательном сопротивлении, основываясь на описанной выше физической картине процесса прохождения тока.

Свойства отрицательного сопротивления можно понять, если проследить, как изменяется V_J по мере увеличения I_e . (Для того чтобы не усложнять это объяснение математическими выкладками, предположим, что $\gamma = 1$, следовательно, дырочный ток равен I_e , и пренебрежем рекомбинацией. В этом случае $\alpha_e = (b + 1)$ и не зависит от токов.) Если бы V_c осталось постоянным, а следовательно, оставался бы постоянным и ток, протекающий через сопротивление r_c эквивалентной схемы, изображенной на фиг. 37, то действие тока I_e привело бы к увеличению I_c на величину $-\alpha_e I_e$. Таким образом, в дополнение к току введенных дырок I_e , протекающему в правой части транзистора (см. фиг. 37, а), через нее будет еще проходить добавочный электронный ток $(\alpha_e - 1)I_e$, обусловленный электронами, двигающимися налево. Добавочные электроны должны проходить через сопротивление r_b , увеличивая ток от его первоначального значения $I_e = 0$ на величину $(\alpha_e - 1)I_e$. Это требует наличия добавочного падения напряжения на r_b , равного $r_b(\alpha_e - 1)I_e$. При этом знак падения напряжения должен быть таким, чтобы V_J изменялось в отрицательном направлении. Следовательно, вхождение положительного

тока в германий в точке контакта e приводит к тому, что соседняя с ним точка не становится более положительной, как это было бы в случае положительного сопротивления, а, наоборот, ее потенциал становится вместо этого более отрицательным. Если бы при этом можно было пренебречь падением напряжения на r_e , то это означало бы, что e обладает отрицательным сопротивлением. Однако r_e есть дифференциальное сопротивление $d(V_e - V_j)/dI_e$ выпрямляющего контакта в прямом направлении. Таким образом, вначале контакт обладает большим положительным сопротивлением, убывающим по мере возрастания I_e . Для малых значений I_e положительное сопротивление r_e значительно больше отрицательного сопротивления $(\alpha_e - 1)r_b$; тем не менее при больших значениях I_e положительное сопротивление r_e может быть меньше $(\alpha_e - 1)r_b$ и входное сопротивление в точке e в связи с уже рассмотренной выше многозначностью перейдет от положительного значения к отрицательному. В этом изложении мы пренебрегаем влиянием изменения V_j на ток, протекающий через r_e . Как будет показано при анализе уравнений эквивалентной схемы, в случае $r_b \ll r_e$ это вносит незначительную поправку в расчеты.

Предположим теперь, что все параметры эквивалентной схемы, за исключением r_e , постоянны, и вычислим дифференциальное сопротивление v_e/i_e , при условии, что V_c поддерживается при постоянном смещении. Условие постоянства V_c требует, чтобы $v_c = 0$, и так как

$$v_c = (r_b + \alpha_e r_e) i_e + (r_c + r_b) i_c = 0, \quad (4.12)$$

то

$$i_c = -\frac{\alpha_e r_e + r_b}{r_c + r_b} i_e. \quad (4.13)$$

Подстановка этого выражения для i_c в уравнение (4.11а) для v_e дает

$$v_e = (r_e + r_b) i_e + r_b \left(-\frac{\alpha_e r_e + r_b}{r_c + r_b} i_e \right). \quad (4.14)$$

Это приводит к следующему выражению для дифференциального сопротивления:

$$\frac{v_e}{i_e} = r_e + r_b - \frac{\alpha_e r_b r_e + r_b^2}{r_c + r_b} = r_e - (\alpha_e - 1) r_b \frac{1}{1 + r_b/r_c}. \quad (4.15)$$

Из (4.15) следует, что входное сопротивление v_e/i_e может быть отрицательным. Если $r_b \ll r_c$, то входное сопротивление отрицательно в том случае, когда $r_e < (\alpha_e - 1)r_b$ — условие, рассмотренное уже ранее ¹⁾.

¹⁾ Точно такое же уравнение было ранее выведено Бардином и Браттейном. Полученное ими уравнение записано в виде $\alpha f' r_b = 1$, где $f' = 1/r_{11}$, так что это эквивалентно $\alpha r_b/r_{11} = 1$. Из определения α и на основании формулы для α_e , приведенной на фиг. 22, можно показать, что оба эти условия равноценны. Как указал Райдер, все эти условия эквивалентны условию $\mu\beta = 1$, если воспользоваться обозначениями теории обратной связи. — *Прим. авт.*

Явления, связанные с временем пролета. С тем фактом, что для прохождения дырок через участок нитевидного транзистора с сопротивлением r_c (см. фиг. 37, б) требуется конечный промежуток времени, связаны два важных эффекта: в течение этого времени дырки рекомбинируют с электронами, что приводит к ослаблению модуляции проводимости, а поскольку модуляция проводимости нитевидного образца германия в любой момент времени определяется током эмиттера за предыдущий интервал времени, то будет наблюдаться уменьшение модуляции, если период сигнала переменного тока сравним с временем пролета или меньше его. Как уже было показано при решении задач гл. 3, при малых значениях плотности тока введенных дырок изменение активной проводимости рабочего участка L_c нитевидного транзистора (см. фиг. 37) зависит от общего числа добавочных дырок, а не от характера их распределения вдоль образца.

Задачи, приведенные в конце этой главы, основаны на предположении, что задается полное число дырок.

В теории малых сигналов влияние времени пролета легко выразить в аналитической форме. Приведенный здесь вывод основан на предположении, что время жизни дырки до рекомбинации ее с электроном равно τ_p . В соответствии с этим предположением доля дырок, введенных в образец в некоторый момент времени t_1 и оставшихся еще не связанными к моменту времени t_2 , будет равна

$$\exp[-(t_2 - t_1)/\tau_p].$$

Это означает, что эффект изменения проводимости, наблюдаемый в нитевидном транзисторе в некоторый момент времени t_2 , является средним и обусловлен всеми дырками, вошедшими до момента времени t_2 , за отрезок времени $t_2/(t_2 - \tau_p)$, где τ_p — время пролета; дырки, вошедшие до момента $t_2 - \tau_p$, выходят из образца к моменту времени t_2 .

Если выразить ток эмиттера как $i_{e0}e^{i\omega t}$, то среднее действующее значение этого тока будет равно

$$i_{e\text{эфф.}}(t_2) = i_{e0} \int_{t_2 - \tau_p}^{t_2} \exp\left[i\omega t_1 - \frac{t_2 - t_1}{\tau_p}\right] \frac{dt_1}{\tau_p}. \quad (4.16)$$

Множитель dt_1/τ_p выбран таким образом, чтобы получилось истинное среднее значение, так как сумма всех интервалов времени dt_1 равна τ_p . Этот интеграл легко вычислить:

$$i_{e\text{эфф.}}(t_2) = i_{e0} e^{i\omega t_2} \frac{1 - \exp[-i\omega\tau_p - (\tau_p/\tau_p)]}{i\omega\tau_p + (\tau_p/\tau_p)}. \quad (4.17)$$

Результат, получаемый для эквивалентной схемы, таков, как если бы мы положили ¹⁾

$$\alpha_e = \gamma(1 + b)\beta, \quad (4.18)$$

где

$$\beta = \frac{1 - \exp[-i\omega\tau_t - (\tau_t/\tau_p)]}{i\omega\tau_t + (\tau_t/\tau_p)}. \quad (4.19)$$

Величина β выражает влияние рекомбинации и угла пролета, $\omega\tau_t$, на уменьшение коэффициента усиления.

Рассмотрим это выражение для двух предельных случаев. Во-первых, если $\omega\tau_t$ очень мало, то новый множитель принимает вид

$$\beta = \frac{\tau_p}{\tau_t} (1 - e^{-\tau_t/\tau_p}). \quad (4.20)$$

Если τ_t значительно больше τ_p , так что дырки успевают рекомбинировать еще до того, как проходят через образец, то экспоненциальным членом можно пренебречь и β становится равным просто τ_p/τ_t . Это означает, что эффективность действия дырок уменьшается в число раз, равное отношению эффективной длины перемещения дырки за время ее жизни к полной длине образца, т. е. в τ_p/τ_t раз. Существенно при этом отметить, что дырки модулируют проводимость только той части образца, в которую они проникают. Время пролета зависит от величины электрического поля, созданного в образце и равного $|V_c - V_J|/L_c$; абсолютное значение здесь взято потому, что V_c отрицательно. Время пролета, таким образом, равно

$$\tau_t = \frac{L_c}{\mu_p |V_c - V_J|/L_c} = \frac{L_c^2}{\mu_p |V_c - V_J|}. \quad (4.21)$$

Для очень малых токов эмиттера $V_c - V_J = r_c V_c / (r_c + r_b)$, так что

$$\tau_t = \frac{L_c^2 (r_c + r_b)}{\mu_p r_c |V_c|}. \quad (4.22)$$

и τ_p обратно пропорционально V_c . При больших значениях V_c τ_t стремится к нулю, а β — к единице. Зависимость β от V_c была исследована путем измерения α и построения графика зависимости α от $1/|V_c|$, как это показано на фиг. 38. Значение коэффициента усиления по току

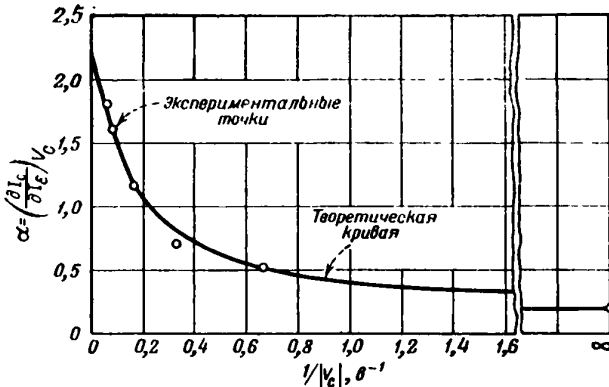
$$\alpha = - \left(\frac{\partial I_c}{\partial I_e} \right)_{V_c} = - \left(\frac{i_c}{i_e} \right)_{V_c} \quad (4.23)$$

¹⁾ Вывод формул (4.11а) и (4.11б), описывающих эквивалентную схему, показывает, что влияние введенных дырок учитывается только членом $\delta r_c i_c$ в формуле (4.9). Это приводит к появлению в формуле (4.11б) выражения $\alpha_e r_c i_c = (1 + b) \gamma r_c i_c$, которое должно быть заменено выражением $(1 + b) \times \gamma r_c i_c^{\text{эфф.}} = (1 + b) \gamma \beta r_c i_c$, приводящим к равенству (4.18). — *Прим. авт.*

легко определить из анализа эквивалентной схемы, если воспользоваться для этого уравнением (4.116). При этом

$$\alpha = \frac{r_b}{r_b + r_c} + \frac{\alpha_e r_c}{r_b + r_c}. \quad (4.24)$$

Для одного исследованного транзистора отношение r_b/r_c при $I_b = 0$ оказалось равным 0,25. Величина коэффициента α , полученная путем



Фиг. 38. Зависимость α от $1/|V_c|$.

экстраполяции данных к $|V_c| = \infty$, оказалась равной 2,2; значение α , рассчитанное для этого случая по формуле, когда $\beta = 1$ и $b = 1,5$, равно

$$\alpha = 0,2 + 0,8 \cdot 2,5 \cdot \gamma, \quad (4.25)$$

откуда следует, что $\gamma = 1,0$; это находится в согласии с результатами анализа графика, показанного на фиг. 29, из которого следует, что почти весь эмиттерный ток переносится дырками¹⁾. Уравнение изображенной на графике теоретической кривой имеет вид

$$\alpha = 0,2 + 0,8 \cdot 2,5 \cdot \left| \frac{V_a}{10} \right| (1 - e^{10/V_a}), \quad (4.26)$$

что соответствует

$$\frac{\tau_t}{\tau_p} = \frac{10}{|V_c|} = \frac{L_a^2 (r_c + r_b)}{\tau_p^2 p r_c |V_c|}. \quad (4.27)$$

На этом основании было сделано заключение, что в исследованном транзисторе время жизни дырки τ_p составляло 0,2 мксек.

¹⁾ Этот результат получился потому, что мы для b приняли значение 1,5, вычисленное на основании подвижности Холта (см. § 9 гл. 12). — Прим. авт.

Если τ_z значительно меньше τ_p , то дырки проходят через весь образец и β принимает вид

$$\beta = \frac{1 - \exp(-i\omega\tau_z)}{i\omega\tau_z} = \frac{\exp(-i\omega\tau_z/2) \sin(\omega\tau_z/2)}{(\omega\tau_z/2)}. \quad (4.28)$$

Для малых значений $\omega\tau_z$ значение β приближается к единице, поскольку $(\sin x)/x$ стремится к единице при x , стремящемся к нулю. При $\omega\tau_z/2 = \pi$ имеем $\beta = 0$. Это и есть условие, согласно которому $\tau_z = 2\pi/\omega = 1/f$. В этом случае длина нитевидного транзистора как раз такова, что модуляция усредняется по времени одного периода входного сигнала и, поскольку это среднее значение включает все фазы, модуляция проводимости исчезает.

Предварительные опыты с нитевидными транзисторами, проведенные в соответствии с изложенными выше принципами, повидимому, подтверждают основные положения теории [2]. В этих опытах было достигнуто усиление в 15 дб, а частотная характеристика коэффициента усиления по току α спадает на 3 дб при частоте 10^6 гц. Измерения уровня шумов показали, что при сравнимых условиях изготовления транзисторы этого типа дают усиление на 10—15 дб больше, чем средний транзистор типа А.

§ 2. ПЕРЕХОДЫ ТИПА $p-n$ И ТРАНЗИСТОРЫ С $p-n$ -ПЕРЕХОДАМИ

Природа тока в $p-n$ -переходах¹⁾. Физическая теория выпрямляющих контактов в транзисторах значительно более проста для случая $p-n$ -переходов, чем для точечных металлических контактов, и поэтому последние рассматриваются в § 3 настоящей главы. Переходы $p-n$ -типа могут быть использованы в качестве коллекторов и эмиттеров, при этом можно показать, как должно быть устроены эмиттерные переходы, чтобы обеспечить высокую эффективность введения в транзистор носителей заряда того знака, которые обычно не присутствуют в данном веществе в значительных количествах.

Переходы $p-n$ -типа, которые будут здесь рассмотрены, встречаются в тех образцах германия или кремния, в которых имеется переменная концентрация доноров и акцепторов и наблюдается непрерывный переход от области с проводимостью p -типа к области с проводи-

¹⁾ Явления на границе между p - и n -полупроводниками исследованы Олем. Работа по изучению $p-n$ -переходов в германии была опубликована группой сотрудников университета в Пардю под руководством Ларк-Горовитца [4—6]. Как сообщается в работах Сосновского [7—9], такие же переходы встречаются в сернистом свинце. Излагаемая здесь теория рассматривалась Гаучером [10, 11] в связи с фотоэлектрическими явлениями в $p-n$ -переходах. Общий обзор исследований по проводимости кремния p - и n -типов приведен в работах [12—14]. В двух последних работах рассматривается также вопрос о фото-э. д. с. барьерах. — *Прим. авт.*

мостью n -типа. Если просто привести в соприкосновение два отдельных куска германия, обладающих проводимостью противоположного типа, то присутствие на поверхности окислов или пленок посторонних веществ, шероховатостей, наличие поверхностных состояний и т. п. препятствует, вообще говоря, образованию истинного $p-n$ -перехода¹⁾.

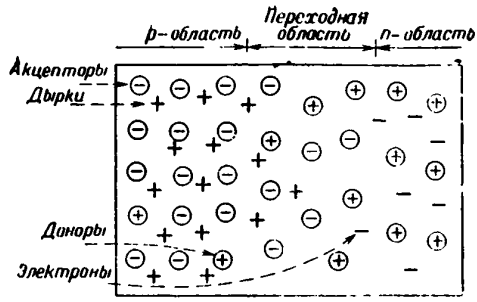
Хорошие $p-n$ -переходы могут быть получены различными способами. Естественным путем они образуются в сплавах относительно чистого кремния благодаря сегрегации примесей при затвердевании [13, 14]. Такие переходы образуются также в германии n -типа, если бомбардировкой нуклонами или нагревом одну часть его превратить в германий p -типа [5, 11]²⁾.

На фиг. 39 изображена схема идеализированного $p-n$ -перехода. На фиг. 39, *a* показано распределение доноров, акцепторов, дырок и электронов, на фиг. 39, *б* — распределение концентраций доноров и акцепторов. Распределение концентраций дырок и электронов, показанное на фиг. 39, *в*, устанавливается само по себе, при условии теплового равновесия, таким образом, что ток электронов или дырок отсутствует, а число рекомбинирующих дырок и электронов в любом малом элементе объема как раз уравнивается числом электронов и дырок, освобождаемых тепловым движением. Так как концентрация дырок больше в области с проводимостью p -типа, то они стремятся диффундировать в область с проводимостью n -типа. Этому стремлению препятствует электростатическое поле; ход электростатического потенциала вдоль образца изображен на фиг. 39, *д*; это поле делает электростатический потенциал в веществе с проводимостью n -типа более положительным, так что дырки уже не имеют возможности проникнуть в n -область, а электроны покинуть ее. Аналогичная ситуация имеет место и для электронов. На фиг. 39, *г* изображено результирующее распределение плотности заряда. (Здесь пренебрегается захватом электронов донорами или дырок — акцепторами. Доказательство справедливости такого предположения для германия приведено в гл. 10.) Такое распределение плотности заряда создает дипольный (двойной) слой, что в свою очередь приводит к возникновению разности потенциалов (см. фиг. 39, *д*). Отыскание распределения потенциала в условиях равновесия в принципе сводится к непосредственному применению теории равновесия, изложенной в гл. 10 и 12³⁾. (Возникновение скачка потенциала является неизбежным следствием требования, чтобы положение уровня Ферми оставалось постоянным для всего образца.)

¹⁾ Обзор по данному вопросу см. [15]. — *Прим. авт.*

²⁾ Более новые данные по этому вопросу приведены в работах [45] и [46]. — *Прим. ред.*

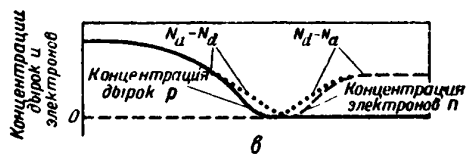
³⁾ Подробности математических выводов см. в работе [16]. — *Прим. авт.*



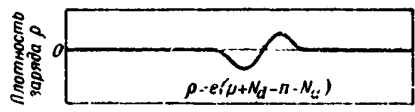
а



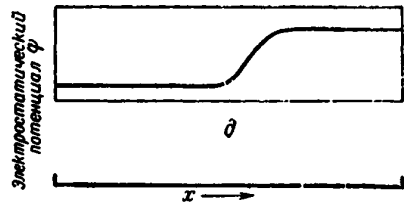
б



в



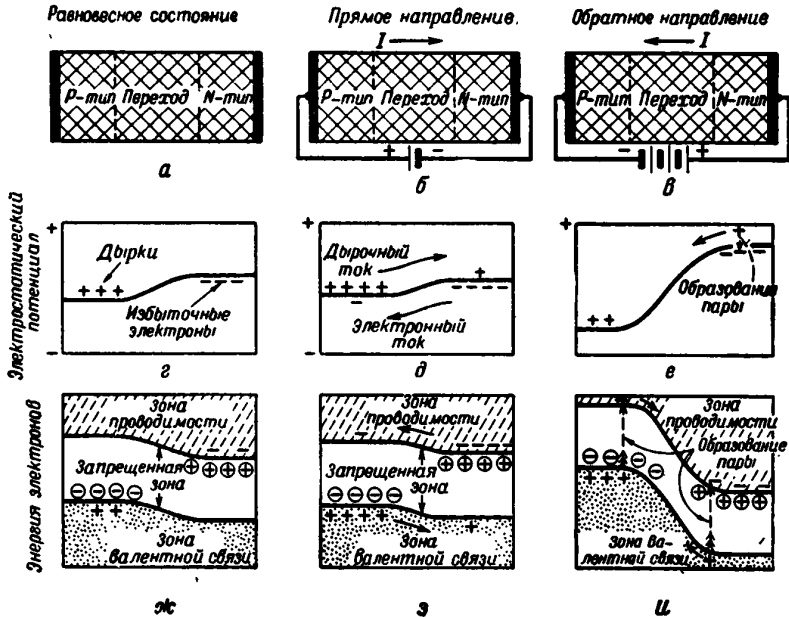
г



д

Фиг. 39. Схема идеализированного *p-n*-перехода.

Рассмотрим теперь вопрос о природе тока через $p-n$ -переход, когда к нему приложена некоторая разность потенциалов. Интересные для нас случаи изображены на рисунках в двух верхних рядах фиг. 40. О нижнем ряде рисунков речь будет идти позже. В равновесном состоянии часть дырок, указанных на фиг. 40, z , благодаря тепловому



Фиг. 40. Распределение дырок и электронов, а также ход энергии в зависимости от положения в $p-n$ -переходе при наличии смещения.

возбуждению будет приобретать достаточную энергию, чтобы преодолеть скачок потенциала и продиффундировать в n -область. В этой области часть их будет рекомбинировать с электронами. Этот дырочный ток точно уравнивается (в соответствии с принципом детального равновесия, см. гл. 11) током дырок, освобожденных тепловым движением в n -области (в результате образования пар дырка — электрон), диффундирующих через переходную область в противоположном направлении и попадающих в p -область. Аналогичным образом текут и электронные токи и точно так же внезапно компенсируют друг друга.

При подаче на $p-n$ -переход смещения в *обратном направлении* к p -области прикладывается отрицательный потенциал, что приводит к увеличению высоты потенциального барьера (см. фиг. 40, e),

При смещении в несколько десятых вольта или больше высота потенциального барьера уже так велика, что лишь незначительное количество дырок обладает достаточной тепловой энергией, чтобы его преодолеть, и, по существу, поток дырок в n -область прекращается. Напротив, поток дырок из n -области практически остается таким же, как и в случае, показанном на фиг. 40, *г*. На фиг. 40, *е* представлено возникновение этого тока как следствие термического возбуждения пары электрон — дырка с последующей диффузией дырки в переходную область и дальнейшим ее переходом в p -область. Этот дырочный ток, как уже отмечалось выше, почти не зависит от приложенного напряжения.

По той же причине обратный дырочный ток стремится к насыщению, т. е. достигает предельного значения по мере возрастания обратного напряжения. Совершенно аналогичные рассуждения применимы и к потоку электронов, так как вполне очевидно, что тот же потенциальный барьер, который удерживает дырки в p -области, удерживает электроны в n -области; в этом смысле обратное напряжение является обратным как по отношению к дырочному, так и по отношению к электронному току.

В прямом направлении высота потенциального барьера уменьшается, и поэтому через него в n -область проходит сравнительно большой дырочный ток. Часть этих дырок диффундирует в обратном направлении; доля таких дырок определяется временем их жизни в n -области.

На фиг. 40 изображены также схемы энергетических зон в применении к p — n -переходу. Хотя зонная теория по существу излагается лишь в конце гл. 5, основные ее положения, необходимые нам в связи с рассмотрением фиг. 40, достаточно просты и могут быть разъяснены непосредственно с помощью этой фигуры. Как уже упоминалось, электрон может быть удален из валентной полосы и может начать двигаться в зоне проводимости. Такой электрон находится в возбужденном состоянии (подобно автомобилю на втором этаже гаража), и энергия его выше, чем в валентной полосе. Этим уровням зоны проводимости соответствует заштрихованная область в нижнем ряду рисунков фиг. 40. Подобная же ситуация имеет место и для уровней валентной полосы. Эти более низкие уровни соответствуют дыркам, являющимся просто уровнями, не занятыми электронами. Доноры и акцепторы изображены в виде положительных и отрицательных зарядов, размещенных на энергетических уровнях в соответствии с их способностью связывать электроны и дырки. В данном случае, исходя из задачи настоящего параграфа, мы предполагаем, что этими связями можно пренебречь, и энергию этих связей учитывать не будем. (Дальнейшее обсуждение этого вопроса см. в гл. 9.) Стремление электронов занять наиболее низкий энергетический уровень эквивалентно для дырок переходу на самый высокий электронный уровень энергии. Так как с точки зрения квантовой механики получить полувозбужден-

ный электрон невозможно (т. е. в гараже нет наклонного въезда на второй этаж), то между состояниями в зоне валентной связи и проводящими состояниями существует зона запрещенных энергетических уровней. Образование пары электрон—дырка в результате теплового возбуждения соответствует переходу электрона через запретную зону, как это показано на фиг. 40, *и*. При увеличении кинетической энергии электрон переходит на более высокие энергетические уровни в зоне проводимости, а дырка—на более низкие уровни энергии в зоне валентных связей. Эти вопросы рассматриваются на основе квантовой теории энергетических зон в конце гл. 7 (см., например, фиг. 73, стр. 226).

Результирующий дырочный ток через $p-n$ -переход может быть вычислен для модели такого перехода, изображенной на фиг. 40. Этот ток выражается хорошо известным в теории выпрямления уравнением¹⁾

$$I_p = I_{ps} \left[\exp\left(\frac{eV}{kT}\right) - 1 \right], \quad (4.29)$$

где V —напряжение (положительные значения V соответствуют прямому направлению), приложенное непосредственно к $p-n$ -переходу, e —заряд электрона, kT —энергия теплового движения. Если V выражено в вольтах, а $T = 300^\circ\text{K}$, что соответствует комнатной температуре, то это уравнение принимает вид

$$I_p = I_{ps} [\exp(39V) - 1] = I_{ps} (10^{17V} - 1). \quad (4.30)$$

Значение I_{ps} определяется только веществом n -типа и не зависит от вещества с проводимостью p -типа. Выражение для I_{ps} можно записать следующим образом:

$$I_{ps} = \frac{e_n p_n D_p}{\sqrt{D_p \tau_p}}, \quad (4.31)$$

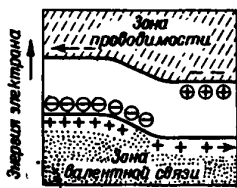
где p_n —концентрация дырок в n -области, τ_p —время жизни дырок в n -области, а μ_p и D_p —подвижность и коэффициент диффузии дырок ($D_p = \mu_p kT/e \approx 1700/39 = 44 \text{ см}^2/\text{сек}$, как это показано в гл. 12).

Для полупроводника в равновесном состоянии закон действующих масс применим к произведению концентраций дырок и электронов [см. уравнение (10.19)]. Следовательно, если вещество с проводимостью n -типа характеризуется большой концентрацией электронов, то концентрация дырок в нем мала и малым будет также значение I_{ps} . Уравнения того же вида применимы и к потоку электронов. Исходя из приведенных выше соображений, легко убедиться в том, что если

¹⁾ См. [16]. Это уравнение впервые опубликовано Вагнером [17] для другой модели и представляет собой выражение максимальной нелинейности, достижимой в случае чисто электронного выпрямления.—Прим. авт.

концентрация дырок в p -области значительно больше концентрации электронов в n -области, то I_{ns} будет меньше I_{ps} и ток через p — n -переход будет преимущественно дырочным.

Существует ряд способов, с помощью которых можно получить только что приведенный результат, и полезно обсудить один из них. На фиг. 41 изображен переход, в котором концентрация дырок в p -области превышает концентрацию электронов в n -области. При смещениях, приложенных в прямом направлении, как электроны, так и дырки должны преодолеть один и тот же потенциальный барьер.



Фиг. 41. Асимметричный p — n -переход.

Схема показывает, каким образом перенос тока (в данном случае прямого) осуществляется в основном преобладающими носителями (дырками).

Если дырок значительно больше, чем электронов, то они в большем числе будут проходить через барьер, причем основная часть тока будет переноситься дырками. Такой p — n -переход будет, очевидно, хорошим эмиттером дырок.

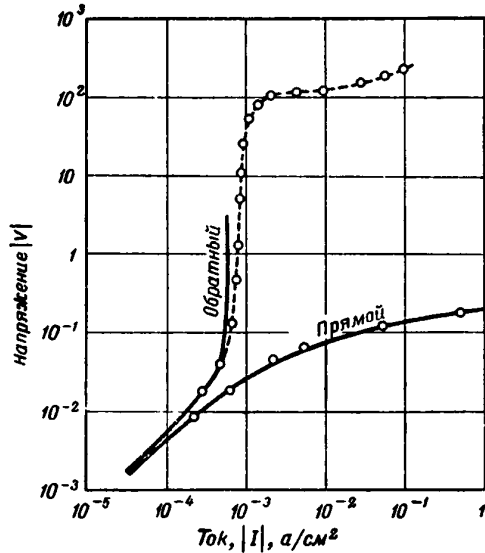
На фиг. 42 сопоставлена теоретическая вольтамперная характеристика p — n -перехода с экспериментальными данными¹⁾. Соответствие, как видно, вполне хорошее, за исключением области больших обратных напряжений, где уже сказываются вторичные эффекты, напоминающие автоэлектронную эмиссию. Следует заметить, что отклонение величины e/kT на 10—20% от ее теоретического значения, равного $39 e^{-1}$, существенно ухудшило бы совпадение данных. Следовательно, совпадение экспериментальных данных с теоретической кривой является дополнительным подтверждением того, что эффективный заряд носителей тока равен e , т. е. заряду свободного электрона.

Транзистор с p — n -переходами²⁾. На фиг. 43 (стр. 122 и 123) схематически изображена идеализированная структура транзистора с двумя p — n -переходами, отделяющими две области с проводимостью p -типа, P_1 и P_2 , от области с проводимостью n -типа, N_b . По сравнению с транзистором типа А описание такого транзистора обладает тем преимуществом, что в этом случае можно наглядно представить себе картину распределения всех участвующих в его работе атомов. Можно провести также весьма поучительную аналогию между транзистором этого типа и вакуумным триодом, изображенным также на фиг. 43.

¹⁾ В работе Эгрейна и др. [47] приводится метод расчета статических характеристик p — n -переходов, а также характеристик p — n — p -транзисторов. — *Прим. ред.*

²⁾ Дополнительная литература по транзисторам с p — n -переходами приведена в Библиографии (стр. 633). См. также статьи Шокли с сотрудниками и Мортон, которые даются в виде приложения к книге. — *Прим. ред.*

На фиг. 43, ж изображено равновесное распределение концентрации дырок и электронов; на фиг. 43, в изображен соответствующий этому случаю ход электростатического потенциала. Дырки, как это схематически показано на диаграмме, стремятся занять положения с самым низким электрическим потенциалом; электроны, заряд которых отрицателен, занимают места с наивысшим потенциалом. Схема энергетических зон изображена на фиг. 43, д. На фиг. 43, б показано, что

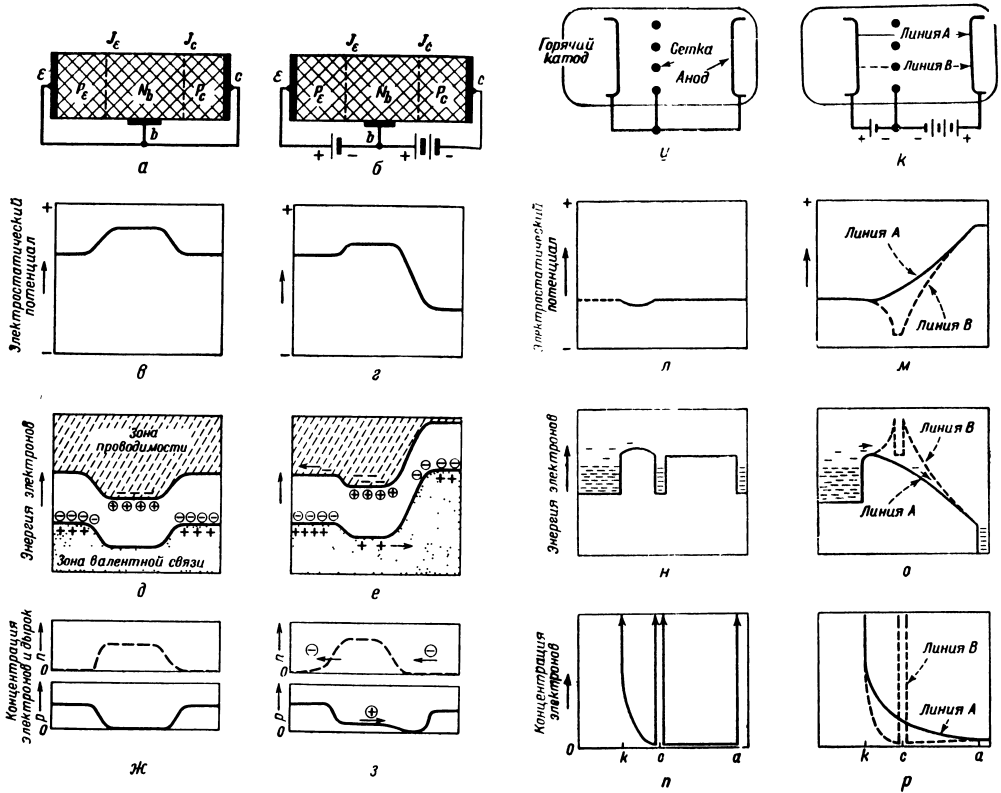


Фиг. 42. Теоретическая кривая выпрямления и экспериментальные данные для $p-n$ -перехода.

— $I = I_s \left[\exp \left(\frac{eV}{kT} \right) - 1 \right]$; ---○--- — экспериментальные данные.

к переходу J_c приложено напряжение в обратном направлении, а к переходу J_a — небольшое напряжение в прямом направлении. Если области с проводимостью p -типа обладают более высокой проводимостью, чем область с проводимостью n -типа, то большая часть тока, протекающего через $p-n$ -переходы, будет переноситься дырками. Более того, если область вещества с проводимостью n -типа, N_b , настолько узка, что дырки могут диффундировать через нее с малой вероятностью рекомбинации, то дырочная составляющая тока, проходящего через переход J_a , будет также проходить и через переход J_c ¹⁾.

¹⁾ Из теории следует, что ничтожно малое число дырок будет создаваться в области перехода J_c с большим градиентом поля при напряжениях, соответствующих удовлетворительному режиму работы. — *Прим. авт.*



Фиг. 43. Сопоставление $p-n-p$ -

транзистора с ламповым триодом.

Если электронные токи, текущие через выпрямляющие переходы, незначительны по своей величине и рекомбинацией электронов и дырок в области N_p можно пренебречь, то в таком транзисторе на основной электрод ток не пойдет. При заземленном эмиттере приложенные к основному электроду переменные напряжения будут вызывать ничтожные токи (по крайней мере, при низких частотах), и свойства такого транзистора будут подобны свойствам катодной лампы, изображенной на фиг. 43, з, за исключением того, что в транзисторе подвижны и положительные носители заряда, а не только отрицательные.

Интересно, кроме того, сопоставить механизм переноса тока электронами в транзисторе и катодной лампе. Действительно, *в обоих случаях ток может переноситься электронами двумя существенно раздельными способами, и именно управление одним током при помощи другого дает возможность получить усиление*. Рассмотрим сначала эти способы переноса тока на примере катодной лампы, а затем сравним их с дырочными и электронными процессами в транзисторе.

В катодной лампе, как это показано на фиг. 43, *n* и *p*, концентрация электронов велика внутри катода, сетки и анода. Значительно меньшая, но все же высокая концентрация испускаемых в процессе термоионной эмиссии электронов имеется вблизи самой поверхности катода. Когда к лампе приложены рабочие напряжения, то ход потенциальной энергии электрона приобретает такой вид, как это изображено на фиг. 43, *o*. У самой сетки потенциальная энергия высока, так как сетка отрицательно заряжена по отношению к катоду. В промежутках между проволочками сетки потенциальная энергия сравнительно низка и электроны проходят через эти области между пиками энергии. Если отводить электроны от проволочек сетки, что уменьшает ее отрицательный заряд, то потенциальная энергия электронов в области сетки понизится и ток в результате этого возрастет. Таким образом, в управляемой области один электронный ток регулируется другим: электроны, входящие в сетку и выходящие из нее, управляют потоком электронов лампы. Эти два тока не смешиваются вследствие того что, во-первых, сетка отрицательна по отношению к катоду, так что на нее не могут попасть электроны, созданные при термоионной эмиссии с катода, и, во-вторых, электроны не могут покинуть сетку ввиду большого значения работы выхода металла, из которого она изготовлена¹⁾.

Напряжение, приложенное между эмиттером и основным электродом в транзисторе, вызывает такой же эффект, как и напряжение между сеткой и катодом в электронной лампе. Переходная область транзистора, соответствующая промежутку катод — сетка в лампе,

¹⁾ Если сетка перегрета, возникает термоионная эмиссия, однако этот неблагоприятный случай нас не интересует. — *Прим. авт.*

обладает некоторой емкостью аналогично емкости сетка — катод. Если, например, к области N_b приложен отрицательный потенциал, то входящий в нее поток электронов будет заряжать эту емкость (а также емкость коллекторного перехода), что приведет к снижению высоты потенциального барьера, через который должны пройти дырки, чтобы достичь коллекторного перехода. Следовательно, в транзисторе имеется такая область, в которой поток избыточных электронов управляется потоком дырок. В этом случае, однако, всегда будет иметь место некоторая рекомбинация, и разделение тока в той степени, в какой это возможно в катодной лампе, затруднительно.

Очевидно, что для характеристики транзистора с $p-n-p$ -переходами, точно так же как и для нитевидного транзистора, можно ввести в рассмотрение коэффициенты γ и β . Так как через переход J_c проходит ток насыщения, то следует ожидать, что изменение коллекторного тока будет просто в $\alpha = \beta\gamma$ раз больше изменения эмиттерного тока. Сопротивления переходов J_e и J_c заменят здесь сопротивления r_e и r_c предыдущего параграфа. Величина r_b носит в данном случае более сложный характер: она представляет собой среднее значение сопротивления участка транзистора, расположенного между контактом из материала n -типа и переходом J_1 .

Из приближенной формулы для усиления по мощности $\alpha^2 r_c / r_e$ видно, что поскольку r_c значительно больше r_e , в рассматриваемом случае будет иметь место большое усиление по мощности, при условии, что большая часть дырок сможет продиффундировать через область вещества с проводимостью n -типа, т. е. что коэффициент α не будет слишком малым. Если значение α близко к единице, то переменный ток основного электрода будет очень малым. В этом случае транзистор может хорошо работать подобно электронной лампе в схеме с заземленным катодом; эмиттер при этом заземляется, а входной сигнал подается на основной электрод. В § 5 будут рассмотрены некоторые видоизменения транзистора с $p-n-p$ -переходами, являющимися коллекторами с большими значениями коэффициента α .

§ 3. О ПРИРОДЕ КОНТАКТА МЕЖДУ МЕТАЛЛОМ И ПОЛУПРОВОДНИКОМ

В транзисторах типа А и в коаксиальных транзисторах в качестве эмиттеров и коллекторов применяются металлические точечные контакты. Через эти контакты, как и в описанных выше $p-n$ -переходах, ток, в зависимости от условий, может переноситься либо дырками, либо электронами. В этом параграфе мы попытаемся в элементарной форме описать явления, происходящие на границе между металлом и полупроводником. Для детального рассмотрения этих вопросов необходимо было бы воспользоваться статистикой

Ферми — Дирака и схемой энергетических зон, а поэтому мы ограничимся здесь лишь поверхностным описанием; более подробное изложение этих вопросов читатель может найти в гл. 12 ¹⁾).

Теория выпрямления в германии отличается от обычной теории выпрямления тем, что в этом случае необходимо принимать во внимание как дырочный, так и электронный токи. В этом параграфе мы в основном будем касаться вопроса эмиссии дырок в германии *n*-типа при токе в прямом направлении и приведем некоторые предварительные соображения относительно механизма усиления по току на контакте коллектора. К этим двум вопросам мы еще вернемся в § 5 при обсуждении способов получения больших значений коэффициента α и теории формовки.

Термин *эмиттер* впервые был введен Бардином и Браттейном, чтобы подчеркнуть ту роль, которую играет этот контакт в создании дырочного тока коллектора; в то время казалось, что главную роль при этом играет прохождение дырочного тока в поверхностном слое.

Термин *введение дырок* был использован для описания процесса проникновения дырок в объем германия. Предположение о введении дырок было сделано в связи с созданием теории *p* — *n*-переходов и *p* — *n* — *p*-транзисторов; независимо от этого явление введения дырок в германий наблюдал Шайв при исследовании транзистора с двойной поверхностью ²⁾.

Для того чтобы объяснить механизм введения дырок в германий, когда эмиттером является точечный металлический контакт, необходимо рассмотреть, какую роль в этих процессах играют поверхностные уровни, уже упоминавшиеся в связи с рассмотрением фиг. 18. Опишем прежде всего то, что получается, когда точечный контакт постепенно приближается к поверхности, а затем касается ее. Для упрощения предположим, что точечный контакт электрически связан с полупроводником. В этом случае между точечным металлическим контактом и полупроводником благодаря наличию контактной разности потенциалов будет существовать электрическое поле. По мере сближения металла с полупроводником это поле становится более интенсивным, индуцированный на поверхности полупроводника заряд возникает главным образом в результате изменения количества электронов, находящихся в поверхностных состояниях. Электроны, находящиеся в этих поверхностных состояниях, столь сильно стабилизируют потенциал поверхности, что поле контактной разности потенциалов *лишь незначительно изменяет распределение потенциала внутри полупроводника даже в том случае, когда металл находится от полупроводника на расстоянии одного-двух атомных размеров*. Кроме этого, весьма

¹⁾ См. также [18]. — *Прим. авт.*

²⁾ Некоторые исторические справки по данному вопросу приведены в § 1 гл. 3 и в работе Бардина и Браттейна [3]. — *Прим. авт.*

возможно, что поверхность германия покрыта тонким (одно- или двухатомным) слоем окиси или адсорбированных газов, и поэтому даже после того, как металл приведен в механический контакт с полупроводником, потенциал внутри полупроводника даже вблизи самой поверхности остается почти таким же, каким он был до того, как эти поверхности пришли в соприкосновение. Больше того, выпрямляющие свойства контактов определяются как раз потенциалом вблизи поверхности полупроводника, и приведенные здесь соображения показывают, что этот потенциал мало меняется при соприкосновении металла с поверхностью полупроводника¹⁾. Этим обстоятельством воспользовался Бардин, чтобы объяснить, почему выпрямляющие свойства германия не зависят от работы выхода металла, применявшегося в качестве точечного контакта²⁾.

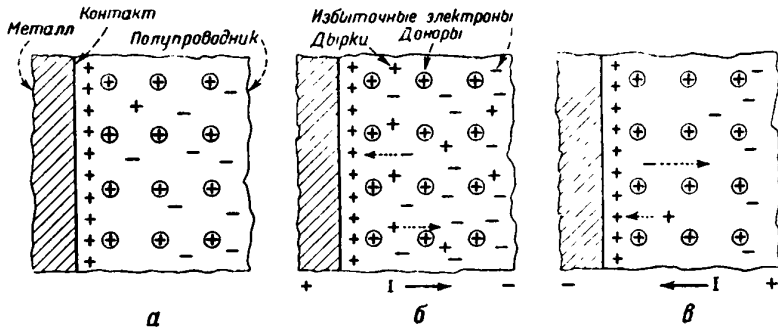
Хотя потенциал поверхности полупроводника не зависит от работы выхода металла, находящегося с ним в контакте, однако *приложение некоторого напряжения к точечному контакту приводит к такому же изменению потенциала внутри полупроводника*. Это кажущееся противоречие объясняется следующим образом: в условиях теплового равновесия концентрация электронов в поверхностных состояниях благодаря обмену с электронами, находящимися в объеме полупроводника, и электронами в металле устанавливается таким образом, что на поверхности полупроводника создается вполне определенный потенциал; однако окончательный результат, как в случае любых равновесных состояний, не должен зависеть от механизма установления равновесия.

С другой стороны, когда к металлическому контакту прикладывается напряжение, состояние перестает быть равновесным. Таким образом, в поверхностных состояниях должен установиться такой заряд, который находился бы в равновесии, с одной стороны, с электронами металла и, с другой — с электронами в объеме полупроводника. Если контакта с металлом нет, то невозможен и обмен электронами с ним. В этом случае преобладает состояние, изображенное на фиг. 18, и поверхность полупроводника находится в равновесии с его объемом при определенном потенциале относительно полупроводника. Однако когда металл находится в контакте с полупроводником, поверхностные состояния очень легко обмениваются электронами с металлом и имеют тот же потенциал относительно металла, который у них был бы в условиях равновесия. Таким образом, поверхность полупроводника приобретает по отношению к металлу такой потенциал, который не зависит от работы выхода металла; эта разность потенциалов остается фиксированной, так что потенциал поверхности полупроводника следует за изменениями потенциала, приложенного к металлу.

¹⁾ Эти вопросы рассмотрены в работе Бардина [15]. — *Прим. авт.*

²⁾ Дополнительные данные о выпрямляющих свойствах германия приведены в работе Бенцера [19]. — *Прим. авт.*

Теория введения дырок из эмиттера, предложенная Бардином и Браттейном, основывается на представлении о важной роли поверхностных состояний. Согласно этим представлениям, в поверхностных состояниях обычно находится большое число электронов. Это приводит к тому, что электроны отталкиваются от поверхности, как показано на схеме поверхностных состояний, приведенной на фиг. 18.



Фиг. 44. Схема точечного контакта транзистора, иллюстрирующая механизм выпрямления.

a — равновесное состояние; *б* — прямое направление тока; *в* — обратное направление тока.

Можно ожидать, что этот эффект должен быть более резко выражен чем это показано на фиг. 18, и из области вблизи поверхности будут выталкиваться не только избыточные электроны (в зоне проводимости), но частично и электроны валентной зоны. Вследствие этого дырки (фиг. 44, *a*) концентрируются непосредственно вблизи металлического контакта. Это во многих отношениях напоминает описанный в предыдущем параграфе *p-n*-переход. Если к контакту напряжение приложено в пропускном направлении (фиг. 44, *б*), то значительная часть тока переносится дырками. Представления, аналогичные описанным в предыдущем параграфе, были использованы Бардином и Браттейном в качестве критерия хорошего дырочного эмиттера. Если концентрация дырок вблизи поверхности значительно больше концентрации электронов в объеме полупроводника, то большая часть тока будет переноситься дырками. К обсуждению этого вопроса мы еще вернемся ниже в связи с рассмотрением теории „формовки“, предложенной Бардином и Пфаном.

Картина прохождения тока в запиорном направлении изображена на фиг. 44, *в*. В этом случае приложенное напряжение имеет то же направление, что и разность потенциалов, создающаяся поверхностными состояниями, и поэтому электроны смещаются еще дальше от точки контакта. Обратный ток обуславливается частично электронами, которые обладают в металле тепловой энергией, достаточной для перехода

в зону проводимости полупроводника, после чего могут в нем двигаться. Этот процесс возбуждения подобен процессу термоионной эмиссии из накаливаемой металлической нити. Так как работа выхода, необходимая для перехода электронов в полупроводник, много ниже работы выхода из металла, то заметный электрический ток в полупроводнике может протекать уже при комнатной температуре.

Другая часть тока переносится дырками, возникающими в объеме полупроводника или вблизи его поверхности благодаря тепловому возбуждению. Эти дырки возникают таким же образом и играют совершенно ту же роль, как и в $p-n$ -переходах. Однако между этими случаями имеется существенное различие. Так как эмиссия электронов из металла определяется работой выхода, которую электроны должны преодолеть, то при любом снижении работы выхода электронная эмиссия возрастает. Дырки, двигающиеся по направлению к металлу, притягивают электроны и облегчают им переход. Таким образом, наличие дырок увеличивает электронный поток.

К обсуждению этого вопроса, являющегося одним из возможных объяснений того, что коэффициент α_i может быть больше единицы, мы еще вернемся в параграфе, где будут рассмотрены вопросы формовки.

Выпрямление, происходящее на границе между металлом и полупроводником (см. фиг. 44), в основном объясняется тем же, что и выпрямление на $p-n$ -переходе; поэтому мы не будем здесь детально повторять обсуждение этого вопроса. Основное отличие заключается в том, что процессы возникновения и рекомбинации носителей тока, имеющие место в p -области $p-n$ -перехода, заменяются здесь процессами эмиссии и поглощения электронов при переходе через потенциальный барьер на поверхности металла и аналогичными процессами для дырок.

В реальном выпрямителе обычно необходимо учитывать скачки потенциала не только на самом выпрямляющем контакте, но и в других частях его. В выпрямителе с точечным контактом это добавочное сопротивление называется *сопротивлением растекания*. Эквивалентная схема выпрямителя с точечным контактом изображена на фиг. 45.

Запирающий слой приводит к появлению сопротивления между точечным металлическим контактом и полупроводником. Сопротивление этого слоя есть контактное сопротивление R_c . Последовательно с ним включено сопротивление растекания R_s полупроводника, через который должен пройти ток, чтобы достичь металлического основания. Площадь контакта последнего обычно настолько велика, что сопротивлением его можно пренебречь даже в том случае, когда он выпрямляет; основной электрод делается обычно таким, чтобы по возможности уменьшить нелинейность его вольтамперной характеристики. Большая часть сопротивления R_s обуславливается областью полупро-

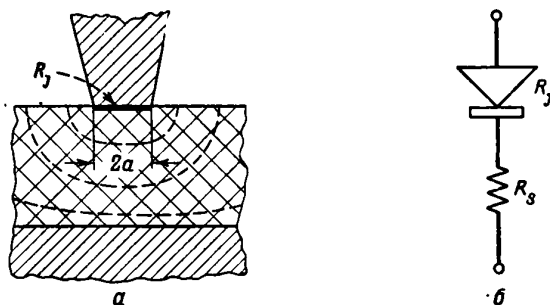
водника, непосредственно примыкающей к точечному контакту эмиттера, где сгущаются силовые линии тока; величина R_S , таким образом, является функцией площади контакта. Если считать, что контакт имеет форму круга с радиусом a , то сопротивление растекания определяется формулой

$$R_S = \frac{1}{4\sigma a}, \quad (4.32)$$

где σ — проводимость вещества. Если контакт имеет форму полушферы, то

$$R_S = \frac{1}{2\pi\sigma a}.$$

В теориях выпрямления тока, не учитывающих внедрения носителей заряда другого знака, величина R_S является постоянной. Сопротивление растекания R_S может быть определено экспериментально при



Фиг. 45. Эквивалентная схема детектора с точечным контактом.

приложении больших напряжений в пропускном направлении, когда R_j становится малым; в этом случае ограничивающим ток сопротивлением становится R_S . Для случая германия найдено, что величина ограничивающего сопротивления в пропускном направлении гораздо меньше, чем это следует из значений a и σ ¹⁾. В большинстве случаев это объясняется тем, что введение дырок в германий n -типа увеличивает

¹⁾ См. работу Брэй, Ларк-Горовитца и Смита [20]. Эти авторы предположили, что данный эффект обусловлен изменением проводимости под действием сильного поля. Впоследствии этот эффект исследовал Брэй [21]. Райдер и Шокли [22] показали, что эффект сильного поля обусловлен вхождением носителей тока из металлических электродов в полупроводник. Это заключение было в дальнейшем подтверждено экспериментами, описанными в § 1 гл. 3. Было также высказано предположение, что малое значение сопротивления растекания может быть объяснено наличием поверхностного слоя с проводимостью обратного знака [23]. — *Прим. авт.*

проводимость области R_g , так что для прямого тока уменьшается как величина R_L , так и величина R_g .

В германиевых выпрямителях n -типа дополнительный эффект введения дырок получается в результате наличия высокой концентрации дырок при прохождении больших токов в пропускном направлении. Если внезапно изменить направление приложенного напряжения, то эти дырки втягиваются точечным контактом и создают импульс дырочного тока. В ряде случаев этот эффект может быть настолько большим, что приводит к повреждению образца ¹⁾.

Мы считали целесообразным сосредоточить основное внимание на рассмотрении двух видов тока, протекающего через выпрямляющий контакт, так как это существенно для уяснения теории транзистора. Следует, однако, отметить, что теории выпрямления, не учитывающие внедрения носителей заряда противоположного знака, за исключением существенного для нас вопроса о дырочном токе, также объясняют большинство особенностей схемы контакта транзистора, приведенных на фиг. 44. Эти теории связаны, главным образом, с именами Мотта [25, 26] и Шоттки [27—29]. Интересные данные относительно слоя объемного заряда, обедненного электронами, могут быть, согласно Шоттки, получены из измерений комплексного полного сопротивления перехода, если рассматривать последний как параллельное соединение сопротивления и емкости. Можно показать, что эта емкость соответствует емкости слоя диэлектрика с толщиной, равной толщине слоя объемного заряда. Толщина этого слоя, как показано на фиг. 44, изменяется с изменением приложенного напряжения, увеличиваясь в запирающем направлении и уменьшаясь в пропускном направлении.

Измерения емкости слоя в зависимости от приложенного напряжения производились многими исследователями, и полученные результаты в основном находятся в хорошем согласии с предсказаниями теории Шоттки ²⁾. Влияние частоты недавно более подробно исследовалось Бардином и им же установлена граница применимости эквивалентной схемы Шоттки [30].

Простое объяснение возникновения скачка потенциала на запертом слое может быть дано, исходя из представлений о локальной проводимости. Для этого запертый слой подразделяют на ряд тонких параллельных подслоев, каждый из которых характеризуется различной концентрацией электронов и, следовательно, имеет различное сопротивление. Можно показать, что полное сопротивление запертого слоя (определяемое как частное от деления приложенного напряжения на ток) является просто суммой сопротивлений отдельных подслоев. Как общая толщина запертого слоя, так и сопротивления отдельных

¹⁾ Эффект такого рода наблюдался Михэмом и Михэзльсом [24]; Вальц и Блер назвали его *эффектом выгорания*. — *Прим. авт.*

²⁾ См. работы Шоттки [27—29], а также [18] и обзорную статью Анжелло [31]. — *Прим. авт.*

подслоев возрастают с увеличением обратного напряжения и убывают с возрастанием прямого напряжения. Этим и объясняется зависимость сопротивления от тока в запиорном слое. Такая картина может быть рассчитана количественно, и можно показать, что она эквивалентна другим приближениям.

Только что описанный метод имеет, однако, серьезный недостаток. Если предположить, что переходной слой состоит из ряда параллельных слоев, то можно рассматривать каждый такой подслой как параллельное соединение сопротивления и емкости и пытаться определить полное сопротивление перехода, соединяя последовательно эти элементы. Можно показать, что такое предположение совершенно неправильно ¹⁾. Однако вопрос этот сложен, и для более подробного ознакомления с ним читателю рекомендуется обратиться к работам Бардина, в которых показано, что получить правильное решение задачи на основе вышеизложенного предположения невозможно.

§ 4. ТЕОРИЯ ТРАНЗИСТОРА ТИПА А

Связь между физической картиной явления и эквивалентной схемой. Для объяснения процессов, происходящих в транзисторе типа А, и получения его эквивалентной схемы необходимо обобщить ряд положений, развитых для более простых случаев.

Были рассмотрены следующие вопросы:

- 1) процесс введения дырок через эмиттер;
- 2) усиление тока коллекторным контактом, характеризующимся собственным значением коэффициента усиления по току α_c ;

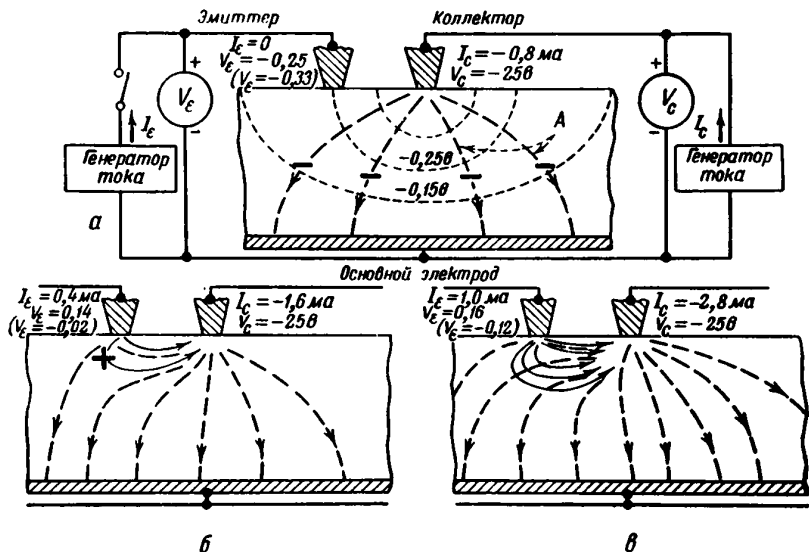
3) свойства нитевидного транзистора и его эквивалентная схема.

Расчеты, связанные с последним вопросом, мы видоизменим так, чтобы они могли быть применены к геометрии транзистора типа А. В результате получим физическую картину прохождения тока через транзистор типа А, а также возможность рассмотреть эквивалентную схему этого транзистора на основе более общих представлений. Для этого воспользуемся фиг. 46, на которой схематически показано распределение токов в интересующем нас случае.

Указанные на фиг. 46 численные значения токов и напряжений, как это следует из изображенных на фиг. 21 статических характеристик, примерно совпадают с обычными данными для транзистора типа А, за исключением значений эмиттерных напряжений, заключенных в круглые скобки. Последние, о чем более подробно речь будет идти ниже, соответствуют увеличению в эквивалентной схеме сопротивления r_{12} на 100 ом, так что входное сопротивление становится отрицательным.

¹⁾ По вопросу об интерпретации результатов измерений таких неоднородных сопротивлений на разных частотах см. работы [48, 49]. — *Прим. ред.*

Транзистор типа А отличается от нитевидного транзистора двумя принципиальными особенностями: 1) высокое сопротивление коллектора в случае транзистора типа А строго локализовано и возникает вследствие контактного сопротивления коллектора; большую часть времени пролета дырка находится в области полупроводника между эмиттером и коллектором с относительно низким сопротивлением;



Фиг. 46. Диаграмма токов в транзисторе.

А — расстояние между соседними линиями, соответствующее 0,2 м.

2) траектории электронов, движущихся от коллектора к основному электроду, не очень строго геометрически ограничены и обладают весьма разнообразными направлениями.

Изображенная на фиг. 46 схематическая картина распределения токов в транзисторе составлена в предположении, что эмиттерный ток целиком обусловлен одними дырками и что все эти дырки доходят до коллектора. В действительности же здесь необходимо было бы ввести в рассмотрение два множителя: γ — доля эмиттерного тока, переносимая дырками (для транзистора p -типа — доля тока, переносимая электронами), и β — та доля общего числа введенных эмиттером дырок (или электронов для транзистора p -типа), которая доходит до коллектора.

Для облегчения справок приведем здесь вторично определения различных значений коэффициента усиления по току α для транзистора n -типа (для транзистора p -типа вместо дырок следует

говорить об электронах):

$$\alpha = - \left(\frac{\partial I_c}{\partial I_e} \right) \text{ при } V_c = \text{const};$$

α_e — значение α для генератора тока или генератора напряжения в эквивалентной схеме фиг. 22 (индекс e означает, что данный параметр относится к эквивалентной схеме);

α_i — собственное значение α , равное увеличению коллекторного тока при увеличении дырочного тока коллектора на единицу при постоянном значении напряжения на коллекторе.

$1 - \alpha_i$, очевидно, равно увеличению электронного тока коллектора при увеличении дырочного тока на единицу.

Дырочный ток, достигающий коллектора, следующим образом связан с коэффициентами β и γ :

$$I_{pc} = -\beta\gamma I_e. \quad (4.33)$$

Здесь знак минус является следствием принятого условия, согласно которому *положительным* направлением для тока считается направление от точечного зонда к германию. Если каждая из этих дырок вызывает эмиссию $(\alpha_i - 1)$ избыточных электронов, то эмиттерный ток приведет к увеличению коллекторного тока на

$$\Delta I_c = -\beta\gamma I_e - (\alpha_i - 1)\beta\gamma I_e = -\alpha_i\beta\gamma I_e = -\alpha I_e. \quad (4.34)$$

Таким образом, величина коэффициента α в действительности обусловлена рядом эффектов и определяется уравнением

$$\alpha = \alpha_i\beta\gamma. \quad (4.35)$$

Полный ток коллектора будет равен поэтому

$$I_c = I_{c0}(V_c) - \alpha I_e. \quad (4.36)$$

Здесь $I_{c0}(V_c)$ — функция V_c , дающая обратную характеристику коллекторного контакта в отсутствие эмиттерного тока. Таким образом, это уравнение связывает три параметра, V_c , I_c и I_e , которые совместно с параметром V_e используются для определения статических характеристик.

Чтобы получить уравнение, содержащее параметр V_e , воспользуемся методами, аналогичными тем, которые были применены выше при выводе уравнения для эмиттерного контакта нитевидного транзистора (см. § 1). Для этого случая нами установлено, что, в соответствии с уравнениями (4.1) и (4.4),

$$I_e = f_e(V_e - V_J), \quad (4.37)$$

где V_J — потенциал некоторой точки в объеме германия, расположенного как раз под контактом эмиттера, определяемый уравнением

$$V_J = (I_e - I_c) R_b. \quad (4.38)$$

Для транзистора типа А величина R_b не имеет столь определенного смысла. Как это видно на фиг. 46, когда $I_e = 0$, эмиттер будет принимать потенциал, соответствующий проходящей через него эквипотенциальной поверхности. На фиг. 46 сопротивление между такой эквипотенциальной поверхностью и основным электродом показано равным $0,25 \cdot 0,8 \text{ ма} = 312 \approx 300 \text{ ом}$. При прохождении эмиттерного тока проводимость лежащей вблизи эмиттера области полупроводника модулируется и линии тока искажаются.

Точно решить эту задачу очень трудно, поэтому введем упрощающее предположение об аддитивности потенциалов. Это предположение, конечно, выполнялось бы точно, если бы можно было пренебречь модуляцией проводимости. При постоянной величине проводимости потенциал непосредственно под контактом эмиттера, который мы снова обозначим через V_J , мог бы быть получен простым наложением потенциалов, создаваемых в этой точке токами I_c и I_e . Согласно данному нами ранее определению R_b , коллекторному току соответствует потенциал $R_b I_c$. Аналогично, току I_e соответствует потенциал $R'_b I_e$. Растекание I_c по большой площади эквипотенциальной поверхности приводит к тому, что R_b становится меньше R'_b . Вводя величины R_b и R'_b и воспользовавшись предположением об аддитивности, можно записать

$$I_e = f_e(V_e - V_J) = f_e(V_e - R'_b I_e - R_b I_c). \quad (4.39)$$

Это выражение определяет соотношение между V_e и двумя токами I_e и I_c ; в совокупности с уравнением для I_c мы имеем систему из двух уравнений, необходимых для определения статических характеристик.

Величину $f_e(V_e - V_J)$ определить нелегко, поскольку V_J нельзя измерить непосредственно. Можно, однако, снять вольтамперную характеристику непосредственно самого эмиттера. Если ток коллектора равен нулю, то имеет место соотношение

$$I_e = f_{e0}(V_e) = f_e[V_e - R'_b f_{e0}(V_e)]. \quad (4.40)$$

Легко показать, что на основании двух последних уравнений мы должны получить ¹⁾

$$I_e = f_{e0}(V_e - R_b I_c). \quad (4.41)$$

Уравнение (4.41) получено Бардином и Браттейном более прямым путем на основании следующих соображений. Если $I_e = f_{e0}(V_e)$ при $I_c = 0$, то при $I_c \neq 0$ возникает действующее смещение, равное $+R_b I_c$; для того чтобы получить заданное значение эмиттерного тока, необходимо приложить добавочное внешнее напряжение, компенсирующее это напряжение смещения. В результате вместо V_e имеем $V_e - R_b I_c$.

¹⁾ Первое уравнение показывает, что между I_e и $V_e - R_b I_c$ имеется функциональная зависимость, которая записывается в виде $I_e = g(V_e - R_b I_c)$. Отсюда следует, что при $I_c = 0$ мы получим $g(V_e) = f_{e0}(V_e)$. — Прим. авт.

Эти соображения эквивалентны приведенным ранее и, в свою очередь, основаны на предположении об аддитивности напряжений, соответствующих токам I_e и I_c . Этот более подробный способ доказательства избран для того, чтобы показать, каким образом предположение об аддитивности напряжений входит в рассматриваемую нами теорию.

Предположение об аддитивности напряжений является хорошим приближением даже в том случае, если имеет место значительная модуляция проводимости. Такая модуляция приводит прежде всего к изменению вольтамперной характеристики эмиттерного контакта для тока прямого направления, и этот эффект будет иметь место, главным образом, в непосредственной близости от точки контакта эмиттера, где, как уже указывалось в § 3, в основном и сосредоточено *сопротивление растекания*. Если в этой области влияние коллектора не приводит к значительному нарушению модуляции проводимости, то изменение эмиттерного напряжения, обусловливаемое током I_e , не будет зависеть от тока коллектора. Кроме того, если модуляция проводимости имеет место лишь в ограниченной области вблизи эмиттера, то значение r_b не будет при этом сильно изменяться. Если эти условия выполнены, то предположение об аддитивности напряжений будет хорошо удовлетворяться. В действительности установлено, что предположение об аддитивности удовлетворительно только для ограниченной области значений токов; в одном случае, например, сопротивление r_b равнялось приблизительно 1 000 ом при $I_e = 0,1$ ма и всего около 400 ом при I_e , лежащем в пределах от 0,2 до 0,8 ма [32]. При полуколичественном сопоставлении эквивалентной схемы с физической картиной явления можно, однако, считать, что предположение об аддитивности выполняется.

Воспользовавшись введенными выше функциями, легко получить уравнения для эквивалентной схемы. Сопротивление коллектора r_{22} определяется из уравнения

$$I_c = I_{c0}(V_c) - \alpha I_e \quad (4.42)$$

путем дифференцирования его левой и правой части. При этом получаем

$$i_c = \frac{1}{r_{22}} v_c - \alpha i_e; \quad \frac{1}{r_{22}} = \frac{dI_{c0}}{dV_c}, \quad (4.43)$$

откуда

$$v_c = \alpha r_{22} i_e + r_{22} i_c = r_{21} i_e + r_{22} i_c \quad (r_{21} \equiv \alpha r_{22}), \quad (4.44)$$

что соответствует правой ветви эквивалентной схемы, изображенной на фиг. 22, а. Из формулы для эмиттерного тока

$$I_e = f_{e0}(V_e - R_b I_e) \quad (4.45)$$

получаем

$$i_e = \frac{1}{r_{11}} v_e - \frac{r_b}{r_{11}} i_c; \quad \frac{1}{r_{11}} = \left[\frac{\partial f_{e0}}{\partial V_e} \right]_{I_c = \text{const}} \quad (4.46)$$

и

$$v_e = r_{11}i_e + r_{12}i_c \quad (r_{12} \equiv r_b), \quad (4.47)$$

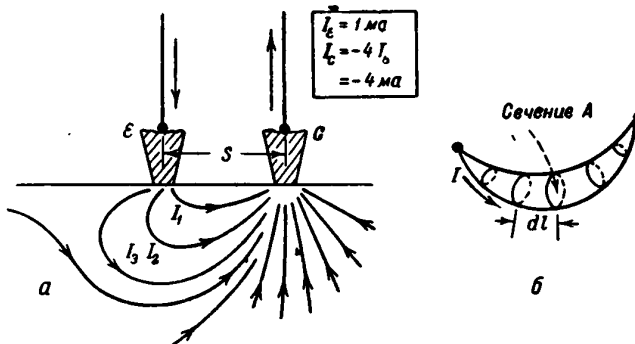
что соответствует левой части фиг. 22, а.

Эти уравнения показывают, в какой мере физическая модель, изображенная на фиг. 46, соответствует эквивалентной схеме. Эквивалентная схема уже достаточно подробно описана выше. Здесь мы коснемся лишь вопроса о неустойчивости и полном входном отрицательном сопротивлении. Как уже было показано в § 1, понижение полного входного сопротивления при увеличении тока обусловлено уменьшением сопротивления входного контакта с увеличением силы тока. Аналогичный эффект имеет место в транзисторе типа А. Положительная обратная связь обусловлена здесь тем, что коэффициент α больше единицы, а также наличием сопротивления r_b . Если коллектор подключен к источнику напряжения с нулевым полным сопротивлением, то в дополнение к сопротивлению r_{11} возникает отрицательное сопротивление, равное αr_b . На фиг. 46 показано, к чему приводит увеличение сопротивления r_b на 100 ом при неизменном значении $f_{\text{ср}}$. Как видно, это приводит к изменению V_e на величину, равную 100 I_c , или соответственно на $-0,08$, $-0,16$ и $-0,28$ v для указанных на графике значений I_c ; новые данные приводятся в скобках. Такого возрастания r_b достаточно, чтобы создать в блоке отрицательное входное полное сопротивление при значениях I_c , лежащих в пределах от 0,4 до 1,0 ма (см. фиг. 46, б и фиг. 46, в).

В случае коаксиального транзистора, за исключением необходимости учета некоторых особенностей, связанных с геометрией, изложенная теория в основном остается без изменений.

Эффекты, связанные с временем пролета. Время пролета носителей заряда играет весьма различную роль в нитевидных транзисторах и в транзисторах типа А. Областью высокого полного сопротивления на входе транзистора типа А является запорный слой толщиной порядка 10^{-5} см , характеризующийся скачком потенциала в 25 v . Для такого слоя влияние времени пролета чрезвычайно мало по сравнению с эффектами, играющими существенную роль в работе транзисторов. Если бы все дырки при своем движении от эмиттера к коллектору проходили по траекториям одинаковой длины и затрачивали на это одинаковое время, то ток, возбуждаемый в коллекторной цепи, в точности соответствовал бы, хотя и с некоторой задержкой во времени, входному сигналу. Если же, как это показано на фиг. 46, траектории дырок неодинаковы, то дырочный ток коллектора будет определяться средним значением эмиттерных токов в различные предшествующие моменты времени. Если разброс времени пролета сравним с периодом переменного входного сигнала, то это приведет к значительному уменьшению коллекторного тока, а следовательно, и к уменьшению усиления.

На фиг. 47, *а* показан способ оценки времен пролета и разброса их значений. Ток, текущий из эмиттера, условно подразделяется на ряд частей, каждая из которых ограничена линиями или поверхностями тока. Для простоты мы будем пренебрегать диффузионными эффектами и модуляцией проводимости. При этих условиях между объемом, через который протекает ток, и средним временем пролета существует простое соотношение, которое иллюстрируется фиг. 47, *б*. Ток I течет по трубке, ограниченной поверхностями тока и обладающей поперечным сечением $A(l)$, где l — расстояние, отсчитываемое вдоль трубки. Линии тока в этом случае легко могут быть найдены



Фиг. 47. Растекание линий тока и схема трубки тока.

на основании электростатической теории [33]. Поле E и скорость дрейфа v в точке l определяются уравнениями

$$\sigma A E = I, \quad E = \frac{I}{\sigma A} \quad \text{и} \quad v = \mu E = \frac{\mu I}{\sigma A}. \quad (4.48)$$

В этом случае время dt , необходимое для того, чтобы дырка прошла отрезок пути dl , равно

$$dt = \frac{dl}{v} = \frac{\sigma A dl}{\mu I} = \left(\frac{\sigma}{\mu I}\right) dV_I, \quad (4.49)$$

где dV_I — элемент объема трубки. Отсюда следует, что время пролета равно

$$t = \frac{\sigma V_I}{\mu I}, \quad (4.50)$$

где V_I — объем, соответствующий току I .

Очевидно, что линиям тока, изображенным на фиг. 47, соответствует большой разброс времен пролета. В качестве грубого приближения примем, что этот разброс сравним со временем пролета электронов группы I_2 . Эта группа будет соответствовать примерно половине эмиттерного тока. Из фиг. 47 следует, что объем трубки

тока, осью симметрии которой является линия $z-c$, равен объему цилиндра длиной S единиц, обладающего средним значением радиуса около $0,6 S$ единиц, усредненным по половине окружности. Объем трубки, таким образом, должен был бы быть равен

$$V_1 = \frac{\pi}{2} (0,6S)^2 S \approx \frac{S^3}{2}. \quad (4.51)$$

Время пролета отсюда было бы равно

$$\tau = \frac{\sigma S^3}{2\mu I_2} = \frac{\sigma S^3}{\mu I_e}. \quad (4.52)$$

Подставив в эту формулу типичные для I_e , μ , σ и S численные значения

$$I_e = 10^{-3} \text{ а}, \quad \mu = 1700 \text{ см}^2/\text{сек}, \quad \sigma = 0,2 \text{ ом}^{-1} \cdot \text{см}^{-1}, \quad (4.53a)$$

$$S = 5 \cdot 10^{-3} \text{ см}, \quad (4.53b)$$

получаем

$$\tau = \frac{0,2 \cdot 125 \cdot 10^{-9}}{1700 \cdot 10^{-3}} = 0,015 \text{ мксек}. \quad (4.54)$$

Если, как предполагалось, разброс времен пролета сравним с этой величиной, то при частоте 60 мггц усиление будет близко к нулю, а при частоте, несколько меньшей 30 мггц , усиление будет порядка 3 дб .

Этот грубый расчет, приведенный здесь лишь для иллюстрации метода расчета, дает для критической частоты примерно в 4 раза большее значение, чем то, которое наблюдалось для промежутка $S = 5 \cdot 10^{-3} \text{ см}$. Такое расхождение, повидимому, в значительной степени обусловлено модуляцией проводимости вблизи контакта эмиттера, повышающей значение σ и увеличивающей время пролета.

Из всего сказанного следует, что те условия, которые будут приводить к сокращению длины линий тока между эмиттером и коллектором, должны вызывать также уменьшение времени пролета. Такие условия будут, например, иметь место, если транзистор находится в магнитном поле, которое, как это показано на фиг. 47, отклоняет по направлению к коллектору траектории дырок, движущихся вниз. Это приводит к сокращению длины трубок тока и к уменьшению времени пролета ¹⁾.

§ 5. О ТЕОРИИ „ФОРМОВКИ“ И ВЫСОКИХ ЗНАЧЕНИЯХ α

В своих ранних работах по исследованию транзисторов Брагтейн заметил, что качество транзистора может быть значительно улучшено, если через коллектор пропустить большой обратный ток. С тех пор

¹⁾ Об этом эффекте сообщил Браун в июле 1949 г. на конференции по электронным приборам в Принстонском университете (см. также работу [34]). — *Прим. авт.*

искусство формовки достигло значительного развития; разработано большое количество методов формовки, дающих возможность применять для этой цели контролируемые по величине токи и напряжения в течение различных промежутков времени.

Мы не будем пытаться здесь дать обзор всего комплекса этих вопросов, а лишь ограничимся рассмотрением некоторых общих результатов наблюдений, касающихся влияния полярности тока и типа проводимости на процесс формовки, и некоторых предположений относительно причин, обуславливающих большие значения коэффициента усиления по току α , например $\alpha > 5$, для коллекторов.

Процесс формовки. Рассмотренные Бардином и Пфаном [35] способы формовки основаны на принципе пропускания через контакт в прямом или обратном направлении кратковременных импульсов тока

Таблица 4

ВЛИЯНИЕ СПОСОБА ФОРМОВКИ

Неформованный	Формовка при положительном потенциале (+ток)	Формовка при отрицательном потенциале (-ток)
Германий <i>n</i> -типа		
(в)	(г)	(д)
<p>Прямой ток (на точечном контакте +): преобладают дырки, хороший эмиттер</p> <p>Обратный ток (на точечном контакте -): мал, хороший выпрямитель</p>	<p>Прямой ток (на точечном контакте +): меняется мало или становится большим</p> <p>Обратный ток (на точечном контакте -): еще меньше</p>	<p>Прямой ток (на точечном контакте +): дырочный ток меньше</p> <p>Обратный ток (на точечном контакте -): возрастает, формируется коллектор</p>
Германий <i>p</i> -типа		
(е)	(ж)	(з)
<p>Прямой ток (на точечном контакте -): мал, мало электронов, плохой эмиттер</p> <p>Обратный ток (на точечном контакте +): большой, плохое выпрямление</p>	<p>Прямой ток (на точечном контакте -): такой же, как у неформованного образца</p> <p>Обратный ток (на точечном контакте +): возрастает</p>	<p>Прямой ток (на точечном контакте -): электронный ток возрастает, формируется эмиттер</p> <p>Обратный ток (на точечном контакте -): большой,</p>

длительностью в 1 сек. или меньше. Вызываемые формовкой изменения частично зависят от обработки поверхности германия, частично же — от сорта материала, применяемого для изготовления „усов“ точечных контактов. В исследованиях Бардина и Пфана для изготовления „усов“ применялась фосфористая бронза, а поверхность германия шлифовалась и затем протравливалась. При этом было установлено, что результаты влияния формовки в зависимости от условий обработки могут быть систематизированы так, как это показано в приведенной выше табл. 4.

Бардин и Пфан указали, что приведенные в этой таблице данные могут быть объяснены с помощью следующего предположения относительно влияния формовки: формовка приводит к изменению высоты потенциального барьера, обратному тому, которое вызывается во время формовки приложенным напряжением. Другими словами, если точечный контакт в процессе формовки находится под отрицательным потенциалом так, что высота потенциального барьера, который должны преодолеть электроны при переходе из полупроводника в металл, увеличивается, то после того, как напряжение формовки снято, высота барьера изменится и станет меньше, чем до формовки.

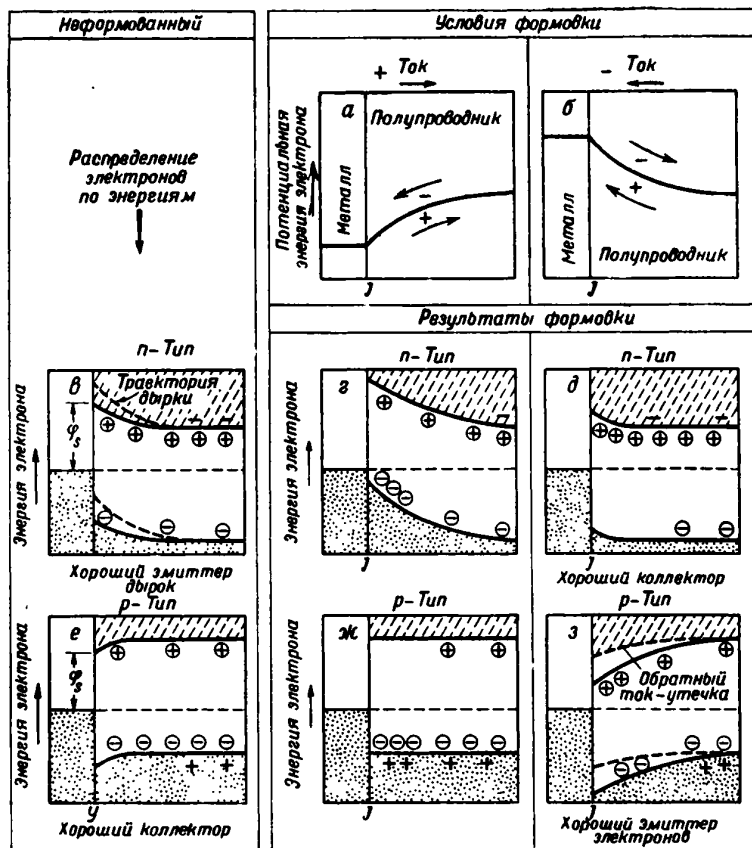
Этот эффект наглядно иллюстрируется при помощи фиг. 48. Изображенные на этой фигуре шесть диаграмм соответствуют шести случаям, рассмотренным в табл. 4. В каждом из этих случаев влияние формовки сводится к изменению высоты потенциального барьера в обратном направлении по отношению к тому изменению, которое вызывает приложенное напряжение формовки, что также показано на чертеже.

в)¹⁾ Предполагается, что нормальный контакт с германием n -типа создает запирающий слой, характеризуемый скачком потенциала ϕ_s , величина которого в основном определяется поверхностными состояниями и очень слабо зависит от природы металлического контакта. Предполагается также, что барьер не вполне однороден и в некоторых точках оказывается более высоким. В таких точках концентрация дырок особенно велика, и когда через контакт идет ток в прямом направлении, то этот ток в основном переносится дырками, которые входят в германий из этих областей высокой концентрации.

е) Поскольку потенциальная энергия поверхностных состояний одинакова как для германия n -типа, так и для германия p -типа, распределение потенциала, изображенное на фиг. 48, в, изменится так, как это показано на фиг. 48, е для германия p -типа. Низкий потенциальный барьер для дырок служит причиной наличия большого дырочного тока, плохого выпрямления и малой электронной эмиссии. Этот результат находится в согласии с тем известным опытным фактом, что контакты с германием p -типа выпрямляют обычно плохо.

¹⁾ Обозначения, применяемые в этом и последующих абзацах, соответствуют обозначениям в табл. 4 и на фиг. 48. — *Прим. ред.*

В случаях, показанных на фиг. 48, а и б, указана величина электрического поля, налагаемого на образец в процессе формовки. Влияние формовки можно объяснить, исходя из предположения, что поле



Фиг. 48. Схема процессов формовки по Бардину и Пфану.

при этом столь велико, что происходят перемещения примесей. Это приводит к изменениям концентраций и потенциалов, указанным на фиг. 48, г, д, ж, з; в результате ϕ_s будет стремиться измениться в направлении, противоположном приложенному потенциалу.

д) Для германия *n*-типа желательно, чтобы обработка приводила к увеличению обратного тока коллектора. Это даст возможность применять более высокие поля и, таким образом, повысить долю собираемых коллектором введенных в германий дырок.

з) В германии p -типа обратный ток достаточно велик и без всякой формовки, однако обычно лишь весьма незначительная доля эмиттерного тока переносится электронами. При понижении φ_n , как это показано на фиг. 48, з, под металлическим точечным контактом образуются области высокой концентрации электронов, которые могут являться хорошими эмиттерами электронов.

Интерпретация высоких значений α . Дырочный ток, протекающий в транзисторе типа А, вызывает увеличение эмиссии электронов из коллектора, как это следует из того факта, что α больше единицы. В § 3 уже упоминалось, что значения коэффициента α , превышающие единицу, можно объяснить образованием объемного заряда, который создается дырками, когда они достигают коллектора. Этот объемный заряд будет вызывать такой же эффект, какой вызывается положительными ионами на фиг. 48, д, повышенная концентрация которых непосредственно перед коллекторным контактом приводит к понижению работы выхода электронов и, следовательно, к увеличению обратного тока.

Эта теория, объясняющая возможность получения высоких значений α , была предложена Бардином и Браттейном в их первой работе [3]. Как будет показано ниже, таким путем может быть объяснено наличие значений α вплоть до 3. Наблюдались, однако, также значения α , намного превышающие 3, и для того, чтобы объяснить это обстоятельство, необходимо тщательно разобраться в вопросе о том, каким образом объемный заряд дырок вызывает увеличение эмиссии электронов. После того как мы исследуем границы применимости простой теории объемного заряда, нами будут изложены также и другие теории этого вопроса.

Покажем теперь, что максимальное значение для коэффициента усиления по току α , получающееся на основе простой теории объемного заряда для транзистора типа А, равно $(1 + b)$ так же, как и для нитевидного транзистора, и в основном по тем же причинам.

Пусть объемный заряд, соответствующий приращению дырочного тока ΔI_p , создает приращение электронного тока, равное ΔI_n . Вычислим, каков был бы общий объемный заряд при условии $\Delta I_n = b\Delta I_p$, что соответствует значению $\alpha = 1 + b$. Покажем, что ΔI_n не может быть столь велико. Очевидно, что ΔI_n будет создавать объемный заряд ρ_n , который будет стремиться компенсировать объемный заряд ρ_p , создаваемый током ΔI_p . Заряд ρ_n не может, однако, скомпенсировать заряд ρ_p полностью, так как если бы это произошло, то результирующий положительный объемный заряд равнялся бы нулю и, следовательно, повышение электронной эмиссии не должно было бы иметь места. Однако если $\Delta I_n = b\Delta I_p$, то ρ_n в точности равно ρ_p , поскольку при одинаковых плотностях заряда электронные и дырочные токи должны находиться в соотношении $b:1$. Отсюда следует, что дырочный ток не может создать электронный ток, равный $b\Delta I_p$,

а следовательно, и α , или, точнее, α_i , не может достигать величины, равной $(1 + b)$. Это заключение основано на трех, неявно принятых нами предположениях: 1) отношение подвижностей не изменяется при наличии сильных полей у коллектора ¹⁾; 2) дырки и электроны перемещаются по одним и тем же траекториям (что, вообще говоря, может быть неверно, если переход „неоднороден“; 3) дырки без затруднений входят в металл и, таким образом, не создают на границе с ним объемного заряда. Хотя эти предположения, вероятно, точно и не выполняются, тем не менее сомнительно, чтобы это могло явиться источником каких-либо серьезных ошибок ²⁾.

Высказывалось также предположение, что усиление тока может быть обусловлено возбуждением вторичных электронов при быстром движении дырок в сильном поле вблизи коллектора. Против этой теории имеется, однако, много возражений. В некоторых случаях большие значения α наблюдались при столь малых напряжениях, например всего в 5 в, что появление вторичных электронов мало вероятно. Кроме того, если бы вторичная эмиссия играла существенную роль, то следовало бы ожидать, что α будет быстро возрастать с увеличением приложенного напряжения. Этот процесс должен был бы быть сходен с лавинообразным процессом при диэлектрическом пробое, и значение α изменялось бы при этом экспоненциально с изменением напряжения. В действительности это не наблюдается, и полученные данные значительно лучше согласуются с предположением о том, что α_i не зависит от приложенного напряжения, как если бы каждая дырка обуславливала эмиссию вполне определенного числа электронов, независимо от величины приложенного к коллектору напряжения. Поэтому мы не будем вдаваться в более подробное исследование теории вторичной эмиссии и перейдем к рассмотрению других теорий, основанных на предположении об образовании объемного заряда дырок [37].

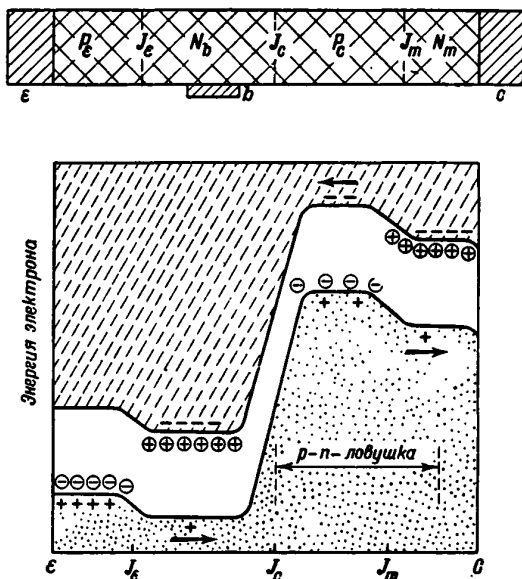
Теории, которые будут нами сейчас рассмотрены, обладают одной общей чертой: предполагается, что движение дырок к металлическому контакту затруднено, так что имеет место образование дырочного объемного заряда, оказывающего влияние на движение электронов.

На фиг. 49 показан транзистор с коллекторной областью усиления тока. Он отличается от транзистора, изображенного на фиг. 43, тем, что непосредственно за областью p -типа (P_c) имеется область „умножения“ из материала n -типа, обозначенная нами через N_m . При этом предполагается, что область P_c настолько узка, что электроны, входящие в нее из области N_m , обладают большой вероятностью пройти

¹⁾ Согласно работе Райдера и Шокли [50], сама подвижность электронов в германии в полях с напряженностью, превышающей $6 \cdot 10^8$ в/см, изменяется пропорционально E^{-1} . — *Прим. ред.*

²⁾ Сходные соображения были высказаны Гунтером [36]. — *Прим. авт.*

через нее, не рекомбинируя при этом с дырками. Благодаря действию эмиттера дырочный ток проходит через J_c , через область N_m и через J_c . Эти дырки благодаря наличию потенциальной „ловушки“, создаваемой переходом J_m , могут оказаться связанными в области P_c . Рассмотрим процесс прохождения тока через J_m . Если в области N_m электронная проводимость играет значительно более существенную



Фиг. 49. Транзистор с коллекторной областью усиления тока.

роль, чем дырочная проводимость в области P_c , ток через переход J_m , как уже указывалось в § 2, будет переноситься в основном электронами. Кроме того, благодаря незначительной ширине области P_c электроны могут переходить справа налево не рекомбинируя; это обстоятельство также должно приводить к уменьшению обратного диффузии в область N_m и, следовательно, к увеличению обратного тока. В результате электронный ток, проходящий через область P_c к основному электроду, будет значительно превышать дырочный ток, текущий через J_m к N_m . Вследствие этого переход J_m действует как усилитель электрического тока, и дырочный ток, доходящий до коллектора, вызывает значительное увеличение электронной эмиссии. Такое объяснение больших значений α принято называть *теорией p-n „ловушки“*.

Следует указать, что существует полная симметрия между правой и левой частями фиг. 49, за исключением перемены знака. Наличие

этой симметрии позволяет высказать несколько иную точку зрения относительно роли той области потенциала, которую мы назвали „ловушкой“. Эта область, по существу, ничем не отличается от области N_m , являющейся эмиттером электронов, и области P_c . Поток дырок к P_c смещает этот эмиттер электронов вперед и обуславливает ввод электронов в область P_c и далее к J_c , в обоих случаях играющего роль коллекторного перехода.

Теория „ловушечного“ усилителя тока с математической точки зрения очень сходна с теорией $p-n-p$ -транзистора, основные математические уравнения которой даны нами ранее [16].

Коллекторные контакты, действие которых, повидимому, основано на только что рассмотренном „ловушечном“ механизме, изготовлены Пфаном, Андерсоном, Райдером, Уоллэсом и Вальдсом. При этом были получены значения α порядка 20 и даже более высокие.

В рассмотренном нами выше случае простого объемного заряда предполагалось, что предельное значение $\alpha_t = (1 + b)$. Если бы величина b могла быть увеличена, то, очевидно, это привело бы к относительному увеличению пространственного заряда дырок, а следовательно, к увеличению α . Один из возможных процессов, которые могли бы вызвать такой эффект, это — „захват дырок“.

Если бы непосредственно вблизи коллектора могла быть создана значительная концентрация центров „захвата дырок“, то дырки накапливались бы в этой области и это привело бы к увеличению объемного заряда. Если мы предположим, что каждая из дырок в среднем проводит долю x времени в связанном состоянии, а долю $(1-x)$ времени движется, обладая нормальной подвижностью, то в результате подвижность дырки должна уменьшиться в $(1-x)$ раз и во столько же раз должно увеличиться эффективное значение b , т. е. мы получим, что $b_{эфф.} = b/(1-x)$. Если $(1-x)$ мало, то этот процесс приведет к очень большим значениям $b_{эфф.}$ и, следовательно, к большим значениям α .

Можно думать, что предложенный „ловушечный“ механизм для объяснения больших значений α находится в лучшем согласии с опытом, чем все остальные теории и что эти представления найдут плодотворное применение в практике создания транзисторов.

§ 6. ФОТОТРАНЗИСТОРЫ И СЧЕТЧИКИ

Шайву [38] удалось сочетать явление усиления тока коллектором с явлением оптического возбуждения дырок и электронов и создать таким образом *фототранзистор*. Последний отчасти сходен с коаксиальным транзистором [39, 40], представляющим собой образец германия в виде тонкой облатки с эмиттерным и коллекторным контактами, расположенными с двух противоположных ее сторон. В фототранзисторе эмиттер отсутствует, и дырки в германии n -типа образуются в результате поглощения света. Эти дырки перемещаются затем

к коллектору и создают усиленный фототок. Таким образом, если каждый фотон создает одну пару электрон—дырка, а дырка, достигая коллектора, обуславливает эмиссию двух электронов, то это соответствует значению $\alpha_1 = 3$. Число зарядов, создаваемых одним поглощенным фотоном, называется *квантовым выходом*; в случае фототранзистора (благодаря большому значению α_1 коллектора) квантовый выход может превышать 100%.

Вопрос о квантовом выходе процесса поглощения света изучен Гаучером [41]. В его опытах измерялось изменение проводимости нитевидного образца германия *n*-типа под действием освещения. Изменение проводимости, как это видно из формулы, приведенной в задаче 3 гл. 3, должно быть пропорциональным общему числу дырок, имеющих в образце. Как легко показать с помощью методов, развитых в § 6 гл. 12, количество дырок равно скорости их образования, умноженной на время их жизни. В свою очередь, скорость образования дырок равна произведению квантового выхода на число фотонов, поглощенных в единицу времени, которое может быть определено, если известны энергия и длина волны света, а также оптические постоянные германия.

В опытах Гаучера время жизни дырок измерялось с помощью методов, применявшихся при изучении процесса введения дырок в германий и описанных нами в § 1 гл. 3 и § 6 гл. 12. Сопоставив наблюдаемые изменения проводимости с числом поглощенных фотонов, Гаучер установил, что при освещении светом с длинами волн, лежащими в пределах от 1,0 до 1,8 μ , каждый поглощаемый фотон создает одну пару электрон—дырка.

Коллекторные переходы применялись также Мак-Кэем [42] для обнаружения дырок, создаваемых в германии α -частицами. Мак-Кэй установил, что чувствительная область обладает диаметром 10^{-2} — 10^{-3} см, а максимальная высота импульса соответствует прохождению 10^6 электронов. Продолжительность импульса была при этом меньше 0,05 мсек. Мак-Кэй предположил также, что *p*—*n*-переходы с большой поверхностью могут быть использованы для создания счетчиков. Это предположение нашло подтверждение в работе Ормана, Фэна, Гольдсмита и Ларк-Горовитца [43].

Переходы *p*—*n*-типа сами по себе являются весьма эффективными фотоэлементами [13, 14]. Для получения тока создание пар электрон—дырка в самом переходе не обязательно; достаточно, чтобы они возникли вблизи перехода так, чтобы вероятность диффузии носителей тока к переходу была велика. Эта вероятность определяется диффузионной длиной, о которой речь будет идти в § 5 гл. 12.

ЗАДАЧИ

1. Предположим, что вероятность рекомбинации введенной дырки за время dt равна dt/τ_p . Показать, что если в момент времени t' введено некоторое количество дырок, то доля дырок, которые

к моменту t останутся несвязанными, равна $\exp(t' - t)/\tau_p$ (аналогичный анализ других физических задач см. в § 3 гл. 8).

2. Предположим, что в нитевидном транзисторе ток эмиттера равен $i_e(t)$ и что доля этого тока γ обуславливается дырками. Считая время пролета равным τ_i , показать, что число дырок в образце равно

$$P_{\text{общ.}} = \left(\frac{1}{e}\right) \int_{t-\tau_i}^t i_e(t') \exp\left[\frac{(t'-t)}{\tau_p}\right] dt'$$

и, следовательно (воспользовавшись уравнением задачи 3 гл. 3), увеличение тока равно

$$\begin{aligned} \delta i_c(t) &= (V_J - V_C) \gamma (\mu_n + \mu_p) L^{-2} \int_{t-\tau_i}^t i_e(t') \exp\left[\frac{(t'-t)}{\tau_p}\right] dt' = \\ &= \gamma (1 + b) \int_{t-\tau_i}^t i_e(t') \exp\left[\frac{(t'-t)}{\tau_p}\right] \frac{dt'}{\tau_i}. \end{aligned}$$

Показать, что этот результат сводится к уравнениям (4.18) и (4.19), так как при $i_e(t) = i_{e0} \exp i\omega t$ имеем $\delta i_c(t)/i_e(t) = \alpha_e = \gamma(1 + b)\beta$.

3. Переход $p-n$ -типа площадью $A = 0,5 \text{ см}^2$ имеет при малых напряжениях сопротивление $R = 400 \text{ ом}$. Предполагая, что ток в основном переносится дырками и воспользовавшись соотношением (4.29), показать, что

$$\left(\frac{e}{kT}\right) I_{ps} A = \frac{1}{R},$$

так что

$$I_{ps} = \frac{kT}{eAR} = \frac{1}{39AR} = 1,28 \cdot 10^{-4} \text{ а/см}^2.$$

Полагая, что время жизни дырки в n -области равно 100 мксек., показать, что на основании уравнения (4.31) это приводит к

$$p_n = \frac{(D_p \tau_p)^{1/2}}{AR e \mu_p} = 1,2 \cdot 10^{12} \text{ см}^{-3},$$

причем эта концентрация соответствует дырочной проводимости, равной $3,3 \cdot 10^{-4} \text{ ом}^{-1} \cdot \text{см}^{-1}$. Это много меньше проводимости германия при комнатной температуре ($2,1 \cdot 10^{-2} \text{ ом}^{-1} \cdot \text{см}^{-1}$) и свидетельствует о том, что высокое сопротивление $p-n$ -перехода обуславливается тем обстоятельством, что прохождение дырочного тока через материал n -типа, в котором число дырок весьма незначительно, затруднительно. Этот вопрос рассматривается более подробно в гл. 12.

ЛИТЕРАТУРА

1. Shockley W., Pearson G. L., Sparks M., Brattain W. H., Phys. Rev., **76**, 459 (1949).
2. Shockley W., Pearson G. L., Haynes J. R., Bell Syst. Techn. Journ., **28**, 344 (1949).
3. Bardeen J., Brattain W. H., Phys. Rev., **75**, 1208 (1949).
4. Benzer S., Phys. Rev., **72**, 1267 (1947).
5. Becker M., Fan H. Y., Phys. Rev., **75**, 1631 (1949).
6. Fan H. Y., Phys. Rev., **75**, 1631 (1949).
7. Sosnowski L., Starkiewicz J., Simpson O., Nature, **159**, 818 (1947).
8. Sosnowski L., Phys. Rev., **72**, 641 (1947).
9. Sosnowski L., Soole B. W., Starkiewicz J., Nature, **160**, 471 (1947).
10. Goucher F. S., Meeting of the American Physical Society, Cleveland, March 10, 1949.
11. Shockley W., Pearson G. L., Sparks M., Phys. Rev., **76**, 180 (1949).
12. Pearson G. L., Bardeen J., Phys. Rev., **75**, 865 (1949).
13. Scaff J. H., Theuerer H. C., Schumacher E. E., Journ. of Metals, **185**, 383 (1949).
14. Pfann W. G., Scaff J. H., Journ. of Metals, **185**, 389 (1949).
15. Bardeen J., Phys. Rev., **71**, 717 (1949).
16. Shockley W., Bell Syst. Techn. Journ., **28**, 435 (1949).
17. Wagner C., Phys. Zs., **32**, 641 (1931).
18. Torrey H. C., Whitmer C. A., Crystal Rectifiers, New York, 1948 (см. перевод: Торрей, Уитмер, Кристаллические выпрямители, М., 1950).
19. Benzer S., Journ. Appl. Phys., **20**, 804 (1949).
20. Bray R., Lark-Horowitz K., Smith R. N., Phys. Rev., **72**, 530 (1947).
21. Bray R., Phys. Rev., **74**, 1218 (1948).
22. Ryder E. J., Shockley W., Phys. Rev., **75**, 310 (1949).
23. Bardeen J., Brattain W. H., Phys. Rev., **74**, 230 (1948).
24. Meacham L. A., Michaels S. E., Phys. Rev., **78**, 175 (1950).
25. Mott N. F., Proc. Roy. Soc., **171A**, 27 (1939).
26. Mott N. F., Gurney R. W., Electronic Processes in Ionic Crystals, Oxford, 1940 (см. перевод: Мотт Н., Герни Р., Электронные процессы в ионных кристаллах, ИЛ, 1950).
27. Schottky W., Spenke E., Wiss. Veröff. aus die Siemens Werken, **18**, (3), 1 (1939).
28. Schottky W., Zs. f. Phys., **113**, 367 (1939); **118**, 539 (1942).
29. Rose F., Spenke E., Zs. f. Phys., **126**, 632 (1949) ¹⁾.
30. Bardeen J., Bell Syst. Techn. Journ., **28**, 428 (1949).
31. Angello S. J., Electr. Eng., **68** (10), 865 (1949).
32. Ryder R. M., Bell Lab. Rec., **27**, 89 (1949).
33. Jeans, Mathematical Theory of Electricity and Magnetism, Cambridge, 1927.
34. Brown C. B., Phys. Rev., **76**, 1736 (1949).
35. Bardeen J., Pfann W. G., Phys. Rev., **77**, 401 (1950).
36. Hunter L. P., Phys. Rev., **77**, 558 (1950).
37. Shockley W., Phys. Rev., **78**, 294 (1950).

¹⁾ Популярный обзор теории Шоттки см. Joffé J., Electr. Communication, **22**, 217 (1945). — *Прим. авт.*

38. Shive J. N., Phys. Rev., **76**, 575 (1949).
39. Shive J. N., Phys. Rev., **75**, 689 (1949).
40. Wallace R. L., Kock W. E., Electr. Eng., **68**, 222 (1949).
41. Goucher F. S., Phys. Rev., **78**, 646 (1950).
42. McKay K. G., Phys. Rev., **76**, 1537 (1949).
43. Orman C., Fan H. Y., Goldsmith G. J., Lark-Horovitz K., Phys. Rev., **78**, 646 (1950).
- 44*. Тønnesen Т. Н., Proc. Phys. Soc., **65B**, 737 (1952).
- 45*. Hall R., Dunlap W. C., Phys. Rev., **80**, 467 (1950).
- 46*. Teal G., Sparks M., Buchler E., Phys. Rev., **81**, 637 (1951).
- 47*. Aigrain P. R., Dugas C. R., Eitzel H. W., *Semiconducting Materials*, London, 1951.
- 48*. Субашиев В. К., Тучкевич В. М., ЖТФ, **17**, 177 (1947).
- 49*. Лашкарев В. Е., ЖТФ, **15**, 598 (1945).
- 50*. Ryder E. J., Shockley W., Phys. Rev., **81**, 139 (1951).

ЧАСТЬ II

ОПИСАТЕЛЬНАЯ ТЕОРИЯ
ПОЛУПРОВОДНИКОВ

**КВАНТОВЫЕ СОСТОЯНИЯ, ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ
И ЗОНЫ БРИЛЛЮЭНА ¹⁾****§ 1. ВВЕДЕНИЕ ²⁾**

В классической теории проводимости носители тока рассматриваются как маленькие частицы, совершающие беспорядочное тепловое движение, на которое накладывается действие электрического и магнитного полей. Предполагается, что, за исключением своих малых размеров, эти частицы почти ничем не отличаются от твердых тел больших размеров. Старым доквантовым теориям не удалось учесть многих особенностей поведения электронов. Это было сделано новой квантовой механикой. Квантовая механика вводит в рассмотрение ряд таких новых понятий, как „волновая функция“, „плотность вероятности“, „уровень энергии“, „спин“, „квантовое состояние“ и „принцип Паули“, — понятий, которые возникли в результате развития математических постулатов теории. Умение владеть математическим аппаратом теории достигается только в результате изучения специальных курсов физики. К счастью, результаты применения квантовой механики к исследованию движения электронов в кристаллах не нуждаются в описании с помощью математического аппарата и могут быть представлены в большинстве случаев достаточно наглядно.

Прежде чем перейти к рассмотрению деталей того или иного вопроса, читателю будет полезно узнать конечный результат вместе с некоторыми указаниями о пути, по которому нужно следовать при его получении. По мнению автора, имеет определенный смысл проследить, хотя бы и поверхностно, отдельные этапы развития теории. Согласно другой точке зрения, целесообразнее перескочить через эти промежуточные ступени и сразу представить себе окончательную физическую картину свойств электронов и дырок так, как это следует из теории. Второй путь нам кажется нежелательным по ряду соображений. Он не дает представления об основных понятиях, на которых основано рассмотрение свойств дырок. По этой причине не исключена возможность возникновения ошибочных представлений. Кроме того, с чисто научной точки зрения интересно показать, как электроны при наличии дырочной проводимости могут симулировать поведение положительных частиц.

¹⁾ Очерки основ зонной теории см., например, в книгах Я. И. Френкеля [9] и Э. И. Адировича [10]. — *Прим. ред.*

²⁾ К гл. 5, 6 и 7. — *Прим. авт.*

В этой и последующей главах дается описание поведения электрона в идеальном кристалле, в котором правильность расположения атомов не нарушена тепловыми колебаниями, атомами примесей или другими причинами. В этом случае электрон обладает некоторыми разрешенными видами движения, квантовыми состояниями. В этих состояниях он обладает определенными значениями энергии и скорости. Однако из волнового уравнения Шредингера следует, что классическое соотношение $\mathcal{E} = mv^2/2$ не применимо к электронам в кристалле и что импульс электрона не равен $m\mathbf{v}$. При этом вводится в рассмотрение так называемый квази-импульс \mathbf{P} , определяемый в § 5 с помощью волновых функций электронов. Как мы увидим в гл. 7, эта векторная величина удовлетворяет уравнению $d\mathbf{P}/dt = \mathbf{F}$, если на электрон действует сила \mathbf{F} . Это соотношение оправдывает введение наименования „импульс“ в соответствии со вторым законом Ньютона. В отличие от импульса свободной частицы квази-импульс электрона в известном смысле ограничен по своей величине. Вследствие повторяемости или периодичности потенциального поля в кристалле, в котором движется электрон, нет необходимости рассматривать значения импульса, большие чем те, которые лежат в определенной области трехмерного \mathbf{P} -пространства, называемого *зоной Бриллюэна*.

Энергия и скорость электрона определяются значением \mathbf{P} внутри зоны Бриллюэна. Однако это определение не однозначно, так что для каждого данного значения \mathbf{P} имеется большое число возможных квантовых состояний, каждое с характерными для него значениями энергии и скорости. Отдельные параграфы гл. 5 посвящены рассмотрению этих понятий; математические же детали вопроса исследуются в части III настоящей книги. Поведение электрона в кристалле описывается посредством определения его энергии, скорости и импульса, однако связь этих величин друг с другом в этом случае иная, чем в классической механике.

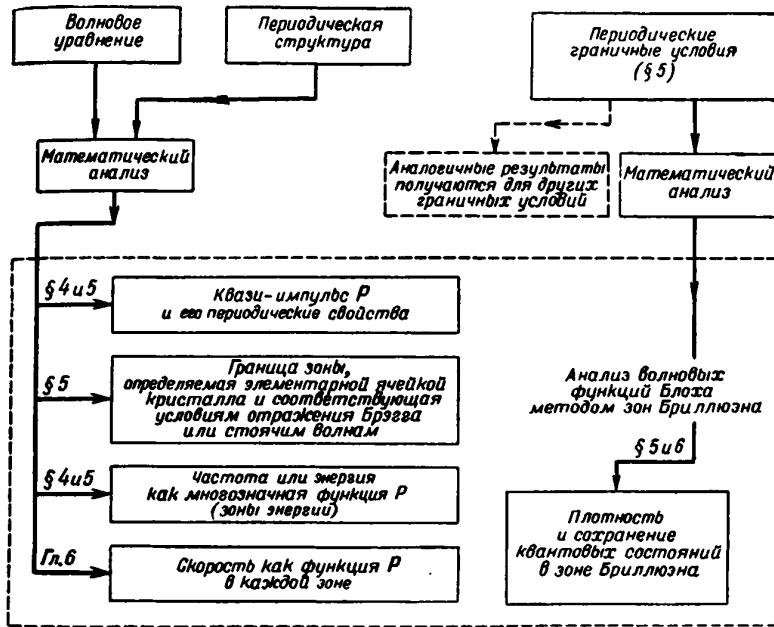
Схемы на фиг. 50, 63 и 64 могут помочь читателю следить за ходом рассуждений. Они дают логическую последовательность этих рассуждений и схему развития понятия зон Бриллюэна при движении электрона в идеальном кристалле¹⁾ в отсутствие внешних сил. На них приведены номера соответствующих параграфов книги.

Если на электрон действует внешняя сила, то из волнового уравнения следует, что квази-импульс электрона изменяется согласно закону²⁾ $d\mathbf{P}/dt \equiv \dot{\mathbf{P}} = \mathbf{F}$. При анализе проводимости и эффекта Холла необходимо рассматривать силы, обусловленные электрическими и маг-

¹⁾ Это требование идеальности кристалла исключает из рассмотрения тепловые колебания или атомы примеси, а следовательно, и случайные процессы, обуславливающие сопротивление. — *Прим. авт.*

²⁾ Иногда для обозначения производной по времени (d/dt , или $\partial/\partial t$) мы будем пользоваться просто точкой над соответствующей величиной, например $\frac{d\mathbf{P}}{dt} \equiv \dot{\mathbf{P}}$. — *Прим. авт.*

нитными полями. На фиг. 68 (стр. 205) указаны последовательные этапы вывода основного закона ускорения вместе с некоторыми следствиями из этого закона. Эти вопросы рассматриваются в гл. 7, посвященной динамике электронов в идеальной решетке.



Фиг. 50. Анализ волновых функций Блоха для электронов в кристаллах по методу зон Бриллюэна.

Зависимости между энергией \mathcal{E} , скоростью v и импульсом P и закон действия силы $\dot{P} = F$ позволяют исследовать влияние приложенных полей на движение электрона. Ситуация здесь совершенно такая же, как и в классическом случае, с той только разницей, что соотношения между \mathcal{E} , v и P носят иной характер. Это различие существенно при объяснении процессов, протекающих в электрических проводниках. Основываясь на новых зависимостях между \mathcal{E} , v и P , можно проанализировать поведение в кристалле избыточного электрона и дырки. Результаты этого анализа оправдывают возможность описания дырок и электронов как классических частиц (см. гл. 1). Этот анализ составляет основное содержание гл. 5, 6 и 7.

В этих главах исследуется роль случайных процессов, вызывающих электрическое сопротивление и устанавливающих тепловое равновесие.

Эти процессы обусловлены нарушениями, делающими кристалл неидеальным, и, следовательно, их рассмотрение автоматически исключено из гл. 5 и 7, в которых мы ограничиваемся рассмотрением идеального кристалла, в котором атомы не совершают теплового движения и расположены точно в соответствующих им местах решетки.

§ 2. КВАНТОВЫЕ СОСТОЯНИЯ, СПИН И ПРИНЦИП ПАУЛИ

Решения волновых уравнений. В настоящее время принято рассматривать электрон как очень маленькую частицу, законы движения которой описываются волновым уравнением Шредингера ¹⁾. Уравнение Шредингера в соединении с принципом Паули удовлетворительно объясняет большинство явлений в физике и химии твердого тела. Мы покажем, как применение этого уравнения приводит к теории энергетических зон, волновым функциям Блоха и зонам Бриллюэна. Но прежде всего в качестве необходимого введения кратко рассмотрим квантовую теорию одноэлектронной системы ²⁾.

При движении электрона в трехмерном пространстве (где его положение определяется координатами x, y, z) под действием сил, которые могут быть охарактеризованы потенциальной энергией $\mathcal{U}(x, y, z)$, возможные значения его энергии \mathcal{E} (не зависящие от x, y, z) и его движения описываются волновым уравнением Шредингера

$$\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} \right) + (\mathcal{E} - \mathcal{U})\psi = 0, \quad (5.1)$$

где \hbar — постоянная Планка, m — масса электрона, \mathcal{E} — энергия электрона, $\psi = \psi(x, y, z)$ — неизвестная волновая функция. Мы обсудим смысл волновой функции ψ несколько позднее, после того как сопоставим это уравнение с некоторыми другими уравнениями математической физики ³⁾. Например, волновое уравнение для скалярного или векторного потенциала электромагнитной волны с частотой ν имеет вид

$$\frac{c^2}{4\pi^2} \left(\frac{\partial^2\mathbf{A}}{\partial x^2} + \frac{\partial^2\mathbf{A}}{\partial y^2} + \frac{\partial^2\mathbf{A}}{\partial z^2} \right) + \nu^2\mathbf{A} = 0. \quad (5.2)$$

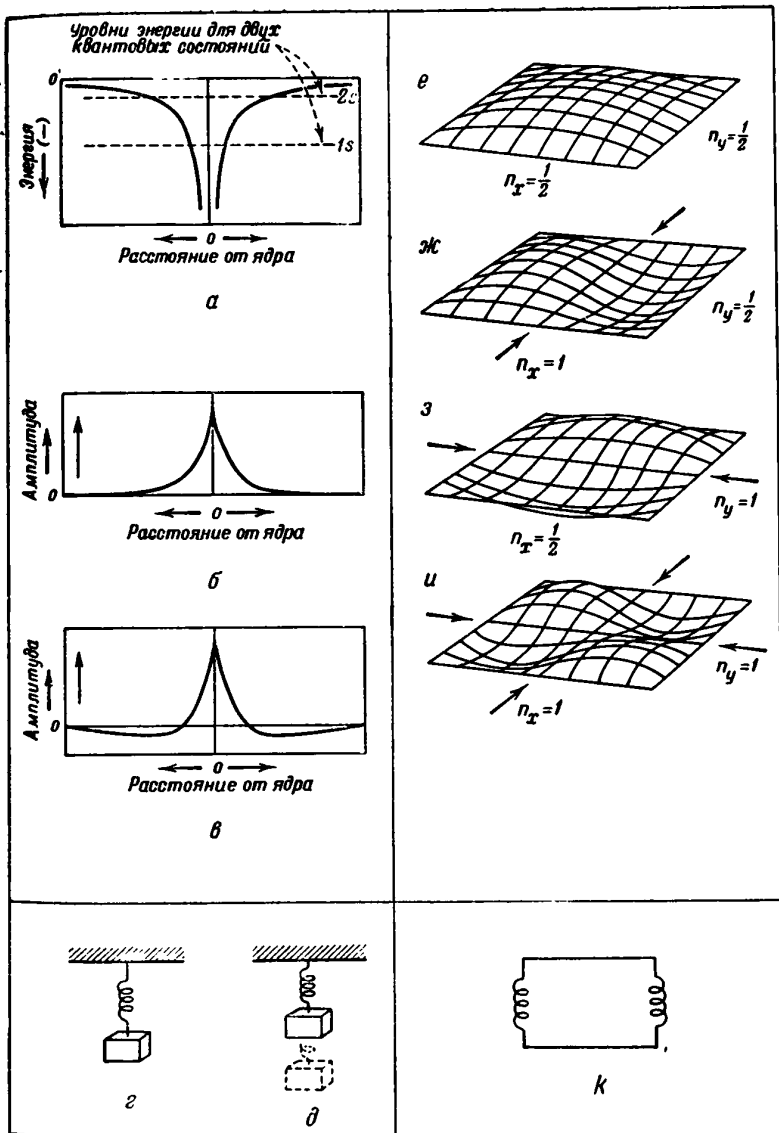
На фиг. 51, $e - u$ показана натянутая мембрана (барабанная мембрана), вертикальное смещение которой ϕ удовлетворяет волновому уравнению

$$\frac{T}{4\pi^2\sigma} \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} \right) + \nu^2\phi = 0, \quad (5.3)$$

¹⁾ Известно, что это уравнение неточно, но оно применимо в тех случаях, когда скорость электрона мала по сравнению со скоростью света. Оно находится в таком же соотношении с уравнением Дирака, как законы движения Ньютона с релятивистскими уравнениями Эйнштейна. — *Прим. авт.*

²⁾ Это рассмотрение представляет собой сокращенную форму теории, изложенной автором в его статье [1]. Большая часть материала, первоначально предназначавшегося для второй части этой статьи, не написанной вследствие войны, изложена в настоящей книге. — *Прим. авт.*

³⁾ Более детальное исследование решений уравнения Шредингера, а также его сопоставление с линией электропередачи дано в гл. 14. — *Прим. авт.*



Фиг. 51. Волновые функции электрона и некоторые механические и электрические аналоги.

a — потенциальная энергия электрона в атоме на разных расстояниях от ядра; *б* — волновая функция для квантового состояния $1s$; *в* — волновая функция для квантового состояния $2s$; *г* и *д* — механическая колебательная система с одним нормальным колебанием; *е* — *и* — прямоугольная барабанная мембрана, закрепленная по краям. Квантовые числа n_x и n_y определяют волновые функции; *к* — электрический контур со многими нормальными колебаниями и результирующими частотами, соответствующими волновым функциям и энергиям электрона.

где σ — масса, приходящаяся на единицу площади, T — поверхностное натяжение и ν — частота. Это уравнение имеет решения, подчиняющиеся граничному условию $\psi = 0$ на прямоугольной границе, где закреплена мембрана, только для некоторых частот, называемых *нормальными колебаниями*, или *собственными частотами* мембраны. Эти нормальные колебания могут быть описаны рядом „квантовых чисел“ n_x и n_y , как это показано на фиг. 51. Квантовые числа могут быть выбраны в известной мере произвольным образом. На фиг. 51, e — u они выбраны в соответствии с числами, используемыми нами ниже при рассмотрении зон Бриллюэна. Эти квантовые числа определяются как числа длин волн, укладывающихся в каждом из двух направлений, так что каждое из них равно $1/2 + 1/2$ (число узловых линий; см. жирные стрелки на рисунке).

Другой системой, нормальные колебания которой могут быть знакомы читателю, является отрезок линии передачи с дросселями между ее концами (см. фиг. 51, k). Эта система и простой гармонический осциллятор (см. фиг. 51, z и d) используются в дальнейшем при исследовании результатов сближения изолированных атомов в процессе образования кристалла.

На фиг. 51, a изображен ход потенциальной энергии \mathcal{U} электрона в атоме водорода. Эта энергия равна нулю, когда электрон находится далеко от протона; следовательно, $\mathcal{U} = -e^2/r$, где e — заряд электрона и r — расстояние от протона. Если это значение потенциальной энергии подставить в уравнение Шредингера, то решения ψ , удовлетворяющие надлежащим граничным условиям (в этом случае $\psi = 0$, если r приближается к бесконечности), могут быть получены только для некоторых значений \mathcal{E} . Эти решения называются „собственными функциями“, а соответствующие энергии — „собственными значениями“, или „собственными энергиями“. Такие волновые функции (две из них показаны на фиг. 51, b и v) могут быть описаны с помощью ряда квантовых чисел, подобно тому как это делается для прямоугольной барабанной мембраны. Например, число 1 в случае состояния $1s$ и число 2 в случае состояния $2s$ как раз на единицу больше числа переходов функции $\psi(r)$ через нуль.

Свойством уравнения Шредингера, которое необходимо отметить, является то, что для введения в рассмотрение квантовых чисел не требуется каких-либо специальных предположений¹⁾. Эти числа получаются точно таким же путем, как и в случае прямоугольной барабанной мембраны, изображенной на фиг. 51, или в случае нормальных

¹⁾ Возможна известная свобода выбора численных значений, приписываемых квантовым числам. Например, на фиг. 51 n_x и n_y могли бы быть выбраны как числа полувольт, укладывающихся вдоль каждого края мембраны. Однако мы хотим особо отметить, что при исходном введении квантовых чисел не делается никаких специальных предположений. Напротив, в старой теории Бора квантовые числа вводятся как специальное предположение. — *Прим. авт.*

колебаний полого резонатора. Математический анализ всех этих задач очень сходен и связан с решением дифференциального уравнения, подчиненного известным граничным условиям. Детальное исследование собственных волновых функций показывает, что в направлениях x , y и z (или в трех других направлениях, в зависимости от граничных условий) укладывается некоторое определенное число длин волн или полуволен. Эти числа длин полуволен и представляют собой, таким образом, естественно возникающие квантовые числа, которые могут быть использованы для характеристики решений волнового уравнения. Для электрона в атоме водорода наиболее подходящие значения квантовых чисел связаны простым образом с числом длин полуволен (или узлов волновой функции) в направлениях r , θ и φ . Для наших целей, однако, нет необходимости детально выяснять смысл квантовых чисел электронов в атомах. Значительно больше нас будут интересовать, однако, квантовые числа, применяемые для описания движения электронов в кристаллах, и их связь с числами длин волн рассматриваемых волновых функций (таким образом мы введем понятие квази-импульса P).

В связи с исследованием движения электронов в кристаллах мы будем иметь в дальнейшем возможность сравнить поведение волновых функций электрона с поведением электромагнитных или механических волн. При этом будет сопоставлен также физический смысл волновых функций в обоих случаях.

Согласно квантовой механике, волновая функция ψ есть „амплитуда вероятности“ (выражение, появившееся в квантовой механике и нигде больше не встречающееся), а квадрат ее абсолютного значения, обозначаемый через $|\psi|^2 = \psi^* \psi$, есть плотность вероятности. Плотность вероятности можно интерпретировать следующим образом: электрон, описываемый волновой функцией, нужно представлять себе как очень маленькую частицу, размеры которой ничтожно малы по сравнению с размерами атома. Он сам при этом не распространяется за объем, на который простирается волновая функция. Однако при своем движении электрон перемещается в пространстве вблизи атома, проводя различные промежутки времени в каждой области. Плотность вероятности в каждой точке пропорциональна времени, в течение которого электрон находится в малом объеме вблизи этой точки. В действительности нормирующий множитель волновой функции обычно выбирается так, что $|\psi|^2 dV$ есть просто доля времени, которую электрон проводит в объеме dV . Следовательно, $|\psi|^2 dV$ есть вероятность нахождения электрона в объеме dV (ψ , таким образом, имеет размерность $см^{-3/2}$). Если эти вероятности суммируются по всему пространству, то сумма равна единице, так как электрон должен где-либо находиться. Говорят, что волновая функция нормирована, если в результате такого суммирования мы получаем единицу.

Следовательно, $|\psi|^2$ дает усредненное по большому промежутку времени распределение плотности заряда, обусловленное электроном.

Именно такие усредненные распределения плотности заряда и изображены на фиг. 2, 3, 5 и 15 (они заменяют боровские орбиты старой квантовой теории).

Электрон так быстро меняет свое положение, что такого рода усреднением можно пользоваться во многих случаях. Однако в других случаях, таких, например, как случай сил Ван дер Ваальса, необходимо учитывать флюктуации положения электрона. Другие характеристики поведения электрона, как, например, его средняя скорость, могут быть получены с помощью волновой функции (см. гл. 6 и 14).

Плотность вероятности ведет себя в некотором отношении аналогично плотности энергии электромагнитных или механических колебательных систем, если эти последние не имеют потерь, так что энергия колебания остается неизменной. В этих случаях напряжения и токи (или электрическое и магнитное поля), а также смещение и скорость являются амплитудами. Плотности энергии пропорциональны квадратам этих амплитуд. Однако имеется и существенное отличие: при рассмотрении электромагнитных или механических систем решения обычно получаются в комплексной форме. В этих случаях вещественная и мнимая части решения, каждая в отдельности, имеют одинаковое самостоятельное значение (скорее в физическом, чем в математическом смысле) как решения задачи. Каждое из этих решений соответствует переходу энергии электрической в магнитную или кинетической в потенциальную и обратно — дважды за каждый цикл. В квантовомеханическом случае волновая функция также является комплексной; ее зависимость от времени имеет вид $e^{i\mathcal{E}t/\hbar}$, где \mathcal{E} — энергия, t — время и \hbar — постоянная Планка. Но ни вещественная, ни мнимая части ψ в отдельности не являются решением уравнения Шредингера. Решением является только сама комплексная волновая функция. Таким образом, значение $|\psi^2|$ не зависит от времени и представляет неизменяющееся распределение вероятности¹⁾. Это вполне естественно, так как в отличие от плотности энергии, которая может существовать в двух формах (как в электромагнитном, так и в механическом случаях), плотность вероятности может быть представлена лишь в одной форме.

Как и в случае классических волн, так и в квантовомеханическом случае иногда бывает необходимо получить новую волновую функцию путем сложения других волновых функций. Это сложение возможно в силу принципа суперпозиции, вытекающего из линейности рассматриваемых дифференциальных уравнений. Исследование волновых функций, полученных путем суперпозиции, особенно существенно для

1) В то время как $|\psi^2|$ никогда не зависит от времени, для нормального вида решений или собственных решений уравнения Шредингера имеются другие решения, при которых $|\psi^2|$ меняется со временем, и распределение вероятности перемещается в пространстве. Такие случаи рассмотрены в связи с исследованием волновых пакетов в § 3 гл. 6. — *Прим. авт.*

понимания законов движения свободного электрона или же движения электрона в кристалле; мы вернемся к этому в § 3 гл. 6.

Важным свойством нормальных колебаний любой системы, такой, например, как барабанная мембрана, является то, что любое свободное колебание этой системы может быть описано с их помощью. Таким образом, если барабанную мембрану ударить и затем предоставить ей возможность свободно колебаться, то ее движение будет описываться не одним из нормальных колебаний (см. фиг. 51), а совокупностью ряда нормальных колебаний, каждое из которых обладает своей собственной частотой, не зависящей от частот других колебаний, и полное смещение определяется суммой составляющих смещений. Возмущение такого рода характеризуется не одной, а рядом частот и удовлетворяет волновому уравнению, содержащему производную по времени

$$\frac{T}{\sigma} \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \right) - \frac{\partial^2 \phi}{\partial t^2} = 0. \quad (5.4)$$

Решение этого уравнения получается путем сложения решений для нормальных колебаний. Если $\phi_1(x, y)$, $\phi_2(x, y)$, ... решения для нормальных колебаний с частотами f_1 , f_2 , ..., то общее решение уравнения, зависящего от времени, будет

$$\phi(x, y, t) = A_1 \phi_1(x, y) \sin 2\pi f_1(t - t_1) + \\ + A_2 \phi_2(x, y) \sin 2\pi f_2(t - t_2) + \dots, \quad (5.5)$$

где A_1 , A_2 , ... и t_1 , t_2 , ... — постоянные. Для каждого из этих членов в отдельности уравнение, зависящее от времени, сводится к уравнению, содержащему f^2 и не содержащему производных по времени.

Соответствующее уравнение Шредингера, зависящее от времени, имеет вид

$$\frac{\hbar^2}{8\pi^2 m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) - \mathcal{U} \psi + \frac{i\hbar}{2\pi} \frac{\partial \psi}{\partial t} = 0. \quad (5.6)$$

Решение этого уравнения также может быть найдено путем суммирования собственных функций, каждая из которых при этом умножается на член, зависящий от времени¹⁾. В гл. 6 будут рассмотрены построенные таким образом волновые функции, в связи с исследованием волновых пакетов, характеризующих движение электронов в кристалле. В этой главе, однако, мы будем иметь дело только с решениями в виде собственных функций.

Спин и квантовые состояния. Помимо перемещения в пространстве, электрон вращается также вокруг собственной оси. Этому

¹⁾ Такие зависящие от времени члены рассматриваются в гл. 14, а также используются в уравнениях (5.8) и (5.11). — *Прим. авт.*

вращению соответствует четвертая степень свободы и четвертое квантовое число — *спин*. На основании соображений, связанных с теорией относительности, найдено, что четвертое квантовое число может иметь только два дозволенных значения: $+\frac{1}{2}$ и $-\frac{1}{2}$. При рассмотрении сил сцепления между атомами большие значения имеют так называемые *обменные* эффекты, связанные со спином. Спин электрона обуславливает также магнитный момент, ответственный за ферромагнетизм. Однако при исследовании законов движения электронов в полупроводниках мы будем учитывать только тот факт, что спин обуславливает дополнительную степень свободы электрона, так что последний может двигаться в соответствии с данной волновой функцией с положительным или отрицательным значением спина.

Введем теперь понятие *квантового состояния*.

Квантовое состояние описывает возможный тип движения электрона. Оно определяется совокупностью четырех квантовых чисел: три из них описывают волновую функцию, а четвертое — спин.

До сих пор речь шла о движении отдельного электрона. Если одновременно имеется несколько электронов, испытывающих электростатическое отталкивание друг от друга, то уже неправильно представлять себе, что каждый электрон движется в постоянном силовом поле, и, следовательно, неправильно говорить о квантовых состояниях отдельных электронов. Вместо этого следует говорить о квантовом состоянии системы в целом. Если число электронов велико, то можно воспользоваться некоторыми приближенными методами расчета. Согласно этим методам, при рассмотрении одного электрона действие других усредняется и вычисляется создаваемое ими силовое поле. Затем исследуется движение электрона в этом усредненном силовом поле и определяется его квантовое состояние. Аналогичный процесс осуществляется для всех остальных электронов. (Этот процесс обычно повторяется до тех пор, пока не будет получено „самосогласованное“ решение, т. е., исходя из совокупности пробных волновых функций электронов, определяется силовое поле, создаваемое другими электронами и действующее на каждый отдельный электрон. Определив эти силовые поля, находим новый ряд волновых функций. Этот процесс повторяется до тех пор, пока не будет получена совокупность волновых функций, соответствующая силовому полю, которое снова дает ту же совокупность волновых функций.) Для той степени приближения, которая принята в теории полупроводников, это усреднение силового поля считается удовлетворительным, и поведение каждого электрона описывается заданием квантового состояния, которое он занимает.

Принцип Паули. Используя понятие квантового состояния, можно сформулировать „принцип Паули“, согласно которому *два электрона не могут занимать одно и то же квантовое состояние.*

Применяя принцип Паули, следует помнить о необходимости правильного определения квантовых состояний. Например, принцип Паули вовсе не требует, чтобы при наличии одного электрона в наименьшем энергетическом квантовом состоянии в каком-либо атоме водорода никакой другой электрон во вселенной не мог бы находиться в том же квантовом состоянии в другом атоме водорода. В действительности следует ожидать, что в газе, состоящем из атомов водорода, половина электронов находится в наименьшем квантовом состоянии с одним спином, а другая половина — в состоянии с противоположным спином. Этот кажущийся парадокс исчезает, если мы учтем, что при наличии двух атомов схема квантовых состояний удваивается. В действительности для определения волновых функций для системы, состоящей из двух ядер, необходим новый ряд квантовых чисел. Однако найдено, что если атомы находятся на большом расстоянии друг от друга, то волновые функции имеют такой же вид, как и в случае изолированного атома, но фигурируют дважды: в одном случае за начало координат выбрано ядро одного атома, в другом — ядро второго атома. Другими словами, два отдельных атома не находятся в одних и тех же квантовых состояниях. Для каждого из них волновые функции имеют совершенно одинаковую форму и соответствуют определенному ряду квантовых состояний.

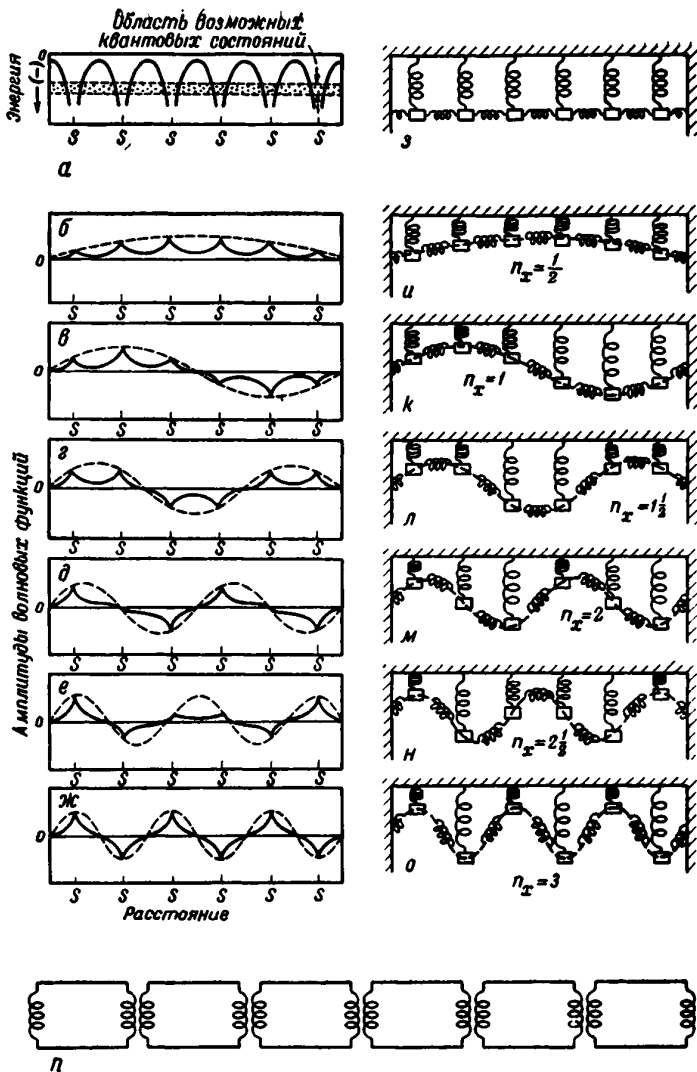
Математическая формулировка принципа Паули заключается в утверждении, что волновая функция, описывающая систему электронов, должна быть *антисимметрична*. Смысл этого термина и различных других математических формулировок принципа Паули рассматривается в § 7 гл. 15. Однако для понимания сущности принципа Паули нет необходимости изучать математическую теорию данного вопроса.

§ 3. КВАНТОВЫЕ СОСТОЯНИЯ И ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ В КРИСТАЛЛАХ

Существуют два основных метода определения системы квантовых состояний твердого тела. Метод усредненного поля, рассмотренный в предыдущем параграфе, применяется для описания движения отдельных электронов в твердом теле. Им необходимо пользоваться при определении импульса, ускорения и скорости носителей зарядов в полупроводниках. Другой метод позволяет найти квантовое состояние в твердом теле, исходя из квантовых состояний изолированных атомов, образующих это тело. Сначала рассмотрим второй метод ¹⁾.

Для этого представим себе, что N_s атомов твердого тела образуют идеально правильную кристаллическую решетку, но расстояния между атомами во много раз превышают нормальную постоянную

¹⁾ В дальнейшем мы в известной мере используем материал, изложенный в нашей прежней работе [1]. — *Прим. авт.*



Фиг. 52. Одномерный кристалл и некоторые механические и электрические аналоги.

В силу законов волновой механики электрон может проходить через область, где его потенциальная энергия больше его полной энергии, как это показано на фиг. а. В этой области волновая функция электрона быстро спадает подобно электромагнитному полю в волноводе при частоте, ниже граничной.

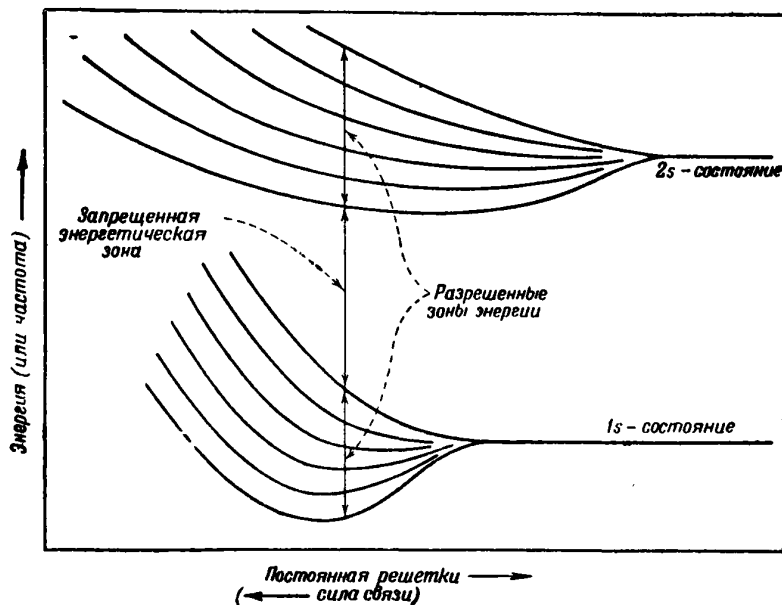
а — потенциальная энергия электрона; б — ж — волновые функции для электрона в кристалле, соответствующие квантовому состоянию 1s; з — шесть связанных механических осцилляторов (в покое); и — о — шесть нормальных колебаний системы связанных механических осцилляторов; л — связанные электрические контуры с шестью нормальными колебаниями в каждой полосе пропускания частот; буквами S обозначены положения ядер.

решетки, так что какое-либо взаимодействие между ними отсутствует¹⁾. Распределение квантовых состояний кристалла представляет собой в этом случае повторенное N_s раз распределение квантовых состояний атома. Это значит, что каждый уровень энергии, соответствующий атомному квантовому состоянию, встречается N_s раз. Однако при уменьшении постоянной решетки положение меняется, и если атомы расположены настолько близко друг от друга, что волновые функции соседних атомов сильно перекрываются, то схема квантовых состояний и энергетических уровней претерпевает сильное изменение. Результат этого изменения представлен на фиг. 52. На фиг. 52, *з* — *о* показано влияние взаимодействия атомов на примере системы одномерных механических осцилляторов. Взаимодействие между атомами представлено здесь слабыми пружинками, связывающими осцилляторы друг с другом. Если связь отсутствует, то каждый осциллятор колеблется со своей собственной нормальной частотой. Частота колебаний всех осцилляторов при этом одинакова. Коль скоро возникает связь, в колебаниях принимают участие уже не отдельные осцилляторы, а вся система в целом. Однако если связь слаба, каждое из колебаний будет иметь почти такую же частоту, как и изолированный осциллятор. По мере того, как связь становится сильнее, полоса частот захватывает все более широкую область. Весьма сходным образом ведут себя атомы. При перекрытии волновых функций квантовые состояния уже относятся не к отдельным атомам, а ко всему кристаллу в целом. Это показано на фиг. 52, *б* — *ж*. Уровни энергии, соответствующие квантовым состояниям, расщепляются подобно тому, как это имеет место для частот колебаний системы осцилляторов. Таким образом, каждый атомный уровень энергии дает полосу уровней энергий, или *энергетическую зону* кристалла. Это показано на фиг. 53, построенной на основании качественного анализа фиг. 52.

В процессе расщепления полное число степеней свободы механических осцилляторов и число квантовых состояний атомов остаются неизменными. Например, если бы шесть атомов, изображенных на фиг. 52, содержали двенадцать электронов, занимавших наинизшие квантовые состояния, то в зоне энергетических уровней кристалла, возникшего в результате взаимодействия этих атомов, число квантовых состояний оказалось бы достаточным, чтобы разместить все эти электроны. Мы будем иметь шесть волновых функций, каждая из которых, благодаря наличию спина, соответствует двум квантовым состояниям.

¹⁾ Кристаллическая решетка дает относительное расположение атомов. Структуру алмаза, изображенную на фиг. 4, имеют углерод, кремний, германий и одна из модификаций олова (серое олово). Обычно под „постоянной решетки“ подразумевают ее значение, наблюдаемое на опыте. Однако в этой главе мы употребляем термин „постоянная решетки“ для обозначения переменного расстояния между атомами, а термин „нормальная постоянная решетки“ — для постоянной, наблюдаемой на опыте. — *Прим. авт.*

Системам с большим числом атомов будет отвечать большое число волновых функций. Однако одна из этих функций, соответствующая наименьшей энергии, будет слабо модулирована, как это изображено на фиг. 52, б пунктирной линией, волновая же функция, соответствующая наибольшей энергии, будет изображаться фиг. 52, ж¹⁾. Следовательно, высшие и низшие значения энергии электронов или значения частот нормальных колебаний системы осцилляторов не будут



Фиг. 53. Зависимость энергетических уровней от постоянной решетки (или частоты колебаний от силы связи).

зависеть от размеров системы, если только число составляющих ее элементов достаточно велико. Только что сделанные выводы можно сформулировать в виде двух основных теорем:

Теорема о ширине зоны. *Ширина энергетической зоны, возникающей из атомного энергетического уровня, не зависит от числа атомов в кристалле²⁾.*

¹⁾ В этой главе нас особенно будут интересовать волны, подобные изображенным на фиг. 52, ж, длина волн которых равна удвоенному периоду структуры a . Такие волны соответствуют краям зоны Бриллюэна. — *Прим. авт.*

²⁾ Если мы представим себе, что постоянная решетки может ме-

ТЕОРЕМА СОХРАНЕНИЯ ЧИСЛА КВАНТОВЫХ СОСТОЯНИЙ. *Число квантовых состояний в энергетической зоне равно числу атомных квантовых состояний, из которых возникла зона.*

Не существует квантовых состояний с энергиями, лежащими между энергетическими зонами. Поэтому область значений энергии, заключенных между верхним краем одной зоны и дном следующей зоны, называется *полосой запрещенных энергий*, или *запрещенной зоной*.

Энергетические зоны, содержащие возможные квантовые состояния, называются *разрешенными энергетическими зонами*.

Там, где не может возникнуть путаницы, они обычно называются просто *зонами энергии*, или *энергетическими зонами*.

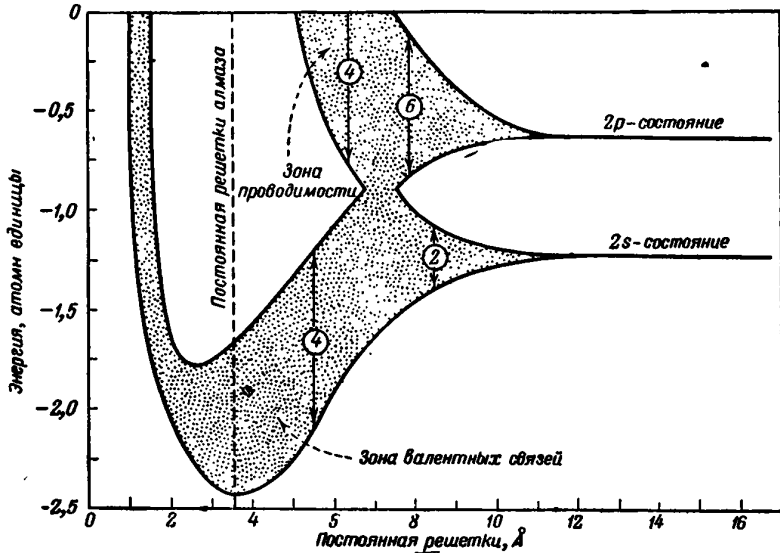
В случае углерода представляющие интерес энергетические зоны возникают из квантовых состояний, занятых валентными электронами.

Эти состояния обозначаются как $2s$ - и $2p$ -состояния и содержат в общей сложности восемь квантовых состояний на атом. Поэтому в изолированных атомах они заняты только наполовину. Каждый атом содержит, кроме того, два электрона в состоянии $1s$. Однако эти электроны сильно связаны с ядром и поэтому волновые функции таких электронов, находящихся в смежных атомах, не перекрываются заметным образом. На этом основании можно считать, что на эти электроны пребывание в кристалле не оказывает влияния. Зоны энергии углерода [2] с решеткой алмаза представлены на фиг. 54.

При больших значениях постоянной решетки нижняя зона заполнена, а верхняя заполнена лишь частично. Однако при меньших межатомных расстояниях восемь состояний, приходящихся на каждый атом, распределяются поровну между двумя зонами, так что нижняя зона оказывается целиком заполненной. Эта нижняя зона соответствует электронам, участвующим в образовании валентных связей. Для возбуждения электрона необходима энергия, достаточная для перехода его в квантовое состояние в верхней зоне. Следует отметить, что среднее значение энергии электрона в нижней зоне, соответствующее истинному значению постоянной решетки, много меньше, чем в случае свободных атомов. Это уменьшение энергии связано с существованием *энергии связи* кристалла. Сравнение полной энергии изолированных атомов углерода с полной энергией тех же атомов, образующих кристалл, показывает, что в результате уменьшения энергии валентных электронов их состояния делаются более устойчивыми.

ваться, как показано на фиг. 53, то ширина зоны будет зависеть от постоянной решетки. Она определяется уровнем энергии, из которого возникла зона. — *Прим. авт.*

Энергия, необходимая для разделения кристалла на свободные атомы, называется *энергией связи*. При рассмотрении энергии связи валентных кристаллов, подобных алмазу, приближения зонной теории дают, вероятно, менее точный результат, чем непосредственное рассмотрение парно-электронной связи. Однако в квантовомеханической теории нас



Ф и г. 54. Структура энергетических зон алмаза в зависимости от постоянной решетки.

За единицу энергии взят 1 ридберг = $13,5 \text{ эв} = \frac{1}{2}$ атомной единицы. Цифрами в кружках указано число квантовых состояний на атоме.

в большей мере интересует проблема электропроводности, чем проблема сил сцепления. При исследовании же электропроводности зонная теория является не только лучшим приближением, имеющимся в настоящее время, но, повидимому, также и правильно отображает истинную картину явления ¹⁾.

¹⁾ Критический анализ основ зонной теории твердого тела, проведенный за последние годы советскими физиками-теоретиками, показал, что к зонной теории следует относиться как к *качественной* теории, описывающей некоторые принципиальные черты кристаллического состояния, но недостаточной для детальных количественных выводов. Исходные предпосылки зонной теории и вопрос о пределах ее применимости рассмотрены в работах [11—19] и др.

Инициатива дальнейшего развития более совершенной теории твердого тела принадлежит также советским физикам [18, 20—24]. — *Прим. ред.*

§ 4. ВОЛНОВЫЕ ФУНКЦИИ БЛОХА В ПЕРИОДИЧЕСКОМ ПОЛЕ

В предыдущем параграфе указывалось, что энергетические уровни системы изолированных атомов расщепляются в энергетические зоны, когда атомы сближаются настолько, что образуется кристалл; были сформулированы также теоремы о ширине зоны и сохранении числа квантовых состояний. В этом параграфе будут рассмотрены волновые функции, полученные при решении волнового уравнения электрона, движущегося в периодическом поле. Это послужит в дальнейшем основанием для рассмотрения скорости и ускорения электронов.

Решения уравнения Шредингера для электрона в кристалле определяются:

1) видом поля, в котором движется электрон, т. е. значением $\mathcal{U}(x, y, z)$ в каждой точке элементарной ячейки кристалла (так как кристалл периодичен, значения потенциала будут повторяться в каждой элементарной ячейке);

2) размерами кристалла;

3) граничными условиями на поверхности кристалла.

Как указывалось в § 2, для решения задачи необходимо ввести ряд квантовых чисел, которые могут быть использованы для определения волновых функций и соответствующих характеристических значений энергии.

Для рассмотрения бегущих электронных волн, переносящих ток через кристалл, введем граничное условие, отличное от того, которое было неявно введено в связи с фиг. 52. Как будет видно ниже, это изменение граничных условий не повлияет на теорему о сохранении числа квантовых состояний.

При граничном условии, принятом на фиг. 52, волновая функция обращается в нуль на концах одномерного кристалла. Такие граничные условия будем называть *граничными условиями исчезновения*. Эти граничные условия математически эквивалентны наличию бесконечно большого поля за пределами кристалла, отражающего электроны внутрь последнего. При этом электронная волна является стоячей волной, которую можно рассматривать как наложение двух бегущих волн. Хотя каждая из бегущих волн в отдельности удовлетворяет уравнению Шредингера для некоторого потенциала \mathcal{U} , она не является, однако, решением, удовлетворяющим граничным условиям. Соответственно нормальные колебания связанных осцилляторов удовлетворяют граничным условиям, согласно которым точки закрепления крайних пружин ряда остаются неподвижными; модулированные волны, обозначенные на рисунках пунктиром, имеют по той же причине узлы на концах. При таких граничных условиях распространение электронных волн (см. фиг. 52, б—ж) не сопровождается каким-либо изменением плотности вероятности вдоль кристалла; аналогичным образом механические колебания не переносят никакой механической энергии

(см. фиг. 51, $u=0$). При рассмотрении собственных функций или нормальных колебаний, соответствующих случаю переноса энергии, удобнее всего пользоваться периодическими граничными условиями.

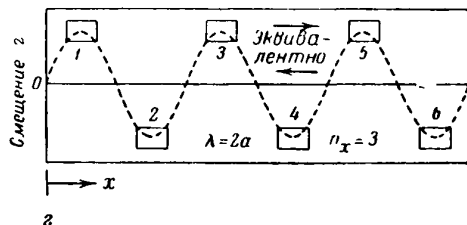
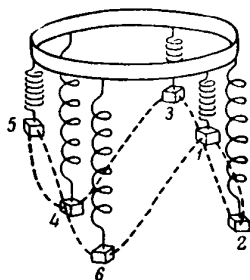
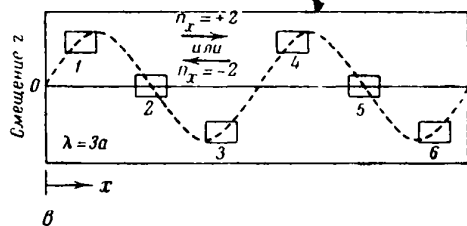
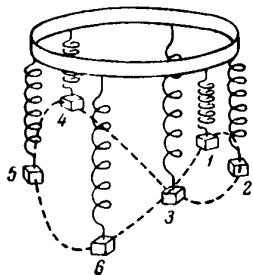
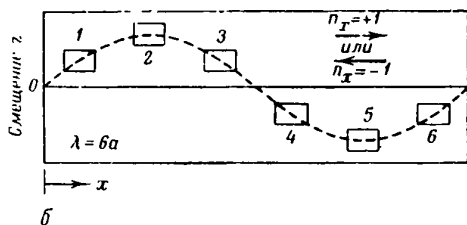
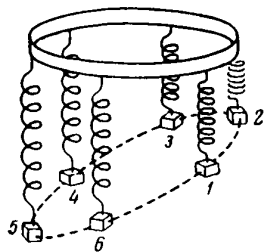
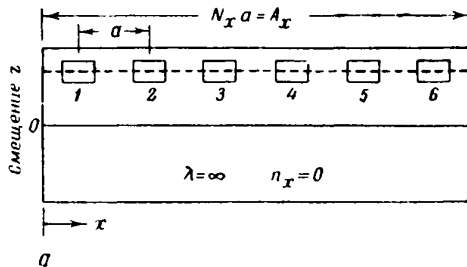
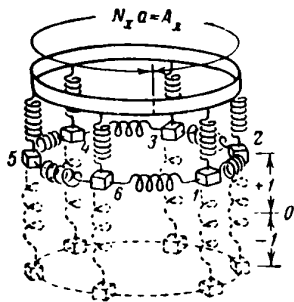
Периодические, или *циклические*, граничные условия сводятся к рассмотрению одномерной модели кристалла, состоящей из атомов, расположенных в виде кольца, или же системы связанных осцилляторов, образующих замкнутое кольцо. Такая модель изображена на фиг. 55. Шесть осцилляторов при помощи пружин подвешены к твердому кольцу и связаны друг с другом также посредством пружинок. Предполагается, что имеется некоторое устройство, позволяющее им перемещаться только в вертикальном направлении. Фиг. 55, *a* соответствует самой низкой частоте колебаний. В этом случае все грузы движутся вместе, и связывающие их пружинки бездействуют. На фиг. 55, *б* движение описывается синусоидальной волновой функцией, длина волны которой равна периметру кольца. Пружины, связывающие грузы 2 и соседние с ним грузы 1 и 3, действуют таким образом, что частота нормального колебания в этом случае будет выше, чем в случае фиг. 55, *a*. (Для упрощения схемы связывающие пружины показаны только на фиг. 55, *a*.) Фиг. 55, *б* соответствуют два нормальных колебания, по одному для каждого направления распространения синусоидальной волны. Частота этих волн будет такой же, как и на фиг. 52, *к*; в самом деле, стоячая волна, изображенная на фиг. 52, *к*, получается вследствие интерференции волн, распространяющихся вправо и влево (см. фиг. 55, *б*). Эти две волны характеризуются квантовыми числами $n_x = +1$ и $n_x = -1$; знак плюс соответствует волне, распространяющейся в направлении положительной полуоси x , а знак минус соответствует волне, распространяющейся в направлении отрицательной полуоси x . Так как в кристалле укладывается только одна длина волны, то $n = 1$. В соответствии с этим условием, мы имеем соотношение $\pm 1/\lambda = n_x/A_x = n_x/N_x a$, которое будет использовано в дальнейшем.

Фиг. 55, *в* отвечает тому случаю, когда вдоль окружности кольца укладываются две длины волны, т. е. $n_x = \pm 2$. При этом также имеются два нормальных колебания (ср. фиг. 52, *б*).

Критическое условие возникает, однако, при $\lambda = 2a$. В этом случае волны, идущие как вправо, так и влево, вызывают одинаковое движение грузов нашей системы; грузы с четными номерами колеблются в противофазе с грузами, обладающими нечетными номерами.

Мы видим, что при периодическом граничном условии фиг. 55, в общей сложности имеется всего шесть нормальных колебаний, т. е. столько же, сколько их было для граничного условия с „закрепленными концами“ на фиг. 52. Однако собственные частоты в этих двух случаях несколько различны¹⁾. Рассмотрим теперь решения уравнения Шредингера, аналогичные описанным бегущим волнам в системе

¹⁾ Некоторые относящиеся сюда задачи приведены в конце этой главы. — *Прим. авт.*



Фиг. 55. Нормальные колебания связанных осцилляторов с циклическими или периодическими граничными условиями.

а — все осцилляторы движутся вверх и вниз; б, в и г — волна может распространяться в любом направлении.

связанных механических осцилляторов. Некоторое представление о связи между этими двумя задачами можно получить, записав уравнение для смещения, вызываемого синусоидальной волной с длиной волны λ , частотой ν , скоростью c (равной $\lambda\nu$) и единичной амплитудой (см. фиг. 55). Смещение z груза в точке с координатой x , отсчитываемой вдоль окружности, выражается следующим образом:

$$z = \sin 2\pi \left(\frac{x}{\lambda} - \nu t \right) = \sin \frac{2\pi}{\lambda} (x - ct) = \sin 2\pi \nu \left(\frac{x}{c} - t \right). \quad (5.7)$$

Если мы увеличим значение x на величину a , т. е. перейдем от одного груза к следующему, то фазовый угол волны увеличится на $2\pi a/\lambda$. Это эквивалентно замене t на $t - a/c$. Другими словами, каждый груз колеблется совершенно так же, как и предыдущий, с той разницей, что колебание его запаздывает во времени на величину a/c , равную времени, за которое волна перемещается на расстояние одного периода структуры. Скорость таких волн меняется с длиной волны λ (см. гл. 15).

Можно ожидать, что решения уравнения Шредингера для электронных волн в кристалле будут одинаковы для каждой элементарной ячейки, но смещены по фазе в соответствии с временем распространения волны от одной ячейки до следующей. Доказательство того, что решения этого типа являются собственными функциями, принадлежит Блоху [3], и соответствующие волновые функции называются *функциями Блоха*¹⁾. Для одномерного кристалла функция Блоха имеет вид

$$\psi_\lambda(x, t) = u_\lambda(x) \exp \left[2\pi i \left(\frac{x}{\lambda} - \nu_\lambda t \right) \right]. \quad (5.8)$$

Здесь функция $u_\lambda(x)$ одинакова для каждой элементарной ячейки; она периодична, так что

$$u_\lambda(x + a) = u_\lambda(x). \quad (5.9)$$

Таким образом, видно, что функция $\psi_\lambda(x, t)$ в точке $x + a$ ведет себя так же, как и в точке x , по истечении времени, определяемого соотношением

$$\frac{a}{\lambda} = \nu_\lambda t \quad \text{или} \quad t = \frac{a}{\lambda \nu_\lambda}; \quad (5.10)$$

другими словами, волна распространяется вправо с фазовой скоростью $\lambda \nu_\lambda$. Периодическая функция $u_\lambda(x)$ и частота ν_λ зависят от λ , что отмечено соответственно индексами. Вообще $u_\lambda(x)$ — комплексная величина с переменным фазовым углом, хотя при $\lambda = \infty$ она становится вещественной. [При $\lambda = 2a$ функция $u_\lambda(x)$ изменяется как $\exp(-2\pi i x/\lambda)$, при этом ψ_λ в функции от x имеет вид, приведенный

¹⁾ Более детально этот вопрос рассмотрен в гл. 14. — *Прим. авт.*

на фиг. 56, в.] Энергия, соответствующая волновой функции, также зависит от λ и связана с частотой ν_λ уравнением ¹⁾

$$\mathcal{E}_\lambda = h\nu_\lambda, \quad (5.11)$$

в чем легко убедиться, вычислив частную производную по времени с помощью уравнения Шредингера, зависящего от времени. Как всегда, для собственных функций энергии значение $|\psi_\lambda(x, t)|^2$ не зависит от времени. [Это следует из того факта, что в выражении (5.8) от времени зависит только множитель $\exp(-2\pi i\nu_\lambda t)$.]

Остальные параграфы этой главы, а также следующая глава посвящены в основном исследованию волновых функций Блоха $\psi_\lambda(x, t)$. Ниже показано, что значения λ , вернее $\pm h/\lambda = hn_{\omega}/A = hn_{\omega}/N_{\omega}a$, являются „квантовыми числами“, характеризующими состояние в энергетической зоне. Особенно существенное значение приобретают условия, иллюстрируемые фиг. 52, ж—о и фиг. 55, з, при которых разность фаз на протяжении одного периода решетки составляет 180° , или, другими словами, $\lambda = 2a$. (Эти условия определяют границы зоны Бриллюэна, рассматриваемой в следующем параграфе.)

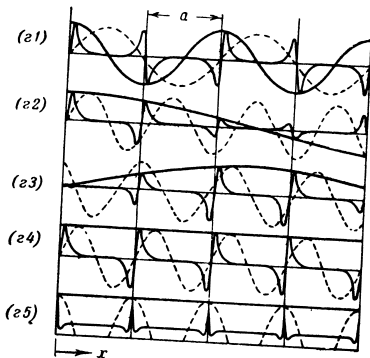
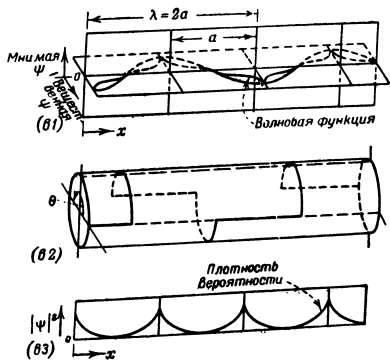
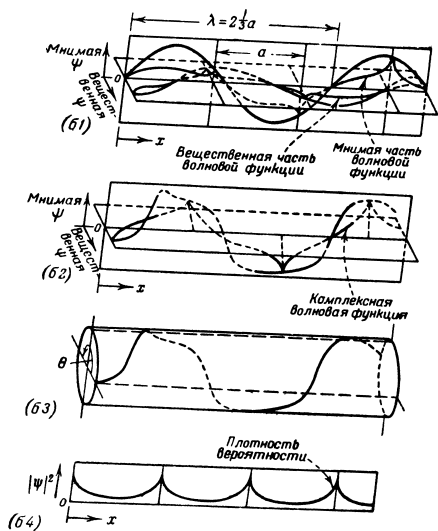
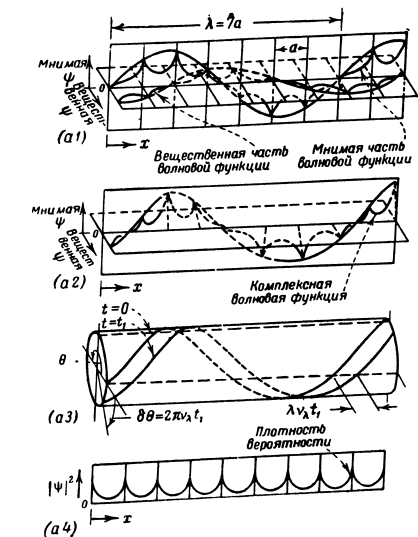
На фиг. 56 (стр. 174—175) изображен ход бегущих волн, описываемых функциями Блоха. Рисунок иллюстрирует свойства волновых функций для трех квантовых состояний, находящихся в наинижней энергетической зоне, и некоторых волновых функций для более высоких энергетических зон.

На фиг. 56, а показан случай относительно большой длины волны, $\lambda = 7a$. Вещественная и мнимая части волновой функции (см. а1) состоят из члена с периодом a , умноженного на синусоидальный множитель. Эти множители связаны друг с другом, как синус и косинус. Из этого следует, что значение $|\psi|^2 = (\text{вещ. } \psi)^2 + (\text{мним. } \psi)^2$ одинаково в каждой элементарной ячейке (см. а4), так как $\sin^2 + \cos^2 = 1$. Кроме того, фазовый угол волновой функции, $\theta = \text{tg}^{-1}(\text{мним. } \psi / \text{вещ. } \psi)$, изменяется линейно от x (см. а3) и поворачивается на 360° через каждые 7 постоянных решеток. С возрастанием t от 0 до t_1 фазовый угол для любого данного значения изменяется на $\delta\theta = -2\pi\nu_\lambda t_1$ (см. а3). Это эквивалентно перемещению волновой функции на $(\delta\theta/2\pi)$ длин волн. Поэтому фазовая скорость будет равна $\lambda(\delta\theta/2\pi)/t_1 = \lambda\nu_\lambda$.

На фиг. 56, б изображена волновая функция при значении $\lambda = (7/3)a$, только немного превышающем критическое значение $\lambda = 2a$. Масштаб графиков увеличен.

Снова синусоидальные множители вещественной и мнимой частей в точках, соответствующих центрам атомов, сдвинуты по фазе на 90° .

¹⁾ Это знаменитое уравнение, являющееся теперь одним из основных уравнений квантовой механики, впервые получено Планком в 1900 г. в связи с исследованием теплового равновесия электромагнитного излучения в замкнутой полости. В 1905 г. Эйнштейн применил его для объяснения фотоэффекта, предполагая, что световые кванты, или фотоны, при фотоэффекте передают электронам энергию порциями, равными $h\nu$. — *Прим. авт.*



Фиг. 56. Волновая функция

$$\psi = u_{\mathbf{P}}(x) \cdot \exp(i\mathbf{P}x/\hbar).$$

Однако между атомами фазовый угол изменяется нерегулярно, оставаясь почти постоянным вблизи каждого атома, а затем резко изменяясь в промежутке между ними, где волновая функция приобретает малое значение. Если бы функции $u_\lambda(x)$ были вещественны или характеризовались постоянным фазовым углом, то фазовый угол для ψ изменялся бы линейно с изменением x . В общем случае, однако, функции $u_\lambda(x)$ комплексны и могут считаться вещественными только при $\lambda = 0$.

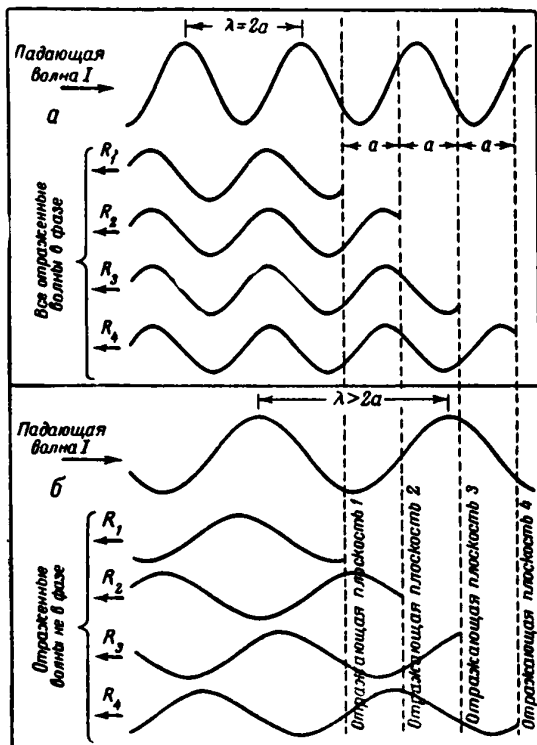
При $\lambda = 2a$ (фиг. 56, в) переход от одного атома к следующему соответствует умножению волновой функции на -1 , что отвечает изменению фазы на 180° . Для изображенного на фигуре случая волновая функция чисто вещественна [в другие моменты времени она будет чисто мнимой благодаря временной зависимости вида $\exp(-2\pi i\nu_\lambda t)$].

В точках, лежащих на середине расстояния между соседними атомами, волновая функция обращается при этом в нуль, так что фаза остается неопределенной. Эти скачкообразные изменения фазы изображены на рисунке (в2), в соответствии с фиг. 56, а и 56, б. Этот рисунок напоминает изображение винта с левой нарезкой. Однако ψ изменялась бы точно так же, если бы наш график был сходен с изображением винта с правой нарезкой. Таким образом, поведение волновой функции при $\lambda = 2a$ одинаково для волн, распространяющихся первоначально в любом направлении. Другими словами, мы имеем дело со стоячей волной. (Об этом уже говорилось в связи с фиг. 55, г.)

На фиг. 56, г показан ход вещественных и мнимых составляющих некоторых волновых функций для электронов в полосе проводимости металлического натрия согласно вычислениям Слэтера [4]. Жирные синусоидальные кривые соответствуют функциям Блоха, а пунктирные линии — плоским волнам свободных электронов, в которые превратились бы функции Блоха, если бы потенциальная энергия u уменьшилась до нуля.

Ввиду особой важности условия $\lambda = 2a$ в теории зон Бриллюэна мы в дополнение к сказанному в связи с обсуждением фиг. 56 получим ниже это условие несколько иным методом. Представим себе, что на фиг. 57, а синусоидальная волна I с $\lambda = 2a$ распространяется по кристаллу слева направо. При прохождении через каждую атомную плоскость волна частично отражается, образуя отраженные волны $R_1 \dots R_4$. Как видно, каждая отраженная волна имеет вначале разность фаз 180° по отношению к волне, отраженной от предыдущей атомной плоскости. Однако к моменту времени, когда, например, волна R_2 доходит до плоскости, от которой отражается волна R_1 , она изменит фазу на 180° и будет теперь находиться в одной фазе с волной R_1 . Поэтому волны $R_1 \dots R_4$ находятся в одинаковой фазе и не уничтожают друг друга. Эта картина весьма отлична от изображенной на фиг. 57, б, где длина волны не равна $2a$. В этом случае отраженные волны $R_1 \dots R_4$ не находятся в фазе и в среднем гасят

друг друга. Вследствие этого в случае фиг. 57, б волна проходит через кристалл, почти не отражаясь. На фиг. 57, а интенсивность отраженной волны равна интенсивности падающей. Отраженные волны вследствие дополнительных отражений превращаются в падающую волну так же быстро, как падающая превращается в отраженные.



Фиг. 57. Фазовые соотношения между отраженными волнами, иллюстрирующие критическое условие при $\lambda = 2a$.

Эта ситуация, при которой в обоих направлениях распространяются волны одинаковой интенсивности, математически эквивалентна стоячей волне. Отсюда следует невозможность передачи энергии или движения электронов в периодических структурах, подобных кристаллам или системам связанных механических или электрических осцилляторов, если длина волны равна удвоенному периоду структуры.

Очевидно, что если вначале мы имеем очень длинные модулированные волны, которые становятся затем все короче и короче, то

естественный предел их распространения достигается при $\lambda = 2a$. В действительности этот интервал длин волн как раз соответствует состояниям в энергетической зоне. Однако чтобы обобщить эти представления на случай трех измерений и перейти к рассмотрению скоростей и импульсов, мы должны найти методы описания модулированных волн в трех измерениях. Вводимое для этой цели квантовое число играет для электронов в кристаллах роль импульса и рассматривается в следующем параграфе.

Прежде чем перейти к обсуждению этого вопроса, нужно упомянуть об очень близкой аналогии между полученными здесь результатами и результатами решения некоторых электрических и акустических задач. Система электрических фильтров, состоящая из одинаковых контуров, соединенных последовательно, характеризуется полосами частот пропускания и полосами частот поглощения. Частота, при которой происходит переход между этими полосами, соответствует условию $\lambda = 2a$, причем токи в соседних контурах смещены по фазе на 180° . В области между полосами пропускания электрический сигнал в фильтре убывает экспоненциально от контура к контуру. Электронные волны, соответствующие энергиям, лежащим вне разрешенных энергетических зон, затухают аналогичным образом и не могут удовлетворять периодическим граничным условиям, рассмотренным выше. Однако они представляют значительный интерес в связи с исследованием свойств поверхности полупроводников, но этот вопрос выходит за рамки настоящей книги. Акустические линии с периодическими структурами ¹⁾ обладают аналогичными свойствами.

§ 5. ЗОНЫ БРИЛЛЮЭНА И КВАЗИ-ИМПУЛЬС Р КАК КВАНТОВОЕ ЧИСЛО

Для характеристики волновых функций Блоха введем в рассмотрение вместо длины волны так называемый *квази-импульс*. Квази-импульс, определение которого будет дано ниже, обладает многими удобными свойствами. Если воспользоваться квази-импульсом как квантовым числом, то возможные значения его описываются очень простой схемой, а критическое условие $\lambda = 2a$ приобретает весьма простой вид. В гл. 6 и 7 мы используем понятие квази-импульса при изучении скорости и ускорения электронов.

Рассмотрим сначала одномерный случай. С этой целью определим значение составляющей квази-импульса P_x вдоль оси x следующим образом:

$$P_x = \pm \frac{\hbar}{\lambda} \quad (5.12)$$

(знак плюс здесь соответствует волнам, распространяющимся вправо, знак минус — волнам, бегущим влево). Так как \hbar имеет размерность

¹⁾ Подобные задачи рассмотрены в книге Бриллюэна [5]. — *Прим. авт.*

действия, то P имеет размерность импульса (масса \times длину/время)¹⁾. Если предполагается, что амплитуда периодического потенциала уменьшается так, что функции Блоха становятся просто плоскими электронными волнами в свободном пространстве (которые квантовая механика интерпретирует как электроны, движущиеся свободно с постоянным значением импульса), то уравнение (5.12) дает правильное соотношение между импульсом и длиной волны. В кристалле, однако, на электрон действуют силы так, что его импульс в процессе движения изменяется, и поэтому, чтобы указать, что P_x не является импульсом в обычном смысле, мы ввели термин *квази-импульс*. (Мы будем обозначать квази-импульс большой буквой P , а обычный импульс — маленькой буквой p .) Так как скорость изменения P_x , как это будет показано ниже, равна приложенной силе, то в ряде случаев P_x можно рассматривать как импульс.

Волновая функция Блоха принимает при этом вид

$$\psi_{P_x}(x, t) = u_{P_x}(x) \exp \left[\frac{2\pi i}{h} (P_x x - \mathcal{E}_{P_x} t) \right]. \quad (5.13)$$

Теперь мы должны выяснить, какие значения P_x возможны. Ответ на этот вопрос зависит от размера кристалла и граничных условий. Рассмотрим кристалл, состоящий из N_x элементарных ячеек с длиной a и имеющий, таким образом, длину $A_x = N_x a$. Примем далее, что имеют место периодические, или циклические, граничные условия, эквивалентные предположению о том, что линейный кристалл согнут в кольцо так, что первый и последний атомы становятся ближайшими соседями. Как следствие этого граничного условия волновая функция в точках x и $x + A_x$ должна иметь одно и то же значение, так как эти два значения x в действительности совпадают. Другими словами, ψ должна быть периодической функцией с периодом A_x . Так как $u_{P_x}(x + a) = u_{P_x}(x)$, то мы имеем также $u_{P_x}(x + A_x) = u_{P_x}(x)$. Следовательно, чтобы ψ_{P_x} была периодической функцией с периодом A_x , должно выполняться условие

$$\exp \left[\frac{2\pi i}{h} P_x (x + A_x) \right] = \exp \left[\frac{2\pi i}{h} P_x x \right],$$

или

$$\exp \left[\frac{2\pi i}{h} P_x A_x \right] = 1, \quad (5.14)$$

или

$$\frac{2\pi i}{h} P_x A_x = 2\pi i n_x.$$

¹⁾ Этот результат может быть получен из соотношения неопределенности $\Delta p \Delta x \geq h/4\pi$, где p — обычный импульс. — *Прим. авт.*

где n_x — целое число, в действительности представляющее просто число длин волн ψ , укладывающихся вдоль циклического кристалла (как показано на фиг. 55). Это приводит к соотношению

$$P_x = \frac{n_x \hbar}{A_x} \quad \text{или} \quad \delta P_x = \frac{\hbar}{A_x}. \quad (5.15)$$

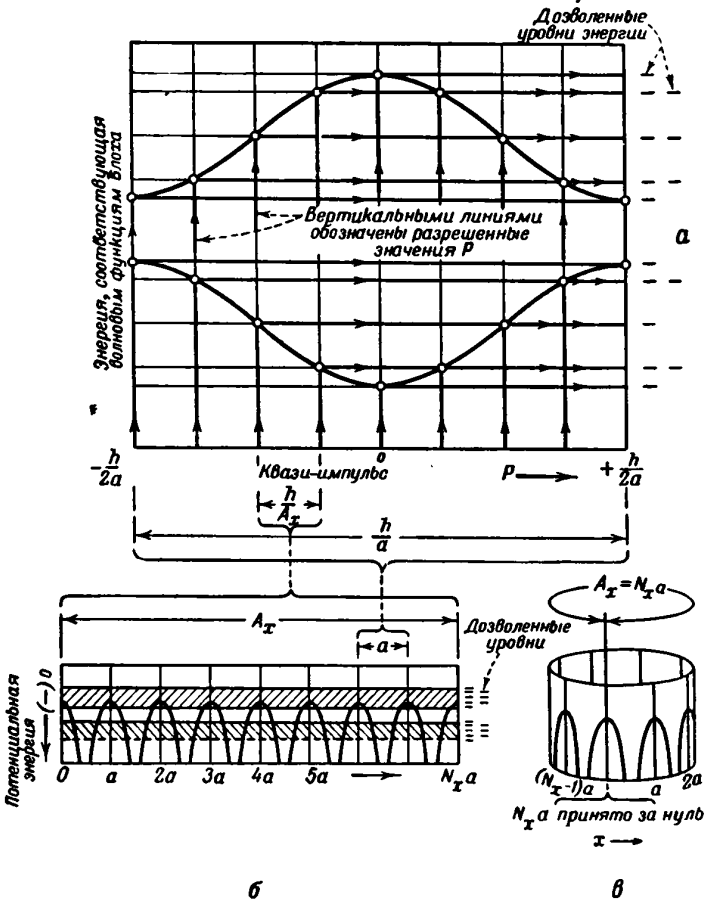
Таким образом, возможные значения P_x отстоят друг от друга на величину \hbar/A_x , как это показано на фиг. 58, а. Значение P_x изменяется в пределах от $-\hbar/2a$ до $+\hbar/2a$, соответственно предельному условию $\lambda = 2a$, которое рассматривалось выше, и волновые функции при этих двух предельных значениях P_x одинаковы. Как мы покажем дальше, значения P_x , лежащие вне интервала от $-\hbar/2a$ до $+\hbar/2a$, эквивалентны значениям P_x , лежащим в этом интервале. Эта область P_x пространства, содержащая всю совокупность возможных значений P_x и имеющая центр в точке $P_x = 0$, определяется как одномерная зона Бриллюэна. Весьма существенно, что в большинстве случаев нам нет надобности рассматривать значения P_x , лежащие вне зоны Бриллюэна, так как все собственные функции для одномерного кристалла соответствуют этой зоне. Действительно, в силу соотношений взаимности:

(Размеры зоны Бриллюэна) \propto (Расстояниям между атомами кристалла) $^{-1}$
 Размеры кристалла \propto (Расстоянию между возможными значениями P_x в зоне Бриллюэна) $^{-1}$,

т. е. в зоне Бриллюэна имеется как раз столько же возможных значений P_x , сколько атомов в кристалле. В этом заключается теорема сохранения квантовых состояний. Эта теорема является следствием того, что число атомов в кристалле равно $A_x/a = N_x$ и число возможных значений P_x также равно $(\hbar'a)(\hbar'/A_x) = A_x/a = N_x$. Учет двух возможных значений спина приводит нас к заключению, что число квантовых состояний в зоне Бриллюэна равно $2N_x$. На фиг. 58, а представлена зависимость энергии от P_x для двух энергетических зон. Это дает возможность построить схему энергетических уровней для разрешенных значений P_x , как это показано в верхнем правом углу фиг. 58, а. Те же самые уровни энергии нанесены и на фиг. 58, б, которая начерчена в уменьшенном масштабе, чтобы показать, как возникают энергетические зоны. Вычисление энергии как функции P_x может быть выполнено в действительности только для идеализированных одномерных кристаллов; вопрос этот рассмотрен в любой современной книге по электронной теории твердых тел и в некоторых книгах по квантовой теории¹⁾. Для электронов в реальных кристаллах вычисления весьма трудоемки и выполнены только в ограниченном числе случаев.

¹⁾ Простой случай рассмотрен в гл. 14. — Прим. авт.

Теперь мы должны обобщить понятие зон Бриллюэна на три измерения. Заменяем P_x вектором \mathbf{P} с составляющими P_x , P_y и P_z .



Фиг. 58. Структура зоны Бриллюэна и энергетические зоны для одномерного циклического кристалла. (Разрешенные значения P_x определяются формулой $P_x = \hbar n_x / A_x$)

a — две энергетические зоны в функции квази-импульса P_x ; б — схема, иллюстрирующая возникновение энергетических зон из разрешенных уровней энергии; в — циклическая или периодическая шкала значений x .

В этом случае мы снова приходим к заключению, что в \mathbf{P} -пространстве следует рассматривать определенный объем, ограниченный плоскостями, соответствующими условию $\lambda = 2a$. Этот ограниченный

объем \mathbf{P} -пространства, который может быть определен как наименьший объем вблизи $\mathbf{P}=0$, содержащий все неэквивалентные значения \mathbf{P} , и есть трехмерная зона Бриллюэна. Рассмотрению подлежат только значения \mathbf{P} , лежащие внутри этой зоны, и только некоторые из этих значений соответствуют собственным функциям, удовлетворяющим граничным условиям. Покажем, что для этих возможных значений \mathbf{P} снова применима теория сохранения квантовых состояний.

Для трех измерений функция Блоха может быть записана следующим образом:

$$\begin{aligned} \psi_{P_x P_y P_z}(x, y, z, t) &= u_{P_x P_y P_z}(x, y, z) \times \\ &\times \exp\left[\frac{2\pi i}{h}(P_x x + P_y y + P_z z - \mathcal{E}_{P_x P_y P_z} t)\right] = \\ &= \psi_{\mathbf{P}}(\mathbf{r}, t) = u_{\mathbf{P}}(\mathbf{r}) \exp\left[\frac{2\pi i}{h}(\mathbf{P} \cdot \mathbf{r} - \mathcal{E}_{\mathbf{P}} t)\right]; \end{aligned} \quad (5.16)$$

в последнем выражении использованы векторные обозначения

$$\mathbf{P} = iP_x + jP_y + kP_z, \quad (5.17)$$

$$\mathbf{r} = ix + jy + kz, \quad (5.18)$$

$$\mathbf{P} \cdot \mathbf{r} = P_x x + P_y y + P_z z, \quad (5.19)$$

где i, j и k — обычные единичные векторы вдоль осей x, y и z (i не следует смешивать с $i = \sqrt{-1}$). Предположим, что мы имеем дело с простым кубическим кристаллом, постоянная решетки которого равна a . (Ниже мы укажем на изменения, которые следует ввести при рассмотрении кристалла с более сложной структурой, например алмаза.) По тем же соображениям, что и в одномерном случае¹⁾, можно показать, что при $P_x = h/2a$ волновые функции соответствуют стоячим волнам в направлении x и что при любых произвольно выбранных значениях P_y и P_z условия $P_x = h/2a$ и $P_x = -h/2a$ приводят к той же самой волновой функции. Таким образом, зона Бриллюэна в пространстве (P_x, P_y, P_z) или в \mathbf{P} -пространстве представляет собой куб, ограниченный плоскостями $P_x = \pm h/2a$, $P_y = \pm h/2a$, $P_z = \pm h/2a$. В случае более сложных решеток нахождения зоны Бриллюэна связано с применением более общих методов. Во всех случаях границы зоны определяются, однако, условиями существования волн, подобными условию $\lambda = 2a$. При этом отражающие плоскости не обязательно должны быть перпендикулярны к осям x, y и z ; следует также учитывать и другие возможные плоскости. Этот метод давно применяется при изучении диффракции рентгеновских

¹⁾ Более подробно с этими соображениями читатель может ознакомиться в книгах Зейтца [6] и Вильсона [7]. — *Прим. авт.*

лучей в кристаллах. Согласно этой теории, рентгеновские лучи дают брэгговские отражения от некоторых определенных плоскостей кристалла, если волновой вектор (\mathbf{P}/\hbar) удовлетворяет условиям, эквивалентным условию $\lambda = 2a$, где a — расстояние между рассматриваемыми плоскостями. Таким путем находятся некоторые определенные плоскости в \mathbf{P} -пространстве, ограничивающие зону Бриллюэна для более сложных решеток. Мы не будем, однако, рассматривать эти более сложные случаи и ограничимся рассмотрением простой кубической модели.

Если исследуемый образец простого кубического кристалла представляет собой прямоугольный параллелепипед с ребрами длиной A_x , A_y и A_z , параллельными осям x , y и z , можно найти возможные значения \mathbf{P} , воспользовавшись, как и раньше, периодическими граничными условиями. В этом случае уже трудно представить себе, что кристалл согнут так, что его противоположные грани соединены вместе, как это показано на фиг. 58, в. Однако требование, чтобы волновая функция была периодической с периодами A_x , A_y и A_z в трех измерениях, математически может быть легко сформулировано. Практика решения задач такого рода в теоретической физике показала, что в общем случае результаты не зависят от точной формулировки граничных условий на поверхности образца ¹⁾. Поэтому мы не будем пытаться дать физическое обоснование наличию периодических граничных условий, а укажем, что эквивалентные результаты получились бы и в том случае, если бы функции обращались в нуль на границе, определяемой полем, удерживающим электроны внутри кристалла. Если волновая функция принимает одно и то же значение в точках x и $x + A_x$, то из тех же соображений, что и раньше, следует

$$P_x = \frac{n_x \hbar}{A_x} \quad \text{или} \quad \delta P_x = \frac{\hbar}{A_x}. \quad (5.20)$$

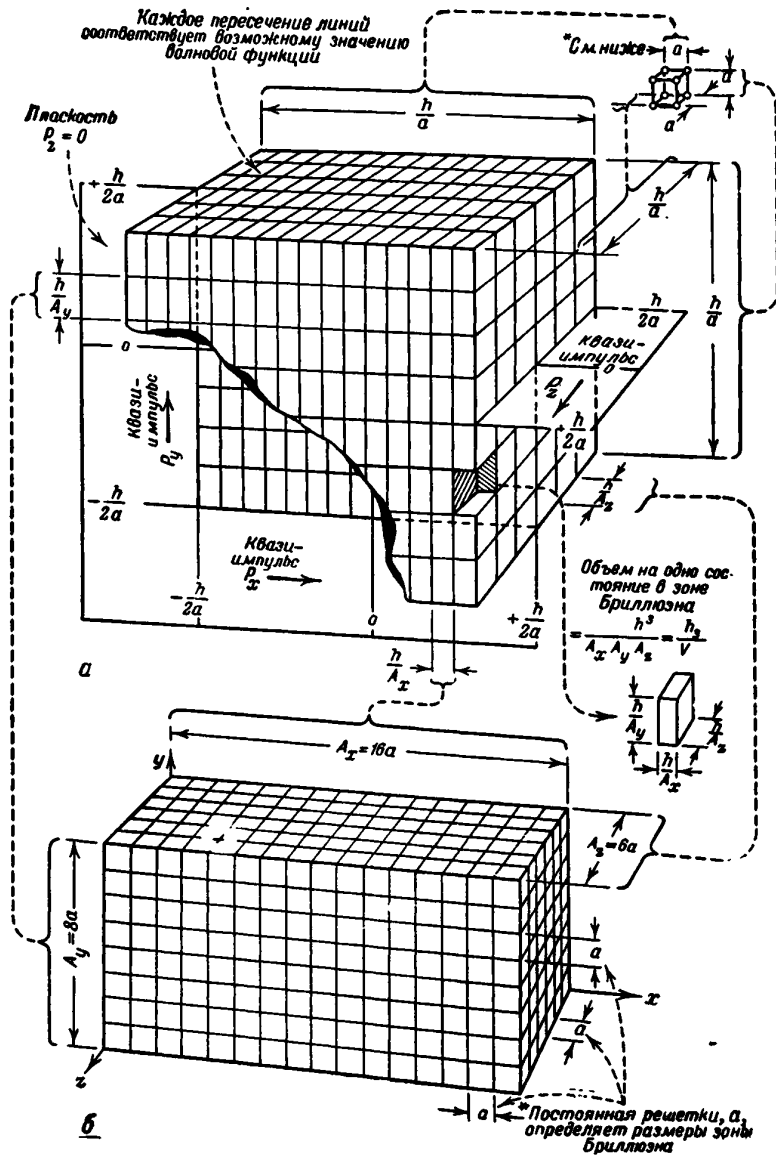
Значения y и z при таком увеличении x остаются неизменными. Аналогично, для составляющих квази-импульса по осям y и z имеем:

$$P_y = \frac{n_y \hbar}{A_y} \quad \text{или} \quad \delta P_y = \frac{\hbar}{A_y}, \quad (5.21)$$

$$P_z = \frac{n_z \hbar}{A_z} \quad \text{или} \quad \delta P_z = \frac{\hbar}{A_z}. \quad (5.22)$$

Таким образом, точки, характеризующие возможные состояния в \mathbf{P} -пространстве, расположены в узлах прямоугольной или орторомбической решетки, элементарная ячейка которой обладает ребрами \hbar/A_x ,

¹⁾ С этим вопросом мы еще встретимся в гл. 7. — *Прим. авт.*



Фиг. 59. Зона Бриллюэна для кристалла с простой кубической структурой в трех измерениях.
 а — зона Бриллюэна; б — схематическое изображение кристалла с кубической решеткой.

h/A_y , h/A_z , так что в \mathbf{P} -пространстве каждой элементарной ячейке или возможной волновой функции отвечает объем $h^3/A_x A_y A_z$ или h^3/V . Таким образом,

(Размеры зоны Бриллюэна) \sim (Размерам элементарной ячейки кристалла) $^{-1}$,
(Размеры кристалла) \sim (Расстояниям между возможными состояниями в зоне Бриллюэна) $^{-1}$.

В дальнейшем мы будем иметь дело с решеткой в \mathbf{P} -пространстве, определяемой уравнениями (5.20) — (5.22) также в связи с другими вопросами, и будем пользоваться термином *основная \mathbf{P} -решетка зоны Бриллюэна* для описания расположения точек, отвечающих отдельным состояниям в зоне Бриллюэна.

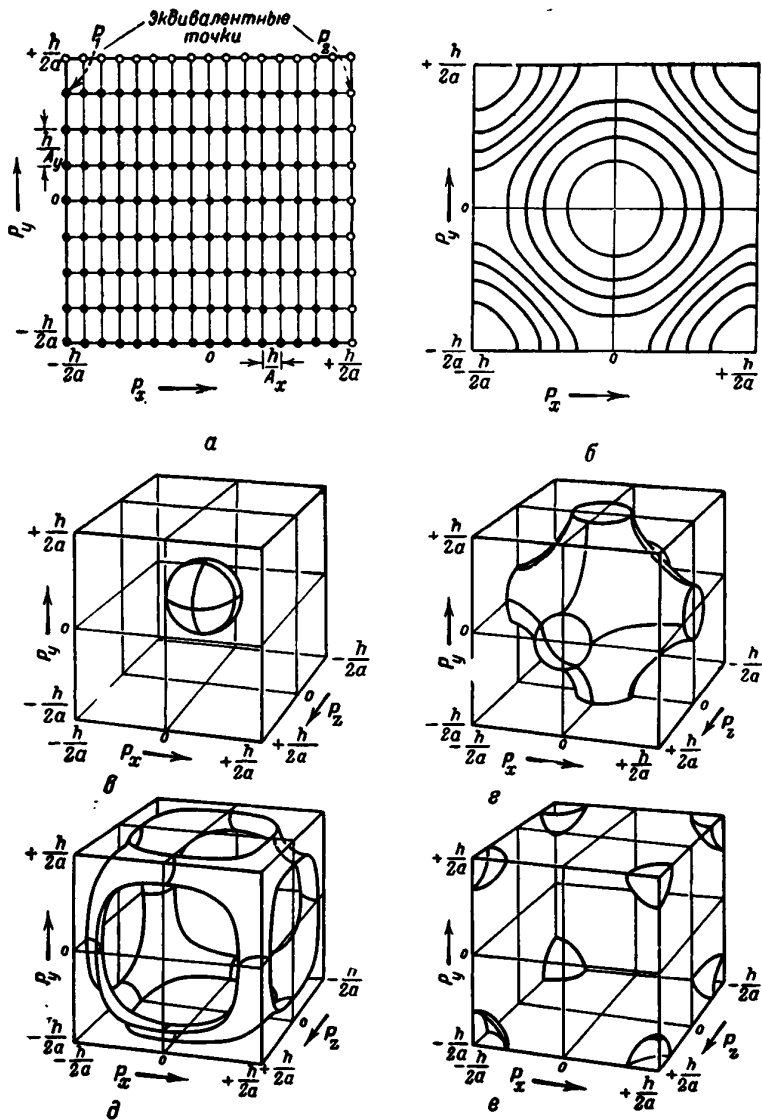
Эта решетка и указанные выше соотношения представлены на фиг. 59. Возможные состояния в плоскости $P_x = 0$ изображены более подробно на фиг. 60, а. Здесь каждое отдельное возможное состояние отмечено жирной точкой. Состояниям, отмеченным кружками (например P_2), на краях зоны Бриллюэна отвечают те же волновые функции, что и другим состояниям (например P_1). Таким образом, число возможных состояний равно числу прямоугольников на плоскости. В трех измерениях число возможных состояний точно так же равно числу элементарных объемов h^3/V . Как показано в § 7 гл. 14, это общий результат, который может быть выражен следующим образом:

$$\text{Плотность возможных состояний в } \mathbf{P}\text{-пространстве} = V/h^3. \quad (5.23)$$

Так как объем зоны Бриллюэна равен $(h/a)^3$, то число возможных состояний равно $(h/a)^3 \cdot (V/h^3) = V/a^3$. Таким образом, мы приходим к формулировке теоремы о сохранении квантовых состояний в рамках теории зон Бриллюэна:

Число возможных состояний в зоне Бриллюэна равно числу элементарных ячеек кристалла.

Отсюда легко установить связь между зоной Бриллюэна и энергетической зоной. Если кристалл получен путем постепенного уменьшения расстояний между атомами, то, как об этом упоминалось в § 3 гл. 5, каждый атомный уровень энергии расщепляется на ряд уровней, образующих энергетическую зону. Вначале на каждый атом для каждого значения спина приходится по одному квантовому состоянию. Число всех квантовых состояний в энергетической зоне равно $2N_s$, где N_s — общее число атомов. Эти $2N_s$ квантовых состояний с учетом спина соответствуют N_s возможным состояниям в зоне Бриллюэна. Если на каждую элементарную ячейку приходится больше одного атома, то картина несколько усложняется. В этом случае уровень энергии будет, вероятно, расщепляться с образованием нескольких перекрывающихся полос; для описания каждой из них необходима отдельная зона Бриллюэна.



Фиг. 60. Зона Бриллюэна, разрешенные квантовые состояния и контуры равной энергии.

а — разрешенные значения импульса в плоскости $P_z=0$ (основная Р-решетка); б — контуры постоянной энергии в плоскости $P_z=0$ (разности энергий между ними одинаковы); в — г — типичные формы поверхностей постоянной энергии для простой кубической решетки.

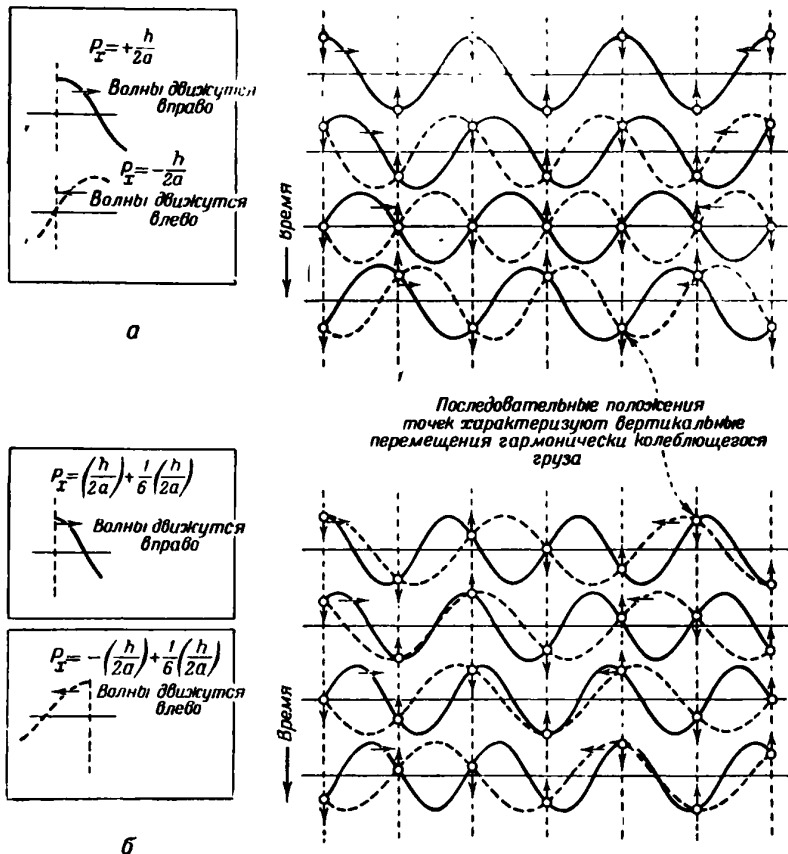
Выше был рассмотрен простой кубический кристалл. Однако доказанная ранее теорема в действительности применима к *любой* кристаллической решетке, если только *элементарная ячейка* выбрана надлежащим образом и представляет собой наименьший параллелепипед, который типичен для кристалла и, будучи многократно перенесен параллельно самому себе, образует кристалл. В структуре алмаза, изображенной на фиг. 4 (стр. 21), элементарная ячейка есть косой параллелепипед (содержащий два атома), объем которого равен $1/4$ объема куба, изображенного на рисунке. Четыре электрона на атом в зоне валентных связей (см. фиг. 54) соответствуют, таким образом, восьми электронам в элементарной ячейке, что как раз достаточно для заполнения четырех зон Бриллюэна, если учесть оба значения спина.

Таким образом, интересующие нас энергетические зоны кристаллов типа алмаза соответствуют зонам Бриллюэна, энергии которых перекрываются сложным образом. В § 5 гл. 7 будет показано, что при рассмотрении этого случая допустимы некоторые приближения. Случай перекрывающихся зон энергии исследуется в гл. 12.

Энергия, отвечающая волновой функции, зависит, конечно, от значения \mathbf{P} . На фиг. 60, *б* изображена зависимость энергии от \mathbf{P} для некоторого гипотетического случая; контуры равной энергии показаны для возможных состояний в плоскости $P_z = 0$, изображенных на фиг. 59. Как видно, рисунок обладает высокой степенью симметрии в соответствии с симметрией рассматриваемой простой кубической модели. Минимум энергии отвечает центру изображенного на рисунке квадрата, а максимальные значения — его углам. Разности значений энергии для двух соседних контуров, изображенных на фиг. 60, *б*, во всех случаях примерно одинаковы. В трехмерном случае мы имеем дело с поверхностями постоянной энергии, почти сферическими вблизи центра зоны и обладающими более сложной формой при больших значениях энергии. Наконец, вблизи углов куба поверхности снова становятся почти сферическими. Наименьшие значения энергии не всегда соответствуют центру зоны. Как мы увидим ниже, когда речь идет об электронах как носителях тока, минимум энергии может с таким же успехом иметь место в углах зоны, а максимум — в ее центре.

Установлено, что состояния, расположенные на противоположных гранях зоны Бриллюэна, как, например, \mathbf{P}_1 и \mathbf{P}_2 на фиг. 60, для которых разности компонент квази-импульсов равны $\Delta P_x = \pm \hbar/a$, $\Delta P_y = \Delta P_z = 0$, характеризуются одной и той же волновой функцией. Таким образом, если P_x превышает $\hbar/2a$, то это означает, что мы проникаем в зону через противоположную грань. Этот результат можно доказать строго, используя волновые функции Блоха. Он может быть также получен в аналогичном случае для системы связанных осцилляторов (см. фиг. 52 и 55).

На фиг. 61 изображены три модулированные волны, распространяющиеся вдоль ряда гармонических осцилляторов согласно схеме, принятой на фиг. 52 и 55. Фиг. 61, а соответствует условиям



Фиг. 61. Значения импульса вне зоны Бриллюэна, эквивалентные значениям импульса, лежащим внутри зоны и полученным путем вычитания h/a из P .

$P_{\omega} = \pm h, 2a$ или $1 \lambda = 1, 2a$. Шпильной линией и пунктиром изображены волны, распространяющиеся соответственно вправо и влево на последовательных диаграммах. На фиг. 61, б изображена волна, для которой $P_{\omega} = (h, 2a) + 1/6(h, 2a)$, т. е. импульс выходит за пределы зоны на $16\%_{10}$. Кроме того, на этой же фигуре рассмотрен случай, когда $P_{\omega} = -(h, 2a) + 1/6(h, 2a)$. Поскольку такая волна

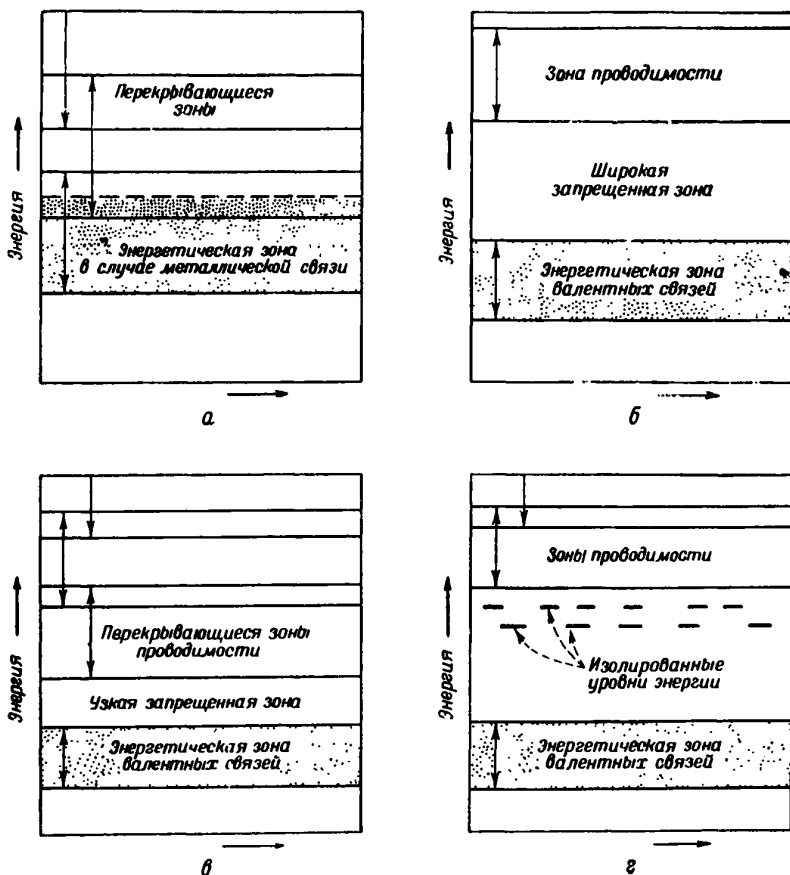
характеризуется отрицательным значением P_x , ее следует рассматривать как распространяющуюся влево. Как видно, перемещение грузов осцилляторов одинаково для обеих волн; это указывает на то, что распространение волн за границу зоны эквивалентно их вхождению в зону через противоположную грань. Следовательно, можно было бы с одинаковым успехом выбрать в качестве границ зоны $P_x = 0$, $P_x = h/a$; $P_y = 0$, $P_y = h/a$, так как в этом случае была бы получена та же совокупность модулированных волн, что и в случае, показанном на фиг. 60, а. Такой выбор границ зоны имеет иногда свои преимущества, и мы им воспользуемся в § 5 гл. 7.

P -пространство периодически в том смысле, что добавление к любой составляющей квази-импульса величины $\pm h/a$ не изменяет волновой функции. Вследствие этого состояния вне зоны Бриллюэна всегда эквивалентны состояниям внутри зоны и в действительности нет никакой необходимости рассматривать значения P , лежащие вне зоны. Однако число квантовых состояний в зоне Бриллюэна достаточно для размещения лишь двух электронов, приходящихся на каждую элементарную ячейку, в то время как для большинства кристаллов число электронов в элементарной ячейке значительно больше двух. Эта трудность может быть устранена, если использовать многократно одни и те же состояния зоны Бриллюэна, указывая в каждом случае, о какой энергетической зоне идет речь. Другой более сложный метод заключается в том, чтобы использовать P -пространство вне данной зоны Бриллюэна для построения других зон. Получающиеся при этом более сложные зоны в P -пространстве называются второй, третьей и т. д. зонами Бриллюэна. (Пунктирные линии на фиг. 56, г иллюстрируют этот метод.) Для наших целей, однако, проще всего пользоваться только первой зоной Бриллюэна, описанной выше. Данному состоянию в этой зоне отвечает определенная модулированная волна, $\exp(2\pi i P \cdot r/h)$, наложенная на атомную волновую функцию. Применение той же модулированной волны и различных атомных функций дает волновые функции, соответствующие разным энергетическим зонам. Некоторые примеры такого рода показаны на фиг. 56, г.

§ 6. ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ МЕТАЛЛОВ, ИЗОЛЯТОРОВ И ПОЛУПРОВОДНИКОВ

Между энергетическими зонами металлов, изоляторов и собственных полупроводников существуют характерные различия, позволяющие объяснить закономерности, наблюдаемые при изучении их электропроводности. Как мы увидим дальше, если все квантовые состояния в зоне Бриллюэна заняты, то зона не может принимать участия в проводимости. В случае металла энергетические зоны перекрываются таким образом, что некоторые из зон Бриллюэна остаются лишь частично заполненными. С другой стороны, у чистого алмаза одна группа зон Бриллюэна полностью занята валентными электронами,

в то время как следующая (зона проводимости) остается совершенно пустой. Запрещенная энергетическая зона в этом случае столь широка, что тепловое возбуждение электронов невозможно, и



Фиг. 62. Схемы энергетических зон для кристаллов разного типа.

а — металл; б — изолятор; в — полупроводник с собственной проводимостью; г — примесный полупроводник.

алмаз поэтому обладает либо целиком заполненными, либо совершенно пустыми зонами Бриллюэна и в тепловом равновесии, при комнатной температуре, является изолятором. У германия зона валентных связей и зона проводимости расположены настолько близко друг от друга, что он обладает заметной проводимостью уже при комнатной

температуре, будучи даже совершенно чистым. Модификация олова, обладающая структурой алмаза (так называемое серое олово), устойчивая при температурах ниже комнатной, является, вероятно, металлическим проводником с перекрывающимися энергетическими зонами или, по крайней мере, собственным полупроводником с очень большой электропроводностью. Ниже мы приводим экспериментальные значения ширины запрещенных зон для четырехвалентных элементов (в эв):

Углерод (алмаз)	Кремний	Германий	Олово серое [8]	Олово метал- лическое	Свинец
6—7	1,11	0,72	0,1	(Перекрывающиеся энергетические зоны)	

Доноры и акцепторы, которые могут связывать дырки и электроны, как это было описано в гл. 1, создают уровни энергии, лежащие в запрещенной зоне. Квантовомеханическое рассмотрение этих энергетических уровней и их связи с энергетическими зонами будет дано в гл. 9.

Схемы энергетических уровней для различных только что рассмотренных случаев изображены на фиг. 62. К обсуждению вопроса о статистическом распределении электронов по этим энергетическим уровням в условиях теплового равновесия мы вернемся в конце гл. 10.

ЗАДАЧИ

1. Произвести анализ нормальных колебаний для системы из четырех и восьми осцилляторов (подобно тому, как это сделано на фиг. 55); рассмотреть также случай, когда число осцилляторов нечетно.

2. Фиг. 55, *a* — *г* напоминают осциллограммы, наблюдаемые на экране осциллографа. Показать, что если на горизонтальные пластины осциллографа подано переменное напряжение с частотой f , а на вертикальные пластины — переменное напряжение с частотой nf , то осциллограммы будут иметь вид, подобный кривым, изображенным пунктирными линиями на фиг. 55.

Если на Y -пластины осциллографа подается, кроме того, еще небольшой сигнал с частотой f , смещенный по фазе на 90° по отношению к сигналу, подаваемому на горизонтальные пластины, то можно получить перспективную осциллограмму. Что произойдет, если частота сигнала, поданного на Y -пластины, будет слегка отлична от nf ?

3. Предположим, что на вертикальные пружины действует постоянная сила и имеется начальное натяжение в связывающих пружинах.

жинах; предположим также, что массы на фиг. 55 скользят без трения по вертикальным стержням. Найти частоту колебаний системы для возмущений с небольшой амплитудой.

4. Сравнить качественно распределение возможных частот для системы, состоящей из нечетного числа связанных осцилляторов в случаях „граничных условий“ с закрепленными концами и периодических граничных условий.

ЛИТЕРАТУРА

1. Shockley W., Bell Syst. Techn. Journ. **28**, 645 (1939).
2. Kimball G. E., Journ. Chem. Phys., **3**, 560 (1935).
3. Bloch F., Zs. Phys., **52**, 555 (1928).
4. Slater J. C., Rev. Mod. Phys., **6**, 209 (1934).
5. Brillouin L., Wave Propagation in Periodic Structures, New York, 1946.
6. Seitz F., Modern Theory of Solids, New York, 1940 (см. перевод: Зейтц Ф., Современная теория твердого тела, М.—Л., 1949).
7. Wilson A. H., The Theory of Metals, Cambridge, 1936 (см. перевод: Вильсон А., Квантовая теория металлов, М.—Л., 1941).
8. Busch G., Wieland J., Zoller H., Conference on the Properties of Semiconducting Materials, University of Reading, 1950.
- 9*. Френкель Я. И., Введение в теорию металлов, 2-е изд., М.—Л., 1950.
- 10*. Адирович Э. И., Некоторые вопросы теории люминесценции кристаллов, М.—Л., 1951.
- 11*. Ансельм А. И., ЖТФ, **21**, 489 (1951).
- 12*. Волькенштейн Ф. Ф., Успехи физич. наук, **49**, 11 (1951).
- 13*. Волькенштейн Ф. Ф., ЖТФ, **21**, 1544 (1951).
- 14*. Волькенштейн Ф. Ф., ЖЭТФ, **22**, 184 (1952).
- 15*. Иоффе А. Ф., Изв. АН СССР, серия физич., **15**, 477, 488 (1951).
- 16*. Ландау Л. Д., Пекар С. И., ЖЭТФ, **18**, 481 (1948).
- 17*. Пекар С. И., ЖЭТФ, **18**, 525 (1948).
- 18*. Пекар С. И., Исследования по электронной теории кристаллов, М.—Л., 1951.
- 19*. Френкель Я. И., Успехи физич. наук, **30**, 11 (1946).
- 20*. Блохинцев Д. И., ЖЭТФ, **6**, 1053 (1936).
- 21*. Вонсовский С. В., Изв. АН СССР, серия физич., **12**, 337 (1948).
- 22*. Ландау Л. Д., Sow. Phys., **3**, 664 (1933).
- 23*. Френкель Я. И., Phys. Rev., **37**, 17, 1276 (1931).
- 24*. Френкель Я. И., ЖЭТФ, **6**, 647 (1936).

Глава 6

СКОРОСТИ И ТОКИ ЭЛЕКТРОНОВ В КРИСТАЛЛАХ

Этой главой завершается рассмотрение волновых функций Блоха с точки зрения зонной теории. Ею мы заканчиваем общее описание движения электронов в периодическом поле идеального кристалла в отсутствие внешних полей. Здесь вводится также новое понятие скорости движения, соответствующей каждому данному квантовому состоянию в зоне Бриллюэна. В руководствах, предназначенных исключительно для физиков, эти вопросы рассматривались бы главным образом на основе общих положений квантовой механики. Хотя мы и ссылаемся на квантовомеханические методы, однако в первую очередь мы стремимся провести аналогию с другими задачами, связанными с волновым движением. В конце этой главы мы подойдем к элементарному рассмотрению теории электропроводности. В гл. 7 будет рассмотрено влияние электрического и магнитного полей на движение электронов, а в гл. 8 введено понятие о процессах, обуславливающих сопротивление.

§ 1. ПОНЯТИЯ СКОРОСТИ И ТОКА

Если движение электрона в кристаллах описывается одной из волновых функций Блоха, то электрон обладает при этом некоторой „средней“, отличной от нуля, скоростью движения, соответствующего наличию электронного тока в кристалле. Как известно, каждому квантовому состоянию в зоне Бриллюэна отвечает определенная средняя скорость, которой электрон обладал бы, если бы занимал это состояние. В настоящей главе обсуждается связь этой средней скорости, обозначаемой вектором v , с другими свойствами волновых функций Блоха.

Как уже указывалось при рассмотрении фиг. 56, по кристаллу распространяются комплексные волны Блоха. Поскольку плотность вероятности для электронов в каждой элементарной ячейке одна и та же, исследование распределения этой плотности в пространстве не дает возможности обнаружить явно ток электронов. Это заставляет нас интерпретировать смысл волновой функции Блоха с помощью более привычного способа описания движения электрона. Мы покажем, что электрон действительно движется в кристалле с определенной средней скоростью. Необходимо сразу же указать, что это

не находится в противоречии с равномерным распределением плотности вероятности. В ряде других случаев при движении некоторого физического тела также не наблюдается изменения в распределении плотности. Например, вращающееся маховое колесо может выглядеть совершенно неподвижным, тогда как железо, равномерно распределенное по его краю, в действительности движется. Плотность энергии вдоль линии электропередачи может быть равномерной и неизменной несмотря на то, что по этой линии передается мощность. (Пример такого рода приведен в § 3 гл. 15.)

Функция Блоха характеризует такое состояние движения электрона, в котором он с одинаковой вероятностью может находиться в любой элементарной ячейке кристалла и в то же время обладает некоторой средней скоростью движения по кристаллу. То, что в результате движения электрон не покидает кристалла по истечении некоторого промежутка времени, объясняется математически наличием определенных периодических граничных условий. Эти условия (как показано на фиг. 58, в для одномерного случая) соответствуют циклическому кристаллу, в котором выход электрона с одного конца кристалла эквивалентен вхождению его с другого конца. Таким образом, электроны все время текут по кристаллу.

Это течение может быть описано с помощью понятия плотности тока в форме, удобной для дальнейшего рассмотрения вопросов электропроводности. В одномерном случае, при средней скорости движения v_x , электрон совершает один оборот по циклическому кристаллу за время A/v_x . Это значит, что заряд электрона $-e$ проходит через кристалл v_x/A раз в секунду и, следовательно, ток равен $I = -ev_x/A = (-e/A)v_x$; последнее выражение показывает, что ток можно представить в виде произведения заряда, проходящего на единицу длины $(-e/A)$, на среднюю скорость электрона. Для случая трех измерений вектор плотности тока или заряд, проходящий через единицу площади за единицу времени, может быть представлен аналогичным образом:

$$I = -\frac{e}{V} \mathbf{v}, \quad (6.1)$$

т. е. как произведение заряда в единице объема на среднюю скорость электрона¹⁾. Это выражение получается следующим образом: если электрон обладает средней скоростью \mathbf{v} , с составляющими v_x , v_y , v_z , то он проходит в направлении x через кристалл v_x/A_x раз за единицу времени; каждый раз он переносит заряд $-e$ через грань кристалла, обуславливая тем самым ток, равный $-ev_x/A_x$. Так как площадь этой грани равна $A_y A_z$, то плотность тока в направлении x

¹⁾ Еще до пересечения поверхностей электрон создает ток смещения, определяемый уравнением (6.1) [см., например, W. Shockley, Journ. Appl. Phys., 9, 635 (1938)]. — *Прим. авт.*

определяется формулой

$$I_x = \frac{-e v_x A_x}{A_y A_z} = -\frac{e}{V} v_x, \quad (6.2)$$

где $V = A_x A_y A_z$ — объем кристалла. Подобные же вычисления производятся и для направлений y и z , в результате чего мы получаем общую формулу для плотности тока (6.1). В действительности приведенный здесь вывод формулы для плотности тока служит лишь для иллюстрации понятия средней скорости и в известном смысле противоположен тому, который обычно применяется в теории для нахождения средней скорости. Это станет очевидным из рассмотрения метода вычисления средней скорости.

§ 2. ВЕКТОР УМОВА — ПОЙНТИНГА ¹⁾ И МЕТОД ПЛОТНОСТИ ПОТОКА

Изучение изменения плотности вероятности не дает возможности исследовать движение электронов, и поэтому для определения скорости электронов следует воспользоваться другими методами.

Таковыми методами располагает квантовая механика, которая дает также определенный способ вычисления среднего тока с помощью волновой функции. Мы не будем здесь останавливаться на этом вопросе ²⁾, а рассмотрим вместо этого аналогичный метод, применяемый для вычисления потока мощности в электромагнитной теории. Схемы вычисления для обоих случаев приведены на фиг. 63. В § 3 гл. 15 подробно исследована искусственная линия передачи с сосредоточенными постоянными, не имеющая потерь, и вычислена энергия, приходящаяся на единицу длины такой линии, когда по ней распространяется бегущая волна.

Выражение, определяющее поток мощности, при этом получается путем умножения значения тока, проходящего слева направо через некоторое сечение линии, на напряжение и соответствующий фазовый множитель.

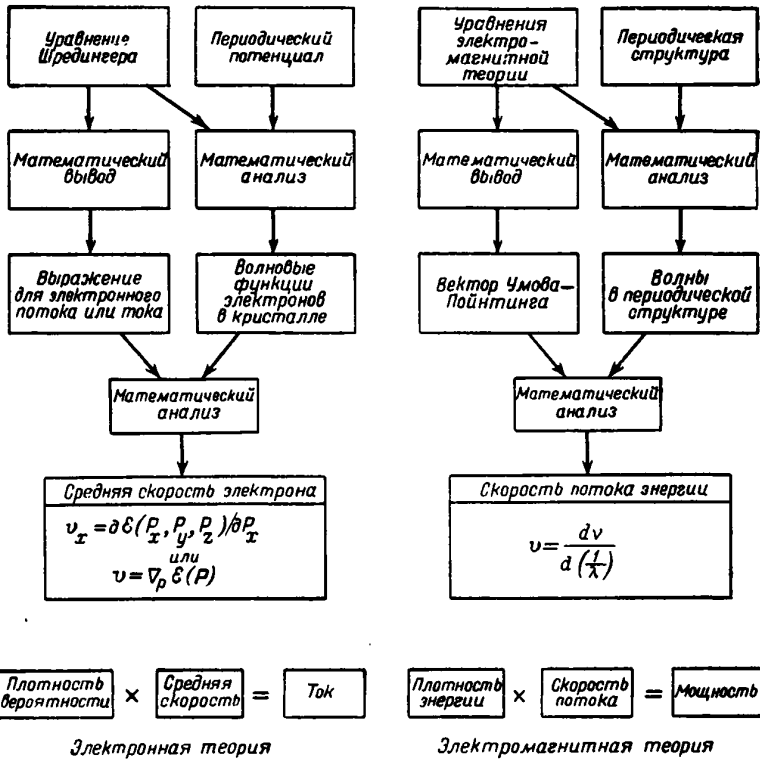
Это знакомое выражение для потока мощности аналогично определению вектора Умова — Пойнтинга для электромагнитных волн или квантовомеханическому выражению для среднего значения тока.

Ниже мы покажем, что поток мощности вдоль линии обусловлен наличием потока плотности энергии. Если скорость течения равна v , то вся энергия, соответствующая отрезку линии длиной v , протекает

¹⁾ Выражение для потока электромагнитной энергии было впервые дано известным русским физиком Н. А. Умовым значительно раньше английского ученого Пойнтинга, которому иностранные авторы упорно приписывают приоритет в этом вопросе. В соответствии с принятой в советской научной литературе терминологией, в переводе мы называем эту величину вектором Умова — Пойнтинга. — *Прим. ред.*

²⁾ См. § 2 гл. 14 и § 4 гл. 15. — *Прим. авт.*

через некоторое сечение этой линии за единицу времени и, следовательно, мощность равна произведению плотности энергии на скорость течения (см. фиг. 63). Если в линии отсутствуют фазовые



Фиг. 63. Схемы вычисления средней скорости электрона и потока электромагнитной энергии.

искажения, т. е. электромагнитные волны всех частот распространяются в ней с одинаковой скоростью или, другими словами, явление дисперсии не наблюдается, то скорость течения всегда оказывается равной скорости распространения волн. Если, однако, скорость распространения волн зависит от длины волны, то скорость передачи мощности не совпадает со скоростью распространения волн, а определяется формулой

$$v = \frac{d\nu}{d(1/\lambda)}. \quad (6.3)$$

Эта скорость равна скорости распространения волн только в том случае, когда последняя постоянна, т. е.

$$v = \frac{c}{\lambda}, \text{ так что } \frac{dv}{d(1/\lambda)} = \frac{dc(1/\lambda)}{d(1/\lambda)} = c, \quad (6.4)$$

где c — скорость распространения волн.

Аналогично, используя выражение, определяющее плотность потока электронов, получим для плотности вероятности точно такую же формулу, как и для плотности электрической энергии в *одномерном случае* (см. фиг. 63). Воспользовавшись соотношениями

$$h\nu = \mathcal{E} \quad \text{и} \quad P_x = \frac{h}{\lambda}, \quad (6.5)$$

можно переписать (6.3) в виде

$$v = \frac{dv}{d(1/\lambda)} = \frac{d\mathcal{E}}{dP_x}. \quad (6.6)$$

По своей форме (6.6) аналогично соответствующему классическому выражению, в котором $p = mv$ и $\mathcal{E} = mv^2/2 = p^2/2m$, так что $d\mathcal{E}/dp = p/m = v$. Однако физический смысл этих соотношений совершенно различен, так как для электронов, движущихся в периодическом поле, \mathcal{E} не равно $p^2/2m$.

То обстоятельство, что как для электромагнитных, так и для электронных волн скорость определяется одной и той же формулой $v = dv/d(1/\lambda)$, может быть объяснено на основании весьма общих соображений. Это объяснение основывается на исследовании волновых пакетов и групповой скорости.

§ 3. ВОЛНОВЫЕ ПАКЕТЫ И ГРУППОВАЯ СКОРОСТЬ

Пользуясь понятиями о волновых пакетах и групповых скоростях, играющих важную роль в волновой механике, исследуем полученное выше выражение, определяющее скорость.

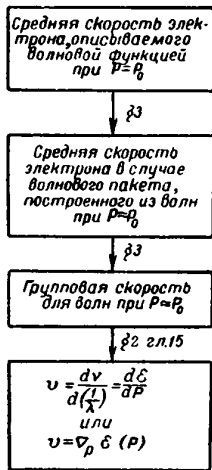
Предположим, что по линии вместо сигнала, характеризующего какой-либо одной частотой, посылаем импульсный сигнал на некоторой несущей частоте. Этот импульс можно рассматривать как суперпозицию некоторого числа волн, каждая из которых характеризуется одной определенной частотой. Положение импульса в пространстве определяется условиями интерференции этих волн. В общем случае эти волны распространяются с различными скоростями и, следовательно, интерференционная картина изменяется. В результате смещения интерференционной картины группы волн, обладающих различными частотами, импульс будет распространяться со скоростью, отличной от скорости распространения отдельных волн. Такой импульс называется *волновым пакетом*, который образовался в результате интерференции *группы* волн. Каждая из этих отдельных волн распространяется со своей

собственной скоростью, называемой *фазовой скоростью*, тогда как волновой пакет движется с так называемой *групповой скоростью*.

Если все волны группы обладают одной и той же фазовой скоростью, то волновой пакет движется с этой же скоростью и его форма при этом не изменяется. Однако если фазовая скорость зависит от длины волны, то интерференционная картина смещается и волновой пакет может двигаться быстрее или медленнее образующих его волн. Математический анализ этой интерференционной задачи (см. гл. 15) показывает, что групповая скорость определяется формулой

$$v = \frac{dv}{d(1/\lambda)}. \quad (6.7)$$

То обстоятельство, что мы получаем одну и ту же формулу как при вычислении потока мощности, так и при вычислении групповой скорости, можно понять на основе следующих соображений: в случае волнового пакета энергия сосредоточена на некотором определенном участке линии передачи и, следовательно, должна передаваться с той же скоростью, с которой перемещается волновой пакет. С другой стороны, при прохождении волнового пакета через некоторое определенное сечение линии в этом месте мы имеем дело с такими же условиями, как и в случае стационарного потока мощности. Энергия будет при этом передаваться вдоль линии с той же скоростью, что и при стационарном режиме. Следовательно, скорость потока мощности в стационарном случае должна определяться формулой для групповой скорости. (Искусственная линия передачи будет исследована более подробно в § 3 гл. 15.)



Фиг. 64. Ход вычислений при выводе формулы для скорости электрона.

Аналогичным образом можно исследовать распространение волн Блоха. Ход вычислений при выводе формулы, определяющей скорость электрона, приведен на фиг. 64. Для того чтобы определить скорость электрона, описываемого некоторой функцией Блоха, мы исследуем суперпозицию этой функции с другими функциями Блоха, которым отвечает приблизительно такое же значение P_{ω} . Таким путем мы получаем локализованный волновой пакет. Такой волновой пакет для электрона показан на фиг. 65. Произведение неопределенностей значений координаты и импульса при $t=0$ для этого пакета имеет минимальное значение $\Delta p \Delta x = h/4\pi$, допускаемое соотношением неопределенностей.

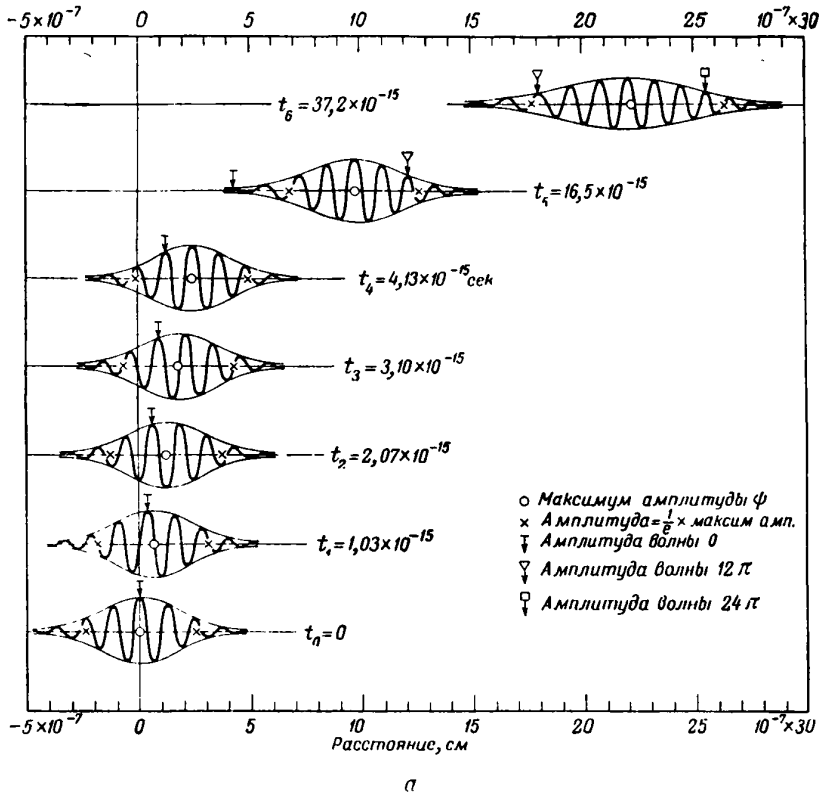
На фиг. 65,а показан ход вещественной части волновой функции $\psi(x, t)$ для различных моментов времени. Так как мы имеем дело с одномерным волновым пакетом, то ψ имеет размерность $см^{-1/2}$. Мнимая часть волновой функции имеет аналогичную форму; функция $\psi(x, t)$ показана огибающими линиями. Скорость группы волн в 2 раза больше фазовой скорости составляющих волн: гребни волн перемещаются на расстояние, равное одной длине волны за время от t_0 до t_1 , группа перемещается за это же время на 2 длины волны.

Значения импульсов, отвечающих пакету, лежат в пределах $\pm 8\%$ вблизи среднего значения импульса. Этот разброс значений составляющих скоростей обуславливает заметное расплывание пакета за время от t_0 до t_1 , при этом фронту пакета с течением времени соответствуют все более высокие значения компонент импульса, т. е. убывающие значения длины волны, а тыльной его части — убывающие значения компонент импульса.

Пунктирные кривые фиг. 65,б характеризуют постепенное перемещение линий, соответствующих постоянному значению комплексного фазового угла. Если угол кратен 2π , то волновая функция вещественна и положительна и отвечает гребням волн, показанным на фиг. 65,а. Максимум амплитуды вероятности перемещается со скоростью, равной удвоенной фазовой скорости; видно также, что линии, проходящие через точки, в которых амплитуда вероятности уменьшилась до $1/e^2$ своего максимального значения, с течением времени расходятся. Изгиб влево линий постоянной фазы вблизи $x=0$ при больших значениях t соответствует движению пакета влево, так как имеется некоторая вероятность того, что электрон будет обладать отрицательным импульсом. Среднее значение скорости соответствует электрону с энергией 1 эв (скорость $5,92 \cdot 10^7$ см/сек и длина волны $12,2 \cdot 10^{-8}$ см). Значение Δx при $t=0$ было произвольно выбрано равным 12,2 А. Это дает $\Delta p = \hbar/4\pi \Delta x = p_1/4\pi$, где p_1 отвечает энергии в 1 эв. Разброс значений импульса приводит к увеличению энергии от $p_1^2/2m$ до $(p_1^2 + \Delta p^2)/2m = p_1^2(1 + (1/16\pi^2))/2m = 1,0063$ эв. Пакет расплывается симметрично по отношению к моменту времени $t=0$ как в сторону положительных, так и отрицательных значений t . Теория волнового пакета рассматривается в § 5 гл. 15.

Легко убедиться в том, что, поскольку значения всех рассматриваемых волновых функций очень близки друг к другу (т. е. отвечающие им значения P_x отличаются весьма мало), они соответствуют примерно одной и той же средней скорости движения электрона, а следовательно, и волновой пакет должен обладать такой же скоростью.

Определение групповой скорости на основании исследования интерференционной картины является уже чисто математической задачей. Результат, который при этом получается, носит весьма общий характер и, как показано в § 2 гл. 15, применим к любой группе интерфери-

Фиг. 65. Движение волнового пакета, для которого $\Delta p \cdot \Delta x$

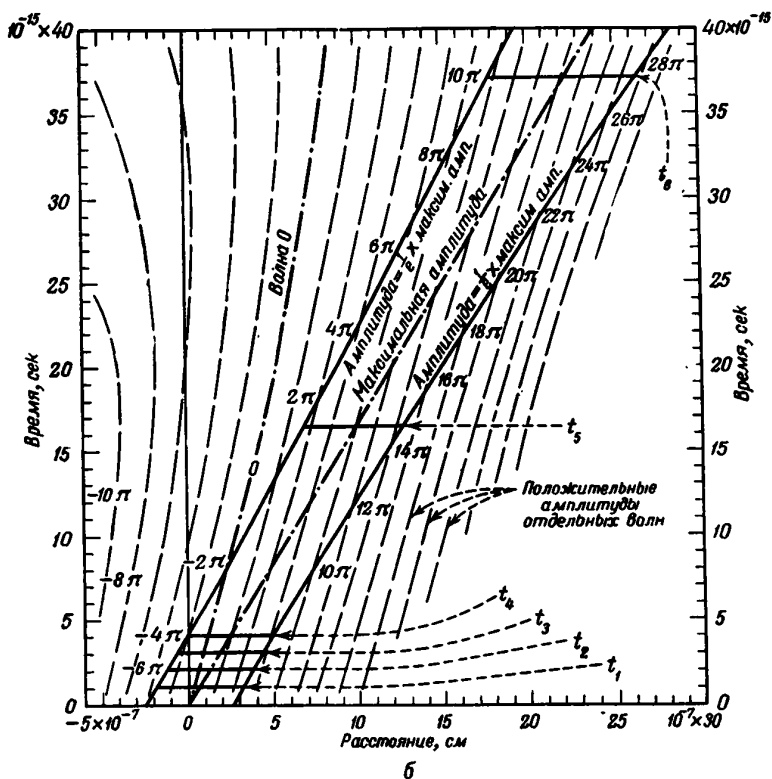
рующих волн. Этот результат согласуется, конечно, с тем, который был получен при рассмотрении потока плотности.

Сейчас мы исследуем полученную выше формулу для одномерного луча.

§ 4. СКОРОСТИ, СООТВЕТСТВУЮЩИЕ КВАНТОВЫМ СОСТОЯНИЯМ В ЗОНЕ БРИЛЛЮЭНА

Ранее уже было указано, что, пользуясь соотношениями $\mathcal{E} = \hbar \nu$ и $P_w = \hbar/\lambda$, можно преобразовать выражение для групповой скорости в одномерном случае к виду

$$v = \frac{d\mathcal{E}}{dP_w}. \quad (6.8)$$



имеет минимальное значение, в отсутствие внешних сил.

Этот результат находится в согласии с классическим выражением для функции Гамильтона. В классической механике энергия записывается в виде $\mathcal{H}(p_i, q_i)$, где q_i — координаты, а p_i — импульсы. Для случая одной частицы координаты и импульсы удобно выбрать следующим образом:

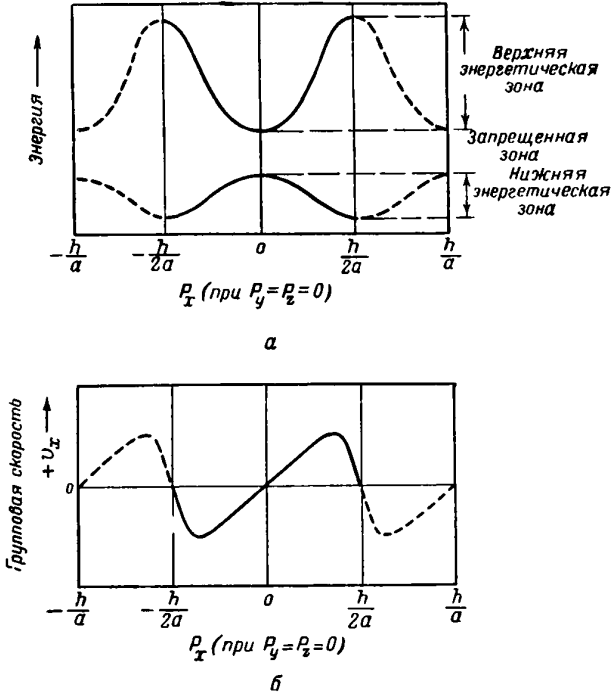
$$\begin{aligned} q_1 &= x, & q_2 &= y, & q_3 &= z, \\ p_1 &= m \frac{dx}{dt}, & p_2 &= m \frac{dy}{dt}, & p_3 &= m \frac{dz}{dt}. \end{aligned}$$

При решении других задач, например таких, как исследование движения планет, обобщенные координаты и импульсы могут быть выбраны другим способом. Соотношение между классическим и квантовомеханическим выражениями для функции Гамильтона исследуется в § 2

гл. 14. Канонические уравнения движения имеют вид

$$q_i = \frac{\partial \mathcal{E}}{\partial p_i} \quad \text{и} \quad \dot{p}_i = -\frac{\partial \mathcal{E}}{\partial q_i}.$$

В случае электрона, движущегося в одном измерении с постоянной потенциальной энергией \mathcal{V}_0 как в квантовой, так и в классической



Фиг. 66. Энергия и скорость в зонах Бриллюэна.

а — энергия как функция квази-импульса; б — групповая скорость для верхней энергетической зоны.

механике, кинетическая энергия определяется как $p^2/2m$, так что полная энергия электрона равна $\mathcal{E} = \mathcal{V}_0 + p^2/2m$. Групповая скорость равна при этом $v = p/m$, откуда $p = mv$, как и в классической механике¹⁾. Значение \mathcal{V}_0 , хотя и связано с частотой, отвечающей волнам Блоха, соотношением $v = \mathcal{E}/\hbar$, не влияет на групповую скорость. Другими словами, групповая скорость не зависит от выбора начала отсчета по шкале энергии, тогда как фазовая скорость зависит от

¹⁾ Мы обозначаем через p импульс в обычном классическом или квантовомеханическом смысле, а через P квази-импульс. — Прим. авт.

этого выбора. Однако поскольку ток определяется только групповой скоростью, последнее обстоятельство для нас несущественно.

Некоторые полученные нами результаты изображены графически на фиг. 66. Здесь энергия и групповая скорость представлены как функции импульса P_x . На фиг. 66,а показан ход энергии для двух энергетических зон, а на фиг. 66,б — групповая скорость для верхней энергетической зоны. В обоих случаях пунктирные линии характеризуют периодическую зависимость энергии и скорости от P_x ; при этом состояния вне зоны эквивалентны состояниям, расположенным внутри зоны, получаемым путем прибавления или вычитания величины \hbar/a из P_x .

Для случая трех измерений необходимо определить вектор скорости так, чтобы его компоненты определялись соотношениями, сходными с $v = d\mathcal{E}/dP_x$ для одного измерения. В гл. 15 показано, что эти компоненты равны:

$$v_x = \frac{\partial \mathcal{E}(P_x, P_y, P_z)}{\partial P_x}, \quad (6.9)$$

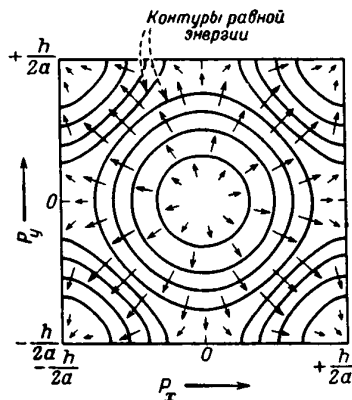
$$v_y = \frac{\partial \mathcal{E}(P_x, P_y, P_z)}{\partial P_y}, \quad (6.10)$$

$$v_z = \frac{\partial \mathcal{E}(P_x, P_y, P_z)}{\partial P_z}, \quad (6.11)$$

т. е. в векторном обозначении

$$\mathbf{v} = \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P}). \quad (6.12)$$

Геометрически это выражение связывает скорость v с поверхностями постоянной энергии в \mathbf{P} -пространстве. Чтобы найти скорость, соответствующую значению \mathbf{P} , в \mathbf{P} -пространстве проводят линию, перпендикулярную к поверхности постоянной энергии, проходящей через конец вектора \mathbf{P} . Затем вдоль этой линии, в направлении возрастающей энергии, вычисляют значение $d\mathcal{E}/d|P|$. Групповая скорость, соответствующая \mathbf{P} , имеет направление и величину получаемого таким путем вектора; при этом, конечно, предполагается, что оси P_x, P_y, P_z зоны Бриллюэна параллельны осям x, y, z кристалла¹⁾. Это построение дает распределение векторов скоростей в плоскости $P_x P_y$ зоны Бриллюэна, подобное изображенному на фиг. 67. Здесь длины стрелок пропорциональны величинам скоростей, соответствующих различным состояниям в \mathbf{P} -пространстве. Так как размерность \mathbf{P} и \mathbf{v} различна,



Фиг. 67. Зависимость скорости от квази-импульса.

Векторы указывают направление и относительную величину скоростей, соответствующих значениям \mathbf{P} . Разности энергий между соседними контурами одинаковы; энергия максимальна в углах зоны.

¹⁾ Мы только выразили словами смысл выражения $\nabla_{\mathbf{P}} \mathcal{E}$ -градиента \mathcal{E} в \mathbf{P} -пространстве. — *Прим. авт.*

то при этом существенны только направления и относительная длина стрелок.

На этом мы заканчиваем рассмотрение свойств волновых функций Блоха методом зон Бриллюэна (см. фиг. 50, стр. 155). Резюмируя коротко изложенное выше, можно сказать, что электрон, описываемый функцией Блоха, движется так, что его с равной вероятностью можно найти в любом месте кристалла. Средняя скорость, с которой он движется при этом циклически по кристаллу, равна v . В связи с этим следует отметить, что если электрон движется по кристаллу в соответствии с распределением скоростей, изображенным на фиг. 67, то значение P при этом не изменяется.

Иначе говоря, векторы, изображенные на фиг. 67, соответствуют скоростям в обычном пространстве, а не в P -пространстве. Значение P изменяется под действием приложенных внешних сил или в результате процессов рассеяния, которые будут рассмотрены нами ниже.

В качестве введения к следующим главам можно теперь сделать определенное заключение, относящееся уже непосредственно к проблеме электропроводности.

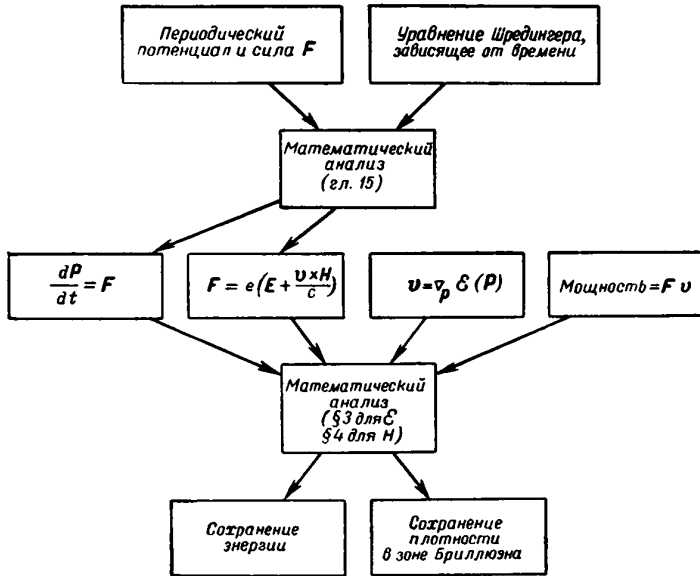
Исследование свойств зоны Бриллюэна показывает, что в случае заполненной зоны результирующий ток равен нулю. Это следует просто из соображений симметрии.

Для всех кристаллических решеток энергия электрона одинакова при значениях квази-импульса, равных P и $-P$, и, следовательно, скорости, соответствующие P и $-P$, равны и направлены противоположно. В заполненной зоне все разрешенные состояния заняты. Каждому электрону, движущемуся с данной скоростью, соответствует электрон, движущийся с такой же скоростью в противоположном направлении, и поэтому средняя скорость электронов равна нулю. Вопрос об аддитивности плотностей заряда и тока будет рассмотрен в § 7 и 8 гл. 15.

Глава 7

ЭЛЕКТРОНЫ И ДЫРКИ В ЭЛЕКТРИЧЕСКИХ И МАГНИТНЫХ ПОЛЯХ

Основная цель этой главы заключается в том, чтобы показать, что электрон в пустой зоне проводимости и дырка в заполненной зоне валентных связей под влиянием электрического и магнитного



Фиг. 68. Вывод соотношения $\dot{P} = F$ и следствия из него.

полей ведут себя подобно отрицательной и положительной частицам соответственно.

Доказательство базируется на основном законе, определяющем скорость изменения квази-импульса электрона под действием силы. Этот закон, аналогичный второму закону Ньютона, $F = ma$, выводится из основных положений квантовой механики, изложенных в гл. 15. В § 3 настоящей главы приводится вывод, хотя и недостаточно строгий, но тем не менее полезный для уяснения физического смысла

этого закона. На фиг. 68 приведена общая схема вывода закона $\dot{\mathbf{P}} = \mathbf{F}$ и некоторых следствий из него, рассматриваемых в различных параграфах настоящей главы.

§ 1. ДЕЙСТВИЕ СИЛЫ $\dot{\mathbf{P}} = \mathbf{F}$

Если на периодическое силовое поле \mathcal{U} , в котором электрон движется в кристалле, накладываются электрическое и магнитное поля, то волновые функции Блоха уже не будут больше решениями уравнения Шредингера, зависящего от времени. Однако если вектор \mathbf{P} , входящий в определение функций Блоха, изменяется со временем, согласно закону

$$\frac{d\mathbf{P}}{dt} \equiv \dot{\mathbf{P}} = \mathbf{F}, \quad (7.1)$$

то эти волновые функции являются решениями уравнения Шредингера. Для определения \mathbf{F} воспользуемся классическим выражением. Электрическое поле \mathbf{E} действует на электрон с силой $-e\mathbf{E}$. Сила, с которой магнитное поле \mathbf{H} действует на электрон, направлена под прямым углом к \mathbf{H} и к направлению средней скорости электрона \mathbf{v} . Это приводит к следующему выражению для силы, действующей на электрон:

$$\mathbf{F} = -e \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{H}}{c} \right), \quad (7.2)$$

где c — скорость света¹⁾.

Сходство между уравнением (7.1) и вторым законом Ньютона, $\mathbf{F} = m\mathbf{a} = d(m\mathbf{v})/dt$, послужило основанием, чтобы назвать \mathbf{P} квазиимпульсом.

При действии на электрон постоянной силы точка в зоне Бриллюэна, соответствующая определенному состоянию электрона, так называемая „изображающая“ точка, движется равномерно в направлении действия силы. Достигая поверхности зоны, она становится эквивалентной точке, расположенной на противоположной стороне зоны, и продолжает движение внутрь зоны от этой эквивалентной точки.

Уравнение $\dot{\mathbf{P}} = \mathbf{F}$ означает, что возможным квантовым состояниям отвечают *все* точки в зоне Бриллюэна, а не только точки основной \mathbf{P} -решетки, определяемой уравнениями (5.20)—(5.22), являющимися следствием наличия периодических граничных условий. Это обстоятельство связано с тем, что наложение электрического и магнитного полей, вообще говоря, изменяет периодический характер граничных условий. Например, в прямом проводе, по которому течет ток, напря-

¹⁾ Для уяснения смысла произведения $\mathbf{v} \times \mathbf{H}$ в некотором частном случае см. фиг. 80, стр. 249. — *Прим. авт.*

жение на одном конце провода не будет равно напряжению на другом, и поэтому бессмысленно требовать выполнения периодических граничных условий для этих двух точек. Однако провод можно согнуть в кольцо и в таком случае периодические граничные условия будут применимы к координате, откладываемой вдоль кольца. В таком кольце можно создать электрическое поле, используя его в качестве коротко замкнутой вторичной обмотки трансформатора. Подобная модель рассматривается в § 9 гл. 15, где будет показано, что хотя индуцируемое электрическое поле и изменяет волновую функцию, тем не менее число длин волн, укладываемых вдоль кольца, остается неизменным. Энергия электрона и переносимый электроном ток изменяются, однако, под действием поля. Это изменение происходит точно так, как если бы периодические граничные условия не принимались во внимание и волновая функция менялась, согласно уравнению $\dot{\mathbf{P}} = \mathbf{F}$.

Более детальный вывод очень сложен в математическом отношении. Однако результат этого вывода можно сформулировать следующим образом: для периодической модели энергии и скорости изменяются как раз так, как если бы решетка изображающих точек, соответствующих возможным квантовым состояниям, смещалась под действием электрического поля, согласно закону $\dot{\mathbf{P}} = \mathbf{F}$. Поэтому влияние полей мы и будем рассматривать с этой точки зрения.

Тот же результат можно получить также и при рассмотрении волнового пакета, построенного из функций Блоха, характеризуемых приблизительно одинаковым значением импульса \mathbf{P} . Электрическое и магнитное поля действуют на волновой пакет так, что среднее значение квази-импульса $\dot{\mathbf{P}}$ изменяется по закону $\dot{\mathbf{P}} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{H}/c)$. Соответствующим образом ведет себя и дырочный волновой пакет (см. § 6 и 8 гл. 15).

В соответствии с этими фактами можно считать, что закон $\dot{\mathbf{P}} = \mathbf{F}$ действительно является обобщением второго закона Ньютона для электронов в кристалле. Так как при применении этого закона ко всем электронам, занимающим квантовые состояния в зоне Бриллюэна, речь идет о перемещении всех занятых состояний в зоне, то мы можем считать его в равной мере применимым и к вакантным состояниям. Это предположение не вносит никакой ошибки, ибо если состояния не заняты, то они нас непосредственно не интересуют, если же они заняты, то их поведение описывается в этом случае правильно. Предположение о том, что вакантные состояния подчиняются тому же закону, что и занятые, имеет ряд преимуществ при рассмотрении движения дырок и используется в § 6 настоящей главы.

Наконец, следует отметить, что при наличии магнитного поля зависимость между классическим импульсом \mathbf{p} и скоростью \mathbf{v} , а также между соответствующими квантовомеханическими операторами уже не имеет вида $\mathbf{p} = m\mathbf{v}$. Зависимость \mathbf{p} от \mathbf{v} для этого случая

рассматривается в § 6 гл. 15. Однако соотношения

$$\xi = \xi(\mathbf{P}) \text{ и } \mathbf{v} = \nabla_{\mathbf{P}} \xi(\mathbf{P}) \quad (7.3)$$

не изменяются в случае наложения магнитного поля, так что совокупность уравнений (7.1)—(7.3) вполне достаточна для описания движения электрона в электрическом и магнитном полях, по крайней мере, с точностью до членов, линейных по отношению к \mathbf{E} и \mathbf{H} .

§ 2. ТЕОРЕМА СОХРАНЕНИЯ КВАНТОВЫХ СОСТОЯНИЙ

Движение изображающих точек в зоне Бриллюэна под действием электрического и магнитного полей подчиняется теореме сохранения квантовых состояний. Рассмотрим сначала влияние электрического поля, а влияние магнитного поля исследуем ниже.

В случае электрического поля результат очевиден. Так как сила, действующая на электрон, равна $-e\mathbf{E}$ и одинакова для всех значений \mathbf{P} , то „решетку“ возможных квантовых состояний можно рассматривать просто как „твердое тело“, совершающее трансляционное движение через зону Бриллюэна. Очевидно, что при таком перемещении полное число возможных изображающих точек в зоне Бриллюэна, а также плотность их распределения не изменяются¹⁾.

§ 3. ПРИНЦИП СОХРАНЕНИЯ ЭНЕРГИИ ПРИ НАЛИЧИИ ЭЛЕКТРИЧЕСКОГО ПОЛЯ

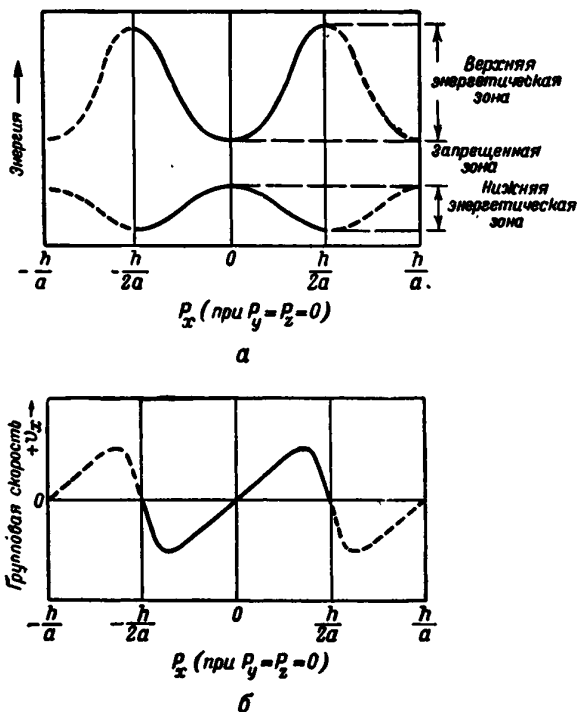
Если ускорение частицы происходит под действием силы \mathbf{F} , то следует ожидать, что увеличение энергии частицы равно работе, произведенной над ней указанной силой. С одной стороны, энергия, получаемая частицей под действием силы за единицу времени, равна $\mathbf{v} \cdot \mathbf{F}$, а с другой стороны, скорость изменения ее энергии определяется соотношением $\dot{\xi} = \nabla_{\mathbf{P}} \xi \cdot \dot{\mathbf{P}}$. Согласно рассмотренным выше законам, $\dot{\mathbf{P}} = \mathbf{F}$ и $\mathbf{v} = \nabla_{\mathbf{P}} \xi$, так что оба эти значения энергии равны и, следовательно, при движении электрона под действием силы закон сохранения энергии выполняется.

Однако поведение электрона в других отношениях существенно отличается от поведения классической частицы в отсутствие внешних сил. Некоторые особые свойства электрона, на описании которых мы вскоре остановимся, в большой степени маскируются случайными процессами рассеяния, всегда имеющими место в реальных кристаллах (за исключением металлов, находящихся в сверхпроводящем состоянии). Тем не менее эти свойства следует рассмотреть подробнее, так как они непосредственно объясняются теорией (см. фиг. 66 и 67) и,

¹⁾ При этом волновые функции также остаются ортогональными.—
Прим. авт.

кроме того, иллюстрируют принципы, которыми мы воспользуемся в дальнейшем.

Поэтому рассмотрим движение электрона в идеальном кристалле под влиянием постоянной силы F , действующей в направлении x .



Фиг. 69. Энергия и скорость в зонах Бриллюэна.

a — энергия как функция квази-импульса; b — групповая скорость для верхней энергетической зоны.

Если этот электрон находился вначале в самом нижнем энергетическом состоянии верхней зоны, соответствующем точке $P_x = 0$ на фиг. 69, то значение P_x с течением времени будет возрастать с постоянной скоростью $\dot{P}_x = F_x$. Скорость электрона вначале также будет возрастать, изменяясь прямо пропорционально импульсу, так что до тех пор, пока кривая, изображенная на фиг. 69, b , может быть аппроксимирована прямой линией, скорость возрастает равномерно со временем, как это имело бы место для классической частицы, например для падающего тела. При этих условиях, как будет показано в дальнейшем, движение электрона может быть действительно описано

с помощью классических представлений. Однако по истечении некоторого промежутка времени скорость электрона достигает значения, соответствующего максимуму на кривой, изображенной на фиг. 69, б, и после этого начинает уменьшаться. Таким образом, в той области значений P , где dv_{ω}/dP_{ω} отрицательно, электрон ведет себя так, как если бы он обладал *отрицательной* массой, т. е. в этой области непрерывное действие силы в направлении положительной полуоси x фактически уменьшает скорость частицы в этом направлении.

Такое поведение электрона не имеет аналогии в классической динамике частиц и обусловлено законами волновой механики, управляющими движением электронов. Если электрическое поле действует на электрон достаточно долго, то P_{ω} достигает значения $\hbar/2a$, после чего скорость электрона становится отрицательной, как это показано на рисунке пунктирной линией, и электрон движется в направлении, противоположном направлению силы. (Такое движение электрона эквивалентно перемещению точки P_{ω} обратно внутрь зоны Бриллюэна от соответствующей точки, расположенной на противоположной стороне зоны, в согласии с периодическим изменением P_{ω} .) Наконец, скорость электрона достигает максимального значения, соответствующего минимуму на кривой фиг. 69, б, в отрицательном направлении и затем уменьшается по абсолютной величине до нуля, после чего цикл снова повторяется. Другими словами, частица совершает колебательное движение. Если бы из волновых функций, которым соответствует примерно одно и то же значение импульса, построить волновой пакет, то такой волновой пакет совершал бы в кристалле колебательное движение.

В действительности положение волнового пакета в зависимости от времени изображалось бы кривой точно такой же формы, как и зависимость его энергии от импульса. Это вытекает из следующих уравнений:

$$\begin{aligned} x &= \int v_x dt = \int \frac{\partial \mathcal{E}}{\partial P_x} dt = \int \frac{\partial \mathcal{E}}{\partial P_x} \frac{dP_x}{F_x}, \\ x &= \frac{\mathcal{E}(P_x)}{F_x} + \text{const} = \frac{\mathcal{E}(F_x t)}{F_x} + \text{const}, \end{aligned} \quad (7.4)$$

так как если сила F_x приложена в момент $t=0$, то $P_x = F_x t$ и $dt = dP_x/F_x$. Работа, совершенная над частицей, равна $x F_x$, так что это уравнение не только определяет положение электрона как функцию времени, но выражает также и закон сохранения энергии. Период движения, равный \hbar/aF_x , есть время, в течение которого точка, соответствующая импульсу P_x , проходит через всю зону Бриллюэна.

Колебательный характер движения электрона в этом случае полезно учесть при вычислении вероятностей возбуждения электрона из одной зоны в другие под действием сильного электрического

поля [1]. Эта задача выходит, однако, за рамки этой книги¹⁾. Приведенные выше результаты следует рассматривать как иллюстрацию теории зон Бриллюэна. Как мы увидим ниже, процессы столкновений, вызывающие случайные переходы электрона из одного состояния в другое, происходят очень часто. Промежутки времени между ними меньше периода движения электрона, и, таким образом, рассмотренные здесь аномалии в поведении электрона не играют существенной роли для явлений обычной электропроводности.

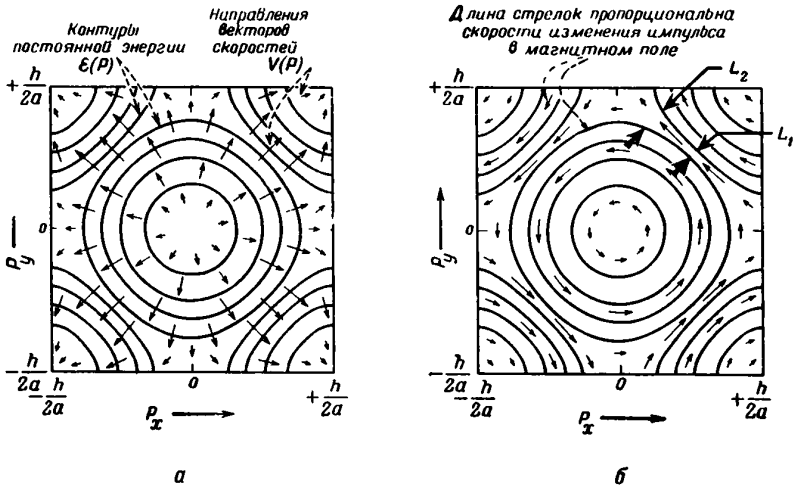
§ 4. ДЕЙСТВИЕ МАГНИТНОГО ПОЛЯ

Рассмотрим теперь силу, с которой магнитное поле действует на движущийся электрон, и изменение импульса электрона, вызванное этой силой. Ограничимся рассмотрением квази-импульса, лежащего в плоскости $P_x = 0$, в случае, когда магнитное поле направлено по оси z . На фиг. 70, а в плоскости $P_x P_y$ изображена двумерная зона Бриллюэна с нанесенными на ней контурами постоянной энергии. Следует напомнить, что скорость электрона пропорциональна градиенту энергии \mathcal{E} в зоне и, таким образом, ее направление перпендикулярно к контурам постоянной энергии. Направления и относительные величины скорости указаны, как и на фиг. 67, стрелками.

Если магнитное поле приложено в направлении $\perp z$, то под действием силы $-e\mathbf{v} \times \mathbf{H}/c$ импульс электрона изменится. Так как сила перпендикулярна направлению скорости \mathbf{v} , то импульс будет изменяться так, как это указано стрелками на фиг. 70, б. В результате возникает перемещение изображающих точек вдоль контуров постоянной энергии. Таким образом, электроны не получают энергии от магнитного поля — результат, согласующийся с известным фактом, что сила, с которой магнитное поле \mathbf{H} действует на электрон, перпендикулярна к направлению его движения и поэтому не совершает работы. Как уже указывалось в связи с рассмотрением принципа Паули, при таком перемещении плотность изображающих точек не изменяется. Этот результат непосредственно вытекает из формулы, определяющей изменение импульса. Таким образом, если мы рассмотрим сечения L_1 и L_2 на фиг. 70, б, то можно показать, что поток частиц в сечении L_1 равен потоку через сечение L_2 ; если, например, сечение L_1 в 2 раза меньше сечения L_2 , то скорость электрона в сечении L_1 будет в 2 раза больше, чем в сечении L_2 , и, следовательно, сила, действующая на электрон, и скорость

¹⁾ Квантовомеханическая задача о движении электронов в периодическом поле кристаллической решетки при наличии сильного электрического поля и дополнительного взаимодействия, вызывающего переходы электронов из одних состояний в другие, рассматривалась Бардином и Шоками [3]. — *Прим. ред.*

изменения его импульса будут также в 2 раза больше. Отсюда следует, что через сечения L_1 и L_2 за одинаковые промежутки времени будут перемещаться одинаковые объемы пространства импульсов. Следовательно, если квантовые состояния были распределены вначале равномерно, то плотность их между сечениями L_1 и L_2 не будет увеличиваться, так как потоки через эти сечения равны. (В случае трех измерений этот результат можно сформулировать в виде равенства $\text{div } \dot{\mathbf{P}} = 0^1$.)



Ф и г. 70. Влияние магнитного поля на энергию и скорость \mathbf{v} в зоне Бриллюэна. а — контуры равной энергии и распределение скоростей в плоскости $P_z = 0$; б — изменение квази-импульса в магнитном поле, приложенном в направлении $+z$. $\dot{\mathbf{P}} = -\frac{e}{c} \mathbf{V} \times \mathbf{H}$ (энергия максимальна в углах зоны).

Отсюда можно сделать общий вывод, что само по себе магнитное поле вызывает только перемещение изображающих точек вдоль соответствующих поверхностей постоянной энергии, не вызывая при этом увеличения их плотности и не нарушая равновесного распределения.

Эти общие результаты, относящиеся к изменению квази-импульса электрона \mathbf{P} в зоне Бриллюэна под влиянием электрического и магнитного полей, будут в дальнейшем использованы при рассмотрении задачи о движении одного электрона в пустой зоне проводимости

¹⁾ $\nabla_{\mathbf{P}} \cdot (\nabla_{\mathbf{P}} \mathcal{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla_{\mathbf{P}} \times \nabla_{\mathbf{P}} \mathcal{E}) + \nabla_{\mathbf{P}} \mathcal{E} \cdot (\nabla_{\mathbf{P}} \times \mathbf{H}) = 0 - 0$. — Прим. авт.

или одиночной дырки в заполненной зоне валентных связей. При этом мы будем предполагать, что электрон и дырка движутся в идеальном кристалле.

§ 5. ПОВЕДЕНИЕ ИЗБЫТОЧНОГО ЭЛЕКТРОНА В КРИСТАЛЛЕ; ЭФФЕКТИВНАЯ МАССА И ПЕРЕСТРОЙКА ЗОНЫ БРИЛЛЮЭНА

При применении только что рассмотренной теории к расчету электропроводности и эффекта Холла необходимо учитывать специфические свойства поверхностей постоянной энергии в зоне Бриллюэна. Расчет формы и размера поверхностей постоянной энергии был выполнен для ряда металлов при помощи приближенных методов. Однако для полупроводников до сих пор такие детальные расчеты произведены не были¹⁾. По этой причине при рассмотрении полупроводников часто приходится довольствоваться рядом предположений. Можно, однако, воспользоваться при этом общими принципами, подобными уже применявшимся нами при определениях формы и размеров зоны Бриллюэна. Эти принципы основываются главным образом на условиях симметрии, которым должны удовлетворять поверхности постоянной энергии, чтобы соответствовать симметрии кристалла. Кроме того, было также выполнено несколько количественных расчетов поверхностей постоянной энергии. Наконец, существуют экспериментальные данные, которые могут быть использованы для проверки законности предположений, касающихся формы этих поверхностей.

На основании всех этих данных мы будем предполагать, что у полупроводников, обладающих структурой алмаза, поверхности постоянной энергии, соответствующие квантовым состояниям с наибольшими энергиями в зоне валентных связей и с наименьшими энергиями в зоне проводимости, имеют сферическую форму²⁾. Только эти и очень близкие к ним квантовые состояния представляют интерес при изучении явлений проводимости. Как это будет более подробно изложено в связи с рассмотрением статистики Ферми—Дирака, это объясняется тем обстоятельством, что при тепловом возбуждении электрона передается энергия всего лишь порядка $0,03 \text{ эв}$, тогда как ширина зоны может составлять несколько эв . Следовательно, интересующие нас квантовые состояния лежат в очень узкой области энергий, вблизи края энергетической зоны. Ширина этой области составляет всего несколько процентов от полной ширины зоны.

¹⁾ Расчет формы и размеров поверхностей постоянной энергии в зонах Бриллюэна был выполнен для ряда полупроводящих веществ в работах [4—6]. — *Прим. ред.*

²⁾ См., например, [2]. Однако имеются веские основания полагать, что одна из зон вырождена в отношении ее граничной энергии, что связано с соответствующими осложнениями формы поверхностей постоянной энергии. Мы вернемся к этому вопросу в гл. 12. — *Прим. авт.*

При исследовании движения электронов и дырок в кристалле мы должны рассматривать как дно зоны проводимости, так и верхний край зоны валентных связей, причем в каждом из этих случаев интересующие нас энергетические состояния могут быть расположены либо в центре, либо в углах зоны Бриллюэна. Как будет показано ниже, все эти случаи можно исследовать с помощью одного и того же метода. Поэтому сначала рассмотрим самый простой из них, а в дальнейшем установим его связь с другими, более сложными случаями.

Предположим, что дну зоны проводимости соответствует энергия \mathcal{E}_0 в центре зоны Бриллюэна, как это показано на фиг. 60 и 67; при этом энергия имеет максимальные значения в углах зоны. Из общих теоретических соображений следует, что уравнение для сферических поверхностей постоянной энергии должно иметь вид

$$\mathcal{E} = \mathcal{E}_0 + K(P_x^2 + P_y^2 + P_z^2) = \mathcal{E}_0 + KP^2, \quad (7.5)$$

где P — абсолютная величина квази-импульса \mathbf{P} . Составляющие скорости частицы равны при этом

$$v_x = \frac{\partial \mathcal{E}}{\partial P_x} = 2KP_x, \quad (7.6a)$$

$$v_y = \frac{\partial \mathcal{E}}{\partial P_y} = 2KP_y, \quad (7.6b)$$

$$v_z = \frac{\partial \mathcal{E}}{\partial P_z} = 2KP_z, \quad (7.6в)$$

или в векторном обозначении

$$\mathbf{v} = 2K\mathbf{P}. \quad (7.7)$$

Если на частицу действует сила \mathbf{F} , то ее скорость изменяется по закону

$$\dot{\mathbf{v}} = 2K\dot{\mathbf{P}} = 2K\mathbf{F}. \quad (7.8)$$

Последнее выражение аналогично второму закону Ньютона, записанному в форме

$$\mathbf{F} = m_n \mathbf{a} \quad \text{или} \quad \dot{\mathbf{v}} = \frac{\mathbf{F}}{m_n}, \quad (7.9)$$

если рассматривать величину $1/2K$ как эффективную массу m_n избыточного электрона. Другими словами, под действием силы скорость электрона изменяется точно так, как если бы его масса была равна $1/2K$.

Значение $1/2K$ для состояний в зоне проводимости (или $1/2K'$ в зоне валентных связей) имеет тот же порядок величины, что и масса свободного электрона, как это будет показано в § 6 гл. 14 при исследовании общих свойств волновых функций. Из экспериментальных данных, полученных для кремния и германия, следует, что эффективная масса электрона в кристалле очень близка к массе

свободного электрона¹⁾. Для большинства вопросов, рассматриваемых в этой главе, различие между эффективной массой и массой свободного электрона не имеет существенного значения, и поэтому нет необходимости вводить для эффективной массы специальное обозначение. Ввиду этого мы будем писать $1/2m$ вместо K^2).

На основании этого предположения можно сделать вывод, что электрон, отвечающий состоянию в зоне Бриллюэна с минимальной энергией, ускоряется под действием силы точно так же, как и свободный электрон, для которого

$$\dot{\mathbf{v}} = \frac{\mathbf{F}}{m}. \quad (7.10)$$

Используем эту зависимость в дальнейшем при рассмотрении вопросов теории электропроводности.

Посмотрим, что произойдет, если состояния с наименьшей энергией отвечают углам зоны Бриллюэна. Для этой цели будет достаточно рассмотреть двухмерный случай, иллюстрируемый фиг. 71, где для большей ясности контуры постоянной энергии, соответствующие большим энергиям, не изображены. Фиг. 71, *a* иллюстрирует результат действия силы, ускоряющей электрон и переводящей его из состояния с квази-импульсом \mathbf{P}_0 в состояние с квази-импульсом \mathbf{P}_g . Каждый раз, когда квази-импульс достигает значения, соответствующего одной из точек на границе зоны, волновая функция становится эквивалентной волновой функции, отвечающей соответствующей точке на противоположной границе зоны. В этом случае необходимо отметить, что когда изображающая точка достигает границы зоны, скорость \mathbf{v} параллельна этой границе, так что, как указывалось выше, в направлении, перпендикулярном к границе, ток не идет. Если для удобства перестроить зону так, как это показано на фиг. 71, *b*, то изменение квази-импульса от значения, равного \mathbf{P}_0 , до значения \mathbf{P}_g изобразится непрерывной прямой линией. Если квази-импульс электрона определяется вектором \mathbf{P}' , начало которого лежит в центре фиг. 71, *b*, то

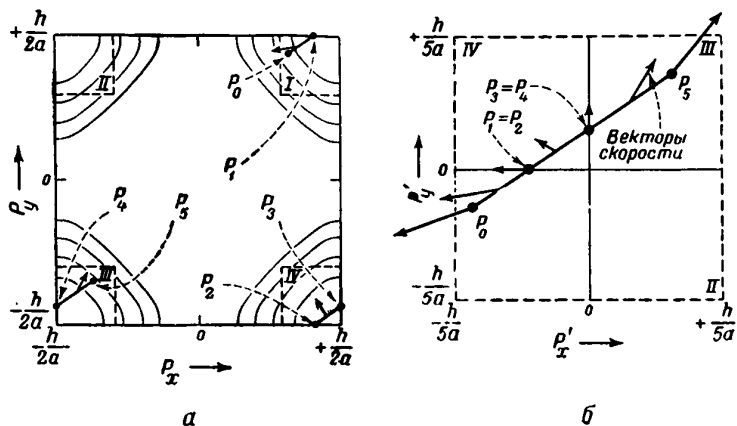
$$\mathcal{E} = \mathcal{E}_0 + K'(P_x'^2 + P_y'^2 + P_z'^2) = \mathcal{E}_0 + K'\mathbf{P}'^2. \quad (7.11)$$

Это следует из общих теоретических соображений, касающихся формы поверхностей постоянной энергии вблизи углов зоны. Так как квази-импульс \mathbf{P}' под действием силы (см. фиг. 71, *b*) изменяется точно

¹⁾ В последнее время опубликовано несколько работ, в которых делались оценки эффективной массы носителей тока в германии [7, 8].— *Прим. ред.*

²⁾ Электронный волновой пакет, построенный из волновых функций, соответствующих квантовым состояниям электрона с очень малой эффективной массой, приобретает в кристалле значительно большее ускорение, чем в вакууме. В § 6 гл. 14 рассматривается, каким образом взаимодействие электрона с кристаллом способствует его ускорению.— *Прим. авт.*

так же, как \mathbf{P} в центре зоны в только что рассмотренном нами случае, то очевидно, что мы приходим здесь к аналогичным результатам и величину $1/2K'$ можно опять рассматривать как эффективную массу. Из рассмотрения фиг. 71 следует, что скорость электрона меняется



Фиг. 71. Зона Бриллюэна, когда энергия в углах минимальна.

a — действие силы, ускоряющей электрон (от \mathbf{P}_0 до \mathbf{P}_2); b — квадранты перегруппированы (увеличено), чтобы яснее показать непрерывность изменения импульса и скорости.

непрерывно с изменением \mathbf{P}' и не претерпевает разрыва непрерывности на краю зоны Бриллюэна.

Таким образом, сила, действующая на электрон, изменяет его квази-импульс, согласно уравнению

$$\dot{\mathbf{P}} \text{ или } \dot{\mathbf{P}}' = \mathbf{F}, \quad (7.12)$$

и так как скорость электрона связана с \mathbf{P} или \mathbf{P}' уравнением

$$\mathbf{v} = 2K\mathbf{P} = \frac{\mathbf{P}}{m},$$

или

$$\mathbf{v} = 2K'\mathbf{P}' = \frac{\mathbf{P}'}{m}, \quad (7.13)$$

то изменение скорости определяется уравнением

$$\dot{\mathbf{v}} = \frac{\mathbf{F}}{m} \text{ или } \mathbf{F} = m\mathbf{a}, \quad (7.14)$$

где $\mathbf{a} \equiv \dot{\mathbf{v}}$ — ускорение.

Энергию можно выразить теперь через \mathbf{P} , \mathbf{P}' или \mathbf{v} :

$$\mathcal{E} = \mathcal{E}_0 + \frac{\mathbf{P}^2 \text{ или } \mathbf{P}'^2}{2m} = \mathcal{E}_0 + \frac{m\mathbf{v}^2}{2}. \quad (7.15)$$

Таким образом, энергия записывается в виде суммы постоянного члена и члена, который формально аналогичен кинетической энергии, соответствующей скорости v . В действительности член $mv^2/2$ характеризует как изменение потенциальной энергии электрона в периодическом поле кристалла, так и изменение его кинетической энергии при изменении волновой функции Блоха, связанном с изменением P . Хотя выражением (7.15) для энергии можно пользоваться так, как если бы этот член определял изменение кинетической энергии избыточного электрона, это обстоятельство следует рассматривать скорее как своего рода математическое упрощение, а не как выражение определенной физической закономерности. Как будет показано в § 6 этой главы, аналогичное упрощение может быть сделано также и в случае дырки.

Покажем сейчас, что под влиянием электрического и магнитного полей электрон в зоне Бриллюэна ведет себя действительно так же, как и классический электрон, поскольку речь идет об его ускорении. Этот результат непосредственно следует из уравнения $F = ma$ и выражения для силы, приведенного в § 7,

$$-e \left(E + \frac{v \times H}{c} \right) = F = m\dot{v} = ma. \quad (7.16)$$

Это уравнение показывает, что для электрона в кристалле скорость изменения v в электрическом и магнитном полях связана с E , H и v точно такой же зависимостью, как и в случае свободного классического электрона.

Этот результат кажется столь очевидным, что, может быть, стоит кратко напомнить о том, каким способом он был получен. В следующем параграфе мы исследуем случай дырки, основываясь при этом на том же методе рассмотрения, и придем к заключению, что перед выражением, определяющим зависимость ускорения a от E , H и v , должен стоять знак плюс.

По существу, мы покажем, что согласно квантовой теории, электрон и в отсутствие внешних сил может находиться в состоянии движения в периодическом силовом поле кристалла, обладая при этом квази-импульсом P и скоростью v , не изменяющимися во времени (кристалл при этом предполагается идеальным). Этот результат сам по себе удивительный, так как с точки зрения классической теории электрон, движущийся в силовом поле кристалла, отклонялся бы вследствие столкновений столь часто, что его скорость принимала бы практически любые значения и в среднем равнялась бы нулю.

Таким образом, исходя из волнового уравнения Шредингера, мы пришли к представлению о том, что электрон, движущийся в кристалле, может обладать квази-импульсом и средней скоростью, не зависящими от времени. Мы показали также, что, как это следует из квантовой механики, $\dot{P} = F$.

Наконец, в качестве некоторого приближения, мы пришли к заключению о том, что вблизи той точки зоны, которая отвечает минимальной энергии, скорость $\mathbf{v} = 2K\mathbf{P}$, и, в качестве следующего приближения, что $\mathbf{v} = \mathbf{P}/m$ ¹⁾. (В следующих главах мы введем обозначения m_n и m_p соответственно для эффективных масс электрона и дырки, а в гл. 12 рассмотрим более сложный случай.) Далее, мы определили \mathbf{F} с помощью классического выражения для силы, действующей на электрон с зарядом $-e$ и скоростью \mathbf{v} , движущийся в полях \mathbf{E} и \mathbf{H} ; этот последний этап следует в известном смысле рассматривать как некоторое дополнительное предположение. На самом деле это выражение, определяющее силу, уже содержится в уравнении Шредингера и, следовательно, не является в действительности новым предположением.

Приведенными выше замечаниями мы заканчиваем рассмотрение вопроса о движении избыточного электрона в зоне проводимости. Из теории зон Бриллюэна следует, что такой электрон движется точно так же, как свободный классический электрон, при условии, что его энергия всегда столь мала, что он остается вблизи дна зоны проводимости. В следующих параграфах мы покажем, что движение дырки в зоне валентных связей аналогично движению положительно заряженного электрона.

§ 6. ДВИЖЕНИЕ ДЫРКИ

Как будет показано ниже, дырка является в действительности абстрактным понятием, которое, однако, весьма удобно ввести в рассмотрение для описания движения электронов. Существенную роль при этом играет тот факт, что в целиком заполненной зоне Бриллюэна, в которой каждое разрешенное состояние занято, результирующий ток равен нулю. Это обстоятельство позволяет найти законы движения дырки непосредственно из рассмотрения движения, которое совершал бы отсутствующий электрон, если бы он находился в кристалле²⁾.

В конце гл. 6 было указано, что среднее значение скорости для всех электронов в зоне Бриллюэна равно нулю, так как каждой из скоростей отвечает равная ей по величине, но противоположно направленная скорость. Это заключение основывалось на соображениях симметрии и следовало из теории основной \mathbf{P} -решетки зоны Бриллюэна (см. фиг. 59). После этого мы предположили, что система

¹⁾ При применении этой теории к практическим случаям следует помнить, что неограниченные изменения \mathbf{P} , вызываемые электрическим полем, лимитируются влиянием столкновений, так что рассмотрению подлежат только малые значения \mathbf{P} , и поэтому наше приближение вполне пригодно. — *Прим. авт.*

²⁾ Математический анализ этой задачи приведен в § 7 и 8 гл. 15. — *Прим. авт.*

изображающих точек может перемещаться через зону Бриллюэна, причем их плотность остается постоянной. Это допущение не вносит, однако, никаких изменений в наше заключение о том, что в случае заполненной зоны ток равен нулю, так как при сложении скоростей всех электронов их значения сокращаются попарно весьма точно. И действительно, чем больше размеры кристалла и чем больше число электронов в зоне, тем меньше будет результирующая скорость. Таким образом, электронным током можно пренебречь по сравнению с током, обусловленным наличием дырки в зоне. Эти соображения лежат в основе исследования движения дырки.

Ход рассуждения при этом следующий. Рассмотрим сначала заполненную зону Бриллюэна. Как указывалось в § 1 гл. 6, каждому электрону в кристалле соответствует плотность тока, равная $(-e/V)\mathbf{v}$. Рассмотрим отдельное квантовое состояние и назовем его состоянием s . Ему отвечает скорость электрона \mathbf{v}_s . Напишем теперь выражение для суммарного тока, создаваемого всеми электронами зоны, но при этом выделим член, соответствующий току, создаваемому электроном, находящимся в состоянии s . Так как зона заполнена, то суммарный ток равен нулю и, следовательно,

$$\sum_{i \neq s} \frac{-e}{V} \mathbf{v}_i + \frac{-e}{V} \mathbf{v}_s = 0, \quad (7.17)$$

откуда

$$\sum_{i \neq s} \frac{-e}{V} \mathbf{v}_i = \frac{e}{V} \mathbf{v}_s. \quad (7.18)$$

Левая часть этого уравнения представляет собой просто суммарный ток, созданный электронами зоны Бриллюэна, в которой заняты все состояния, за исключением состояния s . Мы видим, что этот ток в точности равен току, который получился бы, если бы в пустой зоне находился лишь один *положительно* заряженный электрон, движущийся со скоростью \mathbf{v}_s . Это лемма, из которой следует, что дырки движутся подобно положительно заряженным электронам. Сформулируем этот результат словесно:

Ток, обусловленный дыркой, соответствует движению положительного заряда $+e$ со скоростью, отвечающей незанятому квантовому состоянию

$$I(\text{дырки}) = \frac{+e}{V} \mathbf{v}_s. \quad (7.19)$$

Наша дальнейшая задача заключается в том, чтобы выяснить, как изменяется ток в зоне под влиянием электрического и магнитного полей. Для этого следует рассмотреть зависимость энергии электрона, находящегося вблизи верха энергетической зоны, от его импульса. Как и в случае электрона, находящегося на дне энергетической зоны, возможностью вырождения энергии, соответствующей верхнему краю

зоны, мы пренебрегаем и будем считать, что поверхность постоянной энергии имеет сферическую форму и энергия изменяется пропорционально P^2 . Перестройка зоны для того случая, когда наибольшее значение энергии отвечает ее углам, осуществляется точно таким же способом, как раньше, и отношение $1/2K$ снова принимается равным массе электрона m . Однако здесь имеется существенное отличие от случая электрона, находящегося на дне зоны проводимости, так как теперь энергия должна уменьшаться по мере удаления \mathbf{P} или \mathbf{P}' от точки, отвечающей максимальному значению энергии. Следовательно,

$$\mathcal{E} = \mathcal{E}_0 - K(P_x^2 + P_y^2 + P_z^2) = \mathcal{E}_0 - \frac{P^2}{2m_p}, \quad (7.20)$$

где m_p — эффективная масса дырки. Мы не делаем здесь различия между значениями m_p и m . Из рассмотрения (7.20) можно сделать заключение, что между скоростью и импульсом существует следующая зависимость:

$$\mathbf{v} = -\frac{\mathbf{P}}{m}, \quad (7.21)$$

а между скоростью и энергией — зависимость

$$\mathcal{E} = \mathcal{E}_0 - \frac{mv^2}{2}. \quad (7.22)$$

Отсюда следует, что если бы в рассматриваемой части зоны Бриллюэна мы имели изолированный электрон, то его поведение было бы аномальным, как на это уже указывалось в § 3 в связи с исследованием периодического движения электрона. Например, если бы изолированный электрон находился у верхнего края зоны, то наложение электрического поля уменьшало бы его энергию, а не увеличивало ее. Однако мы имеем дело не с изолированным электроном, а с вакантным квантовым состоянием в верхней части заполненной зоны. Это соответствует, очевидно, условию минимума энергии для зоны в целом, так как все электроны находятся при этом в состояниях с энергиями, более низкими, чем энергия вакантного состояния. Если дырка приобретает скорость, то это осуществляется за счет перемещения электрона из более низкого состояния в более высокое. Мы должны теперь убедиться в том, что происходящие изменения энергии и тока происходят так, как это должно было бы иметь место для частицы с зарядом $+e$ и массой $+m$.

Прежде всего покажем, что поведение дырки, находящейся вначале в покое, соответствует поведению частицы с положительной, а не с отрицательной массой. Предположим поэтому, что вакантное квантовое состояние s соответствует наивысшему значению энергии в зоне валентных связей, так что $\mathcal{E} = \mathcal{E}_0$ и $\mathbf{v}_s = 0$. Под влиянием приложенного электрического поля \mathbf{E} изменение импульса происходит по закону $\dot{\mathbf{P}} = -e\mathbf{E}$. Следовательно, скорость, соответствующая

вакантному квантовому состоянию для частицы с отрицательной массой, будет меняться, согласно уравнению

$$\dot{\mathbf{v}}_s = -\frac{\dot{\mathbf{P}}}{m} = +\frac{e\mathbf{E}}{m}, \quad (7.23)$$

а ток будет меняться как

$$\dot{\mathbf{I}} = +\frac{e}{V}\dot{\mathbf{v}}_s = +\frac{e^2}{m}\mathbf{E}. \quad (7.24)$$

Из этого выражения следует, что ток возрастает в направлении приложенного электрического поля. Если в момент времени $t=0$ скорость $\mathbf{v}_s=0$, то выделяемая в образце мощность определялась бы выражением

$$\mathbf{E} \cdot \mathbf{I} = \frac{e^2}{m}\mathbf{E}^2 t, \quad (7.25)$$

так что при ускорении дырки энергия должна была бы поглощаться. Допустим, что мы хотим описать поведение дырки, предполагая при этом, что она обладает зарядом e_p и массой m_p . В этом случае мы должны были бы сделать заключение, что

$$\dot{\mathbf{v}} = \frac{e_p \mathbf{E}}{m_p} \quad (7.26)$$

и

$$\mathbf{E} \cdot \mathbf{I} = \frac{e_p^2}{m_p}\mathbf{E}^2 t. \quad (7.27)$$

Эти уравнения дадут тот же результат, что и уравнения (7.23) и (7.25), очевидно, только в том единственном случае, если значения e_p и m_p выбрать так, чтобы $e_p = e$, а $m_p = m$.

Выбор значения $m_p = -m$ исключается по той причине, что в этом случае перед правой частью уравнения (7.25) стоял бы знак минус, т. е., иными словами, частица с отрицательной массой должна была бы поставлять энергию полю, а не поглощать ее.

Дальше мы выведем более общие соотношения, учитывающие также влияние магнитных полей. Для этого случая

$$\dot{\mathbf{P}} = \mathbf{F} = -e\left(\mathbf{E} + \frac{\mathbf{v}_s \times \mathbf{H}}{c}\right) \quad (7.28)$$

и, как и раньше, $m\dot{\mathbf{v}}_s = -\dot{\mathbf{P}}$. Из этих уравнений следует, что

$$m\dot{\mathbf{v}}_s = +e\left(\mathbf{E} + \frac{\mathbf{v}_s \times \mathbf{H}}{c}\right). \quad (7.29)$$

Это уравнение есть уравнение $\mathbf{F} = m\mathbf{a}$ для частицы, обладающей положительным зарядом. Кроме того, мы показали, что плотность тока, как и в уравнении (7.19), определяется выражением

$$\mathbf{I} = \frac{e}{V}\mathbf{v}_s. \quad (7.30)$$

Из двух уравнений (7.29) и (7.30) можно для случая дырки, находящейся в зоне валентной связи, получить уравнение, связывающее $\dot{\mathbf{r}}$ с \mathbf{E} , \mathbf{H} и \mathbf{I} :

$$\dot{\mathbf{r}} = \frac{e^2}{m} \left[\frac{\mathbf{E}}{V} + \frac{\mathbf{I} \times \mathbf{H}}{ec} \right], \quad (7.31)$$

где e — абсолютное значение заряда электрона.

Рассматривая аналогичным методом случай электрона, находящегося на дне зоны проводимости, используя при этом вместо уравнения (7.29) уравнение (7.16) и заменяя $(+e/V)\mathbf{v}_s$ на $(-e/V)\mathbf{v}$, получим

$$\dot{\mathbf{r}} = \frac{e^2}{m} \left[\frac{\mathbf{E}}{V} - \frac{\mathbf{I} \times \mathbf{H}}{ec} \right], \quad (7.32)$$

где e — снова абсолютное значение заряда электрона (заряд электрона равен $-e$).

Очевидно, что эти уравнения отличаются друг от друга именно так, как этого и следовало ожидать при перемене знака перед e .

Предположение об отрицательном значении массы привело бы нас к недопустимому заключению о том, что при $\mathbf{H} = 0$ ток течет в направлении, противоположном направлению поля \mathbf{E} (см. задачу 7). Уравнения (7.31) и (7.32) записаны в форме, удобной для рассмотрения вопросов электропроводности, а также эффекта Холла, и применяются для этой цели в § 7 гл. 8.

Уравнение, выражающее закон сохранения энергии для случая дырки в зоне валентных связей, можно получить путем рассмотрения всех электронов, находящихся в зоне. Так как для каждого из них $d\mathcal{E}/dt = \mathbf{F} \cdot \mathbf{v}$, скорость изменения полной энергии должна быть равна энергии, доставляемой электрическим полем в единицу времени.

Нагляднее, однако, представить имеющие при этом место изменения энергии как связанные с „псевдокинетической энергией движения дырки“. К этому представлению мы приходим, рассматривая зависимость энергии электронов от энергии, отвечающей вакантному квантовому состоянию. Если скорость дырки изменяется от $\mathbf{v}_s = 0$ до $\mathbf{v}_s = \mathbf{v}_1$, то очевидно, что электроны при этом перераспределяются так, как если бы энергия одного электрона изменилась от

$$\mathcal{E}_0 - \frac{mv_1^2}{2} \quad \text{до} \quad \mathcal{E}_0, \quad (7.33)$$

т. е. энергия увеличилась бы на $mv_1^2/2$. Конечно, перераспределение осуществляется путем перемещения квантовых состояний¹⁾, в соот-

¹⁾ Из сказанного в § 2 очевидно, что о перемещении квантовых состояний следовало бы говорить независимо от того, заняты ли они электронами или нет. — *Прим. авт.*

ветствии с законом $\dot{\mathbf{P}} = \mathbf{F}$, так что электрон, вначале имевший скорость \mathbf{v}_1 , не обладает в результате скоростью $\mathbf{v} = 0$. Однако увеличение энергии всех электронов равно как раз $m\mathbf{v}_1^2/2$. Другими словами, энергия, приобретенная электронами, равна кинетической энергии дырки, обладающей массой $+m$ и движущейся со скоростью \mathbf{v}_1 .

Наконец, следует добавить, что плотность заряда в кристалле, отвечающая наличию одной дырки в зоне валентных связей, может быть вычислена с помощью электронных волновых функций для занятых состояний. Кроме того, из волновых функций можно построить волновой пакет для дырки. Оказывается, что этот волновой пакет обладает групповой скоростью \mathbf{v}_g , соответствующей скоростям вакантных квантовых состояний, для которых построен пакет (см. § 8 гл. 15). Таким образом, с волновыми пакетами для положительно заряженных дырок можно оперировать точно так же, как и с волновыми пакетами для электронов. Представлением о таких дырочных волновых пакетах и следует пользоваться для описания движения дырок в транзисторах, сделанных из германия *n*-типа, а также в опыте Хайнса — Шокли, описанном в § 1 гл. 3.

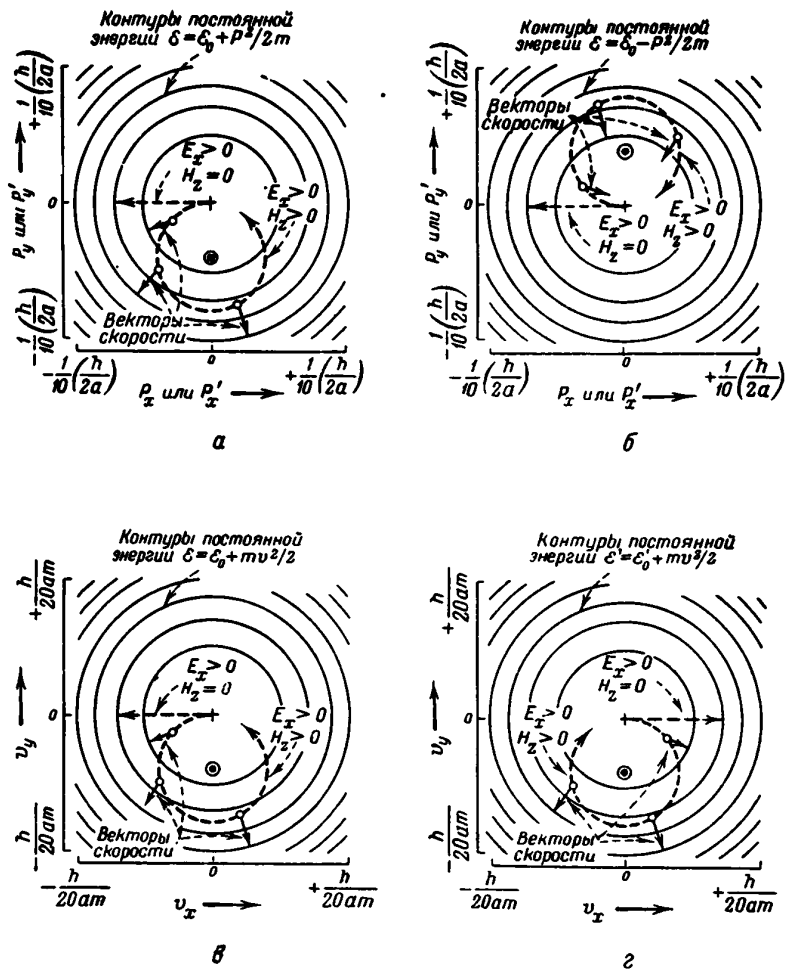
§ 7. ВЫВОДЫ

Исходя из выведенного на основе квантовой механики закона, что $\dot{\mathbf{P}} = \mathbf{F}$, и некоторых приближений, касающихся свойств поверхностей постоянной энергии в зоне Бриллюэна, было показано, что поведение единственного электрона в зоне проводимости и одной дырки в зоне валентных связей описывается классическими уравнениями; в первом случае для электрона с отрицательным, а во втором — с положительным зарядом.

Кроме того, было показано, что понятием дырки мы пользуемся главным образом для сокращенного описания поведения всех электронов.

Некоторые существенные выводы относительно движения электрона и дырки иллюстрирует фиг. 72. На фиг. 72, *a* представлена зона Бриллюэна, соответствующая полосе проводимости. Для большей наглядности специально выделена лишь небольшая область вблизи состояния с минимальной энергией. Скорость и ускорение избыточного электрона в присутствии магнитного поля будут в этой области такими же, как это показано на центральной части фиг. 70, *a* и *b*. Фиг. 72, *b* аналогичным образом изображает изменение квантового состояния с наивысшей энергией в зоне валентных связей; в этом случае скорость и ускорение дырки в магнитном поле будут такими, как это показано в углах фиг. 70, *a* и *b*.

При наличии только одного электрического поля E_x все изображающие точки в зоне перемещались бы влево, согласно уравнению $\dot{\mathbf{P}}_x = -eE_x$, а скорости электрона и дырки изменялись бы соответственно как $\mathbf{v} = \mathbf{P}/m$ и $\mathbf{v} = -\mathbf{P}/m$.



Фиг. 72. Поведение избыточного электрона или дырки при одновременном действии электрического и магнитного полей.

Пунктирные стрелки характеризуют изменения, происходящие с течением времени.

а — изменение квантового состояния электрона, находившегося вначале в состоянии с наименьшей энергией; б — изменение квантового состояния электрона, находившегося вначале в состоянии с наибольшей энергией; в — изменение скорости и энергии электрона в полосу проводимости; г — изменение скорости дырки в зоне валентных связей и изменение энергии ε' всех электронов в зоне.

На фиг. 72 эти четыре различных случая приведены для электрона и дырки, находившихся первоначально в состоянии с $\mathbf{v} = 0$. Энергия квантового состояния, занятого электроном, должна была бы в случаях, показанных на фиг. 72 *a* и *b*, увеличиваться с течением времени, а энергия состояния, занятого дыркой (т. е. вакантного состояния), — уменьшаться. Однако полная энергия всех электронов зоны валентных связей увеличивалась бы так, как показано на фиг. 72, *г*.

При наличии только одного магнитного поля вся совокупность квантовых состояний просто вращалась бы, подобно твердому телу, вокруг оси P_z с угловой скоростью, равной $\pm eH/mc$ (этот случай рассматривается в следующей главе). Направление вращения было бы при этом таким же, как и на фиг. 70.

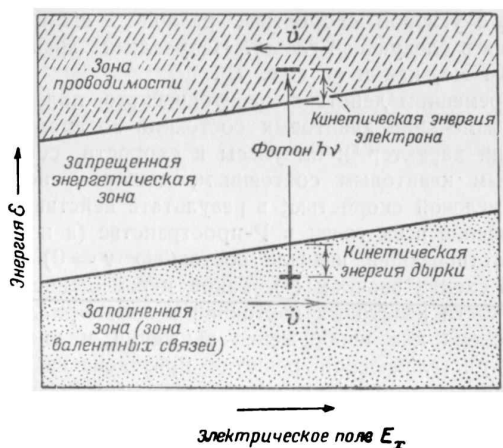
При одновременном действии электрического поля E_x и магнитного поля H_z изменение квантовых состояний во времени будет носить следующий характер¹⁾: импульсы и скорости, соответствующие рассматриваемым квантовым состояниям, движутся по окружностям с постоянной угловой скоростью; в результате действия обоих полей возникает вращение всех точек в P -пространстве (а не только одной точки, соответствовавшей вначале состоянию $\mathbf{v} = 0$) вокруг точки, отмеченной на схемах значком \odot . Такое перемещение, очевидно, не изменяет плотности квантовых состояний в зоне Бриллюэна.

Условная схема для изображения движений и энергий дырок и электронов в кристаллической решетке приведена на фиг. 73. Рассмотрим с помощью этой схемы образование пары электрон — дырка при поглощении фотона кристаллом (см. также гл. 1).

Фотон возбуждает электрон из состояния в нижней зоне в одно из состояний верхней зоны. Вообще говоря, ни одно из этих состояний не находится у самого края запрещенной зоны, так что и дырка и электрон обладают некоторой, отличной от нуля, скоростью и их кинетические энергии добавляются к энергии кристалла. Если положение возбуждающего фотона может быть точно локализовано (например, при освещении кристалла фокусированным пучком света), то волновые функции дырки и электрона представляют собой волновые функции, описывающие локализованные пакеты. Под влиянием электрического поля энергетические зоны приобретают некоторый наклон, так как электрон, находящийся на дне зоны проводимости, будет обладать более низкой энергией в той стороне кристалла, которая имеет более положительный потенциал. В отсутствие электрического

¹⁾ Этот интересный случай не имеет для нас существенного значения, так как формулы для эффекта Холла могут быть получены более простым путем. Решение при этом легко получить в комплексных обозначениях; полагая, что $w = x + iy$, находим решение уравнения $m\dot{w} = \pm e(E_x - iwH_z/c)$ в виде $w = -icE_x/H_z + \text{const} \exp(\pm ieHt/mc)$, так что состояние \odot соответствует $v_y = -cE_x/H_z$. Это выражение легко проинтегрировать и найти значение $w(t)$. — *Прим. авт.*

поля наиминимум энергетическим состоянием возбужденной пары является то состояние, при котором электрон находится на дне, а дырка — у верхнего края соответствующих зон. Общее правило, заключающееся в том, что дырки стремятся занять более высокие уровни на энергетической схеме, конечно, равносильно утверждению, что электроны стремятся занять самые нижние уровни. Действие электрического поля сводится к ускорению волновых пакетов в указанных на фиг. 73 направлениях. Мы вернемся еще к более обстоятельному рассмотрению таких схем энергетических уровней в гл. 12,



Фиг. 73. Изменение энергии электрона и дырки в зонах под влиянием электрического поля.

где они используются при анализе явлений выпрямления на $p-n$ -переходах и других аналогичных вопросов.

Для дальнейшего исследования электропроводности и эффекта Холла необходимо рассмотреть влияние столкновений на систему, содержащую большое число дырок и электронов. Это будет сделано в следующей главе.

§ 8. СВЯЗЬ МЕЖДУ КВАНТОВОЙ И КЛАССИЧЕСКОЙ МЕХАНИКОЙ

На основании материала, изложенного в этой главе, можно высказать предположение, что результаты, даваемые квантовой механикой, могут совпасть в предельном случае с результатами классической механики. Как уже указывалось, можно построить волновые пакеты, движущиеся с некоторыми определенными групповыми скоростями. Если электрон находится вне кристалла (так что периоди-

ческое силовое поле кристалла на него не действует), то квантовая механика дает точно такую же зависимость энергии от импульса, как и классическая теория электрона. Следовательно, групповая скорость, импульс и энергия связаны между собой в этом случае такой же зависимостью, как и в случае свободной частицы. Далее, импульс электрона в электрическом и магнитном полях изменяется под действием силы точно так же, как если бы электрон был классической частицей. В результате электронный волновой пакет будет вести себя совершенно так же, как свободный электрон.

Можно провести более подробный анализ вопроса в этом направлении. Если это сделать, то в результате можно прийти к выводу, что пока размеры области, в которой происходит движение электрона, очень велики по сравнению с длиной его волны, можно считать, что электронный волновой пакет движется, как свободная классическая частица. Однако это требование не является очень жестким. Например, достаточно, чтобы электронные волновые пакеты в самых маленьких вакуумных лампах были в тысячу раз меньше расстояний между проволочками сетки, чтобы с ними можно было обращаться, как с частицами.

В этом смысле классическая механика может рассматриваться как предельный случай квантовой механики, точно так же как геометрическая оптика есть предельный случай физической оптики, имеющей дело с явлениями дифракции света. В обоих случаях, если размеры рассматриваемых структур очень велики по сравнению с длиной волны, можно применять соответствующие предельные приближения. Это обстоятельство необходимо иметь в виду и при работе с длинными и сантиметровыми радиоволнами. При расчете антенны и эффектов малого масштаба явление дифракции имеет существенное значение. При изучении же эффектов большого масштаба, например отражения радиоволн от ионосферы, или для работы с радиолокационными установками, вполне достаточно геометрической оптики.

ЗАДАЧИ

1. Основываясь на соображениях симметрии (см. фиг. 72), можно убедиться в том, что под действием электрического поля $E_x = E_y$ электрон совершает периодические движения в направлениях x и y с одинаковым периодом, проходя один и тот же путь. Что можно сказать о движении электрона, если $E_x = 2E_y$?

2. С помощью метода, указанного в примечании на стр. 225, найти траектории дырок и электронов, обладающих начальными скоростями v_x и v_y в полях E_x и H_z .

3. Доказать, что если начальные скорости в момент времени $t = 0$ равны $v_x = v_0 \cos \theta$ и $v_y = v_0 \sin \theta$, то к моменту времени t смещение

электрона, усредненное по θ , не зависит от v_0 . Это показывает, что влиянием тепловых скоростей, рассматриваемым в § 6 гл. 8, можно пренебречь.

4. Для некоторых кристаллов (например для висмута и цинка) энергия электронов в существенных для нас участках зоны Бриллюэна определяется выражением

$$\mathcal{E} = \mathcal{E}_0 + K_x P_x^2 + K_y P_y^2 + K_z P_z^2.$$

Найти уравнение движения, которое в этом случае заменяет уравнение $m\dot{a} = F$, и проинтегрировать его, при условии, что сила F обусловлена наличием магнитного поля H_z .

5. Что произойдет с массой кристалла в целом при рекомбинации дырки с электроном (см. фиг. 73)?

6. Используя уравнения (7.31) и (7.32), показать, что мощность электрического поля $E (= \int IE dV)$ идет на изменение энергии электрона или дырки.

7. Показать, что частице с отрицательной массой, находящейся вначале в покое, так что при $t=0$ ток $I=0$, соответствует отрицательное сопротивление; при этом за время t выделяется энергия

$$\frac{e^2 E^2 t^2}{2|m|}.$$

Показать, что частица с положительной массой поглощает такое же количество энергии.

ЛИТЕРАТУРА

1. Zener C., Proc. Roy. Soc., 145A, 523 (1934).
2. Seitz F., Phys. Rev., 78, 549 (1948).
- 3*. Bardeen J., Shockley W., Phys. Rev., 80, 69 (1950).
- 4*. Агеев И. В., Природа химической связи в металлических сплавах, Изд. АН СССР, 1947.
- 5*. Wilson A. H., The Theory of Metals, Cambridge, 1936 (см. перевод: Вильсон А., Квантовая теория металлов, М. — Л., 1941).
- 6*. Конторова Т. А., ЖТФ, 18, 1478 (1948).
- 7*. Landsberg P. T., Proc. Phys. Soc., 65B, 397 (1952).
- 8*. Estermann J., Friedberg S. A., Phys. Rev., 85, 715 (A) (1952).

ВВЕДЕНИЕ В ТЕОРИЮ ЭЛЕКТРОПРОВОДНОСТИ И ЭФФЕКТ ХОЛЛА¹⁾

§ 1. ВВЕДЕНИЕ

Законы движения электронов и дырок, выведенные в гл. 7, применяются в настоящей главе к анализу электропроводности и эффекта Холла. Настоящее изложение основывается на упрощенном представлении о „среднем времени свободного пробега“²⁾. Такое упрощенное рассмотрение имеет целью иллюстрировать связь между экспериментальными результатами и отвлеченными понятиями, обсуждавшимися в гл. 5, 6 и 7. Это необходимо сделать прежде, чем перейти к теоретическим вопросам трех следующих глав, трактующих схемы энергетических уровней, статистику Ферми — Дирака и процессы столкновений. Кроме того, это упрощенное рассмотрение служит введением к последующим главам.

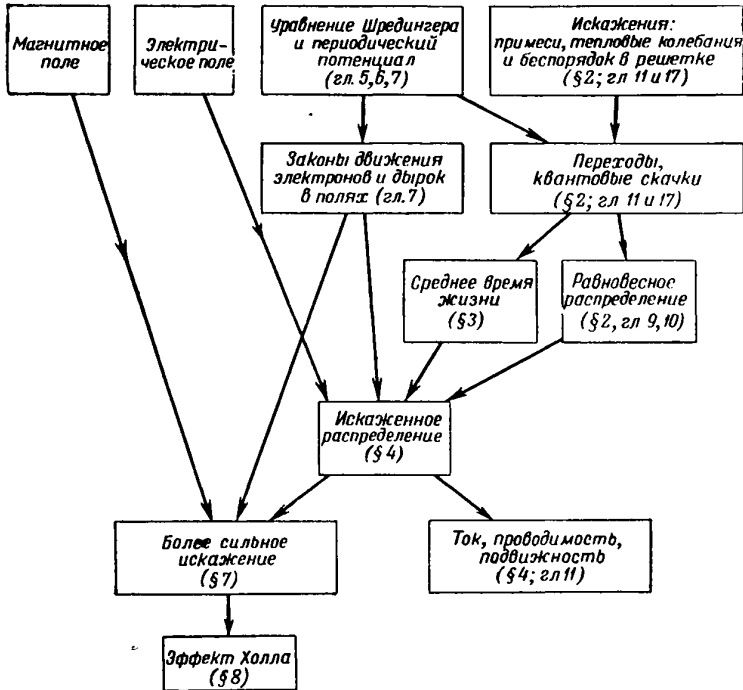
В настоящей главе, наряду с понятием о среднем времени свободного пробега, мы введем понятие о равновесном термодинамическом распределении электронов и дырок и о случайных процессах, приводящих к этому равновесию. Последние два понятия войдут, однако, в уравнения только через выражение среднего времени свободного пробега, обозначаемого через τ .

Существует два метода рассмотрения вопроса о плотности тока в полупроводнике. Хотя в большинстве задач они математически эквивалентны, связанные с ними представления несколько различны, и поэтому здесь описываются оба метода. Согласно первому методу, предполагается, что электроны и дырки находятся в определенных квантовых состояниях в зоне Бриллюэна и переносят токи ($\mp e/V$) \mathbf{v} в согласии с уравнениями (6.1) и (7.19). Этот метод очень удобен для вычисления токов с помощью схем, показывающих, как приложенные электрическое и магнитное поля нарушают распределение электронов в пространстве импульсов зоны Бриллюэна. Согласно другому методу, дырки и электроны рассматриваются как классические частицы, имеющие в каждый момент времени определенные положения и скорости. Такое рассмотрение соответствует теориям, изложенным в гл. 6 и 7 и показывающим, что волновые пакеты дырок и электронов движутся подобно классическим частицам, эффективные

¹⁾ См. обзорную статью [6] и монографии [7—11]. — *Прим. ред.*

²⁾ Более обстоятельное обсуждение вопроса о среднем времени свободного пробега имеется в гл. 11. — *Прим. авт.*

массы которых могут отличаться от массы свободного электрона. Согласно этому методу, электроны и дырки в электрическом и магнитном полях движутся между столкновениями по искривленным



Фиг. 74. Схема рассуждений и вычислений при выводе выражений для проводимости и эффекта Холла.

траекториям. (В гл. 11 используется более общий метод рассмотрения, применимый в том случае, когда понятие эффективной массы должно быть изменено.)

Однако имеются случаи, когда должен применяться метод волновых пакетов, так как он соответствует физической сути дела; в то же время метод индивидуальных состояний в зоне Бриллюэна использоваться быть не может. Последний метод применим к случаям, когда волновая функция дает равную вероятность нахождения электрона (или дырки) в любой элементарной ячейке кристалла. В опыте Хайнса и Шокли (см. § 1 гл. 3), однако, дырки вводятся в германий в определенном месте и непосредственно после этого имеют ничтожную вероятность находиться где-либо в других местах кристалла. Только по прошествии некоторого промежутка времени вероятность находже-

ния дырки у острия коллектора становится отличной от нуля. Для описания этого случая следует пользоваться методом волновых пакетов или частиц, и процессы введения и движения дырок по германию надо представлять себе так, как это описано в гл. 3. Математическим оправданием для таких представлений является то обстоятельство, что волновые пакеты ведут себя как классические частицы.

Фиг. 74 схематически показывает логическую связь между различными понятиями, обсуждаемыми в этой главе. Номера параграфов, приведенные на рисунке, указывают те места текста, где эти понятия обсуждаются. К рассмотрению фиг. 74 следует вернуться еще раз после прочтения текста § 2—7.

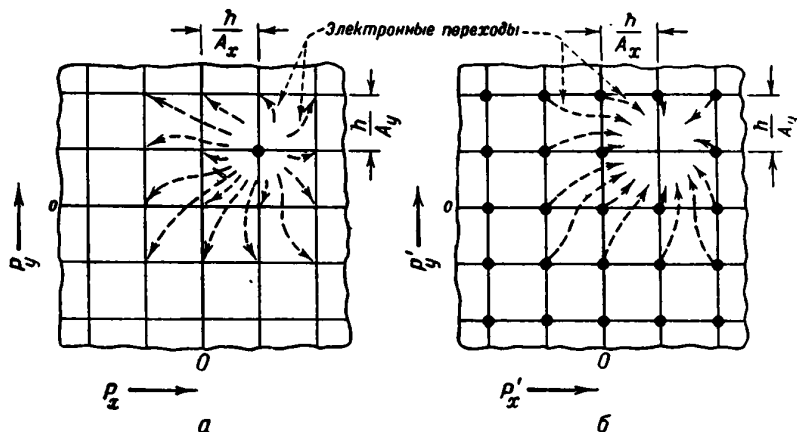
§ 2. СЛУЧАЙНЫЕ ПРОЦЕССЫ И ВЕРОЯТНОСТИ ПЕРЕХОДОВ

Как указывалось во введении к этой главе, теория среднего времени свободного пробега основывается на понятии случайных процессов. Само название „среднее время свободного пробега“, которое уже упоминалось несколько раз, но определения которого не давалось, указывает, что оно связано со временем, в течение которого электрон движется свободно, не испытывая случайных влияний. Рассмотрим сначала, в чем заключается сущность случайных процессов, а затем покажем, как они могут привести к равновесному распределению. В следующем параграфе мы определим среднее время свободного пробега и найдем его связь с этими случайными процессами.

Случайные процессы обусловлены несовершенством кристалла. Если бы кристалл был идеальный и силовое поле, в котором движется электрон, было строго периодическим, то электрон оставался бы в определенном квантовом состоянии неопределенно долго, а при действии на него электрического и магнитного полей менял бы свой импульс, согласно закону $\dot{\mathbf{P}} = \mathbf{F}$. Однако в реальном кристалле имеются нарушения правильности кристаллической решетки и силовое поле в нем не является строго периодическим. Как указано на фиг. 74, в таком кристалле имеются три основных вида нарушений правильности решетки. Тепловая энергия в кристалле проявляется в колебаниях атомов. При отсутствии тепловых колебаний нарушения правильности кристаллической решетки могут быть обусловлены наличием в ней примесных атомов (как доноров, так и акцепторов), рассмотренных в гл. 1, или наличием свободных мест, образовавшихся вследствие удаления атомов из их нормальных положений в правильной решетке. Эти удаленные атомы могут внедряться в другие места кристалла, где раньше нарушений структуры не было. Кроме того, могут иметь место дополнительные нарушения в виде „сдвигов“.

Описание движения электронов в неидеальном кристалле основывается на квантовомеханическом представлении о вероятностях переходов. Если кристалл неидеальный, электрон не может оставаться

неопределенно долго в одном и том же состоянии внутри данной зоны Бриллюэна: по прошествии некоторого промежутка времени он совершит квантовый переход в другое состояние в этой зоне. Этот процесс иллюстрирован фиг. 75, *a*, на которой показаны разрешенные состояния с энергиями, близкими к самой низкой энергии зоны Бриллюэна, когда одно из них занято „лишним“ электроном. Вследствие нарушения



Фиг. 75. Возможные переходы для электрона и дырки.

a — возможные переходы электрона в состоянии с малой энергией; *b* — возможные переходы дырки в состоянии с высокой энергией.

периодичности поля имеется вероятность перехода электрона из состояния, которое он занимает, в одно из других состояний. Эта вероятность перехода сложным образом зависит от значений \mathbf{P} в начальном и конечном состояниях, причем переходы электрона не ограничиваются такой небольшой группой состояний, как показано на фиг. 75, *a* ¹⁾. На фиг. 75, *b* изображена соответствующая картина для дырки, находящейся вблизи верха зоны валентных связей. Переход дырки происходит как следствие попадания в нее электрона (показано стрелками на фиг. 75, *b*). В этом случае на переходы оказывает влияние запрет Паули, так как электрон может совершать переходы только в состояния, еще не занятые другими электронами. Это значит, что для электронов возможны лишь те переходы, которые кончатся в дырке, и как бы такой переход ни происходил, результат эквивалентен переходу дырки. Хотя процессы, показанные на фиг. 75, *a* и *b*, весьма различны, в конечном итоге оказывается, что вероятности перехода для дырки

¹⁾ Переходы рассматриваются более обстоятельно в гл. 11. — *Прим. авт.*

и электрона приблизительно одинаковы. Это можно показать следующим образом.

На фиг. 75, *а* показан электрон, который может совершать переходы, скажем, в любое из *A* состояний. На фиг. 75, *б* изображено несколько электронов, каждый из которых может совершить переход только в одно состояние; однако в этом случае имеется *A* таких электронов. Кроме того, вероятность перехода в свободное состояние примерно одинакова для электрона, находящегося вблизи верхнего края зоны, и для электрона, находящегося вблизи дна зоны (см. гл. 11). Следовательно, вероятность того, что дырка (фиг. 75, *б*) совершит переход, соответствует вероятности перехода *A* электронов, из которых каждый имеет возможность совершить только один переход, тогда как вероятность того, что электрон (фиг. 75, *а*) совершит переход, соответствует вероятности перехода одного электрона в *A* возможных состояний. Таким образом, указанные вероятности переходов сравнимы, а это, как мы более детально покажем в дальнейшем, объясняет тот факт, что дырки и электроны имеют приблизительно одинаковую подвижность ¹⁾.

В настоящей главе предполагается, что только небольшая часть состояний в зоне проводимости занята электронами. Таким образом, мы можем пренебречь невозможностью перехода электрона вследствие того, что конечное состояние уже занято. Подобное предположение делается и для дырок. В математической теории без этих предположений легко обойтись, однако рассмотрение более общего случая в этой главе внесло бы ненужные осложнения.

Если переходы, совершаемые электроном, вызваны тепловыми колебаниями, то (согласно теории квантовых переходов) электрон будет получать или отдавать энергию при каждом таком переходе. Таким образом, он может обмениваться тепловой энергией с колеблющимися атомами кристалла и придти в состояние теплового равновесия. Конечно, бессмысленно говорить об одном электроне, находящемся в тепловом равновесии, так как тепловое равновесие есть статистическое понятие; в условиях теплового равновесия электроны имеют различные энергии. Однако если в зоне проводимости находится большое число электронов, то их равновесное распределение будет вполне определенным, как мы покажем в следующих двух главах. Следует отметить, что термодинамическое равновесное распределение электронов в кристалле, атомов в газе (или в любом аналогичном случае) не зависит от природы переходов, в силу которых оно наступает. Другими словами, независимо от природы переходов, поскольку они являются средством обмена энергий, система будет приходить в одно и то же конечное состояние

¹⁾ В гл. 1 мы показали аналогию между движением дырок и электронов. Здесь мы указали на сходство влияния на них случайных процессов. В гл. 10 мы покажем, что их статистические распределения в тепловом равновесии одинаковы. — *Прим. авт.*

теплового равновесия. Однако скорость, с которой она будет приближаться к равновесному состоянию, зависит от природы переходов, и это обстоятельство позволит нам рассмотреть понятие о среднем времени свободного пробега.

§ 3. СРЕДНЕЕ ВРЕМЯ СВОБОДНОГО ПРОБЕГА

Как указывалось выше, теория переходов или столкновений, изложенная в этой главе, основывается на некоторых упрощающих предположениях:

1. *Вероятность того, что электрон (или дырка) совершит переход в течение любого малого интервала времени dt , равна dt/τ , где τ — постоянная величина.*

Это предположение, в свою очередь, включает две независимые гипотезы: а) вероятность перехода не зависит от времени, протекшего после последнего перехода; б) вероятность перехода не зависит от того, какое квантовое состояние занято электроном или дыркой. Обе эти гипотезы вытекают из утверждения, что „параметр вероятности τ^{-1} “, имеющий размерность времени (секунды), есть постоянная величина, не зависящая от факторов, упомянутых в гипотезах, приведенных в п. „а“ или в п. „б“. Под „постоянной“ мы подразумеваем, конечно, постоянную при данных условиях, т. е. заданной температуре и химическом составе. Кроме того, значение τ неодинаково для электронов и дырок.

Как мы покажем ниже, τ есть среднее время свободного пробега, однако естественное определение среднего времени свободного пробега формулируется иначе, чем это сделано в предположении I. По этой причине мы не будем называть τ средним временем свободного пробега, пока с помощью уравнения (8.1) не определим среднее время свободного пробега, обозначенное через \bar{t} , и не докажем, что оно равно τ в уравнении (8.5).

II. *Конечное состояние при переходе не зависит от начального состояния, и вероятность перехода в какое-нибудь определенное конечное состояние пропорциональна вероятности того, будет ли это конечное состояние занято при термодинамическом равновесном распределении.*

Это предположение означает, что, независимо от того, как распределены электроны между квантовыми состояниями в момент $t=0$, коль скоро каждый из них совершит переход, их распределение будет соответствовать тепловому равновесию. (Другими словами, не существенно, каково было их распределение до последнего столкновения.)

На основании предположений I и II и уравнений гл. 7 выведем уравнения для проводимости и эффекта Холла. Эти уравнения соответствуют физической картине, которую мы сейчас грубо опишем только для того, чтобы показать важность этих предположений для дальнейшего рассмотрения. Согласно предположению I, электроны

ускоряются электрическим и магнитным полями приблизительно в течение времени τ ; затем они претерпевают столкновения. Если эти столкновения производят только небольшие изменения направления скорости частиц, то условия после столкновения зависят от действия полей на частицу до ее столкновения. Однако, согласно предположению II, столкновение уничтожает всякую связь между состояниями частицы до и после столкновения. Таким образом, движение частиц под действием полей — электрического и магнитного — будет, по существу, происходить так, как оно происходило бы τ секунд спустя после мгновенного приложения этих полей¹⁾.

Прежде чем дать определение среднего времени свободного пробега и доказать его совпадение с параметром вероятности τ , входящим в предположение I, попытаемся показать законность этого предположения, рассматривая процессы столкновений с помощью корпускулярной теории. Согласно этой теории, электрон движется как классическая частица по прямолинейному пути (или по криволинейному, если на него действуют электрическое и магнитное поля) в течение некоторого времени и затем внезапно отклоняется, после чего процесс повторяется снова. Эти внезапные отклонения аналогичны столкновениям между молекулами в газе и могут также называться столкновениями.

На языке корпускулярной теории предположение I формулируется следующим образом: *вероятность того, что электрон претерпит столкновение, не зависит от времени, прошедшего с момента его последнего столкновения*. Для движения молекул в газе этот вывод представляется вполне законным.

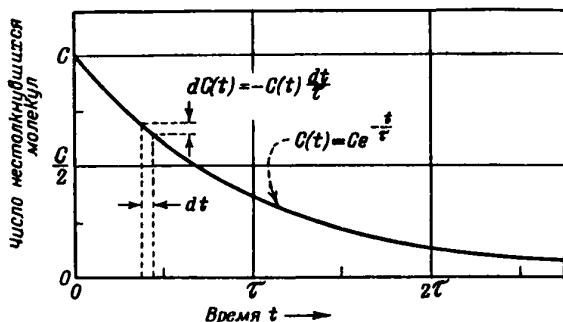
Рассмотрим произвольно выбранную молекулу и назовем ее молекулой a . Хотя молекула a и претерпела столкновение в некоторый предшествующий момент времени, область пространства, к которой она сейчас движется, содержит молекулы, на состояние которых это ее последнее столкновение не оказало влияния. Таким образом, величина проежутка времени, в течение которого молекула a двигалась после ее последнего столкновения, не влияет на молекулы, среди которых она движется. Вероятность того, что она претерпит столкновение в следующий бесконечно малый промежуток времени dt , не

¹⁾ Как мы покажем в гл. 11, ни одно из предположений I и II не находится в хорошем согласии с общепризнанными теориями. Время свободного пробега зависит существенным образом от скорости электрона (или дырки), и после перехода его энергия и направление скорости не будут, в противоположность предположению II, в среднем случайными. Однако, если время свободного пробега усреднено надлежащим образом с помощью методов, рассматриваемых в гл. 11 (с учетом зависимости времени свободного пробега от скорости и скорости после столкновения от скорости до столкновения), тогда полученное таким образом среднее значение τ может быть подставлено в выражение dt/τ , согласно предположению I, и использовано вместе с предположением II. В этом случае рассмотрение, приведенное в этой главе, дает правильные результаты, совпадающие с результатами, полученными посредством более сложных методов гл. 11. — *Прим. авт.*

должна зависеть от времени, в течение которого она двигалась свободно ¹⁾.

Квантовая теория переходов приводит к аналогичному выводу: вероятность квантового перехода равна dt/τ , независимо от того, как долго электрон находился в данном квантовом состоянии.

Теперь мы в состоянии дать определение понятия среднего времени свободного пробега и рассмотреть его подробнее. Сначала сделаем это для случая молекулы a в газе, чтобы иметь определенную физическую картину. Мы будем следить за движением молекулы a в течение длительного периода времени T , который начинается и кончается



Ф и г. 76. Зависимость числа молекул, не испытавших столкновения, от времени.

в моменты столкновения. В течение времени T молекула a испытает C столкновений (не считая исходного). Пусть интервалы времени между столкновениями равны t_1, t_2, \dots, t_C . Эти интервалы называются *временами свободных пробегов*. В течение этих интервалов молекула движется свободно, без воздействия случайных процессов. Среднее значение для таких интервалов представляет среднее время свободного пробега. Это приводит к следующему определению:

Среднее время свободного пробега \bar{t} есть (невзвешенное) среднее из времен свободных пробегов:

$$\bar{t} = \frac{t_1 + t_2 + t_3 + \dots + t_C}{C} = \frac{T}{C}. \quad (8.1)$$

(Слово „невзвешенное“ введено в определение, чтобы показать читателю, что может быть применен и другой способ усреднения. С таким случаем мы встретимся в § 7, где будет показано, что рассматриваемый там парадокс объясняется различием в способах усреднения.)

¹⁾ Аналогичная проблема встречается при бросании монеты. Независимо от того, сколько раз подряд монета упала лицевой стороной вверх, вероятность того, что она при следующем бросании снова упадет той же стороной, всегда равна $1/2$. — *Прим. авт.*

Теперь мы должны изучить распределение времен свободных пробегов от t_1 до t_C . Как мы увидим, их среднее значение равно τ ; однако в распределении встречаются все значения от нуля до значений, во много раз превышающих τ .

Рассуждения упрощаются, если представить себе не C последовательных времен свободного пробега для одной молекулы, а одно время свободного пробега у каждой из C молекул. Согласно предположению 1, любое время свободного пробега, независимо от того, чем оно определяется, будет с точки зрения статистики одинаково. Поэтому в обоих случаях получится одинаковое распределение времен свободных пробегов. Предположим, что случайно все C молекул одновременно претерпевают столкновения (попарно) с другими молекулами в момент времени $t=0$ ¹⁾. Проследим каждую из этих молекул до ее следующего столкновения, считая времена свободных пробегов равными $t_1, t_2 \dots$ до t_C ; мы можем не рассматривать дальнейшее движение молекул, так как мы выбрали для изучения только время их свободного пробега непосредственно после столкновения в момент $t=0$. Существенные статистические особенности этой картины могут быть получены при рассмотрении числа молекул $C(t)$, не испытавших столкновения за время t , после их столкновения в момент $t=0$. Выясним, как число молекул $C(t)$ зависит от времени. На фиг. 76 показано, как выводится эта зависимость. В момент времени t имеется $C(t)$ молекул, которые не испытали столкновения с момента $t=0$. Вероятность того, что в последующий промежуток времени dt каждая из них испытает столкновение, равна dt/τ (предположение 1). Следовательно, число молекул, сталкивающихся за время dt , будет равно $C(t) dt/\tau$ и число нестолкнувшихся молекул уменьшится на эту же величину. Математически это запишется в виде

$$dC(t) = -C(t) \frac{dt}{\tau}. \quad (8.2)$$

Выражение (8.2) представляет дифференциальное уравнение для $C(t)$, которое может быть проинтегрировано следующим образом:

$$\begin{aligned} -\frac{dt}{\tau} &= \frac{dC(t)}{C(t)} = d \ln C(t), \\ \ln C(t) &= -\frac{t}{\tau} + a, \\ C(t) &= e^a e^{-t/\tau} = C e^{-t/\tau}. \end{aligned} \quad (8.3)$$

Постоянная a определена так, что $C(t) = C$ при $t=0$. Из выражения (8.3) можно найти число свободных пробегов, заканчивающихся

¹⁾ Или, что совершенно эквивалентно, рассмотрим C столкновений, распределенных произвольно, и измерим для каждой молекулы время свободного пробега от момента столкновения. Или рассмотрим одну молекулу и C промежутков между ее столкновениями, измеряя продолжительность каждого из них от момента столкновения, с которого оно начинается. — *Прим. авт.*

в интервале времени dt в момент t . Как показано выше, это число равно $C(t)dt/\tau$, и, таким образом, число молекул, испытывающих столкновения в промежутке времени от t до $t+dt$, равно

$$Ce^{-t/\tau} \frac{dt}{\tau}. \quad (8.4)$$

Следовательно, между всеми рассмотренными C молекулами указанное число молекул имеет время свободного пробега, равное t , и поэтому вносит в сумму $t_1 + t_2 + \dots + t_C$ часть, равную $tCe^{-t/\tau} dt/\tau$. Суммирование по всем столкновениям приводит к среднему времени свободного пробега, определяемому выражением

$$\bar{t} = \frac{t_1 + t_2 + \dots + t_C}{C} = \frac{1}{C} \int_0^{\infty} tCe^{-t/\tau} \frac{dt}{\tau} = \tau; \quad (8.5)$$

(интеграл

$$\int_0^{\infty} te^{-t/\tau} \frac{dt}{\tau} = \tau \int_0^{\infty} e^{-t/\tau} \frac{t}{\tau} d \frac{t}{\tau} = \tau \int_0^{\infty} e^{-x} x dx = \tau \quad (8.6)$$

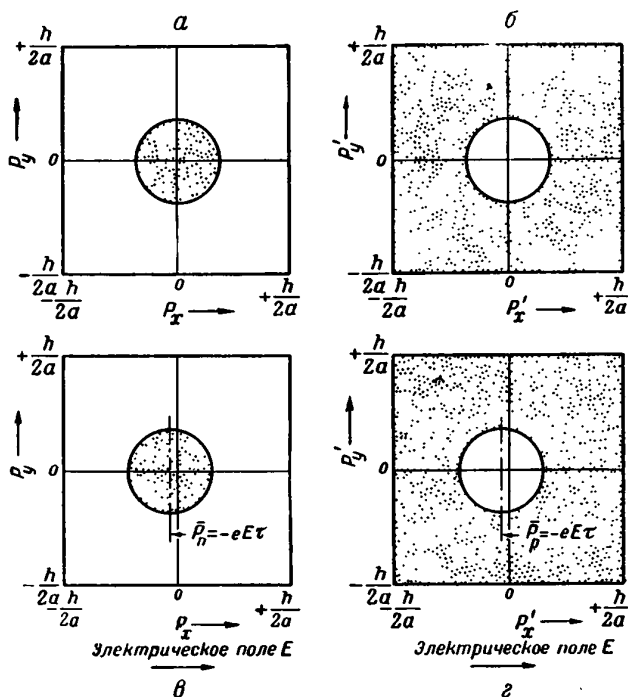
легко взять по частям или из таблиц).

В заключение коротко повторим ход рассуждений, которые привели к уравнению (8.5). Во-первых, было введено предположение I (для модели, подобной газу, было показано, что предположение I законно; параметр вероятности τ постоянен и не зависит от времени, протекшего после предыдущего столкновения). При этом τ еще не определялось как среднее время свободного пробега. Во-вторых, было дано определение среднего времени свободного пробега \bar{t} . В-третьих, было вычислено распределение времен между столкновениями, получающееся в результате предположения I, и из этого распределения было определено значение \bar{t} . Затем было найдено, что $\bar{t} = \tau$.

Эти же математические результаты могли быть получены из предположения I независимо от выбранной для рассмотрения модели. Мы могли с одинаковым успехом рассматривать электроны, описанные методом зон Бриллюэна или при помощи корпускулярного метода. Если бы мы начали с рассмотрения C столкнувшихся электронов в момент $t=0$, то в обоих случаях с помощью того же самого рассуждения нашли бы формулы (8.2)—(8.5). Аналогичные выводы применимы и к дыркам. До сих пор мы не использовали предположения II. Однако выводы, полученные в следующем параграфе, будут основаны на этом предположении. Как указывалось раньше, детальное рассмотрение процессов, приводящих к столкновениям, будет дано в одной из последующих глав.

§ 4. РАССМОТРЕНИЕ СРЕДНЕЙ СКОРОСТИ, ПОДВИЖНОСТИ И ПРОВОДИМОСТИ ПРИ ПОМОЩИ МЕТОДА ЗОН БРИЛЛЮЭНА

Мы должны теперь применить понятие среднего времени свободного пробега и распределения времен свободных пробегов, определяемого уравнением (8.3), к вычислению тока, вызванного электрическим полем. Имеется несколько методов рассмотрения этой задачи,



Ф и г. 77. Влияние электрического поля E_x на распределение электронов и дырок в зоне Бриллюэна.

a — электроны в зоне проводимости; $б$ — дырки в зоне валентных связей;
 $в$ — влияние поля E на электроны; $г$ — влияние поля E на дырки.

но так как все они включают представления, полезные для понимания процесса проводимости, то стоит сравнить их между собой.

Первым мы рассмотрим метод, который можно назвать *методом зон Бриллюэна*, или *методом распределения квази-импульсов*. Этот метод очень удобен для детального математического расчета в том случае, когда принимается во внимание зависимость τ от начального

состояния. Однако мы будем применять его здесь, считая, в согласии с предположением I, что τ — постоянная величина. В отсутствие электрического поля электроны занимают ряд квантовых состояний вблизи состояния с наименьшей энергией. Распределение электронов по состояниям будет детально описано в гл. 9 и 10.

Эта картина для электронов схематически показана на фиг. 77, а, а для дырок — на фиг. 77, б. В рассматриваемых здесь случаях мы полагаем, что число электронов или дырок очень мало. Тогда они занимают только небольшую часть состояний внутри заштрихованных областей и незначительное число их имеет энергии, соответствующие состояниям вне этих областей. В тепловом равновесии это распределение симметрично, так что вправо движется столько же электронов, сколько влево, и средние значения P и v равны нулю.

Сейчас рассмотрим случай, когда в кристалле находится n электронов. Объем V кристалла примем равным единице, так что n представляет концентрацию электронов (т. е. число их в единице объема). Нам будет удобно иметь дело с полным квази-импульсом n электронов, который определяется следующим образом:

$$P_{\text{полн.}} = P_1 + P_2 + \dots + P_n. \quad (8.7)$$

Средний импульс электрона, очевидно, равен $P_{\text{полн.}}/n$, и соответствующая ему средняя скорость \bar{v} равна

$$\bar{v} = \frac{v_1 + v_2 + \dots + v_n}{n} = \frac{P_{\text{полн.}}}{nm}. \quad (8.8)$$

Последнее следует из соотношения $P = mv$, справедливого для электрона, находящегося вблизи дна зоны проводимости, как отмечалось в § 5 гл. 7. Плотность тока, обусловленная каждым отдельным электроном (см. § 1 гл. 6), равна $(-e/V)v$, так что ток, переносимый n электронами, равен

$$I = \frac{-e}{V}(v_1 + v_2 + \dots + v_n) = -\frac{en}{V}\bar{v} = (-ne)\bar{v} \quad (\text{для } V=1). \quad (8.9)$$

Под действием электрического поля, составляющая которого по оси x есть E_x , все электроны изменяют свои импульсы, согласно уравнению $\dot{P} = -eE$. Это сдвинет распределение влево, как показано на фиг. 77, в. В то же время столкновения заставляют электроны возвращаться к хаотическому распределению, показанному на фиг. 77, а. Стационарное состояние будет достигнуто, когда эти два процесса уравновесят друг друга, в результате чего получится некоторый результирующий сдвиг распределения влево.

Уравнение, описывающее стационарное состояние, будет получено, если вычислить изменение $P_{\text{полн.}}$ за время dt , вызванное электрическим полем, и приравнять его изменению, вызванному столкновениями

электронов. Мы должны, конечно, рассматривать только составляющую по оси x . В таком случае изменение полного импульса, обусловленное электрическим полем, равно

$$dP_{x \text{ поля}} = -neE_x dt. \quad (8.10)$$

Согласно предположению I, в течение того же времени dt часть электронов dt/τ испытывает столкновения и, согласно предположению II, возвращается в беспорядочное состояние. Так как любой из электронов может испытать столкновение (предположение I), то изменение полного импульса будет просто пропорционально числу столкнувшихся электронов, т. е. значение $P_{x \text{ поля}}$ уменьшится на величину

$$dP_{x \text{ поля}} = -P_{x \text{ поля}} \frac{dt}{\tau}. \quad (8.11)$$

Приравнивая нулю сумму этих двух изменений, находим условие стационарного состояния

$$-P_{x \text{ поля}} \frac{dt}{\tau} - neE_x dt = 0; \quad P_{x \text{ поля}} = -neE_x \tau, \quad (8.12)$$

из которого легко получить для общего случая

$$P_{\text{поля}} = -neE\tau, \quad (8.13a)$$

$$\bar{P} = \frac{P_{\text{поля}}}{n} = -eE\tau, \quad (8.13b)$$

$$\bar{v} = -\frac{e\tau}{m} E, \quad (8.14a)$$

$$I = (-ne)\bar{v} = ne\frac{e\tau}{m} E. \quad (8.14b)$$

Эти уравнения определяют ток через постоянные τ и m , описывающие движение электрона в зоне проводимости, и соответствуют смещению распределения электронов в зоне Бриллюэна, показанному на фиг. 77, в.

Уравнения (8.14) могут быть выражены не только через τ и m , но и через *проводимость* и *подвижность*. По определению, подвижность есть отношение средней скорости носителя заряда к электрическому полю *без учета его знака*. Таким образом, имеем

$$\mu_n = \frac{e\tau_n}{m_n} \quad (8.15)$$

и

$$\bar{v}_n = -\mu_n E,$$

где индекс n показывает, что подвижность вычислена для отрицательных носителей, т. е. избыточных электронов. Проводимость, по определению, есть отношение величины I к E , т. е.

$$\sigma = \frac{ne^2\tau_n}{m_n} = ne\mu_n \quad (8.16)$$

и, таким образом, представляет полную плотность зарядов, умноженную на подвижность. Знак при этом не учитывается.

Совершенно аналогичные зависимости применяются и для дырок. Обозначим через p концентрацию, или плотность дырок и используем букву p также в качестве индекса для дырок. Изменение импульса дырки вычисляется, как и для электронов, из закона $\dot{P} = F$. Полный импульс $P_{\text{полн.}}$, равный сумме импульсов P , соответствующих незанятым состояниям или дыркам, будет изменяться под влиянием электрического поля на величину $-epE dt$ и вследствие столкновений на величину $-P_{\text{полн.}} dt/\tau_p$. Следовательно, стационарное состояние будет определяться формулой, аналогичной уравнению (8.12) для электронов. Для дырок, как можно легко убедиться, $P = -m_p v$ и $I = (+e/V)v$ (см. § 6 гл. 7), что приводит к

$$\mu_p = \frac{e\tau_p}{m_p}, \quad \bar{v}_p = +\mu_p E, \quad (8.17)$$

$$\sigma = pe\mu_p. \quad (8.18)$$

Если одновременно присутствуют дырки и электроны, что имеет место в области собственной проводимости, то полная плотность тока равна сумме дырочного и электронного токов и проводимость

$$\sigma = ne\mu_n + pe\mu_p, \quad (8.19)$$

где n — концентрация отрицательных носителей, а p — концентрация положительных носителей;

$$\begin{aligned} n &= \text{число электронов в } 1 \text{ см}^3, \\ p &= \text{число дырок в } 1 \text{ см}^3. \end{aligned} \quad (8.20)$$

Здесь читателю может быть полезно рассмотреть фиг. 74 и с ее помощью проследить за ходом рассуждения.

В настоящем рассмотрении применялись абсолютные электростатические единицы. Полезно перевести их в практические единицы и выразить в них некоторые типичные значения встречающихся величин¹⁾. Для обозначения практических единиц введем индекс L . Подвижность, выраженная в абсолютных электростатических единицах, имеет размерность $\text{см/сек на абс. эл.-ст. ед. напряжения/см}$. Таким образом, имеем

$$\mu = \frac{e\tau}{m} \text{ см}^2/\text{абс. эл.-ст. ед. напряжения} \cdot \text{сек}. \quad (8.21)$$

На практике напряженность электрического поля измеряется в в/см , и так как $1 \text{ абс. эл.-ст. ед. напряжения} = 300 \text{ в}$, то значение μ

¹⁾ Более полный перечень формул дается в § 8, формулы (8.58)—(8.67).—
Прим. авт.

§ 4. Рассмотрение средней скорости, подвижности и проводимости 243

в практических единицах равно

$$\begin{aligned} \mu_L \text{ см}^2/\text{в} \cdot \text{сек} &= \mu (\text{см}^2/\text{абс. эл.-ст. ед. напряжения} \cdot \text{сек}) (1 \text{ абс.} \\ &\text{эл.-ст. ед. напряжения}/300 \text{ в}) = (\mu/300) \text{ см}^2/\text{в} \cdot \text{сек} = \\ &= e\tau/300 \text{ т см}^2/\text{в} \cdot \text{сек}. \end{aligned} \quad (8.22)$$

Если $\mu_L = 10^8 \text{ см}^2/\text{в} \cdot \text{сек}$ или $\mu = 10^8 \cdot 300$, то среднее время свободного пробега

$$\tau = 10^8 \cdot 300 \frac{m}{e} = 3 \cdot 10^6 \cdot 9,1 \cdot 10^{-28} / 4,8 \cdot 10^{-10} = 5,7 \cdot 10^{-13} \text{ сек}. \quad (8.23)$$

Это время так мало, что релаксационные явления не должны наблюдаться даже в области миллиметровых волн.

Типичное значение проводимости для германия равно $1 \text{ ом}^{-1} \cdot \text{см}^{-1}$. Если мы выразим e в практических единицах, $e_L = 1,60 \cdot 10^{-19}$ кулона, то проводимость в тех же единицах будет

$$\sigma_L = ne_L \mu_L, \quad (8.24)$$

а соответствующая ей концентрация электронов

$$n = \frac{\sigma_L}{e_L \mu_L} = \frac{1}{1,6 \cdot 10^{-19} \cdot 10^8} = 6 \cdot 10^{15} \text{ см}^{-3}. \quad (8.25)$$

Следует напомнить, что проводимость $1 \text{ ом}^{-1} \cdot \text{см}^{-1}$ так велика, что можно пренебречь, во всяком случае до частот порядка микроволн, токами смещения в массе полупроводника (обусловленными меняющимся диэлектрическим смещением) по сравнению с токами проводимости (обусловленными движением дырок и электронов). Это обстоятельство не имеет места в выпрямляющем слое с большим сопротивлением. В абсолютных эл.-ст. ед. ток смещения равен

$$\frac{1}{4\pi} \chi \omega E = \frac{\chi f E}{2}, \quad (8.26)$$

где $\omega = 2\pi f$ — угловая частота и χ — диэлектрическая постоянная; для германия $\chi = 19$. В тех же единицах ток проводимости равен

$$\sigma E = 9 \cdot 10^{11} \sigma_L E. \quad (8.27)$$

Следовательно,

$$\frac{\text{Ток смещения}}{\text{Ток проводимости}} = \frac{\chi f}{2} \cdot 9 \cdot 10^{11} \sigma_L = 1,05 \cdot 10^{-11} f. \quad (8.28)$$

Отсюда следует, что ток смещения имеет ничтожное значение вплоть до частоты $f = 10^{10} \text{ сек.}^{-1}$, соответствующей длине волны в вакууме, равной 3 см. Тот же результат может быть получен в единицах MKS, в которых проводимость σ_m получается из σ_L следующим образом:

$$\sigma_L \text{ ом}^{-1} \cdot \text{см}^{-1} \cdot \frac{100 \text{ см}}{1 \text{ м}} = 100 \sigma_L \text{ ом}^{-1} \cdot \text{м}^{-1} = \sigma_m \text{ ом}^{-1} \cdot \text{м}^{-1}. \quad (8.29)$$

Отношение тока смещения к току проводимости в этой системе единиц равно

$$\frac{\partial D/\partial t}{\sigma_M E} = \frac{\chi \epsilon_0 \omega}{\sigma_M}. \quad (8.30)$$

При $\sigma_M = 100 \text{ ом}^{-1} \cdot \text{см}^{-1}$ или, что то же, $1 \text{ ом}^{-1} \cdot \text{см}^{-1}$, $\chi = 16$ и $\epsilon_0 = 8,85 \cdot 10^{-12}$; это приводит к уравнению (8.28). То значение ω , при котором отношение токов равно единице, равно обратной величине *времени релаксации*, в течение которого происходит выравнивание концентрации зарядов. Это можно видеть из вычисления скорости ее изменения в единицах MKS:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot I = -\nabla \cdot \sigma_M E = -\frac{\sigma_M}{\chi \epsilon_0} \nabla \cdot D = -\frac{\sigma_M}{\chi \epsilon_0} \rho. \quad (8.31)$$

Вычисление приводит к скорости уменьшения $\ln \rho$, равной $(\partial \rho / \partial t) / \rho = -\sigma_M / \chi \epsilon_0$, в согласии с полученным выше результатом.]

§ 5. РАССМОТРЕНИЕ СКОРОСТИ ДРЕЙФА, ПОДВИЖНОСТИ И ПРОВОДИМОСТИ НА ОСНОВЕ КОРПУСКУЛЯРНЫХ ПРЕДСТАВЛЕНИЙ

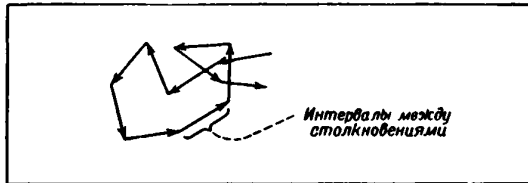
Соотношения, полученные в § 4, выводятся в этом параграфе с помощью теории волновых пакетов или классической корпускулярной теории. Как будет видно ниже, здесь необходимо проявить особую осторожность при усреднении продолжительности жизни во избежание ошибки, могущей исказить результат в 2 раза.

Мы опять будем рассматривать только случай электронов и затем укажем, как те же результаты могут быть распространены на случай дырок. Пусть в полупроводнике имеется n свободных электронов. Вероятность столкновения каждого электрона с нарушением решетки, согласно предположению I, равна dt/τ . (Это не относится к столкновению с другими электронами. Однако для многих задач теории полупроводников столкновением электрона с электроном можно пренебречь.) В согласии с предположением II, каждое столкновение электрона, выражаясь на языке статистики, приводит его в состояние, соответствующее тепловому равновесию, так что его направление и скорость движения не зависят от его направления и скорости до столкновения. Таким образом, электрон совершает беспорядочное движение и его траектория имеет вид, изображенный на фиг. 78, а. Каждый отрезок этой траектории представляет „свободный пробег“, а среднее из таких отрезков равно средней длине свободного пробега. Между средней длиной свободного пробега и средним временем свободного пробега существует определенная зависимость. Действительно, если средняя скорость теплового движения равна v_0 , то

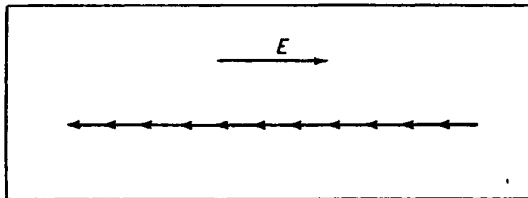
$$\text{Средняя длина свободного пробега} = v_0 \tau. \quad (8.32)$$

К вопросу о средней длине свободного пробега мы вернемся в гл. 11.

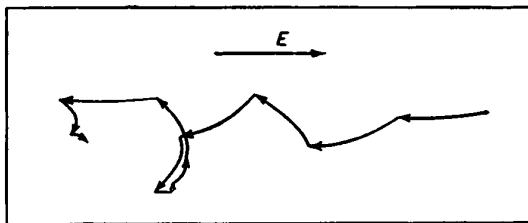
Под влиянием электрического поля траектории электронов искривляются и в действительности имеют форму отрезков параболы,



а



б



в

Ф и г. 78. Влияние электрического поля, вызывающего дрейф, налагающийся на беспорядочное движение. а — беспорядочное движение электрона в кристалле; б — смещения электрона, обусловленные электрическим полем; в — сочетание случаев „а“ и „б“: движение электрона в электрическом поле.

подобно траекториям свободно падающего тела. Среднее отклонение за время t_1 , вызванное ускорением ($a = -eE/m$) вследствие наличия поля, как и для свободно падающего тела, равно $\frac{1}{2}at_1^2$. Если бы время между двумя последовательными столкновениями равнялось t_1 , то смещение на каждом отрезке было бы равно $\frac{1}{2}at_1^2$, а после C столкновений общее смещение равнялось бы $(C/2)at_1^2$. Траектория электрона, показанная на фиг. 78, а, приняла бы вид, изображенный

на фиг. 78, в. Однако отрезки времени между столкновениями неодинаковы и распределяются статистически, согласно уравнению (8.4); таким образом полное смещение будет равно

$$\frac{a}{2} (t_1^2 + t_2^2 + \dots + t_0^2). \quad (8.33)$$

Согласно уравнению (8.4), число промежутков времени свободного пробега, значения которых лежат в интервале времени от t до $t + dt$, равно $C \exp(-t/\tau) dt/\tau$. Это добавляет к сумме (8.33) член

$$t^2 C e^{-t/\tau} \frac{dt}{\tau}. \quad (8.34)$$

Суммирование по всем возможным временам свободного пробега дает

$$\begin{aligned} t_1^2 + t_2^2 + \dots + t_0^2 &= \int_0^{\infty} t^2 C e^{-t/\tau} \frac{dt}{\tau} = C \tau^2 \int_0^{\infty} \left(\frac{t}{\tau}\right)^2 e^{-t/\tau} d\frac{t}{\tau} = \\ &= C \tau^2 \int_0^{\infty} x^2 e^{-x} dx = 2C \tau^2. \end{aligned} \quad (8.35)$$

Таким образом, среднее значение t^2 для каждого столкновения равно $2\tau^2$ и \bar{t}^2 не является просто квадратом \bar{t} , который равнялся бы τ^2 . Общеизвестно, что среднее значение квадрата некоторой величины больше квадрата ее среднего значения. Например, на линии с переменным напряжением, имеющим 60 периодов, средний квадрат напряжения равен $(110)^2$, в то время как среднее напряжение и его квадрат равны нулю. (Другие аналогичные примеры рассматриваются в § 2 гл. 14.) Таким образом, смещение, вызванное C столкновениями, будет равно

$$\frac{a}{2} 2C \tau^2 = Ca \tau^2. \quad (8.36)$$

Эти C столкновений потребуют в среднем время $C\tau$, как это следует из определения среднего времени свободного пробега, данного уравнением (8.1). Таким образом, средняя скорость в электрическом поле равна

$$v_d = \frac{aC\tau^2}{C\tau} = a\tau, \quad (8.37)$$

где v_d — средняя скорость, с которой частицы „дрейфуют“ в направлении поля.

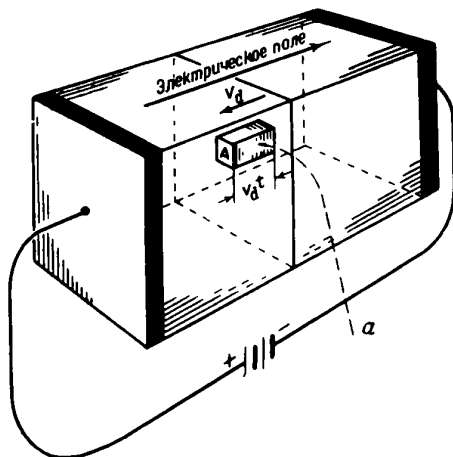
Если ускорение вызвано электрическим полем E , то его значение $a = -eE/m$ и

$$v_d = -\frac{e\tau}{m} E, \quad (8.38)$$

что эквивалентно выражению (8.14а), полученному для \bar{v} в § 4.

Пользуясь корпускулярными представлениями, можно получить аналогичное выражение и для тока, что будет сделано с помощью фиг. 79.

Рассмотрим число электронов, проходящих через площадку A за время t . Все электроны, находящиеся в объеме $A|v_d|t$, пересекут эту площадку, причем каждый перенесет заряд $(-e)$ справа налево или $(+e)$ слева направо. Наряду с этим существует некоторый поток электронов в рассматриваемый объем и из него вследствие



Фиг. 79. Соотношение между скоростью дрейфа и током.

a — объем, соответствующий электронам, проходящим через сечение A за время dt ; заряд $enAv_d dt$; плотность тока env_d .

наличия беспорядочного движения электронов. Однако в среднем этот поток равен нулю, так как одно беспорядочное движение электронов не дает результирующего тока. Таким образом, если концентрация электронов равна n , то результирующий заряд, перенесенный в направлении поля, будет $en|v_d|At$ и плотность тока в направлении E равна

$$I = \frac{en v_d A t}{A t} = en v_d = \frac{ne^2 \tau}{m} E. \quad (8.39)$$

(Плотность тока есть заряд, переносимый в единицу времени через единицу поперечного сечения.) Это выражение совпадает с уравнением (8.146). Аналогичные результаты, очевидно, могут быть получены и для дырок.

Здесь необходимо указать на трудность, возникающую при нестрогом применении понятия среднего времени свободного пробега к вычислению скорости дрейфа. Будем рассуждать следующим образом.

В среднем частица совершает свободный пробег за время τ . Предположим, что мы наблюдаем ее в какой-то произвольный момент времени. В среднем мы найдем частицу на середине отрезка пути между столкновениями, т. е. по прошествии времени $\tau/2$ после ее последнего столкновения. Таким образом, при случайном наблюдении скорость частицы окажется равной $a\tau/2$, а не $a\tau$.

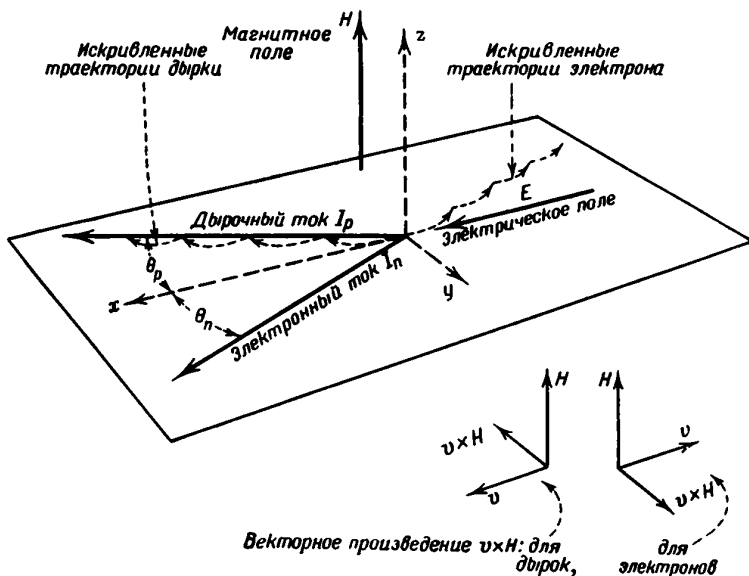
Такое рассуждение ошибочно. Если мы наблюдаем частицу в произвольный момент времени, то в среднем с момента ее последнего столкновения прошло время τ и до момента ее следующего столкновения пройдет такое же время τ , поэтому в среднем она будет иметь скорость дрейфа, равную $a\tau$, что соответствует уравнению (8.37). Однако может оказаться, что, согласно последнему правильному рассуждению, среднее время свободного пробега равно 2τ . Множитель 2 появляется в результате применения различных способов усреднения. В формуле (8.1), определяющей среднее время свободного пробега, каждому промежутку времени между столкновениями приписывается одинаковый вес. Однако отбор времен свободного пробега в произвольные моменты времени автоматически увеличивает вес длинных средних времен свободного пробега, и полученное таким образом среднее оказывается равным 2τ . Это определение статистического веса в математическом отношении похоже на вычисление \bar{v} , в результате которого также получается множитель 2.

§ 6. РАССМОТРЕНИЕ ЭФФЕКТА ХОЛЛА С ПОМОЩЬЮ КОРПУСКУЛЯРНЫХ ПРЕДСТАВЛЕНИЙ

Эффект Холла сыграл решающую роль в выяснении механизма проводимости в полупроводниках. Одновременные измерения эффекта Холла и проводимости в примесных полупроводниках позволяют определить как концентрацию, так и подвижность носителей тока. Измерение же одной электропроводности позволяет определить только произведение np_n или pr_p .

Эффект Холла, открытый в 1879 г., наблюдается в проводнике с током, помещенным в поперечное магнитное поле. При этих условиях наряду с имеющимся продольным электрическим полем в проводнике возникает поперечное электрическое поле, так что ток (который продолжает течь вдоль проводника) и электрическое поле не будут больше параллельны друг другу. В результате на образце появляется поперечная разность потенциалов. Причиной возникновения эффекта Холла является отклонение электронов магнитным полем. До развития зонной теории твердых тел оставалось неясным, почему электроны — носители тока — иногда отклонялись так, как должны были бы отклоняться положительные заряды. Однако, как будет показано ниже, для дырок надо было ожидать именно такого эффекта. Дадим сначала вводное, иллюстративное описание эффекта Холла, основанное на представлении о носителях заряда как частицах. Для

этой цели тепловыми скоростями частиц пренебрежем, считая, что они приходят в состояние покоя после каждого столкновения, так что движение частиц полностью определяется приложенными полями. Частицы рассматриваются нами как классические, без учета их волновых свойств. (Можно показать, что и здесь скорости беспорядочного движения в среднем компенсируют друг друга, как в случае,



Фиг. 80. Электронный и дырочный токи в электрическом и магнитном полях.

только что рассмотренном в § 5.) При таких предположениях движение положительных и отрицательных частиц в одновременно приложенных электрическом и магнитном полях изображается на фиг. 80. Положительная частица, начинающая движение от начала координат, ускоряется электрическим полем. По мере увеличения ее скорости на нее все сильнее действует отклоняющая ее сила со стороны магнитного поля, согласно векторной формуле

$$F = \frac{e}{c} \mathbf{v} \times \mathbf{H}. \quad (8.40)$$

Траектория частицы искривляется до тех пор, пока частица не испытывает столкновения, после чего процесс начинается сначала. В результате ток положительных носителей заряда отклонится от направления электрического поля на угол θ_p . Электрон отклоняется полем в ту же сторону, что и дырка. Это следует из того, что хотя

электрон приобретает скорость, направленную противоположно скорости дырки, но поскольку действие магнитного поля пропорционально произведению скорости на заряд (который у электрона имеет знак, противоположный знаку заряда дырки), его отклонение происходит в ту же сторону, что и у дырки. Электронный же ток отклонится от направления электрического поля в противоположном направлении на угол θ_p . На основе этой модели углы отклонения в абсолютных эл.-ст. ед. и эл.-магн. ед. определяются уравнениями

$$\theta_p = \frac{e\tau_p H}{mc} = \frac{\mu_p H}{c}; \quad \theta_n = \frac{e\tau_n H}{mc} = \frac{\mu_n H}{c}. \quad (8.41)$$

Мы не приводим здесь вывода этих уравнений, так как получим ту же формулу более простым способом в следующем параграфе. Эта формула знакома нам из теории магнетрона; eH/mc представляет собой угловую скорость электрона в магнитном поле.

Как будет показано ниже, из опытов по эффекту Холла могут быть найдены угол θ (или эквивалентная ему величина) и угол между направлениями электрического поля и тока. Эти опыты позволяют определить знак носителей и оценить подвижность μ на основании уравнения (8.41). Зная μ , из данных по проводимости можно найти концентрацию носителей.

В дальнейшем мы покажем, что и теория зон Бриллюэна также объясняет отклонение тока от электрического поля и приводит к той же формуле для угла θ . После того как мы определим постоянную Холла, будет описан экспериментальный метод ее измерения, и результаты будут выражены в практических единицах.

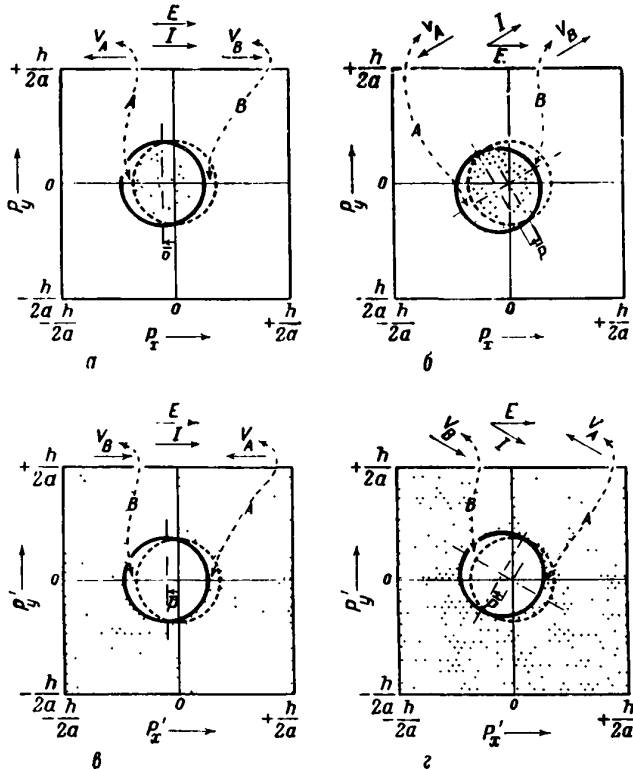
§ 7. РАССМОТРЕНИЕ ЭФФЕКТА ХОЛЛА С ПОМОЩЬЮ ТЕОРИИ ЗОН БРИЛЛЮЭНА. ПОДВИЖНОСТЬ, ОПРЕДЕЛЕННАЯ ИЗ ЭФФЕКТА ХОЛЛА, ПОДВИЖНОСТЬ, ОПРЕДЕЛЕННАЯ ПО СКОРОСТИ ДРЕЙФА, И МИКРОСКОПИЧЕСКАЯ ПОДВИЖНОСТЬ

Рассмотрим одновременное действие электрического и магнитного полей на распределение электронов и дырок по состояниям в зоне Бриллюэна. Это действие схематически иллюстрировано фиг. 81.

Фиг. 81, *а* и *б* соответствует полупроводнику *n*-типа с малым числом электронов в зоне Бриллюэна. В условиях равновесия, в отсутствие поля, эти электроны занимают состояния, распределенные вблизи центра зоны. В соответствии с фиг. 77, электроны изображены точками и постепенное уменьшение их концентрации по мере увеличения энергии занимаемых ими квантовых состояний не показано.

Мы представили распределение так, как будто все состояния до некоторого значения энергии равномерно заполнены, а все состояния выше этой энергии — свободны. Такое упрощение не вносит ошибки в качественное изложение, проводимое в этом и следующем параграфах.

(В дальнейшем при математическом изложении вопроса мы не будем пользоваться этой схемой.) На фиг. 81, *a* показан результат действия одного электрического поля. Предположим, что электрическое поле направлено вдоль положительной оси x ; тогда на электрон действует сила, направленная в сторону отрицательной оси x .



Фиг. 81. Влияние электрического поля E_x и магнитного поля H_z на распределение электронов и дырок в зоне Бриллюэна.

a и *в* — влияние электрического поля; *б* и *г* — влияние обоих полей.

В результате, как уже указывалось выше, происходит сдвиг распределения электронов в зоне Бриллюэна, так что новое стационарное состояние изобразится теперь заштрихованной площадью. Если все квантовые состояния внутри заштрихованной площади заполнены, а вне ее свободны, то этот сдвиг увеличивает число электронов, соответствующих области *A*, и уменьшает число электронов, содержащихся в области *B*. Однако поскольку распределение электронов

диффузно, получается увеличение числа электронов в квантовых состояниях вблизи области A и уменьшение их вблизи области B .

Как показывает рисунок, скорости электронов вблизи области A имеют составляющую, направленную влево, а скорости электронов вблизи области B — составляющую, направленную вправо, так что в результате сдвига происходит нарушение равновесного распределения скоростей, и электроны получают в среднем некоторый импульс P , направленный влево. Так как электроны имеют отрицательный заряд, то это эквивалентно току, текущему вправо, в направлении приложенного поля.

Если одновременно приложено магнитное поле, то все квантовые состояния перемещаются в соответствии с фиг. 70, б. Вследствие этого точки в области A повернутся против часовой стрелки и распределение будет иметь вид, изображенный на фиг. 81, б. Таким образом, вектор средней скорости не совпадает теперь с направлением электрического поля, а образует с ним некоторый угол, изображенный в верхней части фигуры. Из схемы легко понять, почему угол между направлениями тока и электрического поля определяется частотой магнетрона, умноженной на среднее время свободного пробега. Если электрон не испытывает столкновений во взаимно перпендикулярных электрическом и магнитном полях, то его скорость будет принимать последовательно все направления в течение времени, равного периоду, соответствующему частоте магнетрона. Это значит, что если бы электрическое поле было внезапно удалено и столкновения отсутствовали, то заштрихованный круг вращался бы вокруг центра рисунка с частотой магнетрона, как указывалось в связи с фиг. 72. При наличии столкновений такое вращательное движение прекращается по истечении среднего времени свободного пробега τ и эффективный угол, на который повернется распределение, как мы в этом убедимся из математического вывода, будет равен времени свободного пробега τ , умноженному на частоту магнетрона.

Формула для угла между направлениями тока и электрического поля легко может быть выведена из предположений I и II (см. § 3). Согласно этим предположениям, полный ток, имеющий в некоторый момент времени значение I , вследствие столкновений за время dt уменьшится на величину $I dt/\tau$, так как за это время часть электронов dt/τ возвращается в состояние теплового равновесия. В то же самое время изменяется ток, переносимый каждым отдельным электроном, так как электроны ускоряются электрическим и магнитным полями. Мы вычислим полное изменение тока, обусловленное действием этих полей, суммируя изменение токов, переносимых отдельными электронами. [В основном ход рассуждения тот же, что и при выводе уравнения (8.10).] Если мы умножим уравнение (7.34), выражающее \dot{I} через I , E и H , на dt , т. е. найдем изменение тока dI , и огнесем это изменение тока к электрону „№ 1“ из числа рассматриваемых нами n электронов, то для dI_1 получится следующее

выражение:

$$dI_1 = \frac{e^2}{mV} E dt - I_1 \times H \frac{e}{mc} dt. \quad (8.42)$$

Суммируя выражение (8.42) по всем n электронам и принимая объем V равным единице, находим для изменения тока I , обусловленного ускорением электронов в течение времени dt в электрическом и магнитном полях, выражение

$$dI = dI_1 + dI_2 + \dots + dI_n = \frac{ne^2}{m} E dt - I \times H \frac{e}{mc} dt. \quad (8.43)$$

(Следует отметить, что влияние беспорядочного движения не отражается в этом уравнении, так как в него входит только полный ток I .) Изменение тока, обусловленное столкновениями, равно

$$dI = -I \frac{dt}{\tau}. \quad (8.44)$$

В стационарном состоянии сумма этих изменений токов должна равняться нулю:

$$\frac{ne^2}{m} E dt - I \times H \frac{e}{mc} dt - I \frac{dt}{\tau} = 0,$$

или

$$I = \frac{ne^2\tau}{m} E - I \times H \frac{e\tau}{mc}. \quad (8.45)$$

Для упрощения алгебраических выкладок при решении этого векторного уравнения выберем систему координат так, чтобы ось x была параллельна направлению тока I , что удобнее, чем направить ее параллельно E , и вычислим направление и величину электрического поля E . При этих условиях векторное произведение $I \times H$ направлено параллельно оси y и имеет величину $-I_x H_z$. Магнитное поле H , как и раньше, параллельно оси z . Таким образом, составляющие тока [уравнение (8.45)] по осям x и y будут равны

$$I_x = \frac{ne^2\tau}{m} E_x = \sigma E_x, \quad (8.46a)$$

$$0 = \sigma E_y - (-I_x H_z) \frac{e\tau}{mc}, \quad (8.46b)$$

откуда

$$E_x = \frac{I_x}{(ne^2\tau/m)} = \frac{I_x}{\sigma}, \quad (8.47a)^1$$

$$E_y = -\left(\frac{I_x}{\sigma}\right) \left(\frac{e\tau H_z}{mc}\right). \quad (8.47b)$$

¹ Так как при выводе уравнения (8.47a) и уравнения (7.34) предположения I и II применялись без каких-либо приближений, то мы считаем, что рассеяние мощности $I_x E_x$, обязанное току I_x , не зависит от магнитного поля H . Отсюда следует, что выражение для изменения сопротивления в магнитном поле не может быть получено из этих предположений. Теория показывает, что это увеличение сопротивления связано с такими изменениями среднего времени свободного пробега, которые исключаются предположением I или же с отступлениями от сферической формы изонергетических поверхностей (см. гл. 12). — *Прим. авт.*

Отсюда видно, что вектор \mathbf{E} в плоскости xu повернут по часовой стрелке по отношению к вектору \mathbf{I} . При небольших углах

$$\theta = \frac{E_y}{E_x} = \frac{evH_z}{mc} = \frac{\mu H_z}{c}. \quad (8.48)$$

Выражение (8.48) совпадает с выражением для θ , полученным раньше из корпускулярных представлений.

Если вместо электронов рассматриваются дырки, то, согласно указаниям, приведенным в конце § 6 гл. 7, следует только изменить коэффициент у члена с H_z и, таким образом, изменить направление вращения. Конечно, при этом должны быть взяты значения τ и m , соответствующие рассматриваемым носителям тока. Однако диаграммы, показывающие смещение электронов в зоне Бриллюэна, имеют в случае дырок совершенно другой вид и изображены на фиг. 81, *в* и *г*. Как указывалось при рассмотрении фиг. 70 и 72, вращение электронов магнитным полем вблизи верха энергетической зоны имеет направление, противоположное направлению вращения их вблизи дна зоны. Этим объясняется противоположное направление вращения для электронов и дырок, показанное на фиг. 81, *б* и *г*.

До сих пор мы рассматривали преимущественно *угол Холла* θ , а не *постоянную Холла* R_H . Это объясняется тем, что постоянная Холла, которую мы рассмотрим в следующем параграфе, измеряет число носителей тока, тогда как угол θ , или вернее θ/H , измеряет их подвижность. Последняя величина значительно более характерна для полупроводников, так как число носителей в них, как и следует ожидать, резко зависит от содержания примесей, и потому чрезвычайно чувствительно к структурным изменениям полупроводника. Кроме того, величина θ/H легко измеряется. [Значение отношения θ/H не меняется в том случае, если поперечное сечение образца uz (фиг. 82) имеет неправильную или неизвестную форму.] Для обозначения подвижности, измеренной посредством эффекта Холла, мы введем термин *подвижность Холла* и символ

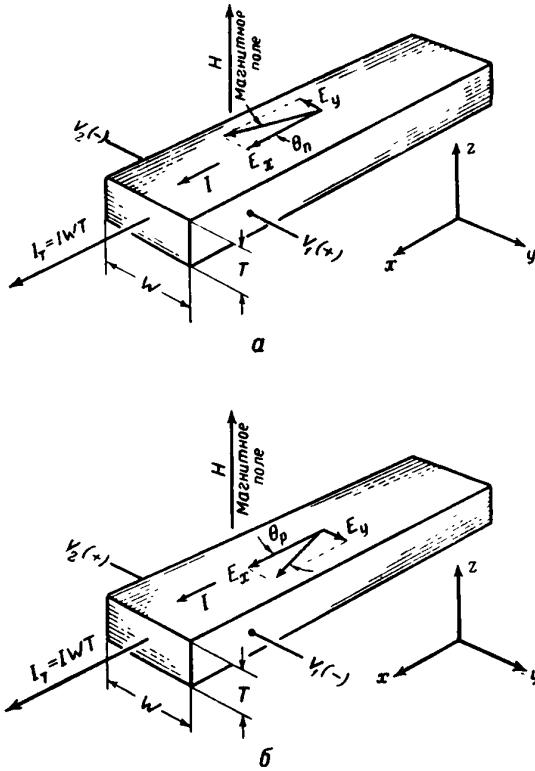
$$\mu_H = c \frac{\theta}{H}. \quad (8.49)$$

Подвижность Холла, вообще говоря, не равна подвижности, входящей в выражение для проводимости. Только недавно стало возможным определять с помощью метода, описанного в § 1 гл. 3, подвижность из измерений *скорости дрейфа*. Мы введем термин *подвижность дрейфа* и определим ее как отношение скорости дрейфа введенных носителей к электрическому полю:

$$\mu_D = \frac{\text{Средняя скорость дрейфа}}{\text{Электрическое поле}}. \quad (8.50)$$

Понятие подвижности дрейфа, полученное из измерений скорости распространения дырочного импульса в нитевидном образце германия,

может отличаться от понятия подвижности дрейфа, рассмотренного в предыдущих параграфах настоящей главы. Если в полупроводнике имеются уровни прилипания, так что дырки могут свободно двигаться только в течение небольшой части времени, то подвижность дрейфа



Ф и г. 82. Поперечное поле, обусловленное эффектом Холла.

а — составляющие электрического поля для электронной проводимости в магнитном поле; *б* — составляющие электрического поля для дырочной проводимости в магнитном поле.

будет составлять лишь часть *микроскопической подвижности*, которую мы определяем как подвижность незахватываемой частицы. Примером может служить дрейф окрашивающих центров в щелочно-галогидных кристаллах. Подвижности дрейфа *F*-центров в них действительно очень малы, порядка 10^{-4} *см*²/*в* · *сек*¹). Очевидно, эффект захвата в таких кристаллах уменьшает подвижность дрейфа, но

¹) См. обзор работ Поля и его сотрудников в книге Мотта и Герни [1], а также обзор Зейтца [2]. — *Прим. авт.*

не влияет на микроскопическую подвижность или подвижность Холла¹⁾). Как это будет показано в гл. 10, в германии при комнатной температуре захватом электронов донорами и акцепторами можно пренебречь. Однако в кремнии при комнатной температуре и в германии при низких температурах захваты должны оказывать заметное действие, приводя к различию между подвижностью дрейфа и подвижностью Холла. В гл. 12 приведен обзор данных по подвижностям.

Следует еще указать на связь микроскопической подвижности и подвижности дрейфа с проводимостью. Предположим, что мы имеем полупроводник *n*-типа с некоторой концентрацией избыточных электронов, которые непрерывно и часто захватываются и затем снова освобождаются. При этих условиях проводимость будет определяться концентрацией незахваченных электронов, так что мы можем написать

$$\begin{aligned} \sigma &= e \times (\text{Микроскопическая подвижность}) \times \\ &\quad \times (\text{Концентрация незахваченных избыточных электронов}) = \\ &= e\mu_D \times (\text{Полная концентрация избыточных электронов}). \end{aligned}$$

Таким образом, в выражении для проводимости может быть использована любая подвижность, при условии, что выбрана соответствующая ей концентрация носителей. Способ определения подвижности из данных о проводимости и полной концентрации доноров рассматривается в § 9 гл. 12.

В следующем параграфе мы опишем стандартный способ измерения эффекта Холла и переведем формулы из единиц абсолютной электромагнитной системы, применявшейся здесь, в другие системы. При этом различие в определении подвижностей не будет принято во внимание. В дальнейшем будет рассмотрено различие между микроскопической подвижностью, обозначенной через μ , и подвижностью Холла и будут приведены формулы, выведенные ниже в гл. 11, связывающие обе подвижности. Дальнейшее рассмотрение различия между ними проводится в § 9 гл. 12.

§ 8. ИЗМЕРЕНИЕ ЭФФЕКТА ХОЛЛА. ПРАКТИЧЕСКИЕ ЕДИНИЦЫ И ЕДИНИЦЫ MKS

Эффект Холла измеряется в образцах, подобных по форме изображенным на фиг. 82. Обычно ток пропускается вдоль образца, длина которого велика по сравнению с поперечными размерами²⁾.

¹⁾ Самые ранние известные автору опыты, в которых было сделано сравнение подвижности Холла и подвижности дрейфа, относятся к фотоэлектронам в хлористом серебре. Они были выполнены Хайнсом, использовавшим метод, описанный в работе [3]. В этих опытах с AgCl при комнатной температуре не было обнаружено захватов, уменьшающих подвижность. — *Прим. авт.*

²⁾ Об эффекте Холла в коротких образцах см. [4]. — *Прим. авт.*

Если к такому образцу приложено магнитное поле, оно отклоняет носители тока и тем самым заставляет ток отклоняться от направления электрического поля. В начальный момент времени, пока устанавливается магнитное поле, ток течет поперек образца и переносит заряды с одной стороны образца на другую. Эти заряды собираются на противоположных гранях, заряжая одну положительно, другую отрицательно, в результате чего создается поперечное электрическое поле. После того как установилось стационарное состояние, это поперечное поле уравнивает действие магнитного поля и на поверхностях не происходит дальнейшего накопления зарядов. При этом ток течет вдоль образца (параллельно оси x на фиг. 82). Электрическое поле теперь не параллельно оси x и имеет составляющую по оси y . Согласно выводам последнего параграфа, угол θ между вектором плотности тока I и вектором электрического поля E определяется уравнением (8.48). На фиг. 82, *a* показана относительная ориентация векторов E , I и H в образце n -типа, а на фиг. 82, *б* — в образце p -типа. Так как θ — малый угол¹⁾, то можно положить $\cos\theta = 1$, а синус и тангенс заменить углом θ . (Для краткости мы не будем писать индексы y и z при величинах I_y и H_z .) Как было показано раньше, $\theta = e\tau H/mc = \mu H/c$. Ввиду малой величины угла θ можно также написать, что плотность тока $I = \sigma E_x$. Таким образом, выражение для поперечного поля упрощается и мы имеем

$$E_y = \theta E_x = \frac{\theta I}{\sigma} = \left(\pm \frac{\mu}{c\sigma}\right) HI, \quad (8.51)$$

где знак плюс относится к дыркам, а знак минус — к электронам. По определению, постоянная Холла R_H есть коэффициент перед произведением HI в уравнении (8.51). Она дает отношение поперечного электрического поля к плотности продольного тока, умноженной на магнитное поле, и определяется уравнением

$$R_H = \frac{\theta}{\sigma H} = \pm \frac{\mu}{c\sigma} = \frac{1}{n(\pm e)c}. \quad (8.52)$$

Из этого уравнения видно, что постоянная Холла является непосредственно мерой числа имеющихся носителей тока, а также определяет их знак. [Если присутствуют одновременно и дырки и электроны, выражение для постоянной Холла будет более сложным и задается уравнением (8.89).]

Постоянная Холла обычно определяется из измерения поперечной разности потенциалов между двумя противоположными точками на образце и вызывающего ее поперечного магнитного поля. Кроме того, обычно измеряется полный ток, протекающий через образец, который мы обозначим через I_T , а не плотность тока I . Эти величины

¹⁾ По крайней мере, поскольку речь идет об эффектах первого порядка. — *Прим. авт.*

связаны с величинами, входящими в уравнение (8.51), следующим образом:

$$I_T = IW, \quad V_2 - V_1 = E_y W, \quad (8.53)$$

где W и T — поперечные размеры образца. Совокупность уравнений (8.51) — (8.53) приводит к соотношению

$$V_2 - V_1 = \frac{R_H H I_T}{T} \quad \text{или} \quad R_H = \frac{(V_2 - V_1) T}{H I_T}. \quad (8.54)$$

Эти уравнения выражены в системе CGS, так что токи и напряжения должны быть измерены в эл.-ст. ед. В остальном уравнение (8.54) ничем не отличается от обычно применяемого для определения R_H в практических единицах. Как уже говорилось в связи с рассмотрением направлений I , E и H , изображенных на фиг. 82, знак эффекта Холла показывает, вызвана ли проводимость электронами или дырками. Случай, когда имеются и электроны и дырки, рассмотрен в § 9.

Практические единицы. Выразим формулы для эффекта Холла в единицах, обычно применяемых в лаборатории:

$см$, $г$, сек. — для механических величин,
вольты (ϕ), кулоны — для электрических величин,
гауссы или эрстеды — для магнитного поля.

Величины, выраженные в единицах этой системы (лабораторной), отметим значком L . Выпишем выведенные выше формулы, куда входят величины, выраженные в единицах CGSE и CGSM:

$$R_H = \frac{\phi}{\sigma H} = \frac{1}{n(\pm e)c}, \quad (8.55)$$

$$c |R_H| \sigma = \mu = \frac{e\tau}{m}, \quad (8.56)$$

$$\tau = mc |R_H| \frac{\sigma}{e}. \quad (8.57)$$

Постоянная Холла в практических единицах, обозначенная через R_L , выражается в $см^3/кулон$, т. е. равна величине, обратной плотности заряда, измеряемой в $кулон/см^3$. Это следует из уравнения (8.55), если заряд электрона в нем выражен в кулонах. Обозначая через e_L заряд электрона в кулонах, имеем $e = ce_L/10$, в согласии с существующим соотношением между этими единицами. Это приводит к следующей зависимости между значениями постоянной Холла в практических единицах и единицах CGSE:

$$R_L = \frac{1}{n(\pm e_L)} = \frac{c}{n(\pm e)10} = \frac{c^2}{10} \cdot \frac{1}{n(\pm e)c} = 9 \cdot 10^{19} R_H. \quad (8.58)$$

Здесь, как обычно, скорость света приближенно принимается равной $c = 3 \cdot 10^{10}$ $см/сек$. Значение постоянной Холла определяется из прак-

тических измерений с помощью формулы (8.54) для R_H , в которой ток и напряжение следует выразить в практических единицах. Для этого воспользуемся известными соотношениями

$$V_2 - V_1 = \frac{(V_2 - V_1)_L}{300}, \quad I_T = 3 \cdot 10^9 I_L \quad (8.59)$$

и

$$R_H = \frac{(V_2 - V_1) T}{H I_T} = \frac{(V_2 - V_1)_L T}{9 \cdot 10^{11} I_L H}; \quad (8.60)$$

тогда, согласно (8.58),

$$R_L = 10^8 \frac{(V_2 - V_1)_L T}{H I_L}. \quad (8.61)$$

В уравнении (8.61) R_L выражено в $см^3/кулон$. С помощью соотношений $E_{yL} = (V_2 - V_1)_L/W$ и $E_{xL} = I_L/\sigma_L W T$ можно преобразовать уравнение (8.60) и получить выражение для угла между направлениями электрического поля и тока

$$\theta = \frac{E_{yL}}{E_{xL}} = \frac{(V_2 - V_1)_L T \sigma_L}{I_L} = 10^{-8} R_L \sigma_L H. \quad (8.62)$$

Найдем значение μ_L , подставляя в формулу для проводимости концентрацию зарядов и подвижность, выраженные в практических единицах:

$$\sigma_L = n e_L \mu_L = \frac{\mu_L}{|R_L|} \text{ ом}^{-1} \cdot \text{см}^{-1}; \quad e_L = 1,60 \cdot 10^{-19} \text{ кулон}, \quad (8.63)$$

откуда

$$\mu_L = \sigma_L |R_L| \text{ см}^2/\text{в} \cdot \text{сек} \quad (8.64)$$

и

$$\theta = 10^{-8} \mu_L H. \quad (8.65)$$

Из уравнений (8.64) и (8.56) легко получить выражение для среднего времени свободного пробега. Для этого уравнение (8.64) следует умножить на 300, чтобы перейти к единицам напряжения в системе CGSE. Таким образом, получаем

$$\tau = \frac{300 m}{e} \mu_L = 5,7 \cdot 10^{-16} \mu_L \text{ сек.} = 5,7 \cdot 10^{-16} R_L \sigma_L \text{ сек.} \quad (8.66)$$

При комнатной температуре тепловая скорость электрона будет порядка 10^7 см/сек (считая его эффективную массу равной массе свободного электрона). Следовательно, среднее расстояние, которое он проходит между двумя последовательными столкновениями, равно

$$l = 10^7 \tau = 0,57 \cdot 10^{-8} R_L \sigma_L \text{ см} = (0,57 \mu_L) A, \quad (8.67)$$

где $A = 1 \text{ \AA} = 10^{-8} \text{ см}$. Выражение (8.67) дает простую зависимость между подвижностью и средней длиной свободного пробега при комнатной температуре. (Средняя длина свободного пробега в кремнии рассматривается в гл. 11.)

Единицы MKS. В системе единиц MKS сила определяется выражением

$$F_M = e_M [E_M + v_M \times B_M], \quad (8.68)$$

где величины выражены в следующих единицах:

$$E_M \text{ в } \text{в}/\text{м}, \quad (8.69)$$

$$B_M \text{ в } \text{вебер}/\text{м}^2, \quad (8.70)$$

$$e_M = 1,60 \cdot 10^{-19} \text{ кулон}. \quad (8.71)$$

Поле в 1 вебер/см² равно 10⁴ гаусс. В этих единицах основные уравнения будут

$$\sigma_M = n_M e_M \mu_M \text{ ом}^{-1} \cdot \text{м}^{-1}, \quad (8.72)$$

$$n_M = \text{носителей}/\text{м}^3, \quad (8.73)$$

$$\mu_M = \text{м}^2/\text{в} \cdot \text{сек}, \quad (8.74)$$

$$R_M = \frac{1}{n_M e_M} = \text{м}^3/\text{кулон}, \quad (8.75)$$

$$\theta \text{ радиан} = \mu_M B_M. \quad (8.76)$$

В системе MKS постоянная Холла определяется следующим образом:

$$R_M = \frac{E_{yM}}{I_M B_M} = \frac{E_{xM} \theta}{E_{xM} \mu_M B_M} = \frac{\mu_M}{\sigma_M} = \frac{1}{n_M e_M}. \quad (8.77)$$

Уравнение (8.76) показывает, что подвижность имеет размерность $1/B_M = \text{м}^2/\text{вебер}$, или, согласно определению вебера, $\text{м}^2/\text{в} \cdot \text{сек}$. Выражения для τ и l , согласно (8.66) и (8.67), в этой системе единиц будут

$$\tau = 5,7 \cdot 10^{-12} \mu_M \text{ сек.}, \quad (8.78)$$

$$l = 5,7 \cdot 10^{-7} \mu_M \text{ м} = 5700 \mu_M \text{ \AA}. \quad (8.79)$$

§ 9. РАЗЛИЧНЫЕ ВИДЫ ФОРМУЛ ДЛЯ ЭФФЕКТА ХОЛЛА

Как показано ранее, предположения I и II (см. § 3) представляют только известное приближение. На самом деле, вероятность перехода зависит сложным образом как от начального, так и от конечного состояний электрона. Если принять во внимание эту зависимость, то в выражения для угла Холла и подвижности войдут другие значения среднего времени свободного пробега, которые мы назовем соответственно τ_H и τ_μ .

Таким образом, угол Холла

$$\theta = \frac{e \tau_H H}{m c}, \quad (8.80)$$

и подвижность

$$\mu = \frac{e \tau_\mu}{m}. \quad (8.81)$$

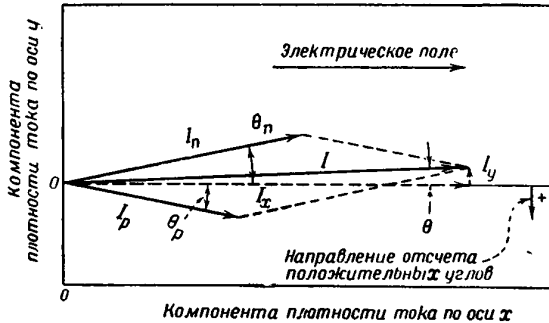
В случае электронов в полупроводниках при температурах настолько высоких, что преобладает рассеяние на тепловых колебаниях, зависимости между временами свободного пробега τ_H и τ_μ и между подвижностями μ_H и μ выражаются формулами, выведенными в § 4 гл. 11,

$$\tau_H = \frac{3\pi}{8} \tau_\mu \quad \text{и} \quad \mu_H = \frac{3\pi}{8} \mu. \quad (8.82)$$

С помощью этих соотношений формула для постоянной Холла (8.55) может быть преобразована к виду

$$R_H = \frac{\theta}{\sigma H} = \frac{\pm e \tau_H H}{m c} \frac{m}{n e^2 \tau_\mu H} = \frac{\tau_H}{\tau_\mu} \frac{1}{n (\pm e) c} = \frac{3\pi}{8} \frac{1}{n (\pm e) c}. \quad (8.83)$$

Этой формулой обычно пользуются при изучении эффекта Холла в полупроводниках. В простых металлах и в примесных полупроводниках с очень большой концентрацией электронов, к которым применима статистика для электронов в металлах, $\tau_H = \tau_\mu$ и формула



Фиг. 83. Эффект Холла, обусловленный дырками и электронами.

Магнитное поле в направлении +z (к читателю).

(8.83) принимает вид $R_H = 1/nec$. Приведенные здесь формулы выводятся в гл. 11.

При наличии в полупроводнике одновременно электронов и дырок постоянная Холла представляется в виде сложного усреднения из постоянных Холла и проводимостей для каждого типа носителей.

На фиг. 83 показано построение, применяемое при вычислениях постоянной Холла для этого случая. Магнитное поле, как и на фиг. 80, направлено перпендикулярно плоскости рисунка и параллельно оси z. При этих условиях дырочный ток I_p и электронный ток I_n отклоняются от направления электрического поля в противоположные стороны. Очевидно, результирующий ток I не должен

иметь составляющую, направленную поперек образца. Такая поперечная составляющая быстро образовала бы заряды на противоположных гранях образца и создала бы поперечное электрическое поле, которое уменьшило бы поперечный ток до нуля. Хотя суммарный поперечный ток равен нулю, поперечные электронный и дырочный токи в отдельности существуют и сносят электроны и дырки к одной стороне образца, как показано в § 2 гл. 3. В электрических полях настолько слабых, что можно пренебречь изменением проводимости, концентрации дырок и электронов будут иметь соответствующие равновесные значения и токи будут такими, как показано на фиг. 83. Таким образом, наша задача состоит в том, чтобы вычислить угол θ из построения, показанного на рисунке, и использовать его значение для определения постоянной Холла по уравнению (8.52): $R_H = \theta/\sigma H$.

Как принято, мы будем считать угол θ_n отрицательным. Если присутствуют оба типа носителей, то, согласно уравнению (8.19),

$$\sigma = e(n\mu_n + p\mu_p). \quad (8.84)$$

Используя приближение $\cos \theta_n = \cos \theta_p = 1$, применяемое для эффекта Холла, находим ток, текущий в направлении E :

$$I_x = I_n + I_p = (ne\mu_n + pe\mu_p) E_x. \quad (8.85)$$

Составляющая тока по y получается с помощью приближений $\sin \theta_n = \theta_n$ и $\sin \theta_p = \theta_p$ и соотношения $I_y/I_x = -\theta$ (см. фиг. 83), следующего из определения эффекта Холла. Таким образом, находим

$$I_y = (I_n\theta_n + I_p\theta_p). \quad (8.86)$$

Значения углов θ_n и θ_p определяются уравнениями (8.41):

$$\theta_n = -\frac{\mu_n H}{c}, \quad \theta_p = +\frac{\mu_p H}{c}. \quad (8.87)$$

Из этих уравнений находим, что угол Холла для полного тока равен

$$\begin{aligned} \theta &= -\frac{I_y}{I_x} = \frac{-(ne\mu_n E_x \mu_n H/c) + (pe\mu_p E_x \mu_p H/c)}{(ne\mu_n + pe\mu_p) E_x} = \\ &= \frac{(-n\mu_n^2 + p\mu_p^2)(H/c)}{n\mu_n + p\mu_p}. \end{aligned} \quad (8.88)$$

Подставляя (8.88) в формулу для R_H , получаем

$$\begin{aligned} R_H &= \frac{\theta}{\sigma H} = \frac{(-n\mu_n^2 + p\mu_p^2)(H/c)}{(n\mu_n + p\mu_p)(en\mu_n + ep\mu_p)H} = \\ &= \frac{-n\mu_n^2 + p\mu_p^2}{ec(n\mu_n + p\mu_p)^2} = \frac{-nb^2 + p}{(nb + p)^2 ec}, \end{aligned} \quad (8.89)$$

где, согласно данным Хайнса для подвижности дрейфа в германии (см. гл. 12),

$$b = \frac{\mu_n}{\mu_p} (= 2,1 \text{ для Ge и } 3,0 \text{ для Si}) \quad (8.90)$$

есть отношение подвижности электронов к подвижности дырок, n — концентрация электронов, p — концентрация дырок, e — заряд электрона и c — скорость света. Соответствующие значения R_H в практических единицах и единицах MKS будут

$$R_L = \frac{-nb^2 + p}{(nb + p)^2 e_L}; \quad R_M = \frac{-n_M b^2 + p_M}{(n_M b + p_M)^2 e_L}, \quad (8.91)$$

где $e_L = 1,6 \cdot 10^{-19}$ кулон. Измеряемое значение R_L определяется из уравнения (8.61). Легко убедиться, что если проводимость осуществляется только электронами (т. е. $p = 0$) или только дырками (т. е. $n = 0$), то формула (8.91) для R_L сводится к уравнению (8.58), или $R_L = 1/n(\pm e_L)$. Если допустить различие τ_H и τ_p , то (8.91) преобразуется к виду

$$R_L = -\frac{3\pi}{8} \frac{nb^2 - p}{(nb + p)^2 e_L}. \quad (8.92)$$

Этот вид формулы часто приводится в литературе. Она применяется, например, в работе Пирсона и Бардина [5].

В заключение укажем на одно простое следствие из формулы (8.91). Пусть образец p -типа, имеющий $n \ll p$ и, следовательно, $R_L > 0$, после прогрева превращается в образец с собственной проводимостью, т. е. $n \approx p$. Так как $b > 1$, то при температуре, при которой $n = p/b^2$, это приведет к перемене знака постоянной Холла. Этот момент перемены знака используется при определении b в задаче 7 [5].

ЗАДАЧИ

1. Вычислить средние значения t , t^2 , t^3 и t^4 для распределения заданного уравнением (8.4).

2. Получить общую формулу $\overline{t^n} = n! \tau^n$ путем рассмотрения интеграла $I(a)$ и его производных:

$$I(a) = \int_0^{\infty} e^{-at} dt = \frac{1}{a}; \quad \overline{t^n} = (-1)^n a \frac{d^n I(a)}{da^n}.$$

3. Найти часть длительного промежутка времени T [уравнение (8.1)], в течение которого совершаются пробеги с временами свободного пробега, лежащими между t и $t + dt$. Эта часть времени, деленная на время T , дает вероятность того, что наблюдение, сделанное в произвольный момент времени, выделит среднее время свободного пробега, лежащее в пределах между t и $t + dt$. Доказать, что среднее из выбранных таким образом средних времен свободного пробега равно 2τ (сравнить с рассуждением в конце § 5).

4. Показать, что если бы все времена свободного пробега имели совершенно одинаковое значение t' (что было бы нарушением предположения 1, см. § 3), то скорость дрейфа электронов была бы $v_d = -eEt'/2m$.

5. Энергетические зоны, соответствующие самым низким энергиям электронов в твердом теле, очень узки, так как волновые функции этих электронов не перекрываются; с другой стороны, зоны, соответствующие высокой энергии, — широки. Предполагая, что среднее время свободного пробега примерно одинаково у частиц, находящихся в зоне валентных связей и в зоне проводимости, решить, чья подвижность будет больше — электрона или дырки?

6. Предположите, что к полупроводнику, помещенному в постоянное электрическое поле E , внезапно прикладывается магнитное поле. Прежде чем ток достигнет значения, соответствующего новому стационарному состоянию, пройдет некоторый промежуток времени. Какой процесс будет определять продолжительность этого промежутка времени?

7. Отношение подвижностей b иногда определяется из выражения $nb^2 - p = 0$, что справедливо для образцов p -типа при температуре T_0 , при которой $R_H = 0$. При несколько более низких температурах все акцепторы ионизированы, так что концентрация дырок p имеет значение „насыщения“ p_s , равное избытку акцепторов над донорами, а концентрация электронов n пренебрежимо мала. По мере приближения к области собственной проводимости электроны и дырки образуются в одинаковом количестве вследствие разрыва валентных связей, так что $p = p_s + n$. Если проводимость при насыщении экстраполировать к T_0 , то для проводимости найдем значение σ_0 , в то время как в действительности проводимость равна σ_0 . Показать, что

$$\frac{\sigma_0}{\sigma_0} = \frac{b}{b-1}, \quad \text{или} \quad b = \frac{\sigma_0}{\sigma_0 - \sigma_0}.$$

8. Проверьте формулы § 8, выраженные в практических единицах, выводя их из формул, выраженных в единицах MKS.

9. Эта задача предназначена для иллюстрации некоторых основных черт влияния магнитного поля на сопротивление. Предположите, что электроны в кристалле могут быть разделены на две группы, в которых их концентрации соответственно равны n и n' , времена релаксации τ и τ' и массы одинаковы. Показать, исходя из уравнений (8.47а) и (8.47б), что углы Холла, соответствующие этим группам

электронов, равны:

$$\theta = \operatorname{tg}^{-1} \frac{e\tau H_z}{mc} = \operatorname{tg}^{-1} \omega\tau,$$

$$\theta' = \operatorname{tg}^{-1} \frac{e\tau' H_z}{mc} = \operatorname{tg}^{-1} \omega\tau'.$$

Произвести вычисление, подобное сделанному с помощью фиг. 83. Показать, что хотя в образце и отсутствует результирующий поперечный ток, в нем текут равные и противоположно направленные поперечные токи, обусловленные отдельными группами электронов. Показать, что если $n'\tau' = n\tau$, то сопротивление образца, определяемое как рассеяние мощности на единицу объема при плотности тока, равной единице, изменяется как

$$\frac{4 + 2(\omega\tau)^2 + 2(\omega\tau')^2}{4 + (\omega\tau + \omega\tau')^2}.$$

Показать, что если $\tau = \tau'$, то это отношение равно единице при всех значениях магнитного поля H_z . Однако если $\tau' = 10\tau$, то его значение изменяется от единицы до $202/121 = 1,67$ при изменении H_z от 0 до ∞ .

10. Проанализировать вопрос о боковом отклонении, испытываемом нитевидным образцом германия n -типа, по которому течет ток, в магнитном поле. Убедитесь в том, что результирующая сила, действующая на электроны со стороны поперечного электрического поля E_y , отнесенная к единице объема, равна и противоположно направлена силе, действующей на механически связанные с образцом доноры. Тогда электроны путем столкновений передают образцу некоторый импульс. Показать, что условие $\tau_\mu = \tau_H$ соответствует случаю, когда столкновения не передают образцу поперечного импульса. Более трудным является детальное рассмотрение задачи о передаче импульса электронами заполненной зоны, определяющей механизм сил, действующих на дырки. Однако она может быть решена методом, подобным изложенному в § 6 гл. 7. Эта задача также показывает, что применяемый иногда для получения значения угла Холла способ приравнивания сил eE_y и $e\tau_\mu H_z/c$ или $e\tau_\mu V_{M_x}$ не является строгим, так как при этом мы пренебрегаем передачей импульса путем столкновений.

ЛИТЕРАТУРА

1. Mott N. F., Gurney R. W., *Electronic Processes in Ionic Crystals*, Oxford, 1948 (см. перевод: Мотт Н., Гурни Р., *Электронные процессы в ионных кристаллах*, ИЛ, 1950).
2. Seitz F., *Rev. Mod. Phys.*, 18, 384 (1946).
3. Haunel J. R., Shockley W., *Report of a Conference on Strength of Solids*, Physical Society, London, 1948.

4. Iscenberg I., Russel B. R., Greene R. F., Rev. Sci. Inst., **19**, 685 (1948).
5. Pearson G. L., Bardeen J., Phys. Rev., **75**, 865 (1949).
- 6*. Давыдов Б. И., Шмушкевич И. М., Успехи физич. наук, **24**, 21 (1940).
- 7*. Дорфман Я. Г., Кикоин И. К., Физика металлов, М.—Л., 1936.
- 8*. Волькенштейн Ф. Ф., Электропроводность полупроводников, М.—Л., 1947.
- 9*. Justi W., Leitfähigkeit und Leitungsmechanismus fester Stoffe, Göttingen, 1948.
- 10*. Seitz F., The Modern Theory of Solids, New York, 1940 (см. перевод: Зейтц Ф., Современная теория твердого тела, М.—Л., 1949).
- 11*. Fowler R. H., Statistical Mechanics, Cambridge, 2 ed., 1936.

РАСПРЕДЕЛЕНИЕ КВАНТОВЫХ СОСТОЯНИЙ ПО ЭНЕРГИЯМ

§ 1. СХЕМА ЭНЕРГЕТИЧЕСКИХ УРОВНЕЙ ДЛЯ ЧИСТЫХ КРИСТАЛЛОВ

Для изучения равновесного распределения электронов по квантовым состояниям кристалла при разных температурах необходимо рассмотреть *распределение квантовых состояний по энергиям*. Эта величина, обозначаемая через $N(\mathcal{E})$, определяется следующим образом. Рассмотрим кристалл, объем которого равен единице. Предположим, что составлен полный перечень всех возможных квантовых состояний и их энергий для всех зон Бриллюэна рассматриваемого кристалла. Из этого перечня выберем и сосчитаем все квантовые состояния, энергии которых лежат в некотором определенном интервале в пределах от \mathcal{E} до $\mathcal{E} + d\mathcal{E}$; обозначим это число через dS . Тогда

$$N(\mathcal{E}) = \frac{dS}{d\mathcal{E}}, \quad \text{или} \quad dS = N(\mathcal{E}) d\mathcal{E}. \quad (9.1)$$

Определение $N(\mathcal{E})$: $N(\mathcal{E})$ есть число квантовых состояний, приходящихся на единицу энергии в единице объема кристалла.

Из теоремы сохранения квантовых состояний, рассмотренной в § 3 и 5 гл. 5, следует, что число квантовых состояний в зоне Бриллюэна пропорционально размерам кристалла, в то время как ширина энергетических зон не зависит от размеров кристалла (см. теорему о ширине энергетических зон в § 3 гл. 5). Следовательно, число квантовых состояний, приходящихся на единицу энергии в кристалле с объемом V , равно $VN(\mathcal{E})$.

Если значение энергии \mathcal{E} попадает в запрещенную зону, т. е. в промежуток между разрешенными энергетическими зонами (см. § 3 и 6 гл. 5), то $N(\mathcal{E}) = 0$. В том случае, когда значение \mathcal{E} лежит в энергетической зоне, значение $N(\mathcal{E})$ принципиально можно определить следующим образом. Предположим, что данной энергетической зоне соответствует одна или несколько зон Бриллюэна. В каждой из них значениям \mathcal{E} и $\mathcal{E} + d\mathcal{E}$ соответствуют две очень близкие поверхности постоянной энергии. В каждой отдельной зоне между этими поверхностями будет заключен некоторый объем dV_p Р-пространства. Объем, приходящийся на каждое возможное состояние в этом пространстве, равен h^3/V , где V — объем кристалла (см. § 2 и 5 гл. 5, § 7 гл. 14 и фиг. 59) и каждое такое состояние отвечает двум квантовым состояниям с разными направлениями спина. Следовательно,

число квантовых состояний в интервале энергии $d\mathcal{E}$ в единице объема кристалла равно

$$dS = \frac{2dV_{\mathbf{p}}/(\hbar^3/V)}{V} = \frac{2}{\hbar^3} dV_{\mathbf{p}}. \quad (9.2)$$

Если некоторые энергетические зоны перекрываются, то полное число состояний в интервале энергий от \mathcal{E} до $\mathcal{E} + d\mathcal{E}$ равно сумме состояний, принадлежащих отдельным зонам. В теории электропроводности целесообразно ввести в рассмотрение вместо элемента объема $dV_{\mathbf{p}}$ элемент площади изоэнергетических поверхностей, имеющий размерность P^2 и обозначаемый через $d\Omega$, и разность энергий $d\mathcal{E}$ между двумя соседними поверхностями. Выражение для $dV_{\mathbf{p}}$ при этом будет иметь вид

$$dV_{\mathbf{p}} = \frac{d\mathcal{E} d\Omega}{v}, \quad (9.3)$$

где $v = |\mathbf{v}| = |\nabla_{\mathbf{p}}\mathcal{E}(\mathbf{p})|$ — скорость, соответствующая рассматриваемой точке \mathbf{p} -пространства. Вывод формулы (9.3) можно найти в § 2 гл. 11 в связи с рассмотрением уравнения (11.5), в которое входит это выражение.

В качестве примера определим число квантовых состояний, приходящихся на единицу энергии вблизи дна энергетической зоны. Как уже указывалось в § 5 и 6 гл. 7, для этой цели можно воспользоваться приближением, согласно которому вблизи верхнего края или дна энергетической зоны энергия изменяется пропорционально P^2 или P'^2 , где $P = |\mathbf{P}|$ и $P' = |\mathbf{P}'|$. Для простоты предположим, что самый низкий уровень энергии расположен в центре зоны, так что в соответствии с уравнением (7.15)

$$\mathcal{E} = \mathcal{E}_0 + \frac{P^2}{2m}. \quad (9.4)$$

Дифференцируя, получаем

$$d\mathcal{E} = \frac{P}{m} dP. \quad (9.5)$$

Объем сферического слоя, заключенного между значениями P и $P + dP$ в \mathbf{P} -пространстве, равен $4\pi P^2 dP$, и, следовательно, в нем содержится

$$dS = \frac{2}{\hbar^3} 4\pi P^2 dP \quad (9.6)$$

квантовых состояний. Величину dS можно выразить через энергию. Определяя P из формулы (9.4) и PdP из формулы (9.5), получаем

$$P = [2m(\mathcal{E} - \mathcal{E}_0)]^{1/2} \quad (9.7a)$$

и

$$PdP = m d\mathcal{E}. \quad (9.7b)$$

После подстановки значений P и dP в выражение (9.6) находим, что

$$dS = \frac{8\pi}{h^3} P^2 dP = \frac{4\pi}{h^3} (2m)^{3/2} (\mathcal{E} - \mathcal{E}_0)^{1/2} d\mathcal{E} = N(\mathcal{E}) d\mathcal{E}. \quad (9.8)$$

Отсюда получаем искомое выражение для

$$N(\mathcal{E}) = \frac{4\pi}{h^3} (2m)^{3/2} (\mathcal{E} - \mathcal{E}_0)^{1/2}. \quad (9.9)$$

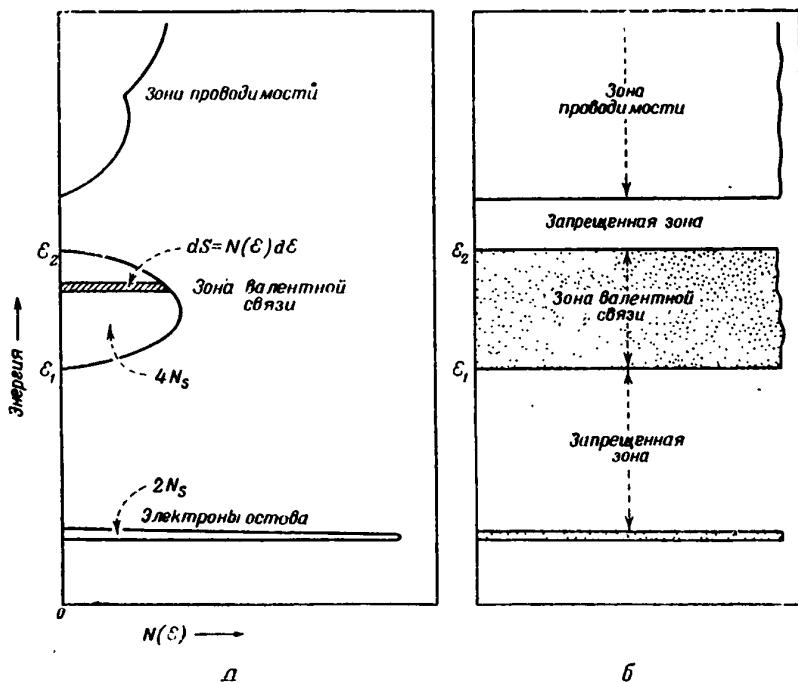
Это выражение определяет число квантовых состояний, приходящееся на единицу объема кристалла и единицу интервала значений энергии в том случае, когда значения энергии лежат лишь несколько выше минимальной энергии в зоне Бриллюэна. Полное число состояний, энергии которых ниже некоторого максимального значения \mathcal{E}_m , получаем, проинтегрировав выражение $N(\mathcal{E}) d\mathcal{E}$ в пределах от \mathcal{E}_0 до \mathcal{E}_m :

$$\int_{\mathcal{E}_0}^{\mathcal{E}_m} N(\mathcal{E}) d\mathcal{E} = \frac{8\pi}{3h^3} (2m)^{3/2} (\mathcal{E}_m - \mathcal{E}_0)^{3/2} = 4,5 \cdot 10^{21} (\mathcal{E}_m - \mathcal{E}_0)^{3/2}; \quad (9.10)$$

численный коэффициент в этом выражении равен числу состояний на 1 см^3 , если энергия выражена в электрон-вольтах. Эта формула часто применяется в теории металлов, где предполагается, что число состояний в энергетической зоне определяется формулой (9.9) и что все состояния с энергиями, лежащими ниже некоторого определенного значения, заняты. Формула (9.10) может быть использована в этом случае для вычисления энергии самого верхнего занятого электронами состояния. К совершенно аналогичным результатам мы приходим при рассмотрении распределения состояний вблизи верхнего края зоны. В этом случае применяется уравнение (7.20) или $\mathcal{E} = \mathcal{E}_0 - P^2/2m$, а уравнение (9.9) преобразуется путем замены $(\mathcal{E} - \mathcal{E}_0)^{1/2}$ через $(\mathcal{E}_0 - \mathcal{E})^{1/2}$. Необходимо упомянуть, что эти результаты основаны на тех же предположениях о простой форме поверхностей постоянной энергии, как и в § 5 гл. 7.

На фиг. 84 дана качественная картина ожидаемого на основе формулы (9.9) и некоторых общих соображений распределения квантовых состояний по энергиям в кристалле алмаза. Мы рассматриваем кристалл с объемом, равным единице, содержащий N_s атомов. На рисунке показаны 3 энергетические зоны. Самая нижняя из них соответствует атомному уровню $1s$ и занята двумя электронами, принадлежащими каждому атомному остатку (см. схему атома углерода на фиг. 3). Как уже указывалось в § 5 гл. 5, каждая элементарная ячейка в структуре алмаза содержит 2 атома, и, следовательно, для размещения четырех электронов атомных остатков этих атомов требуются 2 зоны Бриллюэна. Соответственно для размещения восьми валентных электронов, приходящихся на элементарную ячейку, необходимо иметь 4 зоны Бриллюэна, которые перекрываются, образуя

энергетическую зону энергий валентных связей. Если предположить, что формула (9.9) справедлива вблизи краев зоны [а формула, подобная (9.9), всегда справедлива, если заменить m через соответствующую эффективную массу], то распределение квантовых состояний по энергиям должно быть таким, как это показано на фиг. 84, а.



Фиг. 84. Распределение квантовых состояний по энергиям.

$N(\epsilon)$ — число квантовых состояний на единицу энергии в единице объема.

Направление отсчета энергии соответствует схеме энергетических зон, представленной на фиг. 84, б. Число состояний в зоне валентных связей будет

$$S = \int_{\epsilon_1}^{\epsilon_2} N(\epsilon) d\epsilon; \quad (9.11)$$

элемент интегрирования на рисунке изображен в виде заштрихованной полоски. Зона, соответствующая электронам атомного остатка, узка и, следовательно, характеризуется высоким значением $N(\epsilon)$.

Распределение квантовых состояний в полосе проводимости изображено как неограниченное со стороны больших значений энергии.

Это находится в соответствии с представлением о том, что более высокие атомные уровни взаимодействуют столь сильно, что возникающие из них зоны проводимости перекрывают все возможные более высокие значения энергии.

Фиг. 84, на которой изображена зона валентных связей и дно зоны проводимости, соответствует чистым кристаллам кремния и германия. Однако если в кристалле имеются примеси, эту схему следует изменить.

§ 2. УРОВНИ ЭНЕРГИИ, ОБУСЛОВЛЕННЫЕ ПРИМЕСЯМИ, — ДОНОРНЫЕ И АКЦЕПТОРНЫЕ УРОВНИ

Рассмотрим сначала случай донорной примеси, когда, например, атом фосфора введен в кремний. В гл. 1 уже указывалось, что атом, подобный атому мышьяка или фосфора, имеет на один положительный заряд больше, чем это необходимо для образования четырех валентных связей. Поэтому такой атом образует положительный заряд в решетке. Он вносит также в решетку избыточный электрон, нейтрализующий этот заряд, но сам не принимает участия в структуре валентных связей. Этому избыточному электрону отвечают волновые функции в зоне проводимости. Как мы видели, электрон, находящийся вблизи дна этой зоны, ведет себя в кристалле почти так же, как свободный электрон, за исключением того, что эффективная масса электрона в кристалле может отличаться от массы свободного электрона. В соответствии с этой картиной, можно было бы ожидать, что этот избыточный электрон будет вести себя в присутствии примесного иона так же, как вел бы себя свободный электрон в поле положительного заряда. Детальное рассмотрение этой задачи с точки зрения квантовой механики показывает, что эти представления в основном правильны¹⁾. Характер движения электрона в полосе проводимости в поле атома примеси весьма сходен с движением электрона в поле протона водородного атома. Однако необходимо указать на следующие существенные различия. В водородном атоме электрон притягивается зарядом протона. В случае примеси фосфора электрон притягивается зарядом той же величины. Но этот заряд находится теперь в диэлектрической среде. Вследствие поляризуемости решетки полупроводника сила притяжения электрона донорной примесью уменьшилась и равна теперь не e^2/r^2 , а $e^2/\chi r^2$, где χ — диэлектрическая постоянная, e — заряд электрона и r — расстояние между донором и электроном.

Второе существенное различие заключается в том, что эффективная масса электрона в кристалле не равна массе свободного электрона. Это обстоятельство имеет, вероятно, несколько менее

¹⁾ См., например, [1—5]. В работе Слэтера [5] интересующий нас вопрос рассматривается более подробно. — *Прим. авт.*

существенное значение. В результате волновые функции, описывающие движение электрона в поле донорной примеси, представляющие собой водородоподобные волновые функции, но для атома с гораздо меньшим эффективным зарядом ядра. Хорошо известное выражение для энергии связи электрона с ядром имеет вид

$$\frac{2\pi^2 m^* e^4 Z^2}{h^2} = 13,6 Z^2 \text{ эв}, \quad (9.12)$$

где e — заряд электрона, m^* — его эффективная масса, которую мы при вычислениях будем принимать равной массе свободного электрона, h — постоянная Планка и Z — заряд ядра. В полупроводнике $Z = 1/\kappa$. Так как диэлектрическая постоянная $\kappa = 11,9$ для кремния и $16,1$ для германия [6—9], то энергия связи для кремния составляет $0,10$ эв, а для германия $0,05$ эв. Радиус области, на которую простирается волновая функция, или радиус электронной орбиты, увеличивается при этом в κ раз и, таким образом, настолько велик, что захватывает большое число атомов, как это показано на фиг. 15.

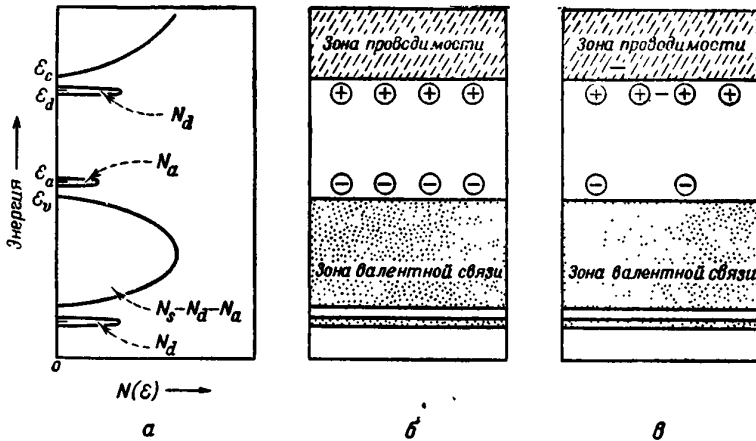
Формула для водородоподобной волновой функции имеет вид

$$\Psi = \frac{1}{\sqrt{\pi}} \left[\frac{1}{\kappa a_0} \right]^{3/2} e^{-r/\kappa a_0}, \quad a_0 = \frac{\hbar^2}{m e^2}, \quad (9.13)$$

где r — расстояние от иона и a_0 — радиус первой боровской орбиты, равный $0,529 \cdot 10^{-8}$ см = $0,529$ А. Это связанное состояние удалено в кремнии от дна зоны проводимости примерно на $0,10$ эв. В результате каждый атом донорной примеси создает квантовое состояние, энергия которого ниже энергии состояний в зоне проводимости. В большинстве практических случаев общее количество имеющейся в полупроводнике примеси очень мало по сравнению с общим числом атомов, так что лишь очень небольшая часть квантовых состояний в зоне проводимости переводится в состояния с более низкими значениями энергии. Атом водорода обладает, кроме нижнего энергетического состояния, целым рядом более высоких энергетических уровней; такие же уровни будут существовать также и в поле донорной примеси. Однако эти уровни будут расположены весьма близко к зоне проводимости и электроны, находящиеся на них, столь легко возбуждаются в зону проводимости, что нам нет необходимости учитывать существование этих уровней.

Атомы донорной примеси будут влиять подобным же образом и на состояния в зоне валентных связей, причем по одному состоянию на дне зоны валентных связей на каждый имеющийся донорный атом будет смещено на более низкий уровень. Однако состояния, расположенные вблизи дна этой зоны, не представляют для нас особого интереса, потому что при всех практически достижимых температурах при работе с полупроводниками эти состояния полностью заняты.

Акцепторные же примеси оказывают значительное влияние на состояние в зоне валентных связей. Эти примеси создают отрицательный заряд, так как заряда их атомных остатков недостаточно, чтобы нейтрализовать заряд четырех электронов, принадлежащих на их долю при образовании структуры валентных связей. Следовательно, они могут притягивать к себе положительный заряд. Как мы видели, дырка в зоне валентных связей обладает свойствами положительного заряда. Квантовая механика показывает, что дырка в зоне валентных



Фиг. 85. Схемы энергетических уровней.

a — распределение квантовых состояний с учетом влияния доноров и акцепторов для полупроводника *n*-типа; *б* — схема энергетических уровней для образца с равным числом доноров и акцепторов; *в* — то же для образца *n*-типа при низкой температуре. $N(\epsilon)$ — число квантовых состояний на единицу энергии в единицу объема.

связей, помимо того, что она ведет себя как положительный заряд при исследовании эффекта Холла и проводимости, будет двигаться в поле акцепторной примеси совершенно так же, как электрон движется в поле донорной примеси. В результате этого некоторые из состояний в зоне валентных связей при наличии акцепторных примесей займут более высокие уровни энергии.

На фиг. 85 показано влияние акцепторной и донорной примесей на распределение квантовых состояний. Мы рассматриваем образец вещества с объемом, равным единице, так что числа квантовых состояний в действительности представляют собой концентрации (число частиц на единицу объема). В зоне валентных связей содержится $4N_s$ квантовых состояний, где, как и раньше, N_s — число атомов в единице объема кристалла. Кроме того, имеется N_a донорных атомов и они переводят N_a состояний из зоны валентных связей и из зоны проводимости на более низкие уровни, как это описано

выше. Благодаря двум значениям спина число этих уровней в действительности будет равно $2N_d$. Однако мы должны принять во внимание то обстоятельство, что если электрон находится в одном из квантовых состояний в поле донора, то он нейтрализует этот атом, так что энергия связи другого захваченного электрона должна быть уже иной. Это означает, что в действительности имеется не $2N_d$, а только N_d состояний. (Это влияние спина на статистику распределения электронов по квантовым состояниям описывается вышеизложенным методом только приближенно. Но для наших целей нет необходимости в более точном рассмотрении этого вопроса; см. задачу 2 гл. 16.) Подобным же образом N_a акцепторных атомов удаляют N_a состояний из зоны валентных связей, оставляя в зоне всего $4N_s - N_d - N_a$ квантовых состояний. Как показано на рисунке, N_a состояний переведено из зоны валентных связей на более низкие уровни.

Схемы, подобные представленным на фиг. 85, б и в, часто применяются для описания и анализа тех случаев, когда существенно знать как распределение в пространстве, так и распределение по энергиям. Значки \oplus и \ominus на этих схемах обозначают энергетические уровни и заряд доноров и акцепторов. Заряд определяется для того случая, когда структура валентных связей завершена и избыточные электроны отсутствуют. В этих условиях каждый донор представляет собой избыточный положительный заряд. Значки, отвечающие донорам, расположены на уровне, соответствующем энергии связи донора с электроном и лежащем несколько ниже края зоны проводимости. Электрон в зоне проводимости и электрон, захваченный донором, изображены на фиг. 85, в. Подобно этому, дырка, обусловленная отсутствием электрона в зоне валентных связей, отвечал бы знак плюс. Так как дырки стремятся перейти на более высокие уровни, то дырка, захваченная акцептором, должна изображаться знаком плюс вблизи акцептора. Очевидно, что для электрической нейтральности системы необходимо, чтобы суммарное значение заряда равнялось нулю. Схемы такого рода используются в гл. 4 и 12.

При высоких значениях концентраций примесей возникают новые эффекты, связанные с явно выраженным перекрытием волновых функций атомов примесей. Это весьма напоминает случай, рассмотренный в § 3 гл. 5, где речь шла об образовании твердого тела при сближении атомов. Как было показано на фиг. 53, при достаточном сближении атомов происходит заметное расщепление атомных энергетических уровней и оказывается более правильным описывать электроны волновыми функциями, соответствующими состояниям в зоне Бриллюэна. Аналогичным образом будут изменяться и волновые функции электронов, движущихся в поле примесей. С увеличением концентрации примеси волновые функции перекрываются все больше и больше и, наконец, образуют примесную энергетическую зону. Электроны в этих состояниях будут проводить ток так же, как электроны

в металле. Действительно, ситуация, с которой мы имеем здесь дело, весьма напоминает случай щелочного металла, который можно рассматривать как решетку из небольших положительных ионов, нейтрализующих пространственный заряд свободных электронов. В полупроводнике ионы доноров (или акцепторов) расположены беспорядочным образом, нейтрализуя пространственный заряд электронов (или дырок), которые движутся, однако, не свободно, так как они занимают определенные состояния в зоне Бриллюэна. Данные, подтверждающие справедливость такого рода представлений, были получены при исследовании кремния и германия *n*- и *p*-типа. Было показано, что образцы с большим содержанием примеси ведут себя подобно металлам, причем при температурах ниже области собственной проводимости число носителей заряда не зависит от температуры. Примеры кривых, характеризующих такую температурную зависимость, представлены на фиг. 11 и 12 и подробно обсуждаются в гл. 11. Из анализа температурной зависимости эффекта Холла и электропроводности определены численные значения ряда энергетических параметров, рассматриваемых в этом параграфе. Они приведены в табл. 5 вместе с поясняющими замечаниями и соответствующими литературными ссылками. Данные для германия предварительные.

Таблица 5

ЭНЕРГИИ (в эв)

Разность энергий		Кремний	Германий
$\mathcal{E}_c - \mathcal{E}_n$	Эксперимент	$1,12 - 3 \cdot 10^{-4} T^a$	$7,2^b$
$\mathcal{E}_c - \mathcal{E}_d$	Теория	$0,08^a - 0,054^2$	$0,05^a$
$\mathcal{E}_a - \mathcal{E}_n$	"	$0,08^a$	$0,05^a$
$\mathcal{E}_c - \mathcal{E}_d$	Эксперимент	$0,06^d$	$0,013^e$
$\mathcal{E}_a - \mathcal{E}_n$	"	$0,075^d$	$0,013^e$
$\mathcal{E}_a - \mathcal{E}_n$	Центры, возникающие при бомбардировке		$0,046 \pm 0,01^{ж}$
$\mathcal{E}_a - \mathcal{E}_n$	Центры, возникающие при нагревании		$0,04^{ж}$

^a [10]. Это значение получено при анализе температурной зависимости *pn*. См. § 3 гл. 16.

^b Неопубликованная оценка, сделанная Бардином.

^c Вычислено на основании уравнения (9,12); m — масса свободного электрона, $x = 13$ для Si, $x = 16$ для Ge.

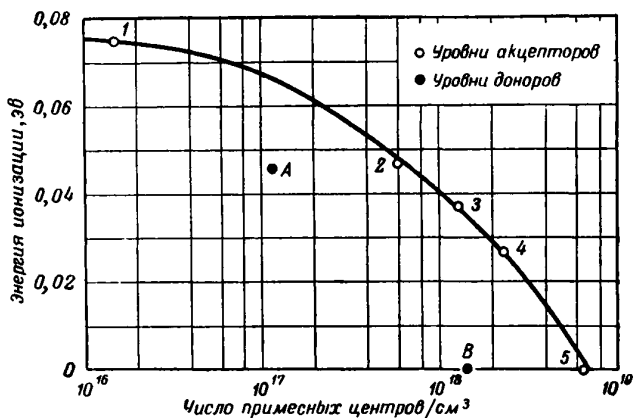
² См. ссылку ^a; $m_n = 0,67 m$.

^d См. ссылку ^a. См. также фиг. 86 и текст.

^e Предварительный анализ ранних данных (см. [11]).

^ж Предварительные результаты, полученные Пирсоном.

Пирсон и Бардин показали, что потенциалы ионизации примесных центров в кремнии $\mathcal{E}_c - \mathcal{E}_d$ и $\mathcal{E}_a - \mathcal{E}_v$ изменяются с концентрацией примеси так, как это изображено на фиг. 86. Они считают, что уменьшение энергии ионизации с увеличением концентрации примеси обусловлено электростатическим притяжением электрона, покинувшего



Фиг. 86. Зависимость энергии ионизации примесных центров от их концентрации.

свой собственный донор, другими донорами. Эти авторы нашли, что экспериментальные данные для дырок, приведенные на фиг. 86, соответствуют формуле

$$\mathcal{E}_a - \mathcal{E}_v = 0,08 - 4,3 \cdot 10^{-8} N_a^{1/3} \text{ эВ}, \quad (9.14)$$

с помощью которой вычислена сплошная кривая, изображенная на этом рисунке. Численное значение коэффициента в этой формуле имеет правильный порядок величины, что можно проверить, вычислив электростатический потенциал, создаваемый одним ионом на расстоянии, равном половине среднего расстояния между ионами. Это расстояние будет приблизительно равно $1/2 N_a^{1/3} = r$. Потенциал на этом расстоянии равен

$$\begin{aligned} \frac{e}{\kappa r} \text{ эл.-ст. ед. напряжения} &= 300 \frac{4,8 \cdot 10^{-10}}{11,9 r} \text{ в} = \\ &= 1,2 \cdot 10^{-8} r^{-1} = 2,4 \cdot 10^{-8} N_a^{1/3} \text{ в}. \end{aligned} \quad (9.15)$$

Применив более точный метод, основанный на вычислении среднего значения потенциала вблизи другого иона, Пирсон и Бардин получили несколько иное значение, равное $3 \cdot 10^{-8} N_a^{1/3}$.

Зейтц¹⁾ указал, что метод, использованный Пирсоном и Бардином, встречает возражения в связи с тем обстоятельством, что волновые функции, использованные ими при расчете, не ортогональны. Исследования Зейтца показывают, что объяснение фиг. 86 может быть связано с совсем иными явлениями.

ЗАДАЧИ

1. Вывести формулу (9.10), рассматривая сферу в пространстве импульсов с радиусом P_m , определяемую уравнением

$$\mathcal{E}_m = \frac{P_m^2}{2m}.$$

2. Вычислить первый и второй потенциалы ионизации примесных центров, образующих твердый раствор замещения, валентность которых отличается на две единицы от валентности кремния и германия, считая при этом, что первый потенциал ионизации гелия равен 24,5 эв. Вычислить радиусы волновых функций для случая связанного электрона и сравнить их с постоянной решетки.

ЛИТЕРАТУРА

1. Wannier G., Phys. Rev., 52, 191 (1937).
2. Tibbs S. R., Trans. Faraday Soc., 35, 1471 (1939).
3. Пекар С., Journ. of Phys., СССР, 10, 431 (1946).
4. James H. M., Phys. Rev., 76, 1602, 1611 (1949).
5. Slater J. C., Phys. Rev., 76, 1592 (1949).
6. Briggs H. B., Phys. Rev., 77, 287 (1950).
7. Mulaney J. F., Phys. Rev., 66, 326 (1944).
8. Briggs H. B., Brattain W. H., Phys. Rev., 75, 1705 (1949).
9. Becker M., Fan H. Y., Phys. Rev., 76, 1530 (1949).
10. Pearson G. L., Bardeen J., Phys. Rev., 75, 865 (1949).
11. Pearson G. L., Shockley W., Phys. Rev., 71, 142 (1947).

1) Частное сообщение — *Прим. авт.*

СТАТИСТИКА ФЕРМИ — ДИРАКА В ПРИМЕНЕНИИ
К ПОЛУПРОВОДНИКАМ

§ 1. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ФЕРМИ — ДИРАКА

В этом параграфе описывается влияние теплового возбуждения на распределение электронов между различными квантовыми состояниями. Сначала рассмотрим это распределение при температуре абсолютного нуля, а затем проследим, как оно изменяется при возбуждении электронов, находящихся в квантовых состояниях с низкой энергией, в квантовые состояния с более высокой энергией. Заполнение электронами квантовых состояний с разными энергиями может быть описано с помощью функции распределения Ферми — Дирака, о которой идет речь в первой части этого параграфа. Эта функция распределения применима вообще к электронам в том смысле, что она не зависит от детального характера распределения квантовых состояний по энергиям. В конце параграфа для иллюстрации общих положений теории приводится пример применения функции распределения Ферми — Дирака при рассмотрении гипотетического случая распределения 20 электронов по 36 квантовым состояниям. За этим примером следует определение распределения квантовых состояний по энергиям в примесном полупроводнике.

При температуре абсолютного нуля электроны находятся в самых низких возможных для них энергетических состояниях и заполняют их, подчиняясь ограничениям, налагаемым принципом Паули. В результате все квантовые состояния, обладающие энергией, лежащей ниже некоторого определенного значения, оказываются занятыми, в то время как состояния с более высокой энергией остаются незанятыми. Однако при наличии теплового возбуждения электроны будут стремиться переходить в состояния с более высокой энергией. На основе статистической механики число таких переходов тем больше, чем больше величина kT , представляющая собой среднюю тепловую энергию одномерного осциллятора при температуре T . Соответственно этому, электроны, энергии которых при абсолютном нуле лежат примерно на величину kT ниже энергий ближайших незанятых состояний, могут вследствие возбуждения перейти в эти более высокие энергетические состояния. Однако для электронов, энергия которых на много kT отличается от энергий незанятых состояний, вероятность возбуждения ничтожно мала. Таким образом, в случае равновесного распределения при более высоких температурах резкая граница между занятыми и незанятыми состояниями размывается в переходную область шириной порядка $2kT$.

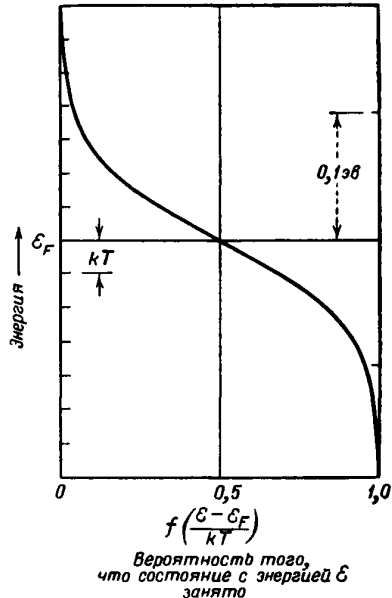
Такое распределение описывается функцией Ферми — Дирака. Эта функция, обозначаемая буквой f , зависит от \mathcal{E} и от T . Вид этой зависимости мы вскоре рассмотрим. Значение функции f изменяется в пределах от нуля до единицы. Эта функция имеет следующий смысл:

Значение функции Ферми — Дирака f есть вероятность того, что при тепловом равновесии квантовое состояние с энергией \mathcal{E} занято электроном. Иначе говоря, это есть среднее значение доли всех квантовых состояний с энергией \mathcal{E} , занятых электронами.

Значение только что описанных свойств функции распределения станет более понятным при рассмотрении приводимых примеров¹⁾. Функция f представлена на фиг. 87, из которой видно, что она изменяется от нуля (при высоких значениях энергии) до единицы (при низких значениях энергии). Основное изменение функции f от значений, почти равных нулю, до значений, почти равных единице, происходит в интервале энергий порядка $4kT$. Энергия \mathcal{E}_F , при которой функция f принимает значение, равное $1/2$, определяет *уровень Ферми*.

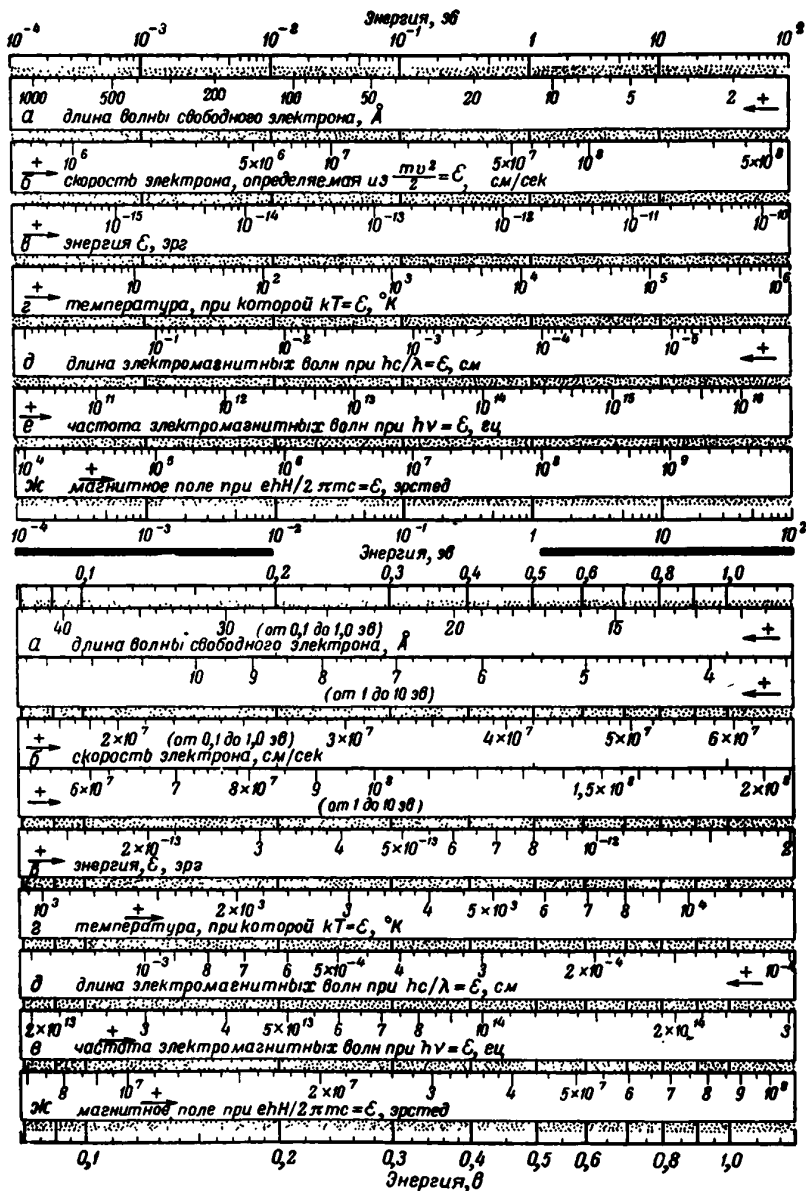
Для сравнения на фиг. 87 нанесена также шкала в электрон-вольтах. При комнатной температуре, т. е. при $T = 300^\circ\text{K}$, kT соответствует энергии $0,026 \text{ эв}$. Иногда бывает также удобно сравнить эти энергии с энергией световых квантов. На фиг. 88 приведена таблица, иллюстрирующая соотношение между различными единицами для измерения энергии.

Следует указать на статистическую природу функции распределения Ферми — Дирака. Функция f определяет *вероятность* того, что квантовое состояние с данной энергией занято. Если, например, рассматриваемая система имеет 6 квантовых состояний с энергией, которой соответствует $f = 1/2$, то в условиях равновесия в этих состояниях в среднем будет находиться 3 электрона. Однако это число



Фиг. 87. Функция распределения Ферми — Дирака.

¹⁾ Основы статистической теории функции распределения Ферми — Дирака даны в гл. 16 — *Прим. авт.*



Фиг. 88. Сравнение различных энергетических шкал.

будет изменяться по мере того, как электроны в результате теплового возбуждения будут переходить из одних состояний в другие; в действительности, в течение $1/64$ части какого-то промежутка времени квантовые состояния будут заполнены шестью электронами. В течение $6/64$ этого промежутка времени они будут содержать 5 электронов, в течение $15/64$ времени — 4 электрона, в течение $20/64$ времени — 3 электрона, в течение $15/64$ времени — 2 электрона, в течение $6/64$ времени — 1 электрон, а в течение $1/64$ времени все состояния будут оставаться пустыми. В этом случае, когда рассматривается лишь небольшое число состояний, статистические флуктуации будут велики. Однако в полупроводнике мы обычно имеем дело с большим числом состояний — 10^{10} или больше. При таком большом числе состояний статистические флуктуации в процентном отношении ничтожны.

Как будет показано ниже, функция распределения Ферми — Дирака обладает полной симметрией в отношении дырок и электронов. Если (см. фиг. 87) уровень энергии, для которого $f = 1/2$, называемый уровнем Ферми, обозначить через \mathcal{E}_F , то значение f для энергии \mathcal{E} дается формулой ¹⁾

$$f \equiv f\left(\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right) = \frac{1}{1 + \exp[(\mathcal{E} - \mathcal{E}_F)/kT]}. \quad (10.1)$$

Как уже было разъяснено выше, значение функции f определяет вероятность того, что квантовое состояние с энергией \mathcal{E} занято. Если \mathcal{E} больше \mathcal{E}_F на величину нескольких kT , то экспоненциальный член в знаменателе велик по сравнению с единицей. Тогда (при $\mathcal{E} > \mathcal{E}_F$) выражение для f приближенно можно записать в виде

$$f \approx \exp\left[-\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right]. \quad (10.2)$$

Это приближение дает то же распределение электронов по квантовым состояниям, что и „классическое“ распределение Максвелла—Больцмана. Мы будем ссылаться на него, а также на соответствующее приближение для дырок, рассматриваемое ниже, как на *классическое приближение*.

С другой стороны, если \mathcal{E} меньше \mathcal{E}_F , то экспоненциальный член в знаменателе мал; в этом случае ($\mathcal{E} < \mathcal{E}_F$) может быть использовано приближение $(1 + e)^{-1} \approx 1 - e$, что дает

$$f \approx 1 - \exp\left[\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right]. \quad (10.3)$$

Таким образом, для энергий, значительно превышающих \mathcal{E}_F , функция f стремится к нулю, как $\exp(-\Delta\mathcal{E}/kT)$, где $\Delta\mathcal{E}$ отсчитывается от \mathcal{E}_F в

¹⁾ Как показывает выражение (10.1), f — функция трех переменных: \mathcal{E} , \mathcal{E}_F и T . В тех случаях, когда значения T и \mathcal{E}_F не меняются, мы будем иногда записывать функцию распределения просто как f или $f(\mathcal{E})$. — *Прим. авт.*

сторону более высоких значений энергии (*вверх*). Ниже \mathcal{E}_F функция f стремится к единице; разность между значением функции f и единицей определяется выражением $\exp(-\Delta\mathcal{E}/kT)$, где $\Delta\mathcal{E}$ отсчитывается теперь *вниз* от уровня \mathcal{E}_F .

Если f — доля квантовых состояний, занятых электронами, то $1-f$, очевидно, доля состояний, остающихся вакантными, или, другими словами, доля состояний, занятая дырками. Приближенная формула (10.3) показывает, таким образом, что вероятность того, что квантовое состояние занято дырками с энергией, лежащей на величину $\Delta\mathcal{E}$ ниже уровня \mathcal{E}_F , в точности равна вероятности встречи квантового состояния, занятого электроном, энергия которого лежит выше \mathcal{E}_F на величину $\Delta\mathcal{E}$. Этот результат справедлив в общем случае, а не только для приближений (10.2) и (10.3). Если записать функцию распределения для дырок в виде $f_p = 1-f$, то путем простого алгебраического преобразования можно показать, что

$$f_p \equiv f_p \left[\frac{\mathcal{E} - \mathcal{E}_F}{kT} \right] = 1 - f = \frac{1}{1 + \exp[(\mathcal{E}_F - \mathcal{E})/kT]}. \quad (10.4)$$

Сопоставляя это выражение с (10.1), можно увидеть, что f_p зависит от энергий, отсчитываемых вниз от уровня \mathcal{E}_F , точно так же, как функция f зависит от энергий, отсчитываемых вверх от этого уровня. Другими словами, функция f (см. фиг. 87) обладает центром симметрии при $\mathcal{E} = \mathcal{E}_F$, когда $f = 1/2$.

Функции f и f_p являются функциями только одной переменной $(\mathcal{E} - \mathcal{E}_F)/kT$. Во многих случаях, однако, \mathcal{E}_F и T — постоянные величины, и единственной переменной является \mathcal{E} или же импульс P , от которого зависит \mathcal{E} . В таких случаях эти функции можно записать сокращенно в виде

$$f = f(\mathcal{E}) = f[\mathcal{E}(P)] = f(P), \quad (10.5a)$$

$$f_p = f_p(\mathcal{E}) = f_p[\mathcal{E}(P)] = f_p(P). \quad (10.5b)$$

Такая форма записи особенно удобна при рассмотрении квантовых состояний в зоне Бриллюэна. Хотя с математической точки зрения такая форма записи не вполне последовательна, однако это не должно нас особенно смущать.

Одна из основных задач при исследовании равновесного распределения электронов по квантовым состояниям связана с определением положения уровня Ферми \mathcal{E}_F . Как будет показано ниже, положение \mathcal{E}_F определяется, во-первых, распределением квантовых состояний по энергиям рассматриваемой системы, во-вторых, общим числом электронов в этой системе, и, в-третьих, значением абсолютной температуры. Температура имеет существенное значение при определении положения уровня Ферми в полупроводниках, в то время как в металлах, в отличие от полупроводников, положение уровня Ферми почти не зависит от температуры. Если при данной температуре положение

уровня Ферми известно, то распределение электронов по энергиям может быть получено путем умножения числа квантовых состояний, приходящихся на единицу энергий, $N(\mathcal{E})$, на функцию $f[(\mathcal{E}-\mathcal{E}_F)/kT]$, определяющую ту долю этих состояний, которая занята электронами. Несколько подробных примеров таких вычислений приведено ниже.

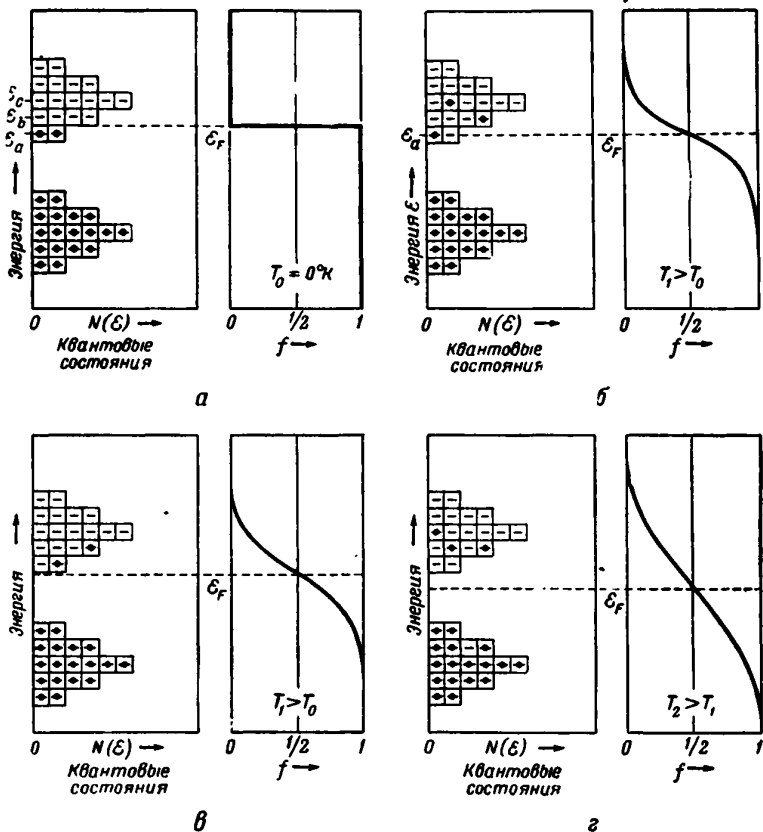
Если \mathcal{E}_F практически не зависит от температуры, то температурная зависимость f для значений энергии \mathcal{E} , отличающихся от \mathcal{E}_F на несколько kT , может быть приближенно найдена с помощью выражений (10.2) или (10.3). Как уже указывалось, из этих выражений следует, что число электронов или дырок изменяется в зависимости от температуры как $\exp(-\Delta\mathcal{E}/kT)$. Зависимость $\ln f$ от $1/T$ изображается в этом случае прямой линией (см. гл. 1). Величина $\Delta\mathcal{E}$ определяет наклон этой прямой или значение энергии активации. Однако для того чтобы выяснить, чему соответствует величина $\Delta\mathcal{E}$ с точки зрения схемы энергетических уровней, необходимо знать положение уровня \mathcal{E}_F . Этот вопрос исследуется ниже.

§ 2. СТАТИСТИКА ФЕРМИ—ДИРАКА В ПРИМЕНЕНИИ К ПРОСТОЙ МОДЕЛИ

Прежде чем приступить к рассмотрению полупроводника, схема распределения квантовых состояний которого представлена на фиг. 85, рассмотрим небольшую систему, состоящую из 20 электронов и характеризующуюся всего лишь 36 состояниями. Преимущество рассмотрения небольшой системы заключается в том, что распределение электронов можно наглядно изобразить графически, не прибегая к математическим расчетам. При таком небольшом числе возможных состояний и электронов статистические флуктуации весьма существенны и их следовало бы принять во внимание. Однако мы не будем их рассматривать, так как общие закономерности распределения электронов будут носить точно такой же характер и для больших систем, где флуктуации уже не имеют существенного значения. Распределение квантовых состояний по энергиям, представленное на фиг. 89, приближенно соответствует двум энергетическим зонам, так что оно имеет некоторое сходство с распределением квантовых состояний в полупроводнике.

На фиг. 89,а представлено распределение электронов по состояниям при температуре абсолютного нуля. При этой температуре электроны будут стремиться занять самые низкие квантовые состояния; однако в силу принципа Паули, согласно которому данное состояние может занимать не более чем один электрон, они будут заполнять состояния и с более высокими энергиями, что и показано на рисунке. При температуре абсолютного нуля переход функции Ферми от значения, равного нулю, к значению, равному единице, будет очень резким, в результате чего все состояния, расположенные выше уровня \mathcal{E}_F , будут пустыми, а все состояния, лежащие ниже \mathcal{E}_F ,

заполнены электронами. Так как рассматриваемая нами система содержит всего 20 электронов, то уровень Ферми должен лежать между уровнями энергии ξ_a и ξ_b . Если бы он был проведен ниже ξ_a , то



Фиг. 89. Распределение квантовых состояний и функция Ферми — Дирака в условиях равновесия.

а — распределение 20 электронов по состояниям с наименьшей энергией при $T_0=0^{\circ}K$; б — при температуре T_1 , случай $\xi_F=\xi_a$ отвечал бы системе из 21 электрона; в — правильный выбор положения ξ_F для системы из 20 электронов при температуре T_1 ; г — дальнейшее смещение ξ_F при $T_2 > T_1$.

это соответствовало бы системе, содержащей максимум 18 электронов, а если бы он был проведен выше уровня ξ_b , то это отвечало бы системе, содержащей минимум 24 электрона.

При более высоких температурах уровень Ферми должен лежать, однако, ниже уровня ξ_a . Это легко понять из рассмотрения фиг. 89,б,

которая соответствует тому случаю, когда kT сравнимо с расстояниями между энергетическими уровнями верхней группы состояний. Уровень Ферми проведен здесь так, как если бы он точно соответствовал энергии \mathcal{E}_a . При таких условиях в среднем половина состояний с энергией \mathcal{E}_a будет занята. Так как электроны в изображенном на рисунке случае возбуждаются также и в состояния, лежащие значительно выше уровня Ферми, то имеется заметная вероятность нахождения электронов в состояниях с энергиями \mathcal{E}_b и \mathcal{E}_c . При этом можно ожидать, что в каждом из этих состояний будет находиться приблизительно по одному электрону. Конечно, распределение, изображенное на фиг. 89,б, является только одним из большого числа возможных распределений, которые описываются функцией Ферми; оно, однако, правильно отражает средние свойства системы, поскольку речь идет о числе электронов, обладающих данной энергией. В верхней группе состояний в среднем находится 3 электрона. С другой стороны, нижняя группа квантовых состояний лежит настолько ниже уровня Ферми, что вероятность нахождения дырки в этой группе ничтожно мала. Следовательно, мы видим, что полное число электронов в этом случае равно 21, т. е. на 1 электрон больше, чем требуется. Для того чтобы получить искомое число электронов, равное 20, необходимо сместить уровень Ферми вниз, как показано на фиг. 89,в. Это смещение уровня Ферми вниз по отношению к распределению квантовых состояний приводит к уменьшению вероятности того, что какое-либо состояние будет занято электроном, как это легко видеть из формы кривой распределения Ферми — Дирака. Следовательно, смещение уровня Ферми вниз на фиг. 89,в по сравнению с фиг. 89,б отвечает уменьшению вероятности того, что состояния верхней группы будут заняты, и, если смещение произведено правильно, вероятность эта такова, что в верхней группе в среднем будет находиться только 2 электрона и, таким образом, будет получено искомое число электронов, равное 20.

Смещение уровня Ферми с повышением температуры является характерным свойством полупроводников. В металлах дело обстоит совершенно иначе — положение уровня Ферми при высоких температурах почти то же, что и при абсолютном нуле.

Распределение при еще более высоких температурах для нашего простого примера показано на фиг. 89,г. В этом случае kT сравнимо с расстоянием между двумя группами квантовых состояний. При этих условиях некоторые электроны будут возбуждены из нижней группы в верхнюю и, в соответствии с условием постоянства общего числа электронов, равного 20, уровень Ферми сдвинут еще дальше вниз — ближе к середине промежутка между двумя группами состояний.

Мы дадим теперь обобщение этих результатов в аналитической форме в приложении к случаю распределения квантовых состояний в полупроводнике.

§ 3. СТАТИСТИКА ФЕРМИ—ДИРАКА В ПРИМЕНЕНИИ К ПОЛУПРОВОДНИКАМ

При исследовании полупроводников существенную роль играет условие электрической нейтральности объема полупроводника. Это условие логически вытекает из хорошо известной теоремы теории электропроводности, заключающейся в том, что в равновесном состоянии результирующая плотность заряда внутри проводника должна быть равна нулю. В применении к полупроводнику это требование эквивалентно сформулированному в § 2 требованию, согласно которому общее число электронов, занимающих квантовые состояния, должно оставаться неизменным. Сначала рассмотрим факторы, определяющие это число, а затем выясним, как должен быть расположен уровень Ферми, чтобы удовлетворить этому требованию.

Так как нас интересуют относительно малые отступления от идеальной структуры валентных связей, то удобнее иметь дело не с полным числом электронов, а только с избытком их по отношению к тому числу, которое необходимо для заполнения структуры валентных связей. Число избыточных электронов в единице объема кристалла определяется четырьмя величинами: n (концентрация избыточных электронов в зоне проводимости, т. е. электронов, не связанных с донорами); n_d (концентрация избыточных электронов, находящихся в связанных состояниях вблизи доноров); p (концентрация дырок в зоне валентных связей, т. е. дырок, не связанных с акцепторами); p_a (концентрация дырок, находящихся в связанных состояниях вблизи акцепторов).

Если все эти концентрации равны нулю, то структура валентных связей завершена и в кристалле нет избыточных электронов. Если же они отличны от нуля, то очевидно, что число избыточных электронов в единице объема кристалла будет равно

$$n + n_d - p - p_a. \quad (10.6)$$

В дополнение к четырем только что определенным значениям концентраций введем еще ряд других величин, которые потребуются при рассмотрении вопросов, обсуждаемых в этой главе. Для удобства приведем их в табл. 6; точное определение величин N_c и N_v будет дано лишь в дальнейшем. Для краткости мы пользуемся термином „концентрация“ вместо выражения „число частиц в единице объема“.

Если структура валентных связей идеальна и избыточные электроны отсутствуют, то кристалл, вообще говоря, все же может и не быть нейтральным вследствие наличия локализованных зарядов доноров и акцепторов. Как указывалось в гл. 1 и 9, в тех случаях, когда вблизи атома донора структура валентных связей идеальна, этот атом представляет собой локализованный положительный заряд, а акцептор соответственно — локализованный отрицательный заряд. При концентрациях доноров и акцепторов, равных N_d и N_a , это означает, что

результатирующая плотность заряда в полупроводнике равна $e(N_d - N_a)$, если структура валентных связей при этом полностью завершена.

Таблица 6

Положительные заряды	Отрицательные заряды	Другие величины
<p>p — концентрация дырок в зоне валентных связей</p> <p>p_a — концентрация дырок, связанных с акцепторами</p> <p>$p_t = p + p_a$</p> <p>N_d — концентрация доноров</p>	<p>n — концентрация избыточных электронов в зоне проводимости</p> <p>n_d — концентрация избыточных электронов, связанных с донорами</p> <p>$n_t = n + n_d$</p> <p>N_a — концентрация акцепторов</p>	<p>N_g — число узлов решетки в единице объема</p> <p>$N_o \equiv N_o(T)$ — эффективная плотность состояний в зоне проводимости</p> <p>$N_v \equiv N_v(T)$ — эффективная плотность состояний в зоне валентных связей</p>

Так как выражение (10.6) определяет число избыточных электронов (по сравнению с числом электронов, образующих структуру валентных связей), то результирующая плотность заряда, которая должна равняться нулю, дается выражением

$$e(N_d - N_a) - e(n + n_d - p - p_a) = 0 \quad (10.7)$$

или

$$n + n_d - p - p_a = N_d - N_a. \quad (10.8)$$

Это выражение показывает, что число избыточных электронов в точности равно избытку числа доноров над числом акцепторов (такой случай уже рассматривался в связи с фиг. 10).

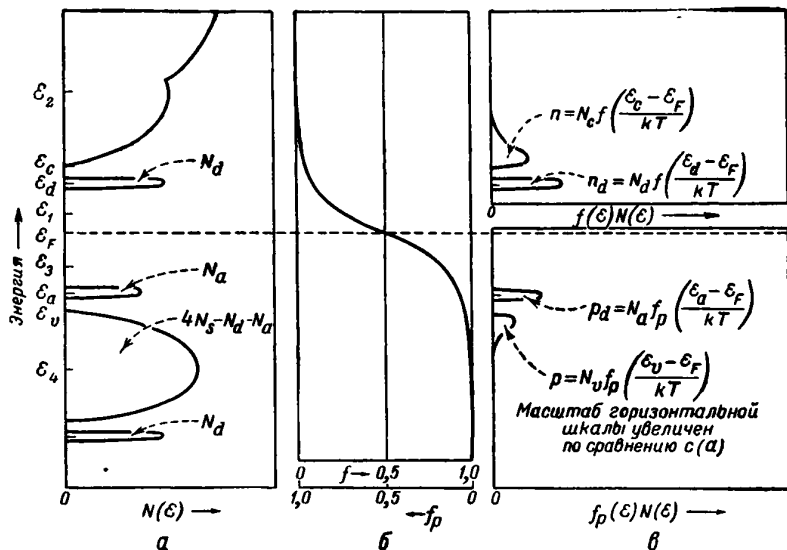
Значения n , n_d , p и p_a определяются: 1) распределением квантовых состояний по энергиям; 2) положением уровня \mathcal{E}_F ; 3) температурой. Соответствующие уравнения получаются следующим образом. В интервале значений энергии $d\mathcal{E}$ содержится $dS = N(\mathcal{E}) d\mathcal{E}$ квантовых состояний. Часть этих состояний, равная $f[(\mathcal{E} - \mathcal{E}_F)/kT]$, занята; следовательно, число электронов с энергиями, лежащими в интервале $d\mathcal{E}$, равно

$$f\left(\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right) dS = f\left(\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right) N(\mathcal{E}) d\mathcal{E}. \quad (10.9)$$

Если проинтегрировать это выражение по некоторой области значений энергий от \mathcal{E}_1 до \mathcal{E}_2 (фиг. 90), внутри которой заключены состояния доноров и зона проводимости, то мы получим величину, равную сумме $(n_d + n)$

$$n_t = n + n_d = \int_{\mathcal{E}_1}^{\mathcal{E}_2} f\left(\frac{\mathcal{E} - \mathcal{E}_F}{kT}\right) N(\mathcal{E}) d\mathcal{E}. \quad (10.10)$$

Точное определение значения верхнего предела \mathcal{E}_2 при этом несущественно, так как функция f быстро приближается к нулю при больших энергиях. Действительно, значение f ничтожно мало повсюду, за исключением самого дна энергетической зоны. Это показано на фиг. 90,б, где графически изображен ход функций $N(\mathcal{E})f$ и $N(\mathcal{E})f_p$.



Фиг. 90. Распределение квантовых состояний и концентрации электронов и дырок в полупроводнике n -типа.

a — распределение квантовых состояний по энергиям в полупроводнике n -типа; $б$ — распределение Ферми—Дирака; $в$ — число электронов и дырок в единице объема на единицу энергии и формулы для концентраций. Масштаб в горизонтальном направлении увеличен.

Как мы видим, существенную роль при интегрировании играют только области вблизи краев зон. Более детально этот вопрос будет рассмотрен в дальнейшем.

Аналогичные вычисления дают для числа дырок

$$p_t = p + p_a = \int_{\epsilon_3}^{\epsilon_4} f_p \left(\frac{\mathcal{E} - \mathcal{E}_F}{kT} \right) N(\mathcal{E}) d\mathcal{E}, \quad (10.11)$$

где область значений энергий от \mathcal{E}_3 до \mathcal{E}_4 захватывает заполненную зону и акцепторные уровни, а $f_p = 1 - f$ [см. уравнение (10.4)] есть вероятность того, что квантовое состояние с энергией \mathcal{E} свободно (т. е. занято дыркой).

При данной температуре T как n_i , так и p_i , определяемые выражениями (10.10) и (10.11), являются функциями только \mathcal{E}_F . Следовательно,

$$n_i - p_i = N_d - N_a \quad (10.12)$$

может рассматриваться как уравнение только с одним неизвестным, — \mathcal{E}_F , так как N_d , N_a и $N(\mathcal{E})$ определяются природой рассматриваемого полупроводника. Этот случай подобен уже рассмотренному нами в связи с анализом фиг. 89. В гл. 16 будет приведено несколько подробных примеров определения \mathcal{E}_F . В этом параграфе мы ограничимся тем, что покажем только, как величины n_i и p_i изменяются в общем случае.

В выражении, определяющем n_i , интегрирование производится по N_d донорным уровням. Предполагается, что все эти уровни соответствуют одному и тому же значению энергии \mathcal{E}_d , при которой $f = f[(\mathcal{E}_d - \mathcal{E}_F)/kT]$. Интегрируя это выражение, получим число электронов, находящихся на донорных уровнях,

$$n_d = N_d f\left(\frac{\mathcal{E}_d - \mathcal{E}_F}{kT}\right), \quad (10.13)$$

что и указано на фиг. 90.

Число электронов в зоне проводимости получается интегрированием по $N(\mathcal{E})$ вблизи дна этой зоны. Интегрирование упрощается благодаря тому обстоятельству, что в большинстве случаев \mathcal{E}_F лежит значительно ниже дна зоны проводимости, так что может быть использовано приближенное выражение (10.2). Это позволяет написать

$$n = e^{\mathcal{E}_F/kT} \int_{\mathcal{E}_c}^{\mathcal{E}_1} N(\mathcal{E}) e^{-\mathcal{E}/kT} d\mathcal{E}. \quad (10.14)$$

Так как экспоненциальный множитель под знаком интеграла убывает в e^{-1} раз при каждом увеличении \mathcal{E} на kT , то в качестве грубого приближения можно произвести интегрирование по области, ограниченной поверхностью постоянной энергии, лежащей внутри зоны Бриллюэна, для которой $\mathcal{E} - \mathcal{E}_c = kT$ или $P^2/2m = kT$. Так как kT очень мало по сравнению с шириной энергетической зоны, то можно использовать приближение $\mathcal{E} = \mathcal{E}_c + P^2/2m$ [см. (7.15)], так что поверхность постоянной энергии представляет собой сферу, объем которой равен $4\pi P^3/3$, где $P^2/2m = \mathcal{E} - \mathcal{E}_c = kT$. Эта сфера содержит $(2/h^3)(4\pi P^3/3) = 8\pi P^3/3h^3$ квантовых состояний на единицу объема кристалла. Так как в этой области $\exp(-\mathcal{E}/kT) \approx \exp(-\mathcal{E}_c/kT)$, то можно аппроксимировать интеграл выражением $\exp(-\mathcal{E}_c/kT) \cdot 8\pi P^3/3h^3$. Записывая $P = (2mkT)^{1/2}$, имеем

$$n \approx e^{-(\mathcal{E}_c - \mathcal{E}_F)/kT} \frac{8\pi}{3} \left(\frac{2mkT}{h^2}\right)^{3/2}. \quad (10.15)$$

Точное интегрирование, выполненное в гл. 16, дает

$$n = e^{-(\mathcal{E}_c - \mathcal{E}_F)/kT} 2 \left(\frac{2\pi m k T}{h^2} \right)^{3/2} = N_c e^{-(\mathcal{E}_c - \mathcal{E}_F)/kT}. \quad (10.16)$$

Это значение n примерно на $1/8$ больше предыдущего. Сравнивая это выражение с выражением (10.13) для n_d и используя для f приближение (10.2), находим, что зона проводимости ведет себя подобно группе состояний с энергией \mathcal{E}_c , число которых равно не N_d , а

$$N_c = 2 \left(\frac{2\pi m k T}{h^2} \right)^{3/2} = 4,82 \cdot 10^{16} T^{3/2} \text{ см}^{-3}. \quad (10.17)$$

[Если эффективная масса отлична от m , то необходимо ввести множитель $(m_n/m)^{3/2}$.] Величину N_c можно рассматривать как *эффективную плотность состояний в зоне проводимости*.

Полученное выше выражение для n справедливо только в том случае, если \mathcal{E}_F меньше \mathcal{E}_d на величину порядка нескольких kT . При комнатной температуре, полагая, что m равно массе свободного электрона, получаем $N_c = 2,41 \cdot 10^{19} \text{ см}^{-3}$, что составляет приблизительно $1/2000$ числа атомов в 1 см^3 . Величина N_c медленно меняется с температурой T по сравнению с показательными функциями, входящими в рассматриваемые формулы, и во многих случаях может считаться постоянной. Совершенно аналогичные вычисления могут быть выполнены для p_d и p :

$$p_d = N_d f_p \left[\frac{\mathcal{E}_c - \mathcal{E}_F}{kT} \right] = N_d f \left[\frac{\mathcal{E}_F - \mathcal{E}_c}{kT} \right], \quad (10.18)$$

$$p = 2 \left(\frac{2\pi m k T}{h^2} \right)^{3/2} e^{-(\mathcal{E}_F - \mathcal{E}_v)/kT} = N_v e^{-(\mathcal{E}_F - \mathcal{E}_v)/kT}. \quad (10.19)$$

Величины N_c и N_v различны, если электроны и дырки обладают различными эффективными массами.

Возможность введения в рассмотрение величин N_c и N_v ограничивается теми случаями, когда функции f и f_p могут быть аппроксимированы простыми показательными функциями выражений (10.2) и (10.3). Если концентрация носителей заряда достаточно велика, эти приближения уже не пригодны и следует пользоваться другими формулами. Это будет показано для случая электронов.

Если число электронов в зоне проводимости очень велико, так что на каждый атом, например, приходится по одному электрону, как это имеет место в одновалентных металлах, натрия и меди, то уровень Ферми будет лежать значительно выше дна зоны проводимости. Для натрия, который изучен наиболее подробно, имеются экспериментальные и теоретические данные, указывающие на то, что вплоть до самых высоких занятых уровней энергия электрона равна $P^2/2m$, так что для $N(\mathcal{E})$ может быть использована формула (9.9). Вычисление показывает, что в этом случае уровни должны быть

заняты электронами вплоть до максимальной энергии \mathcal{E}_m , лежащей примерно на 3,16 эв выше дна зоны проводимости. Эта энергия настолько больше kT , что область, в которой функция f претерпевает существенное изменение, в сравнении с ней ничтожно мала, и значение \mathcal{E}_F может быть определено просто путем решения (9.10) относительно \mathcal{E}_F :

$$n = \int_{\mathcal{E}_c}^{\mathcal{E}_F} N(\mathcal{E}) d\mathcal{E} = \frac{8\pi}{3h^3} [2m(\mathcal{E}_F - \mathcal{E}_c)]^{3/2}, \quad (10.20a)$$

$$\mathcal{E}_F - \mathcal{E}_c = \mathcal{E}_m = \frac{h^2 (3n/8\pi)^{2/3}}{2m} = 21,6 \cdot 10^{-16} n^{2/3} \text{ эв}, \quad (10.20b)$$

где $\mathcal{E}_F = \mathcal{E}_c + \mathcal{E}_m$ и m принято равным массе свободного электрона. Случаи такого рода, когда температура уже не оказывает заметного влияния на распределение электронов по состояниям, называются *вырожденными*, а система электронов называется при этом *вырожденным электронным газом*.

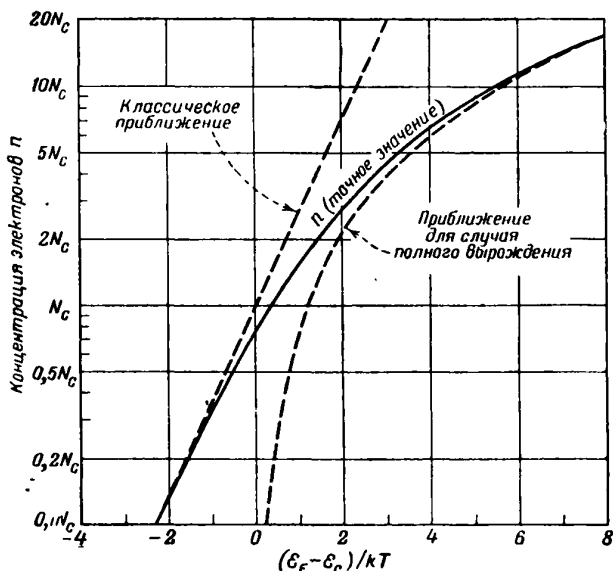
При любой данной температуре, по мере увеличения числа электронов в зоне, мы будем иметь дело с переходом от классического случая к вырожденному. Граница между ними условно определяется концентрацией вырождения $n_{\text{вырожд.}}$ или температурой вырождения $T_{\text{вырожд.}}$. Эти две величины связаны уравнением

$$T_{\text{вырожд.}} = \left[\frac{3}{\pi} \right]^{3/2} \frac{h^2}{8km} n_{\text{вырожд.}}^{2/3} = 4,2 \cdot 10^{-11} n_{\text{вырожд.}}^{2/3}, \quad (10.21)$$

которое интерпретируется следующим образом. При температуре абсолютного нуля электроны находятся в состояниях с самой низкой энергией и занимают все состояния вплоть до состояния с некоторой энергией \mathcal{E}_m , определяемой уравнением (10.20б). Температура вырождения удовлетворяет равенству $kT_{\text{вырожд.}} = \mathcal{E}_m$. Другими словами, если температура T достаточно высока, чтобы перевести электроны со дна зоны на уровни, лежащие выше уровня \mathcal{E}_m , вырождение будет в значительной степени нарушено.

Уравнение (10.21) с одинаковым успехом пригодно как для определения температуры вырождения при заданной концентрации носителей, так и для определения концентрации носителей при данной температуре. Если n сравнимо с $n_{\text{вырожд.}}$, то зависимость n от T и $(\mathcal{E}_F - \mathcal{E}_c)$ не может быть описана с помощью простой приближенной формулы, и значение n должно быть определено путем точного интегрирования выражения $N(\mathcal{E})f(\mathcal{E})d\mathcal{E}$ [1—4]. Полученная в результате такого точного интегрирования зависимость n от $(\mathcal{E}_F - \mathcal{E}_c)$ и T изображена на фиг. 91, на которой нанесены также кривые для предельных случаев (10.16) и (10.20). Точка, соответствующая $n = n_{\text{вырожд.}}$, дается равенством $\mathcal{E}_m = kT$ и, следовательно, на кривой для случая полного вырождения отвечает значению n при $\mathcal{E}_F - \mathcal{E}_c = kT$;

это значение $n \approx 0,75 N_c$. Точное решение при $n = 0,75 N_c$ соответствует малому отрицательному значению разности $\mathcal{E}_F - \mathcal{E}_c$, так что даже состояния на дне зоны в этом случае заполнены меньше



Фиг. 91. Точная зависимость n от $(\mathcal{E}_F - \mathcal{E}_c)$ и T ; сравнение ее с классической зависимостью и зависимостью для случая полного вырождения.

Классическое приближение: $n = N_c \cdot \exp [(\mathcal{E}_F - \mathcal{E}_c) / kT]$.

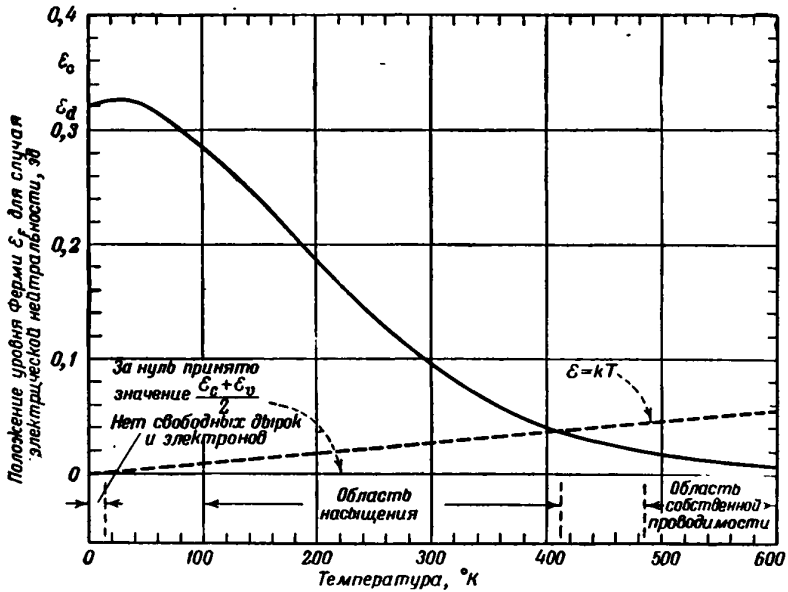
Приближение для случая полного вырождения:

$$n = \frac{8\pi}{3h^3} [2m(\mathcal{E}_F - \mathcal{E}_c)]^{3/2} = \frac{4N_c}{3V\pi} \left[\frac{\mathcal{E}_F - \mathcal{E}_c}{kT} \right]^{3/2}.$$

чем наполовину и распределение не является сильно вырожденным. Распределение будет, однако, полностью вырожденным, когда $(\mathcal{E}_F - \mathcal{E}_c)$ равно или больше $2kT$. Это соответствует значениям n , примерно в три раза превышающим $n_{\text{вырожд.}}$, для данной температуры T , или же температурам, составляющим $3^{-2/3} = 0,48$ температуры вырождения $T_{\text{вырожд.}}$, отвечающей данному значению n .

Случай вырожденного электронного газа будет рассмотрен в гл. 11. Для иллюстрации метода нахождения равновесного распределения электронов в полупроводнике проще, однако, рассмотреть невырожденный электронный газ. Поэтому мы вернемся к выражениям для n , n_a , p и p_a , определяющим эти величины в виде простых функций от \mathcal{E}_F . Эти выражения подставим в уравнение (10.8), выражающее

условие электрической нейтральности системы. Последнее приобретает при этом вид уравнения, которое может быть решено относительно ξ_F . Как уже указывалось, графический метод решения этого



Ф и г. 92. Зависимость положения уровня Ферми от температуры для случая электрической нейтральности в образце германия с $N_d = 10^{15} \text{ см}^{-3}$, $N_a = 10^{14} \text{ см}^{-3}$ и $\xi_c - \xi_d = 0,04 \text{ эв}$.

уравнения будет рассмотрен в гл. 16. Ниже приведены результаты, полученные при анализе частного примера.

Рассмотрим образец германия, для которого, как и в образце кремния А (см. фиг. 10 и 13), $N_d > N_a$. В частности, положим, что $N_d = 10^{16} \text{ см}^{-3}$, а $N_a = 10^{14} \text{ см}^{-3}$. Начало отсчета значений энергии ведется от середины запрещенной зоны, причем $\xi_c = 0,36$, $\xi_d = 0,32$, $\xi_a = -0,32$ и $\xi_v = -0,36 \text{ эв}$. Зависимость положения уровня ξ_F от температуры T показана для этого случая на фиг. 92. Ход соответствующих вычислений приведен в гл. 16. Качественное описание некоторых характерных особенностей фиг. 92 дано ниже.

При очень низких температурах

$$n_d = N_d - N_a, \quad (10.22)$$

так как дырки отсутствуют и все избыточные электроны находятся в связанных состояниях вблизи доноров. Следовательно, заполнена

доля n_d/N_d донорных уровней. Уровень \mathcal{E}_F должен находиться на расстоянии порядка нескольких kT от уровня \mathcal{E}_d , так как иначе все состояния были бы или заполнены или пусты, в соответствии с изображенным на фиг. 89,а ходом функции f при низких температурах. Так как уровень \mathcal{E}_F расположен близко от уровня \mathcal{E}_d , то число электронов в зоне проводимости будет изменяться как $\exp[-(\mathcal{E}_c - \mathcal{E}_d)/kT]$, согласно формуле (10.15). Таким образом, энергия активации, определенная из графика зависимости $\ln n$ от $1/T$, представляет собой в этом случае энергию связи электрона $(\mathcal{E}_c - \mathcal{E}_d)$ в донорном состоянии, а не половину этого значения. Область температур, для которой справедливо это приближение, распространяется примерно только до 10°K , как это можно видеть из графика зависимости положения уровня \mathcal{E}_F от температуры T , изображенной на фиг. 92.

По мере дальнейшего возрастания температуры в зону проводимости переходит в результате возбуждения значительное число электронов и уровень \mathcal{E}_F перемещается вниз, так что значение n_i остается постоянным за счет уменьшения числа электронов на донорных уровнях. В области температур, в которой доноры полностью ионизированы, справедливо уравнение

$$n \approx N_d - N_a. \quad (10.23)$$

Эта область температур называется *областью насыщения*, так как все имеющиеся избыточные электроны находятся при этом в зоне проводимости. Область насыщения достигается в том случае, когда уровень \mathcal{E}_F смещается ниже уровня \mathcal{E}_d на величину порядка нескольких kT , так что донорные уровни остаются в основном пустыми. На фиг. 92 нанесена зависимость величины kT , выраженной в электрон-вольтах, от температуры; как видно из графика, при температуре, близкой к 100°K , значение разности $\mathcal{E}_d - \mathcal{E}_F > 2kT$, так что при более высоких температурах доноры ионизированы полностью. В этой области повышение температуры должно сопровождаться понижением уровня \mathcal{E}_F по тем же причинам, которые были указаны выше — при анализе графиков, изображенных на фиг. 89, б и в.

При этом \mathcal{E}_F достигает, наконец, уровня, лежащего на несколько kT выше середины запрещенной зоны. Как только это произойдет, образуется заметное число дырок. При еще более высоких температурах уровень \mathcal{E}_F еще больше приближается к середине запрещенной зоны, так как в противном случае в сильной степени преобладали бы количественно либо дырки, либо электроны, и уравнение (10.8), выражающее условие электрической нейтральности, не могло бы быть выполнено. В результате этой стабилизации положения уровня \mathcal{E}_F концентрации n и p возрастают экспоненциально, как $\exp[-(\mathcal{E}_c - \mathcal{E}_F)/kT]$ и $\exp[-(\mathcal{E}_F - \mathcal{E}_v)/kT]$. Так как уровень \mathcal{E}_F лежит посередине между \mathcal{E}_c и \mathcal{E}_v , то оба выражения сводятся к $\exp[-(\mathcal{E}_c - \mathcal{E}_v)/2kT]$, так что, как указывалось в гл. 1, энергия активации равна половине ширины запрещенной зоны. {Область собственной проводимости

выбрана на фиг. 92 так, что для нее $n/p < e$, т. е. это есть область, для которой уровень \mathcal{E} расположен выше середины запрещенной зоны меньше чем на $1/2 kT$.

Когда N_a больше N_d , в полупроводнике всегда имеется избыток дырок. Как это следует из (10.3), (10.12) и (10.19), вследствие симметрии между выражениями, определяющими концентрации дырок и электронов, совершенно аналогичный результат получается в том случае, когда уровень \mathcal{E}_F постепенно перемещается от положения \mathcal{E}_a , которое он занимал при низких температурах, до середины запрещенной зоны $(\mathcal{E}_c + \mathcal{E}_v)/2$ при высоких температурах.

Таким образом, в этом параграфе в несколько более формальном виде рассматривается физическая картина явления, уже описанная в гл. 1. В гл. 16 более подробно исследуется изменение положения уровня \mathcal{E}_F с температурой, а также характер изменения концентрации электронов и дырок.

Резюмируя содержание этого параграфа, можно сказать, что если известно распределение квантовых состояний по энергиям и температура задана, то может быть найдено равновесное распределение электронов и дырок. Положение уровня Ферми \mathcal{E}_F при этом определяется условием электрической нейтральности системы [см. уравнение (10.8)]. Решение этой задачи может быть несколько упрощено, если мы введем в рассмотрение N_c и N_v — эффективные числа квантовых состояний в зоне проводимости и в зоне валентных связей. Однако задача остается все же весьма сложной, и только при низких температурах и для области собственной проводимости может быть дано простое толкование значениям энергии, определяемым графически из наклонов прямых, характеризующих зависимость $\ln n$ от $1/T$.

Следует отметить то обстоятельство, что имеется полная симметрия между соотношениями, даваемыми статистической теорией для дырок и электронов, как это видно из уравнения (10.4), а также из следующих уравнений этого параграфа. Отсюда, а также из результатов, полученных в гл. 7 и 8, следует, что единственным существенным различием между дырками и электронами в полупроводниках является различие знаков их эффективных зарядов¹⁾.

Важным следствием приближенных выражений (10.16) и (10.19) для n и p является то, что их произведение зависит только от T и не зависит ни от n , ни от p в отдельности:

$$\begin{aligned} np &= N_c N_v \exp\left(\frac{\mathcal{E}_c - \mathcal{E}_v}{kT}\right) = N_c N_v \exp\left(-\frac{\mathcal{E}_G}{kT}\right) = \\ &= 2,33 \cdot 10^{31} T^3 \exp\left(-\frac{\mathcal{E}_G}{kT}\right) \text{ см}^{-6}. \end{aligned} \quad (10.24)$$

¹⁾ Математическое исследование вопроса об эквивалентности дырок и электронов занимает значительную часть гл. 15 и 17. — *Прим. авт.*

В собственном полупроводнике концентрации n и p равны и каждая из них, в свою очередь, равна корню квадратному из уравнения (10.24). Мы используем это обстоятельство в следующем параграфе, в § 4 гл. 12 и § 4 гл. 16, а также в связи с анализом данных Пирсона и Бардина об электрических свойствах сплавов кремния. Соотношение (10.24) может быть получено и на основании статистической теории детального равновесия, согласно которой скорость рекомбинации дырок и электронов (пропорциональная np) должна равняться скорости их образования, зависящей от T , так же, как правая часть уравнения (10.24), и не зависящей ни от n , ни от p .

§ 4. СЛУЧАЙ КОМНАТНОЙ ТЕМПЕРАТУРЫ

В связи с тем, что германий, обладающий при комнатной температуре высоким удельным сопротивлением, представляет большой практический интерес, мы дадим упрощенный анализ рассмотренных в предыдущем параграфе вопросов применительно к этому случаю. В § 3 гл. 16 эта задача исследуется более общими методами; при этом значение $(\mathcal{E}_c - \mathcal{E}_v)$ полагается равным 0,72 эв, а значения $(\mathcal{E}_a - \mathcal{E}_v)$ и $(\mathcal{E}_c - \mathcal{E}_d)$ — равными 0,04 эв. Здесь мы ограничимся анализом некоторых экспериментальных данных. Экстраполяция графика температурного хода сопротивления германия в области собственной проводимости дает для удельного сопротивления при комнатной температуре значение, равное 48 ом·см¹). С другой стороны, сопротивление сравнительно чистого образца германия n -типа может быть порядка 2,4 ом·см. (Принимая для подвижности электронов в германии значение 2600 см²/в·сек, получаем для концентрации $n = \sigma_L / e_L \mu_L = 1/2,4 \cdot 1,6 \cdot 10^{-19} \cdot 2600 = 1,0 \cdot 10^{16}$ электронов/см³, в согласии с примером, приведенным в гл. 16.) Отношение подвижностей носителей заряда в германии равно²⁾ $b = 1,5$. Из этих данных можно определить непосредственно отношение концентраций n' (один штрих для образца с сопротивлением 2,4 ом·см) к n'' (два штриха для образца с собственной проводимостью) следующим образом. Для образца с собственной проводимостью, по определению, $p'' = n''$, и поэтому его проводимость будет определяться выражением

$$\sigma'' = e(n''\mu_n + p''\mu_p) = e\mu_p(1 + b)n''. \quad (10.25)$$

1) После того как был написан этот параграф, были получены более чистые образцы германия с удельным сопротивлением при комнатной температуре порядка 60 ом·см. В результате опытов Хайнса были получены также и новые значения для подвижности μ_n и μ_p . Однако в тексте книги эти данные не использованы, и, таким образом, численные примеры, рассмотренные в этом параграфе, а также в гл. 16, в настоящее время несколько устарели. — *Прим. авт.*

2) См. в § 9 гл. 12 обзор имеющихся данных. — *Прим. авт.*

Для образца с удельным сопротивлением $2,4 \text{ ом} \cdot \text{см}$, как мы увидим ниже, ρ' ничтожно мало и

$$\sigma' = e\mu_n n' = e\mu_p b n'. \quad (10.26)$$

Разделив второе выражение на первое, имеем

$$\frac{\sigma'}{\sigma''} = \frac{48}{2,4} = 20 = \frac{b n'}{(1+b) n''}, \quad (10.27)$$

откуда

$$\frac{n'}{n''} = \left(\frac{1+b}{b}\right) 20 = 33 = e^{3,5}. \quad (10.28)$$

Поскольку, согласно (10.16),

$$n = N_c e^{-(\mathcal{E}_c - \mathcal{E}_F)/kT}, \quad (10.29)$$

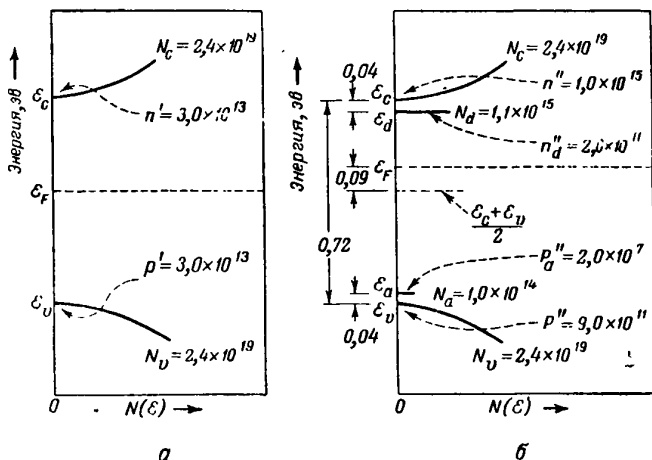
то сразу находим, что для образца с удельным сопротивлением $2,4 \text{ ом} \cdot \text{см}$ уровень \mathcal{E}_F должен лежать на $3,5 \text{ кТ}$ выше, чем у образца германия с собственной проводимостью. Так как у чистого образца германия уровень \mathcal{E}_F расположен посередине между зонами, то, следовательно, для образца германия с удельным сопротивлением $2,4 \text{ ом} \cdot \text{см}$ уровень \mathcal{E}_F должен лежать на $3,5 \text{ кТ}$, или $0,09 \text{ эв}$, выше середины запрещенной зоны. Это иллюстрирует фиг. 93, на которой показано также возможное распределение примесей. Этот же пример в более строгой форме рассмотрен в гл. 16 (см. фиг. 131).

Следует отметить, что для случая, рассмотренного на фиг. 93, $n' p' = n'' p''$. Это является прямым следствием соотношения (10.24).

График для образца кремния с удельным сопротивлением $2,4 \text{ ом} \cdot \text{см}$ выглядел бы почти так же, как график на фиг. 93, и уровень Ферми лежал бы в этом случае примерно на том же расстоянии (около $0,27 \text{ эв}$) от зоны проводимости. Однако вследствие того, что ширина запрещенной зоны у кремния составляет $1,12 \text{ эв}$, уровень Ферми лежал бы на $0,29 \text{ эв}$ выше середины запрещенной зоны, и число дырок было бы весьма малым.

При комнатной температуре алмаз является изолятором. Удельная электропроводность его равна $10^{-13} \text{ ом}^{-1} \cdot \text{см}^{-1}$ или еще меньше. Подвижность электронов и дырок в алмазе, по видимому, примерно в 10 раз меньше подвижности носителей заряда в германии, и, следовательно, концентрация носителей будет составлять только 10^8 см^{-3} . Если положить $N_c = 10^{19}$, то это значит, что $f = 10^{-17}$ и, следовательно, уровень \mathcal{E}_F находится от зоны проводимости или от зоны валентных связей на расстоянии, превышающем 1 эв . Эта разность энергий порядка 1 эв превышает предполагаемое значение энергии связи электронов на донорных уровнях ($0,35 \text{ эв}$), и поэтому можно думать, что в алмазе имеются уровни какого-то другого типа, удаленные от краев запрещенной зоны по меньшей мере на 1 эв . Дополнительным

доказательством существования таких уровней могут служить данные о поглощении алмаза в ультрафиолетовой области спектра. Граница поглощения изменяется от образца к образцу, смещаясь примерно в пределах от 7 до 4 эВ, что указывает на существование уровней, лежащих на расстоянии нескольких вольт от краев запрещенной зоны. Еще одно доказательство существования таких глубоко лежащих уровней было получено в результате изучения процесса освобождения



Фиг. 93. Теоретическое распределение квантовых состояний по энергиям; эффективные значения концентраций и концентрации электронов и дырок в германии.

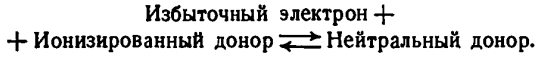
Приведенные значения носят иллюстративный характер и неточны.

электронов и дырок с уровнями захвата в опытах по исследованию электропроводности, вызванной электронной бомбардировкой [6].

Рассмотренные здесь статистические методы были применены также к тем случаям, когда распределение дырок и электронов изменяется при переходе от одной части полупроводникового образца к другой. Основные уравнения, относящиеся к этому вопросу, приведены в § 4 гл. 12. Особый интерес представляет приложение теории к изучению контакта металла с полупроводником. Этот вопрос рассматривается в работах, цитируемых в § 3 гл. 4. Кроме того, было изучено распределение потенциала в полупроводнике и его зависимость от температуры при учете влияния поверхностных состояний [7]; аналогичные вычисления были выполнены для случая p — n -переходов [8]. Обзор этих вопросов читатель может найти в статье Слэтера [9], где особое внимание уделено рассмотрению явлений на контакте металла с полупроводником.

ЗАДАЧИ

1. Условия равновесия при температурах ниже области собственной проводимости могут рассматриваться как результат „псевдохимической“ реакции:



Три входящие сюда концентрации обозначены соответственно через n , $N_d - n_d$ и n_d . В силу закона действующих масс должно было бы иметь место уравнение

$$\frac{n(N_d - n_d)}{n_d} = K.$$

Вывести это уравнение, используя приближение (10.15), и показать, что

$$K = N_c \exp \left[\frac{\mathcal{E}_d - \mathcal{E}_c}{kT} \right].$$

[Это соответствует уравнениям (22) и (23) в работе Пирсона и Бардина [5], с той только разницей, что наше значение K в два раза больше, чем у них. Разница обусловлена тем обстоятельством, что электрон, захваченный донором, может обладать либо положительным, либо отрицательным значениями спина, и поэтому правильное выражение для n_d будет иметь вид $N_d f[(\mathcal{E}_d - 2 \ln kT - \mathcal{E}_F)/kT]$; см. задачу 2 гл. 16.]

Если закон действующих масс сочетается с условием электрической нейтральности, показать, что

$$\frac{n(n + N_a)}{N_d - N_a - n} = K$$

и что

$$(n + K)(n + N_a) = N_d K.$$

Это квадратное уравнение может быть решено относительно n и найдена зависимость n от температуры T . Эту формулу Пирсон и Бардин использовали при построении теоретической кривой фиг. 13; при вычислении были взяты следующие значения: $N_d = 12 \cdot 10^{16}$, $N_a = 1,5 \cdot 10^{16}$, $\mathcal{E}_c - \mathcal{E}_d = 0,045$ эв, $m_n = 0,33$ массы свободного электрона и введен упомянутый выше множитель $1/2$ в выражение для K . Вычислить несколько значений концентрации n по этой формуле и сравнить полученные данные с данными фиг. 13.

Воспользовавшись формулами (16.69) и (16.70), вычислить значение p , показанные на фиг. 13.

2. Получить выражение для N_c или N_v , интегрируя функцию f , непосредственно по \mathbf{P} -пространству. Необходимые указания приведены в § 2 гл. 16. Какое значение имеет показатель степени $3/2$ в выражении (10.17)?

3. Две группы квантовых состояний с числом состояний N_1 и N_2 и с энергиями \mathcal{E}_1 и \mathcal{E}_2 заняты n_1 - и n_2 -электронами. Предположите, что электроны из группы N_1 могут совершать переходы в любое вакантное состояние в группе N_2 с вероятностью перехода в единицу времени, равной T_{12} . Скорость переходов из N_1 в N_2 будет равна в этом случае

$$n_1 T_{12} (N_2 - n_2).$$

Из общей статистической теории следует, что вероятность перехода из группы 2 в группу 1 есть

$$T_{21} = T_{12} e^{(\mathcal{E}_1 - \mathcal{E}_2)/kT}.$$

Показать, что если n_1 и n_2 заданы распределением Ферми—Дирака, то скорость перехода из группы состояний 1 в группу 2 в точности уравнивается скоростью перехода в обратном направлении — из группы 2 в группу 1:

$$n_2 T_{21} (N_1 - n_1).$$

Показать, что если третья группа квантовых состояний, которой соответствуют значения N_3 , \mathcal{E}_3 и n_3 , находится в равновесии с \mathcal{E}_1 , то она также должна быть заполнена в соответствии с тем же распределением Ферми—Дирака.

ЛИТЕРАТУРА

1. McDougall J., Stoner E. C., Trans. Roy. Soc., **237A**, 67 (1938).
2. Müller-Pouillet, Lehrb. d. Phys., **4** (4), 271 (1934).
3. Sommerfeld A., Zs. f. Phys., **47**, 1 (1928).
4. Nordheim L., Ann. d. Phys., **9**, 607 (1931).
5. Pearson G. L., Bardeen J., Phys. Rev., **75**, 865 (1949).
6. McKay K. G., Phys. Rev., **77**, 816 (1950).
7. Markham J. J., Miller P. H., Phys. Rev., **75**, 959 (1949).
8. Shockley W., Bell Syst. Techn. Journ., **28**, 435 (1949).
9. Slater J. C., Phys. Rev., **76**, 1592 (1949).

**МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ПРОВОДИМОСТИ
И ЭФФЕКТА ХОЛЛА¹⁾**

§ 1. ВВЕДЕНИЕ

В этой главе дан вывод основных формул для подвижности и эффекта Холла. Он разделен на две части. В первой части показано, что любое нарушение равновесного распределения электронов с течением времени исчезает вследствие „рассеяния“ электронов, и дается оценка влияния этого обстоятельства на ток. Во второй части выводятся выражения для скорости нарастания тока, вызванного электрическим и магнитным полями.

Если ограничиться линейным случаем, то скорость установления тока при действии электрического поля может быть вычислена из равновесного ненарушенного распределения, причем оказывается, что влияние электрического поля на нарушенную часть распределения зависит от величины поля квадратично. Однако при определении действия магнитного поля необходимо учитывать нарушение распределения электронов, вызванное электрическим полем. Решение для стационарного состояния находится из условия равенства скорости нарастания и скорости убывания тока. Эта задача сложнее, чем задача, рассмотренная в гл. 8, так как для различных групп электронов скорость убывания различна и при расчете должно быть выполнено соответствующее усреднение.

Рассмотрению релаксационных процессов посвящены § 2 и 3; в § 2 рассматриваются в общем виде процессы релаксации и выводятся формулы, которые могут быть применены в частных случаях.

При этом показано, что на скорость релаксационных процессов не влияет то обстоятельство, что, согласно принципу Паули, переходы электронов в уже занятые состояния невозможны. В § 2 мы опираемся на выводы тех частей гл. 17, в которых дается подробная квантовая теория переходов.

В § 3 рассматриваются два наиболее важных механизма, вызывающих переходы: рассеяние ионами, подобное резерфордскому рассеянию α -частиц атомами (впервые послужившее доказательством строения атомов из ядер и электронов), и рассеяние на тепловых

¹⁾ По вопросам современной теории электропроводности и гальваномагнитных явлений в полупроводниках и металлах см. [19—28]. — *Прим. ред.*

колебания атомов¹⁾). Первый механизм рассматривается при помощи обычной приближенной модели. Что касается вопроса о рассеянии тепловыми колебаниями, то в § 3 этот механизм рассмотрен весьма кратко и упрощенно, с целью дать лишь иллюстрацию основных принципов подхода к его решению. Более детально этот вопрос разобран в гл. 17.

В § 4 рассматривается процесс релаксации в присутствии внешних полей, оценивается влияние электрического поля на равновесное распределение и затем с помощью законов релаксации, выведенных в § 2 и 3, определяется распределение электронов в стационарном состоянии. Далее рассматривается влияние магнитного поля на стационарное состояние, установившееся в результате действия электрического поля, и определяется распределение в новом, измененном стационарном состоянии. Из распределений в стационарных состояниях находят выражения для подвижности, проводимости и эффекта Холла. Эти общие выражения подробно анализируются в ряде особо важных случаев, имеющих место в полупроводниках.

В гл. 8 были приведены формулы для собственных полупроводников, у которых дырочная и электронная составляющие тока сравнимы друг с другом. Эти формулы повторяются в § 4 настоящей главы.

В § 5 проводится сравнение между теорией и опытом.

§ 2. ВРЕМЯ РЕЛАКСАЦИИ

В гл. 8 было проведено упрощенное рассмотрение процесса рассеяния, основанное на двух предположениях:

I. Вероятность перехода электрона (или дырки) за любой малый промежуток времени dt равна dt/τ , где τ — постоянная величина.

II. Вероятность перехода в какое-либо определенное, конечное состояние не зависит от начального состояния и прямо пропорциональна вероятности того, что конечное состояние будет занято в условиях равновесного распределения.

В гл. 17 произведен анализ этих предположений на основании уравнения Шредингера. Для полупроводников с кубической решеткой (таких, как кремний и германий) при обычном предположении, что поверхности постоянной энергии обладают сферической формой, найдено, что вероятность перехода не зависит ни от направления движения, ни от времени, прошедшего после последнего перехода, но зависит, вообще говоря, лишь от скорости движения. Таким образом, вместо τ (в предположении I) нужно писать $\tau(v)$, где $v = |\mathbf{v}|$ есть скорость, или $\tau(\mathcal{E})$, где \mathcal{E} — энергия.

Показано также, что предположение II ошибочно и что при столкновении электрон может получать или отдавать только неболь-

¹⁾ По вопросу о механизме рассеяния см. также [29]. — *Приж. ред.*

шую часть своей энергии. Эти изменения энергии существенны при учете превращения электрической энергии, переданной электронам (I^2R или V^2/R). Однако эти изменения являются квадратичной функцией поля, и ими можно пренебречь при развитии линейной теории подвижности и эффекта Холла. Причина небольшого изменения энергии электрона при столкновении подробно рассматривается в § 6 гл. 17. Вкратце она заключается в том, что при одном столкновении электрон может получить или потерять лишь энергию $h\nu$, где ν — частота атомного колебания, с которым электрон взаимодействует. Такой квант энергии называется *фононом*. Электрон взаимодействует только с теми колебаниями, длины волн которых сравнимы с длиной волны электрона. Это приводит к следующей зависимости:

для электрона

$$mv = \frac{h}{\lambda}, \quad (11.1)$$

для фонона

$$\nu = \frac{c}{\lambda} \quad \text{или} \quad h\nu = \frac{ch}{\lambda} = mv c, \quad (11.2)$$

где c — скорость звука в кристалле. При комнатной температуре отношение энергии фонона к энергии электрона равно

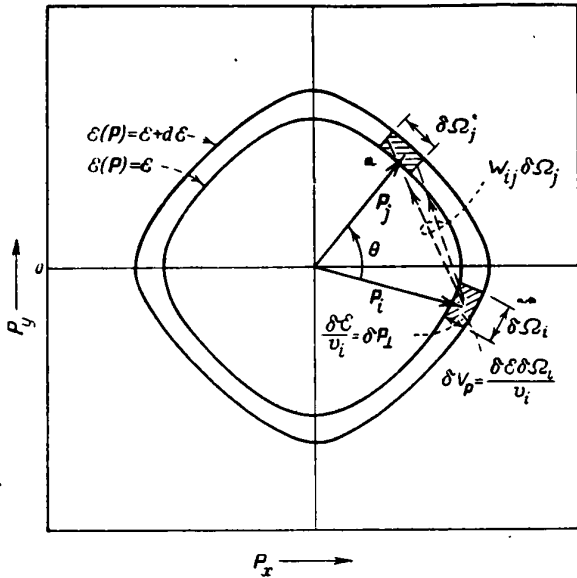
$$\frac{h\nu}{mv^2/2} = \frac{2c}{v} = \frac{10^6 \text{ см/сек}}{10^7 \text{ см/сек}} = 10^{-1}. \quad (11.3)$$

Таким образом, при одном столкновении электрон может изменить свою энергию примерно на 10%. Это изменение относительно мало. Кроме того, всегда имеют место компенсирующие друг друга процессы, вследствие чего электроны, падающие на низкие уровни энергии, заменяются другими электронами, возбуждаемыми на более высокие уровни энергии. По этим причинам применение обычного приближения теории проводимости, основанного на предположении о том, что при переходах энергия не меняется, приводит лишь к ничтожным ошибкам. Это предположение противоречит предположению II, согласно которому энергия после перехода не зависит от энергии до перехода.

Предположение II несправедливо также с точки зрения распределения направлений движения электронов после столкновения. Как будет показано в § 3, при рассеянии электронов ионизированными донорами и акцепторами может иметь место высокая степень корреляции между направлениями их движения до и после столкновения. В этом параграфе мы выведем общую формулу, учитывающую такую корреляцию, а в дальнейших параграфах применим ее к рассеянию на примесях.

Фиг. 94 иллюстрирует случай, который мы сейчас рассмотрим. На рисунке изображены две поверхности постоянной энергии в пространстве импульсов, разделенные небольшим энергетическим интервалом $\delta\epsilon$. Элемент объема слоя, заключенного между этими двумя

поверхностями, может быть определен с помощью соответствующего элемента поверхности $\delta\Omega$, имеющего размерность P^2 , т. е. квадрата импульса. (В этом параграфе мы обозначим через δ бесконечно малые



Фиг. 94. Обозначения, использованные для описания переходов из состояний, лежащих в области $\delta\mathcal{E}\delta\Omega_i$, в состояния в области $\delta\mathcal{E}\delta\Omega_j$.

В случае сферических поверхностей постоянной энергии переход характеризуется изменением угла θ .

$$\delta s_i = \frac{2V}{h^3} \delta V_p = 2P_i \delta\mathcal{E} \delta\Omega_i; \quad \delta n_i = f_i \delta s_i [= (f - f_{if}) \delta s_i].$$

элементы поверхности.) Электрон из состояния с квази-импульсом P_i может переходить в другие состояния с почти такой же энергией. В результате этих переходов электроны (точнее, состояния, занятые электронами) равномерно распределяются в слое между двумя поверхностями постоянной энергии, так что вероятность того, что любое данное состояние будет занято электроном, становится одинаковой для всех состояний и приобретает значение f , определяемое функцией распределения Ферми — Дирака (см. гл. 10) при рассматриваемом значении энергии.

Число состояний в элементе $\delta\mathcal{E}\delta\Omega_i$ вычисляется следующим образом. Объем элемента в пространстве импульсов равен произведению $d\Omega$ на толщину слоя δP_{\perp} . Очевидно,

$$\delta\mathcal{E} = |\nabla_{P\mathcal{E}}(P)|_i \delta P_{\perp} = v_i \delta P_{\perp}; \tag{11.4}$$

последнее равенство следует из формулы для групповой скорости. Следовательно, объем элемента будет $\delta P_{\perp} \delta \Omega_i = \delta \mathcal{E} \delta \Omega_i / v_i$. Плотность состояний в Р-пространстве равна $2V/h^3$, где V , как и в уравнении (5.23), — объем кристалла. Отсюда число состояний в рассматриваемом элементе объема

$$\delta s_i = \frac{2V}{h^3} \frac{\delta \mathcal{E} \delta \Omega_i}{v_i} = 2\rho_i \delta \mathcal{E} \delta \Omega_i, \quad (11.5)$$

где

$$\rho_i = \frac{V}{h^3} v_i \quad (11.6)$$

есть плотность состояний с данным спином, приходящихся на единицу энергии и на единицу поверхности.

В случае почти пустой зоны, когда ограничениями, накладываемыми принципом Паули, можно пренебречь, переходы могут быть описаны с помощью параметра W_{ij} . Этот параметр используется для определения полной вероятности перехода из состояния i во все состояния j , лежащие в области $d\Omega_j$. При таких переходах электроны могут не только получать, но и терять энергию. По причинам, указанным выше, этими изменениями энергии мы пренебрегаем и полагаем, что все состояния, в которые электрон может перейти из состояния P_i , лежат в слое толщиной $\delta \mathcal{E}$, так что мы можем говорить о вероятности перехода, приходящейся на единицу поверхности слоя. Соответственно, вероятность того, что электрон за время dt перейдет из состояния с квази-импульсом, близким к P_i , в состояние, лежащее в области $\delta \Omega_j$ вблизи P_j , будет

$$W_{ij} \delta \Omega_j dt. \quad (11.7)$$

Полное число таких переходов получится умножением этой вероятности на число электронов в элементе $d\Omega_i$, а это число равно числу состояний ds_i , умноженному на вероятность f_i того, что эти состояния заняты. Следовательно, полная скорость перехода из элемента объема $d\mathcal{E} d\Omega_i$ в элемент объема $d\mathcal{E} d\Omega_j$ будет

$$(\delta \mathcal{E} \delta \Omega_i \rightarrow \delta \mathcal{E} \delta \Omega_j) = f_i \delta s_i W_{ij} \delta \Omega_j, \quad (11.8)$$

$$= f_i 2\rho_i W_{ij} \delta \mathcal{E} \delta \Omega_i \delta \Omega_j. \quad (11.9)$$

Аналогично переходы из элемента j в элемент i даются выражением

$$(\delta \mathcal{E} \delta \Omega_j \rightarrow \delta \mathcal{E} \delta \Omega_i) = f_j 2\rho_j W_{ji} \delta \mathcal{E} \delta \Omega_j \delta \Omega_i. \quad (11.10)$$

Было показано, что эти выражения согласуются с *принципом детального равновесия* [1], требующим, чтобы в условиях равновесия прямой и обратный процессы происходили с одинаковой частотой. При тепловом равновесии $f_i = f_j = f(\mathcal{E})$, т. е. f_i и f_j совпадают с функцией распределения Ферми — Дирака, соответствующей энергии слоя \mathcal{E} ,

Кроме того, как показано в § 2 гл. 17, квантовомеханические формулы для W_{ij} и W_{ji} таковы, что

$$\rho_i W_{ij} = \rho_j W_{ji}; \quad (11.11)$$

поэтому две вычисленные ранее скорости равны друг другу. Этот результат будет использован ниже при исследовании влияния принципа Паули на релаксационный процесс в случае, когда большая часть состояний в зоне занята.

Полная вероятность перехода из состояния P_i равна сумме вероятностей переходов во все элементы $d\Omega_j$. Обозначим ее через W_i , причем

$$W_i = \int_{\Omega(\mathcal{E})} W_{ij} d\Omega_j = \frac{1}{\tau_c}. \quad (11.12)$$

Эта вероятность равна обратному значению среднего времени между столкновениями, обозначенному через τ_c , так как, согласно соображениям, приведенным в гл. 8, вероятность столкновения должна была бы быть равна dt/τ_c , что, очевидно, равно $W_i dt$. Однако ввиду того, что предположение II неправильно, τ_c не может быть связано со скоростью спадания тока, и для времени релаксации должна быть выведена новая формула.

Рассмотрим сейчас закон спадания тока, вызванного нарушением равновесного распределения. Сначала остановимся на случае, представляющем принципиальный интерес, в котором предполагается, что поверхности постоянной энергии обладают сферической формой и все направления движения эквивалентны. Позднее будет кратко рассмотрен другой случай, когда τ_c зависит от направления движения. В первом случае по соображениям симметрии очевидно, что в условиях равновесия результирующий ток равен нулю. Предположим сначала, что система подвергается некоторому возмущению. Рассмотрим ток, обусловленный группой электронов δn_i , находящихся в элементе объема $d\mathcal{E} d\Omega_i$. Плотность создаваемого ими тока равна

$$\delta I = -\frac{ev_i}{V} \delta n_i. \quad (11.13)$$

За время dt часть электронов $W_{ij} d\Omega_j dt$ рассеивается в некотором новом направлении, приобретая скорость v_j , так что изменение тока будет составлять

$$d\delta I = dt \int_{\Omega(\mathcal{E})} \frac{-e(v_j - v_i)}{V} \delta n_i W_{ij} \delta\Omega_j, \quad (11.14)$$

где $\Omega(\mathcal{E})$ — поверхность постоянной энергии. В случаях, представляющих особый интерес, имеют место такие переходы, для которых W_{ij} является функцией только угла θ между направлениями начальной и конечной скоростей, т. е. $W(\theta)$. При этих условиях можно выбрать

такую систему сферических координат, в которой направление $\theta = 0$ параллельно \mathbf{v}_i .

В новых координатах можно написать:

$$\mathbf{v}_i = i_z v_i, \quad (11.15a)$$

$$\mathbf{v}_j = (i_z \cos \theta + i_x \sin \theta \cos \varphi + i_y \sin \theta \sin \varphi) v_i, \quad (11.15б)$$

$$d\Omega_j = (m_n v_i)^2 \sin \theta d\theta d\varphi, \quad (11.15в)$$

где i_x, i_y, i_z — взаимно ортогональные единичные векторы. С помощью этих уравнений находим

$$d\delta \mathbf{I} = dt \delta n_i \left(-\frac{e}{V} \right) i_z v_i \int_0^\pi (\cos \theta - 1) W(\theta) (m_n v_i)^2 \sin \theta d\theta 2\pi, \quad (11.16)$$

так как в результате интегрирования составляющие, содержащие i_x и i_y , обращаются в нуль. Последнее выражение можно переписать в следующем виде:

$$\begin{aligned} d\delta \mathbf{I} &= -dt \delta \mathbf{I} \int (1 - \cos \theta_j) W(\theta_j) d\Omega_j = \\ &= -dt \delta \mathbf{I} W_0 (1 - \cos \theta), \end{aligned} \quad (11.17)$$

где

$$W_0 = \int W(\theta) d\Omega_j = \frac{1}{\tau_c} \quad (11.18)$$

и

$$(1 - \cos \theta) = \frac{1}{W_0} \int (1 - \cos \theta) W(\theta) d\Omega. \quad (11.19)$$

Как мы вскоре покажем, из этих уравнений следует, что ток $\delta \mathbf{I}$ постепенно убывает, причем «время релаксации» равно $\tau = \tau_c / \langle 1 - \cos \theta \rangle$, где $\langle 1 - \cos \theta \rangle$ — среднее значение $1 - \cos \theta$ для всех столкновений. Полезно отметить, что если бы переходы во все элементы поверхности сферы были одинаково вероятны, то величина $\langle \cos \theta \rangle$ равнялась бы нулю, так что τ_c и τ были бы равны. С другой стороны, если бы в результате переходов происходили только незначительные изменения в направлении движения электронов, то $\langle \cos \theta \rangle$ было бы близко к единице и время релаксации τ было бы много больше τ_c .

Для того чтобы связать полученные выше уравнения с релаксацией тока, рассмотрим ток, обусловленный произвольным распределением f_i , предполагая, однако, что $f_i \ll 1$. Электроны в каждом элементе слоя с поверхностью $\delta\Omega_i$ создадут ток $\delta \mathbf{I}_i$. За время dt эти электроны будут совершать переходы, и изменение тока, соответствующего элементу $\delta\Omega_i$, будет

$$d\delta \mathbf{I}_i = \frac{-\delta \mathbf{I}_i dt}{\tau}. \quad (11.20)$$

Полный ток $I = \sum \delta I_i$ изменяется в таком случае на величину

$$dI = \sum d\delta I_i = - \frac{(\sum \delta I_i) dt}{\tau} = - \frac{I dt}{\tau}, \quad (11.21)$$

или

$$\dot{I} = - \frac{I}{\tau}. \quad (11.22)$$

Это выражение показывает, что спадание полного тока описывается простым релаксационным уравнением, решение которого имеет вид

$$I = I_0 e^{-t/\tau}. \quad (11.23)$$

Таким образом, τ есть время релаксации для спадания тока и, как мы увидим, является именно той величиной, которая используется при вычислении подвижности. Мы будем называть его *временем релаксации*, или *средним временем между столкновениями*.

В случае тепловых колебаний вероятность перехода не зависит от θ и $\langle \cos \theta \rangle = 0$, так что $\tau = \tau_c$. При рассеянии на примесях значение $\langle \cos \theta \rangle$ приближается к единице при больших энергиях электронов, так что $\tau \gg \tau_c$; в этом случае τ_c входит в теорию только как промежуточная ступень.

О другом случае, который рассматривался различными авторами, мы упомянем здесь только для полноты изложения. В этом случае τ_c является функцией квази-импульса P_i и $\langle \cos \theta \rangle = 0$, если поверхности постоянной энергии имеют сферическую форму. При более сложных формах поверхностей постоянной энергии предполагается, что механизм рассеяния таков, что после одного столкновения электрон с равной вероятностью может попасть в любое состояние в рассматриваемом энергетическом слое. Скорость изменения тока в этом случае определяется формулой

$$\dot{I} = - \frac{\sum \delta I_i}{\tau_i}, \quad (11.24)$$

где суммирование производится по всем элементам слоя, а τ_i есть значение τ_c , соответствующее состоянию, близкому к P_i .

Роль принципа Паули. Если значительная часть состояний в энергетическом слое заполнена электронами, то некоторые переходы не могут осуществляться, так как конечные состояния уже заняты. Эти неразрешенные переходы очень легко учесть, в результате чего мы приходим к неожиданному выводу, заключающемуся в том, что на процесс релаксации не влияет то обстоятельство, что заметная часть энергетических состояний занята электронами.

Наиболее просто этот результат можно получить следующим образом. Если переход из состояния P_i в состояние P_j невозможен, так как состояние P_j занято, то в то же самое время *невозможен* и переход из состояния P_j в состояние P_i . Предположим, что мы не

принимая во внимание ограничения, налагаемых принципом Паули, и допускаем, что осуществляются переходы обоого типа; в таком случае эти переходы в среднем будут попарно взаимно компенсироваться и процесс релаксации будет протекать с той же скоростью, с какой он протекал бы и при учете ограничений, налагаемых принципом Паули. Таким образом, принцип Паули не оказывает влияния на скорость релаксации тока и, следовательно, время релаксации, соответствующее данному значению энергии, не будет зависеть от числа электронов, занимающих состояния в тонком слое вблизи поверхности постоянной энергии. Ниже мы придем к тому же выводу аналитическим путем.

Рассмотрим теперь снова переходы из элемента объема $\delta\mathcal{E}\delta\Omega_i$ в элемент объема $\delta\mathcal{E}\delta\Omega_j$. Так как доля состояний, определяемая значением функции f_j , в элементе $\delta\mathcal{E}\delta\Omega_j$ занята, то число возможных переходов уменьшается в $(1 - f_j)$ раз и мы имеем

$$(\delta\mathcal{E}\delta\Omega_i \rightarrow \delta\mathcal{E}\delta\Omega_j) = f_i(1 - f_j) 2\rho_i W_{ij} \delta\mathcal{E}\delta\Omega_i \delta\Omega_j. \quad (11.25)$$

[В § 7 гл. 17 мы исследуем эту задачу с точки зрения антисимметричных волновых функций для электронов и придем к выводу, что коэффициент $(1 - f_j)$ входит в выражение для вероятности переходов при строгом рассмотрении вопроса.] Скорость перехода из $\delta\mathcal{E}\delta\Omega_j$ в $\delta\mathcal{E}\delta\Omega_i$ определяется аналогичным выражением, содержащим множитель $f_j(1 - f_i)$. Как указывалось выше, квантовомеханическое рассмотрение задачи позволяет сформулировать принцип *детального равновесия* вследствие того, что

$$\rho_i W_{ij} = \rho_j W_{ji}, \quad (11.26)$$

так что результирующая скорость переходов из элемента $\delta\mathcal{E}\delta\Omega_i$ в элемент $\delta\mathcal{E}\delta\Omega_j$ будет

$$\begin{aligned} (\delta\mathcal{E}\delta\Omega_i \leftrightarrow \delta\mathcal{E}\delta\Omega_j) &= [f_i(1 - f_j) - f_j(1 - f_i)] 2\rho_i W_{ij} \delta\mathcal{E}\delta\Omega_i \delta\Omega_j = \\ &= (f_i - f_j) 2\rho_i W_{ij} \delta\mathcal{E}\delta\Omega_i \delta\Omega_j. \end{aligned} \quad (11.27)$$

Таким образом, результирующая скорость зависит только от разности $f_i - f_j$, а не от f_i и f_j в отдельности. Из этого, очевидно, следует, что если равновесное состояние, характеризуемое функцией f , нарушено, так что $f_i = f + f_{1i}$, то скорость переходов будет определяться только значением f_{1i} , т. е. на скорость восстановления равновесия степень заполнения состояний не влияет, а следовательно, и время релаксации остается неизменным.

В полупроводнике, в котором число электронов меняется в зависимости от содержания примеси, изменение числа электронов будет сопровождаться изменением механизма рассеяния. При таких условиях время релаксации будет, конечно, зависеть от концентрации примеси. Это обстоятельство не мешает, однако, использованию полученного нами вначале результата. При этом необходимо лишь, чтобы для данного ряда значений W_{ij} время релаксации не зависело от функции f ,

Сходство между дырками и электронами. Сходство в поведении дырок и электронов вытекает непосредственно из того факта, что скорость переходов электронов между элементами $\delta\mathcal{E}_i$ и $\delta\mathcal{E}_j$ зависит от $f_i - f_j$. Очевидно, что каждый переход электрона можно с одинаковым успехом описать как переход дырки в обратном направлении. Кроме того, часть состояний, занятых дырками, равна $1 - f_i = f_{pi}$ и $1 - f_j = f_{pj}$. Следовательно, скорость переходов дырки есть

$$\begin{aligned} (\delta\mathcal{E}_i \leftrightarrow \delta\mathcal{E}_j)_p &= -(\delta\mathcal{E}_i \leftrightarrow \delta\mathcal{E}_j) = \\ &= -(f_i - f_j) 2\rho_i W_{ij} \delta\mathcal{E}_i \delta\mathcal{E}_j = \\ &= (f_{pi} - f_{pj}) 2\rho_i W_{ij} \delta\mathcal{E}_i \delta\mathcal{E}_j. \end{aligned} \quad (11.28)$$

Это выражение формально идентично аналогичному выражению для скорости переходов электронов. Таким образом, мы приходим к выводу, что скорость релаксации в данном энергетическом слое не зависит от степени его заполнения, лишь бы только это заполнение было частичным.

В § 7 гл. 17 мы подробнее рассмотрим тот случай, когда в энергетическом слое вблизи поверхности постоянной энергии находится несколько дырок, и вычислим вероятности переходов. В результате мы приходим к уже полученному в настоящем параграфе выводу о том, что формально дырки могут рассматриваться теми же методами, как и электроны. Кроме того, будет показано, что вероятности их переходов того же порядка величины, что и для электронов.

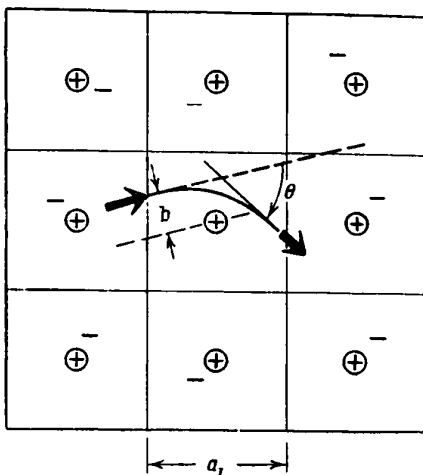
§ 3. ПЕРЕХОДЫ, ОБУСЛОВЛЕННЫЕ ТЕПЛОВЫМ ВОЗБУЖДЕНИЕМ И РАССЕЯНИЕМ НА ПРИМЕСЯХ

Рассмотрим переходы, обусловленные рассеянием как на тепловых колебаниях решетки, так и на ионизированных примесных центрах ¹⁾. Сперва рассмотрим второй случай, так как он несколько проще. Для этой цели воспользуемся классической моделью и рассмотрим электростатическую задачу о рассеянии электрона при его прохождении мимо заряженного иона. Что же касается переходов, вызванных взаимодействием электрона с тепловыми колебаниями атомов решетки, то они представляют собой типичное интерференционное явление. Для обоих случаев получим выражения для времени свободного пробега $\tau(\vartheta)$, которые будут использованы в следующем параграфе при выводе формул для подвижности и эффекта Холла.

¹⁾ Несколько более упрощенное рассмотрение вопроса, сопровождаемое многочисленными приложениями к случаю металлов, читатель найдет в статье Вайскопфа [2]. — *Прим. авт.*

Рассеяние ионами; формула Конуэлла — Вайскопфа ¹⁾. Задача рассеяния электрона заряженными донорами или акцепторами или же ионами других примесей может быть рассмотрена с помощью чисто классических методов: дырки и электроны при этом трактуются как волновые пакеты. Это позволяет нам вычислить рассеяние, пользуясь представлением об отклонении заряженных частиц, движущихся в поле заряженных ионов. Для того чтобы выразить результат в математической форме, необходимо ввести некоторые упрощающие предположения. Результаты, полученные классическим методом, являются хорошим приближением к квантовомеханическим результатам, и в случае рассеяния отдельными ионами формулы для рассеяния в обоих случаях одинаковы.

Рассмотрим сначала фиг. 95, представляющую схематическое изображение полупроводника с электропроводностью *n*-типа и концентрацией полностью ионизированных доноров N_d . На рисунке нанесена траектория электрона, испытавшего отклонение при прохождении вблизи одного из доноров. Для упрощения задачи предполагаем, что доноры расположены в узлах простой кубической решетки с постоянной a_I . Очевидно,



Фиг. 95. Упрощенная схема рассеяния электрона ионом донора.

$$N_d = \frac{1}{a_I^3}. \quad (11.29)$$

Кроме того, мы считаем, что другие электроны равномерно распределены вокруг всех остальных доноров, так что последние хорошо заэкранированы и рассматриваемый нами электрон находится в поле как бы только одного донора. В таком случае можно утверждать, что его отклонение обусловлено только этим ионом, расположенным в центре элементарной ячейки. На самом же деле электрон не будет полностью заэкранирован от действия других ионов и отклонение его до некоторой степени будет происходить под действием сил, исходящих

¹⁾ Кинетические процессы — электропроводность, термоэдс и эффект Холла в атомных полупроводниках при учете рассеяния носителей тока на ионах примеси — рассматриваются в работах [29—31]. — *Прим. ред.*

и со стороны других ячеек. При вычислении мы считаем, что прохождение сквозь каждую ячейку эквивалентно одному столкновению, и, кроме того, полагаем, что среднее время свободного пробега равно $\tau = a_I/v$. Даже и в таком виде задача все же слишком трудна, и мы введем дальнейшее упрощение, приняв, что отклонение электрона на угол θ , вызванное прохождением его через одну из ячеек, будет точно таким же, какое испытал бы электрон, двигаясь по бесконечной ветви гиперболы в присутствии того же заряда. Это приближение приводит нас в результате к некоторой переоценке величины отклонения, вызванного центральным ионом, и тем самым несколько компенсирует ошибку, связанную с тем, что мы не учли влияния других ионов.

Прежде чем перейти к математическому выводу формулы, определяющей рассеяние ионизированными примесными центрами, следует оценить порядок величины этого эффекта для типичного случая германия n -типа с концентрацией доноров $N_d = 10^{15} \text{ см}^{-3}$, значением a_I , равным 10^{-5} см , или 10^8 \AA , и диэлектрической постоянной $\kappa = 19^1$). Потенциальная энергия на расстоянии $a_I/2$ равна

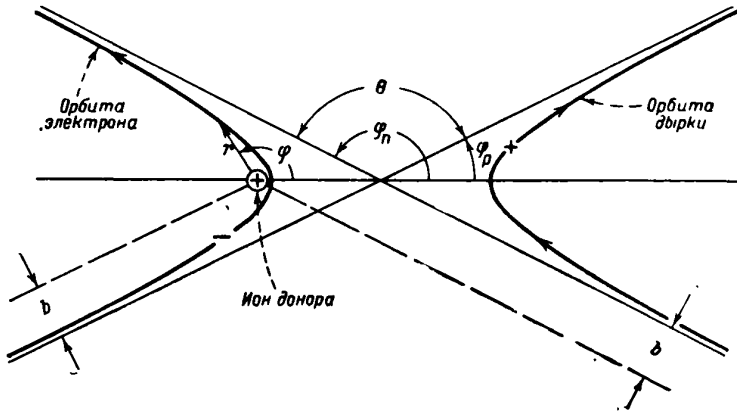
$$\begin{aligned} \mathcal{E}_I &= \frac{e^2}{\kappa (a_I/2)} = e \frac{4,8 \cdot 10^{-10}}{19 (a_I/2)_A \cdot 10^{-8}} \text{ абс. эл.-ст. ед.} = \\ &= \frac{0,76}{(a_I/2)_A} \text{ эв} = \frac{30}{(a_I/2)_A} \text{ кТ при } 300^\circ \text{ К} = \\ &= 0,06 \text{ кТ при } 300^\circ \text{ К}. \end{aligned} \quad (11.30)$$

Это вычисление показывает, что когда электрон находится в центре грани одной из кубических элементарных ячеек, энергия взаимодействия с ионом составляет всего $1/16$ его тепловой энергии kT . Таким образом, сила взаимодействия относительно слаба, и только те электроны испытывают значительные отклонения, которые проходят мимо ионов на расстояниях, меньших $1/16 a_I/2$. С другой стороны, при низких температурах, например порядка 20° К , энергия взаимодействия с ионом электрона, находящегося в углу ячейки, будет сравнима с kT , и электрон, который по какой-либо причине потеряет энергию вблизи иона, может двигаться вокруг него по эллиптической орбите. Поэтому в области таких низких температур настоящая теория будет достаточно плохим приближением. Во всяком случае, она представляется наилучшей из опубликованных ко времени написания книги, а для дальнейшего ее улучшения, вероятно, придется ожидать получения более точных экспериментальных данных. Однако существует область температур, в которой сделанное выше приближение хорошо

¹⁾ После того как этот параграф был написан, Бриггс получил более точное значение диэлектрической постоянной германия $\kappa = 16$; см. гл. 9, — *Прим. авт.*

оправдывается и в которой, в то же самое время, рассеяние на примесях является доминирующим процессом. Сравнение между теорией и опытом для этой области температур проводится в дальнейших параграфах.

Сейчас рассмотрим отклонение электрона, начавшего двигаться из бесконечности со скоростью v по прямолинейной траектории. Расстояние, на котором прошел бы электрон мимо иона, если бы он не отклонялся последним, равно b . На фиг. 96 нанесена гиперболическая траектория электрона, по которой он движется. На этом же



Фиг. 96. Положительная и отрицательная частицы с равными массами и энергиями отклоняются на один и тот же угол θ , двигаясь по гиперболическим орбитам.

рисунке нанесена также траектория дырки, начавшей движение с той же скоростью и с тем же прицельным параметром b .

Из уравнений траекторий движения электрона и дырки, как будет показано, следует, что при равенстве их эффективных масс будут равны и углы отклонения θ .

Траектория заряженной частицы, движущейся в центральном кулоновом поле положительного заряда, определяется уравнением

$$\frac{1}{r} = \pm \frac{e^2 m}{\chi J^2} + A \cos \varphi, \quad (11.31)$$

где r — радиус, φ — угол.

Знак плюс соответствует силам притяжения (электрон на фиг. 96), а знак минус — силам отталкивания (дырка). Через J обозначен момент количества движения, который может быть записан как

$$J = m r^2 \dot{\varphi} = \chi p_y - y p_x = m b v. \quad (11.32)$$

Это уравнение является следствием сохранения момента количества движения и того обстоятельства, что при движении по асимптотам величина этого момента равна mbv . Так как в формулу входит только J^2 , то знак J нас не интересует.

Задача заключается в том, чтобы вычислить зависимость угла θ от прицельного параметра b . Очевидно, что $\theta = 2\varphi_n - \pi$ и $\theta = \pi - 2\varphi_p$, где φ_n и φ_p — углы, образуемые асимптотами с осью гиперболы. Значения φ_n и φ_p определяются следующим образом: когда частицы удаляются в бесконечность, углы φ стремятся к своему максимальному значению, и мы имеем

$$0 = \frac{e^2 m}{\chi J^2} + A \cos \varphi_n, \quad 0 = -\frac{e^2 m}{\chi J^2} + A \cos \varphi_p. \quad (11.33)$$

Вблизи максимума правые части этих выражений можно разложить в ряд по $\varphi - \varphi_n$ или $\varphi - \varphi_p$, что дает

$$\frac{1}{r} = -(\varphi - \varphi_n) A \sin \varphi_n; \quad \frac{1}{r} = -(\varphi - \varphi_p) A \sin \varphi_p. \quad (11.34)$$

Эти уравнения можно решить относительно величин $(\varphi_n - \varphi)r$ и $(\varphi_p - \varphi)r$, которые, очевидно, равны прицельному расстоянию b при больших значениях r . Это дает

$$b = \frac{1}{A \sin \varphi_n}, \quad b = \frac{1}{A \sin \varphi_p}. \quad (11.35)$$

Теперь можно исключить величину A , что дает

$$\left. \begin{aligned} \operatorname{ctg} \varphi_n &= -\frac{e^2 m b}{\chi J^2} = -\frac{e^2}{\chi m v^2 b}; \\ \operatorname{ctg} \varphi_p &= \frac{e^2 m b}{\chi J^2} = \frac{e^2}{\chi m v^2 b}. \end{aligned} \right\} \quad (11.36)$$

При $v^2 b = \infty$, $\operatorname{ctg} \varphi_n$ и $\operatorname{ctg} \varphi_p$ равны 0, что соответствует $\varphi_n = \varphi_p = \pi/2$ и $\theta = 0$, и, следовательно, отклонение при этом места не имеет. Для конечных значений $v^2 b$ углы φ_n и φ_p отличаются от $\pi/2$ на одну и ту же величину, причем $\varphi_n > \pi/2$, а $\varphi_p < \pi/2$, так что углы отклонения θ также равны. Зависимость между θ и φ_n и φ_p приводит в обоих случаях к выражению

$$\operatorname{tg} \frac{\theta}{2} = \frac{e^2}{\chi m v^2 b}. \quad (11.37)$$

Для того чтобы электрон был отклонен от своего первоначального направления на угол, лежащий в интервале между θ и $\theta + d\theta$, соответствующий телесному углу $2\pi \sin^2 \theta d\theta$, его прицельный параметр должен лежать в интервале между b и $b + db$, где

$$db = \frac{e^2}{\chi m v^2} d \operatorname{ctg} \frac{\theta}{2} = -\frac{e^2}{2\chi m v^2} \operatorname{cosec}^2 \frac{\theta}{2} d\theta. \quad (11.38)$$

Иначе говоря, электрон должен пролетать внутри кольца, образованного окружностями с радиусами b и $b + db$. Площадь $2\pi b db$ этого кольца, образованного асимптотами к гиперболам, в тех случаях, когда отклонения имеют место внутри телесного угла $2\pi \sin^2 \theta db$, определяется как $\sigma(\theta) 2\pi \sin^2 \theta db$, где величина $\sigma(\theta)$ называется *дифференциальным сечением рассеяния*. Из этого определения следует, что

$$\sigma(\theta) = \frac{2\pi b db}{2\pi \sin^2 \theta db} = \left(\frac{e^2}{2\chi m v^2}\right)^2 \operatorname{cosec}^4\left(\frac{\theta}{2}\right). \quad (11.39)$$

Это и есть закон рассеяния, полученный Резефордом. Это же соотношение получается и при решении уравнения Шредингера ¹⁾.

Из формулы (11.39) следует, что при небольших углах рассеяния σ стремится к бесконечности; это отвечает бесконечно большому поперечному сечению для больших прицельных расстояний, когда отклонения ничтожно малы. При значении угла θ , приближающемся к π , что соответствует изменению направления движения на противоположное, $\sigma = r_\pi^2/16$, где r_π — значение радиус-вектора, на котором потенциальная энергия равна $m v^2/2$.

Для нашей задачи особое значение представляет величина $1 - \cos \theta$. Ее можно выразить через b и v следующим образом:

$$\begin{aligned} 1 - \cos \theta &= 1 - \cos(2\varphi_n - \pi) = 1 + \cos 2\varphi_n = 2 \cos^2 \varphi_n = \\ &= \frac{2}{1 + \operatorname{tg}^2 \varphi_n} = \frac{2}{1 + \left(\frac{\chi m v^2 b}{e^2}\right)^2}. \end{aligned} \quad (11.40)$$

Эта формула применима как к дыркам, так и к электронам.

Усредним теперь $(1 - \cos \theta)$ по всем столкновениям. Для того чтобы это сделать, припишем каждому прицельному параметру b вес, пропорциональный площади соответствующего кольца, и для простоты заменим квадрат со стороны a_I окружностью с радиусом $a_I/2$. Затем предположим, что вероятность того, что асимптота направлена так, что она лежит в интервале db , равна $(2\pi b db)/\pi (a_I/2)^2$. В таком случае значение

$$\begin{aligned} \langle 1 - \cos \theta \rangle &= \frac{4}{\pi a_I^2} \int_0^{a_I/2} (1 - \cos \theta) 2\pi b db = \\ &= 2 \left(\frac{\mathcal{E}_I}{2\mathcal{E}}\right)^2 \ln \left[1 + \left(\frac{2\mathcal{E}}{\mathcal{E}_I}\right)^2 \right], \end{aligned} \quad (11.41)$$

где

$$\mathcal{E} = \frac{m v^2}{2}, \quad \mathcal{E}_I = \frac{2e^2}{\chi a_I}. \quad (11.42)$$

¹⁾ См., например, в книге Шиффа [3] формулу (20.11). — Прим. авт.

В частности, для электрона с энергией $\mathcal{E} = kT$ [см. пример, рассмотренный в связи с формулой (11.30)]:

$$(1 - \cos \theta) = 2 \left(\frac{1}{32} \right)^2 \ln [1 + (32)^2] \approx \frac{1}{65}. \quad (11.43)$$

Таким образом, в рассматриваемом примере при прохождении состояния a_I средняя длина свободного пробега электрона будет равна $65 a_I$, или $6,5 \cdot 10^{-4}$ см. Из формулы (8.68) видно, что эта средняя длина свободного пробега примерно в 40 раз больше средней длины свободного пробега теплового электрона. Поэтому в полупроводниках с малым содержанием примесей рассеяние на примесях при комнатной температуре не имеет существенного значения.

Время релаксации τ_I при рассеянии на примесях, выраженное через τ_c и $(1 - \cos \theta)$, определяется уравнением

$$\tau_I = \frac{\tau_c}{\langle 1 - \cos \theta \rangle} = \frac{a_I}{2v} \left(\frac{2\mathcal{E}}{\mathcal{E}_I} \right)^2 \frac{1}{\ln \left[1 + \left(\frac{2\mathcal{E}}{\mathcal{E}_I} \right)^2 \right]}, \quad (11.44)$$

которое было использовано Конуэллом и Вайскопфом¹⁾ при вычислении влияния рассеяния на примесях на подвижность носителей заряда. Вследствие некоторой разницы в положенных в основу вычисления приближениях их выражение отличается от уравнения (11.44) на множитель 4π и имеет вид

$$\tau_I = \frac{2a_I}{v\pi} \frac{(2\mathcal{E}/\mathcal{E}_I)^2}{\ln [1 + (2\mathcal{E}/\mathcal{E}_I)^2]}. \quad (11.45)$$

К этой формуле мы вернемся в следующем параграфе и, усреднив ее соответствующим образом по энергиям, получим формулу Конуэлла — Вайскопфа для подвижности.

Если в полупроводнике содержатся примеси различного типа, заряд которых различается как по знаку, так и по величине, то можно воспользоваться теми же приближениями. В этом случае a_I следует вычислять, пользуясь значением суммарной концентрации рассеивающих центров, без учета знака или величины их заряда; кристалл подразделяется при этом на кубические ячейки так, чтобы в каждой ячейке находилось по одному рассеивающему центру. Рассеяние, обусловленное каждым центром, приблизительно пропорционально \mathcal{E}_I^2 , как это можно видеть из формулы для $(1 - \cos \theta)$.

¹⁾ См. [4]. Задача эта в математическом отношении аналогична задаче рассеяния электронов в ионизированном газе, которая была решена в том же приближении Чэпменом [5]. — *Прим. авт.*

Так как для однократно заряженных ионов $|\mathcal{E}_I|$ пропорционально e^2/x , то для ионов с зарядом Z_0 величина \mathcal{E}_I^2 будет пропорциональна Z^2 . Таким образом, средняя величина заряда, которую нужно подставить в формулу (11.45), будет примерно пропорциональна корню квадратному из среднего значения квадрата заряда рассеивающего иона, а число ионов, входящее в формулу, будет равно их общему числу.

Доноры и акцепторы играют весьма существенную роль при рассеянии даже в том случае, когда их заряды нейтрализованы связанными дырками и электронами.

Механизм рассеяния на донорах и акцепторах рассматривался Пирсоном и Бардином [6], которые пришли к заключению, что в кремнии при температуре около 100°K рассеяние, вызываемое ионизированным центром, играет такую же роль, как и рассеяние, вызываемое нейтральным центром¹⁾.

При более низких температурах ионизированные центры рассеяния могут быть более эффективными. Рассеяние связано с обменными эффектами между падающим и связанным электронами. Эта задача аналогична уже исследованной задаче определения сечения столкновения атомов водорода [7]. Эти важные теоретические соображения не нашли еще в настоящее время столь широкого применения к полупроводникам, как уже рассмотренная ранее теория рассеяния на других видах примеси.

Рассеяние на тепловых колебаниях решетки. Формула для потенциала деформации. Тепловые колебания атомов решетки искажают периодическое силовое поле, в котором движутся электроны. Вследствие этого волновые функции Блоха уже не являются больше точными решениями уравнения Шредингера; электрон, обладающий вначале некоторым данным импульсом P , подвергается рассеянию, совершая в результате переходы в другие состояния. Этот процесс аналогичен процессам отражения и затухания, которые могут происходить в неоднородной линии передачи, не имеющей потерь. В тех местах линии, где имеет место изменение величины полного сопротивления, происходит рассогласование полных сопротивлений, и волна отражается.

При тепловом движении атомы кристалла беспорядочно колеблются по отношению к своим соседям. Описание этой картины и исследование вопроса о том, какая ее сторона наиболее существенна для явления рассеяния электронов, дано в гл. 17. Здесь мы дадим обобщенную и упрощенную физическую картину явления рассеяния электронов, основанную на результатах этого анализа.

Прежде всего существенными для рассеяния будут те нарушения в расположении атомов, которые распространяются на расстояния

¹⁾ Более строгое решение задачи о рассеянии на нейтральных атомах примеси содержится в работе Эргинсоа [32]. — *Прим. ред.*

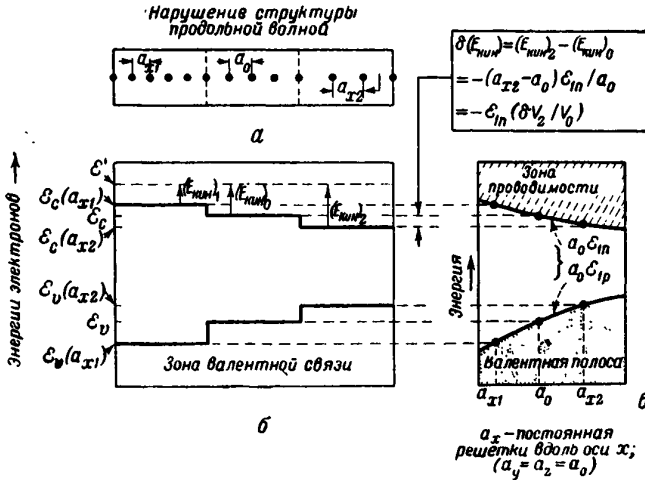
порядка $1/4 \lambda$ (λ — длина волны электрона). Этот результат хорошо известен из волновой теории.

В таких веществах, как, например, в стеклах или жидкостях, имеется большое количество нарушений порядка атомных размеров. Длина волны видимого света соответствует в них нескольким тысячам атомных расстояний, поэтому взаимодействие отдельных атомов со световой волной усредняется по областям, содержащим примерно 1000^3 атомов. При таком большом числе атомов статистические флуктуации ничтожны, вещество кажется однородным и заметного рассеяния света не происходит. С другой стороны, если оптическая прозрачность среды на протяжении длины волны изменяется лишь незначительно, то в этом случае явление протекает в основном как в линии передачи переменного сечения или в волноводе переменного сечения. В таких случаях происходит саморегулирование распространяющейся волны и отражения почти не происходит. В квантовой теории такие случаи удобнее всего исследовать, исходя из приближения Вентцеля — Крамерса — Бриллюэна, которое рассматривается в большинстве учебников по квантовой механике. Однако в том случае, когда изменение свойств материала происходит на протяжении области, размеры которой равны $1/4 \lambda$, имеет место относительно сильное отражение. При этих условиях волны, отраженные от передней и задней границ области нарушения, будут складываться, так как отражение от двух граничных поверхностей дает разность фаз в 180° , а при прохождении через область нарушения фаза запаздывает еще на 180° ; это дает в сумме разность фаз в 360° .

При точном квантовомеханическом расчете задача определения размеров нарушения, обуславливающих наибольшее рассеяние, может решаться с помощью вычислительных машин. Решение задачи сводится, по существу, к разложению деформаций в ряд Фурье. Каждая составляющая этого разложения соответствует возможной нормальной компоненте атомных колебаний (см. фиг. 52 и 55). Каждое из этих атомных колебаний возбуждено термически и при температурах, представляющих для нас наибольший интерес, характеризуется средней энергией, равной kT . Рассеяние электронов в кристалле рассматривается при этом как результат суммарного действия отдельных колебаний. Наиболее существенную роль играют те колебания, длины волн которых несколько меньше длины волны электрона. Такая продольная волна будет создавать в решетке чередующиеся области сжатия и расширения, протяженность которых несколько меньше $1/2 \lambda$, что согласуется с нашим заключением о том, что наиболее эффективны нарушения с размерами порядка $1/4 \lambda$.

Кроме того, необходимо отметить, что скорость теплового электрона составляет около 10^7 см/сек, тогда как скорость звуковой волны равна примерно $5 \cdot 10^5$ см/сек, т. е. в 20 раз меньше. Следовательно, деформация кристалла во время прохождения по нему электрона может рассматриваться как статическая. Относительно медленной

смещение картины деформации будет, однако, обуславливать „эффект Допплера“ для рассеянной электронной волны и, таким образом, изменять энергию электрона. Как указывалось в § 2, этот эффект мал и для наших целей им можно пренебречь.



Фиг. 97. Потенциалы деформации для электронов и дырок.

Изменение расстояния между атомами, вызванное распространением продольной волны, эквивалентно введению переменной потенциальной энергии.

Мы должны теперь спросить, какое влияние будет оказывать деформация решетки на прохождение электронной волны. Ответ можно получить из рассмотрения положения границ энергетических зон в зависимости от постоянной решетки, как это показано на фиг. 97. Здесь изображена одномерная модель кристалла. Распространяющаяся в нем в положительном направлении x продольная волна представлена для наглядности так, как если бы вызванная ею деформация сжатия изменялась скачкообразно; в момент прохождения деформации кристалл оказывается при этом подразделенным на три области с постоянными решетками, равными соответственно a_1 , a_0 и a_2 . Представим себе, что это состояние „заморожено“, и рассмотрим влияние его на движение электрона. На фиг. 97, б представлена зависимость положения границ энергетических зон от a_x — постоянной решетки вдоль оси x ; при этом предполагается, что $a_y = a_z = a_0$, где a_0 — значение постоянной решетки в состоянии равновесия, или истинная постоянная решетки. (Этот график отличается от фиг. 53 и 54 тем, что там переменной являлась постоянная решетки $a = a_x = a_y = a_z$.) Изменения постоянной решетки будут, таким образом, изменять структуру энергетических зон кристалла, как это показано на фиг. 97, б. Изменение

положения границ энергетической зоны фактически эквивалентно введению переменной потенциальной энергии электрона; переменную часть потенциальной энергии мы будем называть *потенциалом деформации*. Если электрон с энергией \mathcal{E}' движется по кристаллу, то его кинетическая энергия будет изменяться, как это показано на рисунке. Скачкообразные изменения кинетической энергии будут сопровождаться резким возрастанием коэффициента отражения электронных волн на переходных границах, что совершенно аналогично отражениям, имеющим место в неоднородной линии электропередачи. Совершенно такие же рассуждения применимы и к случаю дырки, и нет необходимости рассматривать его отдельно.

Рассмотрим процесс отражения электрона на примере упрощенной одномерной задачи, в которой электронная волна, обладая в области a кинетической энергией $(E_{\text{кин.}})_a = P_a^2/2m$, попадает в область b , где ее кинетическая энергия равна $(E_{\text{кин.}})_b = P_b^2/2m$; при этом электронная волна частично проходит через границу этих областей, частично же отражается. Предположим, что граница раздела лежит при $x=0$, и запишем волновые функции следующим образом:

$$\left. \begin{array}{l} \text{Падающая волна } e^{iP_a x/\hbar} \\ \text{Отраженная волна } R e^{-iP_a x/\hbar} \end{array} \right\} \text{ область } a, x < 0 \quad (11.46a)$$

$$\left. \begin{array}{l} \text{Прошедшая волна } T e^{iP_b x/\hbar} \end{array} \right\} \text{ область } b, x > 0 \quad (11.46b)$$

Плотность электронного тока (как читатель может проверить с помощью уравнений § 4 гл. 15) пропорциональна $1 - |R|^2$, так что величину $|R|^2$ нужно рассматривать как отраженную часть входящего тока. Граничные условия при $x=0$ выражают требование непрерывности функций ψ и $d\psi/dx$ и приводят, таким образом, к уравнениям

$$1 + R = T, \quad \frac{iP_a}{\hbar}(1 - R) = \frac{iP_b}{\hbar}T, \quad (11.47)$$

а отсюда к формуле

$$|R|^2 = \left(\frac{P_a - P_b}{P_a + P_b} \right)^2, \quad (11.48)$$

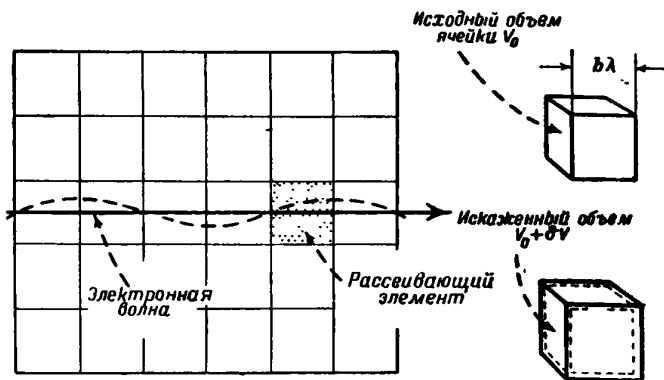
вполне аналогичной формуле $(Z_1 - Z_2)^2 / (Z_1 + Z_2)^2$ для отражения в месте соединения двух линий передачи с полными сопротивлениями Z_1 и Z_2 соответственно. [Требование непрерывности функций ψ и $P\psi$ заменяется в теории линии передачи требованием непрерывности для напряжения V и тока $I = \pm(1/Z)V$.] Для небольших изменений энергии $\delta\mathcal{E}_0$ можно написать

$$\delta\mathcal{E}_0 = \frac{P_b P_a}{m} \quad (11.49)$$

и

$$|R|^2 = \frac{(\delta P)^2}{4P^2} = \frac{m_n^2 (\delta \mathcal{E}_c)^3}{4P^4}. \quad (11.50)$$

Мы должны теперь определить значение потенциала деформации $\delta \mathcal{E}_c$, ожидаемое для области с размерами порядка $1/4 \lambda$. Как показано на фиг. 98, мы произвольно разделили кристалл на кубические ячейки



Фиг. 98. Для вычисления потенциала деформации кристалл разделен на блоки кубической формы.

Предполагается, что эти блоки расширяются и сжимаются вследствие теплового возбуждения. Каждый блок рассеивает электронную волну независимо от остальных.

$$\text{Запасенная энергия} = \frac{B(\delta V)^2}{2V_0} = \frac{1}{6} (c_{11} + 2c_{12}) \frac{(\delta V)^2}{V_0} = \frac{1}{2} kT.$$

с длиной ребра $b\lambda$ и объемом $(b\lambda)^3 = V_0$. Предположим сначала, что каждый куб сжимается и расширяется независимо от других, а в дальнейшем покажем, как может быть исправлено это предположение. Наша задача заключается, таким образом, в определении вероятности отражения волны, падающей на некоторую кубическую ячейку. Для оценки этого эффекта мы должны выразить $\delta \mathcal{E}_c$ через изменение объема. Очевидно, что в кристалле с кубической симметрией $\delta \mathcal{E}_c$ должно зависеть симметрично от a_x, a_y и a_z . Единственная линейная симметричная функция может быть записана как

$$\delta \mathcal{E}_c = \frac{(a_x + a_y + a_z - 3a_0) \mathcal{E}_{1n}}{a_0}, \quad (11.51)$$

где при небольших изменениях a величина \mathcal{E}_{1n} постоянна. Эта формула эквивалентна выражению

$$\delta \mathcal{E}_c = \mathcal{E}_{1n} \frac{\delta V}{V_0}, \quad (11.52)$$

где V_0 — первоначальный объем рассматриваемой кубической ячейки, а $V_0 + \delta V$ — объем ее после деформации. (В случае дырок мы имеем по аналогии

$$\delta \mathcal{E}_v = \mathcal{E}_{1p} \frac{\delta V}{V_0}. \quad (11.53)$$

Более полное рассмотрение обоих случаев проводится в гл. 17.)

Кубическая ячейка, изображенная на фиг. 98, обладает большим числом степеней свободы; число их в 3 раза больше числа содержащихся в ячейке атомов, и на каждую степень свободы приходится тепловая энергия, равная kT . Из всех возможных видов деформации наиболее существенное значение имеет тот, который вызывает расширение (или сжатие) ячейки в целом и, таким образом, определяет среднее значение $\delta \mathcal{E}_c$. Все остальные независимые типы деформаций будут давать в среднем расширение, равное нулю, и соответствовать таким ничтожно малым флуктуациям потенциальной энергии, что мы ими пренебрегаем. Упругая энергия, запасенная при деформации, равна

$$-\frac{1}{2} p \delta V = \frac{1}{2} B \frac{\delta V}{V_0} \delta V, \quad (11.54)$$

где $p = -B \delta V / V_0$ — давление и B — объемный модуль упругости, или величина, обратная сжимаемости. Эта энергия в среднем должна равняться $1/2 kT$, так что

$$(\delta \mathcal{E}_c)^2 = \mathcal{E}_{1n}^2 \left(\frac{\delta V}{V_0} \right)^2 = \mathcal{E}_{1n}^2 \frac{kT}{V_0 B}. \quad (11.55)$$

Отсюда для вероятности отражения получаем

$$|R|^2 = \frac{m_n^2 \mathcal{E}_{1n}^2 kT}{4P^4 V_0 B} = \frac{m_n^2 \mathcal{E}_{1n}^2 kT}{4P^4 (b\lambda)^3 B}. \quad (11.56)$$

Воспользовавшись теперь выражением для $|R|^2$, оценим среднюю длину свободного пробега l . По определению, вероятность того, что электрон будет рассеян при прохождении расстояния δx , равна $\delta x / l$. При прохождении через кубическую ячейку, вероятность рассеяния равна примерно $2|R|^2$, ввиду того что вероятности рассеяния как от передней грани ячейки, так и от ее задней грани будут равны $|R|^2$. Таким образом, имеем

$$\frac{b\lambda}{l} = 2|R|^2$$

или

$$l = \frac{b\lambda}{2|R|^2} = \frac{2(b\lambda)^4 P^4 B}{m_n^2 \mathcal{E}_{1n}^2 kT} = \frac{2b^4 h^4 B}{m_n^2 \mathcal{E}_{1n}^2 kT}; \quad (11.57)$$

последнее равенство следует из соотношения $\lambda P = h$.

Точное выражение для l , выведенное в гл. 17, имеет вид

$$l = \frac{\pi \hbar^4 c_{11}}{m_n^2 \mathcal{E}_{1n}^2 k T}, \quad (11.58)$$

где c_{11} — упругая постоянная для продольной волны, распространяющейся в направлении $[110]$; c_{11} примерно на 25% больше чем c_{11} . Отношение этих двух значений l при $B = 1/3 (c_{11} + 2c_{12}) \approx 1/2 c_{11}$ и $b = 1/4$ равно

$$\frac{l_{\text{прибл.}}}{l_{\text{точное}}} = \frac{2(2\pi b)^4 B}{\pi c_{11}} \approx 2. \quad (11.59)$$

Трудно было бы ожидать лучшего согласия с результатами точной теории при использовании нашей простой модели.

Из строгого рассмотрения вопроса следует, что рассеяние в основном сферическое, так что направление движения электрона после столкновения не зависит от направления движения его до столкновения. В этом случае, как показано в § 2, $\langle \cos \theta \rangle = 0$. В согласии с этими результатами, имеем

$$\tau = \tau_0 = \frac{l}{v} \quad \text{и} \quad \langle 1 - \cos \theta \rangle = 1. \quad (11.60)$$

Следует указать на некоторые недостатки использованной нами модели. Неправильно, например, рассматривать кубические ячейки как независимые друг от друга, так как наличие окружающего вещества приблизительно удваивает их жесткость и, следовательно, увеличивает значение l . Однако рассмотрение рассеяния при помощи модели одной кубической ячейки настолько грубо, что не стоит пытаться улучшить определение величины B без введения дальнейших уточнений. Рассмотрение этого вопроса на основе более точной модели будет проведено в гл. 17, к которой мы и отсылаем читателя, интересующегося строгой теорией.

Формула (11.58) для средней длины свободного пробега будет использована в следующем параграфе для получения формулы, определяющей подвижность [см. (11.92)]. Существенной проверкой теории может служить определение значения \mathcal{E}_{1n} и соответствующей величины для дырок \mathcal{E}_{1p} . Эти величины могут быть определены из измеренных значений подвижностей и упругих постоянных и предполагаемых значений эффективной массы. Сумма этих двух величин сопоставляется затем с величиной

$$\mathcal{E}_{1G} = \frac{d(\mathcal{E}_c - \mathcal{E}_v)}{d \ln V}, \quad (11.61)$$

где V — объем кристалла и \mathcal{E}_{1G} — изменение ширины запрещенной зоны при расширении, равном единице. В гл. 12 мы рассмотрим различные способы непосредственного определения \mathcal{E}_{1G} . Если знаки \mathcal{E}_{1n}

и \mathcal{E}_{1p} противоположны, как это показано на фиг. 97, то справедливо равенство

$$|\mathcal{E}_{1G}| = |\mathcal{E}_{1n}| + |\mathcal{E}_{1p}|. \quad (11.62)$$

Сравнение \mathcal{E}_{1G} с $|\mathcal{E}_{1n}| + |\mathcal{E}_{1p}|$, проведенное в гл. 12, показывает, что теория рассеяния, основанная на представлении о потенциалах деформации, в общих чертах, вероятно, правильна.

§ 4. СРЕДНИЕ ЗНАЧЕНИЯ ПОДВИЖНОСТЕЙ, СРЕДНЕЕ ВРЕМЯ СВОБОДНОГО ПРОБЕГА И ПОСТОЯННЫЕ ХОЛЛА ¹⁾

В результате рассуждений, проведенных в § 2 и 3, можно предположить, что существует некоторое время релаксации $\tau(\mathcal{E})$, связанное с состояниями в энергетическом слое, ограниченном поверхностями постоянной энергии \mathcal{E} и $\mathcal{E} + d\mathcal{E}$, и что это время не зависит от степени заполнения этих состояний. В согласии с представлением о времени релаксации, любой ток, возникающий в результате нарушения равновесия между состояниями, будет уменьшаться за время dt на величину $dt/\tau(\mathcal{E})$. Мы используем этот результат для вычисления стационарных токов.

Электропроводность и другие интересующие нас величины определяются путем вычисления тока, обусловленного движением электронов в частично выполненной зоне Бриллюэна, когда их равновесное распределение нарушается в результате приложения полей. Как показано в § 4 гл. 7, одно лишь магнитное поле вызывает в зоне Бриллюэна только перемещение изображающих точек вдоль поверхностей постоянной энергии. При этом вдоль данной поверхности постоянной энергии плотность занятых состояний остается постоянной и равна произведению функций $f(\mathcal{E})$ на плотность состояний; следовательно, поток, вызванный магнитным полем H , не нарушает равновесия. Электрическое поле изменяет распределение состояний, и на возникшее таким образом неравновесное распределение может уже влиять и магнитное поле (см. гл. 8).

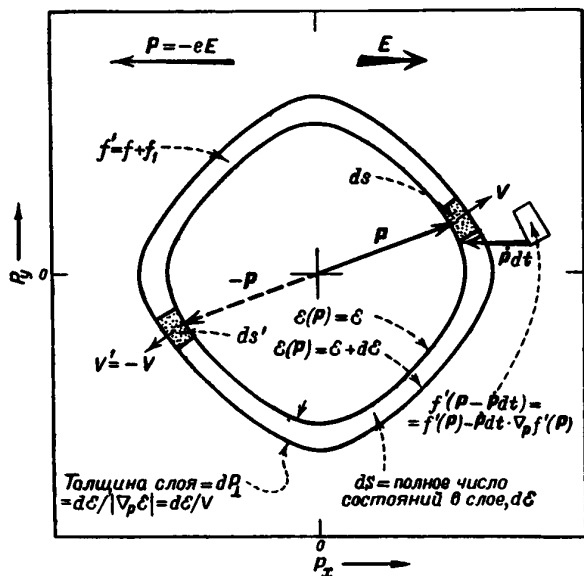
Обозначим часть квантовых состояний, занятых электронами в неравновесных условиях и имеющих квази-импульс, близкий к P , через $f'(P)$. Для малых полей функция f' будет лишь весьма незначительно отличаться от функции распределения Ферми — Дирака $f(P) = f[\mathcal{E}(P)]$, и мы можем написать

$$f'(P) = f[\mathcal{E}(P)] + f_1(P), \quad (11.63)$$

где f_1 представляет собой малое нарушение распределения, произведенное полем.

¹⁾ См. также обзор [20]. Теория хорошо проводящих полупроводников с учетом явления вырождения развита в работах [33—35]. — *Прим. ред.*

Мы используем фиг. 99 при исследовании вопроса об изменении распределения под влиянием приложенного электрического поля. В P -пространстве изображены две поверхности постоянной энергии, соответствующие энергиям \mathcal{E} и $\mathcal{E} + d\mathcal{E}$. Полное число квантовых состояний в слое между этими двумя поверхностями обозначим через dS , а число состояний в любом малом элементе слоя — через ds .



Фиг. 99. Смещение равновесия электрическим полем.

Число электронов, занимающих состояния в элементе ds , равно $f'(\mathbf{P}) ds$. Эти электроны создают ток, плотность которого, вычисленная согласно уравнению (6.1), равна

$$dI_s = \frac{-e}{V} \mathbf{v} f' ds, \quad (11.64)$$

так как каждый из них в элементе ds обладает скоростью $\mathbf{v} = \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P})$. Плотность тока, обусловленного всеми электронами, находящимися в рассматриваемом слое, равна

$$dI_s = \frac{-e}{V} \int_{\text{слой } d\mathcal{E}} \mathbf{v} f' ds. \quad (11.65)$$

В условиях равновесия этот интеграл равен нулю вследствие симметрии поверхностей постоянной энергии. Для всех зон Бриллюэна, независимо от структуры кристалла, $\mathcal{E}(\mathbf{P}) = \mathcal{E}(-\mathbf{P})$; следовательно,

поверхности постоянной энергии имеют центр симметрии и каждому элементу ds , расположенному в точке \mathbf{P} и характеризующемуся скоростью \mathbf{v} , всегда соответствует равный ему элемент ds' , расположенный в точке $-\mathbf{P}$, которому соответствует равная по величине и противоположно направленная скорость $\mathbf{v}' = -\mathbf{v}$; члены, соответствующие этим элементам, взаимно уничтожаются при интегрировании. Поэтому ток обусловлен всецело наличием члена f_1 .

Нарушение равновесия, обусловленное электрическим полем, характеризуется на рисунке вектором $\dot{\mathbf{P}} = -e\mathbf{E}$. За время dt сила $\dot{\mathbf{P}}$ вызывает такое перемещение электронов, в результате которого вместо $f'(\mathbf{P}) ds$ электронов в элементе ds имеем $f'(\mathbf{P} - \dot{\mathbf{P}} dt) ds$ электронов в объеме, смещенном из ds на $-\dot{\mathbf{P}} dt$. В результате число электронов в элементе ds изменяется (под действием поля \mathbf{E}) со скоростью, равной

$$\left[\frac{d}{dt} f' ds \right]_{\mathbf{E}} = \left[\frac{d}{dt} f' \right]_{\mathbf{E}} ds = -\dot{\mathbf{P}} \cdot \nabla_{\mathbf{P}} f'(\mathbf{P}) ds = \\ = -[\dot{\mathbf{P}} \cdot \nabla_{\mathbf{P}} f(\mathbf{P}) + \dot{\mathbf{P}} \cdot \nabla_{\mathbf{P}} f_1(\mathbf{P})] ds. \quad (11.66)$$

Второй член в этом выражении содержит произведение силы $\dot{\mathbf{P}}$, пропорциональной полю \mathbf{E} , на функцию f_1 , которая также пропорциональна \mathbf{E} и, таким образом, характеризует эффект второго порядка, которым в линейной теории мы можем пренебречь. Поэтому для определения условия стационарности распределения мы вычислим скорость возрастания тока, обусловленного членом $\nabla_{\mathbf{P}} f(\mathbf{P})$, и затем приравняем ее к скорости спада тока вследствие столкновений. Так как $f(\mathbf{P})$ зависит только от $\mathcal{E}(\mathbf{P})$, то можно записать

$$\nabla_{\mathbf{P}} f[\mathcal{E}(\mathbf{P})] = \frac{\nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P}) df(\mathcal{E})}{d\mathcal{E}} = v \frac{-1}{kT} (1-f) f, \quad (11.67)$$

где мы использовали соотношение

$$\frac{df(\mathcal{E})}{d\mathcal{E}} = \frac{d}{d\mathcal{E}} \left[\frac{\exp(\mathcal{E} - \mathcal{E}_F)}{kT} + 1 \right]^{-1} = -\frac{1}{kT} f(1-f). \quad (11.68)$$

Отсюда получаем

$$\left(\frac{d}{dt} f' \right)_{\mathbf{E}} = -\dot{\mathbf{P}} \cdot \nabla_{\mathbf{P}} f(\mathbf{P}) = \frac{(-e)}{kT} \mathbf{E} \cdot \mathbf{v} f(1-f). \quad (11.69)$$

Скорость изменения dI_s , обусловленная полем \mathbf{E} , будет, очевидно,

$$\left[\frac{d}{dt} I_s \right]_{\mathbf{E}} = \left[\frac{-e}{V} \right] v \left[\frac{d}{dt} f' \right]_{\mathbf{E}} ds,$$

или, используя предшествующие зависимости, это выражение можно представить в виде

$$\left[\frac{d}{dt} I_s \right]_{\mathbf{E}} = \frac{e^2}{kTV} v(\mathbf{E} \cdot \mathbf{v}) f(1-f) ds. \quad (11.70)$$

Необходимо указать, что в целях наглядности вектор $\dot{\mathbf{P}} dt$ изображен на фиг. 99 очень большим. На самом же деле мы использовали этот рисунок для вычисления производной, а это соответствует тому случаю, когда $\dot{\mathbf{P}} dt$ стремится к нулю. При этих условиях скорость изменения $f' ds$ определяется значениями функции f' , соответствующими двум поверхностям рассматриваемого слоя. (На рисунке приведена также формула для толщины слоя. Она понадобится нам при решении задач, помещенных в конце этого параграфа.)

То обстоятельство, что зависимость скорости изменения тока в выражении (11.70) от степени заполнения состояний выражается простым множителем $f(1-f)$, объясняется следующим образом: если значение функции f очень мало, как это, например, имеет место в высокоомном полупроводнике n -типа, то $f(1-f) ds$ приблизительно равно $f ds$, т. е. просто равно числу электронов, занимающих состояния в элементе ds ; поэтому член, характеризующий участие состояний, заключенных в элементе ds в токе, просто пропорционален числу электронов в этих состояниях. Если, однако, значительная часть состояний занята, то эффективность этих состояний уменьшается; в частности, если бы состояния в элементе ds были полностью заняты, то не произошло бы никакого изменения соответствующего им электронного тока и влияние электрического поля свелось бы просто к замещению одних электронов другими. Коэффициент, характеризующий эффективность состояния, будет, очевидно, равен $(1-f)$, так как эта величина определяет вероятность того, что ускоряемый электрон перейдет в иное состояние, а не просто заместит другой электрон. Аналогичная ситуация имеет место и для случая дырок.

Как это показано в гл. 10, функция $(1-f) = f_p$ есть вероятность того, что состояние занято дыркой; следовательно, выражения $f(1-f)$ и $(1-f_p)f_p$ равны, а формула (11.72) симметрична по отношению к дыркам и к электронам. В случае дырок коэффициент эффективности будет равен аналогично $(1-f_p) = f$. В дальнейшем, при рассмотрении интегральных выражений для тока и электропроводности, мы будем интерпретировать $f ds$ как число электронов, а $(1-f) ds = f_p ds$ как число дырок.

Скорость изменения полного тока, обусловленного электронами слоя $d\mathcal{E}$, мы получим, проинтегрировав уравнение (11.70) по всем квантовым состояниям этого слоя. В общем случае результаты носят более сложный характер: например, если электрическое поле приложено в направлении x , то изменение тока может иметь составляющую в направлении y ; это будет иметь место в том случае, если среднее значение $E_x v_x \cdot \mathbf{v}$ имеет составляющую по оси y . Общую формулу, относящуюся к таким случаям, приведем при рассмотрении задач, а здесь ограничимся исследованием сферических поверхностей постоянной энергии. В случае таких поверхностей среднее значение произведения $v_x v_y$ по поверхности сферы вследствие симметрии равно

нулю, а средние значения v_x^2 , v_y^2 и v_z^2 равны между собой также вследствие симметрии; кроме того, среднее значение $v_w^2 = v^2/3$, так как $v_x^2 + v_y^2 + v_z^2 = v^2$, где $v = |v|$. В этом случае интегрирование по слою $d\mathcal{E}$ приводит к следующему выражению для составляющей тока по оси x , возникшей под действием поля E_x :

$$\left[\frac{d}{dt} dI_{Sx} \right]_{\mathcal{E}} = E_x \frac{e^2}{kTV} \frac{v^2}{3} f(1-f) dS. \quad (11.71)$$

В стационарном состоянии возрастание тока в слое $d\mathcal{E}$, обусловленное полем E , должно точно компенсироваться его убыванием. Скорость спада тока равна просто $(dI_S)/\tau(\mathcal{E})$, так что имеем

$$0 = \left[\frac{d}{dt} dI_{Sx} \right]_{\mathcal{E}} + \left[\frac{d}{dt} dI_{Sx} \right]_{\text{спад}} = E_x \frac{e^2}{kTV} \frac{v^2}{3} f(1-f) dS - \frac{1}{\tau} dI_{Sx},$$

откуда

$$dI_{Sx} = E_x \frac{e^2}{3kTV} \tau v^2 f(1-f) dS, \quad (11.72)$$

где dI_S — ток, соответствующий состояниям dS в слое $d\mathcal{E}$, а dI_{Sx} — его составляющая по оси x . Полный ток получается интегрированием выражения (10.72) по всем значениям энергии.

Невырожденные полупроводники. В случае полупроводников элемент интегрирования для удобства можно заменить, воспользовавшись для этой цели приближенными выражениями (10.2) и (10.4).

Для полупроводника n -типа

$$\frac{f(f-1) dS}{V} \approx f \frac{dS}{V} = dn_{\mathcal{E}}. \quad (11.73a)$$

Для полупроводника p -типа

$$\frac{f(f-1) dS}{V} \approx f_p \frac{dS}{V} = dp_{\mathcal{E}}, \quad (11.73b)$$

где $dn_{\mathcal{E}}$ и $dp_{\mathcal{E}}$ соответственно концентрации электронов и дырок, находящихся в слое $d\mathcal{E}$.

Проинтегрируем выражение для тока и электропроводности в случае полупроводника. Рассмотрим случай, когда малое магнитное поле H_z приложено под прямым углом к электрическому полю E_w и пренебрежем всеми членами, пропорциональными H_z^2 . При наличии только одного электрического поля E_w слою $d\mathcal{E}$ будет соответствовать ток dI_S . Если поверхности постоянной энергии предполагаются сферическими, то направление этого тока параллельно E и его составляющая по оси x будет равна

$$dI_{Sx} = E_x \frac{e^2}{3kT} v^2 \tau(\mathcal{E}) dn_{\mathcal{E}}, \quad (11.74)$$

При включении магнитного поля вся картина распределения состояний вращается с угловой скоростью $\omega = eH_z/mc$, в результате чего возникает ток dI_{Sy} , скорость нарастания которого равна ωdI_{Sx} . Эта компонента тока, как и всякая другая компонента тока, возникающего в слое $d\mathcal{E}$, релаксирует со скоростью, равной произведению ее значения в стационарном состоянии на $1/\tau$, и, следовательно,

$$dI_{Sy} = \omega \tau dI_{Sx}. \quad (11.75)$$

(Влияние магнитного поля H_z на составляющую тока по оси x характеризуется членами, пропорциональными H_z^2 , и поэтому мы можем пренебречь этим эффектом.) Таким образом, составляющие полного тока в направлениях x и y будут равны

$$I_x = \frac{E_x e^2}{3kT} \int v^2 \tau(\mathcal{E}) dn_{\mathcal{E}} \quad (11.76a)$$

и

$$I_y = \frac{E_x e^2 \omega}{3kT} \int v^2 \tau^2(\mathcal{E}) dn_{\mathcal{E}}. \quad (11.76b)$$

Эти интегралы представляют собой суммы величин $v^2 \tau$ и $v^2 \tau^2$ для всех электронов, содержащихся в единице объема кристалла, и равны, таким образом, просто средним значениям этих величин, умноженным на концентрацию электронов n . Обозначая средние значения $v^2 \tau$ и $v^2 \tau^2$ через $\langle v^2 \tau \rangle$ и $\langle v^2 \tau^2 \rangle$, можно написать выражение, связывающее проводимость σ и подвижность μ (опуская, для краткости, значок n , так как понятно, что речь здесь идет об электронах):

$$\sigma = \frac{I_x}{E_x} = ne \frac{e \langle v^2 \tau \rangle}{3kT} = ne\mu. \quad (11.77)$$

Подвижность μ_H и среднее время свободного пробега τ_H , определенные из эффекта Холла (см. § 7 гл. 8), выражаются следующими формулами:

$$\mu_H = \frac{c\theta}{H_z} = \frac{cI_y}{H_z I_x} = \frac{\omega \langle v^2 \tau^2 \rangle}{H_z \langle v^2 \tau \rangle} = \frac{e \langle v^2 \tau^2 \rangle}{m \langle v^2 \tau \rangle}, \quad (11.78a)$$

$$\tau_H = \frac{\theta}{\omega} = \frac{m c \theta}{e H_z} = \frac{\langle v^2 \tau^2 \rangle}{\langle v^2 \tau \rangle}, \quad (11.78b)$$

где θ — так называемый „угол Холла“.

Теперь произведем необходимые усреднения выражений (11.77) и (11.78). Число электронов, скорости которых лежат в интервале $d\mathbf{v}$ вблизи скорости \mathbf{v} , определяется с помощью приближенного выражения для функции распределения f [см. уравнение (10.2)] и пропорционально элементу объема $4\pi P^2 dP$ ($\sim v^2 dv$) в пространстве импульсов. Таким образом, относительная вероятность того, что

скорость электрона лежит в интервале от v до $v + dv$, будет равна

$$e^{-mv^2/2kT} v^2 dv, \quad (11.79)$$

и среднее значение любой функции $g(v)$ есть

$$\langle g(v) \rangle = \frac{\int_0^{\infty} e^{-mv^2/2kT} g(v) v^2 dv}{\int_0^{\infty} e^{-mv^2/2kT} v^2 dv}; \quad (11.80)$$

знаменатель служит здесь для нормировки вероятностей. Если $g(v) = v^n$, то интегралы сводятся к стандартному виду, и мы имеем

$$\langle v^n \rangle = v_T^n \frac{\left(\frac{n+1}{2}\right)!}{1/2!}, \quad (11.81)$$

где v_T — скорость теплового движения (фактически средняя квадратичная скорость по направлению одной из осей)

$$v_T = \sqrt{\frac{2kT}{m}} \quad (11.82)$$

и $x! = \Gamma(x+1)$ для нецелых значений x . Численные значения интересующих нас интегралов можно найти с помощью следующих трех формул:

$$(x+1)! = (x+1)x!, \quad 1! = 1, \quad \frac{1}{2}! = \frac{\sqrt{\pi}}{2}. \quad (11.83)$$

Сначала проверим формулу (11.81), вычислив среднее значение кинетической энергии электрона $\frac{1}{2}mv^2$; из общей теоремы о равномерном распределении энергии по степеням свободы известно, что на каждую степень свободы приходится энергия, равная $\frac{1}{2}kT$, так что искомое среднее должно равняться $\frac{3}{2}kT$; на этом основании мы, очевидно, должны были бы найти, что $\langle v^2 \rangle = 3kT/m$. Из формулы (11.81), действительно, следует, что

$$\langle v^2 \rangle = v_T^2 \frac{3/2!}{1/2!} = \frac{2kT^{3/2} \cdot 1/2!}{m \cdot 1/2!} = \frac{3kT}{m}. \quad (11.84)$$

Подставляя $m \langle v^2 \rangle$ вместо $3kT$ в (11.77), получаем

$$\sigma = ne \frac{e \langle v^2 \tau \rangle}{m \langle v^2 \rangle} = ne \frac{e \tau_\mu}{m}; \quad \mu = \frac{e \langle v^2 \tau \rangle}{m \langle v^2 \rangle} = \frac{e}{m} \tau_\mu, \quad (11.85)$$

где τ_μ — среднее время свободного пробега, определяющее подвижность. Мы видим, что τ_μ — среднее взвешенное значение времени релаксации по всем рассматриваемым частицам, где каждой частице приписан статистический вес v^2 . Видно также, что τ_H равно частному

от деления среднего взвешенного значения τ^2 на среднее взвешенное значение τ . Поэтому выведенную в гл. 8 формулу для коэффициента Холла можно записать в следующем виде:

$$R_H = \frac{\tau_H}{\tau_\mu} \frac{1}{n(\pm e)c} = \frac{1}{n(\pm e)c} \frac{\langle v^2 \tau^2 \rangle \langle v^2 \rangle}{\langle v^2 \tau \rangle^2}, \quad (11.86)$$

так как из нашей формулы следует, что τ_H/τ_μ равно

$$\frac{\mu_H}{\mu} = \frac{\tau_H}{\tau_\mu} = \frac{\langle v^2 \tau^2 \rangle \langle v^2 \rangle}{\langle v^2 \tau \rangle^2} = \frac{\langle v^2 \tau^2 \rangle}{\langle v^2 \rangle} \div \left(\frac{\langle v^2 \tau \rangle}{\langle v^2 \rangle} \right)^2. \quad (11.87)$$

Вторая форма записи уравнения (11.87) показывает, что τ_H/τ_μ есть частное от деления среднего взвешенного значения τ^2 на квадрат среднего взвешенного значения τ . Так как среднее квадрата всегда больше, чем квадрат среднего, как показано в § 2 гл. 14 для другого примера, то подвижность, определенная по эффекту Холла, всегда больше подвижности дрейфа. Этот вывод, однако, основан на предположении, что поверхности постоянной энергии имеют сферическую форму и среднее время свободного пробега зависит только от энергии. В § 9 гл. 12 приведены экспериментальные данные, которые ставят под сомнение справедливость этих упрощающих предположений.

Теперь вычислим эти средние значения для нескольких различных случаев функциональной зависимости τ от энергии \mathcal{E} и затем получим формулы, относящиеся к случаю вырождения.

Формулы, относящиеся к невырожденным полупроводникам

τ — постоянная величина. В простейшем случае среднее время свободного пробега постоянно и равно τ_0 . При этом имеем

$$\tau_\mu = \tau_H = \tau_0, \quad (11.88a)$$

$$\mu = \frac{e\tau_0}{m}, \quad (11.88b)$$

$$R_H = \frac{\tau_H}{\tau_\mu n(\pm e)c} = \frac{1}{n(\pm e)c}. \quad (11.88в)$$

Тепловые колебания. Следующий случай относится к рассеянию на тепловых колебаниях. При этом средняя длина свободного пробега постоянна и

$$\tau = \frac{l}{v}. \quad (11.89)$$

Используя обозначения, принятые в выражении (11.77), согласно которому скобки $\langle \rangle$ есть значение величины, заключенной в скобки,

усредненное по всем электронам, в предположении, что всем электронам отвечает одинаковый статистический вес, находим

$$\tau_{\mu} = \frac{\langle lv \rangle}{\langle v^2 \rangle} = \frac{lv_T l}{v_T^2 \frac{3}{2}!} = \frac{l}{v_T \frac{3\sqrt{\pi}}{4}} = \frac{4l}{3\sqrt{2\pi kT/m}}, \quad (11.90a)$$

$$\mu = \frac{e\tau_{\mu}}{m} = 0,753 \frac{el}{mv_T} = \frac{4el}{3\sqrt{2\pi mkT}}, \quad (11.90б)$$

$$\frac{\mu_H}{\mu} = \frac{\tau_H}{\tau_{\mu}} = \frac{\langle l^2 \rangle \langle v^2 \rangle}{\langle lv \rangle^2} = \frac{1/2! \cdot 3/2!}{(1!)^2} = \frac{3\pi}{8} = 1,18, \quad (11.90в)$$

$$R_H = \frac{3\pi}{8n(\pm e)c} = \frac{1,18}{n(\pm e)c}. \quad (11.90г)$$

Формула для l , выведенная методом потенциалов деформации и приведенная в § 3 гл. 11, для электронов будет иметь вид

$$l = \frac{\pi \hbar^4 c_{11}}{\mathcal{E}_{1n}^2 k T m_n^2}, \quad (11.91)$$

где c_{11} — упругая постоянная для продольных акустических волн в направлении $[110]$, \mathcal{E}_{1n} — производная от \mathcal{E}_c , взятая по логарифму объема элементарной ячейки, и m_n — эффективная масса электрона. Подстановка выражения (11.91) в формулу для μ дает

$$\mu_n = \frac{2\sqrt{2\pi}}{3} \frac{\hbar^4 c_{11}}{\mathcal{E}_{1n}^2 m_n^{5/2} k^{3/2}} T^{-3/2}. \quad (11.92a)$$

Выражая μ в $см^2/в \cdot сек$, c_{11} в $дин/см^2$, \mathcal{E}_{1n} в эв и T в °К, получаем

$$\mu = \frac{3,2 \cdot 10^{-5} c_{11} (m/m_n)^{1/2}}{\mathcal{E}_{1n}^2 T^{3/2}}. \quad (11.92б)$$

Аналогичные выражения получаются и для дырок.

Рассеяние на примесях; формула Конуэлла — Вайскопфа. В этом случае зависимость τ от v определяется сложным выражением (11.45):

$$\tau_I = \frac{(2a_I/\pi v) (2\mathcal{E}/\mathcal{E}_I)^2}{\ln [1 + (2\mathcal{E}/\mathcal{E}_I)^2]}, \quad (11.93)$$

где

$$a_I = N_I^{-1/2}, \quad \mathcal{E}_I = \frac{2e^2}{\kappa a_I}, \quad \mathcal{E} = \frac{m_n v^2}{2} \quad (11.94)$$

(m_n — эффективная масса электрона; N_I — концентрация рассеивающих примесных центров; κ — диэлектрическая постоянная).

Выражение (11.93) можно переписать в виде

$$\tau_I = \frac{A \mathcal{E}^{3/2}}{\ln [1 + (2\mathcal{E}/\mathcal{E}_I)^2]}. \quad (11.95)$$

Для типичного высокоомного образца германия при комнатной температуре величина $(2\mathcal{E}/\mathcal{E}_I)$ равна примерно 30, так что знаменатель этого выражения равен 6,8. Численное значение выражения, стоящего в знаменателе, мало меняется в зависимости от \mathcal{E} . Изменение энергии \mathcal{E} в $e^{\pm 1}$ раз изменяет знаменатель в пределах от 4,8 до 8,8. По этой причине не будет большой ошибки, если заменить логарифмический член постоянной величиной, соответствующей его значению в наиболее существенной части области при усреднении $v^2\tau$. Но среднее значение $v^2\tau$ есть то же самое, что и среднее значение $2m_n\mathcal{E}\tau$, а число электронов в интервале $d\mathcal{E}$ пропорционально $\mathcal{E}^{1/2} \exp(-\mathcal{E}/kT) d\mathcal{E}$. Таким образом, при вычислении $\langle v^2\tau \rangle$ интегрируемое выражение содержит множители

$$(\mathcal{E}^{3/2} \text{ от } \tau_I) (\mathcal{E} \text{ от } v^2) \mathcal{E}^{1/2} \cdot e^{-\mathcal{E}/kT} = \mathcal{E}^3 e^{-\mathcal{E}/kT}. \quad (11.96)$$

Последнее выражение имеет максимум при $\mathcal{E} = 3kT$. Следовательно, в логарифмическом члене можно заменить \mathcal{E} на $3kT$ и записать τ_I в виде

$$\tau_I = Bv^3, \quad \text{где } B = \frac{m_n^2 x^2 / 2\pi N_I e^4}{\ln [1 + (6kT/\mathcal{E}_I)^2]}. \quad (11.97)$$

Так как τ связано теперь с v степенной зависимостью, то применимы уже ранее выведенные нами формулы и, следовательно, ¹⁾

$$\tau_\mu = \frac{B \langle v^5 \rangle}{\langle v^2 \rangle} = Bv_T^3 \frac{3!}{3/2!} = \frac{8}{\sqrt{\pi}} Bv_T^3, \quad (11.98a)$$

$$\mu = \frac{e}{m_n} \tau_\mu = \frac{8}{\pi} \sqrt{\frac{2}{\pi}} \frac{x^2 (kT)^{3/2}}{N_I e^3 \cdot m_n^{1/2}}, \quad (11.98b)$$

$$\mu = \frac{e}{m_n} \tau_\mu = \frac{8}{\ln \left[1 + \left(\frac{3xkT}{e^2 N_I^{1/2}} \right)^2 \right]}, \quad (11.98c)$$

$$\frac{\mu_H}{\mu} = \frac{\tau_H}{\tau_\mu} = \frac{B^2 \langle v^8 \rangle \langle v^2 \rangle}{B^2 \langle v^5 \rangle^2} = \frac{9/2! 3/2!}{3! 3!} = \frac{315\pi}{512} = 1,93, \quad (11.98d)$$

$$R_H = \frac{1,93}{n(\pm e)c}. \quad (11.98e)$$

Собственная проводимость и область, близкая к собственной проводимости. В гл. 8 были выведены выражения для электропроводности и постоянной Холла для случая, когда в полупроводнике одновременно присутствуют как дырки, так и электроны.

¹⁾ Формула для μ получена Конуэллом и Вайскопфом [4]. — *Прим. авт.*

Эти выражения содержат μ и θ . Перепишем полученные там формулы в обозначениях настоящей главы:

$$\sigma = ne\mu_n + pe\mu_p, \quad (11.99)$$

$$R_H = \frac{\theta}{\sigma H} = \frac{-(ne\mu_n e\tau_{Hn} H/m_n c) + (pe\mu_p e\tau_{Hp} H/m_p c)}{\sigma^2 H} = \\ = \frac{-n\tau_{\mu n}\tau_{Hn}/m_n^2 + p\tau_{\mu p}\tau_{Hp}/m_p^2}{ec(n\tau_{\mu n}/m_n + p\tau_{\mu p}/m_p)^2}. \quad (11.100)$$

Если отношение τ_H/τ_μ имеет одинаковое значение как для дырок, так и для электронов, то последнее выражение принимает вид

$$R_H = \frac{\tau_H}{\tau_\mu} \frac{-nb^2 + p}{(nb + p)^2 ce}, \quad (11.101)$$

где $b = \mu_n/\mu_p$. При рассеянии на тепловых колебаниях решетки $\tau_H/\tau_\mu = 3\pi/8$ [см. (8.92)].

Общий случай. Аналогичные выражения можно получить для вырожденного полупроводника, но, как указывалось раньше, эффективность электронов в этом случае характеризуется множителем $(1-f) = f_p$, а эффективность дырок — множителем $(1-f_p) = f$. Соответствующие общие выражения, выведенные на основании (11.72) и слегка обобщенных выражений (11.76а) и (11.76б), будут для полупроводника n -типа иметь вид

$$\sigma = ne \frac{e\langle f_p v^2 \tau \rangle}{3kT} = ne \frac{e\tau_\mu}{m} \quad (11.102)$$

и

$$\tau_H = \frac{\langle f_p v^2 \tau^2 \rangle}{\langle f_p v^2 \tau \rangle}, \quad (11.103)$$

где при усреднении каждому электрону проводимости снова приписывается одинаковый статистический вес. Из выражений (11.102) и (11.103) следует, что в случае вырождения τ_μ есть взвешенное среднее значение τ по всем электронам, причем каждый электрон обладает статистическим весом, равным $v^2 f_p$. Это имеет место в случае сферических поверхностей постоянной энергии, так как при этом $3kT = m \langle v^2 f_p \rangle$ (см. задачи) и можно написать

$$\tau_\mu = \frac{\langle \tau v^2 f_p \rangle}{\langle v^2 f_p \rangle}, \quad \frac{\tau_H}{\tau_\mu} = \frac{\langle \tau^2 v^2 f_p \rangle \langle v^2 f_p \rangle}{\langle \tau v^2 f_p \rangle^2}. \quad (11.104)$$

Необходимо заметить, что (11.102) является общим выражением, в то время как (11.104) относится к частному случаю. Совершенно аналогичные выражения получаются и для дырок при замене f_p на f и усреднении по всем дыркам.

Простые металлы и вырожденные полупроводники. В случае сильно вырожденного полупроводника или металла большинство состояний в интересующей нас зоне или заполнено [чему соответствует $f=1$ и $f(1-f)=0$] или же не заполнено [т. е. $f=0$, и тогда снова $f(1-f)=0$]. Таким образом, при суммировании по энергиям наиболее существенную роль играет сравнительно узкая область значений энергии, ширина которой порядка $2kT$ и центр которой отвечает $\mathcal{E}=\mathcal{E}_F$; при этом $f(1-f)\approx(1/2)^2=0,25$. В этой узкой области остальные функции от \mathcal{E} можно считать постоянными. В соответствии с этим выражение для σ , полученное путем интегрирования и деления на E_x выражения (11.70), будет иметь вид

$$\sigma = \frac{e^2}{3kTV} \tau v^2 \int f(1-f) dS; \quad (11.105)$$

здесь v^2 можно было вынести за знак интеграла, так как поверхности постоянной энергии предполагаются сферическими. Этот интеграл может быть непосредственно преобразован к интегралу по энергии при помощи выражения (9.9):

$$dS = \frac{4\pi}{h^3} (2m)^{3/2} (\mathcal{E} - \mathcal{E}_0)^{1/2} d\mathcal{E}. \quad (11.106)$$

Однако проще воспользоваться для этой цели другим методом вычисления. Пусть $\mathcal{E}_1 = \mathcal{E} - \mathcal{E}_0^1$, тогда в зоне Бриллюэна число состояний S с энергией, меньшей чем \mathcal{E}_1 , будет пропорционально P^3 или $\mathcal{E}_1^{3/2}$. Вследствие этого для наивысших занятых состояний, т. е. для состояний с энергией $\mathcal{E}_1 = \mathcal{E}_F - \mathcal{E}_0$, имеем

$$\frac{dS}{d\mathcal{E}_1} = \frac{3}{2} \frac{S}{\mathcal{E}_1} = \frac{3}{2} \frac{nV}{\mathcal{E}_1}, \quad \text{или} \quad dS = \frac{3nV}{2\mathcal{E}_1} d\mathcal{E}_1, \quad (11.107)$$

так как полное число занятых состояний равно концентрации электронов n , умноженной на V . Интегрирование выражения $f(1-f)d\mathcal{E}$ осуществляется при помощи формулы (11.68)

$$\int \frac{1}{kT} d\mathcal{E}_1 f(1-f) = \int -df = 1 \quad (11.108)$$

в том случае, если область интегрирования содержит уровень Ферми \mathcal{E}_F , т. е. функция f меняется в этой области от единицы до нуля. Таким образом, для σ имеем

$$\sigma = \frac{e^2}{3V} \tau v^2 \frac{3nV}{2\mathcal{E}_1} = ne \frac{e\tau}{m}, \quad \mu = \frac{e\tau}{m}. \quad (11.109)$$

Второе из этих равенств является следствием того, что $\mathcal{E}_1 = mv^2/2$. Это выражение для σ уже было получено в гл. 8 на основании рассмотрения единичного времени свободного пробега. Физический смысл

¹⁾ В этих преобразованиях не следует путать \mathcal{E}_1 с \mathcal{E} в случае рассеяния на тепловых колебаниях атомов. — *Прим. авт.*

этого простого выражения для σ заключается в том, что полем ускоряются все электроны, так что $\dot{\mathbf{i}} = ne^2\mathbf{E}/m$, как уже указывалось в гл. 5, в то время как релаксация тока вызывается только электронами самого верхнего энергетического слоя зоны, так как все внутренние слои зоны заполнены. Таким образом, в выражение для проводимости входит концентрация n всех электронов, но время релаксации τ только тех из них, которые обладают наивысшей энергией. Составляющая тока в поперечном направлении, обусловленная магнитным полем, также убывает с постоянной времени τ , и, следовательно, $\tau = \tau_H$. Так как τ_μ и τ_H равны, то формула для постоянной Холла примет простой вид

$$R_H = -\frac{1}{nec}. \quad (11.110)$$

Джонсон и Ларк-Горовитц [8] обобщили формулу Конуэлла — Вайскопфа на случай вырождения¹⁾. При этом мы имеем дело только с одним значением энергии, и поэтому нет необходимости усреднять τ_I по энергиям. Рассматриваемая энергия \mathcal{E} соответствует наивысшему занятому состоянию вырожденного электронного газа и вычисляется следующим образом:

$$n = \frac{2}{h^3} \frac{4\pi}{3} P^3, \quad P = \frac{h}{2} \left(\frac{3n}{\pi} \right)^{1/3}, \quad \mathcal{E} = \frac{h^2}{8m} \left(\frac{3n}{\pi} \right)^{2/3}. \quad (11.111)$$

Выражение для времени релаксации [см. также (11.45)] имеет вид

$$\tau_I = \frac{4}{\pi} \frac{\tau_c}{\langle 1 - \cos \theta \rangle} = \frac{4}{\pi} \frac{a_I}{v} \frac{(2\mathcal{E}/\mathcal{E}_I)^2}{2 \ln [1 + (2\mathcal{E}/\mathcal{E}_I)^2]}. \quad (11.112)$$

Величина $(2\mathcal{E}/\mathcal{E}_I)$ для кремния и германия соответственно равна²⁾

$$\begin{aligned} \frac{2\mathcal{E}}{\mathcal{E}_I} &= 2\mathcal{E} \frac{\kappa a_I}{2e^2} = \frac{h^2}{8m} \left(\frac{3n}{\pi} \right)^{2/3} \frac{\kappa}{e^2 n^{1/3}} = \frac{(3/\pi)^{2/3}}{8} \frac{\kappa h^2 n^{1/3}}{me^2}, \\ &= 3,4 \cdot 10^{-7} n^{1/3} \text{ (для Si)}, \\ &= 5,0 \cdot 10^{-7} n^{1/3} \text{ (для Ge)}. \end{aligned} \quad (11.113)$$

Здесь m — масса электрона, а κ — диэлектрическая проницаемость, равная соответственно 13 и 19³⁾.

В предыдущем уравнении мы считали, что $n = n_I$, так как в „насыщенном“ вырожденном образце число электронов и доноров должно быть почти одинаковым. Из этого уравнения следует, что

¹⁾ Эти авторы приводят также общее выражение для интеграла, относящееся к промежуточным случаям. — *Прим. авт.*

²⁾ $2\mathcal{E}/\mathcal{E}_I$ является такой же функцией от n , как и u в уравнении (62) в работе Пирсона и Бардина [6]. — *Прим. авт.*

³⁾ Более новые значения для κ приведены в гл. 9. — *Прим. авт.*

$2\mathcal{E}/\mathcal{E}_I = 1$ при $n \approx 3 \cdot 10^{19}$. Исследуя предельную форму выражения, определяющего усредненное значение времени свободного пробега. Джонсон и Ларк-Горовитц пришли к заключению, что отношение $2\mathcal{E}/\mathcal{E}_I$ будет мало¹⁾. На этом основании они упростили полученное ими выражение, разложив логарифмический член в ряд, так что член $(2\mathcal{E}/\mathcal{E}_I)^2$ в числителе при этом сократился; это эквивалентно тому, что величина $\langle 1 - \cos \theta \rangle$ принята равной 2. Физический смысл этого результата заключается в том, что при $\mathcal{E} \ll \mathcal{E}_I$ электрон будет двигаться в поле иона почти по параболической орбите и выбрасываться из ячейки в обратном направлении, так что $\theta = 180^\circ$ или $\cos \theta = -1$. В этом случае формула для подвижности принимает вид

$$\mu = \frac{e}{m} \tau_I = \frac{e2a_I}{m\pi v} = \frac{2en^{-1/3}}{\pi P} = \frac{4}{\pi} \left[\frac{\pi}{3} \right]^{1/3} \frac{e}{\hbar n^{2/3}}, \quad (11.114)$$

а для удельного сопротивления

$$\rho = \frac{1}{ne\mu} = \frac{\pi}{4} \left[\frac{3}{\pi} \right]^{1/3} \frac{\hbar}{e^2} n^{-1/3}. \quad (11.115)$$

Значение, которое получили Джонсон и Ларк-Горовитц, в π раз меньше вследствие различия в принятых приближениях. В практических единицах формула (11.115) дает

$$\rho = 6270 \cdot n^{-1/3} \text{ ом} \cdot \text{см}, \quad (11.116)$$

где n — концентрация (в см^{-3}).

§ 5. СРАВНЕНИЕ С ОПЫТОМ

Наиболее обстоятельное сравнение теории с опытом для тех типов полупроводников, теория которых была рассмотрена в этой главе, содержится в статье Пирсона и Бардина, на которую мы неоднократно уже ссылались²⁾. Образцы, изученные ими, могут быть грубо разделены на два класса: вырожденные и невырожденные. Это разделение можно произвести для полупроводника n -типа, согласно приведенным ниже условиям [см. уравнение (10.21)]:

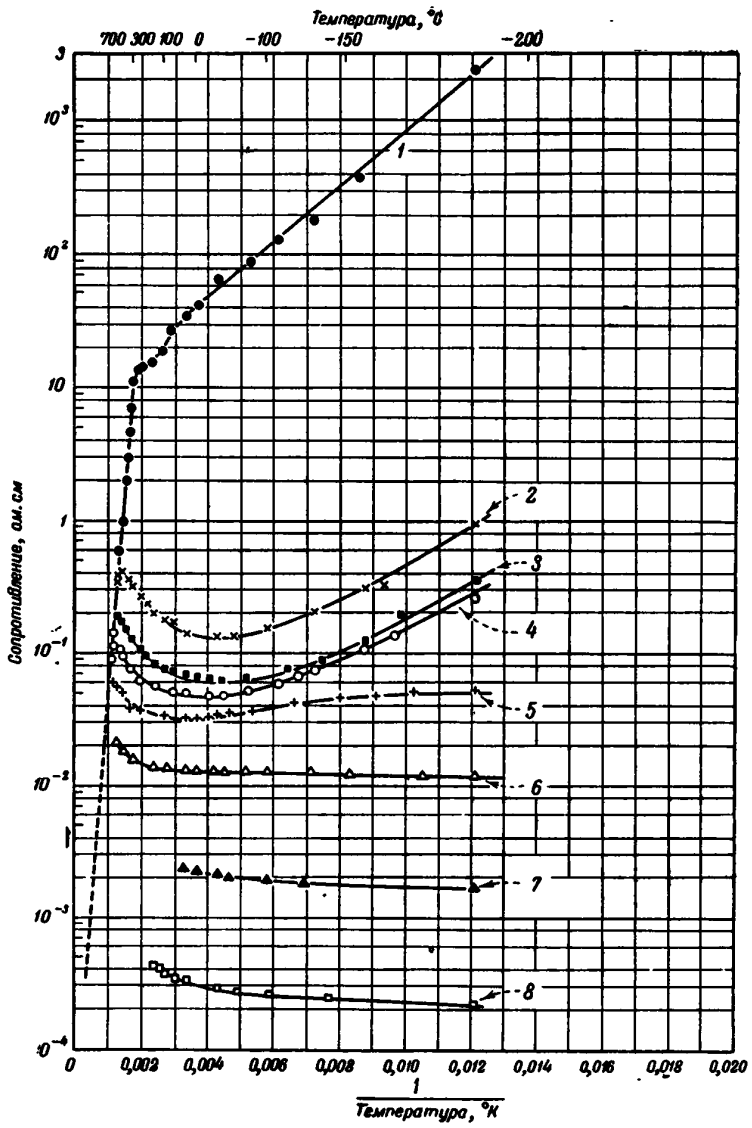
$$T = \left[\frac{3}{\pi} \right]^{2/3} \frac{\hbar^2}{8km_n} n^{2/3} = 4,2 \cdot 10^{-11} \left[\frac{m}{m_n} \right] n^{2/3} \text{ }^\circ\text{K}, \quad (11.117a)$$

$$kT = 21,6 \cdot 10^{-16} \left(\frac{m}{m_n} \right) n^{2/3} \text{ эв}. \quad (11.117b)$$

Такие же уравнения можно написать и для полупроводника p -типа.

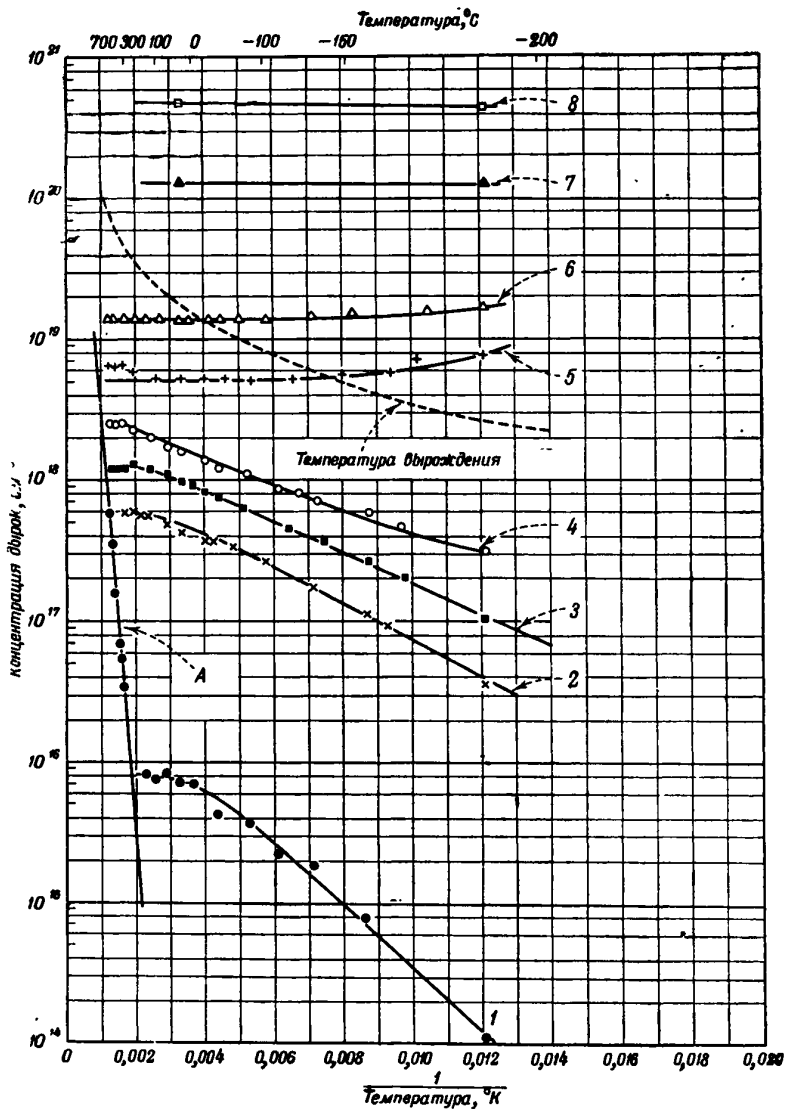
¹⁾ Как указывалось раньше, заключение это, повидимому, несправедливо, пока n не будет меньше $3 \cdot 10^{19}$. — *Прим. авт.*

²⁾ Широкое изучение свойств германия в рассмотренном здесь направлении было выполнено Ларк-Горовитцем с сотрудиниками; краткое сообщение о результатах их исследований приведено в работе [9]. Дополнительные исследования германия опубликованы в работе [10]. Опубликован ряд работ, имеющих общий интерес, по изучению свойств германия и кремния [8, 11—16]. Значительная часть экспериментального материала, опубликованного в указанных выше работах, содержится в обзоре [17]. — *Прим. авт.*



Фиг. 100. Удельное сопротивление сплавов кремния с бором в зависимости от температуры.

Концентрация атомов бора в расплаве равна соответственно в (см⁻³):
 Образец 1 — очень чистый кремний. Образец 2 6,7 · 10¹⁹
 " 3 1,3 · 10¹⁷ " 6 1,3 · 10¹⁹
 " 4 2,7 · 10¹⁹ " 7 1,3 · 10¹⁹
 " 8 1,3 · 10²¹



Фиг. 101. Зависимость концентрации дырок в сплавах кремния с бором от температуры.

Концентрация атомов бора в расплаве равна соответственно (в см^{-3}):
 Образец 1 — очень чистый кремний. Образец 5 $6,7 \cdot 10^{18}$
 : 2 $6,7 \cdot 10^{17}$: 6 $1,3 \cdot 10^{18}$
 : 3 $1,3 \cdot 10^{19}$: 7 $1,3 \cdot 10^{20}$
 : 4 $2,7 \cdot 10^{19}$: 8 $1,3 \cdot 10^{21}$

В случае *вырожденных* полупроводников скорость электрона, входящая в расчетные формулы, соответствует наивысшему занятому электронному уровню и с температурой заметно не изменяется. В этих случаях концентрация примесей столь велика, что рассеяние на них преобладает над рассеянием на тепловых колебаниях решетки. В результате удельное сопротивление и постоянная Холла практически постоянны в широкой области температур. Образцы 6, 7 и 8 сплавов кремния, представленные на фиг. 100 и 101 (заимствованных из работы Пирсона и Бардина), являются как раз примерами такого поведения. Электропроводность этих образцов — *p*-типа; они содержат примесь бора, концентрация которого для каждого образца указана в подписи к рисункам. То обстоятельство, что содержание добавленного бора в образце 8 превосходит число носителей, объясняется (на основании других опытов) выпадением второй, богатой бором, фазы на границах зерен сплава, что и приводит, таким образом, к уменьшению эффективной концентрации бора. При более высоких температурах образцы кремния становятся невырожденными; нанесенная на фиг. 101 пунктирная кривая зависимости температуры вырождения $T_{\text{вырожд.}}$ от концентрации носителей заряда является демаркационной линией, разделяющей область вырождения от невырожденной области. При вычислении этой зависимости отношение m/m_n для упрощения было принято равным единице. (При еще более высоких температурах образцы будут уже характеризоваться собственной проводимостью.) Подвижность носителей заряда в образцах 6, 7 и 8 равна соответственно 31, 30 и 54 $\text{см}^2/\text{в} \cdot \text{сек}$. Это увеличение подвижности с возрастанием содержания примеси противоречит теории даже в том случае, если мы воспользуемся для этой области температур формулой Конуэлла — Вайскопфа, а не приближенной формулой Джонсона и Ларк-Горовитца. Пирсон и Бардин исследовали также формулу, выведенную Моттом для сплавов. Хотя все эти формулы дают в общем удовлетворительное согласие с опытом в отношении порядка величины подвижностей, однако ни одна из них, по видимому, не дает правильной в деталях зависимости подвижности от концентрации n_I . Образец 5, для которого предполагается, что вырождение наступает ниже -150°C , ведет себя аномально и не имеет постоянной подвижности при низких температурах.

Джонсон и Ларк-Горовитц сравнили также свою формулу $\rho = 6270 n^{-1/2}$ с данными для некоторых образцов германия с удельным сопротивлением, лежащим в пределах от $2,9 \cdot 10^{-8}$ до $5,1 \cdot 10^{-8} \text{ ом} \cdot \text{см}$, и получили хорошее согласие расчета с опытом.

Применение метода Конуэлла — Вайскопфа к вырожденным образцам встречает некоторые трудности. Длина волны самых быстрых электронов для вырожденного электронного газа определяется формулой

$$\lambda = \frac{h}{p} = 2 \left[\frac{\pi}{3n} \right]^{1/2} = 2 \left[\frac{\pi}{3} \right]^{1/2} a_I. \quad (11.118)$$

Поэтому размеры элементарного куба, рассмотренного в качестве модели рассеивающего центра при вычислении рассеяния, составляют только $1/2 \lambda$, а его „радиус“ — около $1/4 \lambda$. При таком соотношении между длиной волны и расстоянием между примесными центрами, приближение с использованием формул Резерфорда становится совершенно непригодным при классическом рассмотрении и очень сомнительным при квантовомеханическом, так как отдельные рассеивающие центры находятся в этом случае очень близко друг от друга. Таким образом, в дополнение к задачам определения эффективного силового поля, в котором движутся электроны, возникнут также задачи, связанные с когерентностью волн при рассеянии.

В случае *невырожденных* образцов следует применять формулы, выведенные в последнем параграфе. Однако в некоторых областях температур оба процесса — рассеяние примесями и тепловыми колебаниями — могут играть значительную роль, и поэтому необходимо рассматривать их совместное влияние. Для электронов с данной энергией оба эти релаксационные процесса аддитивны, и, следовательно, нарушенное равновесие будет восстанавливаться за время dt на долю

$$\frac{dt}{\tau_I} + \frac{dt}{\tau_T} = \left[\frac{1}{\tau_I} + \frac{1}{\tau_T} \right] dt = \frac{1}{\tau} dt, \quad (11.119)$$

где

$$\tau = \frac{\tau_I \tau_T}{\tau_I + \tau_T}. \quad (11.120)$$

Эта величина, следовательно, должна заменить τ при усреднении, используемом для вычисления τ_μ и τ_H . Если τ_μ и τ_L одинаковым образом зависят от ν , то для τ_μ получим выражение

$$\tau_\mu = \frac{\tau_{\mu I} \cdot \tau_{\mu T}}{\tau_{\mu I} + \tau_{\mu T}}. \quad (11.121)$$

Аналогичное выражение можно написать для τ_H . Ошибки, связанные с различным характером зависимости τ_I и τ_T от ν , могут достигать 50% и больше. Мы здесь не будем исследовать этот вопрос подробно и ограничимся приближением, обычно применяемым при сравнении теории с опытом, и примем, что рассматриваемые времена релаксации аддитивны.

Таким образом, приближенно напишем

$$\frac{1}{\tau_\mu} = \frac{1}{\tau_{\mu I}} + \frac{1}{\tau_{\mu T}}, \quad (11.122)$$

и отсюда

$$\frac{1}{\mu} = \frac{1}{\mu_I} + \frac{1}{\mu_T}, \quad (11.123)$$

где μ_I и μ_T — подвижности, соответствующие рассеянию на примесях

и на тепловых колебаниях решетки в отдельности. Формула для μ_I имеет вид

$$\mu_I = \frac{\frac{8}{\pi} \sqrt{\frac{2}{\pi}} \frac{x^2 (kT)^{3/2}}{N_I e^3 m_n^{1/2}}}{\ln \left[1 + \left(\frac{3xkT}{e^2 N_I^{1/2}} \right)^2 \right]}, \quad (11.124)$$

что дает для кремния, диэлектрическая проницаемость которого $\kappa = 13$,

$$\mu = \frac{5,5 \cdot 10^{17} T^{3/2}}{N_I \ln(1 + x^2)} \text{ см}^2/\text{в} \cdot \text{сек} \quad (N_I \text{ см}^{-3}; T^\circ\text{К}), \quad (11.125a)$$

$$x = \frac{2,3 \cdot 10^4 T}{N_I^{1/2}} \quad (11.125b)$$

и аналогично для германия с диэлектрической проницаемостью $\kappa = 19$

$$\mu = \frac{1,2 \cdot 10^{18} T^{3/2}}{N_I \ln(1 + x^2)}, \quad (11.126a)$$

$$x = \frac{3,4 \cdot 10^4 T}{N_I^{1/2}}. \quad (11.126b)$$

Формула, определяющая μ_T , имеет вид

$$\mu_T = \frac{2 \sqrt{2\pi}}{3} \frac{\hbar^4 \cdot c_{II}}{\mathcal{E}_{1n}^3 \cdot m_n^{3/2} k^{3/2}} = \text{const } T^{-1/2}. \quad (11.127)$$

Ниже приведены экспериментально определенные значения μ_T для электронов и дырок в кремнии и германии:

$$\left. \begin{aligned} \mu_{nT} &= 15 \cdot 10^5 T^{-1/2} \text{ см}^2/\text{в} \cdot \text{сек} \\ \mu_{pT} &= 5 \cdot 10^5 T^{-1/2} \text{ см}^2/\text{в} \cdot \text{сек} \end{aligned} \right\} \text{ кремний;} \quad (11.128a)$$

$$\left. \begin{aligned} \mu_{nT} &= 19 \cdot 10^6 T^{-1/2} \text{ см}^2/\text{в} \cdot \text{сек} \\ \mu_{pT} &= 8,9 \cdot 10^6 T^{-1/2} \text{ см}^2/\text{в} \cdot \text{сек} \end{aligned} \right\} \text{ германий.} \quad (11.128b)$$

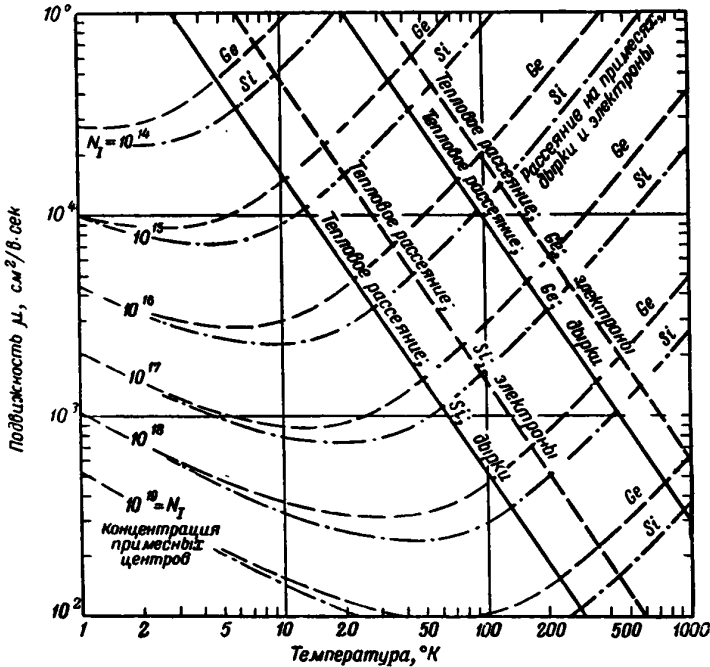
Температурные зависимости μ_I и μ_T совершенно различны; μ_I изменяется приблизительно как $T^{3/2}$, а μ_T — как $T^{-1/2}$. На фиг. 102 в двойном логарифмическом масштабе нанесена температурная зависимость подвижности дырок и электронов в кремнии и германии¹⁾.

¹⁾ Значения диэлектрической проницаемости и скоростей дрейфа в германии, полученные уже после того как фиг. 102 была изготовлена, лишь слегка отличаются от приведенных здесь (см. гл. 9 и 12). — *Прим. авт.*

Пирсон и Бардин нашли, что данные для подвижности в невырожденных образцах (см., например, фиг. 103) достаточно хорошо согласуются с формулой вида

$$\frac{1}{\mu} = \frac{a}{T^{3/2}} + bT^{3/2}, \quad (11.129)$$

что соответствует замене (на фиг. 102) кривых, изображающих ход μ_T , прямыми линиями. Полученные ими значения a с точностью

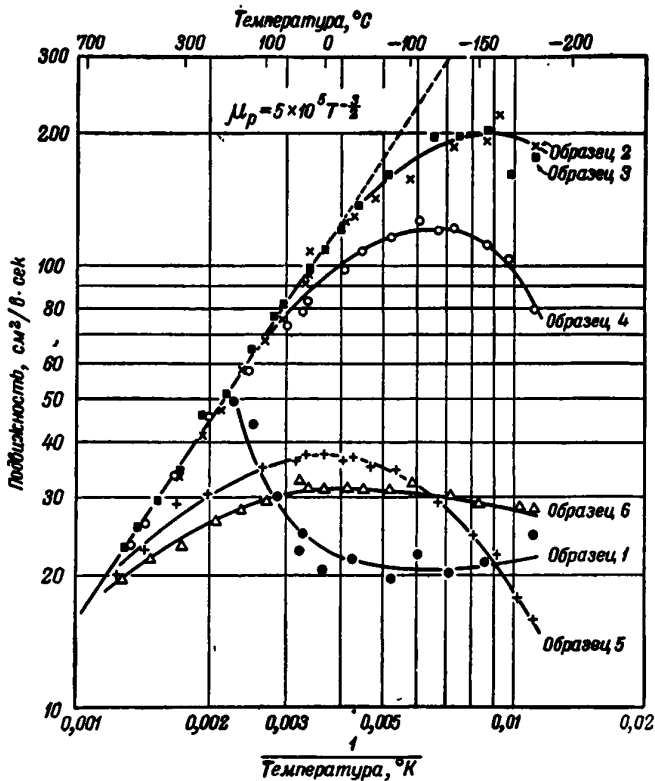


Фиг. 102. Рассеяние электронов и дырок на примесях и на тепловых колебаниях решетки.

до множителя порядка 2 соответствуют значениям a , вычисленным из уравнения для μ_T и данных о содержании примесей, определенных для области насыщения. Аналогичный результат получается и для данных, приведенных в гл. 1.

Кривая 1 (фиг. 103) заслуживает специального пояснения; она соответствует очень чистому образцу кремния. Аномальное поведение очень высокоомных образцов часто наблюдается у кремния и германия и приписывается неоднородностям. В таких образцах небольшие различия концентрации от места к месту создают пропорционально большие

различия в проводимости. При этих условиях путь, по которому проходит ток через образец, становится извилистым, в результате чего для эффекта Холла и подвижности получаются неестественно низкие значения.

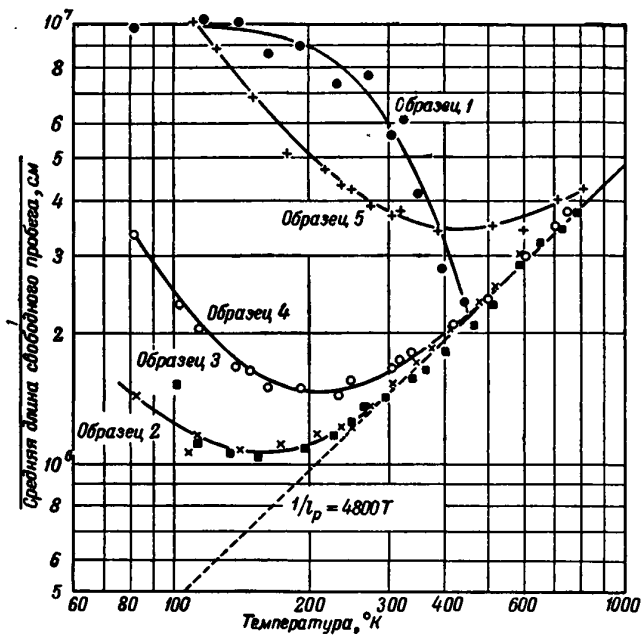


Фиг. 103. Зависимость подвижности дырок в сплавах кремния с бором от температуры.

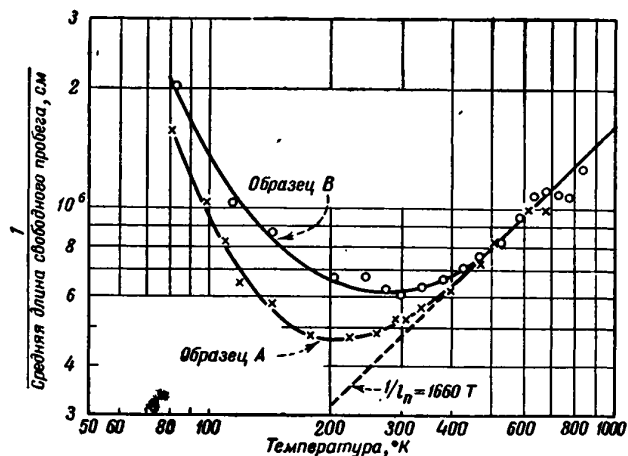
Из уравнения (11.906) следует, что средняя длина свободного пробега электрона в той области температур, где она определяется в основном рассеянием на тепловых колебаниях решетки, будет определяться выражением

$$l = 3 \sqrt{2\pi m_n k T} \frac{\mu}{4e} = 4,12 \cdot 10^{-10} \left(\frac{m_n T}{m} \right)^{1/2} \mu_L \text{ см.} \quad (11.130)$$

Значения средней длины свободного пробега электрона, полученные из этой формулы, при условии, что $m_n/m = 1$ и $m_p/m = 1$, приведены на фиг. 104 и 105.



Фиг. 104. Зависимость средней длины свободного пробега в сплавах кремния с бором от температуры.



Фиг. 105. Зависимость средней длины свободного пробега в сплавах кремния с фосфором от температуры.

Образец А — в 1 см^3 расплава введено $4,7 \cdot 10^{17}$ атомов фосфора.

Образец В — в 1 см^3 расплава введено $2,7 \cdot 10^{18}$ атомов фосфора.

При низких температурах, когда преобладает рассеяние на тепловых колебаниях решетки, можно вычислить длину свободного пробега, подставляя выражение (11.986) в (11.130), что дает

$$l = \frac{\frac{4\pi^2 (kT)^2}{\pi N_I e^4}}{\ln \left[1 + \left(\frac{3\pi kT}{e^2 N_I^{1/2}} \right)^2 \right]} = \frac{a_I 16\pi \left(\frac{kT}{\mathcal{E}_I} \right)^2}{\ln \left[1 + \left(\frac{6kT}{\mathcal{E}_I} \right)^2 \right]}, \quad (11.131)$$

где в обозначениях § 3 гл. 8

$$a_I = N_I^{-1/2} \quad \text{и} \quad \mathcal{E}_I = \frac{2e^2}{\pi a_I}. \quad (11.132)$$

Формулой (11.131), так же как и кривыми фиг. 103, можно воспользоваться для оценки концентрации ионов из данных о подвижности.

ЗАДАЧИ

1. Вывести уравнение, описывающее движение электрона по гиперболическим орбитам. Это можно сделать, если приравнять радиальное ускорение электрона разности центробежной силы и силы притяжения

$$m\ddot{r} = mr\dot{\theta}^2 - \frac{Ze^2}{\chi r^2}$$

и затем исключить $\dot{\theta}$ из этого выражения, воспользовавшись законом сохранения момента количества движения в центральном поле сил

$$mr^2\dot{\theta} = J.$$

Получающееся в результате уравнение для \ddot{r} можно преобразовать (в два приема) в уравнение, определяющее $d^2u/d\theta^2$, где $u = 1/r$:

$$\frac{dr}{dt} = \frac{dr}{d\theta} \dot{\theta} = \frac{J}{m} \frac{1}{r^2} \frac{dr}{d\theta} = -\frac{J}{m} \frac{du}{d\theta};$$

затем следует произвести подобное же преобразование и для $(d/dt)(dr/dt)$. Для того чтобы найти зависимость u от θ , полученное уравнение следует проинтегрировать. В результате интегрирования получаем

$$u = \frac{1}{r} = \frac{Ze^2 m}{\chi J^2} + A \cos(\theta + \theta_0),$$

где A и θ_0 — постоянные интегрирования.

Вывод уравнения для \ddot{r} можно найти в стандартных учебниках по теоретической физике (см. [18]).

2. Применить выражение для плотности тока, выведенное в гл. 15, к проходящей волне, описываемой уравнением (11.46в), и убедиться

в том, что проходящий ток равен

$$\frac{4P_a P_b}{(P_a + P_b)^2} \sim \left| \frac{4Z_1 Z_2}{(Z_1 + Z_2)^2} \right|.$$

Правая часть этого выражения записана применительно к электрическим линиям передачи. Проверить, что сумма отраженного и прошедшего токов равна входящему току.

3. Оценить изменение энергии электрона при его взаимодействии с фононом, учтя при этом эффект Допплера, обусловленный перемещением картины деформации. Сначала рассмотрите случай движения вдоль оси x . Пусть волновая функция падающего электрона будет $\exp i(P_1 x - \mathcal{E}_1 t)/\hbar$, а волновая функция отраженного электрона $\exp i(-P_2 x - \mathcal{E}_2 t)/\hbar$. Пусть также картина деформации перемещается со скоростью c , так что ее положение определяется уравнением $x - ct = 0$. Показать, что требование постоянства разности фаз между падающей и отраженной электронными волнами в точке $(x - ct)$ приводит к тому, что $P_2 - P_1 = 2mc$, если $\mathcal{E} = P^2/2m$, и, следовательно, $\Delta\mathcal{E} \approx m\omega c$ при $v_1 = P_1/m > c$.

Показать далее, что если перемещающаяся картина деформации описывается периодическим выражением вида $\sin 2\pi(x - ct)/\lambda_c$, то условие интерференции с усилением волн, отраженных от последовательных плоскостей равной деформации, будет $(P_2 + P_1)\lambda/\hbar = 2\pi$, и отсюда $\mathcal{E}_1 - \mathcal{E}_2 = 2\pi\hbar c/\lambda_c = \hbar\nu$, независимо от вида функциональной зависимости \mathcal{E} от P .

Значительно более трудную задачу представляет обобщение этих результатов на случай трех измерений и их сопоставление с результатами, полученными в § 6 гл. 17 (см. фиг. 144).

4. Показать, что в случае металла выражение (11.102) эквивалентно выражению (11.105).

5. Вывести формулы, необходимые для вычисления $\langle v^n \rangle$ для целых значений n , полагая $1/v_T^2 = \alpha$, записывая

$$I_n \doteq \int_0^\infty e^{-\alpha v^2} v^{n+2} dv = \frac{1}{\alpha^{(n+3)/2}} \int_0^\infty e^{-x^2} x^{n+2} dx$$

и вычисляя $dI_n/d\alpha$ путем дифференцирования под знаком интеграла. Это приводит к формуле (11.83). Вычислить интеграл I_{-1} непосредственно, а интеграл I_{-2} , рассматривая выражение

$$(I_{-2})^2 = \int_0^\infty e^{-\alpha x^2} dx \int_0^\infty e^{-\alpha y^2} dy = \int_0^{\pi/2} d\theta \int_0^\infty e^{-\alpha r^2} r dr = \frac{\pi}{4\alpha}.$$

6. Доказать, что если поверхности постоянной энергии приближенно предполагаются сферическими, так что $dS = av^2 dv$, где a — постоянная величина, то

$$\langle v^2 f_p \rangle = 3kTm,$$

если усреднение проведено по всем электронам. Это может быть доказано, исходя из определения $\langle v^2 f_p \rangle$,

$$\langle v^2 f_p \rangle = \frac{\int_0^{\infty} v^2 f_p f a v^2 dv}{\int_0^{\infty} f a v^2 dv},$$

путем интегрирования числителя по частям с помощью формулы (11.68).

7. Получить формальное выражение, определяющее тензор проводимости для некубического кристалла, в предположении, что среднее время свободного пробега τ зависит только от \mathcal{E} . Скорость изменения тока в элементе $d\Omega$, в слое $d\mathcal{E}$, вызванная электрическим полем \mathbf{E} , будет при этом

$$\left[\frac{d}{dt} dI_s \right]_{\mathbf{E}} = \frac{2e^2}{h^3 k T} \mathbf{v} (\mathbf{E} \cdot \mathbf{v}) f (1 - f) d \frac{\Omega d\mathcal{E}}{v},$$

где $\mathbf{v} = |\mathbf{v}|$. Показать, что в этом случае

$$\sigma = \frac{2e^2}{h^3 k T} \int \mathbf{v} \mathbf{v} \tau \cdot f (1 - f) \frac{d\Omega d\mathcal{E}}{v},$$

где $\mathbf{v} \mathbf{v}$ — тензорный оператор, для которого

$$(\mathbf{v} \mathbf{v}) \mathbf{E} = \mathbf{v} (\mathbf{v} \cdot \mathbf{E}).$$

8. Имеется кажущийся парадокс, связанный с наличием $\langle v^2 \tau \rangle$ в выражении для τ_{μ} . В гл. 8 было показано, что для отдельного электрона $\mu = e\tau/m$. Следовательно, если различные электроны обладают различными значениями τ , то нужно было бы ожидать, что $\mu = e \langle \tau \rangle / m$. Ошибочность последнего выражения может быть приписана тому обстоятельству, что изменение среднего времени свободного пробега электрона связано с изменением его скорости в процессе ускорения. Задача состоит в том, чтобы показать, что если этот эффект принять во внимание, то правильная формула для проводимости может быть получена из рассмотрения траекторий отдельных электронов, как это было сделано в § 5 гл. 8.

Предположите, что после столкновения электрон с одинаковой вероятностью может попасть в любое состояние в энергетическом слое $d\mathcal{E}$, но что $\tau(\mathbf{P})$ в этом слое может меняться от точки к точке. Показать, исходя из принципа детального равновесия и условия симметрии $\mathcal{E}(\mathbf{P}) = \mathcal{E}(-\mathbf{P})$, что на скорость, с которой электроны в результате рассеивания попадают в объем $dV_{\mathbf{P}}$, электрическое поле

в первом приближении не влияет и что эта скорость равна

$$\delta g_t = \frac{2}{h^3} \frac{f[\mathcal{E}(\mathbf{P}_i)] dV_{\mathbf{P}_i}}{\tau(\mathbf{P}_i)}.$$

Если среднее смещение каждого из этих электронов до его следующего столкновения равно $\bar{\mathbf{r}}_i$, то показать, что среднее значение $\sum \mathbf{v}$ для всех электронов будет

$$\sum \mathbf{v} = \sum \delta g_i \bar{\mathbf{r}}_i = \frac{2}{h^3} \int \frac{1}{\tau_i} \bar{\mathbf{r}}_i f[\mathcal{E}(\mathbf{P}_i)] dV_{\mathbf{P}_i}.$$

Под действием электрического поля, которое для простоты предположим направленным по оси x , т. е. равным $i_x E_x$, на каждый электрон действует ускоряющая его сила F_x . В результате среднее время свободного пробега и скорость электрона меняются как

$$\begin{aligned} \tau &= \tau_i + \frac{\partial \tau}{\partial P_x} F_x t, \\ \mathbf{v} &= \mathbf{v}_i + \frac{\partial \mathbf{v}}{\partial P_x} F_x t. \end{aligned}$$

Показать, что вероятность того, что электрон не испытает столкновений за время t , равна

$$c(t) = \left(1 + \frac{\partial \tau}{\partial P_x} F_x \frac{t^2}{2\tau_i^2}\right) e^{-t/\tau_i}$$

и, следовательно, что его среднее смещение будет

$$\bar{\mathbf{r}}_i = \mathbf{v}_i \tau_i + \left(\mathbf{v}_i \frac{\partial \tau}{\partial P_x} + \tau_i \frac{\partial \mathbf{v}}{\partial P_x}\right) F_x \tau_i.$$

(Заметьте, что учет принципа Паули не изменяет интегральной формы этих выражений.) Проинтегрируйте это выражение по \mathbf{P} -пространству, записав $dV_{\mathbf{P}} = dP_x dP_y dP_z$ и интегрируя dP_x по частям так, чтобы получить

$$\sum \mathbf{v} = \frac{2F_x}{kTh^3} \int \mathbf{v} v_x \tau f(1-f) dV_{\mathbf{P}}.$$

Показать, что это эквивалентно результату решения задачи 7.

9. Показать, что если поверхности постоянной энергии обладают сферической формой и τ зависит только от v , то из уравнения для \mathbf{r}_i (см. задачу 8) следует, что „подвижность“ электронов в энергетическом слое равна

$$\frac{e\tau(v)}{m} \left[1 - \frac{1}{3} \left(\frac{d \ln \tau}{d \ln v}\right)\right].$$

Усреднить это выражение по максвелловскому распределению скоростей, принимая, что $\tau(v) = l/v$, и получить в результате формулу (11.906) (стр. 332).

ЛИТЕРАТУРА

1. Bridgman P. W., Phys. Rev., **31**, 101 (1928).
2. Weisskopf V., Amer. Journ. of Phys., **11**, 1 (1943).
3. Schiff L. I., Quantum Mechanics, New York, 1949.
4. Conwell E., Weisskopf V. F., Phys. Rev., **69**, 258A (1946), **77**, 388 (1950).
5. Chapman S., Monthly Notices, R. A. S., **82**, 292 (1922).
6. Pearson G. L., Bardeen J., Phys. Rev., **75**, 865 (1949).
7. Massey H. S. W., Mohr C. B. O., Proc. Roy. Soc., **132A**, 605 (1947); **196A**, 288 (1932).
8. Johnson V. A., Lark-Horovitz K., Phys. Rev., **71**, 374 (1947).
9. Lark-Horovitz K., Middleton A. E., Miller E. P., Walerstein I., Phys. Rev., **69**, 258A (1946).
10. Bardeen J., Brattain W. H., Phys. Rev., **75** (8), 1208 (1949); Bell Syst. Tech. Journ., **28** (2), 239 (1949).
11. Lark-Horovitz K., Johnson V. A., Phys. Rev., **69**, 258A (1946)
12. Johnson V. A., Lark-Horovitz K., Phys. Rev., **71**, 374 (1947); **72**, 531 (1947).
13. Lark-Horovitz K., Middleton A. E., Miller E. P., Scanlon W. W., Walerstein I., Phys. Rev., **69**, 259 (1946).
14. Johnson V. A., Lark-Horovitz K., Phys. Rev., **69**, 259 (1946).
15. Ringer W., Welker H., Zs. f. Naturforsch., **3A**, 20 (1948).
16. Dunlap W. C., Jr., Gen. Electr. Rev., **52**, 9 (1949).
17. Pearson G. L., Lark-Horovitz K., Jamison N. C., Electr. Eng., **68**, 1047 (1949).
18. Page L., Introduction to Theoretical Physics, New York, 1935.
- 19*. Ландау Л., Компанеец А., Электропроводность металлов, М.—Л., 1935.
- 20*. Давыдов Б. И., Шмушкевич И. М., Успехи физич. наук, **24**, 21 (1940).
- 21*. Писаренко Н. Л., Известия ОМЭН, № 5—6, 631 (1938).
- 22*. Френкель Я. И., Введение в теорию металлов, 2-е изд., М.—Л., 1950.
- 23*. Гуревич Л. Э., ЖЭТФ, **16**, 416 (1946).
- 24*. Рудницкий В., ЖЭТФ, **9**, 262 (1939), **10**, 774 (1940).
- 25*. Дмитриев В. А., ЖЭТФ, **11**, 666 (1941).
- 26*. Кикоин И. К., ЖЭТФ, **10**, 1242 (1940).
- 27*. Рожанский Н., Journ. of Phys. СССР, **4**, 171 (1941).
- 28*. Стильбанс Л. С., ЖЭТФ, **22**, 77 (1952).
- 29*. Ансельм А. И., Клячкин В. И., ЖЭТФ, **22**, 297 (1952).
- 30*. Jones V. A., Phys. Rev. **81**, 149 (1951).
- 31*. Klahr C. N., Phys. Rev., **82**, 109 (1951).
- 32*. Erginsoy C., Phys. Rev., **79**, 1013 (1950).
- 33*. Шифрин К. С., ЖТФ, **14**, 40 (1944).
- 34*. Шифрин К. С., ЖТФ, **14**, 43 (1944).
- 35*. Шифрин К. С., ЖТФ, **14**, 49 (1944).

ПРИЛОЖЕНИЕ ТЕОРИИ К ЭЛЕКТРОНИКЕ ТРАНЗИСТОРОВ

§ 1. ВВЕДЕНИЕ

Теория поведения дырок и электронов в полупроводниках, изложенная в гл. 5—10, носила в значительной степени описательный характер. Влияние электрического и магнитного полей на движение дырок и электронов детально рассматривается на основе статистической теории в гл. 11. Квантовомеханическое обоснование полученных там результатов приводится в III части книги.

Прежде чем описание свойств дырок и электронов, составляющее содержание первых глав II части книги, облечь в форму строгой теории и применить ее к транзисторам и другим полупроводниковым приборам, описанным в I части, необходимо остановиться на некоторых дополнительных вопросах. Эти вопросы, а также некоторые примеры приложения теории рассмотрены в этой главе.

Другая цель настоящей главы заключается в критическом пересмотре приведенных ранее экспериментальных данных. В частности, в последних параграфах главы мы касаемся вопроса о значениях подвижностей дырок и электронов и величины \mathcal{E}_{1G} .

Многие из наиболее надежных экспериментальных данных были получены уже после того, как эта книга была написана, причем оказалось, что некоторые численные значения, использованные нами в том или ином примере, неточны. Так как эти данные были нужны скорее для иллюстрации принципов, чем для вычисления определенных результатов, то они не были заменены более новыми, полученными в то время, когда книга находилась уже в стадии подготовки к печати. Вместо этого в последних параграфах главы приведены новые, наиболее надежные, численные значения экспериментальных величин.

Основным уравнением, необходимым для рассмотрения многих задач, касающихся движения частиц, является *уравнение непрерывности*, которое формулирует закон сохранения вещества в следующем виде: *скорость возрастания числа частиц в некотором элементе объема равна результирующему потоку частиц внутрь этого элемента через его поверхность*. В случае дырок и электронов эту формулировку следует обобщить, принимая во внимание рекомбинацию и образование пар дырка—электрон в рассматриваемом элементе объема, что и будет сделано в § 2.

При изучении механизма действия большинства полупроводниковых устройств необходимо рассмотреть беспорядочное диффузионное

движение носителей зарядов, вызванное тепловыми колебаниями решетки, а также их систематический дрейф в электрическом и магнитном полях (см. фиг. 78, иллюстрирующую оба вида движения). Диффузионное движение характеризуется коэффициентом диффузии, связанным с подвижностью уравнением Эйнштейна: $D_n = kT\mu_n/e$ для электронов или, соответственно, $D_p = kT\mu_p/e$ для дырок. Это уравнение можно получить непосредственно с помощью статистических методов расчета, приведенных в гл. 11, или же на основе рассмотрения принципа детального равновесия. В § 3 приведен этот вывод.

В § 4 введены в рассмотрение так называемые *квази-уровни Ферми*. Это понятие оказалось, весьма полезным при анализе процессов, протекающих в полупроводниках в тех случаях, когда концентрации дырок и электронов не находятся в равновесии.

В § 5 теория диффузионных явлений применяется для получения характеристик упрощенной модели $p-n$ -перехода, в том числе для выпрямления и зависимости полной проводимости от частоты.

В § 6 речь идет о диффузии и рекомбинации носителей заряда в нитевидном образце германия. Во многих случаях рекомбинация имеет место главным образом на поверхности образца и ограничена скоростью диффузии к поверхности.

В § 7 рассматривается движение импульса дырок (или электронов) в веществе с электропроводностью противоположного знака и указывается причина искажения формы импульса.

В § 8 дается обзор различных способов определения величины ξ_{10} , представляющей собой изменение ширины запрещенной зоны при изменении объема на единицу, и найденные значения сравниваются со значениями, полученными из измерений подвижности.

В § 9 приводится обзор данных о подвижности и влиянии магнитных полей на электропроводность, в том числе и наиболее надежных данных для значений подвижности, определенной из эффекта Холла, и наблюдений скорости дрейфа носителей заряда в германии, которые были получены ко времени сдачи книги в печать.

В § 10 приведены некоторые экспериментальные результаты и соображения, относящиеся к вопросу о происхождении шумов.

Заключительный параграф этой главы содержит общую оценку современного состояния теории полупроводников и экспериментальных работ в этой области, а также анализ перспектив развития этой области физики.

§ 2. ПРИНЯТЫЕ ОБОЗНАЧЕНИЯ; УРАВНЕНИЕ НЕПРЕРЫВНОСТИ

В этой главе приняты следующие обозначения (определение некоторых из упомянутых здесь терминов будет дано в § 3):

$$\begin{aligned}
 &\psi \text{ — электростатический потенциал,} \\
 E = -\nabla\psi &\text{ — напряженность электрического поля,} \\
 &\varphi \text{ — уровень Ферми,}
 \end{aligned}$$

φ_n — квази-уровень Ферми для электронов,
 φ_p — квази-уровень Ферми для дырок,
 $+q$ — заряд дырки ($=e$).

Предположим также, что любые отклонения концентраций дырок и электронов от их значений в условиях теплового равновесия выравниваются со скоростью, характеризуемой постоянными времени или временами жизни:

τ_p — время жизни дырок в веществе с проводимостью n -типа¹⁾;
 τ_n — время жизни электронов в веществе с проводимостью p -типа.

В связи с рассмотрением теории p — n -переходов введем для концентраций носителей тока в условиях теплового равновесия следующие обозначения:

n_n и p_n — концентрации электронов и дырок в веществе n -типа;
 n_p и p_p — концентрации электронов и дырок в веществе p -типа.

Основным уравнением, необходимым для рассмотрения движения дырок и электронов в условиях, когда концентрации носителей тока есть функции координат и времени, является *уравнение непрерывности*. Для случая дырок можно записать его в виде:

$$\frac{\partial p}{\partial t} = (g - r) - \frac{1}{q} \nabla \cdot I_p. \quad (12.1)$$

Это уравнение выводится следующим образом. Рассмотрим малую область dx , dy , dz , центр которой находится в точке x , y , z . Скорость изменения числа дырок в элементе объема $dx dy dz$ равна

$$\frac{\partial p}{\partial t} dx dy dz. \quad (12.2)$$

Эти дырки появляются в рассматриваемом элементе объема вследствие преобладания процесса их образования над процессом рекомбинации, а также их притока через поверхность элемента. Первый процесс описывается выражением

$$(g - r) dx dy dz, \quad (12.3)$$

где g — скорость образования дырок в единице объема (под действием света, рентгеновских лучей, теплового возбуждения и т. д.), а r — скорость уменьшения их концентрации, обусловленная рекомбинацией с электронами.

Так как p равно числу свободных незахваченных дырок, то при определении r следует также учитывать дырки, захватываемые без рекомбинации, а при определении g — дырки, освобожденные в результате реэмиссии. Член, содержащий $\nabla \cdot I_p$, представляет собой поток дырок через поверхности, ограничивающие рассматриваемый

¹⁾ Во избежание недоразумений отметим, что в гл. 8 и 11 буквой τ обозначено время релаксации. Время релаксации зависит от скорости v . — *Прим. авт.*

элемент объема, и получается следующим образом. Плотность тока в направлении x в центре грани $dy dz$ с координатой $x = (dx/2)$ равна

$$I_{px}(x, y, z) = \frac{\partial I_p(x, y, z)}{\partial x} \frac{dx}{2}. \quad (12.4)$$

Это есть среднее значение плотности тока для данной грани, так что поток дырок через всю грань $dy dz$ *внутри* объема $dx dy dz$ будет равен

$$\frac{1}{q} \left[I_{px} - \frac{\partial I_{px}}{\partial x} \frac{dx}{2} \right] dy dz. \quad (12.5)$$

Поток дырок *из* элемента объема $dx dy dz$ через грань $dy dz$ с координатой $x = (dx/2)$ будет отличаться только тем, что перед вторым членом этого выражения вместо знака минус будет стоять плюс.

Результирующий поток через обе грани $dy dz$ в элемент объема $dx dy dz$ равен, таким образом,

$$\begin{aligned} \frac{1}{q} \left[I_{px} - \frac{\partial I_{px}}{\partial x} \frac{dx}{2} \right] dy dz - \frac{1}{q} \left[I_{px} + \frac{\partial I_{px}}{\partial x} \frac{dx}{2} \right] dy dz = \\ = - \frac{1}{q} \frac{\partial I_{px}}{\partial x} dx dy dz. \end{aligned} \quad (12.6)$$

Поступая аналогичным образом для граней $dx dy$ и $dx dz$, получаем полный результирующий поток дырок *внутри* элемента объема $dx dy dz$

$$- \frac{1}{q} \left[\frac{\partial I_{px}}{\partial x} + \frac{\partial I_{py}}{\partial y} + \frac{\partial I_{pz}}{\partial z} \right] dx dy dz \equiv - \frac{1}{q} \nabla \cdot I_p dx dy dz, \quad (12.7)$$

где величина, стоящая в квадратных скобках, есть, по определению, $\nabla \cdot I_p$, т. е. $\text{div } I_p$. Приравнявая результирующую скорость изменения числа дырок в элементе объема $dx dy dz$ к сумме скорости их образования и результирующего потока через поверхность *внутри* этого элемента объема, получаем уравнение непрерывности (12.1), умноженное на общий множитель $dx dy dz$, который может быть опущен. Концентрации введенных носителей p и n стремятся к равновесным значениям p_n и n_p ; предположим, что этот процесс характеризуется определенными значениями времен жизни, не зависящими от концентрации носителей и способов введения последних в кристалл. По этой причине будем считать, что *результирующая скорость изменения концентрации введенных дырок в материале n -типа, обусловленная рекомбинацией и спонтанным образованием*, будет равна

$$\frac{p_n - p}{\tau_p}. \quad (12.8)$$

Для электронов в материале с проводимостью p -типа соответствующее выражение будет иметь вид

$$\frac{n_p - n}{\tau_n}. \quad (12.9)$$

Уравнение непрерывности для дырок запишется в таком случае как

$$\frac{dp}{dt} = \frac{p_n - p}{\tau_p} - \frac{1}{q} \nabla \cdot I_p + g'_p, \quad (12.10)$$

где g'_p — результирующая скорость образования дырок в рассматриваемом элементе объема под действием фотонов или других *внешних* возбуждающих факторов. Аналогичное уравнение для электронов в материале с электропроводностью p -типа имеет вид

$$\frac{dn}{dt} = \frac{n_p - n}{\tau_n} + \frac{1}{q} \nabla \cdot I_n + g'_n. \quad (12.11)$$

Этими уравнениями, записанными для случая, когда $g'_p = g'_n = 0$, мы воспользуемся для исследования движения дырок и электронов в p — n -переходах и нитевидных образцах германия (см. § 5 и 6).

§ 3. СООТНОШЕНИЕ ЭЙНШТЕЙНА, СВЯЗЫВАЮЩЕЕ ПОДВИЖНОСТЬ С КОЭФФИЦИЕНТОМ ДИФфуЗИИ¹⁾

Рассмотрим процесс диффузии носителей тока только для того случая, когда концентрация последних столь мала, что можно пренебречь ограничениями, налагаемыми принципом Паули. Можно предполагать, что в этих условиях каждый носитель тока дрейфует в электрическом поле и диффундирует независимо от остальных носителей. Предположим также, что электрическое поле настолько мало, что не нарушает заметным образом равновесного распределения скоростей носителей тока. Это предположение позволило в гл. 11 пренебречь членами, содержащими E в степени выше первой. При этих условиях законно считать, что дырочный ток (в нашем примере мы рассматриваем дырки) складывается из тока проводимости и диффузионного тока. Последний прямо пропорционален²⁾ градиенту концентрации дырок ∇p . Таким образом,

$$I_p = q\mu_p p E - qD_p \nabla p, \quad (12.12)$$

$$I_n = q\mu_n n E + qD_n \nabla n. \quad (12.13)$$

Знак минус перед вторым членом в выражении (12.12) стоит потому, что в отсутствие электрического поля результирующая диффузия дырок происходит в направлении, противоположном направлению возрастания концентрации дырок p . Так как заряд электрона отрицателен, то соответственно перед вторым членом в выражении (12.13) будет стоять знак плюс.

¹⁾ О теории движения электронов и дырок в германии и других полупроводниках опубликована работа [42]. Вопрос о диффузии носителей тока в полупроводниках со смешанной проводимостью рассмотрен в работах [43—46]. — *Прим. ред.*

²⁾ Это можно показать с помощью статистических методов, подобных рассмотренным в гл. 11. — *Прим. авт.*

Предположим, что электрическое поле приложено к кристаллу в условиях теплового равновесия, как, например, в случае p — перехода, рассмотренном в § 2 гл. 4, или контакта металл — полупроводник, рассмотренного в § 3 той же главы. При этих условиях статистический принцип детального равновесия [1] требует, чтобы каждой дырке, проходящей в единицу времени данную площадку в одном направлении, соответствовала другая дырка, проходящая через эту же площадку в обратном направлении и с такой же скоростью. Принцип детального равновесия справедлив также и при наличии магнитных полей [2]; этим обстоятельством мы скоро воспользуемся. Таким образом, в условиях равновесия I_p должно обращаться в нуль. Однако при равновесии концентрация дырок определяется *больцмановским множителем* $\exp(-q\psi/kT)$, характеризующим тенденцию дырок не находиться в областях с высокой потенциальной энергией¹⁾. Формула для p была уже приведена нами раньше [см. (10.19)],

$$p = Ae^{-q\psi/kT}, \quad (12.14)$$

где A — постоянная, содержащая N_v и \mathcal{E}_F ; рассмотрим ее более подробно в связи с вопросом о квази-уровнях Ферми. Подставив (12.14) в уравнение (12.13), вычисляя ∇p и полагая, что $I_p = 0$, находим

$$0 = q\mu_p p E - qD_p \left(-\frac{q}{kT} \nabla\psi\right) Ae^{-q\psi/kT} = q \left(\mu_p - \frac{q}{kT} D_p\right) p E. \quad (12.15)$$

Это уравнение удовлетворяется только в том случае, если справедливо *соотношение Эйнштейна* [3, 4]

$$\mu_p = \frac{q}{kT} D_p. \quad (12.16)$$

Аналогичный вид имеет соотношение Эйнштейна и для электронов:

$$\mu_n = \frac{q}{kT} D_n. \quad (12.17)$$

(В случае электронов нужно помнить, что величина μ_n положительна и, следовательно, положительно также и значение D_n .)

Размерность D_p — (длина)²/время. В системе CGS величина D_p выражается в $\text{см}^2/\text{сек}$. Если подвижность дырок μ_p выражена в $\text{см}^2/\text{в} \cdot \text{сек}$, то (12.16) можно записать следующим образом:

$$\mu_p \text{ см}^2/\text{в} \cdot \text{сек} = \frac{q}{kT} D_p \text{ см}^2/\text{сек}. \quad (12.18)$$

Здесь q/kT имеет размерность в^{-1} ; при 300°K $kT/q = 0,026 \text{ в}$, так что в практической системе единиц и в системе MKS имеем соответственно

$$D_p = D_{pL} = 0,026 \mu_{pL}, \quad (12.19)$$

$$D_{pM} = 0,026 \mu_{pM}. \quad (12.20)$$

¹⁾ Связь $q\psi$ с \mathcal{E}_0 обсуждается более подробно в следующем параграфе.—
Прим. авт.

Величина $D_p \tau_p$, встречающаяся в последующих параграфах этой главы, имеет размерность длины в квадрате и обозначается символом L_p^2 , причем L_p называется *диффузионной длиной*. Диффузионная длина (см. § 5, 6 и задачи) есть среднее расстояние вдоль любого данного направления, например вдоль оси x , которое проходит дырка при диффузии, прежде чем успеваеет рекомбинировать, если время ее жизни равно при этом τ_p .

В неравновесных условиях выражение для тока может быть записано в виде суммы двух членов: члена, обусловленного током проводимости, и диффузионного члена [см., например, выражение (12.13)]. В общем случае эта сумма отлична от нуля.

При наличии магнитного поля направление тока проводимости не совпадает обычно с направлением электрического поля, а направление диффузионного тока не совпадает с направлением ∇p . Полная проводимость и коэффициенты диффузии в этом случае становятся тензорами, которые мы можем определить через тензор сопротивления ρ . Рассмотрим простейший пример кубического кристалла в предположении, что магнитное поле \mathbf{H} направлено вдоль оси z (направление [001]). В этом случае компоненты напряженности электрического поля запишутся в виде

$$E_x = \rho_t (I_x - I_y \operatorname{tg} \theta), \quad (12.21)$$

$$E_y = \rho_t (I_x \operatorname{tg} \theta + I_y), \quad (12.22)$$

$$E_z = \rho_l I_z. \quad (12.23)$$

Здесь ρ_t — сопротивление в направлении, перпендикулярном к \mathbf{H} [плотность мощности рассеяния равна $\rho_t (I_x^2 + I_y^2)$], ρ_l — сопротивление в направлении, параллельном \mathbf{H} , и θ — угол Холла, т. е. угол между направлениями \mathbf{E} и \mathbf{I} в плоскости, перпендикулярной направлению магнитного поля. Эти выражения аналогичны выражениям для эффекта Холла, приведенным в гл. 8 и 11, с той только разницей, что мы записали их в форме, пригодной и при больших значениях \mathbf{H} и θ . При этой форме записи угол θ положителен для случая дырок и отрицателен для электронов. Выражение (12.21) и (12.22) можно записать в виде одного комплексного выражения

$$E_x + iE_y = \frac{\rho_t e^{i\theta}}{\cos \theta} (I_x + iI_y). \quad (12.24)$$

Для определения коэффициента диффузии следует решить выражения (12.21)—(12.23) относительно I_x , I_y и I_z и записать их, как $I_x = q\mu v_x$ и т. д., где v_x и т. д. есть скорости дрейфа. Введем обозначения

$$\mu_t = \frac{1}{q\rho\rho_t}, \quad \mu_l = \frac{1}{q\rho^2}. \quad (12.25)$$

В таком случае получим

$$v_x + iv_y = e^{-i\theta} \mu_t \cos \theta (E_x + iE_y) \quad (12.26)$$

или

$$v_x = \mu_t \cos^2 \theta (E_x + E_y \operatorname{tg} \theta) \equiv \mu_{xx} E_x + \mu_{xy} E_y, \quad (12.27)$$

$$v_y = \mu_t \cos^2 \theta (-E_x \operatorname{tg} \theta + E_y) \equiv \mu_{yx} E_x + \mu_{yy} E_y, \quad (12.28)$$

$$v_z = \mu_l E_z \equiv \mu_{zz} E_z. \quad (12.29)$$

Диффузионный ток определяется на основе принципа детального равновесия и закона распределения Больцмана (12.14), которые справедливы также и при наличии магнитного поля. В результате получаем соотношения

$$I_x = -q \left(D_{xx} \frac{\partial p}{\partial x} + D_{xy} \frac{\partial p}{\partial y} \right), \quad (12.30)$$

$$I_y = -q \left(D_{yx} \frac{\partial p}{\partial x} + D_{yy} \frac{\partial p}{\partial y} \right), \quad (12.31)$$

$$I_z = -q D_{zz} \frac{\partial p}{\partial z}. \quad (12.32)$$

Здесь каждое значение коэффициента диффузии D в (kT/q) раз больше соответствующего значения μ в (12.25)—(12.29). Аналогичные соотношения справедливы и для электронов.

§ 4. ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ. ЭЛЕКТРОСТАТИЧЕСКИЙ ПОТЕНЦИАЛ И КВАЗИ-УРОВНИ ФЕРМИ

На фиг. 106 изображена схема энергетических уровней для p - n -перехода. Подобные схемы уже рассматривались в гл. 4, а также в связи с обсуждением фиг. 62 и 73. Теперь им будет дано аналитическое обоснование. Схема, изображенная на фиг. 106, обозначает, что электрон, занимающий самое низкое состояние в зоне проводимости, или, лучше, волновой пакет, локализованный вблизи этого состояния, будет обладать различной энергией в разных частях кристалла. Эта схема относится к случаю наличия макроскопического электрического поля, так как энергия взаимодействия электрона с локальным атомным полем не зависит от местоположения электрона в кристалле, если только распределение его волновой функции в пространстве во всех случаях одинаково. Таким образом, если электростатический потенциал в одной из точек кристалла равен ψ_A , а в другой ψ_B , то, поскольку заряд электрона обозначен как $-q$, энергия \mathcal{E}_c , отвечающая наименьшему состоянию в зоне проводимости, будет иметь значение

$$\mathcal{E}_{cB} = \mathcal{E}_{cA} + q(\psi_A - \psi_B) \quad (12.33)$$

или, если точка B перемещается, в то время как точка A остается неподвижной,

$$\mathcal{E}_{cB} = -q\psi_B + \text{const.} \quad (12.34)$$

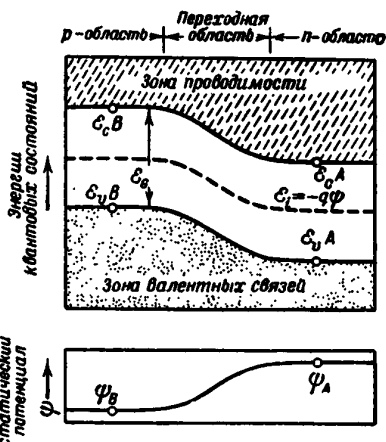
Аналогичное соотношение справедливо и для волнового пакета вблизи верхнего края зоны валентных связей, и энергия электрона будет в этом случае меняться как $-q\psi$. Отсюда можно заключить, что энергия дырки должна меняться как $q\psi$. Рассмотрим этот вопрос подробнее.

Потенциальная энергия дырки может быть определена как работа электростатических сил, совершаемая при перемещении дырки. Эту энергию можно вычислить при помощи следующего процесса: предположим, что дырка находится в точке B . Можно перевести ее в точку A , возбудив для этого электрон, находящийся в точке A , из зоны валентных связей в зону проводимости, на что должна быть затрачена энергия, равная \mathcal{E}_G ; можно затем перевести электрон в точку B , затратив на это работу, равную $-q(\psi_B - \psi_A)$, и далее в результате рекомбинации его с дыркой в точке B получить энергию \mathcal{E}_G . Таким образом, перемещение дырки из B в A связано с затратой работы, равной $q(\psi_A - \psi_B)$, которая как раз равна работе перемещения заряда $+q$ из точки B в точку A . Соответственно можно принять, что энергия дырки меняется как $+q\psi$. Этот результат полностью соответствует результатам, полученным в гл. 7, 8 и 11, согласно которым можно считать, что в электрическом поле дырка ведет себя под действием силы qE , как частица с положительным зарядом и с положительной массой. Те же выводы можно сделать и на основании более детального рассмотрения вопроса, проведенного в § 8 гл. 15.

Для наших целей удобно связать электростатический потенциал с уровнем энергии, лежащим примерно посередине запрещенной зоны. В настоящем параграфе рассмотрим только математическую сторону этой задачи, а на ее физической интерпретации подробно остановимся несколько позже. Обозначим этот уровень через \mathcal{E}_i (индекс i указывает на то, что речь идет о собственной проводимости) и определим его как

$$\mathcal{E}_i = \frac{1}{2} \left[\mathcal{E}_c + \mathcal{E}_v + kT \ln \frac{N_v}{N_c} \right]. \quad (12.35)$$

Отдельные обозначения имеют здесь тот же смысл, что и в гл. 10,



Ф и г. 106. Соотношение между уровнями энергии и электростатическим потенциалом.

Электростатический потенциал равен в этом случае

$$\psi = -\frac{\mathcal{E}_i}{q} \quad \text{или} \quad \mathcal{E}_i = -q\psi, \quad (12.36)$$

а энергии \mathcal{E}_c и \mathcal{E}_v равны соответственно

$$\mathcal{E}_c = -q\psi + \frac{1}{2} kT \ln \frac{N_c}{N_v} + \frac{1}{2} \mathcal{E}_G, \quad (12.37a)$$

$$\mathcal{E}_v = -q\psi + \frac{1}{2} kT \ln \frac{N_c}{N_v} - \frac{1}{2} \mathcal{E}_G, \quad (12.37b)$$

где

$$\mathcal{E}_G = \mathcal{E}_c - \mathcal{E}_v. \quad (12.38)$$

Определим также уровень Ферми через электростатический потенциал из уравнения

$$\varphi = -\frac{\mathcal{E}_F}{q} \quad \text{или} \quad \mathcal{E}_F = -q\varphi \quad (12.39)$$

и будем называть φ и \mathcal{E}_F *уровнем Ферми*. При таком определении этих величин и при условии, что $\varphi = \psi$, концентрации дырок и электронов одинаковы, как в этом можно убедиться, подставляя (12.39) в (10.16) и (10.19). Это дает

$$n = N_c e^{-(\mathcal{E}_c - \mathcal{E}_F)/kT} = n_i e^{q(\psi - \varphi)/kT}, \quad (12.40a)$$

$$p = N_v e^{-(\mathcal{E}_F - \mathcal{E}_v)/kT} = n_i e^{q(\varphi - \psi)/kT}, \quad (12.40b)$$

где

$$n_i = \sqrt{N_v N_c} e^{-\mathcal{E}_G/2kT} = \sqrt{p n}, \quad (12.41)$$

в согласии с результатами, даваемыми выражением (10.24). Величина n_i представляет собой концентрацию дырок и электронов в образце с собственной проводимостью, для которого условие равенства пространственного заряда нулю влечет за собой равенство $p = n$.

Выясним теперь физический смысл полученных выше уравнений. Для этой цели рассмотрим полупроводник, содержащий область p — n -перехода, где электростатический потенциал и значения концентраций p и n меняются от точки к точке. В переходной области имеется, очевидно, точка, в которой концентрации p и n равны и, таким образом, каждая из них равна n_i , т. е. концентрации носителей в образце с собственной проводимостью. Для этой точки образца уровень Ферми лежит приблизительно посередине между заполненной и свободной зонами. Однако если бы уровень Ферми лежал точно посередине запрещенной зоны, то значения $f = \exp[(\mathcal{E}_F - \mathcal{E}_c)/kT]$ и $f_p = \exp[(\mathcal{E}_v - \mathcal{E}_F)/kT]$ были равны друг другу, и отношение p/n равнялось бы отношению N_c/N_v .

Когда $N_c > N_v$, уровень \mathcal{E}_F должен лежать ниже середины запрещенной зоны, так что при этом n уменьшается, а p увеличивается

до тех пор, пока они не становятся равными друг другу. Смещение запрещенной зоны вверх на величину $\Delta\mathcal{E}$ уменьшает концентрацию n в $\exp(-\Delta\mathcal{E}/kT)$ раз и увеличивает p в $\exp(\Delta\mathcal{E}/kT)$ раз, так что n/p уменьшается в результате в $\exp(-2\Delta\mathcal{E}/kT)$ раз. Приравнявая этот последний множитель отношению N_v/N_c , находим

$$\Delta\mathcal{E} = \frac{kT}{2} \ln \frac{N_c}{N_v}.$$

Следовательно, в той точке полупроводника, где $n = p = n_i$, уровень энергии, лежащий на $\Delta\mathcal{E}$ ниже середины запрещенной зоны, будет совпадать с уровнем \mathcal{E}_F . Так как положение этого уровня задано выражением

$$\begin{aligned} \frac{1}{2}(\mathcal{E}_c + \mathcal{E}_v) - \Delta\mathcal{E} &= \frac{1}{2}(\mathcal{E}_c + \mathcal{E}_v) - \frac{1}{2}kT \ln \frac{N_c}{N_v} = \\ &= \frac{1}{2}(\mathcal{E}_c + \mathcal{E}_v + kT \ln \frac{N_v}{N_c}), \end{aligned}$$

то это есть уровень \mathcal{E}_i , определяемый формулой (12.35).

В тех местах p — n -перехода, где уровень \mathcal{E}_i лежит выше уровня Ферми \mathcal{E}_F , концентрация n равна произведению n_i на множитель $\exp(\mathcal{E}_F - \mathcal{E}_i)/kT$, т. е. уменьшается, а концентрация p равна произведению n_i на множитель $\exp[-(\mathcal{E}_F - \mathcal{E}_i)/kT]$, т. е. увеличивается.

Если определить электростатический потенциал из выражения $-q\psi = \mathcal{E}_i$, а потенциал Ферми из выражения $-q\varphi = \mathcal{E}_F$, то на основе предшествующих рассуждений получим формулы (12.40) для концентраций n и p . Если $\varphi = \psi$, то концентрации соответствуют случаю собственной проводимости. Отклонения ψ от φ приводят к появлению множителей $\exp \pm q(\psi - \varphi)/kT$. Знак плюс относится к электронам, так как увеличение ψ соответствует более положительным значениям потенциала и, таким образом, более низким потенциальным энергиям электронов, а следовательно, более высоким концентрациям последних; знак минус соответствует дыркам, которые стремятся уйти из области с положительным потенциалом.

Разности значений потенциала Ферми φ и есть те величины, которые мы измеряем вольтметром. Для иллюстрации этого заключения представим себе, что два проводника A и B соединены через батарею, но что цепь остается разомкнутой или же обладает столь большим сопротивлением, что через нее протекает лишь ничтожный ток. При этих условиях каждый проводник находится по существу в тепловом равновесии и распределение электронов в зоне по энергетическим уровням соответствует распределению по Ферми—Дираку. Однако система в целом не находится в тепловом равновесии, и если металлы A и B привести в контакт друг с другом, то пойдет ток и в системе будут иметь место необратимые тепловые эффекты. Источником энергии, конечно, служит батарея, не являющаяся сама по себе равновесной системой, Если металлы A и B одинаковы и

обладают одинаковыми значениями работы выхода, то в пространстве между ними, очевидно, будет существовать электрическое поле, создающее разность потенциалов между точками, расположенными как раз у самой поверхности каждого металла, и равную напряжению батареи. Так как оба металла обладают одинаковыми свойствами, то для каждого из них разность между уровнем Ферми и уровнем, отвечающим неподвижному электрону, находящемуся вне металла¹⁾, как раз у самой его поверхности, будет одной и той же. Следовательно, разность между уровнями Ферми будет равна потенциалу батареи. Этот результат справедлив также и в том случае, если проводники различны. В этом можно убедиться следующим образом. Предположим, что A' и B' — два различных проводника, причем A соединен с проводником A' , а B приведен в контакт с B' . Тогда ввиду того, что проводники A' и A находятся в локальном равновесии, они будут обладать одним и тем же уровнем Ферми; то же самое можно сказать относительно проводников B и B' . Следовательно, разность потенциалов Ферми для проводников A' и B' будет также равна потенциалу батареи. Другой пример, иллюстрирующий заключение о том, что изменение потенциала Ферми эквивалентно изменению приложенной разности потенциалов, приведен в конце этого параграфа. Статистическая теория, позволяющая рассматривать уровень Ферми как *химический потенциал* электронов, изложена в конце § 1 гл. 16.

Теперь воспользуемся формулами (12.40), чтобы определить распределение потенциала в полупроводнике с неравномерным содержанием примесей, находящемся в состоянии теплового равновесия. В таком полупроводнике концентрации доноров и акцепторов являются функциями $N_d(r)$ и $N_a(r)$ положения r . Предположим также, что доноры и акцепторы полностью ионизированы, как это имеет место в германии при комнатной температуре, когда расстояние уровня Ферми от краев запрещенной зоны превышает 0,1 эв. При этих условиях плотность заряда ρ равна

$$\begin{aligned} \rho &= q(p - n + N_d - N_a) = \\ &= q \left[N_d(r) - N_a(r) + 2n_i \operatorname{sh} q \frac{(\varphi - \psi)}{kT} \right]. \end{aligned} \quad (12.42)$$

Здесь значение φ можно рассматривать как заданное, а ψ — как неизвестную функцию расстояния r . [Для читателя полезно сопоставить формулы (12.42) с фиг. 131 и 135 (стр. 542 и 546), где разность между кривыми, изображенными сплошной и пунктирной жирными линиями, равна значению ρ , построенному как функция от $q(\psi - \varphi)$.] Зависимость ψ от r определяется из уравнения Пуассона, которое в электростатической системе единиц или в системе MKS имеет, соответ-

¹⁾ Эта разность и есть *истинная работа выхода* поверхности. Подробнее см., например, [5]. — *Прим., авт.*

ственно, вид

$$\nabla^2\psi = -\frac{4\pi\rho}{\kappa}, \quad \text{или} \quad \nabla^2\psi = -\frac{\rho q}{\kappa\epsilon_0}. \quad (12.43)$$

Это уравнение становится уравнением с одним неизвестным ψ , если мы подставим в него ρ в форме (12.42) и предположим, что заданы функции $N_d(r)$ и $N_a(r)$, а также граничные условия для ψ .

Нетрудно показать, что это уравнение всегда имеет решения. Одномерный случай подробно рассмотрен в специальной литературе [6, 7]. Не вдаваясь в детали математического анализа, попытаемся выяснить физическое содержание выражений (12.42) и (12.43). Предположим, что нам известно решение для того случая, когда значения N_d и N_a не зависят от r . (Такие решения приведены в гл. 10 и 16.) Предположим затем, что мы увеличиваем N_d в пределах некоторой небольшой области полупроводника R , сохраняя при этом значения n и p постоянными и равными n_0 и p_0 . При таком изменении N_d электростатический потенциал в области R увеличивается от значения ψ_0 до значения ψ_1 ; ψ_1 здесь более положительно, так как увеличение N_d соответствует появлению дополнительного положительного заряда. Представим теперь, что ψ поддерживается постоянным и равным ψ_1 и что концентрации электронов и дырок могут самопроизвольно изменяться, причем дырки покидают область R , а электроны входят в эту область. В результате такого перемещения электронов и дырок концентрация отрицательного заряда в области R увеличивается в соответствии со знаком члена, содержащего ψ [формула (12.42)]. Изменение концентрации зарядов, обусловленное потоком носителей, стремится, таким образом, компенсировать первоначальное изменение концентрации, связанное с увеличением N_d . Эта тенденция играет весьма существенную роль. Если мы изменяем N_d , то n и p меняются так, чтобы компенсировать изменение плотности заряда. Таким образом, перемещение дырок и электронов стремится стабилизировать потенциал и нейтрализовать изменение концентрации примесей.

Автор предполагает, что сделанные выше замечания послужат читателю поводом для размышлений и углубят его понимание процессов, связанных с изменением плотности зарядов. На основании предыдущего можно вполне строго показать, что значения ψ , а следовательно, также n и p могут быть найдены для произвольных значений $N_d(r)$ и $N_a(r)$.

Перейдем теперь к исследованию $p-n$ -перехода, изображенного на фиг. 107. Предположим, что слева от перехода, в p -области, разность $N_a - N_d$ положительна и постоянна, имея значение, равное $p_p - n_p$. В этой области концентрация электронов n_p очень мала по сравнению с концентрацией дырок, так что условие нейтральности пространственного заряда приближенно дает

$$N_p - N_d = p_p \quad (12.44)$$

или, если ψ_p есть значение ψ в p -области, то

$$p_p = n_i e^{-q(\psi - \varphi)/kT}. \quad (12.45)$$

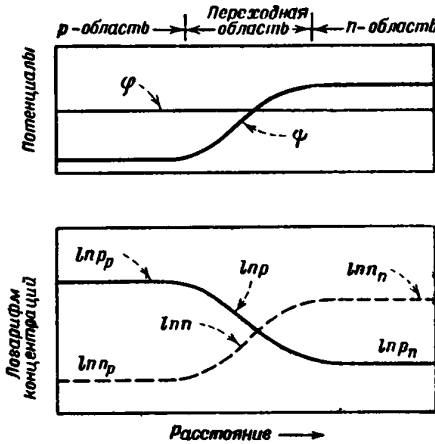
Разделив это выражение на n_i и логарифмируя, получим

$$\psi_p = \varphi - \frac{kT}{q} \ln \frac{p_p}{n_i}. \quad (12.46)$$

Далее предположим, что в n -области разность $N_d - N_a$ положительна и постоянна; обозначим эту разность через n_n . Значение ψ в n -области, следовательно, равно

$$\psi_n = \varphi + \frac{kT}{q} \ln \frac{n_n}{n_i}. \quad (12.47)$$

Таким образом, ψ_n более положительно, чем ψ_p , при постоянном значении φ , соответствующем условию равновесия. Аналогичный случай был уже рассмотрен в гл. 4. Сейчас мы дадим аналитическую формулировку сделанного в гл. 4 заключения о том, что диффузия дырок из p -области, где их концентрация имеет наибольшее значение, компенсируется потоком их в обратном направлении, обусловленным наличием электрического поля. Для этой цели введем понятие о квази-уровнях Ферми.



Фиг. 107. Зависимость n и p от потенциала ψ и положения уровня Ферми для $p-n$ -перехода в случае более высокой проводимости в p -области.

Совокупность процессов диффузии и электропроводности в полупроводниках удобно рассматривать с помощью представления о квази-уровнях Ферми, которые будем обозначать через φ_n для электронов и φ_p для дырок. Значения φ_n и φ_p могут быть выражены через n , p и ψ для неравновесных условий при помощи выражений

$$n = n_i e^{q(\psi - \varphi_n)/kT}, \quad \varphi_n = \psi - \frac{kT}{q} \ln \frac{n}{n_i}; \quad (12.48a)$$

$$p = n_i e^{-q(\psi - \varphi_p)/kT}, \quad \varphi_p = \psi + \frac{kT}{q} \ln \frac{p}{n_i}. \quad (12.48б)$$

Очевидно, что при условии теплового равновесия, являющегося единственным условием, при котором уровень Ферми φ действительно может быть определен, как φ_n , так и φ_p равны φ .

Воспользовавшись выражениями для φ_n и φ_p , а также соотношением Эйнштейна, позволяющим исключить из рассмотрения D_p и D_n ,

результатирующие плотности токов можно записать в виде

$$I_p = q\mu_p p E - qD_p \nabla p = -q\mu_p p \nabla \psi - q \frac{kT}{q} \mu_p p \nabla \frac{-q(\psi - \varphi_p)}{kT} = \\ = -q\mu_p p \nabla \varphi_p \quad (12.49a)$$

и

$$I_n = -q\mu_n n \nabla \varphi_n. \quad (12.49б)$$

Эти уравнения формально аналогичны уравнениям для *токов, обуславливаемых электрическими полями* — $\nabla \varphi_p$ и $-\nabla \varphi_n$ в *веществах с удельной проводимостью* $q\mu_p p$ и $q\mu_n n$.

Таким образом, компенсация двух конкурирующих процессов, иллюстрируемых фиг. 107, объясняется просто как следствие того, что φ_n и φ_p постоянны и равны φ , так что токи диффузии и проводимости равны и противоположны. В следующем параграфе будут рассмотрены $p-n$ -переходы, не находящиеся в условиях теплового равновесия, и показано, как можно использовать понятие о квази-уровнях Ферми при анализе происходящих в них процессов.

В качестве дополнительного примера, характеризующего изменения φ_n , φ_p и ψ , рассмотрим длинный однородный полупроводниковый стержень, по которому течет ток в направлении x ; предположим при этом, что на концах стержня имеются омические контакты, не изменяющие концентрацию дырок и электронов в полупроводнике. В этом случае разности $\psi - \varphi_n$ и $\psi - \varphi_p$, определяющие концентрации n и p , не зависят от координат и имеют те же значения, как и в условиях теплового равновесия, когда $\varphi_n = \varphi_p = \varphi$. Кривые, изображающие зависимость φ_n , φ_p и ψ от расстояния, имеют такой же вид, как и в случае равновесия, с той разницей, что к значениям φ_n , φ_p и ψ добавлено значение потенциала $-E_x x$. Уравнения (12.49) для плотностей тока сводятся к стандартному виду: $I_x = (q\mu_p p + q\mu_n n) E_x$; при этом изменения электростатического потенциала или напряжения равны смещению квази-уровней Ферми.

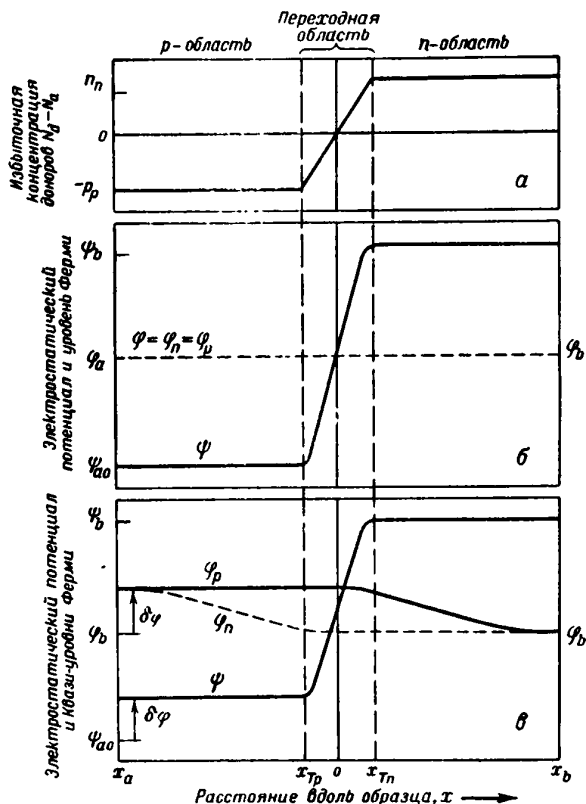
§ 5. ТЕОРИЯ $p-n$ -ПЕРЕХОДА ¹⁾

В этом параграфе рассмотрим упрощенную модель $p-n$ -перехода, и при ее анализе используем представление о квази-уровнях Ферми ²⁾. На фиг. 108 изображен ход значений φ_n , φ_p и ψ в $p-n$ -переходе в условиях теплового равновесия, а также при наложении на p -область, расположенную слева, смещающего потенциала $\delta\varphi$ в прямом направлении, в то время как *правая сторона*, n -область, поддерживается при *постоянном потенциале*. При этих условиях дырки вводятся

¹⁾ По этому вопросу см. [47—54]. — *Прим. ред.*

²⁾ Вопросы, изложенные в этом параграфе, подробно рассмотрены в работе [6]. Ссылки на дополнительную литературу приведены в этой статье, а также в § 2 гл. 4. — *Прим. авт.*

в n -область, а электроны — в p -область. Мы предполагаем, что однородные области столь широки, что введенные в них носители рекомбинируют, прежде чем они достигают сечений x_a и x_b . Вследствие



Фиг. 108. Упрощенная модель $p-n$ -перехода.

a — распределение доноров и акцепторов; b — распределение потенциала при тепловом равновесии; c — влияние потенциала $\delta\phi$, приложенного в прямом направлении.

этого в левой части образца полный ток ¹⁾ I переносится почти исключительно дырками, а в правой его части — почти исключительно электронами.

Таким образом, дырочный ток I_p должен равняться I при $x = x_a$ и постепенно спадать до нуля при $x = x_b$, в то время как электронный

¹⁾ Ниже мы предполагаем, что I , I_p и I_n — плотности токов в направлении x . — Прим. авт.

ток $I_n = I - I_p$ соответственно возрастает. Механизм изменения характера тока можно понять из рассмотрения дырочного тока, проходящего через сечение x_{Tn} ; этот ток уменьшается вследствие того, что дырки, обладающие в n -области определенным временем жизни τ_p , рекомбинируют с электронами, так что дырочный ток убывает по мере возрастания x . Аналогичный эффект имеет место и для тока электронов, входящих в p -область. В дальнейшем рассмотрении примем, что переходная область столь узка, что рекомбинацией в ней можно пренебречь; следовательно, I_p будет иметь одинаковое значение в сечениях x_{Tp} и x_{Tn} .

В сечениях x_a и x_b концентрации дырок и электронов имеют, таким образом, значения, характерные для однородного материала, и, следовательно, при $x = x_a$, например, φ_p и φ_n одинаковы и отличаются от ψ на ту же величину, как и в случае равновесия. Это объясняет, почему при построении рисунка мы положили, что

$$\varphi_p(x_a) = \varphi_n(x_a) = \varphi_p(x_b) + \delta\varphi = \varphi_n(x_b) + \delta\varphi \quad (12.50)$$

($\delta\varphi$ — приложенный потенциал), а также почему при $x = x_a$ имеем $\delta\psi = \delta\varphi_p = \delta\varphi_n$, а при $x = x_b$, соответственно, $\delta\psi = 0$. Предположим далее, что $\delta\varphi$ мало по сравнению с kT/q , так что наложение потенциала вызывает только ничтожные изменения n и p .

На основании этих предположений докажем теперь, что для $p-n$ -переходов значение φ_p почти постоянно в p -области и уменьшается на $\delta\varphi$ в n -области, тогда как φ_n почти постоянно в n -области и изменяется на величину $\delta\varphi$ в p -области. Для доказательства этого утверждения предположим, что проводимости обеих однородных областей $p-n$ -перехода значительно больше, чем собственная проводимость полупроводника, так что концентрации электронов в p -области и дырок в n -области, обозначенные через n_p и p_n , соответственно, много меньше, чем концентрации дырок в p -области, p_p , и электронов в n -области, n_n . Эта зависимость следует из условий

$$n_p = \frac{n_i^2}{p_p} \quad \text{и} \quad p_n = \frac{n_i^2}{n_n} \quad (12.51)$$

[см. § 2 и уравнение (12.41)].

Выразим теперь отношение значений $\nabla\varphi_p$ в сечениях x_{Tp} и x_{Tn} через концентрации n_p и p_n , используя при этом для плотности тока выражение (12.49а).

В сечении x_{Tp} имеем

$$I_p = -q\mu_p p_p \frac{d\varphi_p(x_{Tp})}{dx}. \quad (12.52a)$$

Так как рекомбинацией носителей тока в переходной области мы пренебрегаем, то тот же самый ток будет протекать и через

сечение x_{Tn} :

$$I_p = -q \mu_p p_n \frac{d\varphi_p(x_{Tn})}{dx}. \quad (12.526)$$

Из (12.52) находим

$$\frac{d\varphi_p(x_{Tn})}{dx} / \frac{d\varphi_p(x_{Tp})}{dx} = \frac{p_p}{p_n} = \frac{p_p n_n}{n_i^2}. \quad (12.53)$$

Отсюда следует, что φ_p изменяется значительно сильнее в n -области, чем в p -области, по крайней мере вблизи перехода. Такие же рассуждения применимы и к φ_n . Следовательно, $\nabla\varphi_n$ и $\nabla\varphi_p$ будут весьма велики там, где I_p приобретает заметную величину в n -области, или там, где I_n приобретает достаточную величину в p -области. Отсюда делаем заключение, что φ_p и φ_n изменяются существенным образом только в той области, где протекают токи, обусловленные введенными носителями заряда.

Этот существенный результат можно сформулировать и несколько иначе: *дырочный ток идет в образце, обладающем большой дырочной проводимостью в p -области и малой дырочной проводимостью в n -области*. Следовательно, градиент должен быть много больше в n -области.

Отсюда можно сделать важное заключение: *сопротивление p — n -перехода значительно больше обычного сопротивления материала, образующего переход¹⁾*.

Это обстоятельство следует из того факта, что дырочный ток, например, должен проходить некоторое расстояние в областях, где концентрация носителей заряда много меньше n_i , а в такой области градиент $\nabla\varphi_p$ будет значительно больше, чем он был бы даже в полупроводнике с собственной проводимостью; другими словами, сопротивление p — n -перехода может быть много больше сопротивления слоя полупроводника с собственной проводимостью, обладающего такой же толщиной, как и переходная область²⁾.

До тех пор пока концентрация введенных носителей тока не станет сравнимой с обычным значением концентрации носителей тока в образце, приведенные выше рассуждения будут оставаться в силе; существенные изменения значений φ_p и φ_n будут происходить лишь в областях с относительно малой концентрацией соответствующих носителей тока. В p -области значение p почти постоянно и равно p_p ; это следует из условия электрической нейтральности, требующего, чтобы изменения p были равны изменениям концентрации n в результате введения носителей в образец, а последние малы по сравне-

¹⁾ По этому вопросу см. [5]. — *Прим. ред.*

²⁾ Общий анализ этого вопроса содержится в работе [6]. — *Прим. авт.*

нию с p_p . Следовательно, в силу уравнения (12.486), значения ψ и φ_p в p -области остаются почти постоянными; аналогично, ψ и φ_n почти постоянны в n -области.

Все эти выводы, иллюстрируемые кривыми для ψ , φ_p и φ_n на фиг. 108, необходимы для получения совокупности граничных условий, которые могут быть затем учтены при исследовании потока дырок, входящего в n -область через плоскость x_{Tn} . Аналогичным способом можно получить выражения, характеризующие поток электронов, входящий в p -область через сечение x_{Tp} . Предположим, что все рассматриваемые величины не зависят от y и z . В соответствии с этим предположением, все токи текут параллельно оси x , и поэтому можно ограничиться рассмотрением лишь x -компонент. Для краткости будем в этом параграфе продолжать опускать индекс x и писать I_p вместо I_{px} .

В дальнейшем влиянием рекомбинации на переход от I_p к I_n в самой переходной области будем пренебрегать, так как предполагается, что последняя узка по сравнению с другими размерами, например диффузионными длинами L_p и L_n , рассматриваемыми ниже. Следовательно, можно считать, что I_p имеет одно и то же значение во всей области от $x = x_{Tp}$ до $x = x_{Tn}$. На этом основании плотность полного тока проводимости, проходящего через сечение $x = 0$, будет равна

$$I = I_n(x_{Tp}) + I_p(x_{Tn}). \quad (12.54)$$

Вычислим $I_p(x_{Tn})$, учтя при этом полученные выше граничные условия, а выражение для $I_n(x_{Tp})$ получим затем по аналогии. Рассмотрим случай прямого тока, соответствующий условию $\delta\varphi > 0$; полученные уравнения будут, однако, в одинаковой степени применимы и при условии $\delta\varphi < 0$, т. е. и для обратного тока.

На правой границе переходной области x_{Tn} будет выполняться соотношение

$$\varphi_p = \varphi_n + \delta\varphi. \quad (12.55)$$

Для удобства напомним еще раз определения различных величин для n -области:

$$\begin{aligned} n_n &= N_d - N_a \text{ — нормальная концентрация электронов,} \\ p_n &= \frac{n_i^2}{n_n} \ll n_n \text{ — нормальная концентрация дырок,} \\ \tau_p &\text{ — время жизни дырки в } n\text{-области.} \end{aligned}$$

Так как мы предположили, что концентрация введенных носителей тока мала, то граничные условия для сечения $x = x_{Tn}$ будут

иметь вид

$$n(x_{Tn}) \approx n_n = n_i e^{q(\psi - \varphi_n)/kT}, \quad (12.56a)$$

$$p(x_{Tn}) = n_i e^{q(\varphi_p - \psi)/kT}. \quad (12.566)$$

При замене множителя $\exp(-q\psi/kT)$ в (12.566) выражением (n_i/n_n) , согласно (12.56a), $\exp(-q\varphi_n/kT)$ дает

$$p(x_{Tn}) = \frac{n_i^2}{n_n} e^{q(\varphi_p - \varphi_n)/kT} = p_n e^{q\delta\varphi/kT} \quad (12.57)$$

Уравнение (12.57) является основным в теории выпрямления. Из него следует, что концентрация дырок в плоскости x_{Tn} изменяется экспоненциально с приложенным напряжением. В случае линейного приближения, которое мы вскоре рассмотрим подробно, следует, что дырочный ток должен изменяться пропорционально отклонению концентрации p от ее равновесного значения, а следовательно, как разность $p - p_n = p_n [\exp(q\delta\varphi/kT) - 1]$. Это сразу приводит нас к правильному виду характеристики постоянного тока, определяемой приведенным ниже уравнением (12.71).

Предположим, что $\delta\varphi$ состоит из двух компонент, отвечающих постоянному и переменному токам

$$\delta\varphi = V_0 + v_1 e^{i\omega t}, \quad (12.58)$$

где v_1 — сигнал переменного тока, предполагаемый столь малым, что можно воспользоваться линейным приближением (т. е. $v_1 \ll kT/q$). Разложение (12.57) в ряд дает

$$p(x_{Tn}) = (p_n e^{qV_0/kT}) \left[1 + \frac{qv_1}{kT} e^{i\omega t} \right].$$

Представим эту концентрацию в виде суммы трех слагаемых: ее равновесного значения p_n , компоненты p_0 , соответствующей постоянному току, и компоненты $p_1 \exp(i\omega t)$, соответствующей переменному току,

$$p(x_{Tn}) = p_n + p_0 + p_1 e^{i\omega t}, \quad (12.59)$$

где

$$p_0 = p_n (e^{qV_0/kT} - 1), \quad (12.60)$$

$$p_1 = \frac{qp_n v_1}{kT} e^{qV_0/kT}. \quad (12.61)$$

До тех пор пока $p(x_{Tn}) \ll n_n$, нормальной концентрации электронов в n -области, время жизни τ_p и коэффициент диффузии дырок D_p не будут существенно зависеть от $\delta\varphi$.

Наша задача заключается в том, чтобы определить значение $p(x, t)$ при $x > x_{Tn}$ и вычислить результирующий дырочный ток $I_p(x_{Tn}, t)$, протекающий через сечение x_{Tn} . Полный ток $I(t)$ равен сумме этого дырочного тока, аналогичного ему электронного тока и тока смещения $\dot{D}/4\pi = (\kappa/4\pi)\dot{E}$ (или $\dot{D} = \kappa\epsilon_0\dot{E}$ в системе MKS). В $p-n$ -переходах ток смещения вообще мал по сравнению с током проводимости, за исключением тех случаев, когда обратные смещающие напряжения велики, а также в случае низких температур¹⁾.

Как указывалось раньше, ток дырок, введенных в n -область, спадает с течением времени, причем этот процесс характеризуется временем релаксации или временем жизни дырки τ_p . В согласии со сказанным выше в § 2 [см. формулу (12.10)], уравнение непрерывности для этого случая имеет вид

$$\frac{\partial p}{\partial t} = \frac{p_n - p}{\tau_p} - \frac{1}{q} \nabla \cdot I_p. \quad (12.62)$$

Так как ψ предполагается постоянным по сравнению с Φ_p , то ток является в основном диффузионным током, причем его плотность равна

$$I_p = -qD_p \frac{\partial p}{\partial x}. \quad (12.63)$$

Подстановка (12.63) в уравнение (12.62) дает

$$\frac{\partial p}{\partial t} = \frac{p_n - p}{\tau_p} + D_p \frac{\partial^2 p}{\partial x^2}. \quad (12.64)$$

Решение этого уравнения при учете граничных условий в сечении x_{Tn} имеет вид

$$p = p_n + p_0 e^{-x'/L_p} + p_1 e^{i\omega t - x(1+i\omega\tau_p)^{1/2}/L_p}, \quad (12.65)$$

где

$$x' = x - x_{Tn}, \quad x' = 0 \quad \text{при} \quad x = x_{Tn}, \quad (12.66)$$

а диффузионная длина для дырок в n -области равна

$$L_p = \sqrt{D_p \tau_p}. \quad (12.67)$$

Понятие диффузионной длины имеет следующий физический смысл: предположим, что при некотором значении $x = x_1$ концентрация дырок равна $p(x_1)$, а напряженность электрического поля ничтожно мала. Анализ члена в выражении (12.65), обусловленного постоянным током, показывает, что, если справа от x_1 источники дырок отсутствуют, то p уменьшается в e раз на протяжении каждого интервала, длина которого равна L_p . К этому вопросу мы еще вернемся в дальнейшем при решении задач.

¹⁾ См. [6, § 4 гл. 4]. — Прим. авт.

Условие малости $\nabla\psi$ по сравнению с $\nabla\varphi_p$, в силу которого преобладающую роль играет диффузионный ток, имеет простое объяснение в том случае, если $p \gg p_n$. При этом условии можно записать: $p = p(x_1) \exp[-(x-x_1)/L_p]$. Согласно определению φ_p , приведенному в § 4 [см. (12.486)], это дает

$$\varphi_p = -\frac{kT(x-x_1)}{qL_p} + \frac{kT}{q} \ln \left[\frac{px_1}{n_i} \right] + \psi. \quad (12.68)$$

Диффузионный член будет играть преобладающую роль в выражении для тока, если в уравнении

$$\frac{\partial \varphi_p}{\partial x} = -\frac{kT}{qL_p} + \frac{\partial \psi}{\partial x} \quad (12.69)$$

первый член kT/qL_p много больше, чем $\partial\psi/\partial x$ (см. задачи).

Воспользовавшись только что полученным значением концентрации p , можно вычислить плотность тока в сечении x_{Tn} .

$$I_p(x_{Tn}) = -qD_p \frac{\partial p}{\partial x} = \frac{qp_0 D_p}{L_p} + \frac{qp_1 D_p e^{i\omega t} (1 + i\omega\tau_p)^{1/2}}{L_p}. \quad (12.70)$$

Компоненту плотности тока, отвечающую постоянному току, получим, подставляя (12.60) вместо p_0 ,

$$I_{p0}(x_{Tn}) = \frac{qp_n D_p}{L_p} (e^{qV_0/kT} - 1) \equiv I_{ps} (e^{qV_0/kT} - 1). \quad (12.71)$$

Прямое направление тока соответствует $V_0 > 0$. В этом направлении ток экспоненциально возрастает с напряжением. Для обратного направления тока экспоненциальный член обращается в нуль и значение I_{p0} приближается к значению *тока насыщения* ($qp_n D_p/L_p \equiv I_{ps}$. Сравнение выражения (12.71) [или, скорее, его более общей формы (12.81)] с опытом приводится в § 2 гл. 4, а также в конце этого параграфа.

Компоненту плотности тока, отвечающую переменному току, получим аналогичным образом, комбинируя член формулы (12.70), содержащий p_1 , и формулу (12.61):

$$\begin{aligned} I_{p1}(x_{Tn}) &= \frac{qp_n^2 p}{L_p} e^{qV_0/kT} (1 + i\omega\tau_p)^{1/2} v_1 e^{i\omega t} \\ &\equiv (G_p + iS_p) v_1 e^{i\omega t} \equiv A_p v_1 e^{i\omega t}. \end{aligned} \quad (12.72)$$

Здесь A_p — полная дырочная проводимость (на единицу площади) для дырок, диффундирующих в n -область; ее вещественная часть G_p и мнимая часть S_p — соответственно активная и реактивная проводимости. Если $\omega\tau_p$ мало, то вещественный член есть просто активная проводимость единицы площади слоя, толщина которого равна L_p , а

$$qp_n e^{qV_0/kT} \frac{p}{L_p} = \frac{q(p_n + p_0) p}{L_p}. \quad (12.73)$$

В том случае, когда напряжение смещения равно нулю, этот результат подтверждает полученный уже ранее вывод, заключающийся в том, что сопротивление дырочному току обусловлено малой дырочной проводимостью $q\rho_n\mu_p$ в n -области.

Выражение (12.73) представляет собой также дифференциальную проводимость, определяемую путем дифференцирования уравнения (22.71) по V_0 .

В этом параграфе мы предполагаем, что время жизни τ_p определяется лишь объемной рекомбинацией. В образцах, поперечные размеры которых в направлениях y и z сравнимы с L_p или L_n , преобладающую роль может играть поверхностная рекомбинация. Можно показать, что в этом случае применима та же теория, если только в качестве τ_p берется действительное время жизни носителей тока в образце [6]. В следующем параграфе рассмотрен вопрос о времени жизни дырки в нитевидных образцах германия.

Совершенно аналогичные выражения можно получить и для электронных токов.

Коэффициенты $q\rho_n\mu_p/L_p$ и $qn_p\mu_n/L_n$, зависящие от свойств материала, могут быть выражены через проводимость n -области σ_n , проводимость p -области σ_p , собственную проводимость σ_i и отношение подвижностей

$$b = \frac{\mu_n}{\mu_p}. \quad (12.74)$$

Выражение для собственной проводимости будет иметь вид

$$\sigma_i = q\mu_p n_i (1 + b), \quad (12.75)$$

и если $p_n \ll n_n$ и $n_p \ll p_p$, то

$$q\mu_p p_n = \frac{b\sigma_i^2}{(1+b)^2} \sigma_n, \quad (12.76)$$

$$q\mu_n n_p = \frac{b\sigma_i^2}{(1+b)^2} \sigma_p. \quad (12.77)$$

Используя эти уравнения, можно переписать выражение (12.71) и соответствующее уравнение для электронного тока, текущего в p -область, так, чтобы определить их зависимость от постоянного смещающего напряжения V_0 и свойств материала в соответствующих областях:

$$\begin{aligned} I_{p0}(V_0) &= \frac{b\sigma_i^2}{(1+b)^2 \sigma_n L_p} \frac{kT}{q} (e^{qV_0/kT} - 1) \equiv G_{p0} \frac{kT}{q} (e^{qV_0/kT} - 1) \\ &\equiv I_{ps} (e^{qV_0/kT} - 1), \quad (12.78) \end{aligned}$$

$$\begin{aligned} I_{n0}(V_0) &= \frac{b\sigma_i^2}{(1+b)^2 \sigma_p L_n} \frac{kT}{q} (e^{qV_0/kT} - 1) \equiv G_{n0} \frac{kT}{q} (e^{qV_0/kT} - 1) \\ &\equiv I_{ns} (e^{qV_0/kT} - 1). \quad (12.79) \end{aligned}$$

Значения G_{p0} и G_{n0} [легко видеть, что это значения активных проводимостей при низких частотах и низких напряжениях ($V_0 \ll kT/q$)] и токи насыщения, текущие в обратном направлении, определяются формулами

$$G_{p0} = \frac{b\sigma_i^2}{(1+b)^2 \sigma_n L_p} \equiv \frac{q}{kT} I_{ps}, \quad (12.80a)$$

$$G_{n0} = \frac{b\sigma_i^2}{(1+b)^2 \sigma_p L_n} \equiv \frac{q}{kT} I_{ns}. \quad (12.80b)$$

Выражение для тока в пропускном направлении приобретает вид

$$\begin{aligned} I_0(V_0) &= (G_{p0} + G_{n0}) \frac{kT}{q} (e^{qV_0/kT} - 1) = \\ &= (I_{ps} + I_{ns}) (e^{qV_0/kT} - 1). \end{aligned} \quad (12.81)$$

Эта зависимость представлена графически на фиг. 42 (стр. 121). Там же приведено ее сопоставление с экспериментальными данными.

Воспользовавшись выражением для G_{p0} , можно записать значение полной проводимости для дырочного тока, рассчитанное на единицу площади, в комплексной форме

$$A_p = G_p + iS_p = (1 + i\omega\tau_p)^{1/2} G_{p0} e^{qV_0/kT}. \quad (12.82)$$

При низких частотах, когда ω много меньше $1/\tau_p$, можно разложить первый из множителей выражения в ряд:

$$G_p + iS_p = G_{p0} e^{qV_0/kT} + i\omega \frac{\tau_p}{2} G_{p0} e^{qV_0/kT}. \quad (12.83)$$

Отсюда следует, что величина $(\tau_p/2) G_{p0} \exp(qV_0/kT)$ играет роль емкости на единицу площади.

Полезно выяснить, что представляет собой эта емкость, когда смещающее напряжение $V_0 = 0$. Для этого случая находим

$$C_p = \frac{\tau_p G_{p0}}{2} = \frac{\tau_p q p_n \mu_p}{2L_p} = \frac{q^2 p_n L_p}{2kT}. \quad (12.84)$$

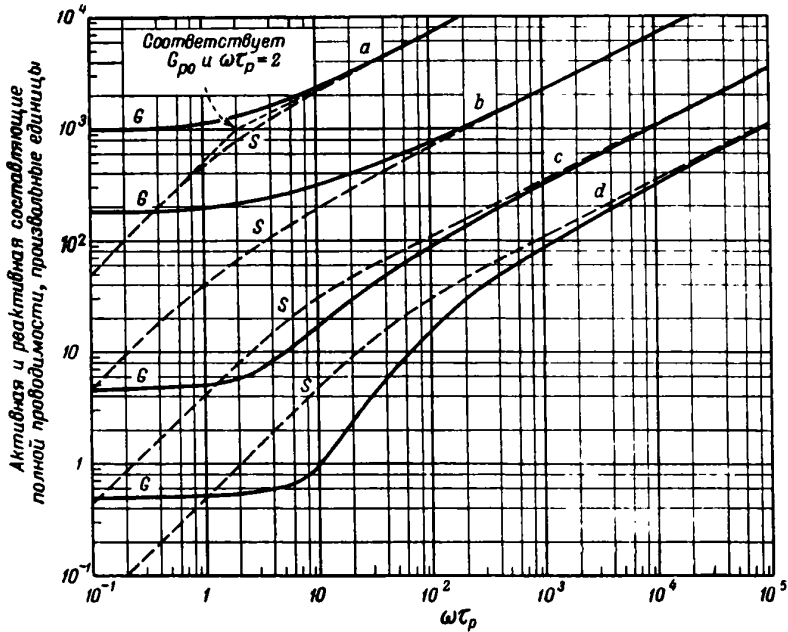
Последнее выражение получается с учетом того, что $\tau_p \mu_p = q\tau_p D_p/kT = qL_p^2/kT$ и имеет простой физический смысл: $q p_n L_p$ есть полный заряд дырок в слое толщиной L_p . При небольшом изменении напряжения v плотность этого заряда должна была бы измениться на (qv/kT) , так что изменение заряда, деленное на изменение напряжения v , будет равно $(q/kT)(q p_n L_p)$. Последнее выражение отличается от выражения для C_p только множителем 2.

При низких частотах величина A_p ведет себя как активная проводимость и емкость, включенные параллельно; при высоких частотах

эта величина ведет себя, однако, совершенно иначе. На фиг. 109 изображен ход функции $(1 + i\omega\tau)^{1/2} = A_p/G_{p0}$. При высоких частотах G_p и S_p равны

$$G_p = S_p = \sqrt{\frac{\tau_p}{2}} G_{p0} \sqrt{\omega} = \frac{bc_1^2 \sqrt{\omega}}{(1+b)^2 \sigma_n \sqrt{2D}}. \quad (12.85)$$

Таким образом, при высоких частотах полная проводимость не зависит от τ_p и определяется только диффузией дырок в n -область и из нее.



Фиг. 109. Активная и реактивная составляющие G и S полной проводимости для потока дырок в n -область.

$a. - 10^6 A_p/G_{p0} = 10^6 (1 + \omega\tau_p)^{1/2}$; b, c и d — соответствуют случаям, в которых свойства n -области изменяются [6].

На фиг. 109 три прямолинейные асимптоты пересекаются в точке $(G_{p0}, \omega\tau = 2)$ — обстоятельство, полезное при определении значения τ на основании таких данных.

Значения величины $(I_{p0} + I_{n0})$, определенные по данным о проводимости материалов, образующих $p-n$ -переход, а также значения τ , которое есть своего рода среднее от τ_p и τ_n , в пределах точности рассматриваемого приближения, находятся в согласии с данными, полученными из статической вольтамперной характеристики выпрямителя,

Во время написания этой книги были произведены опыты по непосредственному определению L_p и L_n и исследованию фотопроводимости с целью определения значений составляющих тока I_p и I_n в отдельности. Сопоставление опытных данных с формулой (12.81) было проведено при анализе фиг. 42.

§ 6. ВРЕМЯ ЖИЗНИ НОСИТЕЛЕЙ ЗАРЯДА В НИТЕВИДНЫХ ОБРАЗЦАХ ГЕРМАНИЯ И ЭФФЕКТ СУЛЯ

Введение. Ряд опытов показал, что процесс рекомбинации носителей заряда в нитевидных образцах германия часто происходит по преимуществу на поверхности и что эта рекомбинация может быть усилена при одновременном наложении на образец продольного электрического и поперечного магнитного полей. Такое усиление рекомбинации на поверхности, как следствие образующейся там высокой концентрации введенных носителей, называется *эффектом Суля*, рассмотренным кратко в § 2 гл. 3.

В настоящем параграфе решено уравнение непрерывности для случая нитевидного образца германия прямоугольного сечения, к которому приложено продольное электрическое поле, и показано, как коэффициенты поверхностной и объемной рекомбинации входят в выражение для времени жизни и *постоянной спада*, характеризующей скорость убывания рассматриваемой величины (обратная времени жизни). Рассмотрим эффект Суля лишь в некотором ограниченном аспекте, так как более полный анализ этого явления, хотя и не представляет затруднений, потребовал бы значительного времени.

В этом параграфе мы ограничимся рассмотрением случая, когда *концентрация дырок, введенных в материал с проводимостью n -типа, мала*. Применяя к этому случаю те же рассуждения, что и в последнем параграфе, приходим к выводу, что поскольку изменения концентрации n малы, то

$$\psi - \varphi_n = \frac{kT}{q} \ln \frac{n_n}{n_i} = \text{const.} \quad (12.86)$$

Токи электронов и дырок, текущие в нитевидном образце в поперечном направлении, равны и противоположно направлены, так как в противном случае на его поверхности образовался бы заряд. Следовательно, отношение градиентов φ_n и φ_p должно быть равно отношению концентраций p/n [см. (12.49)]. Поэтому значения φ_n и ψ в поперечном сечении нитевидного образца изменяются мало, и дырочный ток в значительной степени осуществляется диффузией.

К такому же заключению можно придти гораздо более простым путем. Предположим, что в образец введено лишь несколько дырок. Они создадут небольшое электрическое поле. Электроны будут втягиваться в область этого поля и нейтрализовать заряд дырок. Однако концентрация электронов много больше концентрации дырок, так что

этот процесс не будет сколько-нибудь заметно изменять распределение электронов. Поле, создаваемое введенными дырками, должно быть поэтому так мало, чтобы электроны двигались почти так же, как если бы дырок не было вовсе. Следовательно, в первом приближении можно считать, что дырки движутся как раз в таком поле, которое существовало бы в образце при полном отсутствии дырок. Влияние электронов, поскольку речь идет о дырках, сводится просто к нейтрализации их заряда, так что в результате взаимное отталкивание дырок не имеет места.

Рассмотрим влияние поверхностной рекомбинации на время жизни в двух случаях. В первом случае (назовем его случаем τ_{p1}) представим себе, что распределение дырок вдоль образца совершенно равномерно. Такое распределение может быть, например, получено при кратковременном освещении нитевидного образца рентгеновскими лучами. После этого покажем, что концентрация дырок уменьшается со временем — независимо от присутствия электрического поля — согласно закону

$$\exp(-t/\tau_{p1}), \quad (12.87)$$

где τ_{p1} — время жизни дырки в нитевидном образце. Время жизни, по определению, есть время, в течение которого концентрация убывает в e раз.

Во втором случае (так называемый случай τ_{p2}) представим себе, что дырки все время вводятся в некоторой точке нитевидного образца и движутся вдоль него. По мере продвижения дырок вдоль образца концентрация их убывает вследствие рекомбинации. В этом случае в рассмотрение вводится коэффициент затухания a и концентрация дырок изменяется как e^{-ax} , где a — коэффициент затухания.

Определим время жизни как время, в течение которого дырка под действием поля проходит расстояние, равное $1/a$, на котором концентрация уменьшается в e раз. Таким образом,

$$\tau_{p2} = \frac{1/a}{\mu_p E_x} = \frac{1}{a\mu_p E_x}. \quad (12.88)$$

Далее будет показано, что в сильных электрических полях, когда движение дырок вдоль образца определяется главным образом влиянием электрического поля E_x , значения τ_{p1} и τ_{p2} равны. Покажем, что удобнее иметь дело не с временами жизни, а с *постоянными спада*ния, определяемыми как $1/\tau_{p1}$ и $1/\tau_{p2}$. Так как эффекты поверхностной и объемной рекомбинаций аддитивны, то мы можем написать

$$\frac{1}{\tau_{p1}} = \frac{1}{\tau_p} + v_s, \quad (12.89)$$

где $1/\tau_p$ — объемная постоянная спада и v_s — поверхностная постоянная спада,

Уравнение непрерывности и граничные условия. Таким образом, мы подошли к исследованию решения уравнения непрерывности [см. (12.10)] для дырок в нитевидном образце германия при наличии однородного электрического поля, направленного по оси x вдоль образца. Предположим, что границам нитевидного образца соответствуют координаты $y = \pm B$ и $z = \pm C$, так что поперечное сечение образца имеет прямоугольную форму и обладает площадью, равной $4BC$. Составляющие дырочного тока, протекающего в образце, равны

$$I_x = q\mu_p p E_x - qD_p \frac{\partial p}{\partial x}, \quad (12.90a)$$

$$I_y = -qD_p \frac{\partial p}{\partial y}, \quad (12.90б)$$

$$I_z = -qD_p \frac{\partial p}{\partial z}. \quad (12.90в)$$

Они должны удовлетворять уравнению непрерывности

$$\begin{aligned} \frac{\partial p}{\partial t} &= \frac{p_n - p}{\tau_p} - \frac{1}{q} \nabla \cdot I_p = \\ &= \frac{p_n - p}{\tau_p} - \mu_p p E_x \frac{\partial p}{\partial x} + D_p \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \right). \end{aligned} \quad (12.91)$$

Очевидно, что значение $p = p_n$ есть решение этого уравнения для того случая, когда дырки в нитевидный образец не вводятся. Обозначим изменение концентрации дырок в результате их введения в образец как

$$p_1 = p - p_n. \quad (12.92)$$

В таком случае (12.91) будет однородным линейным дифференциальным уравнением в частных производных, так как каждый член его зависит от p_1 линейно. Кроме того, переменные разделяются, и решения его могут быть записаны как произведения функций, зависящих от t , x , y и z в отдельности. Одним из таких решений будет

$$p_1 = e^{-\nu t - ax} \cos by \cos cz. \quad (12.93)$$

Для того чтобы удовлетворялось уравнение непрерывности (12.91), постоянные ν , a , b и c должны удовлетворять соотношению

$$\nu - \frac{1}{\tau_p} + \mu_p E_x a + D_p (a^2 - b^2 - c^2) = 0. \quad (12.94)$$

Наиболее общее решение уравнения (12.91), соответствующее произвольным начальным условиям при $t=0$, может быть представлено как сумма ряда членов вида (12.93); в общем случае в этот ряд входят также члены, содержащие синусоидальные функции от переменных y и z , а коэффициент a может принимать комплексные или мнимые значения,

В первом из рассмотренных выше случаев (случай τ_{p1}), когда нитевидный образец германия освещается одинаково со всех сторон, так что концентрация p_1 имеет повсюду одинаковое значение, достаточно учесть только симметричные члены, содержащие косинусы. Необходимо, однако, взять ряд таких членов, чтобы их сумма определяла концентрацию дырок при $t=0$. Изменение каждого члена такого рода с течением времени будет при этом характеризоваться своим собственным временем жизни, так что уменьшение общего числа дырок в нитевидном образце не следует простому экспоненциальному закону, а описывается суммой экспоненциально убывающих членов с различными временами жизни. Возникает, естественно, вопрос, что следует подразумевать в этом случае под временем жизни дырки в нитевидном образце?

Ответ заключается в том, что один из членов вида (12.93) характеризуется настолько большим значением времени жизни по сравнению с остальными членами, что играет доминирующую роль.

Именно это наибольшее время жизни мы и будем в основном рассматривать в дальнейшем. Остальные члены убывают настолько быстро, что скорость наблюдаемого процесса определяется только одним упомянутым выше членом ряда.

Аналогичное рассуждение применимо и ко второму случаю (случай τ_{p2}). Можно себе представить, что при $x=0$ распределение дырочного тока в поперечном сечении образца описывается суперпозицией многих членов вида $\cos by \cos cz$, а также членов, содержащих синусоидальные функции. Все эти члены, за исключением одного, убывают очень быстро, по мере того как дырки перемещаются вдоль по нитевидному образцу под действием поля, в то время как один из членов, как мы покажем ниже, характеризуется временем жизни τ_{p1} .

[Задача разложения произвольной функции переменных x , y и z в ряд по собственным функциям, удовлетворяющим уравнению (12.91) и граничным условиям на поверхности, аналогична многим другим задачам математической физики¹⁾. Решение этой задачи для случая тока дырок через $p-n$ -переход в нитевидном образце квадратного сечения было уже опубликовано²⁾.]

Предположим, что на границах образца $y = \pm B$ и $z = \pm C$ избыток числа рекомбинирующих над числом образующихся дырок прямо пропорционален концентрации p_1 , причем коэффициент пропорциональности равен s . Следовательно, ток по направлению к поверхности образца удовлетворяет условиям

$$\frac{1}{q} I_y = \pm s p_1, \quad y = \pm B, \quad (12.95a)$$

$$\frac{1}{q} I_z = \pm s p_1, \quad z = \pm C. \quad (12.95b)$$

¹⁾ Аналогичный квантовомеханический случай рассмотрен в § 4 гл. 14. См. также [8—10]. — *Прим. авт.*

²⁾ См., например, [6]. — *Прим. авт.*

Величина s имеет размерность скорости; мы будем называть ее *скоростью поверхностной рекомбинации*. Из условий (12.95) следует, что скорость рекомбинации как раз равна току дырок с плотностью p_1 , протекающему со средней скоростью s через поверхность образца.

Если p_1 задается выражением (12.93), то граничные условия при $y = \pm B$ и $z = \pm C$ эквивалентны и выражения (12.95) сводятся к

$$bB \operatorname{tg} bB = \frac{sB}{D_p} \equiv \chi_y = \eta \operatorname{tg} \eta, \quad (12.96a)$$

$$cC \operatorname{tg} cC = \frac{sC}{D_p} \equiv \chi_z = \zeta \operatorname{tg} \zeta, \quad (12.96b)$$

где

$$\eta \equiv bB \quad \text{и} \quad \zeta \equiv cC. \quad (12.97)$$

Для любого значения χ_y существует лишь одно решение для значений η , лежащих в интервале $0 \leq \eta \leq \pi/2$, а также по одному решению для каждого интервала от $n\pi$ до $(n+1/2)\pi$. Это справедливо и для любых значений χ_z и ζ . Член, характеризующий наибольшим значением времени жизни, как мы увидим ниже, соответствует наименьшим численным значениям решений уравнений (12.96), которые мы обозначили через η_0 и ζ_0 . Коэффициенты b и c равны при этом соответственно η_0/B и ζ_0/C . Уравнение (12.94) можно рассматривать в таком случае как уравнение с двумя неопределенными параметрами a и v , которое может быть преобразовано к виду

$$v + D_p a^2 + \mu_p E_x a = \frac{1}{\tau_p} + D_p \left(\frac{\eta_0^2}{B^2} + \frac{\zeta_0^2}{C^2} \right) \equiv v_f. \quad (12.98a)$$

Члены второго из этих выражений определяются постоянными, характеризующими нитевидный образец, т. е. величинами τ_p , B , C и s . Ниже будет показано, что сумма этих членов, обозначенная через v_f , есть постоянная спадания для нитевидного образца и равна $1/\tau_{p1}$ и $1/\tau_{p2}$. Основываясь на этом равенстве, можно ввести новое понятие — *время жизни в нитевидном образце*, определяемое формулой

$$\tau_f = \frac{1}{v_f}. \quad (12.98b)$$

На фиг. 110 показано распределение концентрации дырок, введенных в нитевидный образец, для решений η_0 , ζ_0 в плоскости $z=0$. Этот рисунок показывает, как меняется p_1 с изменением y для трех различных значений χ . При $s=0$ распределение концентрации по сечению однородно, так как теперь для компенсации поверхностной рекомбинации нет необходимости в диффузии дырок к поверхности. При $s=\infty$ концентрация дырок на поверхности равна 0, и вследствие большого градиента концентрации возникает поток дырок к поверхности, где они и рекомбинируют. При промежуточном значении χ величина p_1 имеет на поверхности конечное значение и градиент

концентрации как раз достаточно велик, чтобы диффузия дырок к поверхности уравнивала их рекомбинацию.

Рассмотрим сначала первый случай (случай τ_{p1}), когда концентрация дырок не зависит от x , так что $a = 0$. В этом случае (12.98а) сводится к

$$v = v_f \quad (12.99)$$

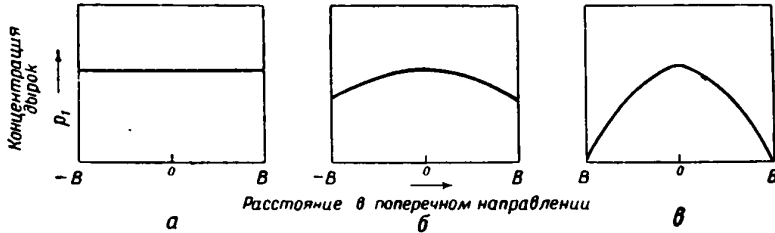
независимо от значения напряженности электрического поля E_x . Решение p_1 может быть записано при этом как

$$p_1 = e^{-v_f t} \cos by \cos cz. \quad (12.100)$$

Это значит, что концентрация дырок убывает экспоненциально со скоростью, характеризуемой постоянной спада v_f , которая равна $1/\tau_{p1}$, в согласии с определением (12.87), так что

$$\frac{1}{\tau_{p1}} = v_f = \frac{1}{\tau_p} + D_p \left(\frac{\eta_0^2}{B^2} + \frac{\zeta_0^2}{C^2} \right) = \frac{1}{\tau_p} + v_s, \quad (12.101)$$

откуда следует, что постоянная спада v_f для нитевидного образца равна сумме „объемной“ постоянной спада $1/\tau_p$ и поверхностной постоянной спада v_s , определяемой уравнением (12.101).



Фиг. 110. Распределение концентрации дырок, введенных в нитевидный образец германия.

$a: \chi = 0, s = 0, \eta_c = 0; \quad б: \chi = (\pi/4)^2, \eta_c = \pi/4; \quad в: s = \infty, \eta_0 = \pi/2.$

Теперь можно показать, почему во многих случаях можно пренебречь рассмотрением членов, содержащих η_n и ζ_n со значением $n > 0$. Для этих членов значения η_n^2 и ζ_n^2 будут по меньшей мере в четыре раза больше, чем η_0^2 и ζ_0^2 , следовательно, такие члены будут уменьшаться быстрее, чем член, содержащий η_0 и ζ_0 , и будут ничтожно малы по сравнению с ним, по крайней мере во всех тех случаях, когда v_s достаточно велико. В формулу, определяющую распределение концентрации дырок по сечению образца, эти члены высшего порядка входят с переменными знаками, а быстрое убывание концентрации является следствием течения избыточных дырок в те области, где их недостает. Форма распределения концентрации дырок для случая $s = \infty$, соответствующего значению $p = 0$ на границе,

сходна с формой прямоугольной барабанной мембраны, изображенной на фиг. 51 (стр. 157). Фиг. 51, e соответствует члену, содержащему η_0 и ζ_0 . Другие части фиг. 51 соответствуют тем случаям, когда в выражение для p_1 , помимо членов, содержащих косинусы, входят также члены, содержащие синусоидальные функции; значения η и ζ для последних удовлетворяют уравнениям вида $\eta \operatorname{ctg} \eta = \chi$. В дальнейшем этими членами высшего порядка мы будем пренебрегать. Значения v_s определяются выражением

$$v_s = D_p \left[\frac{\eta_0^2}{B^2} + \frac{\zeta_0^2}{C^2} \right], \quad (12.102)$$

где v_s — функция D_p , B , C и s , причем s входит в рассмотрение через зависимости η_0 и ζ_0 от $\chi_y = sB/D_p$ и $\chi_s = sC/D_p$. Это выражение принимает простой вид при больших и малых значениях s . При $s \rightarrow \infty$ как η_0 , так и ζ_0 равны $\pi/2$, что соответствует равенству $p_1 = 0$ на границе. В этом случае

$$v_s = \frac{\pi^2 D_p}{4} \left(\frac{1}{B^2} + \frac{1}{C^2} \right), \quad s \rightarrow \infty. \quad (12.103a)$$

При малых значениях s выражение для v_s также принимает простой вид

$$v_s = s \left(\frac{1}{B} + \frac{1}{C} \right), \quad s \rightarrow 0. \quad (12.103b)$$

Для нитевидного образца с квадратным сечением величина v_s может быть представлена как функция трех переменных s , B и D_p при помощи уравнения

$$\frac{v_s B^2}{D_p} = 2 [\eta_0(\chi_y)]^2 = 2 \left[\eta_0 \left(\frac{sB}{D_p} \right) \right]^2, \quad (12.104)$$

где $\eta_0(\chi_y)$ — решение уравнения (12.96a).

Эта функциональная зависимость изображена на фиг. 111.

График, приведенный на фиг. 111, в сочетании с уравнением (12.101), можно использовать для вычисления значений τ_r по заданным размерам нитевидного образца B , „объемного“ времени жизни τ_p и коэффициента поверхностной рекомбинации s . Этот график может быть также использован для определения s по известным значениям τ_p и τ_r . Подсчеты такого рода были произведены рядом исследователей. В табл. 10 приведены приближенные значения коэффициента s , полученные для различно обработанных поверхностей германия.

Покажем, что такое же значение τ_r получается и при исследовании зависимости концентрации дырок от расстояния вдоль нитевидного образца в присутствии сильного электрического поля. Расстояния вдоль образца можно выразить через время, воспользовавшись для этой цели определением скорости дрейфа как $E_{x\mu_p} = v_x$. В этом случае решение является решением для *стационарного состояния*, не зависящим от времени, так что в (12.98) следовало бы положить равным нулю.

Таблица 10

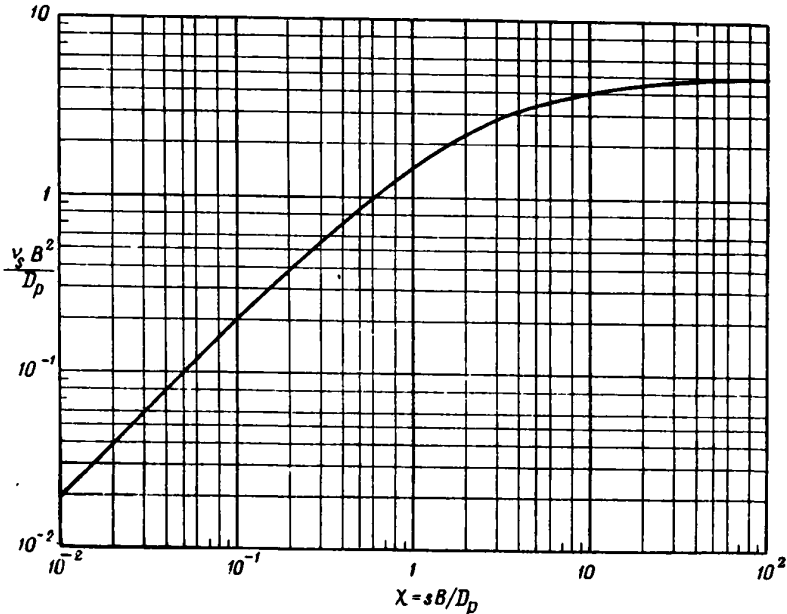
КОЭФФИЦИЕНТЫ ПОВЕРХНОСТНОЙ РЕКОМБИНАЦИИ В ГЕРМАНИИ

Способы обработки поверхности	Коэффициент рекомбинации, см/сек
Поверхность, обработанная пескоструйным аппаратом	~ 10 ⁴
Поверхность, подвергнутая травлению	~ 2000
Специальная обработка поверхности	~ 200

В этом случае значение a будет равно

$$a = \frac{-v_x \pm \sqrt{v_x^2 + 4D_p\gamma_f}}{2D_p} \quad (12.105)$$

Интересующему нас решению соответствует знак плюс. Знак минус соответствует уменьшению концентрации дырок в направлении отрицательной полуоси x или диффузии дырок против поля (см. задачи).



Фиг. 111. Зависимость постоянной v_s от скорости рекомбинации на поверхности s .

При больших значениях $v_x^2/D_p\gamma_f$, встречающихся в опытах с нитевидными образцами германия, можно разложить подкоренное выражение

в (12.105) в ряд или пренебречь членом, содержащим D_p в выражении (12.98а), тогда

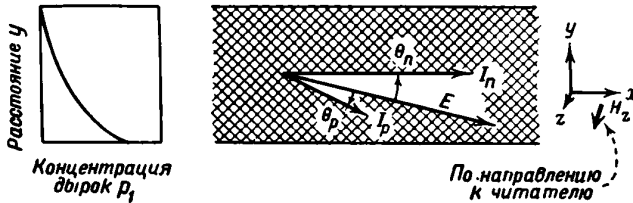
$$a \approx \frac{v_f}{v_x}. \quad (12.106)$$

Отсюда τ_{p2} , определяемое формулой (12.88), равно

$$\tau_{p2} = \frac{1}{av_x} = \frac{1}{v_f} = \tau_f. \quad (12.107)$$

Таким образом τ_{p2} , так же как и τ_{p1} , равно τ_f [см. (12.98б)].

Эффект Суля. Рассмотрим теперь совместное влияние электрического и магнитного полей на время жизни введенных в германий дырок. При этом снова можно пренебречь влиянием введенных дырок



Фиг. 112. Увеличение концентрации дырок вблизи поверхности вследствие эффекта Суля.

на электрическом поле. Однако этот случай отличается двумя особенностями от случая, рассмотренного ранее. Теперь мы имеем дело с поперечным электрическим полем, обусловленным эффектом Холла. Ограничимся малыми значениями магнитного поля H_z , так что учитывать будем лишь линейные члены. При малых значениях напряженности магнитного поля H_z напряжение поперечного электрического поля $E_y = -\theta_n E_x$, как это было показано в гл. 8. Кроме того, на фиг. 112 вектор скорости дрейфа повернут по отношению к направлению поля E на угол Холла θ_p . Наконец, направление диффузионного тока образует угол θ_p с направлением градиента концентрации [см. (12.31) и (12.32)]. В рассматриваемом ниже случае вследствие вращения тока диффузии возникает эффект, зависящий от $\cos(\theta_p + \theta_n)$. При предполагаемой малости углов этот эффект ничтожен.

Если составляющая электрического поля E_y достаточно велика, то поле поперек образца будет настолько сильным, что в основном все дырки сконцентрируются на одной стороне нитевидного образца. Мы ограничимся рассмотрением этого случая (назовем его *сильным эффектом Суля*) и выберем координаты так, чтобы начало отсчета значений y лежало на той стороне образца, по направлению к которой отклоняются дырки. Этот случай показан на фиг. 112.

Для краткости ограничимся рассмотрением случая, аналогичного случаю τ_{p1} , и предположим, что концентрация p_1 зависит только от y и t ; отсутствие зависимости p_1 от z соответствует бесконечно большим размерам образца в направлении z . Эти предположения позволяют пренебречь составляющей дырочного тока по оси x , так как эта составляющая не зависит от x и не определяет величины дивергенца. Составляющая дырочного тока по оси y будет в таком случае равна

$$I_y(y) = -q\mu_p p_1 E_x (\theta_n + \theta_p) - qD_p \frac{\partial p_1}{\partial y}; \quad (12.108)$$

при $y=0$, что соответствует нижней границе, по направлению к которой сносятся дырки, она должна удовлетворять граничному условию

$$-I_y(0) = qsp_1(0). \quad (12.109)$$

Покажем, что концентрация введенных дырок определяется при этом выражением

$$p_1 = e^{-v_t - by}. \quad (12.110)$$

Как показано на фиг. 112, это выражение убывает в направлении $+y$ и представляет собой концентрацию дырок вблизи одной из поверхностей образца. Для сокращения введем обозначение

$$v_t = \mu_p E_x (\theta_p + \theta_n) = \mu_p E_t, \quad (12.111)$$

где v_t — скорость дрейфа дырки в поперечном направлении, которой она обладала бы в случае, рассмотренном на фиг. 112. Таким образом, E_t — «эффективное поперечное поле». Подставляя p_1 и v_t в предыдущие уравнения, получим следующее выражение для граничного условия (12.109):

$$v_t - D_p b = s \quad (12.112)$$

или

$$b = \frac{v_t - s}{D_p}. \quad (12.113)$$

Если $v_t > s$, то дырки накапливаются вблизи поверхности $y=0$, так как они переносятся вниз быстрее, чем успевают рекомбинировать. Рассмотрим только такой случай, когда наблюдается сильный эффект Суля. Случаи, когда $v_t = s$ и $v_t < s$, являются промежуточными, соответствующими значениям H_z , лежащим между $H_z=0$ и большими значениями H_z , и носят гораздо более сложный характер.

Градиент концентрации (см. фиг. 112) обуславливает диффузию дырок от поверхности образца и, таким образом, противоположен направлению градиента на фиг. 110, соответствующему диффузии дырок к поверхности. В первом случае диффузия дырок от поверхности связана с тем, что $v_t > s$. Величина s имеет размерность скорости, и при $v_t > s$ дырки сносятся к поверхности полем E_t столь

быстро, что успевают там накапливаться, и при этом возникает диффузионный ток в обратном направлении.

Значение v определяется путем подстановки выражения (12.110) для p_1 в уравнение непрерывности (12.10)

$$\frac{\partial p_1}{\partial t} = -\frac{p_1}{\tau_p} - \frac{1}{q} \nabla \cdot I_p, \quad (12.114)$$

что дает

$$\begin{aligned} v &= \frac{1}{\tau_p} + v_t b - D_p b^2 = \frac{1}{\tau_p} + (v_t - D_p b) b = \\ &= \frac{1}{\tau_p} + sb \equiv \frac{1}{\tau_p} + v_s. \end{aligned} \quad (12.115)$$

Член $v_s = sb$, характеризующий влияние рекомбинации на поверхности на постоянную спада, как мы видим, прямо пропорционален коэффициенту b ; последний, в свою очередь, быстро возрастает пропорционально

$$E_t = (\theta_p + \theta_n) E_x = 10^{-8} (\mu_{E_p} + \mu_{E_n}) H_z E_x \quad (12.116)$$

(в обозначениях § 8 гл. 8 для практической системы единиц).

Равенство $v_s = sb$ [см. (12.115)] имеет относительно простой физический смысл. Решение

$$p_1 = e^{-vt-by} \quad (12.117)$$

позволяет определить полное число дырок на единицу поверхности в плоскости xz

$$\int_0^{2B} e^{-vt-by} dy = \frac{e^{-vt}}{b} (1 - e^{-2Bb}). \quad (12.118)$$

Величина $2Bb$ есть эффективное напряжение поперечного электрического поля, обусловленного действием магнитного поля, выраженное в единицах kT/q . При $v_t \gg s$ можно записать

$$2Bb = \frac{2B\mu_p E_t}{D_p} = \frac{2qBE_t}{kT}. \quad (12.119)$$

В экспериментальных условиях влияние члена, содержащего v_t , на величину $2Bb$ может быть столь велико, что членом $\exp(-2Bb)$ можно пренебречь. (Если этим членом пренебречь нельзя, то следует учесть влияние границы $y = 2B$.) Скорость рекомбинации на поверхности равна

$$p_1 s = s e^{-vt}, \quad (12.120)$$

т. е. просто равна произведению sb на число дырок, определяемое выражением (12.118). Так как постоянная спада есть отношение скорости убывания числа дырок к их полному числу, то это показывает, что $sb = v_s$.

Более просто этот результат можно сформулировать следующим образом: *в сильных поперечных полях движение дырок ограничивается слоем с эффективной толщиной, равной $kT/qE_t \approx 1/b$* , так что их полное число на единицу площади равно $p_1(0) \cdot (1/b)$, скорость их рекомбинации равна $p_1(0) \cdot s$ и постоянная спадающая $s \cdot (1/b) = sb$.

Исследованный выше случай соответствует случаю τ_{p1} , уже рассмотренному в § 6, так как при этом величина p_1 предполагалась не зависящей от x .

Так же как и в случае τ_{p2} , можно учесть влияние магнитного поля. Здесь снова $\tau_{p1} = \tau_{p2}$, если только углы θ_p и θ_n малы. Если эти углы велики, то магнитное поле будет влиять на сопротивление нитевидного образца и при расчете необходимо вводить в рассмотрение полные тензоры подвижности и диффузии (см. § 3).

Методы, аналогичные рассмотренным здесь, были использованы в § 2 гл. 3 при вычислении коэффициента поверхностной рекомбинации s [11].

§ 7. РЕШЕНИЕ УРАВНЕНИЯ НЕПРЕРЫВНОСТИ ДЛЯ СЛУЧАЯ БОЛЬШИХ ПЛОТНОСТЕЙ ВВЕДЕННЫХ НОСИТЕЛЕЙ ТОКА (Решение Херинга)

Уравнение непрерывности, сформулированное в § 2, соотношение Эйнштейна и формула для определения положения квази-уровней Ферми применимы в случаях любых концентраций дырок и электронов (если только не имеет место вырождение). Однако в § 5 и 6 рассматривались лишь те случаи, когда концентрация дырок была очень мала по сравнению с концентрацией электронов. При этих условиях дифференциальные уравнения линейны относительно p , переменные разделяются и можно получить простые решения, позволяющие исследовать влияние рекомбинации, диффузии и приложенных к образцу электрического и магнитного полей. Когда концентрация дырок сравнима с концентрацией электронов, уравнения становятся нелинейными, решения их и интерпретация рассматриваемых процессов становятся значительно более сложными.

Изысканное решение нелинейной задачи было получено Херингом [12] для случая изменения концентраций в значительных пределах. Рассмотрим здесь простейшую из исследованных им задач, предполагая, что

$$1) \quad n = p + n_e = p + N_d - N_a \quad (12.121)$$

(это означает, что суммарный заряд равен нулю и что доноры и акцепторы полностью ионизированы);

2) рекомбинация ничтожно мала;

3) условия однородны в направлениях y и z , так что все величины являются функциями только x и t ;

4) диффузионными токами можно пренебречь по сравнению с токами проводимости, обусловленными электрическим полем E_x .

В соответствии с предположением 2 мы имеем здесь дело с совершенно иными условиями, чем в двух последних параграфах, где процесс рекомбинации играл главную роль, обуславливая спадающие концентрации введенных носителей до их равновесных значений. В данном случае равновесные значения концентрации не представляют для нас интереса, так как благодаря предположению 2 они не могут быть достигнуты. По этой именно причине в уравнении (12.121) вместо n_n написано n_e , где индекс e указывает на то, что мы имеем дело с избытком числа доноров над числом акцепторов¹⁾. Значение проводимости $\sigma(0) = q\mu_n n_e$ не есть ее равновесное значение, а просто соответствует случаю $p = 0$. В предельном случае вещества с собственной проводимостью, когда $n_e \rightarrow 0$, мы имеем $\sigma(0) = 0$ и $E_0 = \infty$.

При этих условиях плотности дырочного и электронного токов будут равны

$$I_p = q\mu_p p E, \quad (12.122)$$

$$I_n = q\mu_n n E. \quad (12.123)$$

Часть тока, переносимая дырками, является функцией p и может быть записана²⁾ как

$$f(p) = \frac{I_p}{I} = \frac{I_p}{I_n + I_p} = \frac{\mu_p (p/n_e)}{\mu_n [1 + (1 + 1/b)(p/n_e)]}; \quad (12.124)$$

величина n исключена из этого выражения с помощью формулы (12.121) и

$$b = \frac{\mu_n}{\mu_p} \quad (12.125)$$

(эти выражения рассмотрены более подробно в § 1 гл. 3). Подставляя $f(p)$ в уравнение непрерывности (12.10), получим

$$\begin{aligned} \frac{\partial p}{\partial t} &= -\frac{1}{q} \nabla \cdot \mathbf{I}_p = -\frac{I}{q} \frac{\partial f(p)}{\partial x} = \\ &= -\frac{I}{q} \frac{df(p)}{dp} \frac{\partial p}{\partial x} \equiv -V(I, p) \frac{\partial p}{\partial x}. \end{aligned} \quad (12.126)$$

В дальнейшем будет показано, что величина $V(I, p)$, определяемая этим выражением, есть скорость перемещения концентрации дырок. Величину $V(I, p)$ можно записать в форме

$$V(I, p) \equiv \frac{I}{q} \frac{df(p)}{dp} = \frac{I}{q\mu_n n_e} \frac{\mu_p}{[1 + (1 + 1/b)(p/n_e)]^2}. \quad (12.127)$$

В некоторых случаях, когда I не зависит от времени, вместо $V(I, p)$ будем писать просто $V(p)$. При небольших концентрациях введенных

¹⁾ e — первая буква английского слова „excess“, что значит „излишек“. — Прим. ред.

²⁾ Обозначение $f(p)$, применяемое только в этом параграфе, не следует путать с обозначениями функций Ферми — Дирака f и f_p . — Прим. авт.

дырок величина $V(I, p)$ сводится к скорости дрейфа дырок $\mu_p E_0$ в электрическом поле

$$E_0 = \frac{I}{qn_e \mu_n} = \frac{I}{\sigma(0)}, \quad (12.128)$$

где I — ток, протекающий в материале с проводимостью n -типа. При больших значениях p величина $V(I, p)$ стремится к нулю. Полезно сравнить скорость $V(I, p)$ со скоростью дрейфа дырки $v_p = E \mu_p$. Значение напряженности поля E определяется из выражения для проводимости

$$\sigma(p) = q(\mu_p p + \mu_n n) = \sigma(0) [1 + (1 + 1/b)(p/n_e)], \quad (12.129)$$

а ток I из выражения $E = I/\sigma$, так что

$$v_p(p) = \mu_p E = \frac{I}{qn_e \mu_n} \frac{\mu_p}{[1 + (1 + 1/b)(p/n_e)]}. \quad (12.130)$$

При $p = 0$ (12.130), подобно (12.127), сводится к $\mu_p E_0$. При больших значениях p имеем

$$v_p(p) = v_p(0) \frac{\sigma(0)}{\sigma(p)} = \frac{v_p(0)}{[1 + (1 + 1/b)(p/n_e)]}, \quad (12.131a)$$

$$V(I, p) = v_p(0) \left[\frac{\sigma(0)}{\sigma(p)} \right]^2 = \frac{v_p(0)}{[1 + (1 + 1/b)(p/n_e)]^2}. \quad (12.131b)$$

Так как, согласно (12.121), концентрации n и p связаны между собой линейно, то n может быть непосредственно определено по значению p , и поэтому нет необходимости рассматривать ее в отдельности. Мы не будем также рассматривать и тот случай, когда электроны вводятся в материал p -типа, так как все необходимые для этого выражения можно получить из уже выведенных нами путем перестановки индексов и замены b на $1/b$.

Сейчас рассмотрим предельный случай, соответствующий введению дырок в материал с собственной проводимостью. Основываясь скорее на непосредственных физических соображениях, чем на (12.127), покажем, что концентрация дырок в объеме образца остается постоянной, хотя скорость их дрейфа и не равна нулю¹⁾. Предположим соответственно, что концентрация дырок в момент $t = 0$ есть некоторая функция $p_0(x)$; так как $n_e = 0$, то при $t = 0$ концентрация электронов n также равна $p_0(x)$. Таким образом, при всех значениях x отношение дырочного тока к электронному остается постоянным и равно $1/b$, а, следовательно, дырочный ток не зависит от x .

¹⁾ После обсуждения физического смысла величины V , проведенного в связи с фиг. 113, читатель должен убедиться в том, что сделанное выше утверждение соответствует тому, что при $n_n \rightarrow 0$ скорость V обращается в нуль, в то время как скорость дрейфа v_p сохраняет конечное значение. — *Прим. авт.*

При таком условии результирующий дырочный ток, втекающий в какой-нибудь слой образца δx с одной стороны, равен дырочному току, вытекающему из него с другой стороны; при этом концентрации дырок и электронов остаются постоянными, хотя как дырки, так и электроны перемещаются. Можно представить себе, что такого рода условия имеют место при фотопроводимости, возникающей при освещении импульсом света ограниченного участка нитевидного образца, изготовленного из материала с собственной проводимостью. В этом случае область с большой концентрацией носителей заряда остается фиксированной в пространстве, в то время как дырки перемещаются через эту область в одном направлении, а электроны — в другом.

Математически картина явления, рассмотренного в предыдущем параграфе, описывается уравнением $df(p)'dp = 0$, из которого следует, что дырки переносят одну и ту же долю общего тока при всех значениях x и, следовательно, $\nabla \cdot \mathbf{I}_p = 0$. Как мы покажем ниже, при малых значениях p в образце n -типа концентрация дырок перемещается со скоростью, равной скорости их дрейфа.

В действительности явления диффузии и рекомбинации должны были бы вызывать соответственно увеличение и уменьшение размеров области с высокой концентрацией, и поэтому рассмотренное нами стационарное состояние является лишь некоторым приближением. Однако исследование стационарного случая послужило иллюстрацией того факта, что причиной смещения картины распределения концентрации является не дрейф дырок, а наличие дивергенца дырочного тока, как это следует из уравнения непрерывности.

С другой стороны, если концентрация дырок p так мала по сравнению с концентрацией электронов n , что электрическое поле при движении дырок изменяется чрезвычайно мало, то дырочный ток будет равен

$$I_p = E_0 q v_p p = q v_p p, \quad (12.132)$$

где v_p — скорость дрейфа дырки в поле E_0 . В этом случае уравнение непрерывности принимает вид

$$\frac{\partial p}{\partial t} = -v_p \frac{\partial p}{\partial x}. \quad (12.133)$$

Решением его будет любая функция $x - v_p t$:

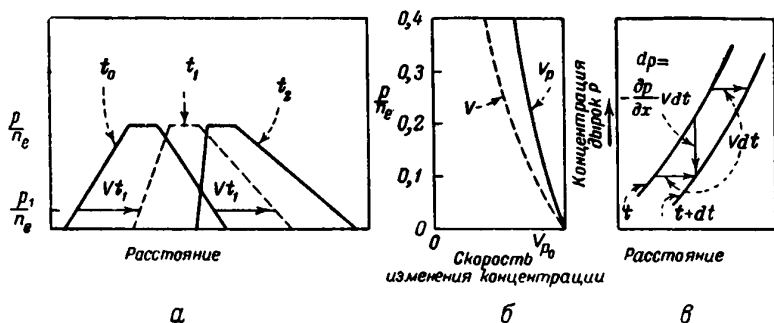
$$p = p_0(x - v_p t),$$

характеризующая смещение картины распределения концентрации дырок в направлении $+x$ со скоростью v_p .

Читатель может убедиться в том, что при $p/n_n \ll 1$ выражения (12.126) и (12.127) сводятся к (12.133), тогда как в случае собственной проводимости они сводятся к уравнению $\partial p/\partial t = 0$.

Херинг указал, что для промежуточных случаев решение уравнения (12.127) можно получить при помощи геометрического построения,

которое мы проведем на примере импульса тока, изображенного на фиг. 113. На этом рисунке импульс дырочного тока при $t=0$ имеет вид равнобедренной трапеции. На фиг. 113, б скорость $V(p)$ нанесена в функции p/n_e в том же масштабе, как и на фиг. 113, а. При этом предполагается, что ток I не зависит от времени. Построение Херинга заключается в следующем: на фиг. 113, а берется любая точка, для которой концентрация p имеет значение p_1 , и из этой точки проводится горизонтальная линия, длина которой равна $V(p_1)t$;



Фиг. 113. Деформация импульса дырочного тока, обусловленная модуляцией проводимости (для случая $b=2$).

конец этой линии будет характеризовать смещение импульса. Из фиг. 113, в следует, что этот процесс удовлетворяет выражению (12.127):

$$\frac{\partial p}{\partial t} = -V(p) \frac{\partial p}{\partial x}. \quad (12.134)$$

Фиг. 113, в показывает, что за время dt увеличение p при постоянном значении x будет равно

$$(dp)_x = [V(p) dt] \left(-\frac{\partial p}{\partial x} \right). \quad (12.135)$$

Разделив (12.135) на dt , получим выражение (12.126). В более общем виде функциональная зависимость между p , x и t может быть представлена как

$$x(p, t) = x(p, 0) + \int_0^t V[I(t'), p] dt'. \quad (12.136)$$

На анализе этого выражения остановимся подробнее при решении задач.

Нужно заметить, что в результате увеличения скорости смещения картины распределения концентрации при малых значениях концентрации дырок области с более низкой концентрацией перемещаются быстрее, чем области с более высокой концентрацией, так что задний

край импульса становится бесконечно крутым; это явление напоминает распространение ударной волны. По мере возрастания крутизны заднего края импульса предположение 4 (см. стр. 387), заключающееся в том, что диффузионными токами можно пренебречь, становится несостоятельным, и следует учесть расширение заднего края импульса за счет диффузии. В результате ширина заднего края импульса достигает предельного значения порядка

$$\Delta x = \frac{2kT}{qE_0} \div \left[1 - \frac{1}{1 + (1 + 1/b)(p_{\max}/n_e)} \right]^2, \quad (12.137)$$

где p_{\max} — максимальное значение p у вершины заднего края. Это уравнение выведено Херингом в предположении, что скорость расширения импульса, обусловленная диффузией, равна скорости его сужения, определяемой разностью в скоростях перемещения областей с предельными значениями концентраций дырок в импульсе (см. задачи).

Херинг рассмотрел также влияние рекомбинации, которое можно учесть относительно простым способом (см. задачи), и диффузии. Когда к нелинейным эффектам, обязанным своим происхождением электропроводности, присоединяются еще и явления, вызываемые диффузией, задача становится значительно более сложной. Ван Роозбрук [13] получил для нее стационарные решения в одномерном случае, сделав при этом приближенное предположение о том, что условие нейтральности $n = p + n_e$ (предположение 1, стр. 387) слегка нарушено.

§ 8. ЗАВИСИМОСТЬ ЭНЕРГИЙ \mathcal{E}_{1p} , \mathcal{E}_{1n} И \mathcal{E}_{1G} ОТ ОБЪЕМА

При изложении метода *потенциалов деформации* (в § 3 гл. 11 и § 6 гл. 17) мы ввели в рассмотрение величины \mathcal{E}_{1p} и \mathcal{E}_{1n} , представляющие собой изменения значений энергий \mathcal{E}_v и \mathcal{E}_c , соответствующих краям зоны при изменении относительного расширения на единицу. Эти величины входят в выражения, определяющие подвижность

$$\mu_n = \frac{2\sqrt{2\pi}}{3} \frac{nc_{II}}{\mathcal{E}_{1n}^2 m_n^{3/2} (kT)^{3/2}}, \quad (12.138a)$$

выведенные в § 4 гл. 11 [см. (11.92)]. Если выразить μ_n в $\text{см}^2/\text{в} \cdot \text{сек.}$, c_{II} — в $\text{дик}/\text{см}^2$, \mathcal{E}_{1n} — в эв и T — в °К, то выражение (12.138a) принимает вид

$$\mu_n = 3,2 \cdot 10^{-6} c_{II} \frac{(m/m_n)^{3/2}}{\mathcal{E}_{1n}^2 T^{3/2}}. \quad (12.138b)$$

Как указывалось в § 3 гл. 11 [см. (11.62)], соотношение

$$|\mathcal{E}_{1G}| = |\mathcal{E}_{1n}| + |\mathcal{E}_{1p}| \quad (12.139)$$

следовало бы применять в том случае, если бы наклон краев запрещенной зоны был таким, как это показано на фиг. 97. Чертеж

сделан в соответствии с результатами теоретического расчета зонной структуры алмаза, выполненного Кимбаллом и обсуждавшегося в связи с рассмотрением фиг. 54 (стр. 168).

Соотношение (12.139) было исследовано Бардином [14] для алмаза, кремния, германия и теллура. Методы, примененные для определения зависимости ширины запрещенной зоны \mathcal{E}_G от объема, основаны на статистической зависимости между \mathcal{E}_G и произведением np концентраций электронов и дырок [см. (10.24)]:

$$np = N_c N_v e^{(\mathcal{E}_v - \mathcal{E}_c)/kT} = 4 \left(\frac{2\pi mkT}{h^2} \right)^3 \left(\frac{m_n m_p}{m^2} \right)^{3/2} e^{-\mathcal{E}_G/kT} =$$

$$= 2,33 \cdot 10^{31} T^3 \left(\frac{m_n m_p}{m^2} \right)^{3/2} e^{-\mathcal{E}_G/kT} \text{ см}^{-6}. \quad (12.140)$$

Наиболее прямой способ изучения зависимости \mathcal{E}_G от объема заключается в изменении объема полупроводника под давлением при постоянной температуре и изучении получающихся при этом изменений величины np .

Таблица 11

СРАВНЕНИЕ ЗНАЧЕНИЙ \mathcal{E}_{1G} , ОПРЕДЕЛЕННЫХ РАЗЛИЧНЫМИ СПОСОБАМИ, СО ЗНАЧЕНИЯМИ СУММЫ $|\mathcal{E}_{1n}| + |\mathcal{E}_{1p}|$

	Алмаз	Кремний	Германий	Теллур
$\rho c^2 = c_{11} (10^{12} \text{ дин/см}^2)$	10,8 ^a	2,0 ^b	1,55 ^b	0,5 ^b
μ_n (электроны) (см ² /в · сек, 295°K)	900 ^c	300 ^d	3 500 ^e	530 ^{ae}
μ_p (дырки) (см ² /в · сек, 295°K)	>200 ^c	100 ^d	1 700 ^e	530 ^{ae}
$\mu_n T^{3/2}$	45 · 10 ⁵	15 · 10 ⁵	180 · 10 ⁵	27 · 10 ⁵
$\mu_p T^{3/2}$	>10 · 10 ⁵	5 · 10 ⁵	86 · 10 ⁵	27 · 10 ⁵
$ \mathcal{E}_{1c} $ (эВ)	8,8	6,5	1,7	2,4
$ \mathcal{E}_{1v} $ (эВ)	<30	11,3	2,4	2,4
$ \mathcal{E}_{1c} + \mathcal{E}_{1v} $	<39	17,8	4,1	4,8
\mathcal{E}_{1G} (эВ)	?	-30 ¹⁾	-5 ²⁾	+5,0 ⁴⁾

^a Продольная скорость в направлении [110], вычисленная из упругих постоянных [15].

^b Вычислено из упругих постоянных [16].

^c Оценено по данным для сжимаемости.

^d См. [17].

^e См. [18].

^{ae} J. R. Naupes, неопубликованные данные.

¹⁾ См. [19].

²⁾ Частное сообщение Мак-Кая.

⁴⁾ См. текст.

Объем полупроводника изменяется также и вследствие теплового расширения, что дает другой метод изучения изменений ξ_G в зависимости от объема. В дальнейшем рассмотрим как эти, так и другие методы исследования изменения величины ξ_G .

Результаты этих исследований приведены в нижнем ряду табл. 11. Значения $|\xi_{1n}|$ и $|\xi_{1p}|$ определены из формулы (12.138а) в предположении, что $m_n = m_p = m$. Их сумма приведена в предпоследнем ряду таблицы.

Следует отметить, что знаки перед числами, приведенными в последней строке таблицы, находятся в согласии с результатами вычислений, проведенных на основе зонной теории. Согласно вычислениям Кимбалла, выполненным для алмаза [20], зоны пересекаются при расстояниях между атомами, превышающих равновесное значение постоянной решетки, так что ξ_G должно было бы уменьшаться при объемном расширении образца и ξ_{1G} должно бы быть соответственно отрицательным. Этот результат, повидимому, справедлив также для кремния и германия. С другой стороны, теллур состоит из длинных цепочек атомов, слабо связанных между собой. По этой причине при равновесном значении постоянной решетки может иметь место относительно малое перекрытие, и зоны могут не пересекаться. Такой случай аналогичен изображенному на фиг. 53, или случаю, отвечающему большим значениям постоянной решетки, приведенному на фиг. 54.

Значение величины ξ_{1G} может быть определено следующими методами:

1) По зависимости собственной проводимости от давления. В области температуры, соответствующей собственной проводимости, концентрации n и p , каждая в отдельности, равны

$$n_i = 4,82 \cdot 10^{16} T^{3/2} \left(\frac{m_n m_p}{m^2} \right)^{3/4} e^{-\xi_G/2kT} \text{ см}^{-3}. \quad (12.141)$$

Так как подвижность носителей меняется с температурой как $T^{-1/2}$, то явная зависимость проводимости от температуры определяется только экспоненциальным множителем и

$$\sigma_i = \sigma_{i0} e^{-\xi_G/2kT}. \quad (12.142)$$

Бардин [22], используя данные Бриджмена [21] для теллура, нашел, что изменения σ_{i0} при изменении объема не играют существенной роли по сравнению с изменениями ξ_G . Аналогичным методом Миллер и Тейлор [23] нашли, что для германия $\xi_{1G} \approx -5,0$ эв.

2) По зависимости сопротивления $p-n$ -перехода от давления. Согласно теории, изложенной в § 5 настоящей главы, ток, протекающий через $p-n$ -переход, равен

$$I = q \left[\frac{D_p p_n}{L_p} + \frac{D_n p_n}{L_n} \right] (e^{qV/kT} - 1). \quad (12.143)$$

Для германия при температурах, близких к комнатной, концентрации p_p и n_n почти постоянны и равны соответствующим разностям между концентрациями доноров и акцепторов. Так как произведение np определяется из формулы (12.140), то n_p и p_n должны изменяться как $T^3 \exp(-\mathcal{E}_G/kT)$. Если предполагается, что изменения величин D_p , D_n , L_p и L_n с давлением малы по сравнению с эффектами, возникающими в результате изменения \mathcal{E}_G , то уравнение (12.143) можно использовать для объяснения зависимости сопротивления от давления. Полученное таким путем значение \mathcal{E}_{1G} для германия ¹⁾ согласуется с упомянутыми выше данными Миллера и Тейлора.

3) По зависимости концентрации от температуры. Экспериментально найдено (подробнее см. § 4 гл. 16), что численное значение постоянного коэффициента в выражении, определяющем произведение np для кремния, в 32,5 раза больше, чем это следует из формулы (12.140). Пирсон и Бардин предположили, что такое расхождение обусловлено температурной зависимостью \mathcal{E}_G , которую они записывают в виде

$$\mathcal{E}_G(T) = \mathcal{E}_G(0) - \beta T. \quad (12.144)$$

Эта формула приводит к

$$e^{-\mathcal{E}_G(T)/kT} = e^{\beta/k} e^{-\mathcal{E}_G(0)/kT}, \quad (12.145)$$

так что в выражении для np появляется множитель $\exp(\beta/k)$. Полагая его равным 32,5, получаем

$$\beta = k \ln 32,5 = 3 \cdot 10^{-4} \text{ эв/}^\circ\text{К}. \quad (12.146)$$

Воспользовавшись значением объемного коэффициента расширения α_v для кремния, находим

$$\mathcal{E}_{1G} = \frac{\beta}{\alpha_v} = \frac{3 \cdot 10^{-4}}{9 \cdot 10^{-6}} = 33 \text{ эв}. \quad (12.147)$$

Для германия соответствующий множитель в выражении, определяющем произведение np , равен 3,5, откуда $\beta = 1,0 \cdot 10^{-4} \text{ эв/}^\circ\text{К}$. И так как $\alpha_v = 19 \cdot 10^{-6}/^\circ\text{К}$, то $\mathcal{E}_{1G} = 5,25 \text{ эв}$ в хорошем согласии с значениями этой величины, полученными другими методами.

Температурную зависимость \mathcal{E}_G можно определить, воспользовавшись для этого данными о зависимости фотоэффекта от длины волны при разных температурах. Такие исследования были проведены Гаучером и Бриггсом ²⁾.

¹⁾ Личное сообщение Бардина, основанное на неопубликованных данных Холла и Пирсона. — Прим. авт.

²⁾ Неопубликованные исследования. — Прим. авт.

§ 9. ПОДВИЖНОСТЬ, ОПРЕДЕЛЯЕМАЯ ПО СКОРОСТИ ДРЕЙФА (ПОДВИЖНОСТЬ ДРЕЙФА), И ПОДВИЖНОСТЬ, ОПРЕДЕЛЯЕМАЯ ПО ЭФФЕКТУ ХОЛЛА (ПОДВИЖНОСТЬ ХОЛЛА)

Пока писалась эта книга, техника исследования движения дырок и электронов в германии была значительно усовершенствована. В опытах, сходных с описанными в § 1 гл. 3, Хайнс, применяя пульсирующее развертывающее поле, смог сократить коэффициент заполнения и повысить таким образом во много раз допустимые значения напряжения поля, развертывающего импульс. Это дало ему возможность получить импульсы вводимых носителей значительно более острой формы, чем это показано на фиг. 33. В результате применения такой техники удалось создать точный метод определения подвижности дрейфа, позволяющий измерять последнюю с ошибкой, не превышающей, вероятно, 5%. Значения подвижности электронов и дырок в германии, полученные с этой усовершенствованной методикой, приведены в табл. 12. Пирсон провел большую серию измерений эффекта Холла. Он нашел при этом, что подвижность, определенная по эффекту Холла, меняется от образца к образцу в значительно больших пределах, чем упомянутая выше подвижность дрейфа.

Таблица 12
ПОДВИЖНОСТЬ ЭЛЕКТРОНОВ И ДЫРОК В ГЕРМАНИИ, см²/в·сек

	Подвижность Холла, помноженная на 8/3π [25]	Подвижность дрейфа *	Подвижность, определенная по электропроводности [26]
Электроны	2 600 ± 300	3 600 ± 180	3 350 ± 400
Дырки	1 700 ± 500 — 100	1 700 ± 90	

* Сообщено Хайнсом на собрании американского физического общества в Чикаго 26 ноября 1949 г. [24].

Это обстоятельство не вполне понятно. Значения подвижности, полученные Пирсоном для тех же образцов, которые были исследованы и Хайнсом, также приведены в табл. 12. Там же указан разброс измеренных значений подвижности. В § 9 гл. 8 и § 4 гл. 11 [см. (11.87) и (11.90в)] показано, что если средняя длина свободного пробега не зависит от скорости электронов или дырок, как это предполагается в теории рассеяния их на тепловых колебаниях, основанной на применении потенциалов деформации, то подвижность дрейфа должна была бы равняться произведению подвижности, определенной из эффекта Холла, на множитель 8/3π. Из табл. 12

видно, что значения подвижности дрейфа не совпадают со значениями подвижности, полученными из эффекта Холла. Расхождение между этими значениями явно превышает возможные ошибки измерения. Ниже мы попытаемся объяснить это расхождение ¹⁾.

Все эти измерения подвижности проводились на монокристаллах германия. Монокристаллы германия получались вытягиванием из расплава по методу Тилля и Литтля [27]. Необходимость применения монокристаллов связана с наличием дополнительных сопротивлений на границах зерен, о чем речь будет идти в следующем параграфе.

Подвижность электронов можно оценить из опытов Пирсона и его сотрудников [24], о чем коротко уже упоминалось в гл. 1. В этих опытах содержание сурьмы в германии с проводимостью n -типа определялось с помощью радиоактивной сурьмы. Полученные данные показывают, что проводимость меняется линейно с содержанием сурьмы, так что можно записать

$$\sigma = N'_d q \mu_n + \sigma_0,$$

где σ_0 — проводимость контрольного образца без примеси сурьмы и N'_d — содержание сурьмы, определенное радиоактивным методом.

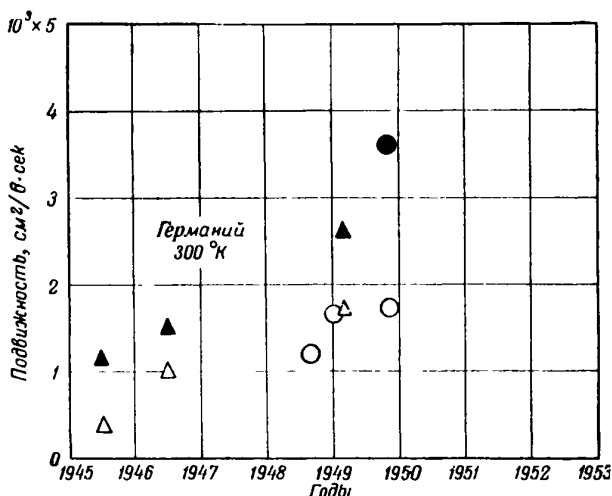
Решая это уравнение относительно μ_n , для ряда образцов можно получить значение подвижности, определенное из данных по электропроводности; эта цифра также приведена в табл. 12. Значительный разброс данных связан частично с тем обстоятельством, что исследование проводилось на поликристаллических образцах. Подвижность, определенная по данным об электропроводности, находится в лучшем согласии со значениями подвижности дрейфа, полученными Хайнсом, чем со значениями подвижности, определенными Пирсоном из эффекта Холла. Этого, конечно, и следовало ожидать, так как и в опытах по определению скорости дрейфа, и в опытах по определению электропроводности измеряется средняя скорость движения носителей тока вдоль нитевидного образца под влиянием приложенного поля. Поскольку захват носителей тока и, возможно, столкновения электронов с дырками не вызывают заметных отклонений подвижности дрейфа от микроскопической подвижности (см. § 7 гл. 8), то в обоих случаях при измерениях мы должны были бы иметь дело с одной и той же физической величиной.

Необходимо, конечно, подчеркнуть, что оба метода измерения подвижности по своему принципу совершенно различны.

Найдено, что значения подвижности носителей заряда в германии, определенные из эффекта Холла, значительно колеблются от образца к образцу. На фиг. 114 нанесены экспериментальные данные для

¹⁾ Это объяснение приведено в сокращенном виде в статье Пирсона [18]. — *Прим. авт.*

подвижности электронов и дырок, полученные разными способами за некоторый период времени. Недавно Дюнлап [28] получил значения подвижностей, существенно отличающиеся от приведенных в табл. 12. По его данным, величина $(8/3\pi)\mu_H$ равна $3400 \text{ см}^2/\text{в}\cdot\text{сек}$ для электронов и $2650 \text{ см}^2/\text{в}\cdot\text{сек}$ для дырок.



Фиг. 114. Данные о подвижности носителей заряда в германии в исторической последовательности.

Поскольку некоторые данные являются неопубликованными „лучшими значениями“, имеющимися в настоящее время, то литературные ссылки не приведены.

- ▲ — подвижность Холла (электроны);
- △ — подвижность Холла (дырки);
- — подвижность дрейфа (электроны);
- — подвижность дрейфа (дырки).

Попытаемся дать возможное объяснение тому очевидному факту, что для электронов в германии n -типа подвижность Холла μ_H , которую мы примем равной $(3\pi/8) 2600 = 3100 \text{ см}^2/\text{в}\cdot\text{сек}$, меньше подвижности дрейфа μ , которая для электронов, введенных в германий p -типа, равна $3600 \text{ см}^2/\text{в}\cdot\text{сек}$. Из (11.87) следует, что

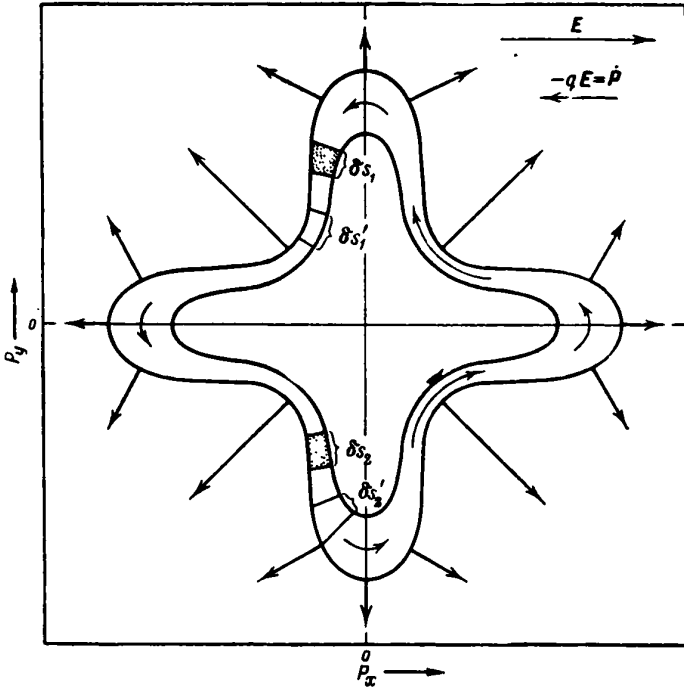
$$\frac{\mu_H}{\mu} = \frac{(\bar{\tau}^2)}{(\tau)^2} \geq 1,$$

где $\bar{\tau}^2$ и $(\tau)^2$ — соответствующие взвешенные средние. В случае постоянства средней длины свободного пробега и невырожденного состояния значение отношения μ_H/μ равно $3\pi/8 = 1,18$, в то время как отношение $3100/3600 = 0,86$. Таким образом, при любых предположениях о зависимости τ (σ) от скорости электронов v экспериментальные

данные не могут быть согласованы с теорией. Значения отношения

$$\frac{\mu_H}{\mu} < 1$$

могут быть, однако, получены, если предположить, что изоэнергетические поверхности в зоне Бриллюэна обладают не сферической



Фиг. 115. Контуры постоянной энергии, которые привели бы к низким значениям μ_H/μ .

формой, отвечающей постоянному значению эффективной массы (на этом предположении мы основывались при выводе всех формул, приведенных в § 4 гл. 11), а формой, изображенной на фиг. 115, на которой показано, как особая форма поверхности постоянной энергии может привести к значительному уменьшению отношения μ_H/μ .

На фиг. 115 изображена плоскость $P_z = 0$ в зоне Бриллюэна. В этом случае длина вектора скорости в различных точках поверхности постоянной энергии будет различной. Наибольшего значения она достигнет в тех точках, где поверхность постоянной энергии ближе всего подходит к началу координат \mathbf{P} -пространства. Если

электрическое поле, как это указано на фиг. 115, приложено в направлении положительной полуоси x , то электроны в образце с электропроводностью n -типа будут ускоряться в противоположном направлении, в результате чего число электронов в состояниях, лежащих между поверхностями постоянной энергии, будет изменяться. Наиболее существенную роль при этом будут играть те области, в которых составляющая скорости v_x велика, так как доля квантовых состояний f' , занятых электронами, возрастает с течением времени по закону

$$\left(\frac{d}{dt} f'\right)_{E_x} = \frac{-q}{kT} f(1-f) v_x E_x$$

[см. (11.69)].

Двум группам квантовых состояний с большими отрицательными значениями v_x на фиг. 115 отвечают элементы площади δs_1 и δs_2 . Поле E_x быстро увеличивает число электронов в этих областях. Если, кроме того, приложено и магнитное поле, то „изображающие“ точки будут перемещаться вдоль поверхностей постоянной энергии в направлении, указанном стрелками, и избыточные электроны в элементах δs_1 и δs_2 будут смещаться по направлению к элементам $\delta s'_1$ и $\delta s'_2$. Такое смещение вызывает увеличение составляющей средней скорости в положительном направлении полуоси y , так как элементам $\delta s'_1$ и $\delta s'_2$ отвечают большие положительные значения v_y , чем элементам δs_1 и δs_2 . В результате вращение ¹⁾ вектора средней скорости происходит по часовой стрелке, а не в противоположном направлении, как это должно было бы быть в случае сферических поверхностей постоянной энергии, сечения которых в плоскости фиг. 115 имели бы, очевидно, форму окружности. Поэтому в том случае, когда в наименее энергетических состояниях зоны проводимости число избыточных электронов мало, знак эффекта Холла может совпадать со знаком, соответствующим обычно дырочной проводимости.

Следует подчеркнуть, что электроны, находящиеся вблизи дна зоны проводимости, в случае фиг. 115 не эквивалентны дыркам. Если бы в зоне находился лишь один избыточный электрон, то ему соответствовал бы отрицательный результирующий заряд и в опытах по изучению дрейфа он двигался бы по направлению к положительному электроду. В этом отношении такой электрон вел бы себя совершенно аналогично избыточным электронам, свойства которых были рассмотрены в предшествующих главах. Однако направление скорости дрейфа такого электрона изменялось бы под влиянием магнитного поля, так что наблюдаемый эффект Холла по знаку соответ-

¹⁾ Выражение для коэффициента Холла для случая сложных поверхностей постоянной энергии см. в книге Вильсона [29], уравнение (236).—
Прим. авт.

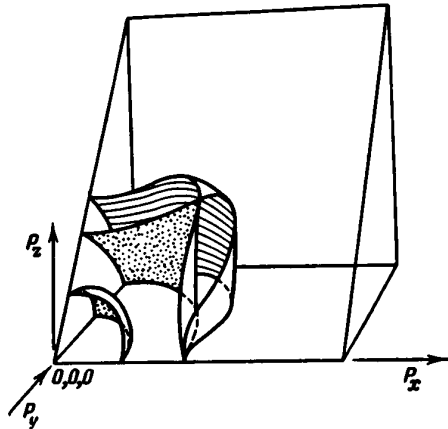
ствовал бы дыркам. Аналогично, дырки в почти заполненной зоне, характеризуемой поверхностями постоянной энергии, изображенными на фиг. 115, вели бы себя как обычные дырки с той только разницей, что знак эффекта Холла соответствовал бы электронам. Для таких электронов и дырок мы введем термины *аномальный электрон* и *аномальная дырка*.

Аномальные дырки дают эффект Холла, соответствующий по знаку электронам и обычно называемый *нормальным эффектом Холла*.

Аномальные электроны дают эффект Холла, соответствующий по знаку положительно заряженным носителям и обычно называемый *аномальным эффектом Холла*. В нитевидном образце, содержащем как аномальные дырки, так и аномальные электроны, носители тока должны были бы концентрироваться на стороне образца, противоположной той, на которой происходит концентрация носителей в случае фиг. 35 (стр. 99). Это должно было бы давать *аномальный эффект Суля*, и направление магнитного поля на графике фиг. 37 (стр. 105), изображающем зависимость полного сопротивления от напряженности магнитного поля, должно было бы быть изменено на противоположное.

В случае поверхностей постоянной энергии, форма которых отличается от сферической в меньшей степени, чем это изображено на фиг. 115, произойдет, вероятно, скорее уменьшение отношения μ_H/μ , чем изменение его знака. Таким путем можно попытаться объяснить наблюдаемое значение отношения μ_H/μ , равное, как было отмечено, 0,86.

Указанное выше изменение знака, связанное с особой формой поверхностей постоянной энергии, в германии места не имеет. На это непосредственно указывают знак эффекта Суля и взаимная согласованность эффекта Холла, направления движения введенных носителей, направления действия поля, зависимости характера проводимости от валентности примеси, направления выпрямления и термоэлектрических явлений. Рассмотрение фиг. 116, на которой поверхности постоянной энергии изображены в трех измерениях, показывает, что предположение о том, что отношение μ_H/μ может быть порядка 0,86, не



Фиг. 116. Поверхности постоянной энергии вблизи края энергетической зоны.

Для хлористого натрия [41].

лишено разумных оснований. Хотя эти поверхности построены для случая другого кристалла, нет оснований считать, что поверхности такой формы не будут иметь места и для германия [30].

Дальнейшее доказательство того, что поверхности постоянной энергии в зоне Бриллюэна не имеют сферической формы, было получено при исследовании изменения сопротивления в магнитном поле. Было обнаружено, что этот эффект по своей величине значительно превышал предсказанный теорией, развитой в предположении, что поверхности постоянной энергии имеют сферическую форму (см. § 4 гл. 11). Последние результаты исследований изменения сопротивления в магнитном поле были опубликованы Пирсоном [31] и Сулем [32]. В этих опытах была обнаружена сильная анизотропия эффекта изменения сопротивления в магнитном поле для монокристаллов германия, что является дальнейшим доказательством недостаточности теории, основанной на предположении, что поверхности постоянной энергии [33] имеют сферическую форму.

Если поверхности постоянной энергии не сферичны, то теория длины свободного пробега, в которой используется понятие „потенциала деформации“, нуждается в значительном пересмотре, так как в этом случае существенную роль может играть рассеяние на поперечных колебаниях. Естественно, однако, предполагать, что рассеяние на поперечных колебаниях будет играть менее существенную роль, чем рассеяние, обусловленное объемным расширением, рассмотренное в гл. 11 и 17, так что согласие между значениями величины \mathcal{E}_{1G} и суммы $|\mathcal{E}_{1n}| + |\mathcal{E}_{1p}|$, отмеченное в § 8, будет, повидимому, иметь место.

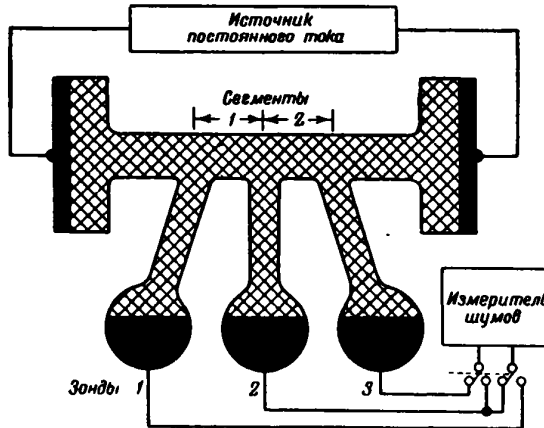
§ 10. К ТЕОРИИ ШУМОВ

В последнее время удалось достичь больших успехов в понимании природы шумов в германии. Монтгомери [34] нашел, что шумы в нитевидных образцах германия характеризуются типичным спектром (1/частота) или спектром контактного шума¹⁾, описанным в § 2 гл. 2.

В этих опытах применялись образцы со столь большими утолщениями на концах, что при этом совершенно исключались какие-либо эффекты, обусловленные контактными явлениями на границе металл — полупроводник. На фиг. 117 изображен один из таких образцов, применявшихся в опытах. Образцы такой, а также и иной сложной формы, необходимые для других целей, изготовлялись с помощью спроектированного Бондом [36] магнитострикционного реза. Некоторые образцы, изготовленные таким способом, изображены на фиг. 1 (стр. 16). В обычных условиях опыта нитевидный образец соединялся с источником постоянного тока, так что флуктуации напряжения между

¹⁾ Название „контактный шум“ связано с тем, что впервые шумы такого типа наблюдались в контактах; см., например, работу Христиансена и Пирсона [35]. — *Прим. авт.*

двумя любыми точками образца могли быть приписаны только флюктуациям сопротивления. Образцы, применявшиеся в опытах по измерению э. д. с. шума, вырезывались вместе с ответвлениями из одного и того же монокристалла; форма такого образца изображена на фиг. 117. Боковые ответвления использовались как зонды. Этот метод изготовления образца исключает дополнительные шумы, часто возникающие на контактах металла с полупроводником. Найдено,



Фиг. 117. Схема опыта Монтгомери для исследования шумов в германии.

что (э. д. с. шума)² складывается в первом приближении из члена, соответствующего шумам теплового возбуждения, и члена, прямо пропорционального I^2 , соответствующего, таким образом, флюктуациям сопротивления. Монтгомери удалось получить убедительное доказательство того, что шум при модуляции сопротивления образца германия n -типа обязан своим происхождением флюктуациям дырочного тока. Предварительная теория этого явления заключается в следующем.

В объеме образца германия или на его поверхности находятся области, характеризующиеся повышенной концентрацией дырок. Такие области действительно наблюдались на опыте Хайнсом. Опыты показали, что наложение на образец пульсирующего разветвляющего поля вызывало в некоторых случаях прохождение импульсов дырочного тока через коллектор, хотя при этом контакт с эмиттером отсутствовал. Хайнс показал, что источники дырок локализованы в некоторых определенных областях нитевидного образца, на вид ничем не отличающихся от окружающего материала. Во всех случаях осталось неясно, были ли эти области включениями p -типа или просто областями n -типа с несколько меньшей электронной прово-

димостью, чем соседние с ними участки кристалла. Предположим, что эти области обладали просто несколько меньшей электронной проводимостью, чем соседние, и, таким образом, в условиях равновесия концентрация дырок в них была несколько выше средней. Однако в присутствии электрического поля эти области высокой концентрации дырок рассасываются. В результате в областях, где концентрация дырок ниже нормальной, последняя возрастает за счет эмиссии дырок, которые затем снова высасываются полем; таким образом, эта область служит источником дырок.

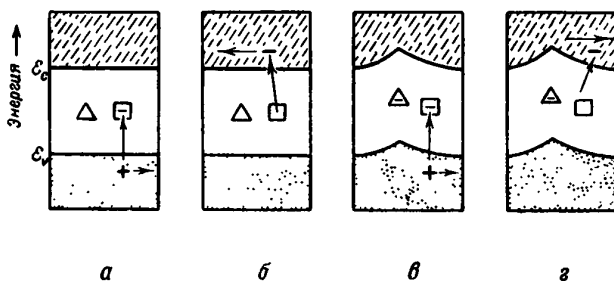
Если бы дырочный ток из этого источника носил беспорядочный характер, то шум в нитевидном образце был бы много меньше шума, наблюдаемого на опыте. Кроме того, такой шум характеризовался бы скорее «белым» спектром частот (мощность шумов на цикл не зависела бы от частоты), а не спектром, отвечающим закону $1/f$ и наблюдающемуся в действительности в нитевидных образцах и транзисторах, о чем уже говорилось в § 2 гл. 2.

Предположим на время, что такие флюктуирующие источники дырочного тока действительно существуют в нитевидном образце, и посмотрим, к каким следствиям приводит такое предположение.

Шум, производимый флюктуациями данного источника дырок, будет пропорционален времени жизни дырок в образце. На самом деле мощность шума будет прямо пропорциональна квадрату этого времени. Как мы указывали раньше в связи с описанием эффекта Суля, время жизни дырок в нитевидном образце германия может быть существенно уменьшено в результате одновременного наложения магнитных и электрических полей. Такое уменьшение шума в нитевидном образце германия действительно наблюдалось Монтгомери, что находится в хорошем количественном согласии с теорией, основанной на указанных выше представлениях. Из опытов Монтгомери следует, что экспериментальные данные лучше всего согласуются с предположением о том, что шумы создаются источниками дырок, расположенными на поверхности образца. Этот результат согласуется и с наблюдениями, указывающими на то, что рекомбинация дырок с электронами происходит в основном на поверхности. В действительности при тепловом равновесии, согласно принципу детального равновесия, любая область, в которой имеет место рекомбинация, должна в такой же мере быть и источником эмиссии. Симметрия такого же типа между эмиссией и рекомбинацией имеет место и в неравновесных условиях. Таким образом, область, характеризующаяся флюктуирующей по интенсивности рекомбинацией дырок и электронов, может создавать шумы точно так же, как и область флюктуирующей эмиссии.

Наряду с эффектом уменьшения шума при наложении магнитных полей Монтгомери наблюдал также *эффект передачи шумов*. Образец, с которым велась работа, имел три боковых электрода, так что, присоединяя различные пары электродов к измерителю шумов, можно

было измерять три напряжения шума V_{12} , V_{23} и V_{13} . Если бы модуляция сопротивления в каждом из двух сегментов не зависела от модуляции в соседнем сегменте, то сумма мощностей шума V_{12}^2 и V_{23}^2 должна была бы равняться V_{13}^2 . Однако в экспериментальных условиях время жизни дырки в нитевидном образце германия так велико, что большая часть дырок, входящих в первый сегмент, проходит и через второй сегмент, не успевая рекомбинировать. Кроме того, время пролета составляет лишь небольшую долю периода переменного тока,



Фиг. 118. Модуляция процесса рекомбинация — эмиссия захваченным зарядом.

△ — модулирующая ловушка; □ — центр рекомбинации и эмиссии.

на котором проводятся измерения. Следовательно, модуляция сопротивления, обусловленная дырочным током, будет приблизительно одинаковой в обоих сегментах, и поэтому должны скорее складываться напряжения шумов, а не их мощности. Это предположение подтверждается опытами Монтомери, который нашел, что

$$V_{13}^2 = V_{12}^2 + V_{23}^2 + 2CV_{12}V_{23}, \quad (12.148)$$

где коэффициент корреляции C равен 0,85. Если поверхность образца обработать так, что время жизни носителей тока в нем станет меньше времени пролета между сегментами, то корреляция между модуляциями сопротивления в двух сегментах в значительной степени ослабляется. Эти опыты, в сочетании с опытом по изучению влияния магнитных полей, дают весьма убедительное доказательство того, что шумы в этих нитевидных образцах своим возникновением обязаны флюктуациям концентрации дырок.

На фиг. 118 изображена схема, иллюстрирующая возможный механизм возникновения шумов описанного выше типа. На этом рисунке изображена некоторая область полупроводника, содержащая два дефекта с описанными ниже гипотетическими свойствами. Один из этих дефектов является *рекомбинационно-эмиссионным центром*; мы будем называть его в дальнейшем просто *центром*. Предполагается, что этому центру соответствует уровень энергии, на который электрон

может быть возбужден термически из зоны валентных связей. После того как электрон возбудился (фиг. 118, *а*), образовавшаяся в этой зоне дырка получает возможность перемещаться; центр создает, таким образом, подвижную дырку. Электрон может быть далее вновь возбужден в зону проводимости, после чего он удалится от центра, ставшего нейтральным. Этот процесс может повторяться, так что центр будет действовать неограниченно долго в качестве источника дырок и электронов. В условиях равновесия столь же часто будет иметь место и обратный процесс: электрон из зоны проводимости захватывается сначала центром, а затем уже рекомбинирует со случайно проходящей вблизи от центра дыркой. Если область полупроводника, изображенная на фиг. 118, *а* и *б*, обладает в обычных условиях относительно высокой концентрацией дырок и если образовавшиеся в ней дырки уносятся приложенным электрическим полем, то можно показать, что этот центр «захвата» действует так, как если бы рассматриваемая область являлась естественным источником дырок.

На фиг. 118 изображен также и другой тип дефекта, способного захватывать электрон. Предположим, что эта ловушка значительно более устойчива, так что электрон остается в ней связанным в течение промежутка времени, который во много раз превышает длительность цикла, иллюстрируемого фиг. 118, *а* и *б*. Назовем эту ловушку *модулирующей ловушкой*, или просто *ловушкой*¹⁾.

Способность центра поглощать или испускать пары дырка — электрон изменяется, если ловушка становится заряженной. Если изображенная на рисунке модулирующая ловушка захватывает электрон, то форма энергетических зон изменяется так, как это показано на фиг. 118, *в* и *г*. В результате отрыв дырки от центра несколько затрудняется, а выход электрона, наоборот, несколько облегчается. В общем случае эти эффекты не компенсируют друг друга полностью, и модуляция одного из процессов будет играть более существенную роль, чем модуляция второго процесса. Следовательно, интенсивность эмиссионной способности центров, находящихся по соседству с модулирующей ловушкой, будет флюктуировать по мере того, как эта ловушка будет захватывать или терять электрон. Если энергетические уровни ловушек распределены в достаточно широком интервале значений энергии, то процесс их зарядки и разрядки при захвате и испускании электронов будет характеризоваться частотным спектром шумов $1/f$ -типа²⁾.

Разобранный только что случай представляет лишь один из большого числа возможных. Опыты Монтгомери с магнитными полями указывают также на то, что источники шума расположены на поверх-

¹⁾ Ловушки этого рода играют большую роль в теории фосфоров, см. например [37]. — *Прим. авт.*

²⁾ На это обстоятельство указал Бардин. — *Прим. авт.*

ности нитевидного образца. В этом случае роль центров могут играть поверхностные состояния; роль ловушек также могут играть поверхностные состояния, но другого типа. Роль модулирующих ловушек могут выполнять ионы или полярные молекулы, перемещающиеся по поверхности образца.

Необходимо подчеркнуть, что рассмотренные здесь источники шумов, подобно другим дефектам, упомянутым ранее в этом параграфе, не должны, повидимому, обязательно присутствовать в германии с проводимостью p - или n -типа. В идеализированной модели германия n - и p -типа, представляющей собой идеальную решетку с включенными в нее по принципу замещения трех- и пятивалентными атомами, казалось бы, не должно существовать источников шумов описанного выше типа. Нужно надеяться, что в дальнейшем будут изготавливаться образцы германия, столь близкие по своим свойствам к их идеализированной модели, что шумы в них окажутся сведенными до минимума, определяемого шумом Джонсона и дробовым эффектом ¹⁾.

§ 11. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

Этот параграф мы посвятим краткому обзору современного состояния электроники транзисторов ²⁾ и укажем возможные направления научного и технического прогресса этой области физики.

Как уже не раз упоминалось, изобретение транзистора Бардином и Браттейном привело к открытию нового круга явлений, протекающих в полупроводниках. Опыты Хайнса по определению подвижности из данных по скорости дрейфа, опыты Пирсона с модуляцией проводимости и опыты Суля с магнитной фокусировкой дырок и электронов сыграют, по мнению автора, весьма существенную роль в дальнейшем развитии общей теории электронной проводимости в полупроводниках. Эти опыты дополнили наши представления о свойствах электронов и дырок в полупроводниках, основанные на измерениях проводимости, эффекта Холла и термоэлектрических эффектов

Нитевидный образец германия играет в электронике транзистора такую же роль, как вакуумная или газоразрядная лампы в обычной электронике. Аналогом термоэлектронного катода здесь является эмиттер, и его эффективность определяется той долей γ его тока в прямом направлении, которая переносится введенными носителями. В гл. 3 приведены численные значения величины γ . После того как были произведены эти измерения, численное значение коэффициента b было исправлено с 1,5 на 2,1. Кроме того, в новых опытах стали

¹⁾ Дробовой шум в монокристаллах германия в области частот от 1 до 1600 *кГц* исследовали Герцог и Ван дер Шиль [55]. — *Прим. ред.*

²⁾ Более свежие данные о технических достижениях в области кристаллических усилителей за рубежом приведены в статье Мортонна, несколько сокращенный перевод которой дается в виде приложения к настоящей книге, стр. 673. — *Прим. ред.*

применяться образцы с боковыми ответвлениями в качестве зондов и была также развита новая методика введения носителей и измерений на переменном токе.

Браттейн, который ввел эти усовершенствования в методику измерения коэффициента γ , работал с германием *n*-типа. Он нашел, что величина γ меняется от значений порядка 0,2 почти до единицы и для данного контакта увеличивается с возрастанием эмиттерного тока. Согласно опытным данным, значение коэффициента γ зависит от способа обработки поверхности, но эти опытные данные не обладают достаточно хорошей воспроизводимостью.

Кажущееся значение γ , определяемое на опыте как отношение малых переменных токов i_p/i_e , возрастает с увеличением напряжения развертывающего поля, и поэтому значения, приведенные Браттейном, были получены путем экстраполяции к бесконечно большому полю.

В дополнение к измерениям коэффициента γ и наблюдениям новых фундаментальных явлений, ставших возможными благодаря применению нитевидных образцов, получены также и другие сведения, касающиеся поведения дырок и электронов. Показано, что обычно рекомбинация происходит главным образом на поверхности образца. Удалось выяснить, что дырка успевает испытать много тысяч или десятков тысяч столкновений с поверхностью, прежде чем она, наконец, рекомбинирует. Повидимому, процесс рекомбинации на поверхности является структурно-чувствительным и зависит как от химической структуры, так и от характера силового поля на поверхности.

Известно также, что в процессе объемной рекомбинации в германии, помимо атомов элементов 3 и 5 групп периодической системы, принимают участие и другие примеси и искажения решетки кристалла. Найдено, что объемная рекомбинация относительно мало чувствительна к присутствию обычных доноров и акцепторов, но что скорость ее может заметно увеличиваться после термической обработки, переводящей германий *n*-типа в германий *p*-типа¹⁾. Было также обнаружено, что центры *p*-типа, образующиеся в результате бомбардировки ядерными частицами, весьма эффективно увеличивают скорость рекомбинации дырок и электронов²⁾.

Нужно надеяться, что природа этих нарушений правильности решетки, способствующих рекомбинации, будет установлена в дальнейшем с помощью исследований при низких температурах, измерений времени жизни носителей тока и изучения шумов и явлений захвата.

Как уже указывалось в § 9, границы зерен в германии создают дополнительные сопротивления. Основные стороны этого явления

¹⁾ Неопубликованные исследования, проведенные Спарксом и Хайнсом.— *Прим. авт.*

²⁾ Образование ловушек для дырок в германии *n*-типа при облучении его быстрыми нейтронами исследовалось в работе Ларк-Горовитца и его сотрудников [56].— *Прим. ред.*

изучены Пирсоном¹⁾. Он нашел, что некоторые границы между зернами действуют как области p -типа, так что их электрические характеристики подобны характеристикам двух p — n -переходов, соединенных последовательно. Это заключение было подтверждено в дальнейшем опытами по измерению фото-э. д. с. Еще не вполне ясно, создается ли p -область на границе зерен вследствие сегрегации химических примесей или же как следствие неупорядоченного состояния, подобного тому, которое возникает при бомбардировке германия нуклонами, но в данном случае являющегося результатом различия ориентации кристаллических зерен по обе стороны границы. Если германий n -типа путем термической обработки превращается в германий p -типа, то граница зерна действует, повидимому, просто как очень тонкий слой с несколько иной проводимостью, и, таким образом, не влияет заметно на сопротивление образца²⁾.

Больших успехов удалось достичь при изготовлении самого германия, что было вызвано необходимостью обеспечить производственную программу по выпуску транзисторов. В результате сейчас можно получать относительно большие химически чистые и весьма совершенные по своей структуре кристаллы, позволяющие легко осуществить опыты по электронике транзисторов, которые не могли быть выполнены в прошлом. Предварительное сообщение о выращивании кристаллов опубликовали Тиль и Литтль [27].

Резюмируем кратко те положения зонной теории проводимости, которые получили подтверждение в связи с работами по транзисторам. Согласно представлениям зонной теории, избыточный электрон 1) обладает зарядом — e ; 2) совершает беспорядочное диффузионное движение; 3) дрейфует в электрическом поле, обладая при этом некоторой определенной подвижностью; 4) отклоняется магнитным полем.

Дырка обладает аналогичными свойствами, отличаясь от электрона лишь знаком заряда. Новые опыты дали всем этим представлениям зонной теории прочное экспериментальное обоснование. Справедливость предположений 1 и 3 (стр. 387) подтверждена опытами по введению носителей. Величина заряда определена с помощью формулы для тока через выпрямляющий p — n -переход, содержащей отношение e/kT . Отклонение в магнитном поле непосредственно наблюдалось при изучении эффекта Суля. Наличие диффузионного движения подтверждается опытами с p — n -переходами, которые показали, что вид частотных характеристик таких переходов согласуется с диффузионной теорией выпрямления. Существование диффузионного движения еще более непосредственно доказывается опытами Гаучера³⁾; в этих опытах возникно-

¹⁾ См. [25]. См. также [38] и [39]. Теория структуры границ зерен у некоторых простейших кристаллов, повидимому, находится в согласии с опытом. — *Прим. авт.*

²⁾ Новые данные приведены в работах Фуллера и др. [57], а также Слехтера и Колба [58]. — *Прим. ред.*

³⁾ Неопубликованное сообщение. — *Прим. авт.*

вление избыточных носителей тока, создаваемых в образце германия под действием света, сопровождается не их дрейфом под влиянием сильного разветвляющего поля, а диффузией их в очень слабом электрическом поле.

Указанные выше свойства дырок и электронов находятся в согласии с теми физическими представлениями о поведении дырок и электронов, которые были рассмотрены в части I книги. Количественная теория, основанная на этих представлениях, используется при исследовании процессов, происходящих в транзисторах. Поэтому важно уяснить себе, что в результате новых опытов, связанных с изучением транзисторов, наиболее существенные стороны этих представлений основаны теперь на непосредственных экспериментальных данных.

Дальнейшим доказательством справедливости зонной теории служит рассмотренная в § 8 корреляция между шириной запрещенной зоны и размерами элементарной ячейки. Еще одним доказательством справедливости этой теории служит тот факт, что значение квантового выхода в опытах по фотопроводимости весьма близко к единице (на каждый поглощенный фотон образуется пара дырка — электрон) [40]. Именно этого и следовало ожидать на основании представлений о механизме возбуждения носителей тока при поглощении фотонов, описанных в гл. 1.

Может быть уместно высказать здесь некоторые предположения о дальнейшем развитии электроники транзисторов. Те, кто интенсивно работал в этой области, вполне разделяют чувство большого оптимизма, испытываемое автором в отношении дальнейших возможностей, открывающихся перед этой областью физики полупроводников. Большинство работающих в этой области полагает, что открыта новая область электроники, сравнимая по своему значению со всей областью вакуумной и газоразрядной электроники. Уже сейчас разработано несколько типов транзисторов, а исследование ряда других типов доведено до такого состояния, что можно с полной уверенностью говорить об их дальнейшем практическом применении. Высказан также ряд новых идей, нуждающихся, правда, еще в соответствующей экспериментальной проверке.

Весьма возможно, что в дальнейшем будут сделаны непредвиденные в настоящее время открытия, основанные на явлениях введения носителей тока, а также действия поля, эффекте Суля и на свойствах выпрямляющих переходов. Очень возможно, что по мере развития наших знаний о полупроводниках для практических целей будут использованы также новые, пока еще не изученные физические явления.

Поводом к написанию этой книги послужило убеждение автора в том, что электронная проводимость полупроводников приобретет вскоре огромное значение как в технике связи, так и в других отраслях электротехники.

Инженеры, работающие в области электроники транзисторов, указывали мне на необходимость развить описательную теорию полупро-

водников, уделив при этом особое внимание исследованию тех вопросов, которые могут быть использованы в качестве основания для технического проектирования. Во второй части книги автор поставил своей главной целью дать именно такую описательную теорию, а также ознакомить читателя с некоторыми математическими приемами, полезными при практических расчетах.

Обзор основных положений общей теории составляет главную задачу части III книги.

Возможно, что некоторые разделы этой части, ввиду их математической сложности, окажутся недоступными многим из читателей. Однако можно надеяться, что у читателя также создается мнение, что важнейшие представления современной физики транзисторов теоретически обоснованы.

ЗАДАЧИ

1. Показать, что если в момент $t=0$ распределение дырок в плоскости $x=0$ таково, что на единицу площади приходится N_p дырок, то (пренебрегая всеми остальными процессами) значение концентрации p в последующие моменты времени будет изменяться в результате диффузии по закону

$$p(x, t) = \frac{N_p}{\sqrt{4\pi Dt}} e^{-x^2/4Dt}.$$

Рассмотреть трехмерный случай и показать, что если при $r=0$ и $t=0$ концентрация дырок равна N_p , то в момент времени t их концентрация будет

$$p(r, t) = \frac{N_p}{(4\pi Dt)^{3/2}} e^{-r^2/4Dt}.$$

(Это легче всего показать, исследуя решения уравнения непрерывности в прямоугольных координатах. Значения встречающихся при этом интегралов приведены в задачах гл. 11.)

2. Показать, что при диффузии носителей тока в одном измерении (см. предыдущую задачу) значение величины $\langle x^2 \rangle$, определяемое как

$$\langle x^2 \rangle = \int_{-\infty}^{+\infty} x^2 p(x, t) \frac{dx}{N_p},$$

будет равно $2Dt$.

Проверить, что для любой функции $p(x, t)$, удовлетворяющей уравнению

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}$$

и соответствующим граничным условиям,

$$\frac{d}{dt} \langle x^2 \rangle = 2D.$$

Обобщить эти результаты на случай двух и трех измерений.

3. Показать, что если дырки в плоскости $x=0$ появляются (например, вследствие фотопроводимости) в момент $t=0$, а время жизни дырки в веществе равно τ_p , то решение уравнения непрерывности для избыточных дырок будет иметь вид

$$p(x, t) = \frac{P_0}{\sqrt{4\pi Dt}} e^{-x^2/4Dt - t/\tau_p}.$$

Показать, что относительное число дырок, рекомбинирующих за время dt , равно $[\exp(-t/\tau_p)] dt/\tau_p$. Показать, что значение x^2 , усредненное для каждой дырки за промежуток времени от момента ее появления до момента рекомбинации, равно

$$\langle x^2 \rangle = 2D\tau_p.$$

Получите этот же результат путем усреднения величины x^2 по той части решения (12.65), которая относится к постоянному току. Исходя из физических соображений, показать, что оба результата должны быть одинаковы. Показать, что среднее расстояние (а не квадрат расстояния), проходимое дыркой при диффузии, для обоих решений равно

$$L_p = \sqrt{D\tau_p}.$$

4. Рассмотреть импульс дырочного тока, образованный в точке $r=0$ в момент $t=0$ в поле E , и предположить затем, что этот импульс расплывается вследствие диффузии и дрейфует в поле. Показать, что распределение дырок в объеме образца в момент t подобно их распределению, рассмотренному в задаче 1, но максимум распределения имеет теперь место при $r = t\mu_p E$. Показать, что отношение увеличения ширины импульса к его смещению при дрейфе равно приближенно

$$\sqrt{\frac{2kT}{eV_0}},$$

где V_0 , равное произведению поля E на дрейфовое расстояние, есть разность напряжений между точками образца, отвечающими начальному положению импульса и его положению в момент t . Эта формула используется для оценки расстояния, которое введенные дырки должны пройти вдоль нитевидного образца, прежде чем они распределятся равномерно по всему сечению, а также и для других подобных расчетов.

5. Сравните члены уравнения (12.69), обусловленные проводимостью и диффузией для дырочного тока в n -области. Для этого рассмотрите некоторую плоскость вблизи переходной области, для которой можно считать, что I_p и I_n примерно одинаковы. Показать, что это дает

$$\frac{d\psi}{dx} \approx \frac{p_p}{n_n} e^{qV_d/kT} \frac{kT}{qL_p}$$

и что, следовательно, дырочный ток осуществляется в основном за счет диффузии до тех пор, пока V_0 не будет так велико, что концентрация введенных дырок станет сравнимой с n_n . Показать, что для дырок в германии, обладающих подвижностью $\mu_p = 1700 \text{ см}^2/\text{в} \cdot \text{сек}$, $kT/qL_p = 3,90 (10^{18} \tau_p)^{1/2} \text{ в/см}$, или $3,90 \text{ в/см}$ при $\tau_p = 1 \text{ мксек}$.

6. Рассмотреть рекомбинацию на поверхности в случае $s \rightarrow 0$. Показать, что формула (12.103б)

$$v_s = s \left(\frac{1}{B} + \frac{1}{C} \right)$$

может быть получена путем вычисления скорости рекомбинации на поверхности в предположении, что $p = \text{const}$ по сечению образца, и вычисления отношения:

(Полная скорость рекомбинации) ÷ (Полное число имеющихся дырок).

7. Рассмотреть решение уравнения (12.105), которым мы пренебрегли и которое описывает процесс диффузии против поля. Показать, что, если пренебречь рекомбинацией, это решение сводится к распределению Больцмана [см. (12.14)].

8. Показать, что при рассмотрении эффекта Суля (§ 6) переменные в уравнении непрерывности (12.114) разделяются, если ввести в рассмотрение множитель $\cos bz$. В результате в (12.115) к члену ν добавится член $D_p \varphi_0^2 / C^2$. Граничные условия для φ_0 при этом такие же, как и при $H_z = 0$.

9. Вычислить поперечное поле E_t и поперечное напряжение в нитевидных образцах германия p - и n -типа с площадью поперечного сечения, равной 20 мм^2 (т. е. $0,05 \text{ см}$ по стороне), при $E_x = 10 \text{ в/см}$ и $H_z = 5000$ эрстед.

10. Предположить, что эффективная масса равна массе свободного электрона, и показать, что число электронов, обладающих скоростями, лежащими в интервале $dv_x dv_y dv_z$ и находящихся в объеме dV , равно (согласно второму приближению § 1 гл. 10):

$$dn = \frac{(2\pi kT/m)^{3/2}}{n} = e^{-m(v_x^2 + v_y^2 + v_z^2)/2kT} dv_x dv_y dv_z dV.$$

Показать, что число электронов с составляющими скорости $v_x v_y v_z$, падающих за время dt на элемент площади $dx dy$, равно числу электронов в объеме $dx dy v_z dt$. Проинтегрировать это выражение по всем скоростям и получить формулу для скорости, с которой электроны достигают площадки $dx dy$:

$$n = \sqrt{\frac{kT}{2\pi m}} = \frac{nv_T}{\sqrt{4\pi}} = 0,282 nv_T.$$

Здесь v_T — скорость, определяемая формулой (11.82); ее зависимость от T показана на фиг. 88.

11. Примените результат, полученный при решении предыдущей задачи, чтобы показать, что число столкновений с поверхностью на один акт рекомбинации равно $0,282 v_T/s$, где s — скорость рекомбинации на поверхности (см. § 6). Показать, что эта формула будет справедлива для значений $s < 10^6$ см/сек. Показать, что при $s = 10^8$ см/сек на акт рекомбинации приходится около 2 800 столкновений.

12. Исходя из геометрических соображений, показать, что в случае, рассмотренном в связи с фиг. 113, полное число дырок не зависит от времени.

13. Вывести уравнение в частных производных:

$$\left(\frac{\partial p}{\partial t}\right)_x = -\left(\frac{\partial x}{\partial t}\right)_p / \left(\frac{\partial x}{\partial p}\right)_t,$$

которое должно быть справедливо всегда, когда x есть функция p и t . Применить это уравнение к тому случаю, когда

$$x(p, t) = x(p, 0) + \int_0^t V(I(t'), p) dt'.$$

Проверить, что эта функция удовлетворяет уравнению (12.126) в том случае, когда ток есть $I(t)$.

14. Вывести уравнение

$$\frac{\partial E}{\partial t} = \frac{E}{T} \frac{dI}{dt} - V(E) \frac{\partial E}{\partial x},$$

где, согласно (12.129),

$$E = \frac{I}{\sigma(p)}$$

и, следовательно,

$$V(E) = E \mu_p \frac{E}{E_0}.$$

Показать, что если E представлено как функция от x , то импульс обладает в момент времени $t = 0$ бесконечно крутым передним фронтом, а при $t > 0$ будет обладать фронтом параболической формы.

15. Предположить, что задний край ударной волны приближенно соответствует решению задачи 1 для случая, когда $x < 0$,

$$p(x, t) = p_0 (4\pi Dt)^{-1/2} e^{-x^2/4Dt}.$$

Показать, что задний край импульса расплывается в направлении, противоположном направлению его движения, со скоростью, приблизительно равной

$$\frac{d\langle x \rangle}{dt} = \sqrt{\frac{2D}{t}} = \frac{2D}{\langle x \rangle}.$$

Приравнять эту скорость к скорости, с которой концентрация от значения $p = 0$ возрастает до значения $p = p_{\max}$, и получить формулу для ширины заднего края импульса, приведенную в конце § 7.

16. Показать, что решение уравнений § 7 можно получить для случая $I = \text{const}$ и $\tau_p = \text{const}$, смещая каждую точку $p(x_0, 0)$ начального распределения концентрации (см. фиг. 113) вдоль кривой, описываемой уравнениями

$$p(x_0, t) = p(x_0, 0) e^{-t/\tau_p},$$

$$x = x_0 + \int_0^t V[I, p(x_0, t')] dt'.$$

17. *Внутренние контактные потенциалы*¹⁾. Имеется нитевидный образец германия n -типа, для которого $s = 0$ (рекомбинация на поверхности отсутствует). Предположите, что этот образец находится в контакте с зондом, изготовленным также из германия n -типа, но что зонд отделен от образца тонкой прослойкой германия p -типа, так что между зондом и образцом имеются два $p-n$ -перехода. Предположите, что в нитевидный образец вводятся дырки, так что $\varphi_p > \varphi_n$.

Используя в этом случае соображения, приведенные в начале § 5, показать, что потенциал зонда будет стремиться к значению, лежащему между φ_p и φ_n . В частности, если φ' есть значение уровня Ферми в объеме зонда, где $\varphi'_p = \varphi'_n$, показать, что $\varphi_n < \varphi' < \varphi_p$. Показать также, что если p -слой превращается в n -слой, скажем, в результате увеличения N_d , то $\varphi' \rightarrow \varphi_n$.

18. В качестве дальнейшего приложения рассмотренных выше методов исследуйте механизм действия транзисторов с $p-n-p$ -переходами [6].

19. Токи коллектора в тех опытах, на основании которых построена фиг. 30 (стр. 91), достаточно велики, так что по сравнению с токами проводимости диффузионные токи ничтожно малы. Показать, что отношение I_n/I_p протекающих через контакт электронного и дырочного токов равно bn/p . Полагая $b = 2$ и используя данные фиг. 30, вычислить „собственный“ коэффициент усиления по току α для каждого из четырех контактов (этот метод предложил Бардин).

ЛИТЕРАТУРА

1. Tolman R. C., The Principles of Statistical Mechanics, Oxford, 1938.
2. Van Vleck J. H., The Theory of Electric and Magnetic Susceptibilities, Oxford, 1932.
3. Einstein A., Ann. d. Phys., 17, 549 (1905).
4. Müller-Pouillet, Lehrb. d. Phys., IV (3), 316 (1933).
5. Herring C., Nichols M. H., Rev. Mod. Phys., 21, 185 (1949) (см. перевод: Херинг К., Никольс М., Термоэлектронная эмиссия, ИЛ, 1950).
6. Shockley W., Bell Syst. Techn. Journ., 28, 435 (1949).
7. Slater J. C., Phys. Rev., 76, 1592 (1949).
8. Schiff L. I., Quantum Mechanics, New York, 1949.
9. Morse P. M., Vibration and Sound, New York, 1936 (см. перевод: Морзе П., Колебания и звук, М.—Л., 1949).

¹⁾ См. [6], где этот вопрос изложен подробно. — *Прим. авт.*

10. Kemble E. C., *The Fundamental Principles of Quantum Mechanics*, New York, 1937.
11. Suhl H., Shockley W., *Phys. Rev.*, **76**, 180 (1949).
12. Herring C., *Bell Syst. Techn. Journ.*, **28**, 401 (1949).
13. Roosbroeck W., *Bell Syst. Techn. Journ.*, **29** (4), (1950).
14. Bardeen J., Shockley W., *Phys. Rev.*, **77**, 407 (1950).
15. Hearmon R. F. S., *Rev. Mod. Phys.*, **18**, 409 (1946), (1950).
16. Bond W. L., Mason W. P., McSkimin H. J., Olsen K. M., Teal G. K., *Phys. Rev.*, **78**, 176 (1950).
17. Klück C. C., Maurer R. J., *Phys. Rev.*, **76**, 179 (1949).
18. Pearson G. L., Bardeen J., *Phys. Rev.*, **75**, 1777 (1949).
19. Johnson V. A., *Phys. Rev.*, **74**, 1255 (1948).
20. Kimball G. E., *Journ. Chem. Phys.*, **3**, 560 (1937).
21. Bridgman P. W., *Proc. Am. Acad. Sci.*, **68**, 95 (1933); **72**, 157 (1938), **74**, 21 (1940).
22. Bardeen J., *Phys. Rev.*, **75**, 1777 (1949).
23. Miller P. H., Taylor J., *Phys. Rev.*, **76**, 179A (1949).
24. Pearson G. L., Haynes J. R., Shockley W., *Phys. Rev.*, **78**, 295 (1950); **81**, 835 (1951).
25. Pearson G. L., *Phys. Rev.*, **76**, 179 (1949).
26. Pearson G. L., Struthers J. D., Theuerer H. C., *Phys. Rev.*, **75**, 344 (1949).
27. Teal G. K., Little J. B., *Bull. of the 298th Meeting of the American Physical Society*, *Phys. Rev.*, **78**, 647 (1950).
28. Dunlap W. C., Jr., *Phys. Rev.*, **77**, 759A (1950).
29. Wilson A. H., *Theory of Metals*, Cambridge, 1936 (см. перевод: А. Вильсон, *Квантовая теория металлов*, М.—Л., 1941).
30. Shockley W., *Phys. Rev.*, **78**, 173 (1950).
31. Pearson G. L., *Phys. Rev.*, **78**, 646 (1950).
32. Suhl H., *Phys. Rev.*, **78**, 646 (1950).
33. Shockley W., *Phys. Rev.*, **71**, 191 (1950).
34. Montgomery H. C., Shockley W., *Phys. Rev.*, **78**, 646 (1950).
35. Christensen C. J., Pearson G. L., *Bell Syst. Techn. Journ.*, **15**, 197 (1936).
36. Bond W. L., *Phys. Rev.*, **78**, 646 (1950).
37. Leverenz H. W., *Luminescence of Solids*, New York, 1950.
38. Taylor W. E., Fan H. Y., *Phys. Rev.*, **78**, 334 (1950).
39. Read W. T., Shockley W., *Phys. Rev.*, **78**, 275 (1950).
40. Goucher F. S., *Phys. Rev.*, **78**, 476, 816 (1950).
41. Shockley W., *Phys. Rev.*, **50**, 754 (1936).
- 42*. Van Roosbroeck W., *Bell Syst. Techn. Journ.*, **29**, 560 (1950).
- 43*. Лашкарев В. Е., *ДАН СССР*, **73**, 929 (1950).
- 44*. Лашкарев В. Е., *ЖЭТФ*, **18**, 917 (1948).
- 45*. Лашкарев В. Е., *ЖЭТФ*, **18**, 953 (1948).
- 46*. Лашкарев В. Е., *Изв. АН СССР, серия физич.*, **16**, 186 (1952).
- 47*. Иоффе А. В., Иоффе А. Ф., *ДАН СССР*, **16**, 77 (1937).
- 48*. Иоффе А. В., *ЖТФ*, **18**, 1498 (1948).
- 49*. Иоффе А. В., *ЖЭТФ*, **10**, 49 (1946).
- 50*. Давыдов Б. И., *ДАН СССР*, **20**, 279 (1938).
- 51*. Губанов А. И., *ЖЭТФ*, **21**, 721 (1951).
- 52*. Губанов А. И., *ЖЭТФ*, **21**, 79 (1951).
- 53*. Губанов А. И., *ЖЭТФ*, **20**, 1287 (1950).
- 54*. Губанов А. И., *ЖЭТФ*, **21**, 304 (1951).
- 55*. Herzog G. B., Van der Ziel A., *Phys. Rev.*, **84**, 1249 (1951).
- 56*. Cleland J. W., Crawford J. H., Lark-Horovitz K., *Phys. Rev.*, **83**, 861 (1951).
- 57*. Fuller S. C., Theuerer H. C., Van Roosbroeck, **85**, 674 (1952).
- 58*. Slichter W. P., Kolb E. D., *Phys. Rev.*, **87**, 527 (1952).

ЧАСТЬ III

НЕКОТОРЫЕ ВОПРОСЫ
КВАНТОВОМЕХАНИЧЕСКОЙ
ТЕОРИИ

ВВЕДЕНИЕ

Физические закономерности, о которых идет речь в этой книге, в принципе (но не практически) могут быть получены непосредственно из основных законов квантовой механики. Если бы все необходимые вычисления можно было осуществить полностью, то строго логичное изложение предмета следовало бы начать с формулировки этих основных законов, сославшись при этом на их историческое происхождение и экспериментальное подтверждение в наиболее простых случаях. В случае германия эти законы следовало бы применить к элементу № 32. Они дали бы возможность определить энергию наиболее устойчивого состояния для большого числа, например для 10^{22} атомов германия. Это состояние соответствовало бы кристаллу, обладающему структурой алмаза; сопоставляя его энергию с энергией изолированных нейтральных атомов германия, мы могли бы вычислить прочность ковалентной связи Ge—Ge. Далее, можно было бы определить энергию всех возбужденных состояний системы и провести классификацию этих состояний, исследуя характер движения электронов в кристалле и колебания атомов решетки.

Применив к этим возбужденным состояниям принципы квантовой статистики, можно было бы исследовать влияние температуры на рассматриваемую систему. Сочетание этих результатов с изучением влияния электрического и магнитного полей привело бы нас к созданию теории электропроводности и эффекта Холла для образцов чистого германия, обладающих собственной проводимостью. Если бы мы исследовали систему, содержащую 10^{15} атомов примеси элемента № 33 (сурьма), то наши результаты относились бы к образцу германия *n*-типа. Если бы в дополнение мы рассмотрели случаи различного пространственного распределения примесей, то это дало бы нам возможность объяснить выпрямляющее действие *p*—*n*-переходов и усилительное действие транзисторов с *p*—*n*-переходами.

Выполнение программы такого рода, однако, не только не может быть осуществлено практически, но и нежелательно. Вычисления энергии основного и некоторых возбужденных состояний были проведены для алмаза, а также для кремния. Полученные при этом численные результаты находятся не в очень благоприятном согласии с опытом и поэтому не могут быть удовлетворительным образом использованы для практических целей. Наиболее важный результат, который

следует из теории, состоит в том, что она дает мысленную схему (conceptual framework), в которой экспериментальные факты можно расположить таким образом, что они получают более глубокое толкование; это позволяет объяснить и предсказать связь между ними¹⁾. Точная теория, дающая точное определение квантовых состояний и собственных функций системы в целом, благодаря значительной сложности получаемых ею результатов была бы гораздо менее пригодной для этой цели, чем приближенная теория, пользующаяся более простыми представлениями и учитывающая их неточность путем введения в рассмотрение различных поправок, которые обычно также имеют простой смысл. Важно показать, как приближенные решения ряда задач могут быть получены на основе общих законов квантовой механики, а также выяснить физический смысл этих решений. Именно с этой целью и написана данная III часть книги.

В последующих главах будет рассмотрен ряд вопросов, имеющих принципиальное значение для квантовомеханической теории полупро-

¹⁾ Это утверждение автора, как и ряд других его высказываний, свидетельствует о том, что, подобно многим современным зарубежным ученым, он придерживается идеалистических философских позиций.

Для Шокли теория является только поставщиком „мысленной схемы“, дающей, по его мнению, лишь возможность удобного „расположения“ опытных фактов. Соотношения квантовой механики, как и соотношения вообще любой физической теории, по Шокли, представляют собой лишь определенные „правила корреляции“ между измеряемыми физическими величинами и их математическими обозначениями.

В действительности, как это особенно четко и ясно было показано В. И. Лениным в его книге „Материализм и эмпириокритицизм“, теория дает нам отражение объективных закономерностей окружающей нас объективной реальности.

Утверждение автора, что все физические закономерности, рассматриваемые в данной книге, могли бы быть получены непосредственно из соотношений квантовой механики (см. стр. 419) показывает, далее, что он отдает также известную дань возрождающимся в настоящее время за рубежом механистическим воззрениям, согласно которым все более сложные явления природы могут быть сведены к элементарным явлениям.

В действительности свойства сложных физических систем, содержащих большое число частиц, не могут быть сведены к движению простейших частиц, подчиняющемуся законам квантовой механики. Как было отмечено еще Энгельсом в „Диалектике природы“, при наличии многих частиц должны сказываться новые, специфические стороны движения, не присущие отдельным простейшим движениям.

Правда, Шокли не смог остаться последовательным в этом вопросе и вынужден был признать, что для объяснения изучаемых им явлений необходимо применить также и законы квантовой статистики. Последняя представляет собой, однако, новую качественную теорию, не вытекающую непосредственно из квантовой теории для отдельной частицы.

При образовании сложных систем должны проявляться все новые и новые качественные особенности. Это является прямым следствием учения Энгельса о движении вообще, а вовсе не результатом того обстоятельства, как это думает Шокли, что *точная* теория только в силу своей математической сложности не пригодна для объяснения физических свойств сложных систем. — *Прим. ред.*

водников. Уровень и методы изложения при этом весьма различны, в зависимости от характера исследуемого вопроса.

Изложение квантовой механики в гл. 14 ведется под чисто практическим углом зрения и, по всей вероятности, представит интерес главным образом для инженеров. В отличие от большинства руководств, основной упор при этом делается на то, чтобы показать, что применение соотношений квантовой механики дает возможность получать количественные результаты. Такой подход представляется существенным, в частности, потому, что механизм электронных процессов в транзисторах может быть исследован только на основе квантовой физики полупроводников.

Помимо формулировки основных законов и количественных соотношений квантовой теории, гл. 14 содержит приложения этих соотношений к некоторым простым задачам. В ней проводится также сопоставление математической формы уравнений квантовой механики с математической формой соответствующих уравнений теории электрических цепей. В § 6 гл. 14 получены и исследованы функции Блоха для одномерного кристалла. В заключение рассмотрен вопрос об обобщении метода Блоха на случай трехмерного кристалла. Построение зон Бриллюэна для различных кристаллов представляет собой трудоемкую задачу и не приводит к каким-либо существенно новым физическим представлениям. Кроме того, эти вопросы рассматриваются подробно в ряде руководств. По этим причинам мы не приводим наглядных примеров построения зон Бриллюэна для кристаллов, обладающих структурой алмаза, или же для других кристаллических структур. Читателю, интересующемуся дополнительными деталями теории этого вопроса, следует воспользоваться соответствующей специальной литературой, список которой приведен в конце гл. 14.

Кинематические и динамические свойства электронов и дырок, рассмотренные в гл. 6 и 7 лишь чисто описательно, исследуются в гл. 15 на основе соотношений квантовой теории. В первых трех параграфах этой главы дан вывод формулы для групповой скорости, для чего использованы сравнительно простые математические приемы. Следующие параграфы гл. 15 носят значительно более сложный характер. В них рассмотрены два метода исследования свойств дырок и электронов. Первый из этих методов изложен в § 6 и основан на представлениях о волновых пакетах; он дает возможность показать, что волновой пакет ведет себя аналогично частице, подчиняясь при этом законам, сформулированным в гл. 7. В двух следующих параграфах речь идет о построении и физическом смысле волновых пакетов, соответствующих дыркам. Получаемые при этом результаты находятся в согласии с теорией, изложенной в гл. 7. В § 9 дан несколько иной метод определения ускорения электрона. В этом случае для одного из направлений движения в рассмотрение вводятся периодические граничные условия, для чего следует представить себе, что тонкий кристалл изогнут так, что образует кольцо. Рассматривая такое кольцо как

замкнутый накоротко виток вторичной обмотки трансформатора, можно определить ускорение электрона, движущегося по кольцу. Этот метод обладает определенными принципиальными преимуществами, так как он дает возможность исследовать свойства системы в терминах строгой квантовомеханической теории. Эта модель представляет собой, таким образом, известное достижение на пути к тому, чтобы в конечном счете непосредственно из квантовомеханической теории получить соотношения, характеризующие физические свойства полупроводников. Эффект Холла на основе этой модели нами не исследован, поскольку теория этого эффекта рассмотрена в § 6 методом волновых пакетов.

Влияние теплового движения атомов кристалла, о котором уже говорилось в гл. 7, исследуется более подробно в гл. 16. Статистическая механика является объектом рассмотрения многих руководств, и мы отнюдь не стремились дать здесь исчерпывающее изложение этого предмета. В гл. 16 приведен простой вывод закона распределения Ферми — Дирака, а также вывод формулы Планка для распределения гармонических осцилляторов по возможным значениям энергии. Изложение этих вопросов основано на использовании приближенного описания квантовых состояний и иллюстрирует высказанное выше утверждение, что точное решение задачи не всегда необходимо или даже желательно. В заключение подробно рассмотрены приложения статистики к некоторым задачам теории полупроводников, представляющим для нас особый интерес.

В гл. 17 речь идет о взаимодействии электронов и дырок с колебаниями атомов кристалла. Для исследования этого вопроса использована квантовомеханическая теория переходов; найдено, что механизм рассеяния электронов и дырок на тепловых колебаниях решетки тесно связан с существованием зависимости ширины запрещенной зоны от постоянной решетки. В последнем параграфе гл. 17 результаты, полученные ранее для кристалла, содержащего один избыточный электрон, обобщены на случай наличия в кристалле одной дырки, а также на промежуточный случай системы, содержащей большое число дырок и большое число электронов.

Как указывает заглавие обсуждаемой части книги, проводимый в ней анализ дает принципиальное обоснование описательной теории, изложенной нами в части II. Это обоснование особенно существенно для математической теории электропроводности и эффекта Холла, изложенной в гл. 11. Полезно указать здесь, в какой мере отдельные главы части III связаны с гл. 11. Напомним, что в гл. 11 электроны (или дырки) предполагаются вначале распределенными по закону Ферми — Дирака; мы можем представить себе, что при этом используется модель кристалла, изогнутого в кольцо, рассматриваемое как виток вторичной обмотки трансформатора. Как показано в § 9 гл. 15, под влиянием электрического поля распределение нарушается и каждый из электронов приобретает ускорение. Принцип детального равновесия (см. § 2 гл. 17) становится при этом неприменимым, и

равновесие стремится к восстановлению. Подробное исследование механизма восстановления равновесия составляет основное содержание гл. 11.

Следует отметить, что обоснование теории, изложенной в гл. 11, не является достаточно полным. В гл. 15 мы исследуем влияние внешних полей на распределение, пренебрегая учетом теплового движения атомов кристалла, тогда как в гл. 17 рассмотрена роль тепловых колебаний решетки в отсутствие внешних полей. В гл. 11 предполагается, что эти эффекты аддитивны. Их аддитивность может быть доказана путем обобщения теории переходов, изложенной в § 2 гл. 17, на тот случай, когда под влиянием электрического поля электроны приобретают ускорение. Получаемые при этом результаты по своей форме несколько отличны от результатов теории Вайскопфа — Вигнера, но их физическое содержание подтверждает возможность сложения двух рассматриваемых эффектов, осуществляемого в гл. 11. Таким образом, этот пробел в общем обосновании теории электропроводности можно восполнить. Изложение этого вопроса не приводит, однако, к каким-либо принципиально новым представлениям, и поэтому для краткости мы его опускаем.

Аналогичные упреки могут быть предъявлены также и к теории, в которой формулы для рассеяния, приведенные в гл. 17, применяются к исследованию движения волновых пакетов, приобретающих ускорение под влиянием поля (см. § 6 гл. 15). И в этом случае можно было бы доказать, что результаты, полученные в гл. 17, могут быть обобщены на случай волновых пакетов, движущихся с ускорением. Однако решение этой сложной и громоздкой задачи, по всей вероятности, представило бы интерес лишь для очень ограниченного круга читателей; с другой стороны, читателям, интересующимся этими вопросами, было бы полезнее провести такое исследование самостоятельно.

ЭЛЕМЕНТАРНАЯ КВАНТОВАЯ МЕХАНИКА

§ 1. МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ЛИНИИ
ПЕРЕДАЧИ БЕЗ ПОТЕРЬ

Рассмотрим электрическую линию передачи без потерь с распределенными постоянными, которые могут меняться от точки к точке. В момент времени t в точке x линии могут быть определены напряжение $V(x, t)$ и ток $I(x, t)$. Кроме того, заданы индуктивность на единицу длины $L(x)$ и емкость на единицу длины $C(x)$. Согласно теории, эти шесть величин должны удовлетворять определенным дифференциальным уравнениям. Ниже будет показано, что эти уравнения дают новые соотношения между измеряемыми величинами, интересные с практической точки зрения. Например, энергия, запасенная на единице длины линии, и поток мощности связаны друг с другом: потеря энергии на любом участке линии эквивалентна разности потоков мощности на концах этого участка.

Получив ряд уравнений, мы поставим далее такой вопрос: какой характер должна была бы носить наша теория, если бы непосредственное измерение тока и напряжения в линии было невозможным и можно было бы измерить только плотность энергии? Если бы это было так, то введение в теорию величин V и I следовало бы рассматривать как дополнительную физическую гипотезу. С этой точки зрения математическая теория линии передачи обнаруживает сходство с квантовой механикой, где для описания поведения электрона в рассмотрении можно было бы ввести две функции, аналогичные функциям $V(x, t)$ и $I(x, t)$, в то время как определена может быть только плотность вероятности¹⁾.

В табл. 13 приведен ряд уравнений, относящихся к данному, а также к следующему параграфам. Для удобства сопоставления эти уравнения записаны в двух параллельных столбцах. В настоящем параграфе будут получены и исследованы соотношения, приведенные в столбце А; соотношения, содержащиеся в столбце В, будут рассмотрены в дальнейшем.

¹⁾ В действительности такие функции, аналогичные функциям $V(x, t)$ и $I(x, t)$, в квантовой механике в рассмотрении обычно не вводятся. Аналогия между математической теорией электрической линии передачи и квантово-механической теорией, которую неоднократно подчеркивает автор, носит чисто формальный характер и не представляет особой ценности.— *Прим. ред.*

Если расстояние x вдоль линии передачи отсчитывается слева направо, причем положительный ток течет вправо, то имеют место уравнения (I.A). Воспользуемся этими уравнениями, чтобы установить связь между потоком мощности и плотностью энергии. При этом

Таблица 13

СОПОСТАВЛЕНИЕ УРАВНЕНИЙ ДЛЯ ЛИНИИ ПЕРЕДАЧИ С ОДНОМЕРНЫМ ВОЛНЫМ УРАВНЕНИЕМ ДЛЯ ЧАСТИЦЫ

	А Линия передачи	В Частица ($\hbar = h/2\pi$)
Дифференциальные уравнения (I)	$\frac{\partial I}{\partial t} = -\frac{1}{L} \frac{\partial V}{\partial x}$ $\frac{\partial V}{\partial t} = -\frac{1}{C} \frac{\partial I}{\partial x}$	$\frac{\partial u}{\partial t} = -\frac{\hbar}{2m} \frac{\partial^2 w}{\partial x^2} + \frac{2uw}{\hbar} =$ $= \frac{\mathcal{E}w}{\hbar}$ $\frac{\partial w}{\partial t} = +\frac{\hbar}{2m} \frac{\partial^2 u}{\partial x^2} - \frac{2uw}{\hbar} =$ $= -\frac{\mathcal{E}u}{\hbar}$
Плотность (II)	Энергия на единицу длины $\frac{1}{2} LI^2 + \frac{1}{2} CV^2$	Вероятность на единицу длины (нормированная) $u^2 + w^2$
Поток плотности (III)	Энергия в единицу времени равна потоку мощности вправо VI	Вероятность, отнесенная к единице времени, равна току вероятности вправо $\frac{\hbar}{m} \left(u \frac{\partial w}{\partial x} - w \frac{\partial u}{\partial x} \right)$
Исключение одной переменной (IV)	$\frac{\partial^2 I}{\partial t^2} = \frac{1}{LC} \left[\frac{\partial^2 I}{\partial x^2} - \frac{d \ln C}{dx} \frac{\partial I}{\partial x} \right]$	$\frac{\partial^2 u}{\partial t^2} = -\frac{1}{\hbar^2} \mathcal{E}^2 u$
Уравнение для частоты (V) $f = \omega/2\pi$	Для случая, когда $\frac{dC}{dx} = 0$, I удовлетворяет уравнению $\frac{\partial^2 I}{\partial x^2} = -LC\omega^2 I$	$\frac{\partial^2 u}{\partial x^2} = -\frac{2m}{\hbar^2} [\omega\hbar - 2] u$

будем исходить из предположения (которое можно рассматривать как *дополнительную физическую гипотезу*), что в любой точке линии поток мощности, распространяющийся вправо, равен значению произведения VI , вычисленного в этой точке. В таком случае результирующая мощность, выделяющаяся на участке линии, заключенном между точками x_1 и x_2 ($x_1 < x_2$), равна

$$P(x_1, x_2) = V(x_1, t) I(x_1, t) - V(x_2, t) I(x_2, t). \quad (14.1)$$

Правая часть этого выражения может быть преобразована к виду

$$\begin{aligned}
 - \int_{x_1}^{x_2} \frac{\partial}{\partial x} (VI) dx &= - \int_{x_1}^{x_2} \left(\frac{\partial V}{\partial x} I + V \frac{\partial I}{\partial x} \right) dx = \\
 &= - \int_{x_1}^{x_2} \left(-L \frac{\partial I}{\partial t} I - VC \frac{\partial V}{\partial t} \right) dx = \\
 &= + \int_{x_1}^{x_2} \frac{\partial}{\partial t} \left(\frac{1}{2} LI^2 + \frac{1}{2} CV^2 \right) dx = \\
 &= \frac{\partial}{\partial t} \frac{1}{2} \int_{x_1}^{x_2} (LI^2 + CV^2) dx. \tag{14.2}
 \end{aligned}$$

Это показывает, что результирующий поток мощности, протекающий через участок (x_1, x_2) , равен скорости изменения интеграла

$\frac{1}{2} \int_{x_1}^{x_2} (LI^2 + CV^2) dx$, взятого вдоль линии. Этот результат остается

в силе также и в том случае, когда L и C являются функциями x (но не t); он не является, конечно, неожиданным, поскольку мы уже знаем, что $\frac{1}{2} LI^2$ и $\frac{1}{2} CV^2$ представляют собой значения энергии, приходящейся на единицу длины линии.

Если линия простирается до $x = \pm \infty$ и если как V , так и I обращаются в нуль при $x = \pm \infty$, то рассмотренная выше теорема показывает, что полная энергия в линии не зависит от времени и определяется численным значением интеграла

$$\int_{-\infty}^{+\infty} \left[\frac{1}{2} LI^2 + \frac{1}{2} CV^2 \right] dx = \mathcal{E}, \tag{14.3}$$

которое остается постоянным [так как $P(x_1, x_2)$ обращается в нуль при $x_1 = -\infty$, $x_2 = +\infty$]. Аналогичный результат получается и в тех случаях, когда линия обладает конечной длиной A и разомкнута ($I=0$) или же замкнута накоротко, так что поток мощности, вытекающий через сечение $x=A$, одновременно втекает через сечение $x=0$.

Из уравнений (I.A) легко получить уравнение (IV.A), а также аналогичное уравнение для $\partial^2 V / \partial t^2$. Если предположить далее, что L может меняться, тогда как C остается постоянным, и рассмотреть волну с частотой ω , так что

$$I(x, t) = f_1(x) \cos \omega t + f_2(x) \sin \omega t, \tag{14.4}$$

то $I(x, t)$, f_1 и f_2 должны удовлетворять уравнению (V.A). Формально это уравнение сходно с уравнением (V.B), которое мы исследуем в дальнейшем.

Предположим теперь, что единственный измерительный прибор, имеющийся в нашем распоряжении, представляет собой сочетание электростатического вольтметра с амперметром переменного тока, так что мы можем измерить только плотность энергии (II.A). Если бы мы могли вызвать возмущения в линии или же если бы такие возмущения почему-либо возникали в связи с условиями опыта, то можно было бы изучить при этом ряд характеристик линии. Если бы наше измерительное устройство обладало достаточной чувствительностью, то мы могли бы шунтировать линию, измерить характеристики стоячих волн и таким путем получить некоторые сведения о длинах волн. Наконец, можно было бы разработать математическую теорию, описывающую все свойства линии, связанные с измерениями средней плотности энергии. Мы предположим, что эта математическая теория дает два уравнения типа (I.A), из которых следуют далее уравнения (II.A) и (III.A), связывающие величины I и V .

Как мы сейчас увидим, роль вещественной и мнимой составляющих волновой функции аналогична роли величин I и V в нашем гипотетическом примере. Необходимо отметить, однако, следующее существенное различие: в квантовомеханической теории опытные основания для интерпретации компонент u и w волновой функции ψ , сходной с интерпретацией I и V , отсутствуют¹⁾.

§ 2. УРАВНЕНИЕ ШРЕДИНГЕРА

Во втором столбце табл. 13 приведен ряд уравнений, связанных с одномерным уравнением Шредингера. Рассматриваемая частица движется вдоль оси x под влиянием сил, сообщающих ей потенциальную энергию $\mathcal{U}(x)$. Две функции $u(x, t)$ и $w(x, t)$, входящие в уравнения (I.B), должны быть вещественными, подобно тому как вещественными должны быть значения I и V в уравнениях (I.A).

Величина $u^2 + w^2$, которая, таким образом, вещественна и положительна, называется *плотностью вероятности*. Смысл этой величины состоит в следующем. Предположим, что мы поставили опыт, с помощью которого определяем положение электрона в некоторый момент времени t , например выясняем, в какое место фотографической пластинки в электронной дифракционной камере²⁾ попадает

¹⁾ См. примечание к стр. 424. — *Прим. ред.*

²⁾ Современная электронная дифракционная камера представляет собой вариант старой дифракционной камеры, с помощью которой Девиссон и Джермер впервые обнаружили, что для электронов имеет место явление дифракции. Эта камера построена аналогично трубке катодного осциллографа. Пучок электронов ($\sim 10^{-8}$ а при 40 кВ) падает на образец, подвергается при этом дифракции и попадает затем на фотографическую пластинку, которая обычно заменяет флуоресцирующий экран. — *Прим. авт.*

электрон, воздействуя при этом на фотографическую эмульсию. В этом случае положение электрона, регистрируемое эмульсией, при больших значениях t не должно зависеть от t . Мы повторяем этот опыт снова и снова со многими электронами, каждый из которых начинает свой путь в камере Вильсона при аналогичных условиях и обладает, по видимому, теми же исходными значениями u и w (последние являются при этом функциями как y и z , так и x и t)¹⁾.

Вероятность нахождения электрона в интервале между x_1 и x_2 равна

$$\int_{x_1}^{x_2} (u^2 + w^2) dx / \int_{-\infty}^{+\infty} (u^2 + w^2) dx, \quad (14.5)$$

т. е. в любой момент t значение $u^2 + w^2$ пропорционально вероятности нахождения электрона в данном месте, отнесенной к единице длины. При повторении опыта эта вероятность дает долю электронов, находящихся в каждом интервале. Теория может быть, таким образом, сопоставлена с опытом. Если провести такое сопоставление, то почернение фотографической пластинки оказывается точно таким, как это следует из теории. Полученные при этом результаты далеко не тривиальны: исследуя явление диффракции электронов, можно определить длину волны электронных волн столь же точно, как и длину волны электромагнитных волн. Далее, по значениям потенциальной энергии \mathcal{U} электронов в кристалле можно вычислить интенсивность электронных волн. Во всех этих отношениях математическая теория, выражаемая уравнениями (I.B) табл. 13 и формулой (14.5), находится в согласии с опытом.

Чтобы упростить вычисления, функции u и w обычно умножают на нормирующий множитель так, чтобы функция $u^2 + w^2$ стала нормированной:

$$\int_{-\infty}^{+\infty} (u^2 + w^2) dx = 1. \quad (14.6)$$

Таким образом, знаменатель выражения (14.5) может быть опущен. В дальнейшем мы будем пользоваться нормированными волновыми функциями.

Как показано в табл. 13, в этом случае можно ввести в рассмотрение поток плотности (III.B), аналогичный потоку мощности или энергии (III.A). Рассмотрим отрезок линии между точками x_1 и x_2 .

¹⁾ В действительности об исходных значениях функций u и w ничего нельзя сказать, так как каждая из них в отдельности не имеет непосредственного физического смысла. — *Прим. ред.*

Скорость изменения вероятности нахождения электрона в этом интервале равна

$$\frac{d}{dt} \int_{x_1}^{x_2} [u^2(x, t) + w^2(x, t)] dx = 2 \int_{x_1}^{x_2} \left(u \frac{\partial u}{\partial t} + w \frac{\partial w}{\partial t} \right) dx. \quad (14.7)$$

Подставляя сюда значения производных по времени из уравнений (1.В) и интегрируя по частям, получаем

$$\left. \begin{aligned} & 2 \int_{x_1}^{x_2} \left[u \left(-\frac{\hbar}{2m} \frac{\partial^2 w}{\partial x^2} + \frac{u}{\hbar} w \right) + w \left(\frac{\hbar}{2m} \frac{\partial^2 u}{\partial x^2} - \frac{u}{\hbar} u \right) \right] dx = \\ & = \frac{\hbar}{m} \int_{x_1}^{x_2} \left(w \frac{\partial^2 u}{\partial x^2} - u \frac{\partial^2 w}{\partial x^2} \right) dx = \\ & = \frac{\hbar}{m} \left(w \frac{\partial u}{\partial x} - u \frac{\partial w}{\partial x} \right) \Big|_{x_1}^{x_2} - \frac{\hbar}{m} \int_{x_1}^{x_2} \left(\frac{\partial w}{\partial x} \frac{\partial u}{\partial x} - \frac{\partial u}{\partial x} \frac{\partial w}{\partial x} \right) dx = \\ & = \frac{\hbar}{m} \left(u \frac{\partial w}{\partial x} - w \frac{\partial u}{\partial x} \right)_{x_1} - \frac{\hbar}{m} \left(u \frac{\partial w}{\partial x} - w \frac{\partial u}{\partial x} \right)_{x_2}. \end{aligned} \right\} \quad (14.8)$$

Это означает, что изменение вероятности в интервале между x_1 и x_2 эквивалентно потоку

$$\left(\frac{\hbar}{m} \right) \left(u \frac{\partial w}{\partial x} - w \frac{\partial u}{\partial x} \right),$$

распространяющемуся в сторону положительной полуоси x , как это показано в уравнении (III.В). Для трехмерного случая этот результат будет получен в § 4 гл. 15. Здесь мы приводим его в порядке иллюстрации рассматриваемого метода.

Помимо объяснения результатов опытов по дифракции электронов, указанная выше интерпретация величины $u^2 + w^2$, будучи использована при исследовании движения электронов в атомах, приводит к значениям плотности вероятности, сходным с упомянутыми для H, H₂, C, Si и Ge в гл. 1. Эти значения плотности в свою очередь используются для предсказания результатов ряда других опытов.

В табл. 13 уравнение Шредингера умышленно записано в неудобной и необычной форме, где волновая функция

$$\psi(x, t) = u(x, t) + iw(x, t) \quad (14.9)$$

обладает вещественной и мнимой компонентами. Величины u и w должны быть вещественными. Назовем их соответственно вещественной

и мнимой компонентами волновой функции ψ . Функцию ψ называют также *амплитудой вероятности*, чтобы отличить ее от *интенсивности*, которая пропорциональна квадрату амплитуды и входит в выражения, определяющие плотность и поток плотности. В рассматриваемом нами одномерном случае u и w обладают размерностью (длина)^{-1/2} (т. е. $см^{-1/2}$), а в трехмерном случае — размерностью $см^{-3/2}$, так как, будучи возведены в квадрат и проинтегрированы по длине ($см$), они должны дать единицу, т. е. просто безразмерное число. Отметим, что u и w всегда встречаются только в виде произведений, например u^2 или $u dw/dx$.

Функция ψ записана нами в комплексной форме, содержащей мнимую компоненту, просто из соображений математического удобства. Правила, применяемые при действиях над двумя вещественными числами u и w , могут быть сформулированы весьма кратко с помощью комплексных чисел, с учетом того, что i^2 всегда может быть заменено на -1 .

Если $\psi = u + iw$, то, по определению

$$\psi^* = u - iw \quad (14.10)$$

и

$$\psi^*\psi = (u - iw)(u + iw) = u^2 - i^2w^2 = u^2 + w^2. \quad (14.11)$$

Точно так же, абсолютное значение ψ равно

$$|\psi| = (\psi^*\psi)^{1/2} = \sqrt{u^2 + w^2}, \quad |\psi|^2 = \psi^*\psi. \quad (14.12)$$

Очевидно, что $\psi^*\psi$ всегда вещественно и положительно, за исключением того случая, когда $\psi = 0$ ¹⁾.

Воспользовавшись комплексными обозначениями, преобразуем уравнение (I.B) табл. 13 к обычному виду. Умножив первое из уравнений (I.B) на $i\hbar$, а второе на $-\hbar$ и сложив их, получаем:

левая часть равенства

$$i\hbar \frac{\partial u}{\partial t} - \hbar \frac{\partial w}{\partial t} = i\hbar \frac{\partial}{\partial t} (u + iw) = i\hbar \frac{\partial \psi}{\partial t}, \quad (14.13)$$

правая часть равенства

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + \mathcal{U} \psi = \left[\frac{1}{2m} \left(\frac{\pm \hbar}{i} \frac{\partial}{\partial x} \right)^2 + \mathcal{U} \right] \psi \equiv \mathcal{H} \psi, \quad (14.14)$$

так что

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H} \psi, \quad (14.15)$$

¹⁾ Другой пример того, как комплексные обозначения облегчают действия над вещественными величинами, приведен в примечании к § 7 гл. 7, а также в § 3 гл. 12. — *Прим. авт.*

где, по определению, \mathcal{H} — линейный оператор, стоящий в (14.14) в квадратных скобках. [Как мы увидим в дальнейшем, при рассмотрении члена $(\hbar/i)\partial/\partial x$ как импульса выясняется, что из двух знаков \pm здесь следует выбрать знак $+$.] Теперь можно утверждать следующее.

Для частицы, движущейся в одном измерении в поле $\mathcal{U}(x)$, уравнение Шредингера, зависящее от времени, имеет вид

$$i\hbar \frac{\partial \psi}{\partial t} = \mathcal{H}\psi, \quad \mathcal{H} \equiv -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \mathcal{U}(x), \quad (14.16)$$

где m — масса частицы, а \hbar — постоянная Планка, деленная на 2π .

Мы получили это уравнение косвенным путем с помощью табл. 13. Проанализируем теперь формулу (14.5) и уравнение (14.16), с тем чтобы подготовить формулировку общего „правила определения средних значений“ (14.25), представляющего собой один из основных постулатов квантовой теории. При этом будем предполагать, что $\psi(x, t)$ имеет форму волнового пакета, хотя соотношения, которые мы таким путем получим, с одинаковым успехом могут быть применены также и в более общих случаях¹⁾. Мы определим среднее положение электрона, среднее квадратичное отклонение его положения и, наконец, вычислим среднюю скорость движения волнового пакета. Эти вычисления будут проведены достаточно подробно с тем, чтобы характер процессов усреднения мог для нас стать вполне ясным.

Предположим, что в нашем распоряжении имеется экспериментальная установка, подобная упомянутой выше электронной дифракционной камере, с помощью которой можно изучать распределение многих (в общей сложности N) электронов, причем все эти электроны ведут себя столь сходным образом, что их движение может быть описано одной и той же волновой функцией

$$\psi(x, t) = u(x, t) + iw(x, t),$$

где время t отсчитывается от начала опыта. (Если бы этот опыт был повторен в дальнейшем для различных электронов, то t следовало бы отсчитывать от различных нулевых точек.) Предположим, далее, что мы определяем положение каждого из электронов в момент времени t . В таком случае, согласно (14.5), предполагая функцию ψ нормированной в соответствии с (14.6), можно найти долю $|\psi(x, t)|^2 dx$ электронов, координаты которых лежат в интервале между x и $x + dx$. В среднем мы получаем значения x , лежащие в этом интервале, как раз $N |\psi(x, t)|^2 dx$ раз. Ряд измеренных значений x_1, x_2, \dots, x_N даст

¹⁾ Математическая теория волновых функций излагается в § 5 и 6 гл. 15; частный случай волнового пакета рассмотрен на фиг. 65. — *Прим. авт.*

нам среднее значение x , равное

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_N)}{N} = \frac{N \int_{-\infty}^{+\infty} x |\psi(x, t)|^2 dx}{N} = \int_{-\infty}^{+\infty} x |\psi(x, t)|^2 dx. \quad (14.17)$$

По определению, эта величина представляет собой среднее значение координаты электронов в момент t (где t — время, прошедшее после начала опыта) для N повторенных опытов. Значения x будут при этом группироваться вблизи \bar{x} . Полезно определить среднюю величину разброса значений x . Это можно осуществить следующим образом. Из значений x , получаемых при каждом из N отдельных наблюдений¹⁾, вычтем \bar{x} с тем, чтобы ввести в рассмотрение отклонения $x_1 - \bar{x}$, $x_2 - \bar{x}$ и т. д., затем возведем эти отклонения в квадрат и найдем среднее значение²⁾. Такое среднее, представляющее собой среднее квадратичное отклонение, обозначается через $(x - \bar{x})^2$ и определяется формулой

$$\overline{(x - \bar{x})^2} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}, \quad (14.18)$$

которая может быть также записана в виде

$$\overline{(x - \bar{x})^2} = \overline{(x^2 - 2x\bar{x} + (\bar{x})^2)} = \bar{x}^2 - 2\bar{x}\bar{x} + (\bar{x})^2 = \bar{x}^2 - (\bar{x})^2. \quad (14.19)$$

По тем же соображениям, которые были использованы выше при вычислении \bar{x} , приходим к заключению, что

$$\bar{x}^2 = \frac{x_1^2 + x_2^2 + \dots + x_N^2}{N} = \int_{-\infty}^{+\infty} x^2 |\psi(x, t)|^2 dx. \quad (14.20)$$

Мы приводим эти вычисления столь подробно для того, чтобы показать, что знание плотности вероятности $|\psi|^2$ дает возможность вычислить среднее значение x и различных функций x (в данном случае среднее значение x^2). В дальнейшем мы встретимся также с вычисле-

¹⁾ В принципе мы производим опыт N раз и N раз регистрируем его результаты, хотя практически усреднение x может быть выполнено с помощью самого измерительного прибора, который мог бы, например, зарегистрировать суммарный заряд, сообщаемый пластине конденсатора N электронами. — *Прим. авт.*

²⁾ Среднее значение самих отклонений равно нулю, так как положительные и отрицательные отклонения встречаются одинаково часто. — *Прим. авт.*

нием средних квадратичных значений ряда других величин. Из соотношений (14.18) и (14.19) следует, что $\overline{x^2} = (\overline{x})^2$ лишь в том случае, если вероятность $|\psi(x, t)|^2$ отлична от нуля только для некоторого частного значения x (так, что $x - \overline{x} = 0$). Во всех остальных случаях $\overline{x^2}$ всегда превышает $(\overline{x})^2$.

Воспользуемся теперь определением \overline{x} , чтобы ознакомиться с одним из методов, дающих принципиальную возможность измерения средней скорости. Величина \overline{x} , которую мы будем в дальнейшем обозначать через $\overline{x}(t)$, зависит от времени t через посредство временной зависимости функции $\psi(x, t)$. Среднее положение электрона меняется при этом со скоростью

$$\begin{aligned} \frac{d\overline{x}(t)}{dt} &= \frac{d}{dt} \int_{-\infty}^{+\infty} x \psi^*(x, t) \psi(x, t) dx = \int_{-\infty}^{+\infty} \left[x \frac{\partial \psi^*}{\partial t} \psi + x \psi^* \frac{\partial \psi}{\partial t} \right] dx = \\ &= \int_{-\infty}^{+\infty} 2x \left[u \frac{\partial u}{\partial t} + w \frac{\partial w}{\partial t} \right] dx. \end{aligned} \quad (14.21)$$

Последняя форма записи этого равенства может быть получена непосредственно из выражения, определяющего $u^2 + w^2$. Значения производных по времени от u и w при этом могут быть определены с помощью уравнений (1.В) табл. 13, а значения производных по времени от функций ψ и ψ^* — из уравнений (14.16). Интегрируя далее по частям (при условии, что при $x = \pm\infty$ функция ψ обращается в нуль, как это должно быть в соответствии с нашими представлениями о волновом пакете), получаем окончательно:

$$\frac{\partial}{\partial t} \overline{x} = \int_{-\infty}^{+\infty} \psi^* \frac{1}{m} \frac{\hbar}{i} \frac{\partial}{\partial x} \psi dx, \quad (14.22a)$$

$$\frac{\partial}{\partial t} \overline{x} = \frac{\hbar}{m} \int_{-\infty}^{+\infty} \left(u \frac{\partial w}{\partial x} - w \frac{\partial u}{\partial x} \right) dx. \quad (14.22б)$$

Два последних выражения эквивалентны друг другу, так как

$$\int_{-\infty}^{+\infty} u \frac{\partial u}{\partial x} dx = \frac{1}{2} u^2 \Big|_{-\infty}^{+\infty} = 0; \quad (14.23)$$

аналогичный результат имеет также место для функции w . Уравнение (14.22a) представляет собой пример одного из общих правил квантовой механики, которое можно представить в следующем виде:

$$p = \frac{\hbar}{i} \frac{\partial}{\partial x}. \quad (14.24)$$

В этом определении p называют оператором, так как $p\psi$ означает применение дифференциальной операции ¹⁾, указанной в правой части (14.24), к функции ψ . Оператор p соответствует классическому импульсу частицы и (как это будет сказано ниже) при усреднении дает среднее значение $m dx/dt$. Другие операторы аналогичным образом соответствуют другим величинам, характеризующим движение электрона. Для всех этих операторов мы имеем следующее

Правило определения средних значений: выбрать оператор \mathcal{R} , соответствующий интересующей нас величине, применить его к волновой функции ψ , полученное выражение умножить на ψ^* и проинтегрировать. Результат представляет собой ожидаемое среднее значение, которое мы получили бы путем ряда измерений рассматриваемой величины с помощью опытов, для каждого из которых поведение электрона описывается волновой функцией ψ :

$$\langle \mathcal{R} \rangle = \int \psi^* \mathcal{R} \psi dx. \quad (14.25)$$

С двумя примерами таких операторов мы уже встретились: координате частицы соответствует оператор x , а ее импульсу p или величине $m dx/dt$ — оператор $(\hbar/i)\partial/\partial x$. Средние значения ²⁾ этих величин определяются выражениями

$$\langle x \rangle = \int_{-\infty}^{+\infty} \psi^* x \psi dx \quad (14.26)$$

и

$$\langle p \rangle = \int_{-\infty}^{+\infty} \psi^* p \psi dx = \int_{-\infty}^{+\infty} \psi^* \frac{\hbar}{i} \frac{\partial}{\partial x} \psi dx. \quad (14.27)$$

Следует отметить, что среднее значение p , по крайней мере в некоторых опытах, представляет собой непосредственно измеряемую величину. Например, в электронных лампах со скоростной модуляцией импульсы электронов, проходящих через пространство дрейфа, обуславливают создание сигналов в выходном контуре. Решая уравнение Шредингера, зависящее от времени, можно сделать определенные заключения о среднем значении величины dx/dt для электрона, движущегося в поле с потенциалом \mathcal{U} . С помощью методов, сходных с рассмотренными в гл. 7, можно показать, что электронный волновой пакет будет при этом перемещаться со скоростью, в точности равной

¹⁾ Следует отметить, что $\bar{x} = \int x |\psi|^2 dx$ может быть также записано в виде $\int \psi^* x \psi dx$, тогда как $\bar{p} = \int \psi^* p \psi dx$ не равно $\int p |\psi|^2 dx$. — *Прим. авт.*

²⁾ Средние значения мы будем обозначать как через $\langle x \rangle$, так и через \bar{x} . — *Прим. авт.*

той скорости, которой электрон должен обладать согласно уравнению движения Ньютона. Если этот же электронный пучок в некоторой точке своего пути встретится с поверхностью кристалла, то при этом будут наблюдаться диффракционные явления, определяемые длиной волны электрона.

Рассмотрим теперь средние значения и операторы, соответствующие потенциальной и кинетической энергии электрона. Поскольку электрон, находящийся в точке x , обладает потенциальной энергией $\mathcal{U}(x)$, среднее значение последней, по соображениям, сходным с теми, которые были использованы выше при вычислении $\langle x^2 \rangle$, равно

$$\langle \mathcal{U} \rangle = \int_{-\infty}^{+\infty} \mathcal{U}(u^2 + w^2) dx = \int_{-\infty}^{+\infty} \mathcal{U} |\psi|^2 dx = \int_{-\infty}^{+\infty} \psi^* \mathcal{U} \psi dx. \quad (14.28)$$

Покажем далее, что оператор, соответствующий кинетической энергии, должен иметь вид

$$E_{\text{кин.}} = \frac{p^2}{2m} = \frac{1}{2m} \left(\frac{\hbar}{i} \frac{\partial}{\partial x} \right) \left(\frac{\hbar}{i} \frac{\partial}{\partial x} \right) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}, \quad (14.29)$$

откуда

$$\bar{E}_{\text{кин.}} = \int \psi^* \left(-\frac{\hbar^2}{2m} \right) \frac{\partial^2 \psi}{\partial x^2} dx. \quad (14.30)$$

Причиной такого выбора оператора $E_{\text{кин.}}$ является то обстоятельство, что можно показать, что сумма $\langle E_{\text{кин.}} \rangle + \langle \mathcal{U} \rangle$ остается постоянной во времени [т. е. доказать, что $d(\langle E_{\text{кин.}} \rangle + \langle \mathcal{U} \rangle)/dt = 0$] независимо от вида решений уравнений (1.В) табл. 13. Этот результат мы получим, если под знаком интеграла возьмем производные по времени, подставим затем их значения из уравнений (1.В) и проинтегрируем полученное выражение по частям. (Эту программу мы осуществим в следующем параграфе при изучении правил действия над операторами.) Таким образом:

$$\frac{d}{dt} (\langle E_{\text{кин.}} \rangle + \langle \mathcal{U} \rangle) = 0. \quad (14.31)$$

Этот результат может быть также сформулирован проще — путем введения оператора, соответствующего полной энергии, который мы обозначим через \mathcal{H} :

$$E_{\text{кин.}} + \mathcal{U} = \frac{p^2}{2m} + \mathcal{U} = -\frac{\hbar^2 \partial^2}{2m \partial x^2} + \mathcal{U} \equiv \mathcal{H}, \quad (14.32)$$

так что

$$\langle E_{\text{кин.}} \rangle + \langle \mathcal{U} \rangle = \langle \mathcal{H} \rangle \quad \text{и} \quad \frac{d \langle \mathcal{H} \rangle}{dt} = 0.$$

Мы стремились показать, что квантовая механика представляет собой математическую теорию, дающую возможность сделать ряд заключений о характере движения электронов. Порядок изложения

при этом был совершенно иным, чем в большинстве руководств по квантовой механике, так как мы предполагаем, что многие читатели имеют некоторое представление о линиях передачи и поэтому рассмотрение аналогии между плотностью энергии и плотностью вероятности могло бы оказаться полезным. По этой же причине в рассмотрение были введены две вещественные функции u и w вместо математически эквивалентной им комплексной функции ψ . На основе сравнительно простой интерпретации $\psi^*\psi$ как плотности вероятности нами было дано определение среднего значения x , которое приводит, в частности, к возможности использовать оператор p для вычисления x .

Методы теории операторов кратко излагаются в следующем параграфе; позднее они будут использованы при введении понятия собственных функций оператора энергии.

§ 3. ОПЕРАТОРНЫЕ МЕТОДЫ ¹⁾

В квантовой механике фундаментальное значение имеет тот факт, что основные операторы являются *эрмитовыми*. Оператор \mathcal{R} называется эрмитовым, если для любых двух волновых функций φ и ψ удовлетворяется соотношение

$$\int_a^b \varphi^* \mathcal{R} \psi dx = \int_a^b (\mathcal{R} \varphi)^* \psi dx. \quad (14.33)$$

Выполнение этого соотношения зависит от поведения волновых функций на пределах интегрирования, определяемых заданными граничными условиями. Рассмотрим два случая: (I) a и b конечны и функция ψ должна обращаться в нуль на обоих пределах ²⁾. Как будет показано в § 5, это соответствует бесконечному значению потенциальной энергии вне интервала ($a - b$), так что электроны отражаются при этом от границ. (II) Значения $a = 0$ и $b = A$ отвечают одной и той же точке циклической или периодической модели (см. фиг. 55 и фиг. 58, стр. 171 и 181), так что ψ принимает одно и то же значение при $x = a$ и $x = b$. В дальнейшем на протяжении данной главы мы будем предполагать, что рассматриваемые волновые функции удовлетворяют одному из этих двух граничных условий.

Очевидно, что эрмитовым является оператор:

$$\mathcal{R} = \text{const}, \quad (14.34)$$

где const — положительное или отрицательное вещественное число, например \hbar/m (хотя применение такого оператора представляет

¹⁾ С вопросами, рассматриваемыми в § 3 и 4, читатель может подробнее ознакомиться по имеющимся курсам квантовой механики [12—15].—Прим. ред.

²⁾ Случай $a = -\infty$ и $b = +\infty$ можно рассматривать, как предельный случай условия (I).—Прим. авт.

собой тривиальную операцию, согласно общей терминологии теории операторов, числа тоже называют операторами), а также оператор

$$\mathcal{R} = \mathcal{R}(x), \quad (14.35)$$

где \mathcal{R} — любая вещественная функция x , например x, x^2 или $\mathcal{U}(x)$.

Результат менее очевиден в случае операторов

$$p = \left(\frac{\hbar}{i}\right) \frac{\partial}{\partial x} \quad (14.36)$$

и

$$E_{\text{кин.}} = \left(\frac{1}{2} m\right) p^2 = \left(\frac{\hbar^2}{2m}\right) \frac{\partial^2}{\partial x^2}. \quad (14.37)$$

Путем непосредственных вычислений можно показать, что оператор p является эрмитовым; исходя из эрмитовых свойств оператора p , можно доказать, что оператор $E_{\text{кин.}}$ также эрмитов. Подставляя в (14.33) вместо \mathcal{R} оператор p и интегрируя по частям, получаем

$$\begin{aligned} \int_a^b \varphi^* p \psi dx &= \frac{\hbar}{i} \int_a^b \varphi^* \frac{\partial \psi}{\partial x} dx = \\ &= \frac{\hbar}{i} \varphi^* \psi \Big|_a^b - \frac{\hbar}{i} \int_a^b \frac{\partial \varphi^*}{\partial x} \psi dx = \\ &= \int_a^b \left(\frac{\hbar}{i}\right)^* \frac{\partial \varphi^*}{\partial x} \psi dx = \int_a^b \left(\frac{\hbar}{i} \frac{\partial \varphi}{\partial x}\right)^* \psi dx = \int_a^b (p\varphi)^* \psi dx, \end{aligned} \quad (14.38)$$

так как, согласно граничным условиям (I) и (II), член $\varphi^* \psi|_a^b$ равен нулю. Переходя к рассмотрению оператора $E_{\text{кин.}}$, отметим, что если операторы \mathcal{R} и \mathcal{S} — эрмитовы, а r и s — два любых вещественных числа, то оператор

$$r\mathcal{R} + s\mathcal{S} \quad (14.39)$$

также эрмитов. Далее, если оператор \mathcal{R} эрмитов, то этим свойством обладает и оператор \mathcal{R}^2 ; полагая

$$\chi_1 = \mathcal{R}\psi \quad \text{и} \quad \chi_2 = \mathcal{R}\varphi,$$

имеем

$$\begin{aligned} \int_a^b \varphi^* \mathcal{R}^2 \psi dx &= \int_a^b \varphi^* \mathcal{R} \chi_1 dx = \int_a^b (\mathcal{R}\varphi)^* \chi_1 dx = \text{и т. д.} = \\ &= \int_a^b (\mathcal{R}^2 \varphi)^* \psi dx. \end{aligned} \quad (14.40)$$

Применяя это соотношение к оператору p^2 , а соотношение (14.39) к оператору $1/2 mp^2$, убеждаемся в том, что оператор $E_{кин.}$ эрмитов. Применяя (14.39) к операторам $E_{кин.}$ и \mathcal{U} , приходим к заключению, что соотношению (14.33) удовлетворяет также и оператор \mathcal{H} .

Методом интегрирования по частям, использованным выше при исследовании свойств оператора p , можно также воспользоваться, чтобы доказать важную теорему, согласно которой среднее значение $E_{кин.}$ никогда не может быть отрицательным. Если состояние электрона описывается волновой функцией ψ , то выражение, определяющее значение $\langle E_{кин.} \rangle$, может быть преобразовано следующим образом:

$$\begin{aligned} \langle E_{кин.} \rangle &= \frac{1}{2m} \int_a^b \psi^* p^2 \psi dx = \frac{1}{2m} \int_a^b \left(\frac{\hbar}{i} \frac{\partial \psi}{\partial x} \right)^* \left(\frac{\hbar}{i} \frac{\partial \psi}{\partial x} \right) dx = \\ &= \frac{\hbar^2}{2m} \int_a^b \left| \frac{\partial \psi}{\partial x} \right|^2 dx. \end{aligned} \quad (14.41)$$

Как вскоре будет показано, можно найти такой оператор \mathcal{S} , для которого $\langle \mathcal{S} \rangle$ будет представлять собой производную по времени от $\langle \mathcal{X} \rangle$, где \mathcal{X} — некоторый другой оператор. В операторной форме это запишется так ¹⁾:

$$\langle \mathcal{S} \rangle \equiv \int \psi^* \mathcal{S} \psi dx = \frac{d}{dt} \langle \mathcal{X} \rangle \equiv \frac{d}{dt} \int \psi^* \mathcal{X} \psi dx. \quad (14.42)$$

Докажем, что для оператора \mathcal{S} , который мы определим как $d\mathcal{X}/dt$, имеет место следующее соотношение:

$$\mathcal{S} \equiv \frac{d\mathcal{X}}{dt} = \frac{1}{i\hbar} (\mathcal{X}\mathcal{H} - \mathcal{H}\mathcal{X}). \quad (14.43) \text{)}$$

Сперва выясним смысл правой части этого операторного равенства для случая $\mathcal{X} = x$; с этой целью применим оператор \mathcal{S} к некоторой функции ψ и упростим получающийся результат

$$\begin{aligned} \mathcal{S}\psi &= \frac{1}{i\hbar} (\mathcal{X}\mathcal{H} - \mathcal{H}\mathcal{X})\psi = \\ &= \frac{1}{i\hbar} \left[x \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \mathcal{U} \right) \psi - \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \mathcal{U} \right) x\psi \right]. \end{aligned} \quad (14.44)$$

Так как

$$\frac{\partial^2}{\partial x^2} x\psi = \frac{\partial \psi}{\partial x} + \frac{\partial}{\partial x} \left(x \frac{\partial \psi}{\partial x} \right) = 2 \frac{\partial \psi}{\partial x} + x \frac{\partial^2 \psi}{\partial x^2}, \quad (14.45)$$

¹⁾ В дальнейшем мы будем опускать пределы интегрирования, подразумевая, что они определяются граничными условиями. — *Прим. авт.*

²⁾ Читатель может убедиться в том, что если операторы A и B эрмитовы, то соотношению (14.33) удовлетворяет также и оператор $i(AB - BA)$. — *Прим. авт.*

получаем

$$\mathfrak{S}\psi = \frac{1}{i\hbar} \left[+ \frac{\hbar^2}{2m} 2 \frac{\partial \psi}{\partial x} \right] = \frac{\hbar}{mi} \frac{\partial \psi}{\partial x} = \frac{p}{m} \psi. \quad (14.46)$$

Этот результат был получен выше в § 2 при вычислении \bar{x} [см. формулы (14.22), (14.24) и (14.25)]. Для того чтобы доказать справедливость соотношения (14.43) в общем случае, запишем

$$\begin{aligned} \langle \mathfrak{S} \rangle &= \frac{d}{dt} \langle x \rangle = \int \left[\frac{\partial \psi^*}{\partial t} x \psi + \psi^* x \frac{\partial \psi}{\partial t} \right] dx = \\ &= \frac{1}{i\hbar} \int [(-\mathcal{H}\psi)^* x \psi + \psi^* x \mathcal{H}\psi] dx = \\ &= \frac{1}{i\hbar} \int \psi^* (-\mathcal{H}x + x\mathcal{H}) \psi dx, \end{aligned} \quad (14.47)$$

что эквивалентно формулам (14.42) и (14.43).

Отсюда следует, что $\int \psi^* \psi dx$ не зависит от времени, как это и должно быть, поскольку $|\psi|^2 dx$ есть вероятность.

Форма соотношения (14.43), а также результаты, полученные при вычислении \bar{x} , свидетельствуют о том, что два оператора не всегда коммутируют друг с другом. По определению:

$$\text{Если } \mathfrak{R}\mathfrak{S} = \mathfrak{S}\mathfrak{R}, \text{ то операторы } \mathfrak{R} \text{ и } \mathfrak{S} \text{ коммутируют.} \quad (14.48)$$

Именно в тех случаях, когда квантовомеханические операторы не коммутируют друг с другом, квантовая механика приводит к результатам, отличным от результатов, даваемых классической механикой.

Если в соотношении (14.43) под оператором \mathfrak{R} подразумевается оператор \mathcal{H} , то скорость изменения $\langle \mathcal{H} \rangle$ равна

$$\frac{d\langle \mathcal{H} \rangle}{dt} = \frac{1}{i\hbar} (\mathfrak{R}\mathcal{H} - \mathcal{H}\mathfrak{R}) = \frac{1}{i\hbar} (\mathcal{H}\mathcal{H} - \mathcal{H}\mathcal{H}) = 0, \quad (14.49)$$

так что среднее значение \mathcal{H} остается постоянным. Этот результат представляет собой квантовомеханическую формулировку принципа сохранения энергии. Вообще любой оператор, коммутирующий с оператором \mathcal{H} , обладает постоянным средним значением.

В гл. 5 большое внимание было уделено исследованию собственных функций уравнения Шредингера. Эти функции удовлетворяют уравнению

$$\mathcal{H}\psi = \mathcal{E}\psi, \quad (14.50)$$

которое будет рассмотрено нами в следующем параграфе. На языке теории операторов значение энергии \mathcal{E} и волновая функция ψ , удовлетворяющие этому уравнению и граничным условиям, называются соответственно собственным значением и собственной функцией оператора \mathcal{H} . Все операторы, рассматриваемые в квантовой механике, обладают собственными функциями и собственными значениями. Например, в случае циклических граничных условий [т.е. в том случае,

когда $\psi(x+A) = \psi(x)$ оператор p обладает собственными функциями, удовлетворяющими уравнению

$$p\psi_s = p_s\psi_s. \quad (14.51)$$

Здесь $\psi_s = \exp(2\pi i s x/A)$, где s — любое целое число, а множитель $p_s = 2\pi\hbar s/A = \hbar s/A$, как в этом легко убедиться, подставив в (14.51) выражение (14.24), определяющее оператор p .

Соотношения, устанавливающие связь между операторами и результатами их применения к различным волновым функциям, играют в квантовой механике весьма существенную роль. Краткий анализ этих соотношений был предпринят нами главным образом с тем, чтобы показать, что операторы \mathcal{H} и p эрмитовы, а также для того, чтобы ввести в рассмотрение оператор $d\mathcal{R}/dt$.

Выше мы ознакомились с методами применения операторов квантовой механики к волновым функциям. В заключение отметим, что возможен совершенно иной метод построения теории, более общий, но в то же время более абстрактный и менее доступный пониманию при первом ознакомлении. Мы имеем в виду теорию, в основе которой лежат соотношения между самими операторами, причем волновые функции в рассмотрение не вводятся. Например, исходя из соотношения

$$px - xp = \frac{\hbar}{i}, \quad (14.52)$$

которое трактуется при этом как основное [тогда как мы ввели бы его с помощью формулы (14.24)], и выражения, определяющего энергию гармонического осциллятора с частотой ω :

$$\mathcal{H} = \left(\frac{1}{2} m\right) (p^2 + m^2\omega^2 x^2), \quad (14.53)$$

Дирак [1] получил соотношения, характеризующие все наблюдаемые свойства осциллятора, имея при этом дело только с операторами и вообще не вводя в рассмотрение волновую функцию $\psi(x, t)$. Такой подход типичен для теории Дирака. Другой метод, при применении которого за основу также берутся сами операторы и соотношения между ними, был разработан Нейманном [2].

§ 4. СОБСТВЕННЫЕ ФУНКЦИИ ОПЕРАТОРА ЭНЕРГИИ

В квантовой механике особую роль играют собственные функции оператора энергии, т. е. такие волновые функции, которые соответствуют некоторым определенным значениям энергии. При этом необходимо, чтобы при измерении энергии в N сходных опытах (см. § 2, где речь шла об измерении x в N опытах) каждое измерение давало бы одно и то же значение $E_{\text{вн.}} + \mathcal{U} = \mathcal{H}$. Правило определения средних значений (14.25) не дает, однако, ответа на вопрос

о том, какие значения энергии должны получаться при отдельных измерениях, но дает только среднее значение $\langle \mathcal{H} \rangle$. На основании этого правила можно также определить среднее значение \mathcal{H}^2 , которое мы обозначим через $\langle \mathcal{H}^2 \rangle$. По тем же соображениям, что и при вычислении $\langle x^2 \rangle$ и $\langle x \rangle$, мы приходим к заключению, что если $\langle \mathcal{H}^2 \rangle = \langle \mathcal{H} \rangle^2$, то при измерениях не должно наблюдаться разброса значений \mathcal{H} . Другими словами, если применение правила вычисления средних значений к операторам \mathcal{H} и \mathcal{H}^2 приводит к соотношению

$$\langle \mathcal{H} \rangle^2 = \langle \mathcal{H}^2 \rangle, \quad (14.54)$$

то это значит, что любое измерение энергии должно дать $\langle \mathcal{H} \rangle$. Мы убедимся в том, что, исходя из (14.54), можно доказать, что

$$\mathcal{H}\psi - \mathcal{E}\psi = 0, \text{ где } \mathcal{E} \equiv \langle \mathcal{H} \rangle = \int \psi^* \mathcal{H} \psi dx. \quad (14.55)$$

Это есть не что иное, как уравнение Шредингера для собственных функций, исследованных в § 2 гл. 5. Для одномерного случая, рассматриваемого в настоящей главе, этот результат может быть получен, если мы воспользуемся свойствами эрмитовости \mathcal{H} , \mathcal{E} и $(\mathcal{H} - \mathcal{E})$ (\mathcal{E} — просто число) при вычислении интеграла

$$\begin{aligned} \int |(\mathcal{H} - \mathcal{E})\psi|^2 dx &= \int [(\mathcal{H} - \mathcal{E})\psi]^* (\mathcal{H} - \mathcal{E})\psi dx = \\ &= \int \psi^* (\mathcal{H} - \mathcal{E}) [(\mathcal{H} - \mathcal{E})\psi] dx = \\ &= \int \psi^* [\mathcal{H}^2 - \mathcal{H}\mathcal{E} - \mathcal{E}\mathcal{H} + \mathcal{E}^2]\psi dx = \\ &= \langle \mathcal{H}^2 \rangle - \langle \mathcal{H} \rangle \mathcal{E} - \mathcal{E} \langle \mathcal{H} \rangle + \mathcal{E}^2 = \langle \mathcal{H}^2 \rangle - \mathcal{E}^2. \end{aligned} \quad (14.56)$$

Крайняя правая часть этого выражения равна нулю, согласно (14.54) и определению \mathcal{E} , как $\langle \mathcal{H} \rangle$. Следовательно, исходный интеграл также равен нулю и подинтегральная функция $|(\mathcal{H} - \mathcal{E})\psi|^2$, которая всегда либо положительна, либо равна нулю, обращается в нуль при всех значениях x . Тем самым уравнение (14.55) доказано.

Временная зависимость собственной функции, удовлетворяющей уравнению (14.55), носит очень простой характер. Действительно, из (14.16) и (14.55) имеем

$$i\hbar \frac{\partial \psi(x, t)}{\partial t} = \mathcal{H}\psi(x, t) = \mathcal{E}\psi(x, t). \quad (14.57)$$

Непосредственное интегрирование этого уравнения дает

$$\psi(x, t) = \psi(x, 0) e^{-i\mathcal{E}t/\hbar}. \quad (14.58)$$

Отсюда следует, что

$$\psi^*(x, t)\psi(x, t) = \psi^*(x, 0)\psi(x, 0),$$

так что плотность вероятности не зависит от времени и, следовательно, является постоянной в пространстве. По этой причине состояния, определяемые собственными функциями оператора энергии, называют также *стационарными состояниями*. Из (14.58) следует, что комплексная фаза волновой функции зависит от угловой частоты $\omega = -\mathcal{E}/\hbar$. Полагая $\omega = -2\pi\nu$, имеем

$$h\nu = \mathcal{E}. \quad (14.59)$$

Это показывает, что соотношение Планка между энергией и частотой справедливо и для частот, входящих в определение собственных функций.

Особенно важным свойством собственных функций является то, что они образуют полную систему функций, через которые может быть выражено любое решение уравнения (14.16). В этом отношении они аналогичны коэффициентам Фурье, через которые может быть выражена любая периодическая во времени функция. Мы не будем здесь касаться вопроса о полноте решений (подробнее см. [5]), но покажем, что собственные функции обладают свойством ортогональности, и выведем формулу разложения любой волновой функции по собственным функциям.

Две волновые функции ψ_1 и ψ_2 называются ортогональными¹⁾, если они удовлетворяют условию

$$\int \psi_1^* \psi_2 dx = 0. \quad (14.60)$$

Если ψ_1 и ψ_2 — собственные функции оператора \mathcal{H} , соответствующие двум различным собственным значениям \mathcal{E}_1 и \mathcal{E}_2 , то они ортогональны. Это можно доказать следующим образом. Пусть

$$\mathcal{H}\psi_1 = \mathcal{E}_1\psi_1, \quad \mathcal{H}\psi_2 = \mathcal{E}_2\psi_2. \quad (14.61)$$

В таком случае

$$\int \psi_1^* \mathcal{H}\psi_2 dx = \int \psi_1^* \mathcal{E}_2 \psi_2 dx = \mathcal{E}_2 \int \psi_1^* \psi_2 dx; \quad (14.62)$$

далее:

$$\int \psi_1^* \mathcal{H}\psi_2 dx = \int (\mathcal{H}\psi_1)^* \psi_2 dx = \mathcal{E}_1 \int \psi_1^* \psi_2 dx. \quad (14.63)$$

Вычитая соответственно левые и правые части этих выражений, получаем

$$(\mathcal{E}_2 - \mathcal{E}_1) \int \psi_1^* \psi_2 dx = 0. \quad (14.64)$$

¹⁾ Термин «ортогональный» связан с представлением о прямом угле. В гильбертовом пространстве волновых функций левая часть (14.60) соответствует скалярному произведению векторов, которое обращается в нуль, если эти векторы образуют друг с другом прямой угол. — *Прим. авт.*

Согласно предположению, $(\mathcal{E}_2 - \mathcal{E}_1) \neq 0$; следовательно, интеграл $\int \psi_1^* \psi_2 dx$ обращается в нуль и условие (14.60) тем самым удовлетворяется. Очевидно, что аналогичный результат может быть получен и для собственных функций любого другого оператора.

Если мы имеем полную систему нормированных собственных функций ψ_1, ψ_2, \dots , то любая волновая функция ψ , удовлетворяющая тем же граничным условиям, что и собственные функции, может быть разложена в ряд по этим функциям:

$$\psi = a_1 \psi_1 + a_2 \psi_2 + \dots = \sum_s a_s \psi_s. \quad (14.65)$$

Коэффициенты a_1, a_2, \dots могут быть найдены следующим образом. Умножим (14.65) на любую собственную функцию ψ_r^* и проинтегрируем по всему пространству:

$$\int \psi_r^* \psi dx = \sum_s a_s \int \psi_r^* \psi_s dx. \quad (14.66)$$

Единственный отличный от нуля член в правой части этого равенства есть $a_r \int \psi_r^* \psi_r dx = a_r$, так как функция ψ_r ортогональна по отношению к остальным функциям ψ . Отсюда следует, что

$$a_r = \int \psi_r^* \psi dx. \quad (14.67)$$

Приведенное выше доказательство ортогональности собственных функций оператора энергии основано на предположении о том, что собственное значение, соответствующее функции ψ_r , отлично от собственного значения, отвечающего функции ψ_s . Во многих случаях две (или большее число) волновые функции обладают, однако, одним и тем же собственным значением, которое называется при этом *вырожденным собственным значением*. Если ψ_r и ψ_s — две такие функции, нормированные, но не ортогональные, то из них можно построить две новые волновые функции

$$\psi_a = a_r \psi_r + a_s \psi_s \text{ и } \psi_b = b_r \psi_r + b_s \psi_s, \quad (14.68)$$

причем коэффициенты здесь могут быть выбраны так, что ψ_a и ψ_b будут нормированы и ортогональны¹⁾. Этот метод может быть обобщен и на тот случай, когда число вырожденных функций ψ больше двух; любой ряд собственных функций таким путем может быть

1) Например, функции ψ_r и $[\psi_s - (\int \psi_r^* \psi_s dx) \psi_r]$ ортогональны, а вторая из них будет нормированной, если мы разделим ее на $(1 - |\int \psi_r^* \psi_s dx|^2)^{1/2}$.
Прим. авт.

приведен в ортонормированную форму, для которой

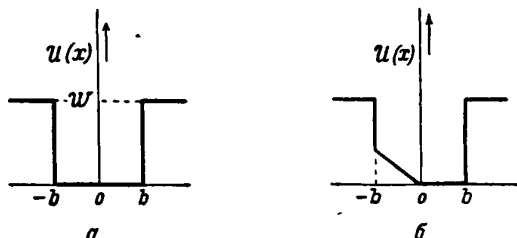
$$\int \psi_r^* \psi_s dx = \begin{cases} 1 & \text{при } r = s \\ 0 & \text{при } r \neq s \end{cases} = \delta_{rs}. \quad (14.69)$$

Знаком δ_{rs} (δ_{rs} называется *дельта-функцией*) часто пользуются для сокращенного обозначения правой части выражения (14.69).

Очевидно, что независимо от того, ортонормированы ли функции ψ_n , коэффициенты разложения (14.65) определяются формулой (14.67).

§ 5. РЕШЕНИЯ ДЛЯ ПРОСТОГО „АТОМА“

Проиллюстрируем теорию энергетических уровней электрона в атоме и в кристалле, рассмотрев в настоящем параграфе простую одномерную модель атома, а в следующем параграфе — периодическую



Фиг. 119. Упрощенная одномерная модель атома.

a — симметричная потенциальная яма;
б — несимметричная потенциальная яма.

последовательность таких атомов. Рассматриваемая модель атома изображена на фиг. 119, *a*; она характеризует область пониженной потенциальной энергии электрона. Эту модель часто называют квадратной потенциальной ямой. Наша задача заключается в отыскании решений уравнения Шредингера для электрона, движущегося в поле, описываемом кривой фиг. 119, *a*. Уравнение Шредингера имеет в этом случае вид

$$\mathcal{H} \psi(x) = -\frac{\hbar^2}{2m} \frac{d^2 \psi(x)}{dx^2} + U(x) \psi(x) = \mathcal{E} \psi(x), \quad (14.70a)$$

причем

$$U(x) = 0, \quad |x| < b, \quad U(x) = W, \quad |x| > b, \quad (14.70б)$$

при условии, что $\psi(x)$ обращается в нуль при $x = \pm \infty$. Это условие представляет собой предельный случай граничного условия (I) (см. § 3). Мы покажем, что любой паре задачных значений глубины ямы W и ее ширины $2b$ отвечает определенный ряд решений, описывающих связанные состояния электрона в яме.

Прежде всего покажем, что в случае симметричной ямы волновая функция ψ может быть либо „четной“, либо „нечетной“ функцией x .

Предположим, что $\psi_1(x)$ есть решение уравнения, соответствующее энергии \mathcal{E}_1 . В таком случае функция

$$\psi_2(x) = \psi_1(-x) \quad (14.71)$$

также является решением уравнения, так как яма симметрична относительно $x=0$; это решение отвечает тому же значению энергии \mathcal{E}_1 . Исследуем теперь функции

$$g(x) = \psi_1(x) + \psi_2(x) \quad (14.72a)$$

и

$$u(x) = \psi_1(x) - \psi_2(x). \quad (14.72b)$$

Если мы разложим ψ_1 и ψ_2 в ряд по степеням x , то в выражении (14.72a) взаимно сократятся все нечетные, а в выражении (14.72b) — все четные члены. Очевидно, что

$$g(x) = g(-x) \quad \text{и} \quad u(x) = -u(-x). \quad (14.72в)^1$$

В отличие от функций χ и φ , вводимых в рассмотрение ниже, функции $g(x)$ и $u(x)$ мы будем относить к любым значениям переменной x . Ниже мы получим явное выражение для $g(x)$ и $u(x)$, найдя с этой целью решения уравнения (14.70) для областей внутри ямы и вне ямы в отдельности, а затем сшив эти решения при $x = +b$ и $x = -b$. При этом мы убедимся, что собственные функции, удовлетворяющие граничным условиям, существуют только при некоторых определенных значениях энергии.

В области ямы волновое уравнение имеет вид

$$\frac{d^2\psi}{dx^2} = -\frac{2m}{\hbar^2} \mathcal{E} \psi. \quad (14.73)$$

Четные и нечетные решения этого уравнения:

$$\chi_g(x) \equiv \cos\left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{1/2} x, \quad -b < x < b, \quad (14.74a)$$

$$\chi_u(x) \equiv \sin\left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{1/2} x, \quad -b < x < b; \quad (14.74b)$$

функции χ_g и χ_u вводятся здесь в рассмотрение только для значений $|x| \leq b$.

Вне ямы волновое уравнение имеет вид

$$\frac{d^2\psi}{dx^2} = \frac{2m(\mathcal{W} - \mathcal{E})}{\hbar^2} \psi. \quad (14.75)$$

При $\mathcal{E} > \mathcal{W}$ решение этого уравнения носит синусоидальный характер и не обращается в нуль на бесконечном расстоянии от ямы. Из таких решений могут быть построены волновые пакеты, описывающие

¹⁾ Обозначения g и u выбраны здесь в соответствии с немецкими терминами „gerade“ и „ungerade“ (четный и нечетный). — *Прим. авт.*

движение свободных частиц вне потенциальной ямы. Эти решения не описывают, однако, интересующих нас связанных состояний электрона в яме. Если $\mathcal{E} < \mathcal{W}$, то частица не обладает энергией, достаточной для того, чтобы покинуть яму, несмотря на то, что волновая функция, как мы увидим, не обращается при этом в нуль вне ямы. В этом случае возможные решения уравнения имеют вид

$$\varphi_- = \exp \left[2m \frac{\mathcal{W} - \mathcal{E}}{\hbar^2} \right]^{1/2} x, \quad (14.76a)$$

$$\varphi_+ = \exp - \left[2m \frac{\mathcal{W} - \mathcal{E}}{\hbar^2} \right]^{1/2} x, \quad (14.76b)$$

где под $[2m(\mathcal{W} - \mathcal{E})/\hbar^2]^{1/2}$ подразумевается положительное значение квадратного корня, а функции φ вводятся в рассмотрение только для значений $|x| > b$. Функция φ_+ возрастает беспредельно при $x \rightarrow -\infty$ и при отрицательных значениях x не может быть использована в качестве решения; она может, однако, являться решением уравнения при положительных значениях x , так как при $x = +\infty$ обращается в нуль. По этой причине функция φ снабжена индексом „плюс“. Ввиду симметрии функций g и u можно ограничиться исследованием решения для $x > 0$, так как решение для $x < 0$, как это будет показано ниже, легко может быть получено путем простой замены x на $-x$.

Исследуем прежде всего решения четного типа для случая $x > 0$. Эти решения имеют, очевидно, вид

$$g = A\chi_g(x) = A \cos \left(\frac{2m\mathcal{E}}{\hbar^2} \right)^{1/2} x, \quad -b < x < b, \quad (14.77)$$

$$g = B\varphi_+(x) = B \exp - \left[2m \frac{\mathcal{W} - \mathcal{E}}{\hbar^2} \right]^{1/2} x, \quad b < x. \quad (14.78)$$

Как следует сшить эти решения при $x = b$? Необходимое требование заключается в том, что в точке b как g , так и dg/dx должны быть непрерывными функциями x . Это следует непосредственно из дифференциального уравнения, согласно которому в любом бесконечно малом интервале δx изменение производной $d\psi/dx$ определяется выражением

$$\delta \left[\frac{d\psi}{dx} \right] = \int_{b-(\delta x/2)}^{b+(\delta x/2)} \frac{2m}{\hbar^2} [\mathcal{U}(x) - \mathcal{E}] \psi(x) dx. \quad (14.79)$$

Поскольку ψ и $[\mathcal{U}(x) - \mathcal{E}]$ конечны, в то время как $\mathcal{U}(x)$ претерпевает разрыв непрерывности при $x = b$, интеграл (14.79) при $\delta x \rightarrow 0$ обращается в нуль, так что при $x = b$ производная $d\psi/dx$ непрерывна. Еще более очевидно, что при $x = b$ должна быть непрерывной сама функция ψ , так как прерывность изменения ψ привела бы к бесконечно большому значению производной $d\psi/dx$.

Таким образом, если A и B могут быть выбраны так, чтобы функции g и dg/dx оставались непрерывными при $x=b$, мы получим волновую функцию, удовлетворяющую уравнению (14.70) и требуемым граничным условиям при $x=+\infty$ и $x=-\infty$. Такая волновая функция, по определению, является собственной функцией уравнения (14.70) и может существовать только при определенных собственных значениях энергии, которые мы найдем, потребовав, чтобы решения (14.77) и (14.78) переходили друг в друга непрерывным образом. Из условия непрерывности функции g следует, что

$$A\chi_g(b) = B\varphi_+(b). \quad (14.80a)$$

Условие непрерывности $g' = dg/dx$ дает

$$A\chi'_g(b) = B\varphi'_+(b). \quad (14.80b)$$

Разделив второе уравнение на первое, получаем

$$\frac{1}{\chi_g} \frac{d\chi_g}{dx} = \frac{1}{\varphi_+} \frac{d\varphi_+}{dx} \text{ при } x=b. \quad (14.80в)$$

Введем сокращенные обозначения:

$$\alpha = \sqrt{2m\mathcal{W}} \frac{b}{\hbar}, \quad (14.81a)$$

$$\beta = \sqrt{2m(\mathcal{W} - \mathcal{E})} \frac{b}{\hbar} = \sqrt{\alpha^2 - \theta^2}, \quad (14.81b)$$

$$\theta = \sqrt{2m\mathcal{E}} \frac{b}{\hbar}. \quad (14.81в)$$

Воспользовавшись ими, имеем

$$\frac{1}{\chi_g} \frac{d\chi_g}{dx} = -\left(\frac{\theta}{b}\right) \operatorname{tg} \theta, \quad (14.82)$$

$$\frac{1}{\varphi_+} \frac{d\varphi_+}{dx} = -\frac{\beta}{b}, \quad (14.83)$$

так что условие существования собственных функций приобретает вид

$$-\theta \operatorname{tg} \theta = -\beta = -\sqrt{\alpha^2 - \theta^2}. \quad (14.84)$$

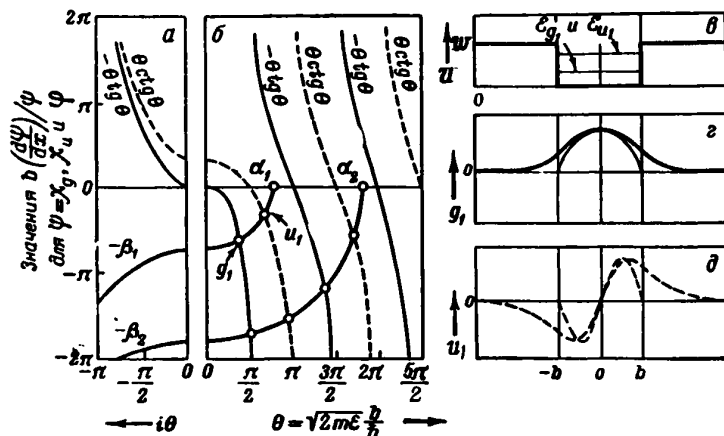
(Сохраняем знаки „минус“ в этом выражении для удобства его сопоставления с выражением, определяющим величины γ и μ в следующем параграфе.)

Аналогичное исследование функции u приводит к соотношению

$$\theta \operatorname{ctg} \theta = -\beta = -\sqrt{\alpha^2 - \theta^2}. \quad (14.85)$$

Как θ , так и β зависят от \mathcal{E} , и задача отыскания собственных значений заключается, таким образом, в отыскании тех значений \mathcal{E} , при которых удовлетворяется написанное выше уравнение.

На фиг. 120 показано, что эта задача может быть решена графически путем построения обеих частей уравнения как функций параметра θ , который в данном случае эквивалентен \mathcal{E} ; отрицательные значения \mathcal{E} соответствуют здесь мнимым значениям θ . Уравнение $\beta = \sqrt{a^2 - \theta^2}$ есть при этом уравнение окружности. Сплошными линиями показан ход функции $-\theta \operatorname{tg} \theta$, пунктирными — ход $\theta \operatorname{ctg} \theta$. При мнимых значениях θ , соответствующих отрицательным значениям \mathcal{E} , решений не существует. (Проверьте это, попытавшись построить функции g и u при отрицательных значениях \mathcal{E} .)



Фиг. 120. Значения волновых функций в области потенциальной ямы.

Волновые функции g_1 и u_1 , ход которых изображен тонкими линиями, соответствуют $\mathcal{W} = \infty$. Масштаб произвольный.

В случае $\alpha_1 \approx 2$, исследованном на графике, существуют только два решения; ход соответствующих волновых функций показан на графике справа. При возрастании α число решений увеличивается и вид функций g_1 и u_1 соответственно изменяется. При $\alpha \rightarrow \infty$ θ принимает значения, равные $\pi/2$ и π ; волновые функции обращаются при этом в нуль на краях ямы и равны нулю вне ямы. Это и есть граничное условие (1), рассмотренное выше в § 3.

Как было отмечено в § 1, волновое уравнение сходно с уравнением для электрической линии передачи, в которой постоянные цепи зависят от энергии. В табл. 13 (пятая строка) показано, что если емкость по отношению к земле C , приходящаяся на единицу длины линии, остается постоянной, то ток, протекающий в линии, аналогичен волновой функции ψ для рассматриваемого нами случая. Мы можем брать решения в форме

$$\psi(x, t) = \psi(x) e^{i\omega t} \quad (14.86)$$

и

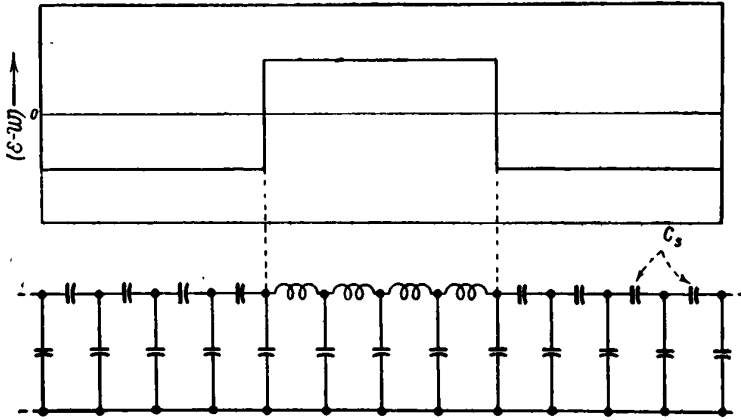
$$I(x, t) = I(x) \cos \omega t. \quad (14.87)$$

Это даст нам уравнения, приведенные в табл. 13:

$$\frac{d^2 I(x)}{dx^2} = -LC\omega^2 I, \quad (14.88)$$

$$\frac{d^2 \psi}{dx^2} = -\frac{2m}{\hbar^2} [\mathcal{E} - \mathcal{U}] \psi. \quad (14.89)$$

В случае решения типа g_1 величина $LC\omega^2$ для $|x| < b$ при этом положительна, а для $|x| > b$ — отрицательна. Отрицательное значение $LC\omega^2$ соответствует отрицательному значению $L\omega$; последнее может

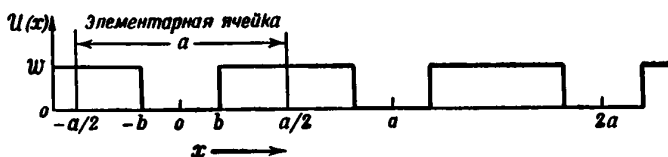


Фиг. 121. Аналогия связанного состояния электрона в потенциальной яме со случаем линии передачи.

быть получено путем замены ряда индуктивностей рядом емкостей C_s , как это показано на фиг. 121, где распределенные постоянные изображены как сосредоточенные. Энергия, накапливаемая в индуктивностях, не может быть потеряна на концах такой цепи, так как емкостное звено ослабителя не может служить передатчиком мощности. Этот случай аналогичен случаю области отрицательной кинетической энергии, внутри которой плотность вероятности отлична от нуля, но в которую электрон может проникнуть только на незначительное расстояние. Случай $\mathcal{W} \rightarrow \infty$ соответствует случаю $L\omega \rightarrow -\infty$, или же тому случаю, когда емкость ряда конденсаторов C_s стремится к нулю. При этом индуктивность линии разомкнута на концах, так что при $x = \pm b$, I обращается в нуль.

§ 6. РЕШЕНИЯ ДЛЯ ПРОСТОГО „КРИСТАЛЛА“

В этом параграфе будут исследованы волновые функции электрона в одномерном кристалле, который получился бы, если бы „атомы“, рассмотренные в предыдущем параграфе, были расположены в периодической последовательности, как это показано на фиг. 122. В этом случае существуют зоны энергии, для которых волновая функция имеет вид функции Блоха, а также зоны, для которых амплитуда волновой функции возрастает экспоненциально в том или ином направлении в кристалле. Последний случай может представлять интерес в связи с изучением поверхностных состояний; при этом мы не получаем, однако, волновых функций, характеризующих движение электрона в периодической решетке и удовлетворяющих



Фиг. 122. Периодический ход потенциала для одномерной модели кристалла.

периодическим граничным условиям; этот случай соответствует значениям энергии, лежащим в *запрещенных* энергетических зонах.

Методы, которыми можно воспользоваться, чтобы показать, как возникают энергетические зоны, весьма разнообразны. Можно исходить из рассмотрения произвольного периодического поля, сосредоточенного в некоторой конечной области пространства, и на основании весьма общих соображений доказать, что при этом должны существовать зоны разрешенных значений энергии [4, 5]. С другой стороны, можно также исследовать какую-либо специальную модель кристалла и произвести для нее подробный расчет энергетических зон [5—7]. Мы выберем промежуточный путь исследования и предположим, что для каждой из элементарных ячеек кристалла кривая потенциальной энергии симметрична относительно центра ячейки, причем центры ячеек расположены в точках $x = 0, \pm a, \pm 2a, \dots$. В таком случае для ячейки, центр которой находится в точке $x = 0$, в качестве волновых функций могут быть выбраны функции типа $g(x)$ и $u(x)$. Мы покажем, что поставленная нами задача сводится к определению двух величин: γ и μ , представляющих собой отношения производных от функций g и u к самим функциям, вычисленные в точке $x = a/2$:

$$\left. \begin{aligned} \gamma &= \frac{(dg/dx)}{g} \\ \mu &= \frac{(du/dx)}{u} \end{aligned} \right\} \text{при } x = a/2. \quad (14.90)$$

Мы исследуем также значения γ и μ как функции энергии и рассмотрим вопрос о том, какова связь между энергетическими зонами электронов в кристалле и энергетическими уровнями электронов в бесконечно удаленных друг от друга атомах¹⁾.

Прежде всего докажем, что любое решение волнового уравнения

$$\frac{d^2\psi}{dx^2} = \frac{2m}{\hbar^2} [\mathcal{U}(x) - \mathcal{E}] \psi, \quad (14.91)$$

где \mathcal{E} — энергия, может быть представлено в виде суммы двух решений ψ_1 и ψ_2 , обладающих следующими свойствами:

$$\psi_1(x+a) = e^{iPa/\hbar} \psi_1(x), \quad (14.92)$$

$$\psi_2(x+a) = e^{-iPa/\hbar} \psi_2(x). \quad (14.93)$$

Здесь P вещественно для разрешенных зон и мнимо для запрещенных зон. Наше доказательство будет основано на исследовании величин γ и μ ; более общее доказательство этого важного результата, который известен как теорема Флоке [9], будет дано в дальнейшем в виде задачи.

Введем в рассмотрение величину β (не имеющую никакого отношения к параметру β какого-либо другого раздела), определив ее посредством уравнения

$$2\beta = \frac{Pa}{\hbar}, \text{ так что } e^{iPa/\hbar} = e^{2i\beta}, \quad (14.94)$$

и будем искать соотношение, связывающее β с γ и μ . При этом мы покажем, что существуют три области возможных значений β :

β вещественно и $-\frac{\pi}{2} < \beta < +\frac{\pi}{2}$, разрешенная зона;

β мнимо, запрещенная зона;

$\beta = \frac{\pi}{2} +$ мнимое число, запрещенная зона.

В случае разрешенной зоны при переходе от одной из ячеек к следующей фаза изменяется меньше, чем на 180° , и ее абсолютное значение одинаково в любой ячейке. В двух других случаях волновая функция убывает экспоненциально в том или ином направлении, подобно функциям φ_+ и φ_- , исследованным выше в § 5 [см. (14.76a) и (14.76b)]. Границы разрешенных зон соответствуют при этом тем значениям энергии, при которых становятся возможными вещественные

¹⁾ Задача отыскания „поверхностных состояний, связанных с периодическим потенциалом“, т. е. задача отыскания волновых функций, которые постепенно затухают в направлениях, перпендикулярных к поверхности конечного кристалла, также может быть сведена к определению величин γ и μ [8]. — *Прим. авт.*

значения величины β . Уравнение, определяющее β , может быть найдено следующим образом.

Для ячейки, расположенной в точке $x = 0$, любая волновая функция, удовлетворяющая уравнению (14.91), может быть записана в виде

$$\psi_1 = Ag(x) + Bu(x). \quad (14.95)$$

Потребуем теперь, чтобы изменение x от $-a/2$ до $+a/2$ соответствовало умножению значений ψ_1 и $d\psi_1/dx$ на $\exp(iPa/\hbar) = \exp(2i\beta)$. В таком случае на левой границе ячейки 1, т. е. ячейки, центр которой расположен в точке $x = a$, волновая функция и ее производная будут равны соответственно произведению множителя $\exp(2i\beta)$ на значения волновой функции и ее производной, вычисленные для левой границы ячейки 0. Поскольку волновое уравнение есть уравнение второго порядка, его решение (при заданном \mathcal{E}) известно, если известны производная от функции и сама функция; следовательно, значение волновой функции в ячейке 1 равно произведению множителя $\exp(2i\beta)$ на значение этой функции в ячейке 0. Аналогичное соотношение справедливо также и для ячейки 2 по отношению к ячейке 1 и т. д.; таким образом, уравнение (14.92) удовлетворяется. Поскольку точка $x = 0$ является центром симметрии ячейки, из свойств симметрии функций $g(x)$ и $u(x)$ [см. (14.72)] следует, что

$$\psi_1\left(-\frac{a}{2}\right) = Ag\left(-\frac{a}{2}\right) + Bu\left(-\frac{a}{2}\right) = Ag\left(\frac{a}{2}\right) - Bu\left(\frac{a}{2}\right), \quad (14.96a)$$

$$\psi_1'\left(-\frac{a}{2}\right) = Ag'\left(-\frac{a}{2}\right) + Bu'\left(-\frac{a}{2}\right) = -Ag'\left(\frac{a}{2}\right) + Bu'\left(\frac{a}{2}\right). \quad (14.96b)$$

В дальнейшем мы будем опускать аргумент $a/2$, подразумевая при этом, что значения g, g', u и u' вычислены при $x = +a/2$. Далее имеем

$$\psi_1\left(\frac{a}{2}\right) = Ag + Bu = e^{2i\beta}\psi_1\left(-\frac{a}{2}\right) = e^{2i\beta}(Ag - Bu), \quad (14.97a)$$

$$\psi_1'\left(\frac{a}{2}\right) = Ag' + Bu' = e^{2i\beta}\psi_1'\left(-\frac{a}{2}\right) = e^{2i\beta}(-Ag' + Bu'), \quad (14.97b)$$

или же

$$(1 - e^{2i\beta})gA + (1 + e^{2i\beta})uB = 0, \quad (14.98a)$$

$$(1 + e^{2i\beta})g'A + (1 - e^{2i\beta})u'B = 0. \quad (14.98b)$$

Эти уравнения могут быть решены относительно A и B в том и только в том случае, если определитель, составленный из коэффициентов при A и B , равен нулю. Это дает

$$\begin{aligned} (1 - e^{2i\beta})^2 gu' - (1 + e^{2i\beta})^2 g'u &= \\ &= 4e^{2i\beta} [-\sin^2 \beta gu' - \cos^2 \beta g'u] = 0. \end{aligned} \quad (14.99)$$

Выражение, стоящее в квадратной скобке, должно обращаться в нуль, откуда следует, что

$$\operatorname{tg}^2 \beta = -\frac{g'u}{gu'} = -\frac{\gamma}{\mu}. \quad (14.100)$$

Решение этого уравнения β может иметь вещественное значение только в том случае, если отношение γ/μ отрицательно. Далее, если β есть решение уравнения (14.100), то решением этого уравнения являются также и $-\beta$ и $\beta + n\pi$, где n — целое число. Таким образом, если $P = 2\beta\hbar/a$ есть одно из решений, то остальные решения, удовлетворяющие условию периодичности, рассмотренному в конце § 5 гл. 5, имеют вид

$$P + \frac{2\pi\hbar}{a} = P + \frac{h}{a}, \quad P + \frac{2h}{a}, \quad P + \frac{3h}{a} \text{ и т. д.} \quad (14.101)$$

Теперь становится ясным, почему P вводится в теорию через посредство формул (14.97а) и (14.97б) в показатель множителя $\exp(iPa/\hbar)$, который остается неизменным при добавлении к P величины, целой кратной h/a .

То обстоятельство, что всегда существует пара решений P и $-P$, показывает, что в зоне Бриллюэна квантовые состояния, соответствующие значениям P и $-P$, обладают одинаковыми энергиями. Этот результат был использован в трехмерной форме в § 4 гл. 6 и § 4 гл. 8, чтобы показать, что в случае заполненной зоны Бриллюэна, а также для частично заполненной зоны в условиях теплового равновесия ток равен нулю. Трехмерная форма решения легко может быть получена и исследуется в дальнейшем (см. задачи).

Независимо от значения отношения γ/μ всегда существуют, однако, два решения волнового уравнения, дающие две функции вида (14.92) и (14.93). При данном значении энергии волновое уравнение может иметь только два линейно независимых решения и любое другое решение может быть представлено в виде их суммы. Особый случай имеет место, когда решением является $P=0$ или $h/2a$, так как при этом функции (14.92) и (14.93) становятся эквивалентными. В этом специальном случае одно из решений есть решение типа (14.92); оно имеет либо одно и то же значение в каждой ячейке, либо при переходе от ячейки к ячейке изменяет свой знак на противоположный. Это решение представляет собой предельное значение решения, получаемого путем сложения функций ψ_{+P} и ψ_{-P} , причем для разрешенной зоны оно дает функцию типа косинуса, а для запрещенной — функцию типа гиперболического косинуса. Второе решение, однако, отлично от решений (14.92) и (14.93) и представляет собой функцию промежуточного типа. При переходе от одной из ячеек к следующей его численное значение изменяется на постоянную величину, так что оно удовлетворяет условию

$$\psi(x + na) = \psi(x) + nC, \quad (14.102)$$

где C — постоянная. Получение этих решений для конкретных случаев не представляет трудностей, они исследуются в дальнейшем (см. задачи).

При положительных значениях отношения γ/μ следует различать два случая. Если $\gamma/\mu < 1$, то решение имеет вид $\beta = \pm i\gamma$; это приводит к уравнению

$$-\text{th}^2 \chi = -\frac{\gamma}{\mu}, \quad (14.103)$$

которое всегда может быть решено при вещественном значении χ . В этом случае при переходе от одной из ячеек к следующей волновая функция умножается на

$$e^{2\chi} = e^{\pm 2\chi}, \quad (14.104)$$

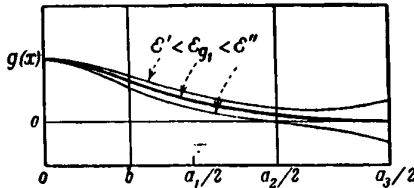
так что одно из возможных решений возрастает в направлении $+x$, а второе убывает. При $\gamma/\mu > 1$ решение имеет вид $\beta = \pm i\chi + \pi/2$; это приводит к уравнению

$$-\text{cth}^2 \chi = -\frac{\gamma}{\mu}, \quad (14.105)$$

и множитель, на который функция умножается при переходе от ячейки к ячейке, равен

$$e^{2i\beta} = e^{\pm 2\chi + \pi i} = -e^{\pm 2\chi}, \quad (14.106)$$

так что решение изменяется при этом так же, как и в



Фиг. 123. Зависимость функции $g(x)$ от изменений значения энергии в уравнении Шредингера (приближенно).

случае (14.104), но при переходе от ячейки к ячейке изменяет свой знак.

Исследуем теперь вопрос о возникновении энергетических зон из атомных энергетических уровней при взаимном сближении атомов, находящихся на бесконечном расстоянии друг от друга (т. е. при переходе от больших значений a к малым значениям a). Этот вопрос мы проанализируем графически, построив кривые, характеризующие возможные значения γ и μ .

На фиг. 123 показан вид функции $g(x)$ для „атомного“ собственного значения \mathcal{E}_{g1} , а также для двух слегка отличных значений энергии. При $\mathcal{E} = \mathcal{E}_{g1}$ величина γ не зависит от a и равна (см. (14.766)):

$$\gamma = \frac{d\varphi_+/dx}{\varphi_+} = -\left[\frac{2m(\mathcal{W} - \mathcal{E}_{g1})}{\hbar^2} \right]^{1/2}. \quad (14.107)$$

При $\mathcal{E}' < \mathcal{E}_{g1}$ волновая функция не является собственной функцией, так как функция типа косинуса, дающая решение в области потенциальной ямы, не может быть сшита непосредственно с экспонен-

циально убывающей функцией φ_+ ; для этого требуется еще некоторая вспомогательная функция φ_- . При больших значениях x доминирующую роль играет член φ_- , при этом γ положительно и равно

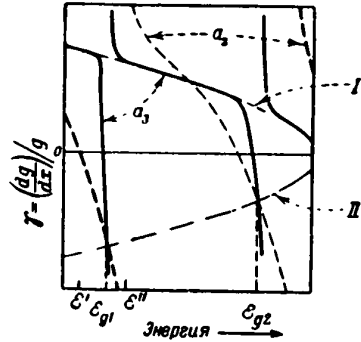
$$\gamma = + \left[\frac{2m(\mathcal{E}'' - \mathcal{E}')}{\hbar^2} \right]^{1/2}. \quad (14.108)$$

Аналогичные условия имеют место и при $\mathcal{E}'' > \mathcal{E}_{g1}$; единственное различие заключается в том, что коэффициент при φ_- имеет противоположный знак; это не влияет, однако, на значение (14.108). Зависимость γ от \mathcal{E} иллюстрируется в этом случае фиг. 124. При больших значениях a , таких, например,

как a_3 , величина γ определяется формулой (14.108) при любых энергиях, за исключением очень узких областей энергии, расположенных вблизи значений \mathcal{E}_{g1} и \mathcal{E}_{g2} ; при этом предполагается, что области ямы соответствуют два четных энергетических уровня. Если \mathcal{E} несколько меньше \mathcal{E}_{g1} , будучи слегка сдвинуто в сторону \mathcal{E}' , то g возрастает и dg/dx обращается в нуль, так что γ при этом возрастает. При незначительном увеличении \mathcal{E} , $g(a_3)$ падает до нуля, в результате чего $\gamma = -\infty$; дальнейшее увеличение \mathcal{E} приводит к большим положительным значениям γ с последующим резким спаданием γ до значений, определяемых формулой (14.108). При меньшем значении постоянной решетки $a = a_2$ изменение энергии оказывает не столь резкое влияние и интервалу энергии от \mathcal{E}' до \mathcal{E}'' соответствует область значений γ от 0 до $\pm\infty$. При всех значениях a , превышающих $2b$ (см. фиг. 122), и условии $\mathcal{E} = \mathcal{E}_{g1}$, γ принимает, однако, значение (14.107); точно такой же результат имеет место и для других энергетических уровней.

Совершенно аналогичные результаты получаются и при анализе возможных пределов изменения величины μ .

Исследуем теперь соотношение между значениями γ и μ и структурой энергетических зон. На фиг. 125 показана структура зон для двух значений a .левой части рисунка отвечает большое значение постоянной решетки $a = a_3$. При $\mathcal{E} = \mathcal{E}_{g1}$ функция g экспоненциально убывает вне ямы, однако функция u ведет себя при этом по-иному. Следовательно, в этой точке μ имеет положительное значение, определяющееся формулой (14.108). При $\mathcal{E} = \mathcal{E}_{u1}$ поведение μ аналогично поведению γ при $\mathcal{E} = \mathcal{E}_{g1}$. Остальные особенности кривых для μ и γ



Фиг. 124. Зависимость γ от энергии на границе ячейки ($x = a/2$) (приближенно).

- I — значения γ определены из уравнения (14.108);
- II — значения γ определены из уравнения (14.107).

При $x = a/2$ функции $\psi(x)$ и $g(x)$ принимают одинаковые значения и имеют также одинаковые производные. Поскольку они удовлетворяют одному и тому же дифференциальному уравнению второго порядка, $\psi(x)$ должно равняться $g(x)$ при любых значениях x . Но это невозможно, так как при $x = 0$ имеем $u = 0$, в то время как функция g при этом отлична от нуля. Таким образом, предположение $\gamma = \mu$ приводит нас к абсурду (этот результат можно обобщить, показав, что при данном значении \mathcal{E} функции γ и μ не могут также одновременно обращаться в бесконечность).

Согласно другой теореме, доказательство которой рассматривается нами в дальнейшем в виде задачи, как $d\gamma/d\mathcal{E}$, так и $d\mu/d\mathcal{E}$ отрицательны, за исключением того случая, когда $\gamma = \pm\infty$ или $\mu = \pm\infty$.

Исследуя при $a = a_3$ значения энергии, близкие к \mathcal{E}_{g1} , мы видим, что разрешенная энергетическая зона возникла из той области, где γ отрицательно (т. е. области, где отрицательно отношение γ/μ). У левого края этой зоны $\gamma = 0$ и $P = 0$, что соответствует центру зоны Бриллюэна. У правого края зоны $\gamma = -\infty$ и $\beta = \pi/2$, а $P = h/2a$, что отвечает границе зоны Бриллюэна; волновая функция меняет при этом знак от ячейки к ячейке. При более высоких значениях \mathcal{E} величина γ положительна и превышает μ . Поскольку кривые для γ и μ не могут пересечься, γ все время больше μ , и отношение γ/μ превышает единицу до тех пор, пока не будут достигнуты значения $\mu = 0$ и $\gamma/\mu = \infty$. При промежуточных значениях γ/μ имеем $\beta = i\chi + \pi/2$, что соответствует экспоненциально убывающему или экспоненциально возрастающему решению с переменным знаком. В следующей разрешенной зоне, центр которой находится в точке \mathcal{E}_{u1} , значение P убывает от h/a до нуля и в ближайшей запрещенной зоне экспоненциальные функции не изменяют своего знака.

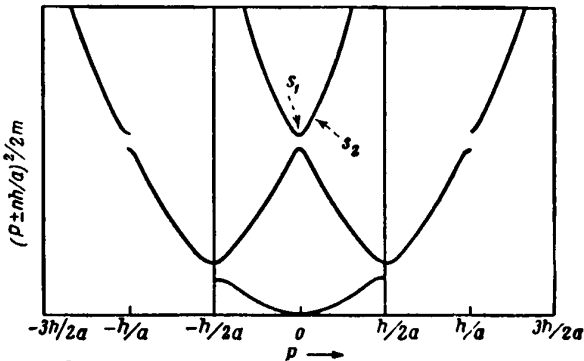
Эти результаты находятся в согласии с графиками, приведенными ранее на фиг. 58 и 66 (стр. 181 и 202), где минимальные значения энергии наблюдаются при $P = 0$ и $P = h/2a$.

Очевидно также, что разрешенные энергетические зоны возникают из атомных уровней, так как энергия \mathcal{E}_{g1} отвечает области отрицательных значений γ и по крайней мере при больших значениях постоянной решетки μ при этом положительно, так что отношение γ/μ отрицательно, что соответствует разрешенной зоне.

Из теоремы, согласно которой $d\gamma/d\mathcal{E}$ и $d\mu/d\mathcal{E}$ отрицательны и изменяются периодически от $+\infty$ до $-\infty$, следует, что зоны разрешенной энергии должны существовать при любых условиях. В этом можно убедиться следующим образом. В любой точке, где $\gamma = 0$, μ либо конечно, либо бесконечно велико. Если μ конечно, то отношение γ/μ меняет знак при $\gamma = 0$, так что эта точка соответствует границе между разрешенной и запрещенной зонами. Если $\mu = \infty$, то это значит, что μ должно изменяться от $-\infty$ до $+\infty$ при переходе от положительных значений γ к отрицательным. Следовательно, и в этом

случае зоны разрешенной энергии существуют по обе стороны точки $\gamma = 0$, так как отношение γ/μ при этом отрицательно.

Такой случай будет иметь место для нашей модели кристалла, если постоянная решетки a равна $2b$, т. е. ширине ямы, так что дно ямы простирается на весь кристалл. При этих условиях функции γ и μ становятся тождественными функциям $b\psi'(b)/\psi(b)$, ход которых изображен на фиг. 120, отличаясь от них лишь отсутствием множителя b . Эти функции содержат в качестве множителей $\operatorname{tg}\theta$ и $\cos\theta$;



Фиг. 126. Энергетические зоны для почти свободных электронов (приближенно).

одна из них проходит через нуль, в то время как другая стремится к бесконечности, и наоборот. Таким образом, при условии $a = 2b$ зон запрещенной энергии не существует и энергетический спектр совершенно непрерывен, как это должно было бы быть для электрона, движущегося в отсутствие внешних сил. В этом случае можно (хотя этот прием и является несколько искусственным) классифицировать волновые функции так же, как и прежде, вводя в рассмотрение зону Бриллюэна для постоянной решетки, равной $2b$. Кривая, иллюстрирующая изменение энергии, приобретает при этом вид простой „параболы“, характерной для свободного электрона

$$\mathcal{E} = \frac{1}{2} m v^2 = \frac{p^2}{2m} \quad (14.111)$$

и расположенной в первой зоне Бриллюэна (фиг. 126 иллюстрирует несколько иной случай, о чем будет сказано ниже).

Если a немного превышает $2b$, то между ямами остаются очень узкие стенки и электрон почти свободен. Однако если P есть целое кратное $\hbar/2a$, то при этом удовлетворяется условие отражения Брэгга и групповая скорость равна нулю. На диаграмме значений γ и μ этому соответствует незначительное смещение точки $\gamma = 0$. Влияние

этого эффекта на кривые, характеризующие изменение энергии свободного электрона, показано на фиг. 126. Характерным является резкое искривление графика вблизи границ зоны. Это искривление можно интерпретировать¹⁾ как соответствующее очень малой эффективной массе электрона, так как в этой области величина

$$v = \frac{d\mathcal{E}}{dP} \quad (14.112)$$

быстро изменяется при незначительных изменениях P , так что

$$\frac{dv}{dP} \gg \frac{1}{m}, \quad (14.113)$$

где m — масса свободного электрона, а $1/m$ — мера кривизны параболы для свободного электрона.

В связи с малостью эффективной массы возникает следующий вопрос. Предположим, что электрон находится в состоянии, соответствующем точке S_1 на фиг. 126, т. е. находится в третьей энергетической зоне, обладая при этом наименьшей возможной энергией. Если к кристаллу приложено поле, то электрон приобретает ускорение и проходит через кристалл за значительно меньшее время, чем это потребовалось бы в случае совершенно свободного электрона, покоившегося вначале неподвижно. Как это может быть? Откуда электрон, находящийся в третьей зоне, получает свою дополнительную энергию? Ответ заключается в том, что электрон уже обладал этой дополнительной энергией раньше. Можно показать, что состояние S_1 отвечает суперпозиции двух бегущих волн, соответствующих значениям $P = +(\hbar/a)$ и $P = -(\hbar/a)$. Таким образом, хотя состояние S_1 и не дает результирующего электронного тока, оно характеризуется значительной энергией. Влияние приложенного поля заключается в том, что поле смещает электрон в состояние S_2 вследствие изменения условий отражения обеих бегущих волн. В результате вся кинетическая энергия переходит в энергию поступательного движения слева направо и электронный волновой пакет приобретает значительное ускорение. Как только состояние S_2 достигнуто, дальнейшая передача кинетической энергии, однако, прекращается и эффективная масса, отвечающая дополнительному ускорению, приобретает свое нормальное значение.

Обычно масса дырок и электронов — порядка массы свободного электрона. При применении приближения почти свободных электронов к исследованию зонной структуры полупроводников создается ошибочное впечатление о крайней малости эффективной массы у краев зоны. Различие между действительным положением дела и случаем почти свободного электрона заключается в следующем. В случае свободного электрона значения $\gamma = 0$ встречаются совместно с бесконечными

¹⁾ Теория групповой скорости излагается в § 2 и 4 гл. 15. — *Прим. авт.*

значениями величины μ , и наоборот. Малое возмущение слегка смещает эти значения γ и μ и приводит к появлению очень узких зон запрещенной энергии, которым соответствуют очень малые значения эффективных масс у границ разрешенных зон. В действительности, в случае волновых функций, отвечающих валентной связи, и волновых функций электронов в зоне проводимости ситуация будет совершенно иной, и в общем случае между величинами γ и μ , или, вернее, между их трехмерными аналогами, не существует какого-либо простого соотношения. При расстояниях порядка атомных размеров ширина разрешенных зон становится сравнимой с энергией связи, достигая нескольких электрон-вольт; ширина запрещенных зон при этом — того же порядка величины. Исходя из соображений размерности, можно, таким образом, утверждать, что эффективная масса будет удовлетворять приближенному соотношению вида

$$\Delta \mathcal{E} = \frac{P_{\max}^3}{2m_{\text{эфф.}}} = \frac{\hbar^2}{8a^2 m_{\text{эфф.}}}$$

Поскольку $\Delta \mathcal{E} \approx 0,1$ атомных единиц и $\hbar/a \approx 2\pi\hbar/4a_0 = = (\pi/2)\hbar/a_0 = \pi/2$ атомных единиц импульса, это дает

$$m_{\text{эфф.}} \approx \frac{(\pi/2)^2}{8 \cdot 0,1} \approx 3 \text{ электронные массы.} \quad (14.114)$$

Мы привели здесь эти рассуждения лишь для того, чтобы показать, что в случае полупроводников с сильной связью эффективные массы порядка массы свободного электрона должны быть скорее правилом, чем исключением ¹⁾.

Что же касается числа собственных функций, отвечающих каждой энергетической зоне, то для его определения можно непосредственно воспользоваться результатами, полученными при исследовании этого вопроса в гл. 5. В случае четного числа элементарных ячеек необходимо учесть то обстоятельство, что волновые функции, соответствующие значениям $P = \hbar/2a$ и $P = -\hbar/2a$, эквивалентны; по этой причине в этом случае существует только одно решение, которое может удовлетворять периодическим граничным условиям, так как второе решение изменяется согласно уравнению (14.102) (см. также задачу 2).

§ 7. ФУНКЦИИ БЛОХА ДЛЯ ТРЕХ ИЗМЕРЕНИЙ

В этом параграфе мы докажем теорему Блоха, согласно которой волновые функции электрона в кристалле имеют вид

$$\psi(\mathbf{r}) = e^{i\mathbf{P} \cdot \mathbf{r}/\hbar} u_{\mathbf{P}}(\mathbf{r}), \quad (14.115)$$

¹⁾ По этому вопросу см., например, работы [16—18]. — *Прим. ред.*

причем множитель $u_{\mathbf{r}}$ (\mathbf{r}) обладает периодичностью решетки¹⁾. Мы ограничимся при этом исследованием кристалла, для которого имеют место периодические граничные условия; в этом случае \mathbf{P} вещественно. Попутно убедимся в том, что объем зоны Бриллюэна равен h^3/V_a , где V_a — объем элементарной ячейки, и что плотность состояний в зоне есть V_A/h^3 , где V_A — объем кристалла, и, следовательно, число состояний в зоне (если пренебречь спином) равно $V_A/V_a = N$, где N — число элементарных ячеек в кристалле.

Основное свойство кристалла заключается в том, что посредством некоторых определенных векторов он может быть подвергнут такой трансляции, в результате которой каждый атом перемещается в положение, занимаемое эквивалентным ему атомом. Если потенциальная энергия электрона есть $\mathcal{U}(\mathbf{r})$, то при такой трансляции, которую мы будем обозначать через \mathbf{R} или \mathbf{R}_a , должно выполняться соотношение

$$\mathcal{U}(\mathbf{r} + \mathbf{R}) = \mathcal{U}(\mathbf{r}). \quad (14.116)$$

Совокупность возможных трансляций \mathbf{R} образует *решетку*. Это означает следующее. Выберем некоторую точку пространства за начало отсчета и рассмотрим все точки, смещенные по отношению к ней на векторы \mathbf{R} . Получаемая таким путем система точек представляет собой решетку, соответствующую структуре кристалла. (Если на фиг. 4 за начало отсчета выбрана вершина большого куба, то атомы, расположенные на гранях куба, находятся в узлах решетки, тогда как атомы, расположенные в объеме куба, находятся вне узлов решетки.) Векторы \mathbf{R} всегда могут быть выражены через ряд *элементарных векторов*:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3. \quad (14.117)$$

Элементарные векторы при этом сами должны являться векторами рассматриваемой решетки, и, далее, любой вектор решетки \mathbf{R} может быть представлен в форме

$$\mathbf{R} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3, \quad (14.118)$$

где α — целые числа, положительные или отрицательные. Элементарные векторы определяют параллелепипед, ребрами которого они являются; каждая вершина этого параллелепипеда служит узлом решетки. Узлы решетки могут при этом находиться только в вершинах параллелепипеда, так как в противном случае по крайней мере одно из чисел α , определяющих положение узла, лежало бы в интервале между 0 и 1, что противоречило бы формуле (14.118). Отсюда следует, что каждому параллелепипеду, ребрами которого служат векторы $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, отвечает только один узел решетки. Очевидно также, что все пространство может быть заполнено такими параллелепипедами

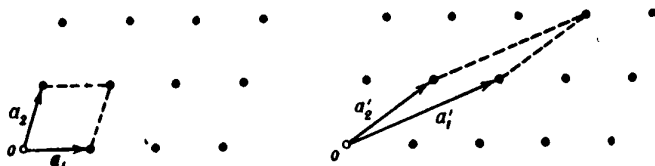
¹⁾ На протяжении всей книги u имеет именно этот смысл. Только в § 5 и 6 гл. 14 u обозначает „нечетный“ (ungerade). — *Прим. авт.*

в результате их повторения путем трансляции типа (14.118). По этой причине отдельный параллелепипед называют *элементарной ячейкой*. Кристалл может быть подразделен на такие элементарные ячейки, каждая из которых обладает одинаковой внутренней структурой; дальнейшее подразделение этих ячеек на эквивалентные части, которые могли бы быть совмещены друг с другом путем трансляции, уже невозможно.

Объем элементарной ячейки равен

$$V_a = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = \mathbf{a}_2 \cdot (\mathbf{a}_3 \times \mathbf{a}_1) = \mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2). \quad (14.119)$$

Здесь предполагается, что векторы \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 образуют правую систему (если они образуют левую систему, то можно ввести в рассмотрение



Фиг. 127. Две группы элементарных векторов в двухмерной решетке.

новый ряд элементарных векторов, положив $\mathbf{a}'_1 = -\mathbf{a}_1$). Из сказанного очевидно, что V_a есть объем, приходящийся на один узел решетки; этот объем должен, следовательно, иметь одно и то же значение для любых элементарных векторов.

На фиг. 127 показаны две возможности выбора элементарных векторов в случае двухмерной решетки. Читатель должен убедиться в том, что с помощью (14.118) векторы \mathbf{a}_1 и \mathbf{a}_2 могут быть выражены через векторы \mathbf{a}'_1 и \mathbf{a}'_2 , и наоборот.

Предположим теперь, что волновые функции удовлетворяют условиям периодичности, причем периоды равны

$$\mathbf{A}_1 = N_1 \mathbf{a}_1, \quad \mathbf{A}_2 = N_2 \mathbf{a}_2, \quad \mathbf{A}_3 = N_3 \mathbf{a}_3, \quad (14.120)$$

так что

$$\psi(\mathbf{r} + \mathbf{A}) = \psi(\mathbf{r}), \quad (14.121)$$

где \mathbf{A} — любая линейная комбинация векторов \mathbf{A}_i с положительными или отрицательными целыми числами в качестве коэффициентов. Объем кристалла, который мы обозначим через V_A , связан с объемом элементарной ячейки V_a соотношением

$$V_a = \mathbf{A}_1 \cdot (\mathbf{A}_2 \times \mathbf{A}_3) = N_1 N_2 N_3 V_a = N V_a. \quad (14.122)$$

Предположим далее, что функция $\psi_0(\mathbf{r})$, удовлетворяющая граничным условиям (14.121), является решением уравнения Шредингера,

соответствующим собственному значению энергии \mathcal{E} . (Существование таких решений может быть доказано без особого труда.) В таком случае очевидно, что функция

$$\psi_{\mathbf{R}}(\mathbf{r}) = \psi_0(\mathbf{r} + \mathbf{R}) \quad (14.123)$$

также является решением этого уравнения при той же энергии \mathcal{E} , так как она представляет собой значение функции ψ_0 , получаемое при смещении на $-\mathbf{R}$, а потенциальные энергии $\mathcal{U}(\mathbf{r})$ и $\mathcal{U}(\mathbf{r} + \mathbf{R})$ одинаковы.

Наша задача заключается в том, чтобы доказать, что если функция $\psi_0(\mathbf{r})$ существует, то она всегда может быть представлена в виде суммы функций Блоха, удовлетворяющих условию

$$\varphi_{\mathbf{B}}(\mathbf{r} + \mathbf{R}) = e^{2\pi i \mathbf{B} \cdot \mathbf{R}} \varphi_{\mathbf{B}}(\mathbf{r}), \quad (14.124)$$

$$\mathbf{B} \equiv \frac{\mathbf{P}}{h} \quad (14.125)$$

и являющихся собственными функциями при значении энергии, равном \mathcal{E} . Это значит, что при наличии периодических граничных условий нет необходимости рассматривать какие-либо функции, отличные от функций Блоха, так как все решения задачи могут быть через них выражены.

Введем в рассмотрение два дополнительных ряда векторов \mathbf{g}_i и \mathbf{G}_i , определив их следующим образом:

$$\mathbf{G}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{V_{\mathbf{a}}}, \quad \mathbf{G}_2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{V_{\mathbf{a}}}, \quad \mathbf{G}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{V_{\mathbf{a}}}. \quad (14.126)$$

Векторы \mathbf{G}_i называют обратными векторами кристаллической решетки. Векторы

$$\mathbf{g}_1 = \frac{\mathbf{A}_2 \times \mathbf{A}_3}{V_{\mathbf{A}}}, \quad \mathbf{g}_2 = \frac{\mathbf{A}_3 \times \mathbf{A}_1}{V_{\mathbf{A}}}, \quad \mathbf{g}_3 = \frac{\mathbf{A}_1 \times \mathbf{A}_2}{V_{\mathbf{A}}} \quad (14.127)$$

являются обратными векторами отдельных периодов решетки и не имеют специального названия. Из этих определений следует, что

$$\mathbf{a}_i \cdot \mathbf{G}_j = \delta_{ij} = \mathbf{A}_i \cdot \mathbf{g}_j \quad (14.128)$$

и что в \mathbf{G} -пространстве или пространстве обратной решетки объемы параллелепипедов, ребрами которых являются векторы \mathbf{G}_i и \mathbf{g}_i , равны соответственно:

$$V_{\mathbf{G}} = \mathbf{G}_1 \cdot \mathbf{G}_2 \times \mathbf{G}_3 \quad \text{и} \quad V_{\mathbf{g}} = \mathbf{g}_1 \cdot \mathbf{g}_2 \times \mathbf{g}_3. \quad (14.129)$$

Отсюда следует также, что между рассматриваемыми векторами существуют соотношения вида

$$\mathbf{a}_1 = \frac{\mathbf{G}_2 \times \mathbf{G}_3}{V_{\mathbf{G}}}, \quad \mathbf{A}_1 = \frac{\mathbf{g}_2 \times \mathbf{g}_3}{V_{\mathbf{g}}}. \quad (14.130)$$

В этом можно убедиться, учитывая, что вектор \mathbf{a}_1 удовлетворяет условию

$$\mathbf{a}_1 \cdot \mathbf{G}_j = \delta_{1j},$$

которое является единственным условием, определяющим \mathbf{a}_1 . Из теоремы векторной алгебры

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w}) \mathbf{v} - (\mathbf{u} \cdot \mathbf{v}) \mathbf{w} \quad (14.131)$$

следует, что если произведение $\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$ выразить через векторы \mathbf{G} , то это даст

$$\begin{aligned} V_{\mathbf{a}} &= \frac{(\mathbf{G}_2 \times \mathbf{G}_3) \cdot [(\mathbf{G}_3 \times \mathbf{G}_1) \times (\mathbf{G}_1 \times \mathbf{G}_2)]}{V_{\mathbf{G}}^3} = \\ &= \frac{(\mathbf{G}_2 \times \mathbf{G}_3) \cdot \{[(\mathbf{G}_3 \times \mathbf{G}_1) \cdot \mathbf{G}_2] \mathbf{G}_1 - [(\mathbf{G}_3 \times \mathbf{G}_1) \cdot \mathbf{G}_1] \mathbf{G}_2\}}{V_{\mathbf{G}}^3} = \\ &= \frac{V_{\mathbf{G}}^3}{V_{\mathbf{G}}^3} = \frac{1}{V_{\mathbf{G}}}, \quad (14.132) \end{aligned}$$

так как $(\mathbf{G}_3 \times \mathbf{G}_1) \cdot \mathbf{G}_1 = 0$. Аналогичное соотношение имеет место также и между объемами $V_{\mathbf{a}}$ и $V_{\mathbf{g}}$:

$$V_{\mathbf{a}} V_{\mathbf{g}} = 1. \quad (14.133)$$

Общее число различных трансляций, подчиняющихся периодическим граничным условиям, равно N ; эти трансляции могут быть представлены в форме

$$\mathbf{R}_{\alpha} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3, \quad (14.134)$$

где α — целые числа, изменяющиеся соответственно от 0 до $N_1 - 1$, $N_2 - 1$ и $N_3 - 1$. При этом не имеет смысла вводить в рассмотрение более высокие значения α , поскольку, например, при $\alpha_1 = N_1 + 2$ трансляция $\mathbf{R}' = \mathbf{R} - \mathbf{A}_1$ при $\alpha'_1 = 2$.

Можно также ввести в рассмотрение аналогичный ряд векторов \mathbf{B} , определяемых соотношениями

$$\mathbf{B}_i = \gamma_1 \mathbf{g}_1 + \gamma_2 \mathbf{g}_2 + \gamma_3 \mathbf{g}_3, \quad 0 \leq \gamma_i \leq N_i - 1. \quad (14.135)$$

Общее число возможных значений \mathbf{B} также равно N . Мы покажем в дальнейшем, что векторы \mathbf{B} определяют функции Блоха.

Воспользовавшись определениями \mathbf{R}_{α} и \mathbf{B}_i в форме (14.134) и (14.135), введем теперь в рассмотрение два ряда новых функций, получаемых из функции $\psi_0(\mathbf{r})$ в результате трансляции; с их помощью мы докажем, что $\psi_0(\mathbf{r})$ может быть разложено в ряд по функциям Блоха. Новые функции, которые мы обозначим через ψ_{α} и φ_i , где как α ,

так и γ есть совокупность трех целых чисел [см. (14.134) и (14.135)], определяются следующим образом:

$$\psi_{\alpha}(\mathbf{r}) = \psi_0(\mathbf{r} - \mathbf{R}_{\alpha}) \quad (14.136)$$

и

$$\varphi_{\gamma} = N^{-1/2} \sum_{\alpha} e^{2\pi i \mathbf{b}_{\gamma} \cdot \mathbf{R}_{\alpha}} \psi_{\alpha} \equiv \sum_{\alpha} S_{\gamma\alpha} \psi_{\alpha}. \quad (14.137)$$

Здесь $S_{\gamma\alpha}$ — матричные элементы матрицы преобразования, связывающие ψ_{α} и φ_{γ} и равные

$$S_{\gamma\alpha} \equiv N^{-1/2} e^{2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}_{\alpha}}. \quad (14.138)$$

Очевидно, каждая из функций φ является собственной функцией при одном и том же значении энергии \mathcal{E} , поскольку этим свойством обладает каждая из функций ψ_{α} .

Докажем далее, что

$$\varphi_{\gamma}(\mathbf{r} + \mathbf{R}) = \exp(2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}) \varphi_{\gamma}(\mathbf{r}), \quad (14.139)$$

т. е. что φ_{γ} имеет вид функции Блоха, причем $\mathbf{P} = 2\pi\hbar\mathbf{B}_{\gamma} = \hbar\mathbf{B}_{\gamma}$. Наконец, убедимся в том, что

$$\psi_{\alpha} = N^{-1/2} \sum_{\beta} \exp(-2\pi i \mathbf{B}_{\beta} \cdot \mathbf{R}_{\alpha}) \varphi_{\beta} \equiv \sum_{\beta} T_{\alpha\beta} \varphi_{\beta}, \quad (14.140)$$

т. е. в том, что любая функция ψ_0 , или же получаемая из нее функция ψ_{α} может быть представлена в виде суммы функций Блоха φ_{γ} .

Чтобы доказать, что φ_{γ} есть функция Блоха, исследуем функцию $\sqrt{N} \varphi_{\gamma}(\mathbf{r} + \mathbf{R})$:

$$\begin{aligned} \sqrt{N} \varphi_{\gamma}(\mathbf{r} + \mathbf{R}) &= \sum_{\alpha} \exp(2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}_{\alpha}) \psi_0(\mathbf{r} + \mathbf{R} - \mathbf{R}_{\alpha}) = \\ &= \sum_{\alpha'} \exp[2\pi i \mathbf{B}_{\gamma} \cdot (\mathbf{R}_{\alpha'} + \mathbf{R})] \psi_0(\mathbf{r} - \mathbf{R}_{\alpha'}) = \\ &= \exp(2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}) \sum_{\alpha'} \exp(2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}_{\alpha'}) \psi_{\alpha'}(\mathbf{r}). \end{aligned} \quad (14.141)$$

Мы заменили здесь $\mathbf{R}_{\alpha} - \mathbf{R}$ на $\mathbf{R}_{\alpha'}$, так что α' принимает также значения, лежащие вне ряда возможных значений α . Каждое $\mathbf{R}_{\alpha'}$, конечно, эквивалентно определенному \mathbf{R}_{α} , отличаясь от него на вектор \mathbf{A} . Замена, например, $\mathbf{R}_{\alpha'} = \mathbf{R}_{\alpha} + \mathbf{A}_1$ на \mathbf{R}_{α} не изменяет численного значения суммы (14.141), так как

$$e^{2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{A}_1} = \exp(2\pi i \gamma_1 \mathbf{g}_1 \cdot \mathbf{A}_1) = \exp(2\pi i \gamma_1) = 1, \quad (14.142)$$

так что экспоненциальный множитель остается при этом неизменным. Далее, если $\mathbf{R}_{\alpha'} = \mathbf{R}_{\alpha} + \mathbf{A}_1$, то $\psi_{\alpha} = \psi_{\alpha'}$. Таким образом, при суммировании α' может быть заменено на α и (14.141) приобретает в результате вид

$$\sqrt{N} \varphi_{\gamma}(\mathbf{r} + \mathbf{R}) = \exp(2\pi i \mathbf{B}_{\gamma} \cdot \mathbf{R}) \sqrt{N} \varphi_{\gamma}(\mathbf{r}). \quad (14.143)$$

Это и есть теорема (14.139), которую требовалось доказать, с той лишь разницей, что выражение (14.139) не содержит множителя $N^{1/2}$.

Покажем теперь, что любая функция ψ_* может быть разложена в ряд по функциям φ_β . С этой целью докажем, что ψ_* равно

$$N^{-1/2} \sum_{\beta} \exp(-2\pi i \mathbf{B}_{\beta} \cdot \mathbf{R}_*) \varphi_{\beta} = \frac{1}{N} \sum_{\beta} \sum_{\gamma} \exp[2\pi i \mathbf{B}_{\beta} \cdot (\mathbf{R}_{\gamma} - \mathbf{R}_*)] \psi_{\gamma}. \quad (14.144)$$

Правая часть этого равенства получается в результате замены φ_{β} согласно (14.137). Показатель экспоненциального множителя может быть записан в форме

$$\begin{aligned} 2\pi i (\beta_1 \mathbf{g}_1 + \beta_2 \mathbf{g}_2 + \beta_3 \mathbf{g}_3) \cdot (\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3) = \\ = 2\pi i \left[\frac{\beta_1 \alpha_1}{N_1} + \frac{\beta_2 \alpha_2}{N_2} + \frac{\beta_3 \alpha_3}{N_3} \right], \end{aligned} \quad (14.145)$$

где α — целые числа (они могут быть и отрицательными), как это следует из определения \mathbf{R}_{γ} и \mathbf{R}_* . При данном значении γ сумма по β эквивалентна тройной сумме по β_1 , β_2 и β_3 и может быть записана в виде

$$\begin{aligned} \sum_{\beta_1} \sum_{\beta_2} \sum_{\beta_3} \exp \left[2\pi i \left(\frac{\beta_1 \alpha_1}{N_1} + \frac{\beta_2 \alpha_2}{N_2} + \frac{\beta_3 \alpha_3}{N_3} \right) \right] = \\ = \left\{ 1 + \exp \left[\left(\frac{2\pi i \alpha_1}{N_1} \right) \beta_1 \right] + \dots + \exp \left[\frac{2\pi i \alpha_1}{N_1} (N_1 - 1) \right] \right\} \times \\ \times \left\{ 1 + \dots + \exp \left[\frac{2\pi i \alpha_2}{N_2} (N_2 - 1) \right] \right\} \times \\ \times \left\{ 1 + \dots + \exp \left[\frac{2\pi i \alpha_3}{N_3} (N_3 - 1) \right] \right\}. \end{aligned} \quad (14.146)$$

Первый член этого выражения равен N_1 , при условии что $\alpha_1 = 0$, или же $\alpha_1 = \pm N_1$. Второе условие, однако, в действительности места не имеет. Если α_1 отлично от нуля, то первый член (14.146) обращается в нуль, как в этом легко убедиться, записав его в форме суммы членов геометрической прогрессии

$$\frac{1 - \exp[(2\pi i \alpha_1 / N_1) \cdot N_1]}{1 - \exp(2\pi i \alpha_1 / N_1)}. \quad (14.147)$$

Числитель равен здесь нулю, в то время как знаменатель отличен от нуля. Таким образом, тройная сумма по β_1 , β_2 , β_3 отлична от нуля только при условии $\alpha_1 = \alpha_2 = \alpha_3 = 0$, т. е. при условии $\mathbf{R}_* = \mathbf{R}_{\gamma}$. Если $\mathbf{R}_* = \mathbf{R}_{\gamma}$, то эта сумма становится равной $N_1 N_2 N_3 = N$, так что формула (14.144) приобретает вид

$$N^{-1/2} \sum_{\beta} \exp(-2\pi i \mathbf{B}_{\beta} \cdot \mathbf{R}_*) \varphi_{\beta} = \sum_{\gamma} \delta_{*\gamma} \psi_{\gamma} = \psi_*. \quad (14.148)$$

Тем самым доказано, что функция $\psi_{\mathbf{r}}$ может быть разложена в ряд по функциям $\varphi_{\mathbf{r}}$, как это утверждалось выше [см. (14.124)].

Далее, при наличии периодических граничных условий, характеризующихся периодами $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$, коэффициенты γ_i , определяющие функции Блоха, могут принимать только целые значения. Действительно, если бы, например, γ_1 было нецелочисленным, то мы бы имели

$$\varphi_{\mathbf{r}}(\mathbf{r} + \mathbf{A}_1) = \exp(2\pi i \gamma_1) \varphi_{\mathbf{r}}(\mathbf{r}); \quad (14.149)$$

при этом коэффициент при $\varphi_{\mathbf{r}}(\mathbf{r})$ был бы отличен от единицы и функция $\varphi_{\mathbf{r}}$ не удовлетворяла бы граничным условиям. Таким образом, все возможные векторы \mathbf{B} , т. е. векторы \mathbf{B} , удовлетворяющие граничным условиям, соответствуют целочисленным значениям коэффициентов γ_i .

Формула (14.148) показывает также, что матрицы $S_{\beta\alpha}$ и $T_{\gamma\beta}$, входящие соответственно в выражения (14.137) и (14.140), взаимны, так как

$$\sum_{\beta} T_{\gamma\beta} S_{\beta\alpha} = \delta_{\gamma\alpha}. \quad (14.150)$$

Кроме того,

$$T_{\gamma\beta} = S_{\beta\gamma}^*, \quad (14.151)$$

так что преобразование является *унитарным*. Это обстоятельство во многих случаях дает возможность упрощать получаемые результаты (см. задачу 3 в § 3 гл. 17).

Покажем теперь, как можно определить зону Бриллюэна посредством векторов \mathbf{G}_i , и исследуем вопрос об эквивалентности различных векторов \mathbf{B} . Поскольку речь идет о трансляционных свойствах этих векторов, покажем, что функции $\varphi_{\mathbf{B}}$ и $\varphi_{\mathbf{B}'}$ эквивалентны, если \mathbf{B} и \mathbf{B}' отличаются друг от друга на целое кратное \mathbf{G}_i , где \mathbf{G}_i — обратные векторы решетки, определяемые формулами (14.126). Этот результат очевиден, так как в выражении

$$\varphi_{\mathbf{B}}(\mathbf{r} + \mathbf{R}) = \exp(2\pi i \mathbf{B} \cdot \mathbf{R}) \varphi_{\mathbf{B}}(\mathbf{r}) \quad (14.152)$$

экспоненциальный множитель при добавлении к \mathbf{B} , например вектора \mathbf{G}_1 , остается неизменным, поскольку

$$\mathbf{G}_1 \cdot \mathbf{R} = \mathbf{G}_1 \cdot (\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3) = \alpha_1, \quad (14.153)$$

так что добавленный в показателе член является целым кратным $2\pi i$. Можно, таким образом, сказать, что два вектора \mathbf{B} , отличающиеся друг от друга на целое кратное \mathbf{G}_i , эквивалентны. Следовательно, можно ограничиться рассмотрением лишь тех векторов \mathbf{B} , которые лежат в элементарной ячейке, обладающей ребрами $\mathbf{G}_1, \mathbf{G}_2$ и \mathbf{G}_3 . Объем такой ячейки равен

$$\mathbf{G}_1 \cdot (\mathbf{G}_2 \times \mathbf{G}_3) = V_{\mathbf{a}} = \frac{1}{V_{\mathbf{a}}}. \quad (14.154)$$

Векторы \mathbf{B} , лежащие вне этой ячейки, эквивалентны векторам \mathbf{B} , лежащим внутри нее. Эта ячейка содержит все N векторов \mathbf{B} , определяемых формулой (14.135), так как при $\gamma_1 = N_1$ и т. д. векторы \mathbf{B} эквивалентны тем векторам \mathbf{B} , для которых $\gamma_1 = 0$, и т. д. Плотность возможных состояний в \mathbf{G} -пространстве равна, таким образом,

$$\frac{N}{V_{\mathbf{G}}} = NV_{\mathbf{a}} = V_{\mathbf{A}}, \quad (14.155)$$

что, разумеется, совпадает с величиной, обратной объему элементарной ячейки \mathbf{g}_i -решетки, в которой заключены все возможные векторы \mathbf{B} .

Поскольку длина вектора \mathbf{P} в \hbar раз больше длины вектора \mathbf{B} , плотность возможных состояний в \mathbf{P} -пространстве равна

$$\frac{V_{\mathbf{A}}}{\hbar^3}, \quad (14.156)$$

или же равна $2V_{\mathbf{A}}/\hbar^3$, если принять во внимание два возможных значения спина. Эта формула совпадает с формулой, полученной ранее в § 5 гл. 5 для простой кубической решетки.

Элементарная ячейка решетки в \mathbf{G} - или \mathbf{P} -пространстве может быть подвергнута такой трансляции, в результате которой ее центром станет точка $\mathbf{P} = 0$; она будет при этом попрежнему содержать N различных возможных векторов \mathbf{B} . Выбор элементарной ячейки является, однако, несколько произвольным, поскольку векторы \mathbf{a}_i и \mathbf{G}_i могут быть выбраны различными способами. Определение первой зоны Бриллюэна не является, тем не менее, произвольным; эта зона строится следующим образом. Рассмотрим в \mathbf{G} -пространстве решетку, образованную точками $n_1\mathbf{G}_1 + n_2\mathbf{G}_2 + n_3\mathbf{G}_3$ при положительных и отрицательных целых значениях чисел n . Эта система точек называется *обратной решеткой*. Каждому узлу такой решетки мы приведем в соответствие тот участок \mathbf{G} -пространства, который расположен к нему ближе, чем к какому-либо другому узлу. В таком случае область пространства, расположенная вокруг начала координат, есть первая зона Бриллюэна. Особые свойства этой зоны заключаются в том, что для векторов \mathbf{B} , концы которых лежат на ее поверхности, удовлетворяются условия отражения Брэгга. В случае почти свободных электронов на этой поверхности имеет место разрыв непрерывности энергии, т. е. ей отвечает запрещенная зона, и составляющая групповой скорости в направлении, перпендикулярном к поверхности, обращается в нуль.

Можно показать (см. задачу 15), что в первой зоне нет двух таких состояний, которые были бы эквивалентны друг другу. Далее, объем этой зоны равен объему элементарной ячейки обратной решетки. Следовательно, в нем должны быть заключены все возможные векторы \mathbf{B} , общее число которых равно N .

Вопрос о построении зон Бриллюэна рассматривается достаточно подробно в ряде монографий [5, 7, 10, 11].

ЗАДАЧИ

1. Установить связь между следующими фактами, имеющими одну и ту же причину: при большом значении a (например, в случае фиг. 123 и 124 при $a = a_0$) 1) разрешенные зоны узки, 2) эффективная масса электрона в нижней зоне велика и 3) скорость перемещения электронного волнового пакета мала.

2. Показать, что на краях и в центре зоны Бриллюэна решения волнового уравнения имеют вид (14.102). Для примера рассмотреть нижнее состояние в первой зоне, соответствующее условию $\gamma = 0$ или $g'(a/2) \equiv g' = 0$. Показать, что для ячейки, центр которой находится в точке $x = na$, одно из решений имеет вид

$$\psi_1(x) = Ag(x - na),$$

что отвечает условию $P = 0$, а второе решение, аналогично,

$$\psi_2(x) = u(x - na) + 2n \left[\frac{u(a/2)}{g'(a/2)} \right] g(x - na).$$

Что произойдет, если к функции ψ_2 добавить член, пропорциональный ψ_1 ? Найти соответствующие решения для краев следующей зоны и показать, что они являются промежуточными между решениями, отвечающими вещественным значениям P , и решениями, соответствующими комплексным или мнимым P .

3. Доказать, что в каждой зоне экстремальные значения энергии имеют место при $P = \pm(\hbar/2a)$ и $P = 0$, и что между этими пределами энергия изменяется монотонно. Для доказательства воспользуйтесь соотношением между скоростью

$$v = \frac{d\mathcal{E}(P)}{dP}$$

и током, рассмотренным в связи с формулой (14.8). Это соотношение может быть записано в виде

$$\frac{\hbar}{im} \left[\psi^* \frac{d}{dx} \psi - \psi \frac{d}{dx} \psi^* \right].$$

Пусть $v = 0$, что отвечает горизонтальному участку кривой, характеризующей зависимость \mathcal{E} от P ; показать, что в этом случае

$$\frac{d \ln \psi}{dx} = \frac{d \ln \psi^*}{dx}.$$

Исходя из этого результата, показать, что P либо равно нулю, либо есть целое кратное $\hbar/2a$. (У большинства изучающих квантовую механику обычно складывается мнение, что эта теорема не может быть

обобщена на случай трех измерений, и что энергия может принимать экстремальные значения внутри зоны Бриллюэна.)

4. Основываясь на результатах, полученных при решении задач 2 и 3, показать, что у края энергетической зоны среднее значение тока равно нулю. Воспользовавшись определением групповой скорости, показать, что отсюда следует, что у края зоны $d\mathcal{E}/dP = 0$.

5. Исследовать кривые для γ и μ для „почти свободного“ электрона. С этой целью перенести центр элементарной ячейки в середину узкой области, высота которой равна \mathcal{W} , а ширина $(a - 2b) = 2c$. Потенциальная энергия принимает при этом следующие значения:

$$\begin{aligned} \mathcal{U}(x) &= \mathcal{W}, & |x| < c, \\ \mathcal{U}(x) &= 0, & c < |x| < \frac{a}{2}. \end{aligned}$$

Для упрощения задачи предположите, что $c \rightarrow 0$ и $\mathcal{W} \rightarrow \infty$, так что $c\mathcal{W} = V$ остается постоянным.

Доказать, что при этом

$$\begin{aligned} u(x) &= \sin\left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{1/2} x, & |x| < a, \\ g(x) &= \cos\left[\left(\frac{2m\mathcal{E}}{\hbar^2}\right)^{1/2} |x| - \alpha\right], & |x| < a, \end{aligned}$$

где

$$\alpha = \sqrt{\frac{2m}{\mathcal{E}}} \frac{V}{\hbar}.$$

Исходя из этого результата, показать, что нижнему энергетическому состоянию ($P = 0$) отвечает энергия $\mathcal{E} = 2V/a$, равная среднему значению потенциальной энергии электрона в элементарной ячейке. Показать, что ширина запрещенной зоны равна $4V/a$. Следует убедиться в том, что кривые, характеризующие зависимость энергии от P , будут иметь в этом случае ту же форму, что и на фиг. 126. Отсюда следует, что эффективная масса — порядка $mV/a\mathcal{E}$, где \mathcal{E} — энергия свободного электрона, отвечающая условию $V = 0$. Вывести точную формулу, определяющую значение эффективной массы у краев зоны. (Постоянство ширины запрещенной зоны является следствием нашего допущения о том, что $c \rightarrow 0$. В реальных условиях ширина запрещенной зоны уменьшается с увеличением энергии.)

6. На фиг. 124 и 125 рассмотрены случаи, когда $\mathcal{E} < \mathcal{W}$. Как будут вести себя кривые для γ и μ при $\mathcal{E} > \mathcal{W}$? Для ответа полезно решить сначала задачи 9—13.

7. Доказать, что в одномерном случае для периодического силового поля, которое при этом не обязательно симметрично по отношению к центру элементарной ячейки, мы имеем $\mathcal{E}(P) = \mathcal{E}(-P)$.

Для этого следует записать функцию ψ_P в виде $\psi_P = \exp(ipx/\hbar) u_P(x)$, где $u_P(x+a) = u_P(x)$. Обобщить этот результат на случай трех измерений.

8. Показать, что в случае симметричного силового поля функция $u_P(x)$, упомянутая в предыдущей задаче, всегда может быть представлена в виде суммы вещественной четной функции и мнимой нечетной функции x .

Примечание. Следующие четыре задачи представляют собой отдельные стадии доказательства того, что кривые для γ и μ , исследованные в § 8, монотонно убывают с энергией от $+\infty$ до $-\infty$. Теорема о том, что эти кривые не могут пересекаться, доказана в § 7. Этим сведений достаточно, чтобы показать, что зонная структура энергетического спектра, иллюстрируемая фиг. 125, всегда имеет место при движении электрона в одномерном симметричном периодическом поле. Решение задачи 13 показывает, как этот же результат может быть получен в общем случае периодического потенциала.

9. Задан периодический симметричный потенциал $\mathcal{U}(x)$, причем $\mathcal{U}(x) = \mathcal{U}(-x)$ и $\mathcal{U}(x+a) = \mathcal{U}(x)$; $\mathcal{U}(x)$ конечно, так что $0 \leq \mathcal{U}(x) \leq \mathcal{W}$. Доказать, что при $\mathcal{E} < 0$ как γ , так и μ [см. (14.90)] положительны, и что функции $g(x)$ и $u(x)$ не имеют узловых точек (нулей) в пределах элементарной ячейки. Доказать, что если \mathcal{E} достаточно велико, порядка $\hbar^2 a^2 m$, то g и u должны иметь узловые точки в пределах ячейки, а если \mathcal{E} очень велико, то число таких узловых точек также велико.

10. Предположим, что волновая функция $\psi_1(x)$, удовлетворяющая уравнению Шредингера при значении энергии, равном \mathcal{E} , в ограниченном в пространстве силовом поле (подобном полю, рассмотренному в предыдущей задаче) имеет две последовательные узловые точки: при $x = a_1$ и при $x = b_1 > a_1$. Исследовать ход волновой функции ψ_2 , которая также соответствует энергии \mathcal{E} и обладает узловой точкой при $x = a_2$, где a_2 несколько меньше a_1 . Доказать, что ψ_2 имеет узловую точку при $x = b_2$, где $b_2 < b_1$. С этой целью исследовать функцию $\eta_1(x) = [d\psi_1(x)/dx]/\psi_1(x)$. Доказать, что функция η_1 изменяется в пределах от $+\infty$ при $x = a_1 + 0$ до $-\infty$ при $x = b_1 - 0$ и остается непрерывной во всей области, за исключением ее границ, где она изменяется как $\eta_1 = 1/(x - a_1)$ и $1/(x - b_1)$. Доказать, что функция $\eta_2(x)$, соответствующая волновой функции ψ_2 , ни при каких условиях не совпадает с функцией $\eta_1(x)$.

Поскольку $\eta_2(a_1 + 0) < \eta_1(a_1 + 0)$, то при $x < b_1$, очевидно, η_2 должно стремиться к $-\infty$.

11. Исследовать функцию ψ_1 задачи 10, соответствующую энергии \mathcal{E}_1 . Пусть функция ψ_2 также имеет узловую точку при $x = a_1$ и является решением волнового уравнения при значении энергии, равном \mathcal{E}_2 , где \mathcal{E}_2 несколько больше \mathcal{E}_1 . Доказать, что функция ψ_2 имеет узловую точку при $b_2 < b_1$.

У к а з а н и е. Построить график, изображающий ход функции $\eta_1(x)$ в пределах от $+\infty$ при $x = a_1$ до $-\infty$ при $x = b_1$. Доказать, что при $x \rightarrow a_1$ мы имеем $\eta_2 = \eta_1 = 1/(x - a_1)$, но что при конечном значении разности $(x - a_1)$ значение $\eta_2(x)$ всегда меньше $\eta_1(x)$. Показать, что если $b_2 \geq b_1$, то при конечном значении разности $(b_1 - x)$ имеем $\eta_2(x) > \eta_1(x)$. Следовательно, b_2 должно быть меньше b_1 . Другими словами, при увеличении энергии расстояние между узловыми точками уменьшается. Для решения этой задачи может оказаться полезным вывести уравнение

$$\frac{d\eta_1}{dx} = \frac{2m}{\hbar^2} [\mathcal{U}(x) - \mathcal{E}_1] - \eta_1^3.$$

12. Основываясь на результатах решения задач 9, 10 и 11, показать, что кривые γ и μ , подобно кривым, изображенным на фиг. 124 и 125, должны спадать монотонно с увеличением энергии. [Для примера исследуйте функцию $u(x)$, имеющую узловую точку при $x = 0$, и предположите, что энергия столь мала, что $u(x)$ не имеет узлов в пределах элементарной ячейки. В таком случае с помощью метода, рассмотренного в задаче 9, можно снова показать, что увеличение \mathcal{E} приводит к уменьшению μ у краев ячейки. Если функция $u(x)$ имеет несколько узловых точек, то представляется интересным исследовать вопрос о влиянии изменения \mathcal{E} на расстояние между узлами, а также на ход кривой η в интервале между узловыми точками.]

13. Доказать, что в общем случае периодического силового поля $\mathcal{U}(x+a) = \mathcal{U}(x)$ (которое может быть также и несимметричным) решения имеют вид

$$\begin{aligned}\psi_1(x+a) &= \lambda\psi_1(x), \\ \psi_2(x+a) &= \lambda^{-1}\psi_2(x).\end{aligned}$$

Обозначить через $s(x)$ и $v(x)$ решения, удовлетворяющие уравнению Шредингера и принимающие значения

$$\begin{aligned}s &= 0 \quad \text{и} \quad \frac{ds}{dx} = 1 \quad \text{при} \quad x = 0, \\ v &= 1 \quad \text{и} \quad \frac{dv}{dx} = 0 \quad \text{при} \quad x = 0.\end{aligned}$$

Решения и производные от них при $x = a$ соответственно обозначьте через s_a, s'_a, v_a и v'_a . Доказать, что для любых двух решений уравнения Шредингера φ_1 и φ_2 имеем

$$\frac{d}{dx} \left[\varphi_2 \frac{d\varphi_1}{dx} - \varphi_1 \frac{d\varphi_2}{dx} \right] = 0$$

и, следовательно,

$$s'_a v_a - s_a v'_a = 1.$$

Показать, что условие периодичности функции $\psi_1 = As(x) + Bv(x)$ приводит к уравнению

$$(\lambda - v_a)(\lambda - s'_a) - s_a v'_a = \lambda^2 - \lambda(v_a + s'_a) + 1 = 0.$$

Отсюда следует, что если λ_1 и λ_2 — корни этого уравнения, то $\lambda_1 = 1/\lambda_2$. Убедитесь в том, что λ вещественно при $|v_a + s'_a| > 2$ или же $|\lambda|^2 = 1$ при $|v_a + s'_a| < 2$. Это указывает на то, что затухающих комплексных волновых функций не существует. Воспользовавшись результатами решения задач 9, 10 и 11, следует построить кривые, характеризующие зависимость функций $s(x)$ и $v(x)$ от x ; показать, что узловые точки этих кривых чередуются и, следовательно, что условие $|v_a + s'_a| < 2$ в некоторой определенной области значений энергии действительно удовлетворяется при любом значении постоянной решетки a .

14. Доказать, что если $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ — три возможные трансляции решетки и $\mathbf{R}_1 \cdot \mathbf{R}_2 \times \mathbf{R}_3 = V_a$, то эти векторы могут быть использованы в качестве элементарных векторов [см. формулу (14.118)].

15. Доказать, что грани первой зоны Бриллюэна (см. конец § 7) представляют собой плоскости, перпендикулярные биссектрисам углов, образованных векторами, проведенными из начала координат к остальным узлам обратной решетки. Доказать, что внутри зоны нет двух таких состояний, которые были бы эквивалентны друг другу.

16. Рассмотреть пространственную решетку, характеризующуюся элементарными векторами $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ и обратными векторами $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$. Рассмотреть семейство плоскостей, определяемое уравнениями

$$\mathbf{r} \cdot \mathbf{H} = n, \quad \mathbf{H} = h_1 \mathbf{G}_1 + h_2 \mathbf{G}_2 + h_3 \mathbf{G}_3,$$

где \mathbf{r} — некоторый вектор, проведенный из начала координат, а h_i и n — целые числа. Показать, что при соответствующем выборе числа n любой вектор

$$\mathbf{R} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3$$

удовлетворяет этим уравнениям, так что все узлы решетки лежат в плоскостях, характеризующихся различными значениями n . Показать

также, что если h_i — сравнительно простые числа (не имеющие общих множителей), то n может принимать любые целые значения. Показать, что в этом случае кристаллографические плоскости отстоят друг от друга на $1/|N|$. Решение этой задачи показывает, что векторы обратной решетки могут быть использованы для систематизации различных групп плоскостей, в которых расположены атомы кристалла.

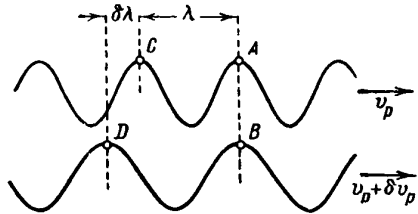
ЛИТЕРАТУРА

1. Dirac P. A. M., The Principles of Quantum Mechanics, Oxford, 1939 (см. перевод: Дирак П., Основы квантовой механики, М.—Л., 1937).
2. Neumann J., Mathematische Grundlagen der Quantenmechanik, Leipzig, 1932.
3. Kemble E. C., Fundamental Principles of Quantum Mechanics, New York, 1937.
4. Kramers H. A., Physica, 2, 483 (1935).
5. Seitz F., The Modern Theory of Solids, New York, 1940 (см. перевод: Зейтц Ф., Современная теория твердого тела, М.—Л., 1949).
6. Rojansky V., Introductory Quantum Mechanics, New York, 1938.
7. Wilson A. H., The Theory of Metals, Cambridge, 1936 (см. перевод: Вильсон А., Квантовая теория металлов, М.—Л., 1941).
8. Shockley W., Phys. Rev., 56, 317 (1939).
9. Whittaker, Watson, Modern Analysis, Cambridge, 1927 (см. перевод: Уиттекер и Ватсон, Курс современного анализа, М.—Л. 1934).
10. Mott N. F., Jones H., The Theory of the Properties of Metals and Alloys, Oxford, 1936.
11. Brillouin L., Wave Propagation in Periodic Structures, New York, 1946.
- 12*. Блохинцев Д. И., Основы квантовой механики, М.—Л., 1951.
- 13*. Френкель Я. И., Волновая механика, т. 1 и 2, М.—Л., 1934—1935.
- 14*. Ландау Л. Д. и Лифшиц Е. М., Квантовая механика, М.—Л., 1948.
- 15*. Фок В. А., Начала квантовой механики, Л., 1932.
- 16*. Ансельм А. И., ЖТФ, 21, 489 (1951).
- 17*. Регель А. Р., ЖЭТФ, 21, 845 (1951).
- 18*. Пекар С. И., Исследования по электронной теории кристаллов, М.—Л., 1951.

**СКОРОСТИ И УСКОРЕНИЯ ЭЛЕКТРОНОВ И ДЫРОК;
ТОК (ТЕОРИЯ)**

§ 1. ГРУППОВАЯ СКОРОСТЬ

На фиг. 128 изображены две синусоидальные волны. Гребень *A* верхней волны и гребень *B* нижней волны находятся в одной и той же фазе. В остальных точках фазы волн не совпадают и волны стремятся погасить друг друга. В том случае когда большое число волн, обладающих близкими значениями длины волны, образует волновой пакет, эти волны в значительной мере гасят друг друга повсюду, за исключением некоторой области пространства, размеры которой могут быть определены математически, если за центр этой области принять точки *A* и *B*. Предположим, что отдельные волны распространяются с фазовой скоростью v_p . Если фазовые скорости волн сколько-нибудь отличны друг от друга, то точка наблюдения будет смещаться так, что группа волн будет двигаться по отношению к отдельным волнам. Например, если скорость нижней волны превышает скорость верхней волны на δv_p , то гребень *D* поравняется с гребнем *C* через время $t = \delta\lambda / \delta v_p$. Если разность длин волн мала, то это выражение есть не что иное, как производная $d\lambda / dv_p$. За это время гребень *A* пройдет расстояние $v_p t$, однако точка совпадения фаз, или центр волнового пакета, переместится при этом в обратном направлении на расстояние λ . Скорость группы волн, таким образом, равна



Фиг. 128. Сложение двух волн при образовании волнового пакета.

Фиг. 128. Сложение двух волн при образовании волнового пакета.

$$v_g = \frac{(v_p t - \lambda)}{t} = v_p - \lambda \frac{dv_p}{d\lambda}. \quad (15.1)$$

Это показывает, что если фазовая скорость не зависит от длины волны, то групповая и фазовая скорости одинаковы. Это справедливо как для электромагнитных волн в пустоте, так и для акустических волн большой длины в воздухе.

Выражение для v_g может быть преобразовано следующим образом. Запишем

$$v_p = \lambda \nu, \quad \frac{dv_p}{d\lambda} = \nu + \lambda \frac{d\nu}{d\lambda}. \quad (15.2)$$

Это дает

$$v_g = v_p - \lambda v - \lambda^2 \frac{dv}{d\lambda} = -\lambda^2 \frac{dv}{d\lambda}. \quad (15.3)$$

Воспользовавшись соотношением $k = 1/\lambda$, получаем

$$dk = -\frac{d\lambda}{\lambda^2}, \quad v_g = \frac{dv}{dk}. \quad (15.4)$$

Последняя формула использована в гл. 6 при $v = \mathcal{E}/\hbar$ и $k = P/\hbar$.

В трехмерном случае мы будем рассматривать волны вида $\exp 2\pi i [\mathbf{k} \cdot \mathbf{r} - v(\mathbf{k}) t]$, из которых может быть построен пакет

$$\psi(\mathbf{r}, t) = \int \{\exp 2\pi i [\mathbf{k} \cdot \mathbf{r} - v(\mathbf{k}) t]\} \omega(\mathbf{k}) d\mathbf{k}. \quad (15.5)$$

Здесь $\int d\mathbf{k}$ означает интегрирование по \mathbf{k} -пространству, а $\omega(\mathbf{k})$ предполагается вещественной, положительной, плавно меняющейся функцией \mathbf{k} , значения которой отличны от нуля только внутри некоторого определенного $\mathbf{k} = \mathbf{k}_0$. Наибольшее значение ψ при любом t отвечает, следовательно, такому значению \mathbf{r} , при котором экспоненциальный множитель почти не зависит от \mathbf{k} , так что при интегрировании по \mathbf{k} фактически должны суммироваться члены, обладающие одинаковыми комплексными фазами. Это требование выполняется наилучшим образом при условии

$$\nabla_{\mathbf{k}} [\mathbf{k} \cdot \mathbf{r} - v(\mathbf{k}) t] = \mathbf{r} - \nabla_{\mathbf{k}} v(\mathbf{k}) t = 0 \quad (15.6)$$

для $\mathbf{k} = \mathbf{k}_0$. Это приводит к соотношению

$$\mathbf{r} = \nabla_{\mathbf{k}} v(\mathbf{k}) t, \quad (15.7)$$

которое эквивалентно определению компонент групповой скорости, приведенному ранее в гл. 6.

§ 2. СЛУЧАЙ ЭЛЕКТРОННЫХ ВОЛН

При применении полученных выше результатов к электронным волнам следует учесть два дополнительных обстоятельства. Во-первых, следует показать, что эти результаты не зависят от выбора нулевой точки, т. е. выбора начала отсчета энергии. Например, если бы к полупроводнику в целом было приложено некоторое поле, то вся система энергетических уровней электронов должна была бы сместиться. Все особенности поведения электронов при этом должны были бы, однако, остаться неизменными. Во-вторых, следует принять во внимание, что электронные волны не являются плоскими волнами, а описываются модулированными атомными волновыми функциями¹⁾.

¹⁾ Автор имеет здесь в виду случай электронов, движущихся в кристалле. Для свободных электронов в одномерном случае мы имели бы дело с плоскими волнами. — *Прим. ред.*

Несущественность выбора начала отсчета энергии легко может быть доказана, например, путем анализа соотношения Планка $\nu = \mathcal{E}/h$. Изменение начала отсчета энергии эквивалентно добавлению некоторой постоянной частоты ν_0 ко всем значениям ν . Поскольку ν_g содержит только производные от ν , добавление постоянного члена ν_0 не окажет при этом никакого влияния на величину ν_g .

Интересующие нас электронные волны описываются функциями Блоха

$$\psi_{\mathbf{P}} = u_{\mathbf{P}}(\mathbf{r}) \exp \left\{ \frac{2\pi i}{h} [\mathbf{P} \cdot \mathbf{r} - \mathcal{E}(\mathbf{P}) t] \right\}. \quad (15.8)$$

Величина $(1/h)\mathbf{P}$ играет здесь роль волнового вектора \mathbf{k} . Если построить волновой пакет из волн с почти одинаковыми значениями \mathbf{P} , то функция $u_{\mathbf{P}}(\mathbf{r})$ будет иметь примерно один и тот же вид для всех волн и поэтому при интегрировании по \mathbf{P} -пространству может быть вынесена за знак интеграла. Это дает

$$\psi = u_{\mathbf{P}_0}(\mathbf{r}) \int \exp \left\{ 2 \frac{\pi i}{h} [\mathbf{P} \cdot \mathbf{r} - \mathcal{E}(\mathbf{P}) t] \right\} w(\mathbf{P}) d\mathbf{P}. \quad (15.9)$$

Из тех же соображений следует, что если $w(\mathbf{P})$ велико только вблизи точки \mathbf{P}_0 , то ψ велико только вблизи

$$\mathbf{r} = [\nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P})]_{\mathbf{P}=\mathbf{P}_0} t. \quad (15.10)$$

Отсюда групповая скорость равна

$$\dot{\mathbf{r}} = \frac{d}{dt} \mathbf{r} = [\nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P})]_{\mathbf{P}=\mathbf{P}_0}. \quad (15.11)$$

§ 3. АНАЛОГИЯ ИЗ ТЕОРИИ ЭЛЕКТРИЧЕСКИХ ЦЕПЕЙ

Для того чтобы установить соотношение между групповой скоростью, плотностью и потоком, исследуем искусственную линию, состоящую из сосредоточенных постоянных индуктивностей и емкостей. Нужно доказать следующее соотношение:

Передаваемая мощность = энергии, приходящейся на единицу длины линии, умноженной на групповую скорость.

Как было показано при рассмотрении фиг. 63, этот результат является частным случаем более общего результата, согласно которому скорости потока, описываемого волновыми уравнениями, могут быть вычислены либо по методу потока плотности энергии, либо с помощью понятия групповой скорости.

На фиг. 129 изображена исследуемая линия и указаны принятые обозначения. Попытаемся получить решение в форме

$$V_n = V_0 e^{2\pi i (na/\lambda - \nu t)} = V_0 e^{2\pi i (nka - \nu t)}, \quad (15.12)$$

где $k = 1/\lambda$, а na — положение n -й емкости вдоль линии. [Множители вида $\exp(-i\omega t)$ введены в эти выражения с тем, чтобы последние

соответствовали квантовомеханическим соотношениям; это обуславливает появление знака минус перед $i\omega L$ и $i\omega C$.] Соответственно, получаем следующие формулы для определения токов:

$$I_n = \frac{(V_n - V_{n+1})}{(-i\omega L)} = \frac{V_n(1 - e^{2\pi ika})}{(-i\omega L)}, \quad (15.13a)$$

$$I_{n-1} = \frac{(V_{n-1} - V_n)}{(-i\omega L)} = -\frac{V_n(1 - e^{-2\pi ika})}{(-i\omega L)}. \quad (15.13б)$$

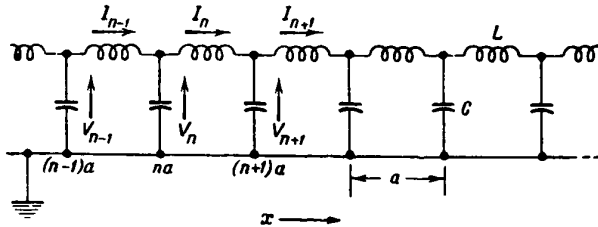
Полагая разность этих токов равной току, идущему на зарядку емкости, имеем

$$\begin{aligned} -i\omega C V_n &= I_{n-1} - I_n = \frac{V_n}{(-i\omega L)} (e^{2\pi ika} - 2 + e^{-2\pi ika}) = \\ &= \frac{V_n}{(-i\omega L)} (e^{\pi ika} - e^{-\pi ika})^2 = \frac{4V_n}{(i\omega L)} \sin^2 \pi ka. \end{aligned} \quad (15.14)$$

Это уравнение может быть решено относительно $\sin \pi ka$, что дает

$$\sin \pi ka = \omega \sqrt{\frac{LC}{4}} = \pi v \sqrt{LC}. \quad (15.15)$$

Отсюда мы заключаем, что $\pi v \sqrt{LC}$ не должно превышать единицы, т. е. что v не может быть больше $1/\pi \sqrt{LC}$, так как более высокие



Фиг. 129. Искусственная линия, используемая для иллюстрации соотношения между передаваемой мощностью, плотностью энергии и групповой скоростью.

значения v отвечали бы комплексным значениям k . В предельном случае $\sin \pi ka = 1$ и $\pi ka = \pi/2$, или 90° .

В полосе пропускания линии групповая скорость равна

$$v_g = \frac{dv}{dk} = \frac{d}{dk} \frac{\sin \pi ka}{\pi \sqrt{LC}} = \frac{a \cos \pi ka}{\sqrt{LC}}. \quad (15.16)$$

Отсюда следует, что групповая скорость изменяется от a/\sqrt{LC} при низких частотах (т. е. при $ka \ll 1$) до нуля при предельном значении частоты, отвечающем условию $\pi ka = 90^\circ$ или $\cos \pi ka = 0$.

Вычислим теперь поток мощности и плотность энергии вдоль линии. Напряжение на n -м конденсаторе равно

$$\text{Re } V_n = V_0 \cos(2\pi nka - \omega t), \quad (15.17)$$

чему соответствует энергия

$$\frac{CV_0^2}{2} \cos^2(2\pi nka - \omega t). \quad (15.18)$$

Пиковое и среднее значения этой энергии равны, соответственно, $CV_0^2/2$ и $CV_0^2/4$. Энергия, отвечающая току I_n , аналогичным образом определяется квадратом амплитуды I_n .

Комплексное выражение для I_n может быть записано в форме

$$\begin{aligned} I_n &= V_n \frac{1 - e^{2\pi ika}}{-i\omega L} = V_n e^{\pi ika} (-2i) \frac{\sin \pi ka}{-i\omega L} = \\ &= \frac{2V_n}{\omega L} e^{\pi ika} \sin \pi ka = \frac{2V_0}{\omega L} \sin \pi ka \cdot e^{2\pi i(nka - \omega t) + \pi ika}. \end{aligned} \quad (15.19)$$

Очевидно, что пиковое значение I_n равно

$$\frac{2V_0}{\omega L} \sin \pi ka \quad (15.20)$$

и смещено по фазе на πka по отношению к V_n . Максимальное значение энергии, запасенной в индуктивности, равно, таким образом,

$$\frac{1}{2} L \frac{4V_0^2}{\omega^2 L^2} \sin^2 \pi ka = \frac{1}{2} L \frac{4V_0^2}{\omega^2 L^2} \frac{\omega^2 LC}{4} = \frac{1}{2} CV_0^2, \quad (15.21)$$

т. е. совпадает с максимальной энергией, накопленной конденсатором. Энергия линии в среднем равна при этом $1/4 CV_0^2$ для каждого конденсатора и каждой индуктивности, что составляет $CV_0^2/2a$ на единицу длины линии.

Если эта энергия передается с групповой скоростью, то передаваемая мощность равна

$$\frac{CV_0^2}{2a} a \frac{\cos \pi ka}{\sqrt{LC}} = \frac{V_0^2 (C/L)^{1/2} \cos \pi ka}{2}. \quad (15.22)$$

Поток мощности через индуктивность при токе I_n может быть вычислен также и непосредственно — по значениям амплитуд и разности фаз между I_n и V_n . Это приводит к среднему значению мощности, равному

$$\begin{aligned} \frac{1}{2} V_{\text{пик}} \cdot I_{\text{пик}} \cos(\text{разность фаз}) &= \frac{1}{2} V_0 \frac{2V_0}{\omega L} \sin \pi ka \cdot \cos \pi ka = \\ &= \frac{V_0^2}{\omega L} \cdot \frac{\omega \sqrt{LC}}{2} \cos \pi ka = \frac{V_0^2 (C/L)^{1/2} \cos \pi ka}{2} \end{aligned} \quad (15.23)$$

в согласии с формулой (15.22).

Исследование этого частного примера позволяет сделать заключение, что вычисления потока мощности: 1) непосредственно с

помощью соответствующих соотношений, определяющих мощность в некоторой точке линии, и 2) в предположении, что накопленная энергия передается вдоль линии с групповой скоростью, приводят к одинаковым результатам.

§ 4. ТОК ВЕРОЯТНОСТИ

В этом параграфе мы изложим теорию оператора тока и вычислим ток, создаваемый электронами, находящимися в зоне Бриллюэна.

Операторные соотношения $p_{\text{op}} = (\hbar/2\pi i)\partial/\partial x$ и $\mathcal{H} = (i\hbar/2\pi)\partial/\partial t$ будут при этом исходными. Рассмотрим некоторый объем V , ограниченный поверхностью S , внутри которого волновая функция электрона $\psi(x, y, z, t) \equiv \psi(r, t)$ удовлетворяет уравнению

$$\mathcal{H}\psi = \frac{i\hbar}{2\pi} \dot{\psi}. \quad (15.24)$$

В этом случае полный ток вероятности из объема V определяется скоростью уменьшения вероятности нахождения электрона в рассматриваемом объеме

$$\begin{aligned} -\frac{d}{dt} \int_V \psi^* \psi \, dx \, dy \, dz &= -\frac{2\pi}{i\hbar} \int_V (\psi^* \mathcal{H}\psi - \psi \mathcal{H}\psi^*) \, dx \, dy \, dz = \\ &= \frac{2\pi i}{\hbar} \left(\frac{-\hbar^2}{8\pi^2 m} \right) \int_V (\psi^* \nabla^2 \psi - \psi \nabla^2 \psi^*) \, dx \, dy \, dz. \end{aligned} \quad (15.25)$$

Здесь

$$\mathcal{H} = -\frac{\hbar^2}{8\pi^2 m} \nabla^2 + \mathcal{U}(x, y, z) \quad (15.26)$$

и члены, содержащие функцию \mathcal{U} , взаимно сократились. Воспользовавшись тем, что

$$\nabla \cdot (A\nabla B - B\nabla A) = A\nabla^2 B - B\nabla^2 A,$$

и направляя $d\mathbf{S}$ вдоль внешней нормали к поверхности S , можно преобразовать этот интеграл к виду:

$$\begin{aligned} -\frac{i\hbar}{4\pi m} \int_S \nabla \cdot (\psi^* \nabla \psi - \psi \nabla \psi^*) \, dx \, dy \, dz &= \\ &= \frac{1}{2m} \frac{\hbar}{2\pi i} \int_S (\psi^* \nabla \psi - \psi \nabla \psi^*) \cdot d\mathbf{S}. \end{aligned} \quad (15.27)$$

Таким образом, вероятность уменьшается так, как если бы через поверхность S протекал ток вероятности, равный

$$\mathbf{I}_{\text{вер.}} = \frac{1}{2m} [\psi^* \mathbf{p}\psi + \psi(\mathbf{p}\psi)^*] = \frac{1}{m} \text{Re}(\psi^* \mathbf{p}\psi). \quad (15.28)$$

(Re означает, что от выражения, стоящего справа, берется вещественная часть.) Для плоской волны с импульсом p_0 это дает $(1/m)p_0|\psi|^2$, как если бы плотность вероятности $|\psi|^2$ перемещалась со скоростью p_0/m .

Чтобы уяснить смысл этого выражения, определяющего ток вероятности, полезно проанализировать следующие результаты. Если $\psi(\mathbf{r}, t)$ есть собственная функция энергии, то она имеет вид $\psi(\mathbf{r}) \exp(2\pi iEt/\hbar)$, так что $|\psi|^2$ не зависит от времени. Это находится в согласии с тем обстоятельством, что в этом случае $\nabla \cdot \mathbf{I}_{\text{вер.}}$ содержит в качестве множителя выражение $(\psi^* \nabla^2 \psi - \psi \nabla^2 \psi^*)$, которое обращается в нуль, если ψ — собственная функция энергии¹⁾. В одномерном случае для волновой функции, ход которой изображен на фиг. 56, в, наличие узловой точки показывает, что в этой точке ток равен нулю; следовательно, поскольку $\nabla \cdot \mathbf{I} = 0$, во всех точках пространства \mathbf{I} обращается в нуль.

Найдем теперь выражение, определяющее ток, протекающий через одну из граней элементарной ячейки, и покажем, что оно сводится к выражению для среднего значения групповой скорости (см. [1]). Поскольку $\mathbf{I}_{\text{вер.}}$ содержит произведения функций ψ и ψ^* , очевидно, что волновая функция Блоха дает одно и то же значение тока для любой из элементарных ячеек, так как члены, содержащие производные от экспоненциальных множителей, в этом случае взаимно сокращаются. Предположим, что элементарная ячейка образована векторами \mathbf{a} , \mathbf{b} , \mathbf{c} . Направим ось x перпендикулярно к грани (\mathbf{ab}) и вычислим ток через эту грань. Это даст нам, очевидно, x -компоненту тока. В этом случае функция Блоха является функцией \mathbf{P} и \mathbf{r} , и \mathcal{E} зависит от \mathbf{P} . Дифференцирование уравнения Шредингера по P_x дает

$$\frac{\partial}{\partial P_x} \left[\frac{p^2}{2m} + \mathcal{U}(\mathbf{r}) - \mathcal{E} \right] \psi = 0, \quad (15.29)$$

или

$$\frac{\partial \mathcal{E}}{\partial P_x} \psi = \left[\frac{p^2}{2m} + \mathcal{U}(\mathbf{r}) - \mathcal{E} \right] \frac{\partial \psi}{\partial P_x}. \quad (15.30)$$

Преобразуем правую часть этого уравнения так, чтобы она содержала ток вероятности. Умножение на ψ^* и интегрирование по элементарной ячейке C дает

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial P_x} \int_C \psi^* \psi dV &= \int_C \psi^* \left[\frac{p^2}{2m} + \mathcal{U}(\mathbf{r}) - \mathcal{E} \right] \frac{\partial \psi}{\partial P_x} dV = \\ &= \int_C \frac{\partial \psi}{\partial P_x} \left[\frac{p^2}{2m} + \mathcal{U}(\mathbf{r}) - \mathcal{E} \right] \psi^* dV + \int_C \left(\psi^* \frac{p^2}{2m} \frac{\partial \psi}{\partial P_x} - \frac{\partial \psi}{\partial P_x} \frac{p^2}{2m} \psi^* \right) dV. \end{aligned}$$

¹⁾ См. § 4 гл. 14. — *Прим. авт.*

Учитывая, что если $\mathcal{H}\psi = \mathcal{E}\psi$, то $\mathcal{H}\psi^* = \mathcal{E}\psi^*$, получаем

$$\left. \begin{aligned} \frac{\partial \mathcal{E}}{\partial P_x} \int_C \psi^* \psi dV &= 0 - \frac{\hbar^2}{8\pi^2 m} \int_C \left(\psi^* \nabla^2 \frac{\partial \psi}{\partial P_x} - \frac{\partial \psi}{\partial P_x} \nabla^2 \psi^* \right) dV = \\ &= - \frac{\hbar^2}{8\pi^2 m} \int_C \nabla \cdot \left(\psi^* \nabla \frac{\partial \psi}{\partial P_x} - \frac{\partial \psi}{\partial P_x} \nabla \psi^* \right) dV = \\ &= - \frac{\hbar^2}{8\pi^2 m} \int_S \left(\psi^* \nabla \frac{\partial \psi}{\partial P_x} - \frac{\partial \psi}{\partial P_x} \nabla \psi^* \right) \cdot dS. \end{aligned} \right\} (15.31)$$

Последний интеграл взят по шести плоским граням элементарной ячейки. Как мы увидим, члены, соответствующие тем граням, которые не перпендикулярны оси x , при этом взаимно сокращаются, так как подинтегральная функция периодична; даже при интегрировании по двум граням (\mathbf{ab}) многие члены также взаимно сокращаются. Чтобы вычислить интересующий нас интеграл, запишем

$$\psi = \exp \left[2\pi i \frac{\mathbf{P} \cdot \mathbf{r}}{\hbar} \right] u_{\mathbf{P}}(\mathbf{r}), \quad (15.32)$$

$$\frac{\partial \psi}{\partial P_x} = \frac{2\pi i x}{\hbar} \psi + \exp \left[2\pi i \frac{\mathbf{P} \cdot \mathbf{r}}{\hbar} \right] \frac{\partial u_{\mathbf{P}}(x)}{\partial P_x}, \quad (15.33)$$

$$\frac{\partial}{\partial x} \frac{\partial \psi}{\partial P_x} = \frac{2\pi i x}{\hbar} \frac{\partial \psi}{\partial x} + \frac{2\pi i}{\hbar} \psi + \frac{\partial}{\partial x} \exp \left[2\pi i \frac{\mathbf{P} \cdot \mathbf{r}}{\hbar} \right] \frac{\partial u_{\mathbf{P}}(x)}{\partial P_x}. \quad (15.34)$$

Вид функции $\partial \psi / \partial P_x$ свидетельствует о том, что подинтегральное выражение периодически по отношению к ребрам ячейки \mathbf{a} и \mathbf{b} (x -компонента которых равна нулю), так что при вычислении можно ограничиться рассмотрением грани (\mathbf{ab}) и брать при этом только составляющую вектора ∇ по оси x . По отношению к грани (\mathbf{ab}) периодичны все члены, за исключением тех, которые содержат множитель x . Поскольку при переходе от одной грани к другой x изменяется на величину $c_{\mathbf{a}}$, окончательный результат имеет вид

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial P_x} \int_C \psi^* \psi dV &= \frac{-\hbar^2}{8\pi^2 m} \int_{\mathbf{ab}} \left(\psi^* \frac{2\pi i c_{\mathbf{a}}}{\hbar} \frac{\partial \psi}{\partial x} - \frac{2\pi i c_{\mathbf{a}}}{\hbar} \psi \frac{\partial \psi^*}{\partial x} \right) dx dy = \\ &= \frac{c_{\mathbf{a}} \hbar}{4\pi i m} \int_{\mathbf{ab}} \left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x} \right) dx dy = \\ &= c_{\mathbf{a}} \int_{\mathbf{ab}} I_{x \text{ в.ер.}} dx dy = c_{\mathbf{a}} \mathbf{a} \cdot \mathbf{b} [\bar{I}_{x \text{ в.ер.}} \text{ по грани } (\mathbf{ab})]. \end{aligned} \quad (15.35)$$

Среднее значение тока вероятности через грань (\mathbf{ab}) равно, таким образом,

$$\frac{\partial \mathcal{E}}{\partial P_x} \int_C \psi^* \psi dV / c_{\mathbf{a}} (\mathbf{a} \cdot \mathbf{b}). \quad (15.36)$$

Этот результат можно интерпретировать так: средняя плотность вероятности, равная частному от деления величины $\int_C \psi^* \psi dV$ на объем ячейки, протекает через грань $x = \text{const}$ со средней скоростью $v_x = \partial \mathcal{E} / \partial P_x$. Аналогичные результаты, получаемые для остальных граней, приводят нас к заключению, что $\mathbf{v} = \nabla_{\mathbf{p}} \mathcal{E}(\mathbf{p})$.

§ 5. НЕКОТОРЫЕ ОБЩИЕ СВОЙСТВА КВАНТОМЕХАНИЧЕСКИХ ВОЛНОВЫХ ПАКЕТОВ

Как мы уже отмечали в гл. 7 и 8, исследование поведения дырок и электронов с точки зрения теории волновых пакетов существенно для понимания механизма действия транзисторов, а также объяснения ряда явлений в полупроводниках. Настоящий параграф написан в качестве введения в изучение некоторых свойств волновых пакетов, которые, хотя и хорошо известны, тем не менее не рассматриваются в обычных учебниках. Прежде всего исследуем такой волновой пакет, для которого произведение $\Delta p \cdot \Delta x$ имеет наименьшее значение. (Мы не будем приводить здесь выводов соответствующих формул, так как они даны в работах, на которые мы сошлемся в дальнейшем.) Затем изложим некоторые общие методы теории волновых пакетов и применим их к изучению волнового пакета, изображенного на фиг. 65.

Одним из наиболее известных соотношений квантовой механики является соотношение неопределенностей¹⁾. Проанализируем смысл этого соотношения с точки зрения представлений о волновых пакетах. Волновой пакет, подобный, например, изображенному на фиг. 65, обладает определенной протяженностью в пространстве. Ему отвечает, соответственно, определенное ожидаемое значение среднего квадратичного отклонения при измерениях его положения. Формулы для вычисления этой величины были приведены в § 2 гл. 14 (стр. 432). Предположим, что, применив их, мы вычислили $\overline{x^2} - (\overline{x})^2 = (\overline{x - \overline{x}})^2 = (\Delta x)^2$, так что величина Δx может быть названа неопределенностью положения электрона. Наряду с этим будет также иметь место некоторый разброс ожидаемых значений импульса вокруг его среднего значения \overline{p} . Если для простоты предположим, что $\overline{p} = 0$, то среднее значение $(p - \overline{p})^2$, которое обозначим $(\Delta p)^2$, будет равняться $\overline{p^2} = 2m\overline{E_{кин}}$. В § 3 гл. 14 мы показали [см. формулу (14.41)], что эта величина всегда положительна, за исключением случая $\psi = \text{const}$, который для волнового пакета не может иметь места. Непосредственные вычисления, основанные на использовании формулы (14.52), приводят к соотношению неопределенностей

$$\Delta p \Delta x \geq \frac{\hbar}{2}, \quad (15.37)$$

¹⁾ Критический анализ физического смысла этого соотношения дан в книге Д. И. Блохинцева [18]. — *Прим. ред.*

справедливого для любой волновой функции. Таким образом, если волновой пакет становится более компактным, так что Δx уменьшается, то Δp должно при этом увеличиваться. Математически это следует из того факта, что если волновая функция становится менее протяженной, например в результате простого сжатия вдоль оси x , то ее производная по x , а также величина $\overline{p^2}$ при этом должны возрастать.

Для электрона в отсутствие внешних сил можно построить такой волновой пакет, для которого в любой данный момент времени, например $t=0$, величина $\Delta p \Delta x$ будет иметь наименьшее значение. Найдено, однако, что с течением времени этот пакет будет постепенно расплываться. Для нормированного волнового пакета, центр которого находится в точке $x=0$ (причем $\overline{p}=0$, а $\Delta p \Delta x$ обладает минимальным значением), волновая функция имеет вид

$$\psi(x, t) = \frac{\exp\left[-\frac{x^2}{4a(a+it)}\right]}{(2\pi)^{1/4}(a+it)^{1/4}}, \quad (15.38)$$

где

$$a = \Delta x \quad \text{при} \quad t=0, \quad (15.39)$$

$$u = \frac{\hbar}{2m \Delta x} = \frac{\Delta p}{m}. \quad (15.40)$$

Величина u имеет здесь размерность скорости. Это уравнение может быть получено по аналогии с уравнением диффузии; при этом для свободной частицы уравнение Шредингера, содержащее время, преобразуется следующим образом:

$$\begin{aligned} i\hbar \frac{\partial \psi}{\partial t} &= \mathcal{H}\psi = \frac{1}{2m} \left(\frac{\hbar}{c}\right)^2 \frac{\partial^2 \psi}{\partial t^2}, \\ \frac{\partial \psi}{\partial t} &= -\frac{\hbar}{2mi} \frac{\partial^2 \psi}{\partial x^2} = D \frac{\partial^2 \psi}{\partial x^2}, \end{aligned} \quad (15.41)$$

где $D = \hbar/2mi$ обладает размерностью постоянной диффузии¹⁾. Более удобный метод решения этой задачи основан на разложении функции $\psi(x, 0)$ (для которой величина $\Delta p \Delta x$ обладает минимальным значением) по собственным функциям импульса, временная зависимость которых характеризуется множителем $\exp(itp^2/2m\hbar)$; таким образом получается искомая временная зависимость функции $\psi(x, t)$ ²⁾.

Докажем теперь некоторые общие теоремы, справедливые для любых волновых пакетов, соответствующих свободным электронам, после чего снимем ограничение $\overline{p}=0$, наложенное нами на пакет (15.38), с тем чтобы можно было исследовать его движение и получить пакет, изображенный на фиг. 65.

¹⁾ Этот метод рассмотрен Дашменом [2]. — *Прим. авт.*

²⁾ См., например, [3]. — *Прим. авт.*

Сначала докажем, что уравнение Шредингера обратимо во времени, т. е. что для любой волновой функции $\psi(x, t)$ может быть найдена другая функция $\psi_-(x, t)$, для которой $|\psi_-(x, t)|^2 = |\psi(x, -t)|^2$. Функция ψ_- дает, таким образом, то же значение плотности вероятности, что и функция ψ , с той, однако, разницей, что время t заменено при этом на $-t$. Эта теорема доказывается очень просто. Обозначим через $\chi(x, t)$ производную от функции $\psi(x, t)$ по времени. Очевидно, мы имеем

$$i\hbar \frac{\partial \psi}{\partial t} = i\hbar \chi(x, t) = \mathcal{H}\psi(x, t). \quad (15.42)$$

Пусть

$$\psi_-(x, t) = \psi^*(x, -t).$$

В таком случае

$$i\hbar \frac{\partial \psi_-(x, t)}{\partial t} = i\hbar \frac{\partial \psi^*(x, -t)}{\partial t} = -i\hbar \chi^*(x, -t). \quad (15.43)$$

Из уравнения, комплексно сопряженного уравнению (15.42), следует, что

$$-i\hbar \chi^*(x, -t) = \mathcal{H}\psi^*(x, -t) = \mathcal{H}\psi_-(x, t). \quad (15.44)$$

Мы использовали здесь соотношение

$$(\mathcal{H}\psi)^* = \mathcal{H}(\psi^*),$$

являющееся следствием того обстоятельства, что \mathcal{H} не содержит мнимых членов. Уравнения (15.44) и (15.43) дают

$$i\hbar \frac{\partial \psi_-}{\partial t} = \mathcal{H}\psi_-, \quad (15.45)$$

что и требовалось доказать. Покажем теперь, что при изменении t величина Δx возрастает от своего минимального значения симметрично по отношению к $+t$ и $-t$. Этот результат мог бы показаться неестественным, если бы уравнение Шредингера не обладало свойством обратимости во времени. Докажем, в частности, что для любого волнового пакета Δx меняется согласно уравнению

$$(\Delta x)^2 = \frac{\bar{p}^2 - (\bar{p})^2}{m^2} (t - t_0)^2 + (\Delta x)_{\min}^2, \quad (15.46)$$

где t_0 — тот момент времени, когда Δx имеет минимальное значение. Это уравнение показывает, что пакет расплывается в пространстве симметрично по отношению к прошедшему и будущему времени, отсчитываемому от момента t_0 . Доказательство формулы (15.46) основано на использовании выражения для $(\Delta x)^2$ [см. формулу (14.19)] и

формулы, определяющей производную по времени от некоторого оператора [см. (14.43)]. Мы имеем

$$\frac{dx}{dt} = \frac{p}{m}, \quad (15.47)$$

$$\frac{d}{dt} \left(\frac{d}{dt} x \right) = \frac{1}{m} \frac{d}{dt} p = \frac{1}{m} \frac{1}{i\hbar} (p\mathcal{H} - \mathcal{H}p) = 0, \quad (15.48)$$

так как

$$\mathcal{H} = \frac{p^2}{2m} \text{ и } \mathcal{H}p = p\mathcal{H} = \frac{p^3}{2m}.$$

Вычислим теперь dx^2/dt :

$$\frac{d}{dt} x^2 = \frac{1}{2i\hbar m} (x^2 p^2 - p^2 x^2) = \quad (15.49a)$$

$$= \frac{1}{2i\hbar m} \left\{ x^2 p^2 - \left[x^2 p^2 + 4 \left(\frac{\hbar}{i} \right) xp + 2 \left(\frac{\hbar}{i} \right)^2 \right] \right\} = \quad (15.49b)$$

$$= \frac{-2(\hbar/i)}{2i\hbar m} \left(2xp + \frac{\hbar}{i} \right) = \quad (15.49b)$$

$$= \frac{1}{m} (xp + px). \quad (15.49c)$$

Равенство (15.49a) следует непосредственно из формулы (14.43); (15.49b) получается в результате простых алгебраических преобразований; (15.49c) вытекает из уравнения (14.52), а равенство (15.49b) получается путем повторных применений уравнения (14.52):

$$\begin{aligned} p^n x &= p^{n-1} p x = p^{n-1} \left(xp + \frac{\hbar}{i} \right) = p^{n-1} xp + \frac{\hbar}{i} p^{n-1} = \\ &= p^{n-2} x p^2 + 2 \frac{\hbar}{i} p^{n-1} = \text{и т. д.} = xp^n + n \frac{\hbar}{i} p^{n-1}. \end{aligned} \quad (15.50)$$

Вычислив аналогичным образом d^2x^2/dt^2 и выразив с помощью (15.50) все члены, содержащие $p^n x$, через xp^n , имеем

$$\begin{aligned} \frac{d^2x^2}{dt^2} &= \frac{d}{dt} \left(\frac{px + xp}{m} \right) = \frac{1}{i\hbar 2m^2} [(px + xp)p^2 - p^2(px + xp)] = \\ &= \frac{1}{i\hbar 2m^2} \left(xp^3 + \frac{\hbar}{i} p^2 + xp^3 - xp^3 - 3 \frac{\hbar}{i} p^2 - 2 \frac{\hbar}{i} p^2 \right) = \\ &= \frac{1}{i\hbar 2m^2} \left(-\frac{4\hbar}{i} \right) p^2 = \frac{2p^2}{m^2}. \end{aligned} \quad (15.51)$$

Поскольку $dp^2/dt = 0$, очевидно, что $d^3x^2/dt^3 = 0$.

Усредняя предыдущие выражения, т. е. вычисляя интегралы типа $\bar{x} = \int \psi^* x \psi dx$ [см. формулу (14.25)], получаем

$$\frac{d}{dt} \bar{x} = \frac{\bar{p}}{m} = \bar{v} \quad (\text{постоянно во времени}), \quad (15.52)$$

$$\frac{d}{dt} \bar{x}^2 = \frac{1}{m} (\overline{px + xp}) = \overline{xv}, \quad (15.53)$$

$$\frac{d^2}{dt^2} \bar{x}^2 = \frac{2\bar{p}^2}{m^2} = 2\bar{v}^2 \quad (\text{постоянно во времени}). \quad (15.54)$$

Производные от \bar{x} и \bar{x}^2 по времени более высокого порядка обращаются в нуль так же, как и все производные по времени от \bar{p} и \bar{p}^2 . Воспользовавшись тем, что $(\Delta x)^2 = \bar{x}^2 - (\bar{x})^2$, можно записать

$$\frac{d}{dt} (\Delta x)^2 = \frac{d\bar{x}^2}{dt} - 2\bar{x} \frac{d\bar{x}}{dt} = \overline{xv} - 2\bar{x}\bar{v}, \quad (15.55)$$

$$\frac{d^2}{dt^2} (\Delta x)^2 = 2\bar{v}^2 - 2(\bar{v})^2 = 2 \frac{\bar{p}^2 - (\bar{p})^2}{m^2} = \frac{2(\Delta p)^2}{m^2}. \quad (15.56)$$

Эти уравнения показывают, что $(\Delta x)^2$ является квадратичной функцией t ; коэффициенты формулы, определяющей зависимость $(\Delta x)^2$ от t , могут быть выражены через значения различных величин в момент времени $t=0$ (например, \bar{x}_0^2):

$$(\Delta x)^2 = \frac{(\Delta p)^2}{m^2} t^2 + (\overline{xv} - 2\bar{x}\bar{v})_0 t + \bar{x}_0^2 - (\bar{x}_0)^2. \quad (15.57)$$

$(\Delta x)^2$ принимает наименьшее значение в момент времени t_m , для которого

$$\frac{d}{dt} (\Delta x)^2 = 2 \frac{(\Delta p)^2}{m^2} t_m + (\overline{xv} - 2\bar{x}\bar{v})_0 = 0. \quad (15.58)$$

Это уравнение можно решить относительно t_m и вычислить затем минимальное значение $(\Delta x)^2$ по формуле

$$(\Delta x)_{\min}^2 = \frac{(\Delta p)^2}{m^2} t_m^2 + (\overline{xv} - 2\bar{x}\bar{v})_0 t_m + \bar{x}_0^2 - (\bar{x}_0)^2. \quad (15.59)$$

В результате для $(\Delta x)^2$ получаем выражение

$$(\Delta x)^2 = \frac{(\Delta p)^2}{m^2} (t - t_m)^2 + (\Delta x)_{\min}^2. \quad (15.60)$$

Полученные соотношения показывают, что центр тяжести волнового пакета перемещается с постоянной скоростью $\bar{v} = \bar{p}/m$ и что пакет расплывается в пространстве, как этого и следует ожидать при наличии размытия скоростей. Величина Δx при этом никогда не может, однако, стать меньше $(\Delta x)_{\min}$. Этот результат следует сопоставить

с формулой (15.38), экспоненциальный множитель которой может быть преобразован к виду

$$\psi(x, t) = \frac{\exp\left[+\frac{ix^2ut}{4a(a^2+u^2t^2)} - \frac{x^2}{4(a^2+u^2t^2)}\right]}{(2\pi)^{1/4}(a+iut)^{1/2}}, \quad (15.61)$$

так что $\psi^*\psi$ равно

$$|\psi(x, t)|^2 = \frac{\exp\left[-\frac{x^2}{2(a^2+u^2t^2)}\right]}{\sqrt{2\pi}(a^2+u^2t^2)^{1/2}}. \quad (15.62)$$

Эта функция представляет собой нормированную функцию ошибок Гаусса, причем $(\Delta x)^2$ определяется в этом случае выражением вида

$$(\Delta x)^2 = (a^2 + u^2t^2) = (\Delta x)_0^2 + \left(\frac{\Delta p}{m}\right)^2 t^2, \quad (15.63)$$

где a и u имеют те же значения, что и в формулах (15.39) и (15.40).

Покажем далее, что в отсутствие внешних полей волновой пакет удовлетворяет преобразованию Галилея¹⁾, т. е. что при поступательном движении пакета решение уравнения Шредингера, содержащего время, остается в силе. Этот результат в общем случае справедлив для любого решения одномерного уравнения:

$$\frac{1}{2m} p^2 \psi(x, t) = i\hbar \frac{\partial}{\partial t} \psi(x, t). \quad (15.64)$$

Интересующее нас преобразование имеет вид

$$\varphi(x, t) = \exp\left[\frac{i}{\hbar}\left(p_1 x - \frac{p_1^2}{2m} t\right)\right] \psi\left(x - \frac{p_1}{m} t, t\right), \quad (15.65)$$

причем дополнительная скорость движения равна p_1/m , как это следует из того обстоятельства, что новое значение плотности вероятности равно

$$|\varphi|^2 = \left|\psi\left(x - \frac{p_1}{m} t, t\right)\right|^2, \quad (15.66)$$

т. е. представляет собой старое значение плотности вероятности, перемещающееся поступательно со скоростью p_1/m . Легко убедиться в том, что функция φ удовлетворяет уравнению Шредингера, зависящему от времени. Записывая для краткости $\varphi = \exp(iA)\psi$ и принимая во внимание, что $p \exp(iA) = \exp(iA)(p_1 + p)$, находим

$$\frac{1}{2m} p^2 \varphi = \frac{1}{2m} p [e^{iA} p_1 + e^{iA} p] \psi = \frac{e^{iA}}{2m} [p_1^2 + 2p_1 p + p^2] \psi. \quad (15.67)$$

При вычислении $i\hbar \partial \varphi / \partial t$ следует учесть, что в формуле, определяющей функцию φ , переменная t встречается трижды, и что произведе-

¹⁾ См., например, [19] и [20]. — *Прим. ред.*

ние величины $i\hbar$ на производную по t от первого аргумента функции ψ дает просто $(p_1/m) p \psi$. В результате имеем

$$i\hbar \frac{\partial \varphi}{\partial t} = e^{iA} \left[\frac{p_1^2}{2m} + \frac{p_1 p}{m} + \frac{p^2}{2m} \right] \psi, \quad (15.68)$$

причем производная от последнего аргумента дала член $(p^2/2m) \psi$, так как функция $\psi(x, t)$ удовлетворяет уравнению Шредингера. Мы видим, таким образом, что $(p^2/2m) \varphi = i\hbar \partial \varphi / \partial t$, так что волновой пакет действительно можно подвергнуть преобразованию Галилея, что и требовалось доказать.

Сопоставим теперь значения $\overline{\mathcal{H}}$ для функций φ и ψ . Имеем

$$\begin{aligned} \overline{\mathcal{H}} &= \int \varphi^* \mathcal{H} \varphi dx = \int e^{-iA} \psi \frac{e^{iA}}{2m} [p_1^2 + 2p_1 p + p^2] \psi dx = \\ &= \frac{1}{2m} \int \psi^* [p_1^2 + 2p_1 p + p^2] \psi dx = \\ &= \frac{1}{2m} [p_1^2 + 2p_1 \bar{p} + \bar{p}^2] = \frac{1}{2m} \overline{(p_1 + p)^2}, \end{aligned} \quad (15.69)$$

где средние значения взяты по функциям ψ . Последнее выражение представляет собой ожидаемое значение энергии в том случае, когда распределение импульсов, характеризуемое функцией ψ , смещено на величину p_1 . Если $\bar{p} = 0$, то дополнительная энергия при этом равна энергии поступательного движения массы m , перемещающейся со скоростью $v = p_1/m$. Отсюда очевидно, что движение пакета не должно вызвать изменения величины $(\Delta p)^2$:

$$\begin{aligned} (\Delta p^2) &= \text{средн. по } \varphi (\overline{p^2} - (\bar{p})^2) = \\ &= \text{средн. по } \psi [\overline{(p_1 + p)^2} - (p_1 + \bar{p})^2] = \\ &= \text{средн. по } \psi [p_1^2 + 2p_1 \bar{p} + \bar{p}^2 - p_1^2 - 2p_1 \bar{p} - \bar{p}^2] = \\ &= \text{средн. по } \psi [\overline{p^2} - (\bar{p})^2]. \end{aligned} \quad (15.70)$$

Так как распределение вероятностей $|\varphi|^2$ [см. (15.66)] есть движущееся распределение вероятностей $|\psi|^2$, то очевидно, что функции φ отвечают те же значения величин $(\Delta x)^2$ и $(\Delta p)^2$, что и функции ψ .

Мы можем теперь применить преобразование (15.65) к рассмотренному выше волновому пакету, для которого величина $\Delta p \Delta x$ имеет наименьшее значение, с тем чтобы получить волновой пакет, изображенный на фиг. 65. Применяя преобразование (15.65) к функции (15.61) и вводя обозначения

$$\frac{p_1}{\hbar} = k, \quad \frac{p_1}{m} = v, \quad x = y + vt, \quad (15.71)$$

так что y — расстояние от центра волнового пакета, находим:

$$\begin{aligned} \varphi(x, t) &= \\ &= \frac{\exp\left[-\frac{y^2}{4(a^2 + u^2 t^2)}\right]}{(2\pi)^{1/2} (a + iut)^{1/2}} \exp i\left[k\left(y + \frac{vt}{2}\right) + \frac{y^2 ut}{4a(a^2 + u^2 t^2)}\right]. \end{aligned} \quad (15.72)$$

Отсюда можно получить аналитически ряд интересных результатов, иллюстрируемых фиг. 65. Вблизи $y = 0$, как мы видим, $p\psi = \frac{\hbar}{i} k\psi = p_1\psi$. Вблизи переднего и заднего фронтов пакета, т. е. соответственно при $y = \pm(a^2 + u^2 t^2)^{1/2}$, имеем, однако,

$$p\psi = \frac{\hbar}{i} \left[k \pm \frac{2ut}{4a(a^2 + u^2 t^2)^{1/2}} \right] \psi, \quad (15.73)$$

так что при $ut \gg a$ применение оператора импульса дает

$$p\psi = \left(p_1 \pm \frac{\hbar}{2a} \right) \psi = (p_1 \pm \Delta p) \psi. \quad (15.74)$$

Это изменение импульса в области пакета иллюстрируется фиг. 65.

§ 6. УСКОРЕНИЕ ЭЛЕКТРОННОГО ВОЛНОВОГО ПАКЕТА ЭЛЕКТРИЧЕСКИМИ И МАГНИТНЫМИ ПОЛЯМИ¹⁾

Выше мы подчеркивали, что дырки и электроны можно рассматривать как частицы, которые могут ускоряться электрическими и магнитными полями и рассеиваться на неоднородностях кристалла. В основе этих представлений лежит тот факт, что движение дырочного и электронного волновых пакетов достаточно хорошо подчиняется классическим законам. Связь между скоростью и импульсом при этом определяется формулой, рассмотренной в гл. 6, а связь между силой и скоростью изменения импульса — формулой, приведенной в гл. 7. В настоящем параграфе мы докажем эти формулы для электронного волнового пакета; в дальнейшем будет также рассмотрен волновой пакет, соответствующий дыркам.

Теорема, которую мы хотим доказать, заключается в следующем: если мы построим волновой пакет для локализованного электрона, волновая функция которого соответствует значению импульса \mathbf{P} в энергетической зоне, то этот волновой пакет будет двигаться со скоростью $\mathbf{v} = \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P})$ (аналогичный результат был получен ранее в § 2),

¹⁾ Доказательство теоремы, рассматриваемой в этом параграфе, было также дано Зоммерфельдом и Бете [4], Джонсом и Зинером [1] и Фрелихом [5]. — *Приж. авт.*

а импульс \mathbf{P} при этом будет изменяться по закону

$$\dot{\mathbf{P}} = (-e) \left[\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right], \quad (15.75)$$

в согласии с соотношениями, рассмотренными в гл. 7.

В этой теореме содержатся результаты, которые частично уже были получены различными авторами другими путями. Один из простейших, хотя и не строгих методов ее доказательства основан на использовании определения групповой скорости и принципа сохранения энергии ¹⁾. Если сила, действующая на электрон, равна \mathbf{F} , то, приравнявая мощность, сообщаемую электрону, увеличению его энергии, получаем

$$\mathbf{v} \cdot \mathbf{F} = \frac{d}{dt} \mathcal{E}(\mathbf{P}) = \dot{\mathbf{P}} \cdot \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P}) = \dot{\mathbf{P}} \cdot \mathbf{v}.$$

Этому уравнению удовлетворяет значение $\dot{\mathbf{P}}$, равное

$$\dot{\mathbf{P}} = \mathbf{F}.$$

Однако решением этого уравнения является и вектор

$$\dot{\mathbf{P}} = \mathbf{F} + \frac{1}{|\mathbf{v}|} \mathbf{F} \times \mathbf{v},$$

или любой другой вектор, у которого компоненты, параллельные \mathbf{v} , также равны \mathbf{F} . Кроме того, использование соотношения

$$\mathbf{F} = (-e) \left[\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right]$$

представляет собой дополнительную гипотезу, так как силы, обусловленные наличием поля \mathbf{H} , не совершают работы.

Слэтером [8] предложен другой вывод закона ускорения в электрическом поле, основанный на использовании собственных функций импульса. Сходное решение задачи с учетом влияния магнитных полей было дано Вильсоном [9]. Ни один из упомянутых авторов не рассматривает, однако, вопроса о движении волновых пакетов в явном виде, тогда как в интересующем нас случае проще всего начать непосредственно с исследования законов движения волнового пакета.

Метод, которым мы будем пользоваться, основан на определении приближенного решения уравнения Шредингера для электрона в электрическом и магнитном полях и на доказательстве того, что это решение является искомым. Этот метод аналогичен методу Хаустона [10] с той, однако, разницей, что мы исследуем также и влияние магнитных полей и имеем при этом дело с волновым пакетом. В качестве введения

¹⁾ Этому методу доказательства следуют в своих монографиях Зейтц [6] и Мотт и Джонс [7]. — *Прим. авт.*

изложим кратко сущность метода Хаустона. Предположим, что волновая функция электрона есть функция Блоха

$$\psi_{\mathbf{P}} = \exp \left[i \frac{\mathbf{P} \cdot \mathbf{r}}{\hbar} \right] u_{\mathbf{P}}(\mathbf{r}),$$

где множитель $u_{\mathbf{P}}(\mathbf{r})$ обладает периодичностью решетки и нормирован к единице объема ¹⁾. Эта волновая функция удовлетворяет уравнению

$$\left[\frac{1}{2m} p^2 + \mathcal{U}(\mathbf{r}) \right] \psi_{\mathbf{P}} \equiv \mathcal{H}_0 \psi_{\mathbf{P}} = \mathcal{E}(\mathbf{P}) \psi_{\mathbf{P}}. \quad (15.76)$$

Если электрическое поле \mathbf{E} , обуславливающее наличие силы $-e\mathbf{E}$ и потенциальной энергии электрона $e\mathbf{E} \cdot \mathbf{r}$, приложено в течение промежутка времени dt , то изменение функции $\psi_{\mathbf{P}}$ при этом равно

$$d\psi_{\mathbf{P}} = \frac{1}{i\hbar} dt \mathcal{H} \psi_{\mathbf{P}} = \frac{dt}{i\hbar} [\mathcal{E}(\mathbf{P}) + e\mathbf{E} \cdot \mathbf{r}] \psi_{\mathbf{P}}, \quad (15.77)$$

так что в момент времени $t + dt$ функция $\psi_{\mathbf{P}}$ имеет вид

$$\begin{aligned} \left\{ 1 - i[\mathcal{E}(\mathbf{P}) + e\mathbf{E} \cdot \mathbf{r}] \frac{dt}{\hbar} \right\} \psi_{\mathbf{P}} &= \exp \left\{ -i[\mathcal{E}(\mathbf{P}) + e\mathbf{E} \cdot \mathbf{r}] \frac{dt}{\hbar} \right\} \psi_{\mathbf{P}} = \\ &= \left\{ \exp \left[-i\mathcal{E}(\mathbf{P}) \frac{dt}{\hbar} \right] \right\} \left\{ \exp \left[i(\mathbf{P} - e\mathbf{E} dt) \cdot \frac{\mathbf{r}}{\hbar} \right] \right\} u_{\mathbf{P}}(\mathbf{r}). \end{aligned}$$

Это показывает, что за время dt величина \mathbf{P} изменилась на $(-e)\mathbf{E} dt$, т. е. как раз так, как если бы мы имели $\dot{\mathbf{P}} = (-e)\mathbf{E}$. Хаустон показал, что $u_{\mathbf{P}}$ при этом также приобретает вид $u_{\mathbf{P} - e\mathbf{E} dt}$, так что волновая функция в целом изменяется в соответствии с соотношением $\dot{\mathbf{P}} = (-e)\mathbf{E}$. За исключением тех случаев, когда поле столь велико, что на расстоянии, равном постоянной решетки, электрон приобретает энергию, достаточную для перехода через энергетический зазор, разделяющий разрешенные зоны, функция $\psi_{\mathbf{P} - e\mathbf{E} dt}$, как показал Хаустон, дает удовлетворительное решение задачи. Наш метод также будет заключаться в отыскании изменения функции $\psi_{\mathbf{P}}$ за время dt (производной от $\psi_{\mathbf{P}}$ по времени) с той, однако, разницей, что мы будем иметь дело с волновым пакетом и исследуем, кроме того, влияние магнитного поля.

При наличии электрического и магнитного полей уравнение Шредингера ²⁾ для электрона с зарядом $-e$ содержит скалярный и

¹⁾ В гл. 17 функция $\psi_{\mathbf{P}}$ нормирована к объему, определяемому границами периода, и равна произведению $1/\sqrt{V}$ на функцию $\psi_{\mathbf{P}}$, рассматриваемую в данном параграфе. — *Прим. авт.*

²⁾ См., например, [3]. — *Прим. авт.* (См. также [18] и [21]. — *Прим. ред.*)

векторный потенциалы φ и \mathbf{A} и имеет вид

$$\left[\frac{1}{2m} (\mathbf{p} - \mathbf{p}_A)^2 + \mathcal{U}(\mathbf{r}) + (-e)\varphi \right] \psi = \mathcal{H}\psi = i\hbar \frac{\partial}{\partial t} \psi, \quad (15.78)$$

где

$$\mathbf{p}_A = \frac{(-e)}{c} \mathbf{A}. \quad (15.79)$$

Скалярный и векторный потенциалы определяют электрическое и магнитное поля \mathbf{E} и \mathbf{H} :

$$\mathbf{E} = -\nabla\varphi - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{A}; \quad \mathbf{H} = \nabla \times \mathbf{A}. \quad (15.80a)$$

Эти потенциалы неоднозначны: к φ может быть добавлен член $(-1/c) \partial \chi / \partial t$, а к \mathbf{A} — член $\nabla \chi$, причем это преобразование не повлечет за собой изменения значений \mathbf{E} и \mathbf{H} . Если мы предположим, что потенциал φ создается системой неподвижных зарядов, а потенциал \mathbf{A} возникает благодаря изменению токов и зарядов вне рассматриваемой области, то функция градиентного преобразования χ может быть выбрана таким образом, что $\nabla \cdot \mathbf{A} = 0$. Мы предположим, что это условие выполняется, и в дальнейшем будем соответственно пользоваться уравнениями

$$\nabla \cdot \mathbf{A} = 0 \quad \text{и} \quad \mathbf{p} \cdot \mathbf{p}_A = \mathbf{p}_A \cdot \mathbf{p}. \quad (15.80b)$$

При наличии магнитного поля волновые функции не могут быть собственными функциями оператора скорости $\dot{\mathbf{r}}$. Действительно, воспользовавшись соотношением

$$p_x x - x p_x = \frac{\hbar}{i},$$

находим, что при наличии магнитного поля оператор, соответствующий $\dot{\mathbf{r}}$, уже не равен \mathbf{p}/m ; новый оператор для $\dot{\mathbf{r}}$ обладает при этом такой формой, что не существует волновой функции, которая являлась бы одновременно собственной функцией всех трех составляющих $\dot{\mathbf{r}}$. С помощью соотношения коммутативности

$$p_x x - x p_x = \frac{\hbar}{i}$$

и формулы (14.43), определяющей производную по времени от некоторого оператора, получаем

$$\dot{\mathbf{r}} = \frac{1}{i\hbar} (\mathbf{r}\mathcal{H} - \mathcal{H}\mathbf{r}) = \frac{\mathbf{p} - \mathbf{p}_A}{m}. \quad (15.81)$$

В случае однородного поля при $\mathbf{H} = \mathbf{i}_z H_z$ можно положить $\mathbf{A} = \mathbf{i}_y H_z x$; очевидно, что при этом

$$\dot{x}\dot{y} = \left[\frac{\hbar}{i} \frac{\partial}{\partial x} \right] \left[\frac{\hbar}{i} \frac{\partial}{\partial y} + \frac{e}{c} H_z x \right] = \dot{y}\dot{x} + \frac{\hbar e}{ic} H_z, \quad (15.82)$$

так что \dot{x} и \dot{y} не коммутируют, и, следовательно, никакая волновая функция не может быть одновременно собственной функцией как \dot{x} , так и \dot{y} . Благодаря этому при наличии магнитного поля из рассмотрения исключаются неограниченно-протяженные волновые функции, подобные тем, которыми пользуется Хаустон; это обстоятельство не играет, однако, существенной роли в случае исследуемых нами волновых пакетов.

Мы должны теперь построить волновую функцию, описывающую движение электрона в магнитном поле и обладающую теми же свойствами, что и волновая функция ψ_P для электрона в разрешенной зоне в отсутствие магнитного поля. Ключ к решению этой задачи дает градиентное преобразование функции ψ , сопровождающее соответствующее преобразование потенциалов A и φ . Если ψ удовлетворяет уравнению

$$\left[\frac{1}{2m} (\mathbf{p} - \mathbf{p}_A)^2 + \mathcal{U} + (-e)\varphi \right] \psi = i\hbar \frac{\partial}{\partial t} \psi,$$

то волновая функция

$$\psi' = e^{i(-e/c)\chi/\hbar} \psi \quad (15.83)$$

удовлетворяет уравнению, в котором

$$A' = A + \nabla\chi, \quad \varphi' = \varphi - \frac{1}{c} \frac{\partial}{\partial t} \chi, \quad (15.84)$$

как это легко проверить путем подстановки. Если центром волнового пакета выбрана точка \mathbf{r}_0 , то можно считать, что наличие члена $\mathbf{p}_A(\mathbf{r}_0, t)$ частично обусловлено градиентным преобразованием при

$$\chi = A(\mathbf{r}_0, t) \cdot \mathbf{r} = \frac{-c}{e} \mathbf{p}_A(\mathbf{r}_0, t) \cdot \mathbf{r}.$$

Это заключение приводит нас к необходимости исследования свойств функции

$$\psi = e^{i\mathbf{p}_{A0} \cdot \mathbf{r}/\hbar} \psi_P(\mathbf{r}), \quad (15.85)$$

где

$$\mathbf{p}_{A0} \equiv \mathbf{p}_{A0}(t) = \mathbf{p}_A(\mathbf{r}_0(t), t).$$

Очевидно, что 1) функция ψ дает то же значение плотности $|\psi_P|^2$, что и функция ψ_P . Докажем также, что 2) вблизи точки \mathbf{r}_0 волновая функция (15.85) приближенно может рассматриваться как собственная функция оператора \mathcal{H} при значении энергии, соответствующем состоянию \mathbf{P} в отсутствие магнитного поля, и, далее, что 3) среднее значение оператора скорости

$$\dot{\mathbf{r}} = \frac{\mathbf{p} - \mathbf{p}_A}{m}$$

для элементарной ячейки, расположенной в точке \mathbf{r}_0 , равно просто значению $\mathbf{v}(\mathbf{P}) = \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P})$ для электрона в состоянии \mathbf{P} .

Таким образом, вблизи точки \mathbf{r}_0 волновая функция (15.85) действительно описывает состояние движения, характеризующееся величинами \mathbf{P} и $\mathbf{v}(\mathbf{P})$. После проверки утверждений 2) и 3) построим из функций (15.85) волновой пакет.

Для доказательства положения 2) вычислим $\mathcal{H}\psi$, применяя при этом все операции только к первому множителю функции (15.85):

$$\begin{aligned} \mathcal{H}\psi = e^{i\mathbf{p}_{A0} \cdot \mathbf{r}/\hbar} & \left[\mathcal{H}_0 + (-e)\varphi + \frac{1}{m}(\mathbf{p}_{A0} - \mathbf{p}_A) \cdot \mathbf{p} + \right. \\ & \left. + \frac{1}{2m}(\mathbf{p}_{A0} - \mathbf{p}_A)^2 \right] \psi_{\mathbf{P}}. \end{aligned} \quad (15.86)$$

При вычислении третьего члена этого выражения мы учли то обстоятельство, что $\nabla \cdot \mathbf{p}_A$ и $\nabla \cdot \mathbf{p}_{A0}$ равны нулю. В точке $\mathbf{r} = \mathbf{r}_0$ разность $(\mathbf{p}_{A0} - \mathbf{p}_A)$ равна нулю, так что (15.86) дает

$$\begin{aligned} \mathcal{H}\psi = e^{i\mathbf{p}_{A0} \cdot \mathbf{r}/\hbar} [\mathcal{H}_0 + (-e)\varphi] \psi_{\mathbf{P}} = e^{i\mathbf{p}_{A0} \cdot \mathbf{r}/\hbar} [\mathcal{E}(\mathbf{P}) + (-e)\varphi] \psi_{\mathbf{P}} = \\ = [\mathcal{E}(\mathbf{P}) + (-e)\varphi] \psi. \end{aligned}$$

Таким образом, в этой точке ψ является собственной функцией оператора \mathcal{H} , а энергия равна при этом сумме энергии $\mathcal{E}(\mathbf{P})$, соответствующей электрону в разрешенной зоне, и энергии $(-e)\varphi$, обусловленной его взаимодействием с дополнительным электростатическим полем. Если \mathbf{r} отлично от \mathbf{r}_0 , то φ изменяется; как мы увидим, это изменение φ оказывает на величину $i\hbar\dot{\psi}$ точно такое же влияние, как и наличие члена $-E \cdot \mathbf{r}$ в теории Хаустона. Кроме того, вследствие магнитных эффектов изменяются также значения разности $(\mathbf{p}_{A0} - \mathbf{p}_A)$. Членом, содержащим $(\mathbf{p}_{A0} - \mathbf{p}_A)^2$, при этом можно пренебречь по сравнению с членом, пропорциональным $(\mathbf{p}_{A0} - \mathbf{p}_A)$, при условии, что размеры волнового пакета не слишком велики. В этом можно убедиться путем приближенной оценки обоих членов, используя в качестве оператора \mathbf{p} его среднее значение для тепловых электронов, принимая радиус пакета равным произведению некоторого числа n на постоянную решетки a и полагая $A = i_y H_z x$. Квадратичный член при этом равен

$$\frac{1}{2m} \left[\frac{eH_z n a}{c} \right]^2 = \frac{m}{2} (\omega n a)^2, \quad \text{где } \omega = \frac{eH_z}{mc}. \quad (15.87)$$

Линейный член мы получим, положив $|\mathbf{p}| = m\mathbf{v}$, где \mathbf{v} — тепловая скорость:

$$\frac{1}{m} \frac{eH_z n a}{c} m\mathbf{v} = m(\omega n a)\mathbf{v}. \quad (15.88)$$

Отношение (15.88) к (15.87) равно

$$\frac{v}{\omega a} = \frac{v/\omega}{\pi a}.$$

Последнее выражение представляет собой отношение радиуса v/ω круговой орбиты электрона в поле H_z к радиусу волнового пакета. Из физических соображений очевидно, что если это отношение мало, то электрон будет двигаться по окружности, размеры которой меньше размеров волнового пакета; такое внутреннее движение может в этом случае оказаться более существенным, чем движение пакета в целом. Для тепловых электронов, однако, $v = 10^7$ см/сек, а в поле 10^8 гаусс $\omega \approx 10^{10}$ сек.⁻¹, так что радиус орбиты равен 10^{-8} см и совпадает с радиусом пакета при $\pi a = 10^{-8}$ см, т. е. при $n \approx 10^5$. Мы будем иметь дело с пакетами значительно меньших размеров и поэтому квадратичным членом будем пренебрегать.

Значение $\dot{\mathbf{r}}$, усредненное по элементарной ячейке, расположенной в точке \mathbf{r}_0 , равно

$$\langle \dot{\mathbf{r}} \rangle_{\mathbf{r}=\mathbf{r}_0} = \frac{\int_{\text{ячейка } 0} \psi^* \frac{1}{m} (\mathbf{p} - \mathbf{p}_A) \psi dV}{\int_{\text{ячейка } 0} |\psi|^2 dV} = \quad (15.89a)$$

$$= \frac{\int_{\text{ячейка } 0} \psi_P^* \frac{1}{m} (\mathbf{p} + \mathbf{p}_{A0} - \mathbf{p}_A) \psi_P dV}{\int_{\text{ячейка } 0} |\psi_P|^2 dV} = \quad (15.89б)$$

$$= \mathbf{v}(\mathbf{P}), \quad (15.89в)$$

где равенство (15.89a) следует из (15.81), равенство (15.89б) получается в результате применения оператора \mathbf{p} к экспоненциальному множителю функции (11.85), а к равенству (15.89в) мы приходим, пренебрегая разностью $(\mathbf{p}_{A0} - \mathbf{p}_A)$ по сравнению с \mathbf{p} и используя результаты, полученные выше в § 4. Утверждение 3), следовательно, правильно, и волновая функция ψ действительно соответствует средней скорости, равной \mathbf{v} .

Волновой пакет. Построим теперь волновой пакет, умножив волновую функцию ψ_P на весовой множитель $\omega(\mathbf{r}, t)$ и на временной множитель $\exp[-if(t)/\hbar]$. Предположим, что как ω , так и f вещественны, так что f определяет только комплексную фазу, а ω — форму пакета. Позднее [см. (15.105) и (15.106)] будет показано, что полученная таким путем функция является решением уравнения Шредингера. Напишем, таким образом:

$$\psi = e^{i[\mathbf{p}_{A0} \cdot \mathbf{r} - f]/\hbar} \omega(\mathbf{r}, t) \psi_P(\mathbf{r}) = e^{i\theta} \omega \psi_P. \quad (15.90)$$

Последней сокращенной формой записи мы часто будем пользоваться в дальнейшем. Условие нормировки функции ψ имеет вид

$$\int |\psi|^2 dV = \int |\psi_{\mathbf{r}}|^2 |\omega|^2 dV = \int |\omega|^2 dV = 1,$$

так как предполагается, что $\psi_{\mathbf{r}}$ нормировано к единице объема, а ω изменяется медленно, так что в подынтегральном выражении $|\psi_{\mathbf{r}}|^2$ может быть заменено его средним значением, равным единице. Предположим, что функция ω обладает колоколообразной формой, например имеет вид $\exp[-(\mathbf{r}-\mathbf{r}_0)^2/(na)^2]$, так что в точке \mathbf{r}_0 имеем $\nabla\omega = 0$ и положение центра пакета определяется равенством

$$\mathbf{r}_0 = \int \mathbf{r} |\omega|^2 dV. \quad (15.91)$$

Для получения волнового пакета можно было бы также воспользоваться иным методом; вместо того чтобы умножать $\psi_{\mathbf{r}}$ на функцию ω , можно было бы просуммировать функции $\psi_{\mathbf{r}}$; оба эти метода приводят к одному и тому же результату, но первый из них более удобен.

Применим теперь к функции ψ оператор \mathcal{H} . Поступая при этом так же, как при выводе (15.86), и пренебрегая членом, пропорциональным $(\mathbf{p}_{A0} - \mathbf{p}_A)^2$, получаем

$$\begin{aligned} \mathcal{H}\psi &= e^{i\theta} \left\{ \frac{1}{2m} [p^2 + 2(\mathbf{p}_{A0} - \mathbf{p}_A) \cdot \mathbf{p}] + \mathcal{U} + (-e)\varphi \right\} \omega \psi_{\mathbf{r}} = \\ &= e^{i\theta} \left\{ \omega [\mathcal{H}_0 + (-e)\varphi] \psi_{\mathbf{r}} + \frac{1}{m} (\mathbf{p}_{A0} - \mathbf{p}_A) \cdot (\omega \mathbf{p} \psi_{\mathbf{r}} + \psi_{\mathbf{r}} \mathbf{p} \omega) + \right. \\ &\quad \left. + \frac{1}{m} \mathbf{p} \omega \cdot \mathbf{p} \psi_{\mathbf{r}} + \frac{1}{2m} \psi_{\mathbf{r}} p^2 \omega \right\} \quad (15.92) \end{aligned}$$

(буквы, которыми отмечены отдельные члены, понадобятся нам в дальнейшем для ссылок). В точке $\mathbf{r} = \mathbf{r}_0$ снова имеем

$$\mathcal{H}\psi = [\mathcal{E}(\mathbf{P}) + (-e)\varphi] \psi;$$

это уравнение не содержит, однако, члена (e), которым, как будет показано ниже, в первом приближении можно пренебречь. Аналогично, среднее значение $\dot{\mathbf{r}}$ для ячейки, расположенной в точке \mathbf{r}_0 , снова приближенно равно \mathbf{v} ; в дальнейшем мы покажем, что весь пакет в целом движется с этой скоростью.

За исключением членов (b) и (d), все остальные члены, определяющие $\mathcal{H}\psi$, представляют собой произведения медленно меняющихся функций от $\dot{\mathbf{r}}$ на функцию $\psi_{\mathbf{r}}$. Члены же (b) и (d), содержащие произведение $\mathbf{p}\psi_{\mathbf{r}}$, в каждой элементарной ячейке ведут себя совершенно иначе, чем сама функция $\psi_{\mathbf{r}}$. Эти члены также представляют собой, однако, функции вида $\exp[i(\mathbf{P} \cdot \mathbf{r})/\hbar]$, умноженные на периодическую функцию от \mathbf{r} . Они могут быть поэтому разложены в ряд по волновым функциям $\psi_{\mathbf{r}}(\mathbf{r})$, соответствующим другим энергетическим

зонам, нежели функции $\psi_{\mathbf{r}}$. Если мы предположим, что функции $\psi_{\mathbf{r}}$ отвечают „нулевой“ зоне, и для простоты опустим индекс 0, так что $\psi_{0\mathbf{r}}(\mathbf{r}) = \psi_{\mathbf{r}}(\mathbf{r})$, то можно записать

$$\mathbf{p}\psi_{\mathbf{r}} = \sum_i (i|\mathbf{p}|0) \psi_{i\mathbf{r}}, \quad (15.93)$$

где

$$(i|\mathbf{p}|0) = \int \psi_{i\mathbf{r}}^* \mathbf{p}\psi_{\mathbf{r}} dV, \quad (15.94)$$

причем интегрирование производится здесь по единице объема. При $i = 0$ имеем

$$(0|\mathbf{p}|0) = m\mathbf{v} = m\nabla_{\mathbf{r}}\mathcal{E}(\mathbf{p}) \quad (15.95)$$

в согласии с результатами, полученными выше в § 4. В дальнейшем мы докажем, что в случае обычных электрических и магнитных полей роль членов, содержащих функции $\psi_{i\mathbf{r}}$, ничтожна. Это доказательство отложим до конца настоящего параграфа, а пока будем пренебрегать всеми эффектами, связанными с учетом членов, содержащих $\psi_{i\mathbf{r}}(\mathbf{r})$.

Покажем теперь, что членами (с) и (е) можно соответственно пренебречь по сравнению с членами (b) и (d). Сохраняя в (b) только член, содержащий $(0|\mathbf{p}|0)$, имеем (по порядку величины):

$$(b) : (c) = \omega\psi_{\mathbf{r}} m\mathbf{v} : \omega\psi_{\mathbf{r}} \frac{\hbar}{na} = m\mathbf{v} : \frac{\hbar}{na}. \quad (15.96)$$

Тепловому электрону соответствует длина волны порядка 50 \AA , и, следовательно, в этом случае $m\mathbf{v} = \hbar/50 \text{ \AA}$. Таким образом, если $na > 50 \text{ \AA}$, т. е. если размеры волнового пакета во много раз больше длины волны, то член (с) много меньше члена (b). Совершенно аналогичные результаты получаются и при сравнительной оценке членов (d) и (е), так что членом (е) также можно пренебречь. Как и в § 5, этот член характеризует расплывание волнового пакета. В данном изложении мы этим эффектом будем пренебрегать. Приближенное значение $\mathcal{H}\psi$ может быть, таким образом, записано в виде

$$\mathcal{H}\psi = e^{i\theta} \left\{ \omega [\mathcal{H}_0 + (-e)\varphi] \psi_{\mathbf{r}} + \frac{\omega}{m} (\mathbf{p}_{A0} - \mathbf{p}_A)^{(b)} \cdot \mathbf{p}\psi_{\mathbf{r}} + \frac{1}{m} \mathbf{p}\omega \cdot \mathbf{p}\psi_{\mathbf{r}} \right\}. \quad (15.97)$$

Член \mathcal{H}_0 оставлен здесь в форме оператора для того, чтобы в дальнейшем формула (15.97) могла быть использована для оценки роли членов, содержащих функции $\psi_{i\mathbf{r}}$.

Вычислим теперь $i\hbar(\partial\psi/\partial t)$. Для этого необходимо исследовать два выражения. Первое выражение имеет вид

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{p}_{A0}(t) \cdot \mathbf{r} &= \frac{\partial}{\partial t} \mathbf{p}_A [\mathbf{r}_0(t), t] \cdot \mathbf{r} = (\dot{\mathbf{r}}_0 \cdot \nabla) \mathbf{p}_A \cdot \mathbf{r} + \\ &+ \frac{\partial}{\partial t} \mathbf{p}_A \cdot \mathbf{r} = \left[(\dot{\mathbf{r}}_0 \cdot \nabla) \mathbf{p}_A + \frac{\partial}{\partial t} \mathbf{p}_A \right]_0 \cdot \mathbf{r}. \end{aligned} \quad (15.98)$$

Последнюю строку этого равенства получаем, записав \mathbf{p}_A как $\mathbf{p}_A(\mathbf{r}, t)$ и произведя вычисления при $\mathbf{r} = \mathbf{r}_0$. Второе выражение имеет вид

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \psi_{\mathbf{r}} &= i\hbar \dot{\mathbf{P}} \cdot \nabla_{\mathbf{r}} \psi_{\mathbf{r}} = i\hbar \dot{\mathbf{P}} \cdot \nabla_{\mathbf{r}} e^{i(\mathbf{P} \cdot \mathbf{r})/\hbar} u_{\mathbf{r}} = \\ &= -\dot{\mathbf{P}} \cdot \mathbf{r} \psi_{\mathbf{r}} + i\hbar e^{i(\mathbf{P} \cdot \mathbf{r})/\hbar} \dot{\mathbf{P}} \cdot \nabla_{\mathbf{r}} u_{\mathbf{r}}(\mathbf{r}). \end{aligned} \quad (15.99)$$

Последний член этого равенства обладает той же периодичностью, что и функция $\psi_{\mathbf{r}}$, и подобно $\mathbf{r} \psi_{\mathbf{r}}$ может быть разложен в ряд по функциям $\psi_{i\mathbf{r}}$:

$$e^{i(\mathbf{P} \cdot \mathbf{r})/\hbar} \nabla_{\mathbf{r}} u_{\mathbf{r}}(\mathbf{r}) = \sum_i (i | \nabla_{\mathbf{r}} | 0) \psi_{i\mathbf{r}}. \quad (15.100)$$

В отличие от формулы (15.94) здесь

$$(i | \nabla_{\mathbf{r}} | 0) = \int u_{i\mathbf{r}}^* \nabla_{\mathbf{r}} u_{\mathbf{r}} dV. \quad (15.101)$$

Эти члены, в свою очередь, могут быть выражены¹⁾ через $(i | \mathbf{p} | 0)$; в этом нет, однако, необходимости, поскольку нас сейчас вполне удовлетворяют приближенные качественные результаты. Верхний предел значения $(i | \nabla_{\mathbf{r}} | 0)$ может быть найден следующим образом: при изменении \mathbf{P} , сопровождающем переход из центра зоны Бриллюэна в одну из точек на ее боковой грани, т. е. при изменении \mathbf{P} на $\hbar/2a$, функция $u_{\mathbf{r}}$ изменяется примерно на 100% (см., например, фиг. 5б). Поэтому $\nabla_{\mathbf{r}} u_{\mathbf{r}}$ будет того же порядка, что и $(2a/\hbar) u_{\mathbf{r}}$. Отсюда следует, что $(i | \nabla_{\mathbf{r}} | 0)$ не должно превышать $2a/\hbar$. В действительности изменения $\nabla_{\mathbf{r}} u_{\mathbf{r}}$ не будут точно следовать изменениям $u_{\mathbf{r}}$ или $u_{i\mathbf{r}}$, так что интеграл, содержащий произведение этих функций, может быть значительно меньше $2a/\hbar$. Эта оценка очень груба, но для наших целей она вполне достаточна. Можно также убедиться в том, что величина $(0 | \nabla_{\mathbf{r}} | 0)$ является чисто мнимой, показав, что она равна своей комплексно сопряженной величине, взятой с обратным знаком. Последнее следует из того обстоятельства, что волновые функции нормированы для всех значений \mathbf{P} , так что

$$\nabla_{\mathbf{r}} \int u_{\mathbf{r}}^* u_{\mathbf{r}} dV = (0 | \nabla_{\mathbf{r}} | 0)^* + (0 | \nabla_{\mathbf{r}} | 0) = 0. \quad (15.102)$$

Теперь можно вычислить $i\hbar (\partial \psi / \partial t)$, воспользовавшись с этой целью соотношениями (15.98) — (15.100):

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \psi &= e^{i\mathbf{t}} \left\{ \dot{f} \omega - \left[(\dot{\mathbf{r}}_0 \cdot \nabla) \mathbf{p}_A + \frac{\partial}{\partial t} \mathbf{p}_A + \dot{\mathbf{P}} \right]_0 \cdot \mathbf{r} \omega \psi_{\mathbf{r}} + \right. \\ &\quad \left. + i\hbar \frac{\partial \omega}{\partial t} \psi_{\mathbf{r}} + i\hbar \dot{\mathbf{P}} \cdot \sum_i (i | \nabla_{\mathbf{r}} | 0) \psi_{i\mathbf{r}} \right\}. \end{aligned} \quad (15.103)$$

¹⁾ См. [9], § 2.81. — Прим. авт.

Это выражение снова содержит волновые функции, соответствующие другим энергетическим зонам; коэффициенты при этих функциях, однако, малы. Сначала мы будем пренебрегать этими членами, а несколько позднее оценим их.

Для того чтобы исследовать изменения параметров f , ω , \mathbf{P} и \mathbf{r}_0 (последний зависит от ω), характеризующих волновой пакет, умножим левую и правую части уравнения

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t}$$

на $\exp(-i\epsilon)\psi_{\mathbf{P}}^*$ и проинтегрируем полученное выражение по элементарной ячейке, расположенной в точке \mathbf{r} , предполагая при этом, что все медленно меняющиеся функции остаются постоянными и обладают значениями, отвечающими точке \mathbf{r} . Если подинтегральная функция равна произведению $\psi_{\mathbf{P}}$ на $\psi_{i\mathbf{P}}$, то интеграл равен нулю; интеграл от произведения $\psi_{\mathbf{P}}$ на $\psi_{\mathbf{P}}$ равен просто объему элементарной ячейки. Таким путем получаем левую и правую части следующего уравнения:

$$\begin{aligned} \omega \left[\mathcal{E}(\mathbf{P}) + (-e)\varphi|_0 + (\mathbf{r} - \mathbf{r}_0) [(-e)\nabla\varphi - \nabla(\mathbf{p}_A \cdot \mathbf{v})|_0 \right] \omega + \frac{\hbar}{i} \nabla\omega \cdot \mathbf{v} = \\ = \dot{f}\omega - \left[(\dot{\mathbf{r}}_0 \cdot \nabla) \mathbf{p}_A + \frac{\partial}{\partial t} \mathbf{p}_A + \dot{\mathbf{P}} \right] \cdot [\mathbf{r}_0\omega + (\mathbf{r} - \mathbf{r}_0)\omega] + \\ + i\hbar \frac{\partial}{\partial t} \omega + i\hbar \dot{\mathbf{P}} \cdot (\mathbf{0} | \nabla_{\mathbf{P}} | \mathbf{0}) \omega. \quad (15.104) \end{aligned}$$

Члены, содержащие \mathbf{v} , получены в соответствии с (15.95) из тех членов выражения (15.97), которые содержат $\mathbf{r}\psi_{\mathbf{P}}$. Члены, содержащие $(\mathbf{r} - \mathbf{r}_0)$, получаются в результате разложения медленно меняющихся функций φ , \mathbf{p}_A и \mathbf{g} в ряд Тэйлора вблизи точки $\mathbf{r} = \mathbf{r}_0$; все частные производные от этих функций появились в связи с тем, что мы представили φ и \mathbf{p}_A как функции \mathbf{r} и t ; значения этих производных взяты при этом в точке $\mathbf{r} = \mathbf{r}_0$, что отмечено в уравнении (15.104) индексом нуль [см., например, (15.98)]. Буквы, которыми отмечены отдельные члены уравнения (15.104), указывают, из каких членов предыдущих уравнений они получены.

В уравнении (15.104) все величины, за исключением мнимой единицы i , вещественны. Это уравнение содержит, таким образом, *две группы вещественных членов*: (I) члены, представляющие собой произведение функции ω на некоторую постоянную, равные соответственно \dot{f} , (a_1) , $(g_0) \cdot (g_1)$ и (k) ; (II) вещественные члены, пропорциональные разности $(\mathbf{r} - \mathbf{r}_0)$. Кроме того, имеются (III) *два мнимых члена*, содержащих производные от ω . Приравнивая первую группу членов нулю, имеем

$$\dot{f} = (a_1) + (g_0) \cdot (g_1) - (k). \quad (15.105)$$

[Можно убедиться в том, что член (к) играет очень малую роль, соответствуя энергии порядка 10^{-4} эв. Это будет показано в конце настоящего параграфа. Во всяком случае, член (к) не должен влиять на плотность заряда и, следовательно, на положение или движение волнового пакета, так как он входит в ψ только через θ .] Полученное соотношение (15.105) имеет следующий смысл. Напомним, что при анализе уравнения

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t}$$

было вычислено значение $\mathcal{H}\psi$ [см. (15.92)], причем было показано, что при $\mathbf{r} = \mathbf{r}_0$ оно равно $\mathcal{E}(\mathbf{P}) + (-e)\varphi(\mathbf{r}_0)$. С другой стороны, совокупность членов \dot{f} , $(g_0) \cdot (g_1)$ и (к) в правой части уравнения (15.104) определяет $i\hbar (\partial\psi/\partial t)$. Таким образом, выбор f в соответствии с уравнением (15.105) означает правильное определение скорости изменения комплексной фазы функции ψ , при котором дифференцирование ψ по времени дает $\mathcal{E}(\mathbf{P}) + (-e)\varphi$. Поскольку произведение (g_0) на (g_1) зависит от выбора начала координат, параметр \dot{f} сам по себе не имеет существенного физического смысла.

Анализируя мнимые члены (III), находим

$$\frac{\partial}{\partial t} w = -\mathbf{v} \cdot \nabla w.$$

Это означает, что

$$w = w\left(\mathbf{r} - \int \mathbf{v} dt\right), \quad (15.106)$$

и показывает, что пакет как целое движется со скоростью \mathbf{v} , обладая при этом неизменной формой. Отсюда очевидно, что

$$\dot{\mathbf{r}}_0 = \mathbf{v}. \quad (15.107)$$

В действительности пакет будет расплываться, причем его расплывание характеризуется членом (е), а также другими малыми членами, которыми мы в нашем приближении пренебрегли.

Совокупность членов (II) может равняться нулю при любых значениях разности $(\mathbf{r} - \mathbf{r}_0)$ только в том случае, если сумма векторов, на которые умножены значения $(\mathbf{r} - \mathbf{r}_0)$, в свою очередь, равна нулю. Это условие приводит к уравнению, которое может быть решено относительно $\dot{\mathbf{P}}$:

$$\dot{\mathbf{P}} = -(-e)\nabla\varphi + \nabla(\mathbf{p}_A \cdot \mathbf{v}) - \left[(\dot{\mathbf{r}}_0 \cdot \nabla)\mathbf{p}_A + \frac{\partial}{\partial t}\mathbf{p}_A \right].$$

Для упрощения члена (b) воспользуемся соотношениями

$$\nabla(\mathbf{p}_A \cdot \mathbf{v}) = (\mathbf{v} \cdot \nabla)\mathbf{p}_A + \mathbf{v} \times (\nabla \times \mathbf{p}_A),$$

$$\nabla \times \mathbf{p}_A = \frac{(-e)}{c} \nabla \times \mathbf{A} = \frac{-e}{c} \mathbf{H},$$

а для вычисления члена $(g_0) \cdot (g_2)$ — равенством

$$\dot{\mathbf{r}}_0 = \mathbf{v}.$$

Используя далее формулы (15.79) и (15.80а)

$$(-e)\mathbf{E} = (-e) \left[-\nabla\varphi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right] = -(-e)\nabla\varphi - \frac{\partial}{\partial t} \mathbf{P}_A,$$

получаем искомое выражение для $\dot{\mathbf{P}}$:

$$\dot{\mathbf{P}} = (-e) \left[\mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{H} \right]. \quad (15.108)$$

Оставляя временно в стороне исследование вопроса о роли других энергетических зон, проанализируем смысл полученных результатов. Мы построили волновую функцию ψ , для которой плотность заряда вблизи точки \mathbf{r}_0 имеет то же значение $|\psi_{\mathbf{p}}|^2$, что и для волновой функции, соответствующей состоянию \mathbf{P} в зоне Бриллюэна. В точке \mathbf{r}_0 эта волновая функция совершенно точно удовлетворяет уравнению

$$\mathcal{H}\psi = [\mathcal{E}(\mathbf{P}) + (-e)\varphi]\psi,$$

соответствуя энергии, равной сумме энергии состояния \mathbf{P} , и дополнительной энергии, обусловленной наличием поля φ . В элементарной ячейке, расположенной в точке \mathbf{r}_0 , среднее значение скорости равно, таким образом, \mathbf{v} и связано с \mathbf{P} обычным соотношением $\mathbf{v} = \nabla_{\mathbf{p}} \mathcal{E}(\mathbf{P})$. Рассматриваемая волновая функция ψ представляет собой, следовательно, пространственно ограниченную часть безгранично протяженной волновой функции $\psi_{\mathbf{p}}$, описывающей состояние электрона в отсутствие электрического и магнитного полей. Изменения этой волновой функции определяются уравнением

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t}.$$

Выше мы связали изменение ψ с величинами \mathbf{P} , f и ω , характеризующими волновой пакет. Следует отметить, что наш метод анализа групп членов (I), (II) и (III) был до некоторой степени произвольным. Таким путем мы нашли, однако, искомое решение, и, поскольку временная зависимость ψ однозначно определяется уравнением

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t},$$

полученное нами значение скорости изменения ψ также должно быть единственно возможным. Найденное решение обладает удобной формой, дающей возможность его непосредственной интерпретации с точки зрения движения пакета в целом и изменений эффективного значения \mathbf{P} . Полученные результаты находятся в согласии с результатами, установленными в гл. 7.

Вернемся теперь к анализу членов, соответствующих другим энергетическим зонам. Волновая функция, полученная путем суперпозиции функций $\psi_{\mathbf{p}}$, не может являться решением уравнения

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t},$$

так как множитель $\rho_{\mathbf{p}}^{\psi}$ в выражениях (b) и (d), а также $\nabla_{\mathbf{p}}$ в выражении (k) содержат члены, соответствующие другим энергетическим зонам. Мы должны поэтому видоизменить эту волновую функцию, добавив к $\psi_{\mathbf{p}}$ сумму членов $a_i \psi_{i\mathbf{p}}$. Коэффициенты a_i (они могут зависеть от \mathbf{r}) можно вычислить, умножив уравнение

$$\mathcal{H}\psi = i\hbar \frac{\partial \psi}{\partial t}$$

на $\exp(-i\epsilon) \psi_{i\mathbf{p}}^*$ и проинтегрировав полученное выражение по элементарной ячейке, расположенной в точке \mathbf{r} . Совокупность членов (1) теперь уже не будет обращаться в нуль, так как для i -й энергетической зоны

$$\mathcal{H}_0 \psi_{i\mathbf{p}} = \mathcal{E}_i(\mathbf{p}) \psi_{i\mathbf{p}}.$$

Это приводит к появлению члена вида

$$[\mathcal{E}_i(\mathbf{p}) - \mathcal{E}(\mathbf{p})] a_i.$$

Все остальные члены уравнения, содержащие a_i в качестве коэффициентов, зависят от размеров волнового пакета или от пространственной протяженности полей и малы по своей величине. Подстановка приближенного значения $\psi_{\mathbf{p}}$ в (b), (d) и (k) дает члены следующего порядка величины:

$$(b) \quad \frac{1}{m} (\mathbf{p}_{A0} - \mathbf{p}_A) \cdot (i | \mathbf{p} | 0) \approx \frac{1}{m} \frac{eH_z n a}{c a} \frac{\hbar}{a} = n \hbar \omega;$$

$$(d) \quad \frac{1}{m} (i | \mathbf{p} | 0) \cdot \frac{\mathbf{p} \omega}{\omega} \approx \frac{1}{m} \frac{\hbar}{a} \cdot \frac{\hbar}{n a} = \frac{1}{m n} \left[\frac{\hbar}{a} \right]^2;$$

$$(e) \quad \hbar \dot{\mathbf{p}} \cdot (i | \nabla_{\mathbf{p}} | 0) \sim \hbar e \mathbf{E} \cdot \frac{2a}{h} = \frac{e \mathbf{E} a}{\pi}.$$

С помощью фиг. 88 (стр. 280) можно убедиться в том, что все эти члены значительно меньше разности $\mathcal{E}_i(\mathbf{p}) - \mathcal{E}(\mathbf{p})$, составляющей несколько электрон-вольт, так что $|a_i| \ll 1$ и влияние других энергетических зон на свойства волнового пакета незначительно. (Вопрос о сходимости добавленной нами суммы членов $a_i \psi_{i\mathbf{p}}$ может быть исследован с помощью правила „f-суммы“; см., например, книги Вильсона [9] и Зейтца [6].) Роль члена (e) была также исследована несколькими иными методами Хаустоном и ранее Зинером. Этому члену отвечает заметная вероятность перехода из одной зоны в другую, если падение напряжения $e\mathbf{E}$ на

расстоянии порядка размеров элементарной ячейки a сравнимо с шириной запрещенной зоны $\mathcal{E}_i(\mathbf{P}) - \mathcal{E}(\mathbf{P})$ в той точке, где разрешенные зоны расположены наиболее близко друг к другу. Это соответствует полям порядка 1 в на элементарную ячейку, или $10^7 - 10^8 \text{ в/см}$. Как показал Хаустон, в случае более слабых полей роль других энергетических зон оказывается незначительной. Мы не будем воспроизводить здесь строгое доказательство этого утверждения и ограничимся качественными результатами, полученными выше.

§ 7. ПРИНЦИП ПАУЛИ И АНТИСИММЕТРИЯ ВОЛНОВЫХ ФУНКЦИЙ

Исследуем теперь в общем виде волновую функцию для системы M электронов, занимающих определенный ряд квантовых состояний, описываемых волновыми функциями $\chi_1, \chi_2, \dots, \chi_M$. Каждая из этих функций зависит от координат $\mathbf{r} = x, y, z$ электрона и его спинового квантового числа s , которое может принимать одно из двух значений: $+1/2$ или $-1/2$. Удобно иметь дело с волновыми функциями, являющимися собственными функциями s , и ввести в рассмотрение спиновые волновые функции $\alpha(s)$ и $\beta(s)$, где

$$\alpha\left(\frac{1}{2}\right) = \beta\left(-\frac{1}{2}\right) = 1, \quad \alpha\left(-\frac{1}{2}\right) = \beta\left(\frac{1}{2}\right) = 0.$$

Волновые функции χ_i при этом могут быть записаны в форме

$$\chi_n(q_i) = \varphi_n(\mathbf{r}_i) \alpha(s_i), \quad (15.109)$$

где q_i — совокупность значений координат и спина. Функция χ_n характеризует, таким образом, как положение электрона, так и его спин. В следующем параграфе будут исследованы некоторые общие свойства функций χ .

Если электрону i отвечает волновая функция χ_n , то вероятность его нахождения в элементе объема $dV_i = dx_i dy_i dz_i$ с данным значением спина s_i (мы будем говорить при этом кратко, что электрон находится в элементе dq_i) равна

$$|\chi_n|^2 dq_i = |\varphi_n(\mathbf{r}_i)|^2 dV_i \alpha^2(s_i). \quad (15.110)$$

Полную вероятность нахождения электрона i в объеме dV_i при любом значении его спина обозначим через $\rho_n(\mathbf{r}_i) dV_i$. Эту вероятность найдем, просуммировав (15.110) по двум возможным значениям s_i :

$$\rho_n(\mathbf{r}_i) dV_i = |\varphi_n(\mathbf{r}_i)|^2 dV_i \left[\alpha^2\left(\frac{1}{2}\right) + \alpha^2\left(-\frac{1}{2}\right) \right] = |\varphi_n(\mathbf{r}_i)|^2 dV_i. \quad (15.111)$$

Аналогичный результат получается также и в случае функций $\varphi_n(\mathbf{r}_i) \beta(s_i)$, так что плотность частиц ρ_n , определяемая функцией χ_n , не зависит от вида спиновой функции. При вычислении средних

значений следует суммировать по всем возможным значениям s_i , а также интегрировать по объему; сокращенно мы обозначим эти действия как $\int dq_i$. Очевидно, что если функции φ_n ортонормированы, то этим свойством обладают также и функции χ_n , и, далее, что волновые функции, отвечающие противоположно направленным спинам, взаимно ортогональны, так как выражение $\alpha(s)\beta(s)$ при суммировании по s дает нуль.

В следующем параграфе будет исследован специальный случай волновых функций φ , относящихся к зоне валентных связей, в которой недостает одного электрона. Здесь мы рассмотрим лишь общие вопросы, связанные с постановкой задачи о системе M электронов, описываемых M волновыми функциями $\chi_1, \chi_2, \dots, \chi_M$. При этом предположим, что как φ_n , так и χ_n образуют ортонормированные системы функций, подобно тому как это имеет место в случае волновых функций, отвечающих зоне Бриллюэна.

Если в каждом из M состояний находится по одному электрону, то волновая функция системы имеет вид ¹⁾

$$\chi_0 = \chi_1(q_1)\chi_2(q_2)\chi_3(q_3)\dots\chi_M(q_M). \quad (15.112)$$

Столь же пригодными будут, однако, и волновая функция, отвечающая тому случаю, когда электрон 1 находится в состоянии, описываемом функцией χ_2 , а электрон 2 — в состоянии, описываемом функцией χ_1 , а также волновые функции, получаемые в результате других аналогичных перестановок. В наиболее общем случае такие функции могут быть записаны в виде

$$\Delta\chi_0 = \chi_1(\Delta q_1)\chi_2(\Delta q_2)\dots\chi_M(\Delta q_M). \quad (15.113)$$

Здесь Δ — оператор перестановок, применение которого приводит к перераспределению значений q_i . Например, условия $\Delta q_1 = q_2$, $\Delta q_2 = q_1$ и $\Delta q_i = q_i$ при $i > 2$ определяют оператор Δ , применение которого приводит к взаимным перестановкам электронов 1 и 2. Всего существует $M!$ перестановок M различных значений q_i и, следовательно, $M!$ различных возможных операторов Δ (в том числе и оператор $\Delta = 1$ — тождественная перестановка, не вносящая никаких изменений). Общее число волновых функций вида $\Delta\chi_0$ равно, таким образом, $M!$.

Исследуем теперь два вопроса, тесно связанные друг с другом: 1) Какая из функций $\Delta\chi_0$, или же какая линейная комбинация этих функций правильно описывает состояние системы M электронов?

¹⁾ Такая волновая функция будет собственной функцией оператора \mathcal{H} только при специальных условиях; см. § 8 гл. 15. Этот вопрос нас здесь, однако, интересовать не будет. — *Прим. авт.*

2) Как следует сформулировать принцип Паули, чтобы из рассмотренных были исключены такие волновые функции, как, например,

$$\chi_2(q_1)\chi_2(q_2)\chi_3(q_3)\dots\chi_M(q_M),$$

где двум электронам отвечает одна и та же функция χ_2 ?

Оба эти вопроса получают удовлетворительное решение, если мы выдвинем следующее требование: волновая функция системы M электронов должна быть такой функцией переменных q [мы обозначим ее через $A(q_1, q_2, \dots, q_n)$], чтобы результат взаимной перестановки любых двух электронов был эквивалентен умножению A на -1 . Таким образом,

$$A(q_2, q_1, q_3, q_4, \dots, q_M) = -A(q_1, q_2, q_3, q_4, \dots, q_M)$$

или, в общем случае,

$$A(\Delta q) = \varepsilon(\Delta) A(q), \quad (15.114)$$

где $\varepsilon(\Delta) = \pm 1$ в зависимости от того, осуществляется ли перестановка Δ в результате четного или же нечетного числа отдельных попарных перестановок электронов. Как известно из теории определителей, данная перестановка Δ может быть осуществлена различными способами путем ряда попарных перестановок отдельных значений q_i . Далее, общее число таких попарных перестановок, которые приводят к перестановке Δ , всегда является либо четным, либо нечетным, в зависимости от вида оператора Δ . Таким образом, операторы Δ можно подразделить на „четные“ и „нечетные“, определив при этом $\varepsilon(\Delta)$ так, как это показано выше. Волновая функция такого типа называется *антисимметричной*; с этим связано и принятое нами для нее обозначение A .

Воспользовавшись этими определениями, можно доказать, что единственная волновая функция типа A , которая может быть получена путем линейной комбинации функции χ_0 и функций $\Delta\chi_0$, имеет вид

$$A = \frac{1}{\sqrt{M!}} \sum \varepsilon(\Delta) \Delta\chi_0. \quad (15.115)$$

Другие эквивалентные волновые функции могут быть получены из функции A в результате ее умножения на некоторую постоянную; каких-либо иных функций типа A , не связанных линейной зависимостью с функцией (15.115), не существует. Множитель $1/\sqrt{M!}$, как будет показано ниже, служит для нормировки волновой функции [см. текст, следующий за выражением (15.121)]. Функция (15.115) может быть записана в виде определителя

$$A = \frac{1}{\sqrt{M!}} \begin{vmatrix} \chi_1(q_1) & \chi_2(q_1) & \dots & \chi_M(q_1) \\ \chi_1(q_2) & \chi_2(q_2) & \dots & \chi_M(q_2) \\ \dots & \dots & \dots & \dots \\ \chi_1(q_M) & \chi_2(q_M) & \dots & \chi_M(q_M) \end{vmatrix}, \quad (15.116)$$

как в этом легко убедиться, разложив этот определитель по минорам. Очевидно, что если какие-либо две из волновых функций, определяющих функцию χ_0 , одинаковы, то два столбца определителя также окажутся при этом одинаковыми и определитель обратится в нуль. Это показывает, что требование антисимметрии A исключает возможность двукратного применения одной и той же волновой функции χ при определении A . Этот результат представляет собой аналитическую формулировку более простого утверждения о том, что два электрона не могут занимать одно и то же квантовое состояние.

Вычислим теперь плотность электронов, отвечающую функции A . Это может быть сделано следующим образом. Согласно основным положениям квантовой механики, вероятность того, что координаты электрона 1 заключены в интервале

$$dq_1 = dx_1 dy_1 dz_1, \text{ причем } s_1 = -\frac{1}{2} \text{ либо } s_1 = +\frac{1}{2},$$

$$\text{т. е. } dq_1 = dV_1 \quad \text{и} \quad s_1 = -\frac{1}{2} \text{ либо } s_1 = +\frac{1}{2},$$

а координаты электрона 2 лежат при этом в интервале dq_2 и т. д. должна определяться выражением вида

$$|A(q_1, q_2, \dots, q_M)|^2 dq_1 dq_2 \dots dq_M. \quad (15.117)$$

Если волновая функция правильно нормирована, то интегрирование этого выражения по всем значениям q_i должно дать единицу. Полную вероятность нахождения электрона 1 в интервале dq_1 , независимо от положений и спинов электронов 2, ..., M , получим, проинтегрировав (15.117) по q_2 и т. д. (т. е. проинтегрировав это выражение по всем координатам и просуммировав по s_2). Это дает

$$\rho_1(q_1) dq_1 = dq_1 \int_{q_2} \dots \int_{q_M} |A|^2 dq_2 \dots dq_M. \quad (15.118)$$

Величина $\rho_1(q_1)$ представляет собой вероятность нахождения электрона 1 в интервале dq_1 . Умножив ее на заряд электрона и просуммировав по s_1 , получим среднее значение плотности заряда, отвечающей этому электрону. Для вычисления ρ_1 можно воспользоваться формулой (15.115), определяющей функцию A . Имеем

$$\rho_1(q_1) = \frac{1}{M!} \int_{q_2} \dots \int_{q_M} \sum e(\Delta) \Lambda \chi_0^* \sum e(\mathcal{R}) \mathcal{R} \chi_0 dq_2 \dots dq_M, \quad (15.119)$$

где первая сумма относится к ряду $M!$ операторов Δ , а вторая означает независимое суммирование по \mathcal{R} . Все функции, входящие в произведение

$$\chi_0 = \chi_1(q_1) \chi_2(q_2) \dots \chi_M(q_M),$$

различны и предполагаются ортонормированными. Поэтому в выражении (15.119) единственными членами, дающими при интегрировании отличные от нуля результаты, являются члены, для которых $\Lambda = \mathcal{R}$. Во всех остальных случаях перестановка значений q_i для функции χ_0^* будет давать иные результаты, нежели для функции χ_0 , и, следовательно, при интегрировании в рассмотрение войдет выражение вида

$$\chi_m^*(q_i) \chi_n(q_i) dq_i = 0.$$

Таким образом, (15.119) сводится к сумме по Λ :

$$\rho_1(q_1) = \frac{1}{M!} \sum \int_{q_2} \dots \int_{q_M} \Lambda \chi_0^* \Lambda \chi_0 dq_2 \dots dq_M. \quad (15.120)$$

В состав этой суммы входит $(M-1)!$ членов, содержащих произведение $\chi_1^*(q_1) \chi_1(q_1)$, так как число возможных перестановок координат $q_2 \dots q_M$ между остальными функциями $\chi_2 \dots \chi_M$ равно $(M-1)!$. Каждая из $(M-1)$ операций интегрирования любого из этих членов дает множитель

$$\int \chi_n^*(q_i) \chi_n(q_i) dq_i = 1,$$

так что при суммировании член $\chi_1^*(q_1) \chi_1(q_1)$ встречается $(M-1)!$ раз.

Аналогичный результат получим для любого из произведений $\chi_n^*(q_1) \chi_n(q_1)$, так что окончательно имеем

$$\rho_1(q_1) = \frac{(M-1)!}{M!} \sum_n \chi_n^*(q_1) \chi_n(q_1) = \frac{1}{M} \sum_n |\chi_n(q_1)|^2. \quad (15.121)$$

Если это выражение мы проинтегрируем по q_1 , то получим в результате единицу, так как функции χ_n предполагаются нормированными. (*Это показывает, что функция A нормирована правильно.*) Можно поэтому интерпретировать ρ_1 как плотность вероятности для электрона 1, причем она просто равна среднему значению плотностей заряда, определяемых M волновыми функциями χ_i . Из симметрии функции A и результата, полученного для $\rho_1(q_1)$, очевидно, что точно такое же значение плотности дает и каждый из остальных электронов. Это приводит к следующему определению среднего значения полной плотности заряда $\rho_T(q_1)$:

$$\rho_T(q_1) dq_1 = \sum_n |\chi_n(q_1)|^2 dq_1; \quad (15.122)$$

здесь $\rho_T(q_1) dq_1$ — среднее число электронов в элементе dq_1 , т. е. сумма вероятности нахождения электрона 1 в элементе dq_1 и аналогичных вероятностей для всех остальных электронов. Результирующая плотность заряда, определяемого функцией A , имеет, таким образом,

в точности то же значение, что и для функции χ_0 . Это значит, что принцип Паули не нарушает аддитивности плотностей зарядов.

Совершенно аналогичные результаты могут быть получены и для операторов тока, отвечающих отдельным электронам. В силу тех же условий ортогональности члены, содержащие произведение функций χ_n^* и χ_m , при $n \neq m$ обращаются в нуль и окончательный результат таков: среднее значение тока, обусловленного M электронами, равно просто сумме M значений токов, определяемых волновыми функциями $\chi_1, \chi_2 \dots \chi_M$.

Покажем теперь, что если временная зависимость каждой из волновых функций χ_i определяется уравнением

$$\mathcal{H}_1 \chi_i(q_1, t) = i\hbar \frac{\partial}{\partial t} \chi_i(q_1, t), \quad (15.123)$$

где \mathcal{H}_1 — оператор Гамильтона для электрона 1 при наличии внешнего поля, то временную зависимость функции A можно найти, выразив A через зависящие от времени функции χ_i .

Уравнение, которому должна удовлетворять функция A , имеет вид

$$\mathcal{H}_T A = (\mathcal{H}_1 + \mathcal{H}_2 + \dots + \mathcal{H}_M) A = i\hbar \frac{\partial}{\partial t} A. \quad (15.124)$$

Покажем, что если A задано в форме (15.115), то каждый член выражения (15.115) в отдельности удовлетворяет уравнению (15.124) и, следовательно, ему удовлетворяет также и функция A . Достаточно, таким образом, исследовать функцию χ_0 . Имеем

$$\begin{aligned} \mathcal{H}_T \chi_0 &= [\mathcal{H}_1 \chi_1(q_1, t)] \chi_2(q_2, t) \dots \chi_M(q_M, t) + \\ &+ \chi_1(q_1, t) [\mathcal{H}_2 \chi_2(q_2, t)] \dots \chi_M(q_M, t) + \\ &\dots \\ &+ \chi_1(q_1, t) \chi_2(q_2, t) \dots [\mathcal{H}_M \chi_M(q_M, t)] = \\ &= i\hbar \dot{\chi}_1(q_1, t) \chi_2(q_2, t) \dots \chi_M(q_M, t) + \\ &+ i\hbar \chi_1(q_1, t) \dot{\chi}_2(q_2, t) \dots \chi_M(q_M, t) + \\ &\dots \\ &+ i\hbar \chi_1(q_1, t) \chi_2(q_2, t) \dots \dot{\chi}_M(q_M, t) = \\ &= i\hbar \frac{\partial}{\partial t} \chi_0, \quad (15.125) \end{aligned}$$

что приводит к искомому результату.

Уравнение Фока. Исследованные выше волновые функции, заданные в форме определителей, представляют собой лишь приближенные решения волнового уравнения, учитывающего взаимодействие электронов. Задача об отыскании более точных значений функций χ для

заданного оператора Гамильтона была исследована Фоком [12], а также Слэтером [13]. Полученные при этом уравнения, которым должны удовлетворять отдельные одноэлектронные функции χ , называются *уравнениями Фока*. При выводе этих уравнений предполагается, что на каждый электрон действует некоторое среднее поле, создаваемое остальными электронами; при этом при усреднении принимается во внимание принцип Паули. Уравнения Фока рассмотрены в монографиях Френкеля [14] и Зейтца [6]; в последней даны специальные приложения этих уравнений к вопросам, рассматриваемым в настоящем параграфе, и используется теорема Купменса [15], чтобы показать, что изменения энергии системы определяются изменениями значений $\mathcal{E}(\mathbf{P})$, отвечающих волновым функциям отдельных электронов.

§ 8. ДЫРОЧНЫЙ ВОЛНОВОЙ ПАКЕТ

В этом параграфе мы будем иметь дело с зоной валентных связей, в которой недостает одного электрона, так что фактически мы будем изучать свойства дырки. Для простоты предположим, что волновая функция при этом полностью определяется состояниями с одним и тем же значением спина, т. е. что все состояния с другим значением спина заняты. Если зона полностью заполнена, то это значит, что все возможные квантовые состояния заняты. Следовательно, электрическое и магнитное поля не могут в этом случае изменить характер распределения электронов по квантовым состояниям, так как, согласно принципу Паули, все состояния уже заняты электронами. Поэтому электрические и магнитные поля не оказывают существенного влияния на электроны, находящиеся в заполненной зоне; эти электроны принимают лишь участие в создании диэлектрической постоянной и диамангнитной восприимчивости. Однако если в зоне валентных связей недостает электрона, то электрическое поле создает ток, который в магнитном поле ведет себя так, как этого следовало бы ожидать в случае положительного заряда. Сказанное выше составляет содержание теоремы, которую мы докажем, основываясь на принципе Паули и уравнении (15.75).

Предположим, что пространственные (но не спиновые) части волновых функций всех электронов, находящихся в зоне валентных связей, удовлетворяют одному и тому же уравнению Шредингера

$$\mathcal{H}_0 \psi_a = \left[\frac{1}{2m} p^2 + \mathcal{U}(\mathbf{r}) \right] \psi_a(\mathbf{r}) = \mathcal{E}_a \psi_a(\mathbf{r}). \quad (15.126)$$

Если имеют место периодические граничные условия, то решениями ψ_a этого уравнения являются N волновых функций $\psi_{\mathbf{p}}$, где N — число элементарных ячеек в кристалле. При любых других граничных условиях также существует N волновых функций, отвечающих той же энергетической зоне. Поскольку одновременный учет наличия периодических граничных условий и влияния магнитных полей встречает

серьезные трудности, мы будем в дальнейшем предполагать, что вне рассматриваемого прямоугольного образца кристалла волновые функции обращаются в нуль. Соответствующие волновые функции будем обозначать через ψ_α .

Прежде всего путем наложения волновых функций ψ_α построим волновую функцию φ_1 , описывающую волновой пакет. Поскольку система волновых функций ψ_α является полной, эти функции могут быть использованы для получения пространственно ограниченного волнового пакета, соответствующего электрону в зоне валентных связей точно так же, как и волновые функции ψ_p , рассмотренные в § 6. Однако когда такой волновой пакет достигнет границы кристалла, он должен от нее отразиться, так как вне кристалла функции ψ_α обращаются в нуль, тогда как волновой пакет, образованный функциями ψ_p при наличии периодических граничных условий, достигнув какой-либо грани кристалла, должен был бы пройти через нее. В дальнейшем мы будем предполагать, что речь идет о волновых пакетах, соответствующих дырке, находящейся вдали от границ кристалла, так что эти трудности не будут нас беспокоить.

Рассмотрим, таким образом, волновую функцию $\varphi_1(\mathbf{r}, t)$, представляющую собой сумму функций ψ_α :

$$\varphi_1(\mathbf{r}, t) \equiv \sum_{\alpha=1}^N a_{1\alpha}(t) \psi_\alpha(\mathbf{r}). \quad (15.127)$$

Предположим, что φ_1 нормировано. Под влиянием электрического и магнитного полей волновой пакет будет приобретать ускорение [в согласии с уравнениями (15.75) и (15.78)] и волновая функция φ_1 будет при этом соответственно изменяться. Ниже мы будем подразумевать под φ_1 зависящую от времени функцию, которая удовлетворяет волновому уравнению

$$\mathcal{H}\varphi_1 = i\hbar \frac{\partial}{\partial t} \varphi_1, \quad (15.128)$$

где оператор \mathcal{H} содержит также члены, обусловленные наличием электрического и магнитного полей.

Определим теперь $(N-1)$ волновых функций $\varphi_n(\mathbf{r}, t)$, отличных от φ_1 , которые также являются решениями уравнения (15.128) и вместе с функцией φ_1 образуют ортонормированную систему функций. Это возможно потому, что в любой момент времени, например при $t=0$, из функции φ_1 и любых $(N-1)$ функций ψ_α , которые вместе с функцией φ_1 образуют линейно независимый ряд решений, всегда может быть построена ортонормированная группа функций¹⁾. Функции φ_n могут быть заданы в виде

$$\varphi_n = \sum_{\alpha=1}^N a_{n\alpha}(t) \psi_\alpha(\mathbf{r}), \quad (15.129)$$

¹⁾ См., например, [16, 17]. — *Прим. авт.*

где зависимость a_{na} от t определяется уравнением (15.128). Если функции φ_n и φ_1 ортонормированы в момент времени $t=0$ и удовлетворяют при этом одному и тому же уравнению (15.128), то они ортонормированы также и во все последующие моменты времени. Действительно, исходя из свойств эрмитовости оператора \mathcal{H} , имеем

$$\begin{aligned} \frac{d}{dt} \int \varphi_n^* \varphi_m dV &= \int \left[\left(\frac{1}{i\hbar} \mathcal{H} \varphi_n \right)^* \varphi_m + \varphi_n^* \frac{1}{i\hbar} \mathcal{H} \varphi_m \right] dV = \\ &= \frac{1}{i\hbar} \int \left[-(\mathcal{H} \varphi_n)^* \varphi_m + \varphi_n^* \mathcal{H} \varphi_m \right] dV = 0. \end{aligned} \quad (15.130)$$

Отсюда следует, что численное значение интеграла $\int \varphi_n^* \varphi_m dV$ остается постоянным, так что условие ортонормированности выполняется.

Плотность заряда и плотность тока в случае дырочного волнового пакета. Воспользуемся теперь $2N$ волновыми функциями, а именно, функциями $\varphi_1(\mathbf{r})\alpha(s), \dots, \varphi_N(\mathbf{r})\alpha(s)$, отвечающими спину одного знака, и функциями $\varphi_1(\mathbf{r})\beta(s), \dots, \varphi_N(\mathbf{r})\beta(s)$, отвечающими спину противоположного знака, для того чтобы в соответствии с соображениями, изложенными в § 7, записать волновую функцию в форме определителя. При этом функция φ_1 , описывающая движение дырки, предполагается построенной из волновых функций, соответствующих области, близкой к верхнему краю зоны валентных связей, где, как это было показано при анализе уравнения (7.20), энергия убывает с увеличением P . Это обстоятельство существенно для характеристики поведения дырки.

Рассмотрим прежде всего волновую функцию, отвечающую заполненной зоне. В этом случае антисимметричная волновая функция имеет вид определителя, образованного всеми функциями φ . Обозначим ее через $A_f(\varphi)$. Функции φ являются линейными комбинациями функций ψ_a , поэтому некоторые столбцы определителя $A_f(\varphi)$ представляют собой линейные комбинации столбцов определителя $A_f(\psi)$, образованного всеми функциями ψ . Поскольку определитель не изменяется при добавлении к одному из его столбцов линейной комбинации других столбцов, отсюда следует, что волновые функции $A_f(\varphi)$ и $A_f(\psi)$ отличаются друг от друга лишь множителем, абсолютное значение которого равно единице.

При наличии электрических полей функция $A_f(\psi)$ будет определять значения диэлектрической и диамagnetной восприимчивости; никакого тока проводимости мы при этом, однако, иметь не будем. Поскольку $A_f(\varphi)$ та же волновая функция, что и $A_f(\psi)$, она будет, разумеется, меняться во времени точно таким же образом и обладать при этом теми же свойствами, что и функция $A_f(\psi)$.

Рассмотрим теперь антисимметричную функцию $A_p(\varphi)$, образованную всеми волновыми функциями за исключением функции $\varphi_1\alpha$.

Соответствующую плотность заряда обозначим через ρ_p . Согласно уравнению (15.122), плотность ρ_p равна сумме плотностей зарядов, отвечающих каждой из волновых функций, входящих в состав функции $A_p(\varphi)$. Она связана поэтому с плотностью электронов ρ_f (но не с плотностью заряда), определяемой функцией $A_f(\varphi)$, соотношением

$$\rho_p(\mathbf{r}) + \rho_1(\mathbf{r}) = \rho_f(\mathbf{r}), \quad \rho_1(\mathbf{r}) \equiv |\varphi_1(\mathbf{r})|^2, \quad (15.131)$$

так как $\rho_p(\mathbf{r})$ отличается от $\rho_f(\mathbf{r})$ только тем, что не содержит члена $|\varphi_1(\mathbf{r})|^2 \equiv \rho_1(\mathbf{r})$. Аналогичным образом, поток частиц (но не электрический ток), определяемый функцией $A_p(\varphi)$, будет отличаться от потока частиц, определяемого функцией $A_f(\varphi)$, только членом, соответствующим волновому пакету φ_1 . Если J — поток частиц (число электронов/см²·сек), то плотность электрического заряда и плотность электрического тока соответственно равны

$$-e\rho_p - e\rho_1 = -e\rho_f, \quad \text{или} \quad -e\rho_p = -e\rho_f + e\rho_1, \quad (15.132)$$

$$-eJ_p - eJ_1 = -eJ_f, \quad \text{или} \quad -eJ_p = -eJ_f + eJ_1. \quad (15.133)$$

Таким образом, добавление дырки будет изменять ток и заряд так, как если бы был добавлен заряд $+e$, движение которого описывается волновой функцией φ_1 .

Чтобы доказать, что поведение такой дырки соответствует поведению частицы, обладающей положительным зарядом и положительным значением массы, следует вычислить ее ускорение. Поскольку волновой пакет построен из волновых функций, отвечающих области, близкой к верхнему краю зоны, значение $\mathcal{E}(\mathbf{P})$ определяется при этом формулой (7.20). Это приводит к соотношению

$$\mathbf{v} = -\frac{\mathbf{P}}{m}, \quad (15.134)$$

где m — положительная величина. Подстановка (15.134) в формулу, определяющую $\dot{\mathbf{P}}$, дает

$$m\dot{\mathbf{v}} = -\dot{\mathbf{P}} = +e \left[\mathcal{E} + \frac{\mathbf{v} \times \mathbf{H}}{c} \right], \quad (15.135)$$

что соответствует положительному заряду.

Таким образом, если в заполненной зоне недостает электрона и функция, характеризующая волновой пакет, соответствует верхнему краю зоны, так что соотношение между \mathbf{v} и \mathbf{P} содержит знак минус [см. (15.134)], то движение остальных электронов происходит так, как если бы под действием поля перемещался волновой пакет, отвечающий положительному заряду и положительной массе.

Эти результаты могут быть сформулированы в более общем виде, если мы введем в рассмотрение импульс и энергию любого из электронов, находящихся в одной из зон. Обозначим через \mathbf{P}_n и \mathbf{P}_p , $\mathcal{E}_c(\mathbf{P}_n)$ и $\mathcal{E}_v(\mathbf{P}_p)$, $\mathbf{v}_n = \nabla_{\mathbf{P}_n} \mathcal{E}_c(\mathbf{P}_n)$ и $\mathbf{v}_p = \nabla_{\mathbf{P}_p} \mathcal{E}_v(\mathbf{P}_p)$ значения импульса,

энергии и групповой скорости соответственно для электрона в зоне проводимости и для дырки в зоне валентных связей. Энергия \mathcal{E}_B и импульс \mathbf{P}_B в любой точке зоны равны $\mathcal{E}_B = \mathcal{E}_0 + \mathcal{E}_c(\mathbf{P}_n)$ и $\mathbf{P}_B = \mathbf{P}_n$ для электрона и равны $\mathcal{E}_B = \mathcal{E}_0 - \mathcal{E}_v(\mathbf{P}_p)$ и $\mathbf{P}_B = -\mathbf{P}_p$ для дырки. Воспользовавшись этими определениями, можно получить симметричные выражения для скорости и ускорения волнового пакета:

Электрон в зоне проводимости

Дырка

$$\begin{aligned} \mathbf{P}_B &= \mathbf{P}_n & \mathbf{P}_B &= -\mathbf{P}_p \\ \mathcal{E}_B &= \mathcal{E}_0 + \mathcal{E}_c(\mathbf{P}_B) & \mathcal{E}_B &= \mathcal{E}_0 - \mathcal{E}_v(-\mathbf{P}_B) = \\ & & &= \mathcal{E}_0 - \mathcal{E}_v(\mathbf{P}_B), \end{aligned} \quad (15.136)$$

$$\begin{aligned} \mathbf{v}_n &= \nabla_{\mathbf{P}_n} \mathcal{E}_c(\mathbf{P}_n) = & \mathbf{v}_p &= \nabla_{\mathbf{P}_p} \mathcal{E}_v(\mathbf{P}_p) = \\ &= \nabla_{\mathbf{P}_B} \mathcal{E}_B(\mathbf{P}_B) & &= \nabla_{\mathbf{P}_B} \mathcal{E}_B(\mathbf{P}_B), \end{aligned} \quad (15.137a)$$

$$\begin{aligned} \dot{\mathbf{P}}_B &= \dot{\mathbf{P}}_n = & \dot{\mathbf{P}}_B &= -\dot{\mathbf{P}}_p = \\ &= -e \left[\mathcal{E} + \frac{1}{c} \mathbf{v}_n \times \mathbf{H} \right] & &= +e \left[\mathcal{E} + \frac{1}{c} \mathbf{v}_p \times \mathbf{H} \right]. \end{aligned} \quad (15.137b)$$

Равенство (15.136) для дырки является следствием того обстоятельства, что для любой энергетической зоны $\mathcal{E}_v(\mathbf{P}) = \mathcal{E}_v(-\mathbf{P})$; равенство (15.137a) для дырки следует из того, что при этом $\mathbf{v}(\mathbf{P}) = -\mathbf{v}(-\mathbf{P})$.

Согласно (15.137), групповая скорость и величина $\dot{\mathbf{P}}_B$, будучи выражены через энергию и импульс любого из электронов, определяются одинаково и теми же уравнениями, различаясь лишь знаком заряда. Обе эти скорости соответствуют положительному значению массы, так как в обоих случаях при возрастании \mathbf{P}_B от значения, отвечающего минимуму энергии, величина $\mathcal{E}_B(\mathbf{P}_B)$ возрастает. Минимальное значение энергии соответствует при этом неподвижной частице.

Дырочный волновой пакет и собственные функции оператора энергии. Волновая функция $A_p(\varphi)$, отвечающая дырочному волновому пакету, не является собственной функцией оператора Гамильтона для $(2N-1)$ электронов, находящихся в зоне валентных связей. Эта функция может быть, однако, получена путем суперпозиции собственных функций каждого из $(2N-1)$ электронов. Как будет показано ниже, исследование этой задачи, а также результаты ее решения дадут возможность ознакомиться с рядом свойств дырок.

Напомним, что функция φ_1 получена путем суперпозиции волновых функций ψ_α :

$$\varphi_1(\mathbf{r}, t) = \sum_{\alpha=1}^N \alpha_{1\alpha}(t) \psi_\alpha(\mathbf{r})$$

[см. (15.127)]. При этом ψ_α являются собственными функциями одно-электронного оператора Гамильтона \mathcal{H}_0 , обладающего собственными значениями \mathcal{E}_α [см. 15.126)]. Отсюда следует, что антисимметричная волновая функция, построенная из функций ψ_α , есть собственная функция для системы $(2N-1)$ электронов с собственным значением $\mathcal{E} = \sum \mathcal{E}_\alpha$ (суммирование производится здесь по всем ψ_α , образующим определитель). Это легко может быть доказано с помощью методов, использованных в § 7 при изучении временной зависимости волновой функции, заданной в форме определителя [см. (15.125)]. Для обозначения таких волновых функций удобно в качестве индекса ввести квантовое число β , отвечающее волновой функции ψ_β недостающего электрона. Волновую функцию системы $(2N-1)$ электронов будем обозначать соответственно через $M_\beta(\psi)$. Приводимое ниже уравнение (15.138) дает правило определения комплексной фазы функции M_β . Энергия, соответствующая волновой функции $M_\beta(\psi)$, если отсчитывать ее от верхнего края зоны валентных связей, равна, очевидно, $-\mathcal{E}_\beta$.

Наша задача заключается теперь в том, чтобы найти связь между антисимметричной функцией $A_p(\varphi)$, образованной всеми функциями φ помимо функции φ_1 , и функцией $M_\beta(\psi)$. Как будет показано ниже, эта связь дается формулой

$$A_p(\varphi) = \sum_\alpha a_{1\alpha}^* M_\alpha(\psi), \quad (15.138)$$

где $a_{1\alpha}^*$ — комплексно сопряженные значения коэффициентов $a_{1\alpha}$, входящих в формулу (15.127).

Соотношение (15.138) может быть получено следующим образом: волновую функцию, отвечающую заполненной зоне, мы разложим в ряд двумя разными способами и приравняем затем получающиеся при этом выражения. Если фазы функций $A_p(\varphi)$ и $A_r(\psi)$ выбраны правильно, то

$$A_r(\psi) = A_r(\varphi). \quad (15.139)$$

Обе волновые функции имеют вид определителей, образованных волновыми функциями $2N$ электронов, причем координаты электрона 1 фигурируют в первой строке определителя [см. (15.116)]. Каждая из них может быть поэтому представлена в виде ряда, содержащего произведения функций, записанных в первой строке на миноры определителя. Миноры определителя $A_r(\psi)$ равны, очевидно, функциям $M_\alpha(\psi)$, представленным как функции координат и спинов электронов 2, 3, ..., $2N$. Таким образом,

$$A_r(\psi) = \sum_\alpha \psi_\alpha(q_1) M_\alpha(\psi; q_2, \dots, q_{2N}). \quad (15.140)$$

Это выражение служит для определения фаз функций M_α . Аналогично

$$A_r(\varphi) = \sum_\alpha \varphi_\alpha(q_1) M_\alpha(\varphi; q_2, \dots, q_{2N}), \quad (15.141)$$

где $M_1(\varphi; q_2, \dots, q_{2N})$, очевидно, отличается от $A_p(\varphi)$ только фазовым множителем. Для того чтобы выразить $M_1(\varphi)$ через $M_\alpha(\psi)$, следует приравнять (15.141) и (15.140), умножить полученное выражение на $\varphi_1^*(q_1)$ и проинтегрировать его по q_1 (что означает также и суммирование по s_1). Поскольку функции φ ортонормированы,

$$M_1(\varphi; q_2, \dots, q_{2N}) = \sum_{\alpha} \int \varphi_1^*(q_1) \psi_{\alpha}(q_1) dq_1 M_{\alpha}(\psi; q_2, \dots, q_{2N}). \quad (15.142)$$

Коэффициент при M_{α} легко найти с помощью соотношения (15.127), умножив последнее на ψ_{α}^* , проинтегрировав и взяв затем комплексно сопряженное. Коэффициенты в разложении (15.142) оказываются при этом равными $\alpha_{1\alpha}^*$. Этот результат эквивалентен соотношению (15.138), которое и требовалось доказать.

Очевидно, что если значения \mathcal{E}_{α} , отвечающие отдельным волновым функциям M_{α} в формуле (15.138), неодинаковы, то это значит, что функции M_{α} соответствуют различным значениям энергии, так что, как об этом было сказано выше, $A_p(\varphi)$ не является собственной функцией оператора энергии.

Разложение $A_p(\varphi)$ в ряд по функциям M_{α} дает нам метод исследования дырок, не связанный с использованием одноэлектронного приближения, вводимого выражением (15.126). Для зоны, в которой недостает одного электрона, существует $(N-1)$ собственных функций, дающих точные решения задачи, в отличие от приближенных функций в форме определителя, рассмотренных нами выше. Если из этих точных решений построить волновую функцию (15.138), то мы получим точное решение, отвечающее недостающему отрицательному заряду. Эти решения дадут, далее, результаты (15.137а) и (15.137б), где \mathbf{v} определяется, как $\nabla_{\mathbf{P}_S} \mathcal{E}_S(\mathbf{P}_S)$, а \mathbf{P}_S и $\mathcal{E}_S(\mathbf{P}_S)$ соответствуют точным решениям задачи для системы в целом. Таким путем поведение дырки можно исследовать на основании данных об энергетической зоне, в которой недостает одного электрона, совершенно не пользуясь при этом волновыми функциями, получаемыми в одноэлектронном приближении. Хотя проведенное выше исследование и было связано с рядом приближений, основные его результаты, касающиеся свойств дырок, носят общий характер и могли бы быть получены также и при точном решении задачи.

§ 9. ПЕРИОДИЧЕСКИЕ ГРАНИЧНЫЕ УСЛОВИЯ И ЭЛЕКТРИЧЕСКИЕ ПОЛЯ

В этом параграфе в известной степени содержатся результаты, уже полученные в § 6. Его содержание, однако, существенно отличается от содержания § 6. В гл. 6 мы вычислили ток, воспользовавшись волновыми функциями Блоха, соответствующими однородному распределению плотности заряда в кристалле. Вычисление электро-

проводности кристаллов часто проводится на основе такой картины; однако этот метод обычно представляется внутренне противоречивым. Противоречие возникает в связи с тем обстоятельством, что волновые функции, удовлетворявшие вначале периодическим граничным условиям, вследствие постепенного изменения значений P под влиянием электрического поля с течением времени перестают удовлетворять этим условиям. Это кажущееся противоречие, однако, устраняется в результате более подробного анализа характера изменения волновых функций под действием поля при правильном выборе периодических граничных условий.

Различие между содержанием настоящего параграфа и содержанием § 6 заключается в том, что здесь мы будем иметь дело с безгранично протяженными волновыми функциями, каждая из которых соответствует одному определенному значению P . Мы будем также учитывать наличие периодических граничных условий и электрических полей. Обобщение этого метода на случай магнитных полей встречает значительные трудности, и мы не будем здесь касаться этого вопроса. Ниже будет показано, что если условия, рассмотренные в § 6, приведены в соответствие с условиями, исследуемыми в настоящем параграфе, то волновые функции ведут себя при этом одинаковым образом.

Рассмотрим модель, сходную с изображенной на фиг. 55 или 58, характеризующуюся периодическими граничными условиями в одном из направлений и граничными условиями „исчезновения“ в двух других направлениях. Предположим, что электрическое поле приложено в том же направлении, в котором имеют место периодические граничные условия, и покажем, что волновые функции изменяются при этом в соответствии с результатами, полученными ранее в § 3 гл. 7, согласно которым поле вызывает смещение значений компоненты квази-импульса P_y , параллельной направлению поля.

Представим себе, что образец кристалла имеет форму тонкой прямоугольной пластинки, изогнутой так, что она образует полый цилиндр. Предположим далее, что периодические граничные условия имеют место в направлении ϕ цилиндрической системы координат. Пусть электрическое поле приложено в этом же направлении, для чего внутрь цилиндра помещен концентрически удлиненный сердечник трансформатора, так что цилиндр образует один короткозамкнутый виток. Если толщина пластинки достаточно мала по сравнению с радиусом цилиндра, то ее изгиб осуществляется упруго и расположение атомов при этом остается идеальным, за исключением того, что решетка несколько искривлена. Если радиус цилиндра r_0 достаточно велик по сравнению с толщиной пластинки δr_0 , то можно показать, что волновое уравнение в этом случае лишь незначительно отличается от волнового уравнения для плоской пластинки, записанного в прямоугольных координатах:

$$x = r - r_0, \quad y = r_0 \phi, \quad z = z. \quad (15.143)$$

Ниже рассматривается предельный случай, когда $\delta r_0/r_0 \rightarrow 0$, так что координаты (15.143) действительно становятся прямоугольными.

Предположим, что электрическое поле изменяется медленно, так что эффектами запаздывания можно пренебречь. Если поток в сердечнике определяется выражением

$$\int B_z dA_z = -c \int V(t) dt, \quad (15.144)$$

где $V(t)$ — разность потенциалов, которая возникла бы на концах разомкнутой вторичной цепи, то поле вне сердечника задано одним только вектор-потенциалом, причем

$$\mathbf{A} = i\phi \frac{c \int V(t) dt}{2\pi r}, \quad \mathbf{E} = -\frac{1}{c} \dot{\mathbf{A}} = i\phi \frac{-V(t)}{2\pi r}, \quad \mathbf{H} = \nabla \times \mathbf{A} = 0, \quad (15.145)$$

где $i\phi$ — единичный вектор направления ϕ .

Оператор кинетической энергии при наличии вектор-потенциала имеет вид

$$\frac{1}{2m} (p^2 - \mathbf{p} \cdot \mathbf{p}_A - \mathbf{p}_A \cdot \mathbf{p} + p_A^2)^2 = \frac{1}{2m} (p^2 - 2\mathbf{p}_A \cdot \mathbf{p} + p_A^2)^2, \quad (15.146)$$

где $\mathbf{p}_A = (-e/c)\mathbf{A}$ — член, отвечающий частице с зарядом $-e$. Поскольку $\nabla \cdot \mathbf{A} = 0$, то $\mathbf{p} \cdot \mathbf{p}_A = \mathbf{p}_A \cdot \mathbf{p}$. В цилиндрических координатах (15.146) имеет вид

$$\begin{aligned} \frac{1}{2m} \left[(-\hbar^2) \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \right) - \right. \\ \left. - 2 \frac{(-e) \int V(t) dt}{2\pi r} \frac{\hbar}{ir} \frac{\partial}{\partial \phi} + p_A^2(r, t) \right], \quad (15.147) \end{aligned}$$

где \mathbf{p} записано как $(\hbar/i)\nabla$. Если мы подставим в это выражение переменные x, y, z , определяемые формулами (15.143), то оно приобретет вид

$$\begin{aligned} \frac{1}{2m} \left[-\hbar^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) - 2 \frac{(-e) \int V(t) dt}{2\pi r_0} \frac{\hbar}{i} \frac{\partial}{\partial y} + p_A^2(r_0, t) \right] + \\ + \frac{1}{2m} \left[-\frac{\hbar^2}{r} \frac{\partial}{\partial x} - \frac{2(-e) \int V(t) dt}{2\pi r_0} \left(\frac{r_0^2}{r^2} - 1 \right) \frac{\hbar}{i} \frac{\partial}{\partial y} + \right. \\ \left. + \left(\frac{r_0^2}{r^2} - 1 \right) p_A^2(r_0, t) \right]. \quad (15.148) \end{aligned}$$

Вторая строка этого выражения представляет собой поправочный член, обусловленный конечностью радиуса кривизны. При этом отношение члена $(\hbar^2/r)\partial/\partial x$ к члену $\hbar^2(\partial^2/\partial x^2)$ первой строки приближенно равно a/r_0 , где a — постоянная решетки. Последние два члена

могут быть сделаны сколь угодно малы, если мы возьмем столь тонкую пластинку, что для расположенных внутри нее точек выполняется условие $(r - r_0) \ll r_0$. В результате оператор кинетической энергии может быть записан в виде

$$\frac{1}{2m} \left[p^2 - 2p_{Ay} p_y + p_{Ay}^2 \right], \quad (15.149)$$

где оператор $(\hbar/i) \nabla$ взят теперь в квази-декартовых координатах x, y, z и

$$p_{Ay} = \frac{(-e) \int V(t) dt}{2\pi r_0} = (-e) \int -E_y(t) dt, \quad (15.150)$$

где

$$E_y(t) = -\frac{V(t)}{2\pi r_0}.$$

Из (15.150) следует, что величина p_{Ay} представляет собой одну из компонент вектора \mathbf{p}_A , рассмотренного в § 6 [формула (15.79)]. Для упруго изогнутой решетки \mathcal{U} можно считать периодической функцией координат x, y, z :

$$\mathcal{U}(x, y, z) = \mathcal{U}(x + n_x a, y + n_y a, z + n_z a),$$

где a — постоянная решетки, n_x, n_y, n_z — целые числа. При таком значении потенциальной энергии волновое уравнение, определяющее волновую функцию $\psi(x, y, z, t)$, приобретает вид

$$\mathcal{H}\psi = \left[\frac{1}{2m} (p^2 - 2p_{Ay} p_y + p_{Ay}^2) + \mathcal{U} \right] \psi = i\hbar \dot{\psi}. \quad (15.151)$$

Удовлетворительное решение этого уравнения может быть найдено на основе решений, полученных при условии $p_{Ay} = 0$. Эти решения соответствуют периодическим граничным условиям в направлении y и граничным условиям „исчезновения“ в направлениях x и z . При любых заданных значениях квантовых чисел для направлений x и z , соответствующих в этом случае величинам $|P_x|$ и $|P_z|$, можно получить решения для различных значений P_y , удовлетворяющие периодическим граничным условиям. Эти решения могут быть записаны в форме

$$\psi_{Py} = e^{iP_y y / \hbar} u_{Py}(x, y, z) \quad (15.152)$$

(здесь опущены члены $|P_x|$ и $|P_z|$, предполагаемые в дальнейшем постоянными). Поскольку ψ_{Py} — собственная функция оператора Гамильтона \mathcal{H}_0 , соответствующего случаю $p_{Ay} = 0$, то

$$\begin{aligned} \mathcal{H}_0 \psi_{Py} &= e^{iP_y y / \hbar} \left[\frac{1}{2m} (p^2 + 2P_y p_y + P_y^2) + \mathcal{U} \right] u_{Py} = \\ &= \mathcal{E}(P_y) \psi_{Py} = e^{iP_y y / \hbar} \mathcal{E}(P_y) u_{Py}. \end{aligned} \quad (15.153)$$

Приближенное решение уравнения (15.151), которое сводится к ψ_{P_y} при $p_{A_y} = 0$, имеет вид

$$\begin{aligned} \psi &= e^{i(P_y y/\hbar)} u_{P_y - p_{A_y}} \exp \int \mathcal{E}(P_y - p_{A_y}) \frac{dt}{\hbar} \equiv \\ &\equiv e^{i(P_y y/\hbar) + i\theta} u_{P_y - p_{A_y}}. \end{aligned} \quad (15.154)$$

Здесь

$$\theta = -\frac{1}{\hbar} \int \mathcal{E}(P_y - p_{A_y}) dt,$$

а P_y — постоянная, не зависящая от t .

Вычислив левую и правую части этого уравнения в отдельности, покажем, что функция (15.154) является приближенным решением уравнения $\mathcal{E}\psi = i\hbar \partial\psi/\partial t$:

$$\begin{aligned} \mathcal{E}\psi &= e^{i(P_y y/\hbar) + i\theta} \left[\frac{1}{2m} (p^2 + 2P_y p_y + P_y^2 - 2p_{A_y} p_y - 2p_{A_y} P_y + p_{A_y}^2) + \mathcal{U} \right] \times \\ &\times u_{P_y - p_{A_y}} = e^{i(P_y y/\hbar) + i\theta} \left\{ \frac{1}{2m} [p^2 + 2(P_y - p_{A_y}) p_y + (P_y - p_{A_y})^2] + \mathcal{U} \right\} \times \\ &\times u_{P_y - p_{A_y}} = e^{i(P_y y/\hbar) + i\theta} \mathcal{E}(P_y - p_{A_y}) u_{P_y - p_{A_y}}, \end{aligned} \quad (15.155)$$

тогда как

$$\begin{aligned} i\hbar \frac{\partial\psi}{\partial t} &= \\ &= e^{i(P_y y/\hbar) + i\theta} \left[-i\hbar p_{A_y} \frac{\partial}{\partial P_y} u_{P_y - p_{A_y}} + \mathcal{E}(P_y - p_{A_y}) u_{P_y - p_{A_y}} \right]. \end{aligned} \quad (15.156)$$

Таким образом, функция (15.154) удовлетворяет волновому уравнению, если не принимать во внимание член, содержащий $\partial u/\partial P_y$. Член такого типа был рассмотрен в § 6. Как было показано, даже в случае электрических полей значительной величины он вносит лишь незначительную поправку. Функцию (15.154) можно поэтому рассматривать, как удовлетворяющее решение уравнения.

Исследуем смысл этого решения. Несмотря на то, что в показателе экспоненциального множителя функции (15.154) сохраняется одно и то же значение P_y , так что при движении электрона в кристалле комплексная фаза все время удовлетворяет периодическим граничным условиям, энергия и скорость изменяются как раз так, как если бы P_y возрастало со скоростью $-eE_y$ (см. § 2 гл. 7). Для энергии этот результат следует непосредственно из уравнения (15.155), которое показывает, что

$$\int \psi^* \mathcal{E}\psi dx = \mathcal{E}(P_y - p_{A_y}) = \mathcal{E}\left(P_y - e \int E_y dt\right). \quad (15.157)$$

Воспользовавшись формулой, определяющей групповую скорость v_y , получаем

$$v_y = \frac{\partial}{\partial P_y} \mathcal{E}(P_y - p_{Ay}) = v_y(P_y - p_{Ay}), \quad (15.158)$$

где $v_y(P_y) = \partial \mathcal{E}(P_y) / \partial P_y$ — значение скорости при условии $p_A = 0$. [Точно такой же результат может быть получен и в том случае, если при наличии вектор-потенциала мы для определения скорости воспользуемся оператором $(p_y - p_{Ay})/m$, см. формулу (15.115).] Таким образом, хотя поле и не изменяет разности фаз волновой функции при переходе от ячейки к ячейке, энергия и скорость при этом меняются как раз так, как если бы P_y изменялось со скоростью $-\dot{p}_{Ay} = -eE_y$, так что хотя представления, изложенные в § 2 гл. 7, не являются аналитически точными, они приводят, тем не менее, к правильным результатам.

Иначе говоря, можно сказать, что при наличии поля квантовое число P_y остается постоянным, но энергия и скорость в различные моменты времени являются при этом различными функциями P_y , причем изменение этих функций соответствует смещению поверхности постоянной энергии и диаграммы скоростей со скоростью $+eE_y$ по отношению к значениям P_y .

Сопоставим теперь волновую функцию (15.154) с волновой функцией, описывающей волновой пакет, исследованный в § 6. Если электрическое поле однородно и задано одним только вектор-потенциалом, то уравнение для P , приведенное в § 7, легко может быть проинтегрировано. Имеем

$$\dot{P} = -eE = (-e) \left(-\frac{1}{c} \frac{\partial}{\partial t} A \right), \quad (15.159)$$

или

$$P = (-e) \left(-\frac{1}{c} A \right) + P_0 = P_0 - p_A, \quad (15.160)$$

где P_0 — постоянная интегрирования. Пространственный множитель волновой функции, описывающей волновой пакет [в соответствии с формулой (15.84) для ψ и формулой (15.108) для P , приведенными в § 6], приобретает при этом вид

$$\psi = w \left[\exp \left(\frac{i p_A \cdot r}{\hbar} \right) \right] \psi_{P_0 - p_A} \quad (15.161)$$

(поскольку мы предполагаем, что E однородно, p_A не зависит от координат и $p_{A0} = p_A$). Функция (15.161) может быть также представлена в виде

$$\begin{aligned} \psi &= w \left[\exp \left(\frac{i p_A \cdot r}{\hbar} \right) \right] \left\{ \exp \frac{i (P_0 - p_A) \cdot r}{\hbar} \right\} u_{P_0 - p_A} = \\ &= w \left[\exp \left(\frac{i P_0 \cdot r}{\hbar} \right) \right] u_{P_0 - p_A}, \quad (15.162) \end{aligned}$$

что эквивалентно функции (15.154), рассмотренной выше. Таким образом, хотя в том случае, когда поле определяется вектор-потенциалом \mathbf{A} , \mathbf{P} изменяется (см. § 6), комплексный фазовый множитель волновой функции, на который умножается периодический множитель u , при этом остается неизменным. Если, однако, электрическое поле определяется также и электростатическим потенциалом ϕ , то при этом изменяется и фазовый множитель волновой функции. При наличии периодических граничных условий, рассматриваемых в настоящем параграфе, поля, обусловленные потенциалом ϕ , исключаются из рассмотрения, так как ϕ должно быть определено однозначно. Для случая, исследованного в § 6, переход от постоянного значения P_0 к переменному P_0 может быть осуществлен в результате изменения значений ϕ , \mathbf{A} и ψ путем градиентного преобразования; физические условия, описываемые волновой функцией, при таком преобразовании остаются неизменными. Таким образом, пространственный множитель волновой функции, описывающей движение волнового пакета в однородном электрическом поле, отличается от безгранично протяженной волновой функции, исследованной в настоящем параграфе, только наличием множителя w .

Упомянем в заключение, что можно представить себе модель кристалла рассмотренного здесь типа, для которой условия периодичности выполняются также и в направлении z .

Для этого решетку кристалла следует продолжить в направлении z на расстояние z_0 , значительно превышающее r_0 . При этом мы получим коаксиальную систему, состоящую из сердечника трансформатора в центре и цилиндрического слоя проводника. Эта система может быть затем изогнута в тороид, так что поверхность кристалла приобретает при этом тороидальную форму. Если $z_0 \gg r_0$, то локальным влиянием искривления кристалла снова можно пренебречь. Сердечник, как и прежде, создает электрическое поле в направлении y . Изменяя величину потока, проходящего через отверстие тороида, можно создать новое поле в направлении z .

ЛИТЕРАТУРА

1. Jones H., Zener C., Proc. Roy. Soc., 144A, 101 (1934).
2. Dushman S., Elements of Quantum Mechanics, New York, 1938.
3. Schiff L. I., Quantum Mechanics, New York, 1949.
4. Sommerfeld A., Bethe G., Handb. d. Phys. XXIV/2 (см. перевод: Зоммерфельд А., Бете Г., Электронная теория металлов, М.—Л., 1938).
5. Fröhlich H., Theorie der Metalle, Berlin, 1937.
6. Seitz F., The Modern Theory of Solids, New York, 1940 (см. перевод: Зейтц Ф., Современная теория твердого тела, М.—Л., 1949).
7. Mott N. F., Jones H., Theory of Metals and Alloys, Oxford, 1936.
8. Slater J. C., Rev. Mod. Phys., 6, 209 (1934).
9. Wilson A. H., The Theory of Metals, Cambridge, 1935 (см. перевод: Вильсон А., Квантовая теория металлов, М.—Л., 1941).
10. Houston W. V., Phys. Rev., 57, 184 (1940).

11. Zener C., Proc. Roy. Soc., **145A**, 523 (1934).
12. Фок В., Zs. f. Phys., **61**, 126 (1930).
13. Slater J. C., Phys. Rev., **35**, 210 (1930).
14. Френкель Я. И., Волновая механика, ч. II, М.—Л., 1934.
15. Коортманс Т., Physica, **1**, 104 (1933).
16. Madelung E., Die Mathematischen Hilfsmittel der Physikers, New York, 1943.
17. Courant P., Hilbert D., Methoden der Mathematische Physik, New York, 1947 (см. перевод: Курант Р., Гильберт Д., Методы математической физики, М.—Л., 1949).
- 18*. Блохинцев Д. И., Основы квантовой механики, 2-е изд., М.—Л., 1951.
- 19*. Ландау Л. Д., Лифшиц Е. М., Теория поля, 2-е изд., М.—Л.; 1948.
- 20*. Иваненко Д. Д., Соколов А. А., Классическая теория поля, 2-е изд., М.—Л., 1951.
- 21*. Френкель Я. И., Волновая механика, ч. 1 и 2, М.—Л., 1934—1935.

СТАТИСТИЧЕСКАЯ МЕХАНИКА В ПРИЛОЖЕНИИ
К ПОЛУПРОВОДНИКАМ§ 1. ВЫВОД ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ДЛЯ ПРОСТЫХ
КВАНТОВАННЫХ СИСТЕМ

Основные понятия ¹⁾. В этом разделе мы дадим краткий обзор приложений статистической механики к некоторым квантованным системам. При этом мы выведем формулу Планка, характеризующую распределение энергий гармонических осцилляторов, и получим закон распределения Ферми—Дирака. Будет также кратко затронут вопрос о связи статистической механики с термодинамикой (этот вопрос рассматривается подробнее в форме задач в конце главы).

В статистической механике квантованной системы понятия *системы*, ее *квантовых состояний*, *распределений* и *априорной вероятности* играют основную роль. Рассмотрим эти понятия поочередно.

Под системой мы будем понимать некоторое количество электронов и ядер, находящихся при определенных граничных условиях. Например, под системой можно было бы подразумевать $4,52 \cdot 10^{22}$ атомов германия, содержащихся в образце объемом 1 см^3 . Обычно в качестве достаточно хорошего приближения можно исследовать свойства некоторой части какой-либо реальной системы и трактовать эту часть так, как если бы она сама по себе являлась системой. Один из примеров такого типа был рассмотрен в § 2 гл. 10, где речь шла о 20 электронах и 36 квантовых состояниях. Другим примером может служить система, состоящая из Q гармонических осцилляторов. Нормальные колебания атомов в кристалле, исследуемые в § 3 гл. 17, являются системой этого типа. Полупроводник содержит системы обоого типа, образующие в совокупности единую систему.

Со статистической точки зрения наиболее существенной характеристикой системы является энергетическая схема ее квантовых состояний, другими словами — полный перечень ее квантовых состояний и соответствующих им значений энергии. Этот перечень может быть также дан в виде схемы энергетических уровней системы при обязательном учете того обстоятельства, что несколько квантовых состояний могут относиться к одному и тому же энергетическому уровню. При этом редко оказывается возможным иметь дело с точной схемой энергетических уровней системы в целом; энергия системы

¹⁾ Вопросы, рассматриваемые в этом, а также в следующих разделах настоящего параграфа, подробно изложены в ряде книг советских авторов, см. [8—12, 14]. — *Прим. ред.*

обычно рассматривается как слагающаяся из энергий ее различных частей, каждая из которых находится в своем собственном квантовом состоянии.

Таким приближением часто приходится пользоваться, и, как показано в гл. 13, оно дает возможность наиболее глубокого анализа разнообразных вопросов. Так, раздельное статистическое исследование движения электронов и колебаний решетки, проводимое в гл. 17, дает возможность рассматривать неучтенное взаимодействие между ними как причину, обуславливающую переходы. Трудно представить себе, как следовало бы формулировать теорию электрического сопротивления, оперируя при этом одними лишь точными квантовыми состояниями системы в целом; такой метод решения задачи почти наверняка оказался бы значительно более сложным. Подобная постановка задачи может, однако, оказаться необходимой для объяснения явления сверхпроводимости¹⁾. В § 2 гл. 5 и в конце § 7 гл. 15 рассмотрены также и другие примеры подразделения системы на отдельные почти не зависящие друг от друга части.

Приближение, при котором квантовые состояния системы в целом рассматриваются как совокупность квантовых состояний подсистем, приводит к очень незначительной ошибке, если неучтенное взаимодействие подсистем невелико. Этот результат, который мы примем здесь как некоторое дополнительное предположение, может быть установлен с помощью квантовомеханических теорем, касающихся инвариантности следа эрмитовой матрицы при унитарном преобразовании волновых функций.

Коль скоро система предполагается подразделенной на ряд подсистем, приобретает смысл понятие *распределения*²⁾. Это понятие мы проиллюстрируем прежде всего на примере системы электронов, подчиняющихся принципу Паули (см. § 2 гл. 5, § 7 гл. 15, а также гл. 9 и 10). Энергию такой системы можно считать равной сумме энергий квантовых состояний, занимаемых электронами. Квантовые состояния мы будем классифицировать по соответствующим им значениям энергии и подразделять их на группы состояний, обладающих примерно одинаковой энергией. При такой классификации распределение будет характеризоваться заданием числа занятых состояний каждой группы. Каждому данному распределению обычно соответствует ряд квантовых состояний системы в целом. Например, если одна из групп состояний содержит два состояния, а распределение таково, что только одно из них должно быть занято, то возможны два способа выбора занятого состояния. В общем случае число таких

¹⁾ Мы имеем здесь в виду современные теории сверхпроводимости Фрелиха [1] и Бардина [2], которые показали, что при определенных условиях нижние энергетические состояния сложной системы, состоящей из электронов и атомов, могут оказаться сверхпроводящими. — *Прим. авт.*

²⁾ Понятие распределения имеет смысл также и тогда, когда, кроме данной системы, рассматриваются еще и другие системы. — *Прим. ред.*

способов для каждой из групп значительно больше, так что число способов осуществления некоторого данного распределения обычно очень велико, причем каждый способ соответствует одному из квантовых состояний системы.

Основной задачей статистической теории является отыскание такого распределения, которое для данной системы с данной энергией может быть достигнуто наибольшим числом способов. Решение этой задачи существенно потому, что это распределение характеризует свойства системы в равновесных условиях. Чтобы показать, что распределение, осуществляемое наибольшим числом способов, характеризует тепловое равновесие, необходимо сделать ряд дополнительных предположений. Первое предположение — о равенстве значений априорных вероятностей для каждого из квантовых состояний системы — заключается в следующем: если исследуемая система подготовлена в лаборатории так, что обладает данной определенной энергией, затем изолирована и оставлена в некотором определенном квантовом состоянии, то в последующие моменты времени ее нахождение в любом из возможных квантовых состояний с такой же энергией равновероятно; другими словами, все квантовые состояния системы обладают одной и той же априорной вероятностью при условии, что энергии этих состояний одинаковы.

Предположение о равенстве значений априорных вероятностей представляется вполне естественным. Оно может быть также обосновано с помощью квантовомеханической *H*-теоремы. Доказательство *H*-теоремы и исследование других методов обоснования статистической механики квантованных систем принадлежат, однако, к числу весьма сложных вопросов, и мы не будем здесь вдаваться в их более подробный анализ.

Возможные квантовые состояния с данной энергией могут, конечно, группироваться различными способами, образуя различные распределения; при этом каждое квантовое состояние будет входить не более чем в одно распределение. Среди всех возможных распределений встретится одно такое, которое будет обладать большим числом квантовых состояний, нежели любое другое распределение. Будут встречаться также и распределения, лишь слегка от него отличные, которые могут быть осуществлены примерно таким же числом способов. Можно показать, что в случае систем, содержащих большие числа частиц, имеет смысл рассматривать только наиболее вероятное распределение. Это связано с тем обстоятельством, что распределения, заметно отличающиеся от наиболее вероятного, содержат значительно меньшее число квантовых состояний. Действительно, если бы мы объединили все такие распределения в одно, то общее число квантовых состояний оказалось бы при этом ничтожно малым по сравнению с числом квантовых состояний, отвечающих наиболее вероятному распределению. Это значит, что вероятность нахождения системы в любом из состояний, отличных от тех, которые принадлежат к наиболее

вероятному распределению или же к распределениям, очень к нему близким, ничтожно мала.

Для уяснения механизма возникновения наиболее вероятного распределения весьма полезным является понятие *микроканонического ансамбля*. Представим себе очень большое число одинаковых систем, каждая из которых занимает одно из квантовых состояний и обладает заданным значением полной энергии. Чтобы обойти трудности, связанные с дискретной природой энергетических уровней, обычно предполагается, что энергия каждой такой системы заключена в малом интервале между \mathcal{E}_1 и \mathcal{E}_2 , причем интервал ($\mathcal{E}_2 - \mathcal{E}_1$) составляет лишь очень малую долю величины \mathcal{E}_1 . Это большое число, или *ансамбль* систем и есть так называемый *микроканонический ансамбль*. Все системы ансамбля, за исключением лишь очень незначительной их части, находятся в квантовых состояниях, принадлежащих к наиболее вероятному распределению. Наблюдаемое значение любой физической величины для некоторой реальной системы в среднем совпадает со значением этой величины, получаемым при ее усреднении по всем системам микроканонического ансамбля. Этот результат является непосредственным следствием предположения о равенстве априорных вероятностей. Благодаря доминирующей роли наиболее вероятного распределения среднее значение любой интересующей нас величины для микроканонического ансамбля совпадает с ее средним значением для наиболее вероятного распределения. В случае системы малого размера, такой, например, как отдельная коллоидная частица, отклонения от средних значений, соответствующих наиболее вероятному распределению, могут, однако, стать существенными. Эти отклонения могут быть вычислены с помощью теории флуктуаций; один из примеров таких вычислений будет рассмотрен нами в качестве задачи.

В дальнейшем мы будем поступать следующим образом: задавать различные возможные распределения и отыскивать затем наиболее вероятное из них. Свойства системы в состоянии равновесия будут соответствовать этому распределению. Такой метод решения задачи несколько произволен постольку, поскольку произвольным является выбор исходных распределений. На исследуемых ниже примерах мы убедимся, однако, в том, что условия, используемые при задании исходных распределений, на конечной стадии решения задачи выпадают из рассмотрения, поэтому вопрос о выборе этих условий не должен нас беспокоить.

Ниже мы исследуем систему, состоящую из независимых друг от друга гармонических осцилляторов, в предположении, что все они обладают одной и той же частотой колебания, и найдем наиболее вероятное распределение для такой системы. Число способов, которыми может быть осуществлено наиболее вероятное распределение, мы будем обозначать в дальнейшем через $W(\mathcal{E})$, где \mathcal{E} — энергия системы ¹⁾.

¹⁾ Число способов, которыми может быть осуществлено данное распределение, называется также *термодинамической вероятностью данного распределения*. — *Прим. ред.*

Для того чтобы иметь возможность определить понятие температуры на основе статистической теории, а также установить связь между статистической механикой и термодинамикой, мы исследуем далее случай двух систем, находящихся в равновесии друг с другом. Если эти системы обладают соответственно энергиями \mathcal{E}_1 и \mathcal{E}_2 , причем $\mathcal{E}_1 + \mathcal{E}_2 = \mathcal{E}_T$, где \mathcal{E}_T — полная энергия, то число способов, которыми величина \mathcal{E}_T может быть подразделена на \mathcal{E}_1 и \mathcal{E}_2 , равно $W_1(\mathcal{E}_1)W_2(\mathcal{E}_2)$. Мы покажем, что температура системы определяется при этом соотношением

$$\frac{d \ln W_1(\mathcal{E})}{d\mathcal{E}} = \frac{1}{kT_1}, \quad (16.1)$$

где k — постоянная Больцмана, а энтропия равна ¹⁾

$$\mathcal{S}_1 = k \ln W_1(\mathcal{E}_1). \quad (16.2)$$

Основываясь на этих результатах, мы выведем в дальнейшем закон распределения Ферми — Дирака.

Распределение гармонических осцилляторов по Планку. Рассмотрим систему, состоящую из большого числа Q гармонических осцилляторов, каждый из которых обладает одной и той же частотой ν и энергетическими уровнями $(n + 1/2)h\nu$, где $n = 0, 1, 2, \dots$ — квантовое число, а h — постоянная Планка. Квантовые состояния отдельных осцилляторов будем называть *O-состояниями*, чтобы отличить их от состояний системы в целом, которые будем называть *S-состояниями*. Задача будет заключаться в отыскании закона распределения осцилляторов, т. е. в определении ряда чисел $Q_0, Q_1, Q_2, \dots, Q_n, \dots$, представляющих собой числа осцилляторов в состояниях, характеризующихся соответственно квантовыми числами $0, 1 \dots$. Очевидно, что

$$Q = \sum_n Q_n. \quad (16.3)$$

Если полная энергия системы равна \mathcal{E} , то

$$\mathcal{E} = \sum_n \left(n + \frac{1}{2}\right) h\nu Q_n = Q \frac{h\nu}{2} + h\nu \sum_n n Q_n. \quad (16.4)$$

Квантовое состояние системы, или *S-состояние*, обычно описывают, указывая, какие осцилляторы находятся в *O-состояниях* с квантовыми числами $0, 1, 2 \dots$. В частности, *S-состояние* может быть, например, определено следующим образом:

$$\left(\frac{0}{abcde}\right) \left(\frac{1}{fgh}\right) \dots \left(\frac{s}{ijk\dots}\right) \dots \quad (16.5)$$

¹⁾ Температура T_2 и энтропия \mathcal{S}_2 системы с энергией \mathcal{E}_2 определяются аналогичными формулами; в состоянии равновесия $T_1 = T_2$. — *Прим. ред.*

Формула (16.5) означает, что осцилляторы a, b, c, d , и e находятся в O -состоянии с $n = 0$, осцилляторы f, g, h — в O -состоянии с $n = 1$ и т. д. В дальнейшем выражение вида (16.5) мы будем называть „записью“ состояния ¹⁾. Этому же состоянию могут отвечать, конечно, и другие „записи“, например

$$\left(\frac{0}{baedc}\right)\left(\frac{1}{ghf}\right)\dots\left(\frac{s}{kji\dots}\right)\dots \quad (16.6)$$

Наша задача будет заключаться в том, чтобы выяснить, сколько S -состояний соответствует некоторому данному распределению, например распределению, при котором $Q_0 = 5, Q_1 = 3 \dots$. Различные „записи“ одного и того же S -состояния при этом не следует рассматривать как различные S -состояния. Покажем прежде всего, что число „записей“ данного S -состояния равно

$$Q_0! Q_1! \dots \equiv \prod_n Q_n! \quad (16.7)$$

Рассмотрим с этой целью любую „запись“; выражение в круглой скобке, отвечающей $n = 0$, может быть записано $Q_0!$ различными способами (так, в приведенном выше примере имеется 5 возможностей выбора первой буквы, 4 возможных варианта выбора второй буквы и т. д.). Аналогичным образом, каждая круглая скобка может быть записана $Q_n!$ способами, так что каждое S -состояние может быть осуществлено в общей сложности $\prod_n Q_n!$ способами. Выясним теперь, каково число неодинаковых „записей“, отвечающих данному распределению. В любой данной „записи“ буквы, относящиеся к отдельным осцилляторам, расположены в определенной последовательности и существует, очевидно, $Q!$ способов их перестановки. Следовательно, всего имеется $Q!$ „записей“. Поскольку каждое S -состояние учитывается при этом $\prod_n Q_n!$ раз, распределение может быть осуществлено

$$W(Q_0, Q_1, \dots) = \frac{Q!}{\prod_n Q_n!} \quad (16.8)$$

способами, т. е. существует $Q!/\prod_n Q_n!$ различных квантовых состояний, соответствующих данному распределению.

Какое из возможных распределений наиболее вероятно, т. е. обладает наибольшим значением W при условии постоянства общего числа осцилляторов Q и полной энергии системы ξ ? Мы будем искать это распределение, исходя из требования, чтобы при любом

¹⁾ Термин „запись“ („writing“) не является общепринятым. Определенный способ осуществления данного состояния иногда называют „комплексией“. См., например, [13]. — *Прим. ред.*

малом изменении чисел Q_n величина W оставалась постоянной. Это условие само по себе, конечно, еще не означает, что W имеет максимум; в равной мере W могло бы при этом иметь также и минимум. Путем дополнительных вычислений можно, однако, показать, что полученное при этом распределение отвечает максимальному значению W (см. задачи). Рассмотрим теперь некоторое возмущение P , действующее на распределение и определяемое условиями

$$P \begin{cases} Q_r \rightarrow Q_r + \delta Q_r, & (16.9a) \\ Q_s \rightarrow Q_s + \delta Q_s, & \delta Q_s = -\delta Q_r, & (16.9b) \\ \mathcal{E} \rightarrow \mathcal{E} + \delta \mathcal{E}, & \delta \mathcal{E} = h\nu(r-s)\delta Q_r. & (16.9b) \end{cases}$$

Это возмущение, как мы видим, соответствует переходу δQ_r осцилляторов из O -состояния r в O -состояние s . Для удобства будем в дальнейшем исследовать не величину W , а ее логарифм и обозначим изменение $\ln W$, вызванное возмущением, через $\delta \ln W$. Рассматриваемое возмущение должно было бы нарушить постоянство полной энергии системы \mathcal{E} , так что следует ввести в рассмотрение дополнительное возмущение, восстанавливающее исходное значение энергии. Предположим, что второе возмущение (будем обозначать его P') вызывает переход δQ_v осцилляторов из O -состояния u в O -состояние v . Ниже мы покажем, что при этом

$$\frac{\delta \ln W}{\delta \mathcal{E}} = \frac{(\delta \ln W)'}{\delta \mathcal{E}'}, \quad (16.10)$$

т. е. что изменение логарифма W , соответствующее изменению энергии на единицу, одинаково для обоих возмущений; поскольку состояния u и v могут быть выбраны произвольным образом, это значит, что любому возмущению должно отвечать одно и то же значение $(\delta \ln W)' / \delta \mathcal{E}'$. Основываясь на этом результате, мы с легкостью получим закон распределения Планка, а в дальнейшем также и закон распределения Ферми — Дирака.

Равенство (16.10) доказывается очень просто. Если W имеет максимум, то $\ln W$ также имеет максимум. Возмущения P и P' вызывают следующие изменения чисел осцилляторов в состояниях r , s , u и v :

$$\delta Q_r = -\delta Q_s = \frac{\delta \mathcal{E}}{h\nu(r-s)}, \quad (16.11a)$$

$$\delta Q_u = -\delta Q_v = \frac{\delta \mathcal{E}'}{h\nu(u-v)}. \quad (16.11b)$$

Эти изменения, в свою очередь, обуславливают изменения $\ln W$, равные $\delta \ln W$ и $(\delta \ln W)'$. Если $\delta \mathcal{E} = -\delta \mathcal{E}'$, то полная энергия системы \mathcal{E} при этом остается постоянной, и если W имеет максимум,

то не должен изменяться также и $\ln W$. Изменение $\ln W$ пропорционально $\delta \mathcal{E}$ и $\delta \mathcal{E}'$, так что

$$\delta \ln W = \frac{\delta \ln W}{\delta \mathcal{E}} \delta \mathcal{E} + \frac{(\delta \ln W)'}{\delta \mathcal{E}'} \delta \mathcal{E}'. \quad (16.12)$$

Для того чтобы это выражение обращалось в нуль при $\delta \mathcal{E} = -\delta \mathcal{E}'$, коэффициенты при $\delta \mathcal{E}$ и $\delta \mathcal{E}'$ должны быть равны между собой, так что можно положить

$$\beta = \frac{\delta \ln W}{\delta \mathcal{E}} = \frac{(\delta \ln W)'}{\delta \mathcal{E}'}. \quad (16.13)$$

В дальнейшем мы покажем, что $\beta = 1/kT$, где k — постоянная Больцмана, а T — абсолютная температура.

Величину $\delta \ln W / \delta \mathcal{E}$ легко выразить через параметры распределения Q_n , воспользовавшись с этой целью формулой, определяющей $\ln W$, и вычислив производную от логарифма факториала. Последняя может быть определена как значение $\delta \ln x! / \delta x$ при наименьшем изменении δx , равном 1; это дает

$$\frac{d \ln x!}{dx} \approx \frac{\ln(x + \delta x)! - \ln x!}{\delta x} = \ln(x + 1) \approx \ln x, \quad (16.14a)$$

$$\delta \ln x! \approx \delta x \ln x. \quad (16.14b)$$

Отсюда, используя выражения (16.11a) и (16.8), получаем

$$\begin{aligned} \delta \ln W &= -\delta (\ln Q_r! Q_s!) = -\delta Q_r \ln Q_r - \delta Q_s \ln Q_s = \\ &= \frac{\delta \mathcal{E} \ln(Q_s/Q_r)}{h\nu(r-s)}. \end{aligned} \quad (16.15)$$

Если представить себе теперь, что квантовые числа u и v остаются постоянными, а квантовые числа r и s могут принимать все возможные значения, то из (16.13) следует, что

$$\beta = \frac{\delta \ln W}{\delta \mathcal{E}} = \frac{\ln(Q_s/Q_r)}{h\nu(r-s)}. \quad (16.16)$$

Отсюда

$$\beta h\nu r + \ln Q_r = \beta h\nu s + \ln Q_s. \quad (16.17)$$

Если $s = 0$, то

$$Q_r = Q_0 e^{-\beta h\nu r} = Q_0 e^{-\gamma r}, \quad (16.18)$$

$$\gamma = \beta h\nu. \quad (16.19)$$

Величины β и Q_0 определяются из условий, которым подчиняются Q и \mathcal{E} :

$$Q = \sum Q_n = Q_0(1 + e^{-\gamma} + e^{-2\gamma} + \dots) = \frac{Q_0}{1 - e^{-\gamma}}, \quad (16.20)$$

$$\begin{aligned} \mathcal{E} &= \frac{h\nu Q}{2} + h\nu \sum nQ_n = \\ &= \frac{h\nu Q}{2} + h\nu Q_0(0 \cdot 1 + e^{-\gamma} + 2e^{-2\gamma} + 3e^{-3\gamma} + \dots) = \\ &= \frac{h\nu Q}{2} + h\nu Q_0 \left(-\frac{d}{d\gamma} \right) (1 + e^{-\gamma} + e^{-2\gamma} + \dots) = \\ &= \frac{h\nu Q}{2} - h\nu Q_0 \frac{d}{d\gamma} \frac{1}{1 - e^{-\gamma}} = \\ &= \frac{h\nu Q}{2} + h\nu Q_0 \frac{e^{-\gamma}}{(1 - e^{-\gamma})^2} = h\nu Q \left[\frac{1}{2} + \frac{1}{e^{\gamma} - 1} \right]. \end{aligned} \quad (16.21)$$

Эти уравнения связывают Q и \mathcal{E} с Q_0 и γ ; в принципе они могут быть решены относительно γ и Q_0 , так что последние в результате будут выражены через Q и \mathcal{E} . Ниже будет показано, что величина $\gamma = \beta h\nu$ может быть отождествлена с $h\nu/kT$; это приводит к следующим значениям средней энергии осциллятора:

$$\langle \mathcal{E} \rangle = h\nu \left[\frac{1}{2} + \frac{1}{e^{h\nu/kT} - 1} \right] \quad (16.22)$$

и среднего значения квантового числа n :

$$\langle n \rangle = \frac{1}{e^{h\nu/kT} - 1}. \quad (16.23)$$

Вероятность того, что любой из осцилляторов находится в состоянии n , равна Q_n/Q и определяется выражением

$$\frac{Q_n}{Q} = e^{-n h\nu/kT} (1 - e^{-h\nu/kT}). \quad (16.24)$$

Совокупность этих соотношений и дает закон распределения Планка для системы гармонических осцилляторов.

Исследуя изменения второго порядка, вызываемые возмущениями δQ_r , δQ_s и т. д., можно показать, что $\ln W$ действительно имеет максимум и что значение $\ln W$ очень быстро убывает при отклонении распределения от наиболее вероятного. Численное значение $\ln W$ для наиболее вероятного распределения в общем случае очень велико. Это обстоятельство приводит к неожиданному и важному результату: численные значения величины $\ln W$ и логарифма общего числа квантовых состояний с энергией \mathcal{E} весьма незначительно отличаются друг от друга. Это удивительно потому, что число квантовых состояний, соответствующих наиболее вероятному распределению, конечно,

составляет лишь небольшую долю общего числа всех возможных квантовых состояний с энергией \mathcal{E} . Этот результат можно проиллюстрировать численным примером. Значение $\ln W$ для 1 грамм-атома осцилляторов при $\gamma \approx 1$ равно примерно 10^{23} . (Эта цифра может быть получена с помощью выражения, определяющего $\ln W$, если при этом воспользоваться приближенной формулой Стирлинга $x! = (x/e)^x \sqrt{2\pi x}$, а также более простым путем — на основании формул, приводимых в задачах к данной главе.) Предположим, что только одно из каж-дых 10^{20} состояний соответствует наиболее вероятному распределению. В таком случае общее число состояний равно $10^{20} \exp(10^{23})$ и логарифм этого числа равен $10^{23} + 46$. Очевидно, что числом 46 здесь полностью можно пренебречь. Ввиду больших численных значений отдельных членов выражения, определяющего $\ln W$, часто оказывается возможным ограничиться весьма приближенными вычислениями, определяя величину W лишь с точностью до больших множителей.

Связь с термодинамикой. Мы показали, что любые возмущения системы гармонических осцилляторов должны приводить к одному и тому же значению величины

$$\frac{\delta \ln W}{\delta \mathcal{E}} = \beta. \quad (16.25)$$

Основываясь на этом результате, докажем теперь, что если $W(\mathcal{E})$ представляет собой число способов, которыми может быть осуществлено наиболее вероятное распределение, то

$$\frac{d \ln W(\mathcal{E})}{d \mathcal{E}} = \beta. \quad (16.26)$$

С этой целью исследуем наиболее вероятные распределения при двух значениях полной энергии, равных \mathcal{E} и $\mathcal{E} + \Delta \mathcal{E}$. От одного из этих распределений можно, очевидно, перейти к другому путем ряда последовательных переходов типа $(Q_r, Q_s) \rightarrow (Q_r + \delta Q_r, Q_s + \delta Q_s)$, рассмотренных выше. Эти переходы будут сопровождаться изменениями энергии системы $\delta \mathcal{E}, \delta \mathcal{E}', \delta \mathcal{E}'' \dots$; изменения $\ln W$ будут при этом равны $\beta \delta \mathcal{E}, \beta \delta \mathcal{E}', \beta \delta \mathcal{E}''$ и т. д. Полное изменение $\ln W$ будет равняться

$$\beta (\delta \mathcal{E} + \delta \mathcal{E}' + \delta \mathcal{E}'' + \dots) = \beta \Delta \mathcal{E},$$

где $\Delta \mathcal{E}$ — полное изменение энергии системы. Таким образом, при переходе от наиболее вероятного распределения при энергии \mathcal{E} к наиболее вероятному распределению при энергии $\mathcal{E} + \Delta \mathcal{E}$ имеем

$$\frac{\Delta \ln W}{\Delta \mathcal{E}} = \beta.$$

Поскольку наиболее вероятное распределение является функцией \mathcal{E} , это отношение представляет собой искомую производную от $\ln W$ по \mathcal{E} .

Предположим теперь, что имеются две системы, 1 и 2, которые могут обмениваться друг с другом энергией. Их можно рассматривать как одну сложную систему, полная энергия которой $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$. Возможность обмена энергией между системами соответствует приведению двух тел в тепловой контакт, когда между ними может иметь место теплообмен и они могут прийти в состояние теплового равновесия, при котором их температуры одинаковы и полная энтропия имеет максимальное значение.

Вследствие наличия взаимодействия ни система 1, ни система 2 не обладают при этом теми энергетическими уровнями, которыми они обладали бы, если бы были изолированы друг от друга. В действительности в этом случае существуют только энергетические уровни системы в целом. Однако если взаимодействие между системами очень мало, то энергия сложной системы лишь очень незначительно отличается от суммы энергий двух изолированных систем. Поэтому естественно рассматривать такую сложную систему как совокупность двух отдельных систем и представлять себе в то же время, что системы 1 и 2 могут взаимодействовать друг с другом в том смысле, что между ними имеет место теплообмен. Эти соображения, однако, недостаточно строги; строгое доказательство того, что поведение двух слабо взаимодействующих систем в термодинамическом отношении вполне эквивалентно поведению двух тел, находящихся в тепловом контакте друг с другом, представляет трудности. Для более подробного ознакомления с этими вопросами читателю рекомендуется обратиться к специальной литературе [3, 4].

При любом данном подразделении полной энергии на части \mathcal{E}_1 и \mathcal{E}_2 наиболее вероятные распределения могут быть осуществлены $W_1(\mathcal{E}_1)$ способами для системы 1 и $W_2(\mathcal{E}_2)$ способами для системы 2. Действительно, как это было показано в конце предыдущего параграфа, при вычислении величины $W_1(\mathcal{E}_1)$ в рассмотрение могут быть включены все квантовые состояния с энергией \mathcal{E}_1 . По тем же соображениям, если это желательно, можно иметь также дело с общим числом состояний системы 1, заключенных в малом интервале значений энергии между \mathcal{E}_1 и $\mathcal{E}_1 + \delta\mathcal{E}_1$. Как в том, так и в другом случае благодаря большому численному значению величины W численное значение $\ln W$ окажется одним и тем же. Логарифм общего числа возможных способов подразделения энергии \mathcal{E} на части \mathcal{E}_1 и \mathcal{E}_2 можно, таким образом, положить равным $\ln W_1(\mathcal{E}_1) W_2(\mathcal{E}_2)$; наиболее вероятным подразделением энергии при этом будет такое, при котором эта величина имеет максимум. (При этом ввиду независимости величины $\ln W$ от точного определения вероятности W нет даже необходимости в том, чтобы выполнялось условие $\mathcal{E}_1 + \mathcal{E}_2 = \mathcal{E}$; достаточно, чтобы $\mathcal{E}_1 + \mathcal{E}_2$ лежало в интервале $\delta\mathcal{E}$ вблизи \mathcal{E} .) Максимум значения $\ln W_1(\mathcal{E}_1) W_2(\mathcal{E}_2)$ соответствует условию

$$\frac{d}{d\mathcal{E}_1} \ln W(\mathcal{E}_1, \mathcal{E}_2) = 0 = \frac{d \ln W_1(\mathcal{E}_1)}{d\mathcal{E}_1} - \frac{d \ln W_2(\mathcal{E}_2)}{d\mathcal{E}_2}. \quad (16.27)$$

Последнее из этих уравнений является следствием того, что $\mathcal{E}_2 = \mathcal{E} - \mathcal{E}_1$. Таким образом, в состоянии равновесия величины

$$\beta_1 = \frac{d \ln W_1}{d \mathcal{E}_1} \quad \text{и} \quad \beta_2 = \frac{d \ln W_2}{d \mathcal{E}_2} \quad (16.28)$$

должны равняться друг другу, так как $\ln W_1 + \ln W_2$ имеет в этом случае максимум. Можно показать, что эти условия эквивалентны отождествлению энтропии \mathcal{S} с величиной $k \ln W$ и равенству $\beta = 1/kT$ (см. задачи). Равенство величин β_1 и β_2 соответствует, таким образом, равенству температур T_1 и T_2 и максимуму величины W при условии максимума энтропии.

Ознакомимся теперь с методом получения термодинамических соотношений, воспользовавшись для этого простым примером. Предположим, что система 2 состоит из гармонических осцилляторов, обладающих сколь угодно низкой частотой. В таком случае величина $\gamma = \beta h\nu$ много меньше единицы и

$$\mathcal{E} = h\nu Q \left[\frac{1}{2} + \frac{1}{e^{\gamma} - 1} \right] \approx h\nu Q \frac{1}{\beta h\nu} = \frac{1}{\beta} Q. \quad (16.29)$$

Из классической статистической механики известно, что среднее значение энергии системы гармонических осцилляторов равно kTQ . Следовательно, для системы 2 величину $1/\beta_2$ можно отождествить с kT ; поскольку $\beta_1 = \beta_2$, значение β вообще может быть положено равным $1/kT$. Этот метод нельзя, однако, считать вполне удовлетворительным, так как равенство $\mathcal{E} = kTQ$ само по себе основано на классической статистической теории. Поэтому в качестве системы 2 лучше было бы рассмотреть идеальный одноатомный газ. В этом случае имеем

$$pV = \frac{N}{\beta_2}, \quad (16.30)$$

где p — среднее значение давления, V — объем, N — число атомов. Поскольку для идеального газа термодинамическая температурная шкала может быть определена соотношением

$$pV = NkT, \quad (16.31)$$

предположение $\beta_2 = 1/kT$ является при этом естественным. В случае квантового газа это соотношение получается почти непосредственно и не вносит ничего существенно нового по сравнению с результатами, уже полученными в гл. 5, а также рассматриваемыми в следующем параграфе настоящей главы. В заключение отметим, что цель данного параграфа — скорее ввести в рассмотрение различные понятия и соотношения, нежели дать их строгое доказательство; по этой причине здесь не приводятся выводы каких-либо формул. Значительно более общий интерес представляет установление зависимости энергии квантовых состояний от объема V или же иного

параметра, характеризующего механические свойства системы. При этом может быть получено уравнение состояния системы, содержащее величины \mathcal{E} и V . Эти вопросы рассматриваются в виде задач в конце главы.

Вывод закона распределения Ферми — Дирака. Выведем теперь формулу, характеризующую распределение Ферми — Дирака, воспользовавшись для этого тем же методом, что и при отыскании закона распределения Планка. Рассмотрим систему, состоящую из N электронов. Квантовые состояния этой системы подразделим на группы так, чтобы все состояния любой данной группы обладали приблизительно одной и той же энергией. Рассмотрим две такие группы состояний с энергиями \mathcal{E}_r и \mathcal{E}_s ; предположим при этом, что всего имеется S_r состояний с энергией \mathcal{E}_r и что N_r из них заняты электронами, а $P_r (= S_r - N_r)$ — дырками. Выясним прежде всего, каково число способов, которыми N_r электронов могут быть распределены между S_r состояниями. Согласно принципу Паули (см. § 7 гл. 15), электроны следует рассматривать как неотличимые друг от друга, так что квантовое состояние системы электронов характеризуется перечнем занятых одноэлектронных состояний. Наша задача заключается, таким образом, в том, чтобы выяснить, сколькими способами можно выбрать N_r занятых состояний из общего числа S_r состояний. Пронумеруем все состояния, обозначив каждое из них отдельной буквой: a, b, c и т. д. В таком случае распределение электронов может быть, например, задано так:

$$\frac{abef}{dcgh\dots} \quad (16.32)$$

Здесь предполагается, что состояния a, b, e, f заняты, а состояния d, c, g, h, \dots — свободны. Очевидно, что каждый определенный выбор занятых состояний может быть осуществлен $N_r!P_r!$ способами, а также что имеется $S_r!$ способов перестановки S_r букв в этом выражении, с сохранением при этом N_r букв над чертой и P_r букв под нею. Следовательно, всего имеется

$$\frac{S_r!}{N_r!P_r!} \quad (16.33)$$

различных способов выбора N_r занятых состояний. Общее число возможных способов осуществления некоторого распределения N_1, N_2, \dots равно, таким образом,

$$W = \prod_r \left(\frac{S_r!}{N_r!P_r!} \right). \quad (16.34)$$

Предположим теперь, что на рассматриваемое распределение действует некоторое возмущение, вызывающее переход δN_r электронов

из группы s в группу r :

$$N_r \rightarrow N_r + \delta N_r, \quad (16.35a)$$

$$N_s \rightarrow N_s + \delta N_s; \quad \delta N_s = -\delta N_r, \quad (16.35b)$$

$$\delta \mathcal{E} = (\mathcal{E}_r - \mathcal{E}_s) \delta N_r. \quad (16.35в)$$

Все соображения, касающиеся значения величины $\delta \ln W / \delta \mathcal{E}$ при таких переходах, приведенные в предыдущих параграфах, справедливы также и в данном случае, так что можно записать

$$\frac{1}{kT} = \beta = \frac{\delta \ln W}{\delta \mathcal{E}}. \quad (16.36)$$

Величину $\delta \ln W$ легко вычислить, принимая во внимание, что

$$\delta \ln N_r! = \delta N_r \ln N_r, \quad (16.37)$$

$$\delta \ln P_r! = \delta P_r \ln P_r = -\delta N_r \ln P_r. \quad (16.38)$$

При этом получаем

$$\beta = \frac{\delta \ln W}{\delta \mathcal{E}} = \frac{1}{\mathcal{E}_r - \mathcal{E}_s} (-\ln N_r + \ln P_r + \ln N_s - \ln P_s). \quad (16.39)$$

Отсюда следует, что

$$\beta \mathcal{E}_r + \ln \frac{N_r}{P_r} = \beta \mathcal{E}_s + \ln \frac{N_s}{P_s}. \quad (16.40)$$

Поскольку это соотношение должно иметь место при любом выборе значений s и r , каждая из его частей не должна зависеть от s и r и может быть положена равной $\beta \mathcal{E}_F$. Вводя в рассмотрение функцию распределения Ферми — Дирака $f(\mathcal{E}_r) \equiv f_r$ с помощью уравнений

$$N_r = f_r S_r \quad \text{и} \quad P_r = (1 - f_r) S_r \quad (16.41)$$

получаем

$$\ln \frac{f_r}{1 - f_r} = \beta (\mathcal{E}_F - \mathcal{E}_r), \quad (16.42)$$

откуда

$$f_r = \frac{1}{1 + \exp \beta (\mathcal{E}_r - \mathcal{E}_F)}. \quad (16.43)$$

Поскольку $\beta \equiv 1/kT$, это и есть искомый закон распределения Ферми — Дирака. Следует отметить, что произвольность подразделения состояний на группы, содержащие по S_r состояний, не сказывается на окончательном ответе и что f является функцией энергии рассматриваемого состояния, уровня Ферми \mathcal{E}_F и величины kT . Вопрос об определении \mathcal{E}_F для полупроводников исследуется в § 3 настоящей главы.

Непосредственно из (16.40) следует, что уровень Ферми можно интерпретировать как *химический потенциал*. Предположим, что

к исследуемой системе добавлено δN электронов. Если эти электроны попадают в состояния группы r , то увеличение свободной энергии системы $\mathcal{F} \equiv \mathcal{E} - T\mathcal{S} = \mathcal{E} - kT \ln W$, как это следует из (16.34), (16.40), а также из определения \mathcal{E}_r , при этом равно

$$\delta \mathcal{F} = \delta \mathcal{E} - kT \delta \ln W = \mathcal{E}_r \delta N + kT \delta N \ln \frac{N_r}{P_r} = \mathcal{E}_F \delta N. \quad (16.44)$$

Как мы видим, изменение свободной энергии системы $\delta \mathcal{F}$ не зависит от того, какую именно группу состояний занимают добавляемые электроны. Величина \mathcal{E}_F играет при этом роль химического потенциала, который может быть определен как изменение свободной энергии системы при добавлении к ней одного атома данного химического элемента ¹⁾.

Этот результат может быть использован для доказательства сделанного в § 4 гл. 12 заключения о том, что разность уровней Ферми представляет собой значение напряжения, отсчитываемое по вольтметру. Предположим, что из проводника A с уровнем Ферми \mathcal{E}_{FA} в проводник B с уровнем Ферми \mathcal{E}_{FB} переходит δN электронов, причем этот переход осуществляется обратимо, так что

$$\text{Высвобождаемая энергия} = -e\delta N(V_A - V_B), \quad (16.45)$$

где V_A и V_B — значения напряжения. Задача заключается в том, чтобы показать, что

$$-e(V_A - V_B) = \mathcal{E}_{FA} - \mathcal{E}_{FB}.$$

Для доказательства этого соотношения применим теорему термодинамики, согласно которой при любом обратимом процессе, не сопровождающемся теплообменом, $\delta \mathcal{S} = 0$. Эта теорема будет доказана при решении задачи 12 для того случая, когда работа совершается давлением. [Мы докажем при этом, что

$$d\mathcal{S} = k\beta d\lambda,$$

где

$$d\lambda = d(\text{энергия системы}) - d(\text{работа, произведенная системой}).]$$

Предположим, что разность напряжений поддерживается с помощью батареи, включенной между проводниками A и B . Если рассматриваемая система изолирована термически так, что приток тепла извне исключен, то при обратимом переходе электронов из проводника A в проводник B $d\mathcal{S} = 0$. Работа $-e\delta N(V_A - V_B)$, произведенная системой, должна поэтому равняться $-\delta \mathcal{E}$, и можно записать

$$-e\delta N(V_A - V_B) = -\delta \mathcal{E} = -(\delta \mathcal{E} - T\delta \mathcal{S}) = -\delta \mathcal{F}. \quad (16.46)$$

¹⁾ В более общем случае — при добавлении одной частицы данного сорта. — *Прим. ред.*

Изменение свободной энергии системы $\delta\mathcal{F}$ может быть, однако, представлено в виде суммы двух членов: члена $\delta\mathcal{F}_A = -\mathcal{E}_{FA}\delta N$, соответствующего удалению δN электронов из проводника A , и члена $\delta\mathcal{F}_B = \mathcal{E}_{FB}\delta N$, соответствующего введению этих электронов в проводник B . Подстановка

$$\delta\mathcal{F} = \delta\mathcal{F}_A + \delta\mathcal{F}_B$$

в (16.46) приводит к искомому соотношению

$$-e(V_A - V_B) = \mathcal{E}_{FA} - \mathcal{E}_{FB}. \quad (16.47)$$

§ 2. ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОГО ЧИСЛА КВАНТОВЫХ СОСТОЯНИЙ В ЭНЕРГЕТИЧЕСКОЙ ЗОНЕ

Для вычисления концентрации электронов в зоне проводимости можно воспользоваться формулой (10.17), определив при этом $N(\mathcal{E})$ по формуле (9.9). Это приводит к выражению вида

$$\begin{aligned} n &= \exp\left(\frac{\mathcal{E}_F}{kT}\right) \int_{\mathcal{E}_c}^{\mathcal{E}_2} \frac{4\pi(2m)^{3/2}}{h^3} (\mathcal{E} - \mathcal{E}_c)^{1/2} \exp\left(-\frac{\mathcal{E}}{kT}\right) d\mathcal{E} = \\ &= \exp\left(\frac{\mathcal{E}_F - \mathcal{E}_c}{kT}\right) \frac{4\pi(2m)^{3/2}}{h^3} \int_0^{\mathcal{E}_2 - \mathcal{E}_c} \exp\left(-\frac{\mathcal{E}_1}{kT}\right) \mathcal{E}_1^{1/2} d\mathcal{E}_1. \end{aligned} \quad (16.48)$$

Мы произвели здесь подстановку $\mathcal{E}_1 = \mathcal{E} - \mathcal{E}_c$. Поскольку интеграл, входящий в (16.48), быстро сходится, его верхний предел может быть заменен на бесконечность. В результате интеграл приобретает стандартный вид (см. [5]); его значение равно $(\pi kT)^{1/2} kT/2$. Подстановка этой величины в (16.48) дает

$$n = 2 \left(\frac{2\pi m k T}{h^2} \right)^{3/2} \exp\left(-\frac{\mathcal{E}_c - \mathcal{E}_F}{kT}\right), \quad (16.49)$$

т. е. формулу (10.14). Эффективное число состояний в зоне проводимости равно, таким образом,

$$\begin{aligned} N_c &= 2 \left(\frac{2\pi m k T}{h^2} \right)^{3/2} = \\ &= 4,82 \cdot 10^{15} T^{3/2} \text{ см}^{-3} = \\ &= 2,41 \cdot 10^{19} \text{ см}^{-3} \text{ при } 20^\circ\text{C}, \end{aligned} \quad (16.50)$$

причем

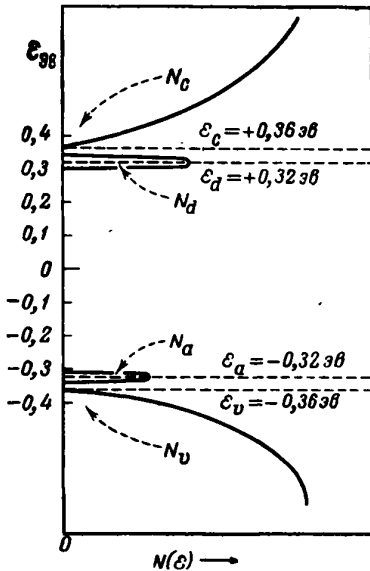
$$N_c^2 = 2,32 \cdot 10^{31} T^3 \text{ см}^{-6}. \quad (16.51)$$

Этот же результат может быть получен путем непосредственного интегрирования по \mathbf{P} -пространству без использования формулы (9.9).

В этом случае следует исходить не из значения $N(\mathcal{E})d\mathcal{E}$, а из значения $(2/h^3)dP_x dP_y dP_z$, причем $\mathcal{E} = (P_x^2 + P_y^2 + P_z^2)/2m$. Интеграл, входящий в формулу (16.48), при этом может быть записан в виде произведения трех интегралов, каждый из которых при вычислении N_c дает множитель $(2\pi mkT/h^2)^{3/2}$.

§ 3. ТЕМПЕРАТУРНАЯ ЗАВИСИМОСТЬ УРОВНЯ ФЕРМИ

В этом параграфе мы исследуем более строго, нежели это было сделано ранее, вопрос о температурной зависимости величины \mathcal{E}_F , а также о концентрации электронов и дырок. При этом мы будем исходить из распределения квантовых состояний, уже рассмотренного в гл. 9. Это распределение иллюстрируется фиг. 130, на которой



Фиг. 130. Распределение квантовых состояний по энергиям.

$N(\mathcal{E})$ — число квантовых состояний на единицу энергии в единице объема.

приведены значения эффективного числа квантовых состояний в заполненной и свободной зонах. Как было показано в гл. 10, требование электрической нейтральности системы заключается в том, что сумма положительных зарядов:

- p — концентрации дырок в заполненной зоне,
- p_a — концентрации дырок, связанных с акцепторами,
- N_a — концентрации доноров

должна равняться сумме отрицательных зарядов:

- n — концентрации электронов в зоне проводимости,
- n_d — концентрации электронов, связанных с донорами,
- N_a — концентрации акцепторов.

Это приводит к уравнению

$$p + p_a + N_a = n + n_d + N_a \quad (16.52)$$

В гл. 10 были получены соотношения, определяющие значения этих концентраций [см. (10.13), (10.16), (10.18), (10.19)]. Последние зависят от природы полупроводника, будучи связаны со значениями энергии \mathcal{E}_v , \mathcal{E}_a , \mathcal{E}_d и \mathcal{E}_c , эффективными массами, входящими в определение N_v и N_c , и концентрациями примесей N_a и N_d . Если эти величины заданы, то задача заключается в отыскании зависимости n и p от температуры T путем решения уравнения (16.52)

относительно \mathcal{E}_F . Для этого, как показано ниже, левая и правая части уравнения (16.52) строятся как функции \mathcal{E}_F ; решение находится затем графически. Рассматриваемый ниже пример соответствует случаю германия *n*-типа, обладающего при комнатной температуре сопротивлением в 2,4 ом·см, как и в случае, исследованном в § 4 гл. 10.

В этом примере приняты следующие численные значения различных величин:

$$\begin{aligned} \mathcal{E}_v &= -0,36 \text{ эв}, & \mathcal{E}_a &= -0,32 \text{ эв}, \\ \mathcal{E}_d &= +0,32 \text{ эв}, & \mathcal{E}_c &= +0,36 \text{ эв} \end{aligned} \quad (16.53)$$

(энергия отсчитывается при этом от середины запрещенной зоны). Значения N_d и N_a положим равными

$$N_d = 10^{16} \text{ см}^{-3}, \quad N_a = 10^{14} \text{ см}^{-3}, \quad (16.54)$$

а массы электрона и дырки будем считать равными массе свободного электрона, так что при этом применимо уравнение (16.50).

Значения концентраций, полученные при этих предположениях, приведены на фиг. 131 в логарифмическом масштабе как функции \mathcal{E}_F при $T = 300^\circ\text{K}$. Ход концентраций положительных зарядов изображен сплошными линиями, а сумма этих концентраций, представляющая собой левую часть уравнения (16.52), — жирной сплошной кривой. Концентрации отрицательных зарядов изображены пунктирными линиями. В точке пересечения жирных кривых удовлетворяется уравнение (16.52) и выполняется условие электрической нейтральности рассматриваемой системы.

Отдельные кривые, характеризующие изменение концентраций отрицательных зарядов, можно интерпретировать следующим образом:

N_a — постоянный отрицательный заряд, создаваемый акцепторами;

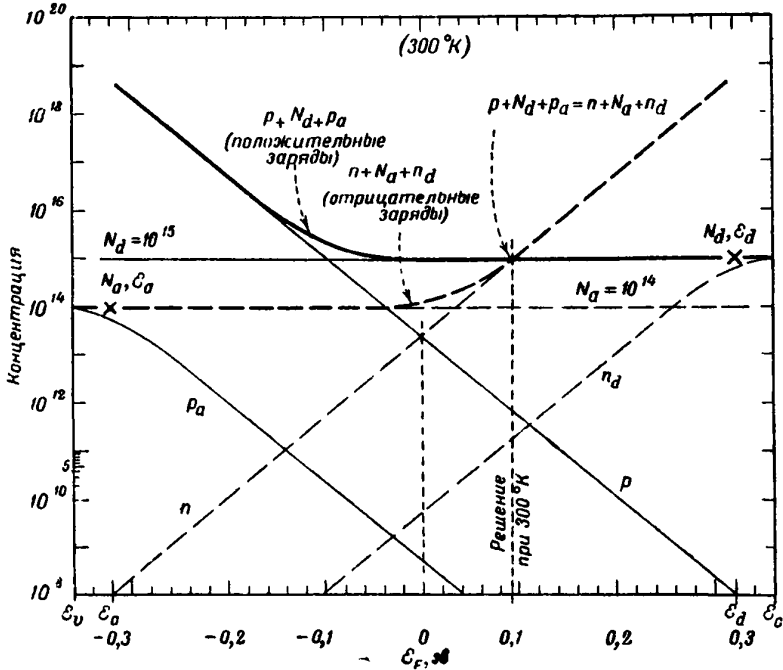
n — концентрация электронов в зоне проводимости. Согласно § 2 настоящей главы, она равна примерно $2,41 \cdot 10^{19} \exp [(-\mathcal{E}_c + \mathcal{E}_F)/kT]$. В электрон-вольтах это составляет $2,41 \cdot 10^{19} \exp(-39 \Delta V)$, где ΔV — разность между \mathcal{E}_F и \mathcal{E}_c , отсчитываемая вниз от дна зоны проводимости, или

$$n = 2,41 \cdot 10^{19} \cdot 10^{-17 \Delta V}. \quad (16.55)$$

В логарифмическом масштабе это дает прямую линию, изображенную на фиг. 131. (В том случае, когда расстояние между \mathcal{E}_F и \mathcal{E}_c становится порядка kT , рассматриваемое приближение уже непригодно и ход $\ln n$ должен отклоняться от прямолинейного, как это показано на фиг. 91, стр. 292.)

n_d — число электронов, связанных с донорами. В том случае, когда \mathcal{E}_F много больше \mathcal{E}_d , все донорные состояния заняты и $n_d = N_d$. При $\mathcal{E}_F = \mathcal{E}_d$ занята половина этих состояний и кривая для n_d лежит на 2 единицы ниже точки (N_d, \mathcal{E}_d) , отмеченной на рисунке крестиком. При более низких значениях \mathcal{E}_F может быть использовано

приближение $f = \exp(\mathcal{E}_F - \mathcal{E}_a)/kT$, дающее для n_d прямую линию с тем же наклоном, что и для n . Если продолжить эту прямую, то она пройдет через точку (N_a, \mathcal{E}_a) . Совершенно аналогичные условия имеют место также и для концентраций положительных зарядов N_d, p и p_a , с той разницей, что речь идет при этом о распределении дырок, а не о распределении электронов.

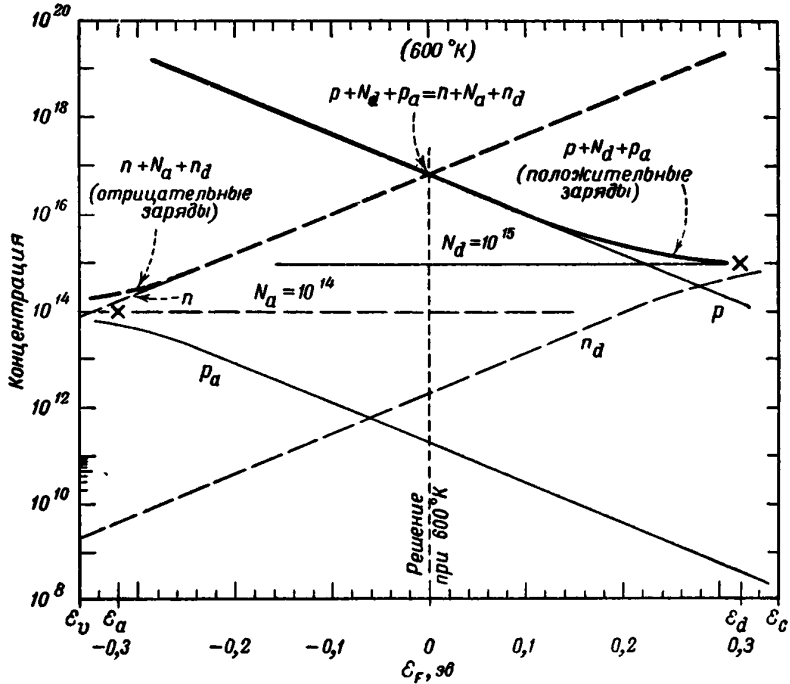


Фиг. 131. Значение \mathcal{E}_F при $N_a = 10^{14}$ см^{-3} , $N_d = 10^{15}$ см^{-3} и $T \approx 300^\circ\text{K}$.

В случае, иллюстрируемом фиг. 131, при всех значениях \mathcal{E}_F n много больше n_d . Это обусловлено тем обстоятельством, что kT сравнимо с энергией связи электрона в кристалле и N_c много больше N_a . В результате наличие большого числа состояний, которые электроны могут занимать в зоне проводимости, играет значительно более существенную роль, чем различие статистических множителей f , являющееся благоприятным для заполнения нижних энергетических состояний. При равновесии (которому отвечает точка пересечения жирных кривых) концентрации дырок p и p_a настолько малы, что ими можно пренебречь. Таким образом,

$$n + N_a = N_d \text{ или } n = N_d - N_a. \quad (16.56)$$

С физической точки зрения это означает, что избыточные электроны, обусловленные избытком числа доноров над числом акцепторов, в результате теплового возбуждения стали свободными и находятся в зоне проводимости. Однако поскольку концентрация p очень невелика, рассматриваемая температура еще недостаточно высока для разрыва значительного числа валентных связей.



Фиг. 132. Значение \mathcal{E}_F при $N_a = 10^{14}$ см⁻³, $N_d = 10^{15}$ см⁻³ и $T \approx 600^\circ\text{K}$.

Случай $T = 600^\circ\text{K}$ иллюстрируется фиг. 132. При этих условиях основную роль играют концентрации n и p , тогда как значения N_a и N_d не влияют сколько-нибудь заметным образом на результаты. Этот случай соответствует температуре, при которой мы уже имеем дело с собственной проводимостью вещества. Из условия нейтральности следует, что при этом

$$\mathcal{E}_F \approx \frac{\mathcal{E}_v + \mathcal{E}_c}{2},$$

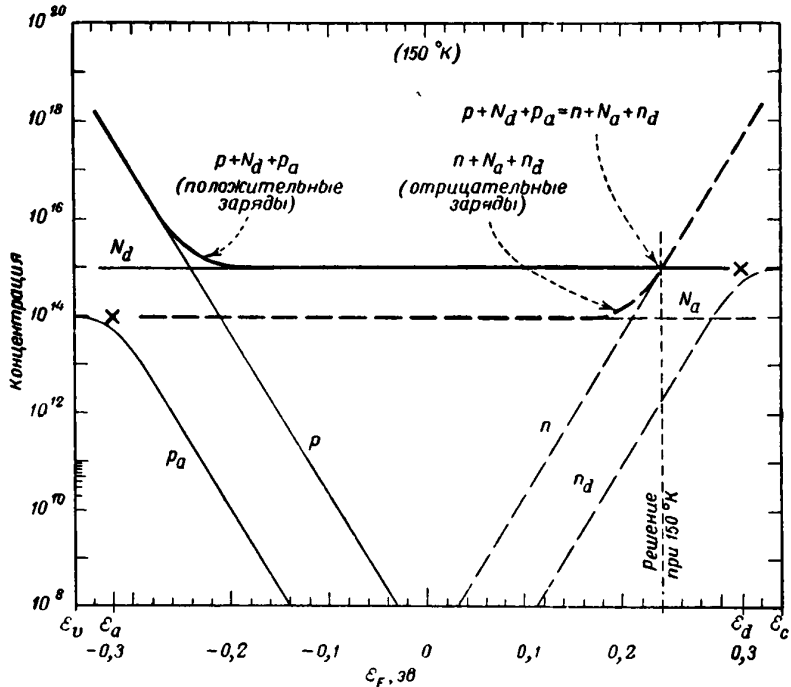
так что $p \approx n$.

В случае, рассматриваемом на фиг. 133, $T = 150^\circ\text{K}$ и при всех значениях величины \mathcal{E}_F концентрация n больше чем в 10 раз

превышает n_d , составляя, таким образом, наибольшую долю переменной части отрицательного заряда. Равновесие осуществляется при условии

$$n + n_a = N_a - N_d. \quad (16.57)$$

Это значит, что электроны, наличие которых обусловлено избытком числа доноров над числом акцепторов, нейтрализуют доноры и акцепторы,



Ф и г. 133. Значение ϵ_F при $N_a = 10^{14} \text{ см}^{-3}$, $N_d = 10^{16} \text{ см}^{-3}$ и $T \approx 150^\circ\text{K}$.

причем большинство таких электронов находится в зоне проводимости и только примерно 1 из 3 000 электронов остается связанным.

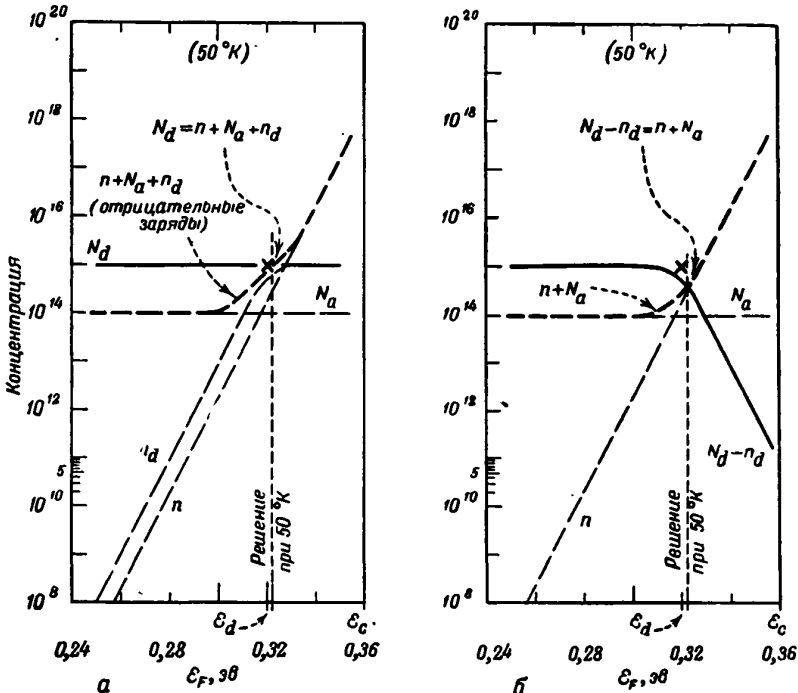
При $T = 50^\circ\text{K}$ (фиг. 134) n_d превышает n при условии, что $\epsilon_F < \epsilon_d$. Если $\epsilon_F > \epsilon_d$, то донорные состояния в основном заполнены, т. е. большинство электронов осталось невозбужденными на донорных уровнях. Концентрации p и p_a в этом случае ничтожно малы и их значения не указаны на рисунке, который иллюстрирует в увеличенном масштабе ход концентраций при энергиях, близких к ϵ_d . Условие равновесия, как и в предыдущем примере, имеет вид

$$n + N_a + n_a = N_d. \quad (16.58)$$

В данном случае, когда донорные уровни почти полностью заполнены, удобнее, однако, рассматривать не электроны, а дырки, локализованные на донорных уровнях. Число таких дырок равно $N_d - n_d$. Уравнение (16.58) можно записать в виде

$$n + N_a = N_d - n_d. \quad (16.59)$$

Здесь $(n + N_a)$ — концентрация отрицательных зарядов, а $(N_d - n_d)$ — концентрация положительных зарядов, т. е. число донорных уровней,

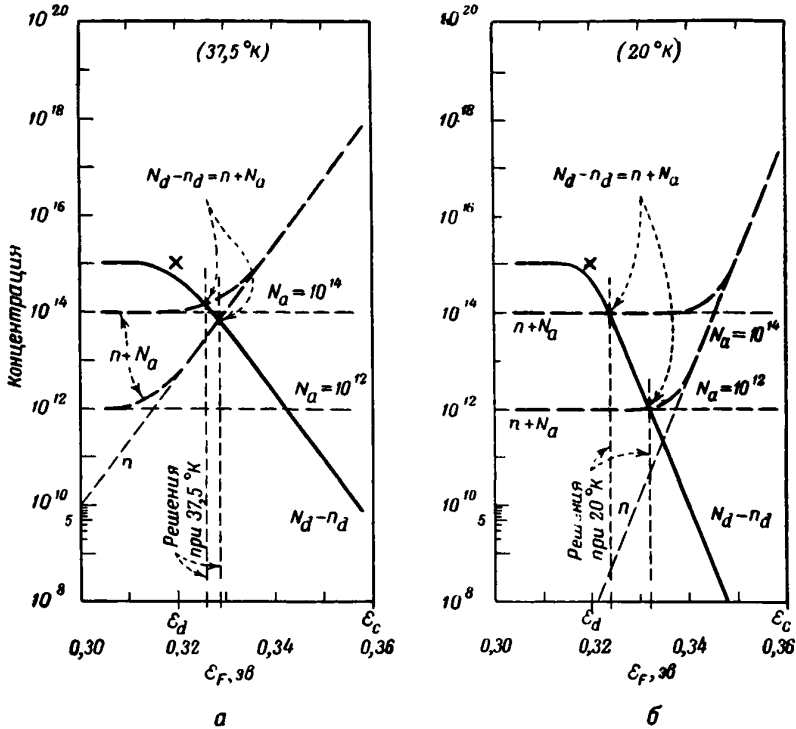


Фиг. 134. Значение ϵ_F при $N_a = 10^{14} \text{ см}^{-3}$, $N_d = 10^{15} \text{ см}^{-3}$ и $T \approx 50^\circ \text{ К}$.

занятых дырками. Зависимость $(N_d - n_d)$ от ϵ_F при этом аналогична зависимости p_a от ϵ_F ; при $\epsilon_F = \epsilon_d$ имеем $(N_d - n_d) = N_d/2$. На фиг. 134, б приведен график зависимости $(N_d - n_d)$ (сплошная кривая) и график зависимости $(n + N_a)$ (пунктирная кривая) от ϵ_F .

На фиг. 135 показан ход концентраций при $T = 37,5^\circ \text{ К}$ и $T = 20^\circ \text{ К}$. Здесь $N_a = 10^{12}$. При этих температурах ϵ_F столь близко к ϵ_d , что при приближении к $T = 0$ донорные уровни остаются частично заполненными. Очевидно, однако, что если бы N_a было

равно нулю, то при $T \rightarrow 0$ \mathcal{E}_F лежало бы посередине между \mathcal{E}_c и \mathcal{E}_d , подобно тому как в области собственной проводимости \mathcal{E}_F лежит посередине между заполненной зоной и зоной проводимости.



Фиг. 135. Значение \mathcal{E}_F при $N_a = 10^{14}$ см^{-3} и 10^{12} см^{-3} , $N_d = 10^{15}$ см^{-3} , $T \approx 37,5$ и 20°K .

В следующем параграфе мы продолжим анализ изменения величины \mathcal{E}_F и исследуем вопрос о соотношении между наклоном прямой, характеризующей зависимость $\ln n$ от $1/T$, и соответствующими значениями энергии.

§ 4. ЭНЕРГИЯ АКТИВАЦИИ

Полученные выше результаты позволяют сделать следующие выводы. По мере того как T приближается к нулю, n_d стремится к $(N_d - N_a)$ и n стремится к нулю. Проанализируем зависимость $\ln n$ от T и попытаемся связать ее с возможными значениями энергии активации. При низких температурах, когда величиной n можно

пренебречь по сравнению с n_a , условие нейтральности системы приобретает вид

$$n_a = N_d - N_a. \quad (16.60)$$

Принимая во внимание, что

$$n_a = f N_d, \quad (16.61)$$

находим

$$f = \frac{N_d - N_a}{N_d}. \quad (16.62)$$

Поскольку предполагается, что $N_d > N_a$, отсюда следует, что значение f лежит в интервале между нулем и единицей. Это уравнение может быть решено относительно величины \mathcal{E}_F . Действительно,

$$f = \frac{1}{1 + \exp[(\mathcal{E} - \mathcal{E}_F)/kT]}, \quad (16.63)$$

откуда

$$\mathcal{E} - \mathcal{E}_F = \frac{kT \ln(1-f)}{f}. \quad (16.64)$$

Подставляя \mathcal{E}_d вместо \mathcal{E} и $(N_d - N_a)/N_d$ вместо f , получаем

$$\mathcal{E}_F = \mathcal{E}_d + kT \ln\left(\frac{N_d - N_a}{N_d}\right). \quad (16.65)$$

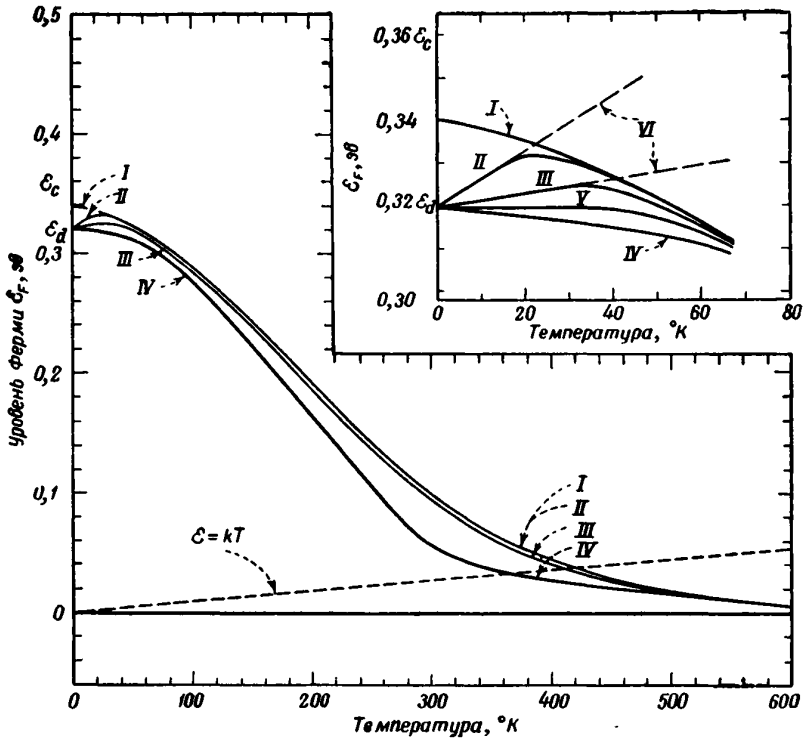
Это показывает, что в тех случаях, когда температура столь низка, что величиной n можно пренебречь по сравнению с n_a (так что n_a можно положить равным $N_d - N_a$), \mathcal{E}_F зависит от T линейно. При $N_d = 2N_a$, N_d состояний при $T=0$ заполнены как раз наполовину и \mathcal{E}_F не зависит от температуры, так как $\ln 1 = 0$. При $N_d > 2N_a$ значение \mathcal{E}_F возрастает с ростом T ; с физической точки зрения это означает, что донорные состояния заполнены больше чем наполовину и \mathcal{E}_F должно, следовательно, лежать выше \mathcal{E}_d . При $N_d < 2N_a$ эти состояния заполнены меньше чем наполовину и \mathcal{E}_F всегда лежит ниже \mathcal{E}_d .

Поскольку \mathcal{E}_F — величина переменная, каково значение энергии активации, получаемое при исследовании зависимости $\ln n$ от $1/T$? Интересно отметить, что оно оказывается равным $(\mathcal{E}_c - \mathcal{E}_d)$. Действительно, формула (16.49) дает

$$\begin{aligned} n &= N_c \exp\left[\frac{\mathcal{E}_F - \mathcal{E}_c}{kT}\right] = \\ &= 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2} \exp\left[\frac{\mathcal{E}_d - \mathcal{E}_c + kT \ln(N_d - N_a)/N_d}{kT}\right] = \\ &= 2\left(\frac{2\pi mk}{h^2}\right)^{3/2} T^{3/2} \frac{N_d - N_a}{N_d} \exp\left[\frac{\mathcal{E}_d - \mathcal{E}_c}{kT}\right]. \end{aligned} \quad (16.66)$$

Отсюда следует, что экспоненциальный множитель содержит лишь разность $\mathcal{E}_d - \mathcal{E}_c$, и что наличие зависимости \mathcal{E}_F от T влияет только

на коэффициент этого выражения. Таким образом, при достаточно низких температурах исследование графика зависимости $\ln(nT^{-3/2})$ от $1/T$ должно дать численное значение разности $(\mathcal{E}_c - \mathcal{E}_d)$.



Фиг. 136. Зависимость положения уровня Ферми от температуры при различных концентрациях акцепторов. $N_d = 10^{16} \text{ см}^{-3}$.
 I. $N_a = 0$; II. $N_a = 10^{15}$; III. $N_a = 10^{14}$; IV. $N_a = 8 \cdot 10^{14}$; V. $N_a = 5 \cdot 10^{14}$; VI — касательные к кривым, определенные из уравнения (16.65).

Этот результат может быть обобщен путем построения графика зависимости \mathcal{E}_F от T , приведенного на фиг. 136. Этот график построен на основании результатов, полученных в § 3, а также в настоящем параграфе. При $N_a = 5 \cdot 10^{14}$, т. е. при $N_a = N_d/2$, начальный участок кривой, характеризующей зависимость \mathcal{E}_F от T , параллелен оси абсцисс. При более высоких значениях N_a величина \mathcal{E}_F убывает с ростом T . При любой зависимости \mathcal{E}_F от T $\ln(nT^{-3/2})$

изменяется пропорционально $(\mathcal{E}_F - \mathcal{E}_c)/kT$. Таким образом,

$$\begin{aligned} \frac{d}{d(1/T)} \left(\frac{\mathcal{E}_F - \mathcal{E}_c}{kT} \right) &= \frac{\mathcal{E}_F - \mathcal{E}_c}{k} + \frac{1}{kT} \frac{d\mathcal{E}_F}{d(1/T)} = \frac{\mathcal{E}_F - \mathcal{E}_c}{k} - \frac{1}{kT} \frac{d\mathcal{E}_F}{dT} = \\ &= \frac{1}{k} \left[\mathcal{E}_F - T \frac{d\mathcal{E}_F}{dT} - \mathcal{E}_c \right] = \frac{1}{k} [\mathcal{E}_F^* - \mathcal{E}_c], \end{aligned} \quad (16.67)$$

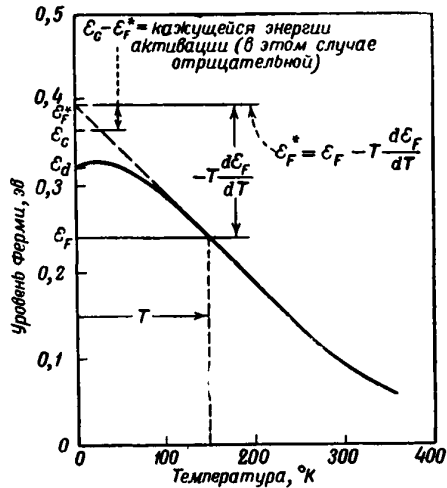
где

$$\mathcal{E}_F^* \equiv \mathcal{E}_F - T \frac{d\mathcal{E}_F}{dT}. \quad (16.68)$$

Это как раз та величина, которую мы получили бы, проведя касательную в некоторой точке кривой вплоть до ее пересечения с осью

ординат, как это показано на фиг. 137. Получаемая таким путем величина \mathcal{E}_F^* представляет собой кажущийся уровень Ферми. В рассматриваемом случае он лежит выше \mathcal{E}_c , так что кажущаяся энергия активации отрицательна. Этот аномальный результат и должен иметь место в области насыщения. В этой области температур n не зависит от T , тогда как N_c возрастает. Таким образом, если трактовать N_c как общее число состояний с энергией \mathcal{E}_c , то повышение T приводит к уменьшению величины f для этих уровней, как если бы уровень Ферми был расположен выше \mathcal{E}_c . Тем самым становится понятным, почему кажущаяся энергия активации оказывается отрицательной, как это показано на фиг. 137.

Эти соображения показывают, что в общем случае анализ кривой зависимости логарифма концентрации от $1/T$ не дает возможности простой интерпретации получаемых таким путем значений энергии активации ¹⁾. Всякий раз необходимо проводить детальное сопоставление теоретических кривых, построенных на основе точных решений уравнений для типичных случаев, примеры которых рассмотрены на фиг. 130—136.



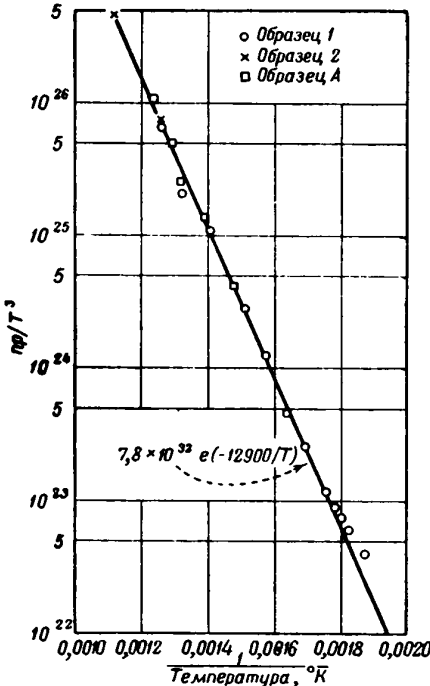
Фиг. 137. Кажущаяся энергия активации при $N_a = 10^{14}$ см⁻³.

¹⁾ Эти замечания автора верны, и их следует учитывать при анализе опытных данных, относящихся к полупроводникам. — *Прим. ред.*

В работе Пирсона и Бардина такого рода анализ осуществлен для ряда сплавов кремния [6]. Наиболее простым при этом оказалось исследование области, в которой как n , так и p принимают достаточно большие значения. Воспользовавшись формулами (10.16) и (10.19), дающими приближенные значения n и p , имеем

$$\begin{aligned} npT^{-3} &= 4 \left(\frac{2\pi mk}{h^2} \right)^3 \exp \frac{\mathcal{E}_v - \mathcal{E}_c}{kT} = \\ &= 2,32 \cdot 10^{31} \exp \frac{(\mathcal{E}_v - \mathcal{E}_c)}{kT} \text{ см}^{-6} \text{ } ^\circ\text{K}^{-3}. \end{aligned} \quad (16.69)$$

Численные результаты, полученные на основе этой формулы, ложатся на прямую линию, как это показано на фиг. 138. Численное значение



Фиг. 138. Сопоставление экспериментальных данных для нескольких образцов кремния с теоретической кривой, вычисленной по формулам (16.69) и (16.70).

данными Пирсона и Бардина и теорией. Этот результат согласуется со значениями β , полученными другими методами (см. § 8 гл. 12). Наблюдавшиеся расхождения между значениями β могут быть обусловлены

предэкспоненциального множителя, определенное на основании экспериментальных данных, однако, примерно в 34 раза больше, нежели это следует из формулы (16.69). Этот результат может быть объяснен в предположении, что ширина запретной зоны $(\mathcal{E}_c - \mathcal{E}_v)$ зависит от температуры и при повышении T уменьшается (повидимому, вследствие теплового расширения), как $\mathcal{E}_G - \beta T$. Если мы учтем это в формуле (16.69), то экспоненциальный множитель изменится следующим образом:

$$\begin{aligned} \exp \frac{(\mathcal{E}_v - \mathcal{E}_c)}{kT} &= \\ &= e^{\beta/k} \exp \left(- \frac{\mathcal{E}_G}{kT} \right). \end{aligned} \quad (16.70)$$

Численное значение коэффициента β было также оценено Гаучером и Бриггсом на основании измерений фотоэ. д. с., возникающей на $n-p$ -переходах в кремнии. Оно оказалось при этом несколько меньшим, чем это нужно для того, чтобы объяснить расхождение между

тем обстоятельством, что эффективные массы электрона и дырки отличны от массы свободного электрона.

ЗАДАЧИ

(См. также гл. 10)

1. Вывести закон распределения Ферми — Дирака, рассмотрев для этого результат воздействия на систему двух возмущений $\delta\mathcal{E}$ и $\delta\mathcal{E}'$, подобно тому как это было сделано в § 1 при выводе закона распределения Планка.

2. Исследовать распределение электронов, захваченных донорами, учтя при этом роль спина. (Из чисто электростатических соображений следует, что каждый донор может захватить не больше одного электрона.) Показать, что имеется

$$\frac{2^{n_d} N_d!}{n_d! (N_d - n_d)!}$$

способов распределения n_d электронов по N_d уровням и, следовательно,

$$n_d = \frac{N_d}{1 + \left(\frac{1}{2}\right) \exp(\mathcal{E}_d - \mathcal{E}_F)/kT}.$$

Этот результат был использован в гл. 10 при решении задачи 1.

3. Показать, что для распределения Планка $\ln W$ действительно имеет максимум. Для этого рассмотрите значение $\delta \ln W$ при наличии большого возмущения, положив $\delta\mathcal{E}' = -\delta\mathcal{E}$.

Примечание. Следующие задачи иллюстрируют один из методов, с помощью которого можно показать, что соотношения термодинамики являются непосредственным следствием статистической механики квантованных систем¹⁾. В этих задачах мы вводим также в рассмотрение функции Z и z , которые широко используются при вычислениях средних значений различных физических величин и их флуктуаций. Основным при этом является предположение о равенстве априорных вероятностей для всех квантовых состояний системы. В задаче 4 вводятся в рассмотрение понятие канонического ансамбля и функция Z и показано, что для канонического ансамбля средние значения совпадают со средними значениями для одного из входящих в его состав микроканонических ансамблей.

Таким образом, все существенные слагаемые, определяющие величину Z , отвечают одному определенному микроканоническому

¹⁾ Наше изложение сходно с изложением, которому следует Шредингер [7], отличаясь от него, однако, тем, что мы обращаем особое внимание на роль контакта с тепловым резервуаром. — *Прим. авт.*

ансамблю. В задаче 5 устанавливается связь между числом состояний в микроканоническом ансамбле и величиной Z . В задаче 6 показано, что слагаемые, определяющие величину Z для канонического ансамбля, отвечают в основном одному определенному распределению и что чем больше размеры системы, тем менее существенную роль играют другие возможные распределения. Это распределение должно быть, конечно, наиболее вероятным. Решение всех этих задач показывает, что при вычислении средних значений различных физических величин, характеризующих свойства рассматриваемой системы, вместо понятия микроканонического ансамбля можно воспользоваться понятием канонического ансамбля. Этот результат обладает тем преимуществом, что дает возможность ввести в рассмотрение функцию Z . В задачах 9 и 10 показано, что понятие канонического ансамбля имеет вполне определенный физический смысл; действительно, канонический ансамбль характеризует распределение не изолированной системы, но системы, находящейся в контакте с некоторым тепловым резервуаром. В задачах 11—14 среднее значение давления для системы, находящейся в равновесии с тепловым резервуаром, выражено через Z ; эти задачи иллюстрируют также эквивалентность статистических и термодинамических свойств системы.

4. Рассмотрев среднее по каноническому ансамблю, показать, что заметные отклонения от равновесного распределения не имеют места. Понятие канонического ансамбля вводится при этом как вспомогательное и определяется следующим образом.

Каждое квантовое состояние системы с энергией \mathcal{E}_α обладает статистическим весом $\exp(-\beta \mathcal{E}_\alpha)$. Пусть

$$Z = \sum_{\text{все } \alpha} e^{-\beta \mathcal{E}_\alpha}$$

(Z называется *статистической суммой*). Взвешенное среднее значение энергии для канонического ансамбля равно

$$\langle \mathcal{E} \rangle = \frac{\sum \mathcal{E}_\alpha e^{-\beta \mathcal{E}_\alpha}}{Z} = - \frac{d}{d\beta} \ln Z.$$

(Скобки $\langle \rangle$ обозначают взвешенное среднее для канонического ансамбля.) Аналогичным образом

$$\langle \mathcal{E}^2 \rangle = \frac{\sum \mathcal{E}_\alpha^2 e^{-\beta \mathcal{E}_\alpha}}{Z} = \frac{1}{Z} \frac{d^2}{d\beta^2} Z.$$

Предположим, что систему можно рассматривать, как состоящую из большого числа одинаковых и не зависящих друг от друга подсистем 1, 2, ..., Q . Если известны энергетические уровни одной из таких подсистем $\epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma, \dots$, то энергия системы в целом может

быть записана как сумма значений энергии всех подсистем. Показать, что

$$Z = z^Q,$$

где z — статистическая сумма для одной из подсистем. Показать, что

$$\langle \mathcal{E}^2 \rangle - \langle \mathcal{E} \rangle^2 = Q [\langle \epsilon^2 \rangle_s - \langle \epsilon \rangle_s^2],$$

где $\langle \rangle_z$ — среднее значение рассматриваемой величины для подсистемы. Отсюда следует, что для большой системы, например при $Q > 10^{10}$, состояния, для которых $|\mathcal{E} - \langle \mathcal{E} \rangle|^2 > 10^{-6} \langle \mathcal{E} \rangle^2$, играют очень незначительную роль при определении Z , при условии, что $\langle \epsilon^2 \rangle_s$ не превышает заметным образом $\langle \epsilon \rangle_s^2$. Как, исходя из этого результата, показать, что среднее значение некоторой физической величины, вычисленное для канонического ансамбля, равно среднему значению этой величины для микроканонического ансамбля при $\mathcal{E} = \langle \mathcal{E} \rangle$?

Показать также, что если мы вычислим z для энергетических уровней $(n + 1/2) h\nu$, то формула

$$\langle \mathcal{E} \rangle = -Q \frac{d}{d\beta} \ln z$$

(используемая в предыдущем доказательстве) приводит к выражению, определяющему энергию гармонического осциллятора [см. (16.22)].

5. Исходя из результатов, полученных при решении предыдущей задачи, показать, что

$$\ln W_m(\mathcal{E}) = \beta \mathcal{E} + \ln Z,$$

где $W_m(\mathcal{E})$ — число состояний в микроканоническом ансамбле, энергия которого лежит в малом интервале вблизи энергии \mathcal{E} . Величины Z и $\langle \mathcal{E} \rangle$ являются функциями еще не определенного нами параметра β , входящего в Z . При этом следует записать Z в форме

$$Z = \sum_{\mathcal{E}} W_m(\mathcal{E}) e^{-\beta \mathcal{E}};$$

состояния подразделены здесь на группы состояний, обладающих примерно одинаковой энергией \mathcal{E} . Показать, что при любом разумном выборе энергетических интервалов, вводимых в рассмотрение при такой группировке состояний, значение $\ln W_m(\mathcal{E})$ не зависит от размеров этих интервалов, при условии, что Q велико. (Показать, например, что если ширина интервала предполагается пропорциональной Q^1 , Q^0 или Q^{-1} , то значение $\ln W_m(\mathcal{E})$ при этом во всех случаях остается одинаковым.)

6. Показать, что Z определяется в основном состояниями, отвечающими одному и тому же распределению. Для этого показать, что существует некоторое среднее распределение, и что отклонения от

этого среднего распределения ничтожны. Соответственно можно определить среднее число $\langle Q_\alpha \rangle$ подсистем с энергией ϵ_α , а также среднее значение квадрата этого числа $\langle Q_\alpha^2 \rangle$. Далее можно показать, что среднее квадратичное отклонение Q_α пропорционально Q , так что относительное значение этой величины пропорционально $1/Q$ и при больших Q ничтожно мало. Этот анализ проводится следующим образом.

Каждое из слагаемых статистической суммы Z имеет вид $\exp[-\beta(Q_\alpha \epsilon_\alpha + Q_\beta \epsilon_\beta + Q_\gamma \epsilon_\gamma + \dots)]$, где Q_α — число подсистем в данном состоянии α . Очевидно, что

$$\langle Q_\alpha \rangle = -\frac{1}{\beta Z} \frac{\partial}{\partial \epsilon_\alpha} Z$$

и

$$\langle Q_\alpha^2 \rangle = \frac{1}{\beta^2 Z} \frac{\partial^2}{\partial \epsilon_\alpha^2} Z.$$

Подставляя z^Q вместо Z , получаем

$$\langle Q_\alpha \rangle = Q \frac{e^{-\beta \epsilon_\alpha}}{z},$$

$$\langle Q_\alpha^2 \rangle = \langle Q_\alpha \rangle^2 + Q \frac{e^{-\beta \epsilon_\alpha}}{z} \left[1 - \frac{e^{-\beta \epsilon_\alpha}}{z} \right].$$

Показать, что при больших значениях Q относительное квадратичное отклонение Q_α ничтожно, так что все существенные слагаемые, определяющие статистическую сумму Z , отвечают одному и тому же распределению.

7. Основываясь на результатах, полученных при решении задач 5 и 6, доказать, что для систем большого размера (и конечных значений параметра β)

$$\ln W(\langle \mathcal{E} \rangle) = \beta \langle \mathcal{E} \rangle + \ln Z,$$

где значение W относится к наиболее вероятному распределению.

8. Показать, что если в качестве подсистем рассматривать отдельные осцилляторы, то решение задачи 6 приводит непосредственно к закону распределения Планка [см. (16.24)]. В общем случае подсистемы могут быть значительно более сложными. (Распределение Ферми — Дирака таким путем, конечно, получено быть не может, так как в силу принципа Паули отдельные электроны нельзя рассматривать как независимые друг от друга подсистемы.)

9. Предположим, что тепловой резервуар можно рассматривать как систему, для которой Q сколь угодно велико. Показать, что при больших изменениях энергии $\delta \mathcal{E}$ для такого резервуара удовлетворяется условие

$$\delta \ln W(\mathcal{E}) = \beta \delta \mathcal{E},$$

У к а з а н и е. Выразите через Q и $z(\beta)$ результат, полученный выше при решении задачи 7, и предположите, что $Q \rightarrow \infty$.

10. Рассматриваемая система находится в равновесии с тепловым резервуаром, характеризующимся параметром β_R . Показать, что относительная вероятность нахождения такой системы в состоянии с энергией \mathcal{E}_a равна $\exp(-\beta_R \mathcal{E}_a)$. Для этого нужно воспользоваться предположением о равенстве априорных вероятностей, методом составных систем, использованным в § 1 настоящей главы, а также результатом, полученным при решении предыдущей задачи. Это показывает, что в случае системы, находящейся в равновесии с тепловым резервуаром, мы имеем дело с каноническим ансамблем, свойства которого были рассмотрены в задачах 4—6.

11. Предположим, что энергия любого состояния системы \mathcal{E}_a есть функция некоторой переменной, характеризующей механические свойства системы, например функция объема V . В таком случае давление может быть определено по формуле

$$p_a = - \frac{d\mathcal{E}_a}{dV}.$$

Статистическая сумма Z будет при этом функцией параметров V и β , т. е. $Z = Z(V, \beta)$. Показать, что среднее значение давления равно

$$\langle p \rangle = \frac{1}{\beta} \frac{\partial}{\partial V} \ln Z.$$

12. Рассмотрим тот случай, когда состояние системы характеризуется параметрами V и β . Предположим, что эти параметры изменились соответственно на δV и $\delta\beta$. Показать, что $\delta \ln W$ (см. задачу 5 или 7) может быть при этом записано в форме

$$\begin{aligned} \delta \ln W &= \beta \left(- \frac{\partial^2 \ln Z}{\partial \beta^2} \delta\beta - \frac{\partial^2 \ln Z}{\partial V \partial \beta} \delta V + \frac{1}{\beta} \frac{\partial \ln Z}{\partial V} \delta V \right) = \\ &= \beta(\delta\mathcal{E}) + \langle p \rangle \delta V = \beta \delta \lambda, \end{aligned}$$

где $\delta \lambda$ — тепло, доставляемое системе тепловыми резервуарами. Из этого выражения следует, что $\beta \delta \lambda$ есть полный дифференциал, поскольку он равен $\delta \ln W$, а $\ln W$ есть функция β и V . Отсюда следует далее, что термодинамический к. п. д. цикла Карно для двух тепловых резервуаров с параметрами β_1 и β_2 равен

$$\frac{d\lambda_1 - d\lambda_2}{d\lambda_1} = 1 - \frac{\beta_1}{\beta_2}.$$

Это показывает, что

$$\beta_1 : \beta_2 = T_2 : T_1$$

в соответствии с определением абсолютной шкалы температур. Постоянную k можно, следовательно, определить как отношение T_1/β_1 . (Доказательство того, что k есть постоянная Больцмана, проводится путем исследования какой-либо конкретной системы, например идеального газа или же системы гармонических осцилляторов.)

Показать, что рассуждения, приведенные в задачах 11 и 12, могут быть отнесены и к электрической системе, такой, например, как аккумуляторная батарея, если мы заменим при этом объем V зарядом, проходящим через батарею, а давление p — напряжением.

13. Показать, что энтропия системы, определяемая как

$$\int \frac{d\lambda}{T} = k \int \beta d\lambda,$$

с точностью до аддитивной постоянной равна

$$\mathcal{S} = k \ln W = \left(\frac{1}{T}\right) (\mathcal{E}) + k \ln Z = -k\beta^2 \frac{\partial}{\partial \beta} \left(\frac{\ln Z}{\beta}\right).$$

Используя соотношение между $\beta d\lambda$ и производными от Z , установленное в задаче 12, показать, что последний член предыдущего выражения определяет энтропию даже в том случае, когда T столь мало, а β соответственно столь велико, что приближение задачи 5 становится непригодным. Показать, что нуль отсчета энтропии может быть при этом выбран так, что при $T=0$ энтропия равна

$$\mathcal{S}_0 = k \ln W(\mathcal{E}_{\min}),$$

где $W(\mathcal{E}_{\min})$ — число квантовых состояний с наименьшим значением энергии. (Равна ли нулю энтропия атома водорода при $T=0$?)

14. Используя данное выше определение энтропии системы \mathcal{S} , показать, что

$$-\left(\frac{1}{\beta}\right) \ln Z = (\mathcal{E}) - T\mathcal{F} \equiv \mathcal{F},$$

где \mathcal{F} — величина, часто называемая свободной энергией. Показать, что

$$\left(\frac{\partial \mathcal{F}}{\partial T}\right)_V = -\mathcal{S}.$$

15. Показать, что для системы, находящейся в равновесии с тепловым резервуаром, среднее квадратичное отклонение энергии $\langle(\mathcal{E} - \langle\mathcal{E}\rangle)^2\rangle = kTc_V T$, где c_V — теплоемкость системы при постоянном объеме.

У к а з а н и е. Следует использовать соотношение

$$c_V = \frac{\partial \langle\mathcal{E}\rangle}{\partial T} = -k\beta^2 \frac{\partial \langle\mathcal{E}\rangle}{\partial \beta}$$

и выразить все величины через Z .

ЛИТЕРАТУРА

1. Fröhlich H., Phys. Rev., 79, 845 (1950).
2. Bardeen J., Phys. Rev., 79, 167 (1950).
3. Fowler R. H., Statistical Mechanics, Cambridge, 1936.
4. Tolman R. C., The Principles of Statistical Mechanics, Oxford, 1938.
5. Pierce, A Short Table of Integrals, New York, 1929.
6. Pearson G. L., Bardeen J., Phys. Rev., 75, 865 (1949).
7. Schroedinger E., Statistical Thermodynamics, Cambridge, 1946 (см. перевод: Шредингер Э., Статистическая термодинамика, ИЛ, 1948).
- 8*. Леонтович М. А., Статистическая физика, М.—Л., 1944.
- 9*. Леонтович М. А., Введение в термодинамику, 2-е изд., М.—Л., 1951.
- 10*. Ландау Л. и Лифшиц Е., Статистическая физика (классическая и квантовая), М.—Л., 1951.
- 11*. Левич В. Г., Введение в статистическую физику, М.—Л., 1950.
- 12*. Френкель Я. И., Статистическая физика, Изд. АН СССР, М.—Л., 1948.
- 13*. Шеффер К., Теория теплоты, ч. II, М.—Л., 1933.
- 14*. Гуревич Л. Э., Основы физической кинетики, М.—Л., 1940.

Глава 17

ТЕОРИЯ ПЕРЕХОДОВ ДЛЯ ДЫРОК И ЭЛЕКТРОНОВ

§ 1. ВВЕДЕНИЕ

Эта глава охватывает наиболее сложный круг вопросов. Здесь рассматривается уравнение Шредингера для системы атомов и электронов в форме, дающей возможность интерпретировать его на языке теории переходов. Исследование распадается на три основные части: 1) отыскание волновых функций, пригодных для описания системы, состоящей из кристалла, атомы которого совершают колебательное движение, и движущегося в кристалле электрона; 2) отыскание закона изменения этих волновых функций с течением времени; 3) объяснение результатов с точки зрения теории переходов.

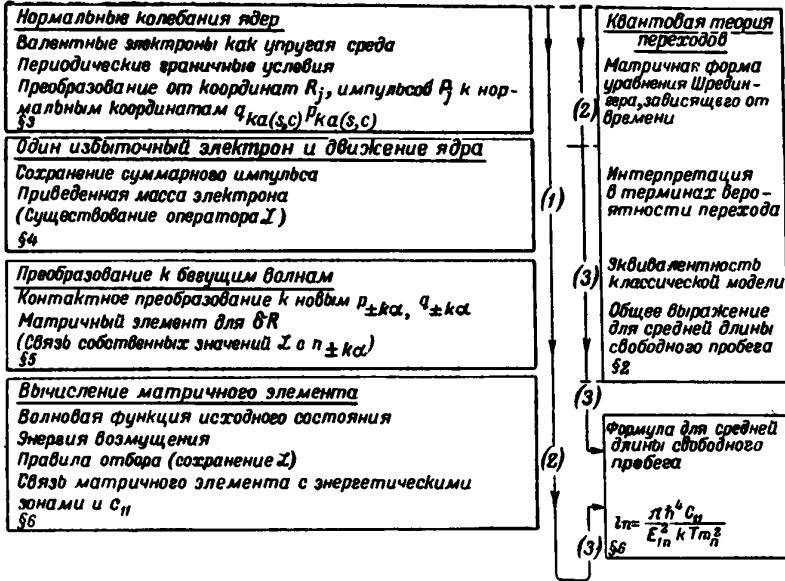
Изложение материала в такой последовательности привело бы, однако, к тому, что он мог бы показаться слишком абстрактным. С педагогической точки зрения правильнее рассмотреть сначала третий вопрос и частично второй и, таким образом, показать в общем виде, как решение уравнения Шредингера дает возможность определить непосредственно вероятности переходов, а также какова при этом роль матричных элементов, вычисленных с помощью исходных значений волновых функций. Соответственно в § 2 прежде всего рассматриваются основные соотношения, определяющие вероятности переходов. При изложении этого вопроса мы следуем методу Вайскопфа — Вигнера, который, насколько известно автору, к вопросам теории электропроводности ранее не применялся. Мы пользуемся здесь этим методом, несмотря на его сложность по сравнению с другими более простыми методами, чтобы выяснить детали тех простых представлений о вероятностях переходов, которые были развиты нами ранее в гл. 8 и 11.

В § 2 будет показано, что вероятности переходов определяются матричными элементами, вычисленными для случая взаимодействия электрона с атомами, колеблющимися в узлах решетки. Волновые функции, необходимые для вычисления этих матричных элементов, определяются в § 3, 4 и 5, а вычисление самих матричных элементов приводится в § 6.

На фиг. 139 приведена схема, показывающая логическую связь между различными частями главы. Подобно всем схемам такого типа, она представляет, вероятно, большую ценность как обзор содержания главы, а не как введение к ней.

В § 7 показано, что задача о дырке эквивалентна задаче об электро-не, и исследован тот случай, когда в кристалле присутствует

большое число дырок и электронов (эта задача рассмотрена также в § 2 гл. 11).



Фиг. 139. Логическая структура гл. 17.

(1) — определение волновых функций; (2) — отыскание закона изменения волновых функций; (3) — интерпретация результатов.

Переходы, обусловленные рассеянием на примесях, были рассмотрены ранее в гл. 11; в данной главе мы к этому вопросу больше не возвращаемся. Следует, однако, отметить, что соотношения, приведенные в § 2, в равной мере могли бы быть применены и к матричным элементам, характеризующим рассеяние на примесях. В § 7 указано, что процесс рассеяния дырок аналогичен процессу рассеяния электронов, независимо от механизма рассеяния.

Ниже приведен перечень обозначений, наиболее часто встречающихся в данной главе:

- Γ_T — положение центра тяжести кристалла, содержащего избыточный электрон (§ 4),
- \mathbf{r} — положение избыточного электрона по отношению к центру кристалла (§ 4),
- m_0 — масса электрона (§ 4),
- m — приведенная масса системы „кристалл + электрон“ (в таком смысле это обозначение используется только в данной главе) (§ 4),

- M_c — масса кристалла (§ 3),
 M_T — масса системы „кристалл + электрон“ (§ 4),
 $M = \frac{1}{2} M_c$ (§ 3),
 N_s — число атомов в кристалле (§ 3),
 Φ — волновая функция системы в целом, т. е. волновая функция кристалла, содержащего избыточный электрон (§ 5),
 χ — сомножитель функции Φ , зависящий от \mathbf{r}_T (§ 4),
 ψ — сомножитель функции Φ , зависящий от \mathbf{r} (§ 4),
 $\varphi_1, \varphi_2, \dots$ — сомножители функции Φ , соответствующие нормальным колебаниям (без учета движения центра тяжести) (§ 5 и 6),
 $\varphi_i = \varphi(n_{ka}, n_{-ka}; q_{kas}, q_{kas})$ — сомножители функции φ_i (§ 5 и 6),
 \mathbf{G}_{ka} — единичный вектор поляризации (§ 3),
 \mathbf{k} — волновой вектор (§ 3),
 α — коэффициент поляризации (§ 3).

§ 2. ТЕОРИЯ ПЕРЕХОДОВ; СВЯЗЬ СРЕДНЕЙ ДЛИНЫ СВОБОДНОГО ПРОБЕГА С МАТРИЧНЫМИ ЭЛЕМЕНТАМИ

В этом параграфе мы выведем формулу для средней длины свободного пробега, воспользовавшись для этого квантовомеханической теорией переходов. Формула для матричных элементов, определяющих вероятности переходов, будет при этом введена в рассмотрение без доказательства; ее доказательство будет дано в § 6 для случая атомных колебаний.

Исследуемая система состоит из ряда атомов, связанных друг с другом валентными электронами, и одного избыточного электрона, находящегося в зоне проводимости. Все электроны, за исключением избыточного, рассматриваются при этом просто как образующие среду, упруго связывающую ядра с их положениями равновесия; ядра могут колебаться около положений равновесия (эта задача будет рассмотрена в § 3 и 5); колебания ядер взаимодействуют с избыточным электроном так, что в результате последний совершает переходы из данного квантового состояния в другие. Из любого данного состояния системы, в котором электрон обладает импульсом P_0 , возможны переходы в состояния, в которых он обладает импульсами P_j ; каждый такой переход может осуществляться двумя способами, в зависимости от того, приобретает электрон энергию или же отдает ее колебаниям решетки. Оказывается, однако, что оба эти перехода возможно трактовать как переход одного и того же типа [см. формулу (17.131)]. По этой причине конечные состояния, в которые электрон переходит из состояния с импульсом P_0 , можно охарактеризовать заданием одного лишь конечного значения импульса P_j . Исходное состояние системы можно, таким образом, описывать функцией Φ_0 , а конечные состояния — функциями Φ_j , где Φ_j — функции координат ядер, а также координат и спина избыточного электрона.

Мы будем предполагать, что волновые функции Φ_j являются собственными функциями оператора \mathcal{H}_0 , который отличается от точного оператора \mathcal{H} рассматриваемой задачи малым возмущающим членом \mathcal{U}_1 . Это предположение, представляющее собой общее требование теории возмущений, будет проверено в дальнейшем.

Таким образом имеем

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{U}_1 \quad (17.1)$$

и

$$\mathcal{H}_0 \Phi_j = \mathcal{E}_j \Phi_j. \quad (17.2)$$

Если вначале в момент времени $t=0$ система находится в состоянии Φ_0 , то наличие возмущения \mathcal{U}_1 будет служить затем причиной ее переходов в другие состояния Φ_j ; наша задача заключается в том, чтобы определить вероятность нахождения системы в любой данной группе состояний как функцию времени. Для этого следует найти решение уравнения Шредингера, зависящего от времени, для оператора \mathcal{H} , представив функцию Φ в виде ряда

$$\Phi = \sum_j a_j \Phi_j,$$

где коэффициенты a_j являются функциями времени, а исходному состоянию при $t=0$ соответствует значение a_j , равное

$$a_j = \delta_{j0}, \quad \text{т. е. } a_0 = 1, \quad a_j \neq 0 = 0.$$

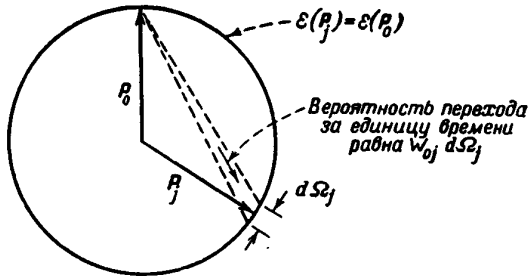
Вероятность нахождения системы в любой группе состояний в некоторый более поздний момент времени равна

$$\sum_j |a_j|^2;$$

суммирование здесь производится по всем интересующим нас состояниям. Прежде чем перейти к исследованию математической стороны этой задачи теории возмущений, рассмотрим вопрос о соотношении между ее постановкой и постановкой аналогичной задачи в гл. 8 и 11.

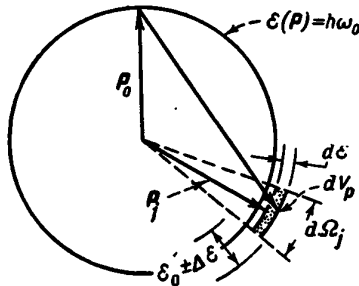
Об эквивалентности квантовой и классической моделей. При взаимодействии с решеткой электрон приобретает энергию или же передает ее колебаниям решетки, и именно таким путем работа электрических сил $I^2 R$ превращается в энергию тепловых колебаний. При любом отдельном столкновении энергия, теряемая или приобретаемая электроном, однако, очень мала по сравнению с его полной энергией, поэтому при изучении рассеяния электронов этими изменениями энергии можно пренебречь. Таким образом, предположим, что электрон, находящийся вначале в состоянии P_0 , может переходить затем в другие состояния с точно такой же энергией, например в состояние P_j (фиг. 140); из этих состояний он может далее перейти в другие состояния и т. д.

В следующем разделе (стр. 564) будет показано, что возмущающее взаимодействие обуславливает некоторую неопределенность значений энергии электрона после перехода. По этой причине конечные



Фиг. 140. Электрон, находящийся в зоне Бриллюэна в состоянии P_0 , совершает переходы в другие состояния с той же энергией (согласно классической модели).

состояния не будут лежать на сферической поверхности постоянной энергии, а будут распределены в тонком слое вблизи этой поверхности (фиг. 141), причем разброс значений энергии определяется



Фиг. 141. Соотношения между элементом объема, числом конечных состояний и скоростью.

$$ds = \frac{V}{\hbar^3} dV_P = v_j d\mathcal{E} d\Omega_j; \quad dV_P = \frac{d\mathcal{E} d\Omega_j}{v_j}.$$

соотношением $\Delta\mathcal{E} \cdot \tau_0 \approx \hbar$, где τ_0 — время жизни исходного состояния, рассматриваемое ниже. Эта неопределенность значения энергии, однако, столь мала, что в большинстве случаев ею можно пренебречь и считать, что энергия электрона при переходе не изменяется.

Согласно пункту а) предположения I, высказанного нами в гл. 8 (стр. 234), вероятность перехода из какого-либо одного состояния в любое другое состояние не должна зависеть от времени. Мы

придадим теперь этому предположению форму математических уравнений, следствия которых (но не сами уравнения), как это позднее будет показано, эквивалентны результатам, получаемым в квантово-механической теории возмущений. Эквивалентность этих двух методов решения задачи оправдывает возможность использования классической картины столкновений частиц и понятия среднего времени свободного пробега для наглядного описания рассматриваемых процессов, а также при вычислениях.

Следуя методу, развитому в гл. 11, рассмотрим (см. фиг. 140) тонкий слой в пространстве импульсов, ограниченный двумя близкими друг к другу поверхностями постоянной энергии. Поскольку мы условились пренебрегать изменениями энергии электрона, такой слой можно трактовать как некоторую поверхность постоянной энергии и говорить о числе состояний или о числе электронов, приходящихся на единицу площади этой поверхности. Таким образом, мы будем рассматривать переходы из исходного состояния P_0 в группу состояний, близких к P_j и отвечающих элементу поверхности $d\Omega_j$, где $d\Omega_j$ имеет размерность квадрата импульса.

Вероятность перехода за единицу времени пропорциональна площади элемента $d\Omega_i$ и может быть записана в виде

$$W_{0j} d\Omega_j. \quad (17.3)$$

Полная вероятность перехода из состояния P_0 за единицу времени равна соответственно

$$W_0 = \int_{\Omega(\mathcal{E})} W_{0j} d\Omega_j, \quad (17.4)$$

где $\Omega(\mathcal{E})$ — поверхность постоянной энергии.

Вероятности переходов из состояний P_j обозначим аналогично

$$W_{ji} \text{ и } W_j = \int_{\Omega(\mathcal{E})} W_{ji} d\Omega_j. \quad (17.5)$$

Воспользуемся теперь этими определениями вероятностей для статистического описания процесса рассеяния.

Удобно выразить вероятности через числа электронов; для этого предположим, что имеется большое число N тождественных систем, в каждой из которых в момент времени $t=0$ один из электронов находится в состоянии P_0 . Соответственно будем пользоваться следующими определениями:

α_0 — число электронов, оставшихся к моменту t в состоянии P_0 ,

β_j — число электронов, приходящихся на единицу площади поверхности постоянной энергии (см. фиг. 140) вблизи состояния P_j и испытавших однократное рассеяние,

γ_i — число электронов, приходящихся на единицу площади поверхности постоянной энергии вблизи состояния P_i и испытавших двукратное рассеяние, и т. п.

Число электронов, которые подверглись однократному рассеянию и находятся в интервале $d\Omega_j$, и общее число β электронов, испытавших однократное рассеяние, равны соответственно

$$\beta_j d\Omega_j \quad \text{и} \quad \int_{\Omega(\mathcal{E})} \beta_j d\Omega_j.$$

Классификация электронов по числам столкновений, которые они испытали, упрощает сопоставление наших результатов с результатами квантовомеханической теории. Из данных выше определений и предположения о том, что вероятности $W_{j\ell}$ не зависят от времени, очевидно, что скорость уменьшения величины α_0 равна $W_0\alpha_0$. Это дает

$$\frac{d}{dt} \alpha_0 = -W_0 \alpha_0, \quad \alpha_0 = e^{-W_0 t}, \quad W_0 = \frac{1}{\tau_0}, \quad (17.6)$$

где τ_0 — среднее время между столкновениями, введенное в рассмотрение в гл. 11. Скорость изменения числа однократно рассеянных электронов в элементе $d\Omega_j$ равна

$$\begin{aligned} \frac{d}{dt} \beta_j d\Omega_j &= \alpha_0 W_{0j} d\Omega_j - W_j \beta_j d\Omega_j, \\ \dot{\beta}_j &= W_{0j} \alpha_0 - W_j \beta_j; \end{aligned} \quad (17.7)$$

аналогично

$$\dot{\gamma}_\ell = \int W_{j\ell} \beta_j d\Omega_j - W_\ell \gamma_\ell.$$

Эти уравнения представляют собой математическую формулировку предположения о том, что для электрона, находящегося в данном состоянии P_j , имеется определенная вероятность перейти за единицу времени в различные другие состояния P_ℓ , причем эта вероятность не зависит от времени. Этот результат находится в согласии с представлениями, использованными нами выше в гл. 11 при вычислении подвижности. Основная задача настоящей главы заключается в том, чтобы: 1) показать, что квантовая механика приводит к тем же результатам, что и полученные выше уравнения, т. е. показать тем самым, что простая физическая картина, использованная в гл. 8 и 11, верна, и 2) вычислить вероятности W . Первый пункт этой программы будет осуществлен в настоящем параграфе; при этом мы убедимся также в том, что величины W_{ij} удовлетворяют принципу детального равновесия, рассмотренному в гл. 11. В дальнейшем в § 6 нами будут получены формулы, позволяющие выразить значения вероятностей W_{ij} через константы материала.

Квантовая теория переходов. Дадим теперь квантовомеханический вывод приведенных выше формул теории столкновений. Метод, которым мы при этом будем пользоваться, основан на теории расширения спектральных линий, разработанной Вайскопфом и Вигнером.

Теория Вайскопфа — Вигнера дает приближенное уравнение, определяющее скорость перехода атома из исходного возбужденного состояния в нормальное состояние, сопровождающегося излучением кванта света.

В нашей задаче речь идет об электроне, который может совершить переход из исходного состояния в состояние с другим значением импульса, причем такой переход будет сопровождаться испусканием фона. В случае задачи об излучении света конечное состояние определяется путем задания величин, характеризующих испускаемый фотон. В нашем случае мы можем воспользоваться с этой целью конечным значением импульса, так как процесс испускания или поглощения фона характеризуется только изменением его импульса. (В несколько упрощенной форме этот вопрос будет рассмотрен нами также в § 6.) Задача о рассеянии электронов сложнее задачи об излучении света атомом, так как непосредственно вслед за одним из переходов при этом могут иметь место и другие переходы. Можно, однако, показать, что в наиболее интересных для нас случаях все последующие переходы также могут быть исследованы с помощью уравнений (17.6) и (17.7). Эти вопросы будут рассмотрены частично в настоящем параграфе, частично — в конце § 6 в виде задач.

Метод Вайскопфа — Вигнера дает значительно более точные результаты, чем другие обычно применяемые методы определения вероятностей переходов. Общепринятые методы дают только начальное значение скорости перехода, которое является линейной функцией времени и представляет собой, таким образом, два первых члена $1 - t/\tau$ уравнения (17.6). Метод Вайскопфа — Вигнера приводит к экспоненциальной зависимости скорости от времени. При этом он дает те же формулы для определения τ , что и более элементарные теории; читателю, незнакомому с элементарной теорией, полезно изучить ее по какому-либо из учебных руководств.

Исходным соотношением теории Вайскопфа — Вигнера является разложение волновой функции в ряд вида ¹⁾

$$\Phi = a_0 e^{-i\omega_0 t} \Phi_0 + \sum' b_j e^{-i\omega_j t} \Phi_j, \quad (17.8)$$

где коэффициенты являются неизвестными функциями времени, удовлетворяющими условиям

$$a_0 = 1, \quad b_j = 0 \quad \text{при} \quad t = 0 \quad (17.9)$$

и

$$\hbar\omega_0 = \mathcal{E}_0, \quad \hbar\omega_j = \mathcal{E}_j. \quad (17.10)$$

В нашем случае функции Φ_j отличаются от функции Φ_0 тем, что им соответствует на единицу большее или же на единицу меньшее

¹⁾ Штрих у знака суммы в (17.8) означает исключение из рассмотрения состояния $j = 0$ при суммировании. — *Прим. авт.*

число фононов; в случае задачи об излучении света этим функциям отвечает один лишний фотон. Мы будем пренебрегать возможностью переходов из одного из состояний Φ_j в остальные состояния Φ_j , а также переходами, которые сопровождаются испусканием или поглощением большого числа фононов. При этих предположениях дифференциальное уравнение, определяющее изменение коэффициента a_0 , может быть получено из уравнения Шредингера, зависящего от времени,

$$\mathcal{H}\Phi = (\mathcal{H}_0 + \mathcal{U}_1)\Phi = (i\hbar)\dot{\Phi}, \quad (17.11)$$

если мы умножим последнее на Φ_0^* , проинтегрируем по всем координатам и воспользуемся тем, что $\mathcal{H}_0\Phi_0 = \mathcal{E}_0\Phi_0$:

$$i\hbar a_0 = \sum' \mathcal{U}_{0j} e^{i(\omega_0 - \omega_j)t} b_j. \quad (17.12)$$

Здесь \mathcal{U}_{0j} — матричный элемент, соответствующий переходу из состояния 0 в состояние j :

$$\mathcal{U}_{0j} = \int \Phi_0^* \mathcal{U}_1 \Phi_j d(\text{всех координат}). \quad (17.13)$$

Матричные элементы \mathcal{U}_1 , как это будет доказано в дальнейшем, являются эрмитовыми, а матричный элемент \mathcal{U}_{00} , опущенный в формуле (17.12), как будет показано в § 6, равен нулю. [В полном согласии со сказанным выше, из формулы (17.126) следует, что возмущение \mathcal{U}_1 мало по сравнению с тепловыми энергиями электронов.] Аналогичным образом умножение уравнения (17.11) на Φ_j^* и последующее интегрирование по всем координатам дает

$$i\hbar \dot{b}_j = \mathcal{U}_{j0} e^{i(\omega_j - \omega_0)t} a_0. \quad (17.14)$$

Мы покажем, что приближенные решения этих уравнений имеют вид

$$a_0 = e^{-\eta t}, \quad (17.15)$$

$$b_j = \frac{\mathcal{U}_{j0}}{i\hbar} \frac{e^{i(\omega_j - \omega_0)t} - 1}{i(\omega_j - \omega_0) - \eta}, \quad (17.16)$$

где η — постоянная, которая будет вычислена ниже.

Вероятность нахождения электрона в нулевом состоянии будет при этом равна

$$|a_0|^2 = e^{-2\eta t},$$

откуда очевидно, что вероятность перехода и среднее время жизни для состояния Φ_0 равны соответственно

$$W_0 = 2\eta \quad \text{и} \quad \tau_0 = \frac{1}{2\eta}.$$

Записанное выше выражение (17.16) является непосредственным следствием уравнения, определяющего a_0 , и дифференциального урав-

нения (17.14). Уравнение для \dot{a}_0 удовлетворяется при условии

$$\eta = \sum' \frac{u_{0j}u_{j0}}{\hbar^2} \frac{1 - e^{[i(\omega_0 - \omega_j) + \eta]t}}{i(\omega_j - \omega_0) - \eta}, \quad (17.17)$$

как в этом легко убедиться путем непосредственной подстановки. [Как будет показано ниже, (17.17) может быть решено относительно η , при этом мы получим постоянную величину, не зависящую от t .] Заменяем теперь суммирование интегрированием в предположении, что $u_{0j}u_{j0} = |u_{0j}|^2$ является медленно меняющейся функцией конечного состояния j (см. § 6). Плотность конечных состояний равна V/\hbar^3 , поскольку изменения спина не имеют места. В качестве переменных интегрирования мы воспользуемся элементом $d\Omega_j$ поверхности постоянной энергии, а также самой энергией. Элемент объема в \mathbf{P} -пространстве равен

$$dV_{\mathbf{P}} = d\Omega_j g(\mathbf{P}_j) d\mathcal{E},$$

где

$$g(\mathbf{P}_j) = \frac{1}{|V_{\mathbf{P}}(\mathbf{P})|} = \frac{1}{v_j}, \quad v_j = |\mathbf{v}_j|$$

есть расстояние между поверхностями постоянной энергии, соответствующее изменению энергии на единицу¹⁾ (как сказано выше, энергией фонона мы пренебрегаем). Для краткости введем в рассмотрение величину ρ_j (являющуюся функцией \mathbf{P}_j), представляющую собой число конечных состояний ds , приходящихся на единицу энергии и единицу площади; по определению имеем

$$\rho_j d\Omega_j d\mathcal{E} = \frac{V}{\hbar^3 v_j} d\Omega_j d\mathcal{E}, \quad \rho_j = \frac{V}{\hbar^3 v_j}. \quad (17.18)$$

Это соотношение, а также некоторые другие закономерности, связанные с интегрированием по \mathbf{P} -пространству, иллюстрируются фиг. 141.

Интегрирование уравнения, определяющего параметр η , может быть выполнено приближенно. Удобно заменить $d\mathcal{E}$ на $\hbar d\omega_j$; при этом (17.17) приобретает вид

$$\eta = \int \frac{|u_{0j}|^2}{\hbar} \frac{1 - e^{[i(\omega_0 - \omega_j) + \eta]t}}{i(\omega_j - \omega_0) - \eta} \rho_j d\omega_j d\Omega_j. \quad (17.19)$$

Множитель, содержащий разность $(\omega_0 - \omega_j)$, имеет наибольшее значение для таких переходов, при которых энергия остается неизменной; в этом случае он равен $(1 - e^{\eta t})/\eta$. При больших значениях разности $(\omega_0 - \omega_j)$ этот множитель рассматривается обычно как очень малый. Если мы будем считать, что интеграл (17.19) в основном определяется столь узкой областью значений ω_j , близких к ω_0 , что остальные множители подинтегрального выражения практически

¹⁾ Определение $dV_{\mathbf{P}}$ см. подробнее в § 2 гл. 11. — *Прим. авт.*

остаются при этом постоянными, то можно далее воспользоваться соотношениями

$$\lim_{L \rightarrow \infty} \int_{-L}^{+L} \frac{1}{i(\omega_0 - \omega_j) - \eta} d\omega_j = \begin{cases} -\pi, & \eta > 0 \\ +\pi, & \eta < 0 \end{cases}$$

$$\lim_{L \rightarrow \infty} \int_{-L}^{+L} \frac{e^{i(\omega_0 - \omega_j) + \eta} t}{i(\omega_j - \omega_0) - \eta} d\omega_j = \begin{cases} -2\pi, & \eta > 0 \\ 0, & \eta < 0 \end{cases}$$

(см. задачи, где указаны способы вычисления этих интегралов и рассмотрен вопрос об их сходимости, и проинтегрируйте по ω_j). В результате (17.19) приобретает вид

$$\eta = \int_{\mathcal{E}(\mathcal{E}_0)} \frac{|Z_{0j}|^2}{\hbar} \pi \rho_j d\Omega_j, \quad (17.20)$$

причем интегрирование осуществляется здесь по поверхности $\mathcal{E}_j = \mathcal{E}_0$. Таким образом, для всех значений t (17.17) удовлетворяется при одном и том же значении η , так что функции (17.15) и (17.16) дают удовлетворительное решение уравнения Шредингера (17.11), зависящего от времени.

Покажем теперь, что отдельные слагаемые, определяющие величину η , можно рассматривать как вероятности переходов, отнесенные к единице площади поверхности постоянной энергии. С этой целью определим вероятность нахождения электрона в элементе $d\Omega_j$ с энергией $\hbar\omega_j$. Эта вероятность представляет собой сумму соответствующих значений $|b_j|^2$ и равна

$$\begin{aligned} d\Omega_j \int |b_j|^2 \rho_j \hbar d\omega_j &= \\ &= d\Omega_j \int \frac{|Z_{j0}|^2}{\hbar} \frac{i + e^{-2\eta t} - 2e^{-\eta t} \cos(\omega_j - \omega_0) t}{(\omega_j - \omega_0)^2 + \eta^2} d\omega_j \rho_j = \\ &= \frac{1}{2\eta} (1 - e^{-2\eta t}) \frac{|Z_{j0}|^2}{\hbar} 2\pi \rho_j d\Omega_j. \end{aligned} \quad (17.21)$$

В области, существенной для интегрирования по ω_j , мы снова считали здесь остальные переменные постоянными и распространили пределы интегрирования до $\pm \infty$; рассматриваемые интегралы приобретают при этом стандартный вид. Как это следует из фиг. 141, существенной для интегрирования является область, для которой $\Delta\mathcal{E} = |\mathcal{E}_j - \mathcal{E}_0| = \hbar\eta$ соответствует значению $|\omega_j - \omega_0| = \eta$. Этот результат находится в согласии с принципом неопределенностей, записанным в форме $\Delta\mathcal{E} \Delta t = \hbar$, а также с соотношением $\Delta t = 1/\eta = 2\tau_0$, где τ_0 , как будет показано ниже, — время жизни состояния Φ_0 . В тех случаях, которые представляют для нас практический интерес, $\Delta\mathcal{E}$ значительно

меньше \mathcal{E}_0 , так что разброс значений энергии не играет существенной роли.

Если при решении классических уравнений мы пренебрежем учетом вторичных столкновений, чтобы иметь условия, соответствующие только что рассмотренной квантовомеханической задаче, то долю электронов, испытавших однократное рассеяние, можно найти, решив уравнение (17.7), записанное в следующей упрощенной форме:

$$\begin{aligned} \dot{\beta}_j &= W_{0j} \alpha_0 = W_{0j} e^{-W_0 t}. \\ \text{Это дает} \quad \beta_j &= \frac{1}{W_0} (1 - e^{-W_0 t}) W_{0j}, \end{aligned} \quad (17.22)$$

что эквивалентно (17.21), если мы положим

$$W_{0j} = \frac{|U_{j0}|^2}{\hbar} 2\pi\rho_j, \quad (17.23)$$

$$\frac{1}{\tau_0} = W_0 = 2\eta = \int_{\Omega(\mathcal{E}_0)} W_{0j} d\Omega_j. \quad (17.24)$$

Зависимость приближенного решения Вайскопфа — Вигнера от времени совпадает, таким образом, с классической, при условии, что вероятность перехода, отнесенная к единице площади поверхности постоянной энергии, определяется формулой (17.23), а среднее время жизни τ_0 — формулой (17.24).

Принцип детального равновесия. Соотношение (17.23) дает возможность убедиться в том, что для исследуемых здесь столкновений справедлив принцип детального равновесия. Рассматривая тонкий слой, заключенный между двумя поверхностями постоянной энергии (см., например, фиг. 94, стр. 304), можно показать, что в условиях равновесия число электронов, совершающих переходы из элемента объема $d\mathcal{E} d\Omega_i$ в элемент объема $d\mathcal{E} d\Omega_j$, в точности равно числу электронов, совершающих обратные переходы из $d\mathcal{E} d\Omega_j$ в $d\mathcal{E} d\Omega_i$. Ранее (§ 2 гл. 11) было высказано предположение о равенстве величин $\rho_i W_{ij}$ и $\rho_j W_{ji}$. Это равенство вытекает непосредственно из формулы (17.23), так как можно написать

$$\rho_i W_{ij} = \frac{\rho_i |U_{ij}|^2 2\pi\rho_j}{\hbar} = \frac{\rho_i \rho_j U_{ji} U_{ij} 2\pi}{\hbar} = \rho_j W_{ji}. \quad (17.25)$$

Третье из этих выражений, симметричное по отношению к индексам i и j , является следствием эрмитовых свойств U , т. е. того, что $U_{ij} = U_{ji}^*$.

Формулировка принципа детального равновесия с учетом принципа Паули рассматривается в § 7 настоящей главы, а также в гл. 11, где используется результат (17.25).

Критика квантовомеханической теории. Вывод формул (15.23) и (15.24) неудовлетворителен в том отношении, что мы произвольно пренебрегли учетом вторичных столкновений. Изложенная выше теория может быть обобщена таким образом, чтобы вторичные столкновения также были приняты во внимание; при этом, однако, возникают дополнительные трудности, связанные со сходимостью исследуемых выражений. Здесь будут изложены лишь основы этого метода, а читателю предоставляется возможность произвести самостоятельно все связанные с ним математические расчеты в порядке упражнений.

Выше было найдено, что из любого данного состояния, описываемого функцией a_0 , возможны переходы в состояния b_j , причем a_0 с течением времени меняется как $\exp(-\eta t)$. Таким образом, под влиянием взаимодействия с b -состояниями a_0 изменяется согласно уравнению

$$i\hbar\dot{a}_0 = i\hbar\eta a_0.$$

Это означает, что при рассмотрении переходов из любого данного состояния можно учесть влияние других состояний, которые при этом возбуждаются и в которые происходят переходы, заменив в уравнении для $i\hbar\dot{a}_0$ [см. (17.12)] сумму членов $\mathcal{U}_{0j} b_j \exp[i(\omega_0 - \omega_j)t]$ членом $-i\hbar\eta a_0$. Если такой простой прием законен, то можно также учесть и влияние состояний, в которые возможны переходы из состояний Φ_j , записав

$$i\hbar\dot{b}_j = \mathcal{U}_{j0} e^{i(\omega_j - \omega_0)t} a_0 - i\hbar\eta_j b_j. \quad (17.26)$$

Член, содержащий множитель η_j , получается здесь точно таким же путем, каким выше был получен член, содержащий η , с той лишь разницей, что исходным в этом случае является не состояние Φ_0 , а состояние Φ_j . В таком приближении для коэффициентов b_j может быть получено другое решение, типа решения Вайскопфа—Вигнера¹⁾:

$$b_j = \frac{\mathcal{U}_{j0}}{i\hbar} \frac{e^{[i(\omega_j - \omega_0) - \eta_j]t} - e^{-\eta_j t}}{i(\omega_j - \omega_0) - \eta + \eta_j}. \quad (17.27)$$

Можно показать, что оно снова приводит к классическим значениям вероятностей переходов W_{jk} из рассматриваемого состояния Φ_j в другие состояния.

Этот прием может быть затем использован снова, так что в результате будут учтены состояния, возникающие при переходах из состояний Φ_j , и т. д. Таким путем аналогия между квантовой и классической теориями столкновений может быть сделана сколь угодно полной.

Возникает, однако, ряд вопросов, связанных с законностью применения этого метода; эти вопросы мы попытаемся разрешить в конце § 6 (см. задачу 4).

¹⁾ Весьма сходное с тем решением, которое эти авторы используют при рассмотрении двух последовательных процессов излучения. — *Прим. авт.*

Один из них — вопрос о законности замены величины $|\mathcal{U}_{0j}|^2 \rho_j$ постоянной величиной при интегрировании по ω_j — мы обсудим здесь. Областью, существенной для интегрирования, является та, для которой $|\omega_j - \omega_0| \approx \eta = 1/2\tau_0$. Это соответствует изменению энергии электрона на $\delta\mathcal{E} = \hbar/2\tau_0$. Для электрона в кристалле германия при комнатной температуре $\tau_0 \approx 10^{-12}$ сек., так что $\delta\mathcal{E} \approx 10^{-16}$ эрг, т. е. примерно $kT/50$. Таким образом, в области, существенной для интегрирования, энергия электрона изменяется всего на несколько процентов и, следовательно, изменениями величины $|\mathcal{U}_{0j}|^2 \rho_j$ при этом можно пренебречь. (См., однако, задачу 2, где речь идет о других трудностях, возникающих при интегрировании.) Зейтц показал недавно, что требование малости интервала значений $\hbar\omega_j$, по которому производится интегрирование, по сравнению с kT приводит к неравенству вида $\mu > 30 \text{ с.м}^2/\text{в} \cdot \text{сек.}$

Вычисление среднего времени свободного пробега и средней длины свободного пробега. Как будет показано в § 6, квадрат модуля матричного элемента \mathcal{U}_{j0} может быть представлен в виде ¹⁾

$$|\mathcal{U}_{j0}|^2 = \frac{AT}{V}, \quad (17.28)$$

где T — абсолютная температура, V — объем кристалла A — постоянная, зависящая от упругих свойств кристалла, массы атомов, а также от некоторых параметров, характеризующих волновые функции, но не зависящая от \mathbf{P}_0 или же от \mathbf{P}_j . Из этой формулы следует, что $|\mathcal{U}_{j0}|^2$ не зависит от направления движения электрона, так что для электрона, испытавшего рассеяние, все направления импульса равновероятны. Поскольку значения \mathcal{W}_{0j} не зависят от направления \mathbf{P}_j , интегрирование по $d\Omega_j$ в выражении (17.24) в этом случае сводится просто к умножению на $4\pi P_0^2$. Воспользовавшись формулой

$$v_0 = \frac{|\mathbf{P}_0|}{m_n},$$

связывающей скорость электрона v_0 с его импульсом и эффективной массой m_n , можно вычислить (17.24):

$$\begin{aligned} \frac{1}{\tau_0} &= \int_{\Omega(\mathfrak{S}_0)} \mathcal{W}_{0j} d\Omega_j = \frac{2\pi}{\hbar} \rho_j |\mathcal{U}_{j0}|^2 4\pi P_0^2 = \\ &= \frac{2\pi}{\hbar} \left[\frac{V}{\hbar^3} \frac{1}{v_0} \right] \frac{AT}{V} 4\pi P_0^2 = \frac{ATm_n^2}{\pi\hbar^4} v_0 = \frac{v_0}{l(T)}. \end{aligned} \quad (17.29)$$

Эта формула показывает, что вероятность перехода пропорциональна v_0 ; отсюда следует, что средняя длина свободного пробега $l(T)$

¹⁾ Это эффективное значение $|\mathcal{U}_{j0}|^2$ получено с учетом процессов, при которых электрон приобретает энергию, а также процессов, при которых электрон теряет ее. — *Прим. авт.*

не зависит от скорости и равна

$$l(T) = \frac{\pi \hbar^4}{ATm_n^2}. \quad (17.30)$$

Стоит отметить, что независимость средней длины свободного пробега от энергии является непосредственным следствием того обстоятельства, что матричный элемент не зависит от исходного и конечного значений импульса электрона. Средняя длина свободного пробега не зависит от энергии также и в том случае, когда зоны вырождены, при условии, что в любом направлении \mathbf{P} -пространства энергия изменяется пропорционально P^2 . Если последнее условие выполняется, то площади поверхностей постоянной энергии будут изменяться также пропорционально P^2 , т. е. пропорционально энергии \mathcal{E} , а расстояние между этими поверхностями будет меняться пропорционально $\mathcal{E}^{-1/2}$. Следовательно, вероятность перехода будет изменяться как $\mathcal{E} \cdot \mathcal{E}^{-1/2}$, или же как $\mathcal{E}^{1/2}$, т. е. будет пропорциональна скорости. Таким образом, даже в том случае, когда поверхности постоянной энергии носят весьма сложный характер, средняя длина свободного пробега для данного направления движения не зависит от скорости.

Как было показано в гл. 11, при постоянном значении средней длины свободного пробега мы получаем следующее выражение для подвижности:

$$\mu = \frac{4l}{3\sqrt{2\pi kT/m_n}}. \quad (17.31)$$

В последующих разделах данной главы будут исследованы колебания атомов решетки и их взаимодействие с электронами, вычислена величина $|\mathcal{U}_{ij}|^2$, показано, что она действительно может быть представлена в виде AT/V , и определена постоянная A .

Вопрос об эквивалентности законов рассеяния дырок и электронов, а также некоторые смежные вопросы будут рассмотрены в § 7.

ЗАДАЧИ

1. Вычислить интегралы (17.19). Они могут быть вычислены путем интегрирования по контуру в комплексной плоскости. Первый интеграл элементарен, а для вычисления второго следует интегрировать по контуру от $-R$ до $+R$ вдоль вещественной оси и по полуокружности в нижней полуплоскости (см. также [6], стр. 166, задача 15).

2. Оба интеграла, входящие в выражение (17.19), являются несобственными интегралами, и их значения зависят от того, насколько переменная ω_j близка к своим предельным значениям $\pm\infty$. Член, содержащий под знаком интеграла экспоненциальный множитель, дает сходящееся выражение, за исключением того случая, когда $|\mathcal{U}_{0j}|^2$

возрастает быстрее разности $|\omega_0 - \omega_j|$. Второй член может, однако, дать мнимое слагаемое вида

$$\lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\omega_0 - \epsilon} + \int_{\omega_0 + \epsilon}^{\infty} \frac{|K_{0j}|^2}{(\omega_j - \omega_0)} d\omega_j.$$

Сопоставить это выражение с энергией возмущения второго порядка [см., например, [3], формула (25.12)] и показать, что оно вносит поправку, учитывающую среднюю энергию системы в целом, в определение частоты, отвечающей состоянию Φ_0 .

3. Подставить найденные выше значения коэффициентов b_j (17.27) в уравнение (17.12) и показать, что оно при этом удовлетворяется, а параметр η определяется формулой (17.20). Пользуясь этими значениями b_j , вычислить β_j и показать, что выражение, заменяющее при этом (17.21), является точным решением уравнения теории столкновений (17.7), при условии $W_t = 2\eta_i$.

4. Если $\Phi(t)$ есть решение уравнения Шредингера (в отсутствие магнитного поля), то функция $\Phi^*(-t)$ также представляет собой решение этого уравнения. Очевидно, однако, что в выражении $a_0 = e^{-\eta t}$, дающем решение уравнения (17.12), переменная t не должна убывать до $-\infty$. Показать, что при $t=0$, когда все коэффициенты b_i равны нулю, вместо условия $\dot{a}_0 = -\eta$ имеем $\dot{a}_0 = 0$ и, следовательно, при $t=0$ приближение Вайскопфа—Вигнера становится непригодным. Какому решению при отрицательных значениях t приближение Вайскопфа—Вигнера отвечает в действительности? Почему решение Вайскопфа—Вигнера обратимо во времени, тогда как решения уравнений теории столкновений (17.6) и (17.7) не обладают этим свойством?

§ 3. НОРМАЛЬНЫЕ КОЛЕБАНИЯ АТОМОВ КРИСТАЛЛА

Как уже было указано выше (см. фиг. 139), для вычисления матричного элемента, отвечающего переходу, связанному с изменениями электронного и колебательного состояний кристалла, необходимо решить ряд вспомогательных задач. В настоящем параграфе будет исследована задача о движении ядер атомов, причем валентные электроны будут рассматриваться просто как среда, в которой действуют силы между атомными остатками. Оправданием этого метода служит то обстоятельство, что масса ядра велика по сравнению с массой электрона. Вследствие такого различия масс валентные электроны движутся настолько быстрее, чем ядра, что волновые функции, описывающие их движение, имеют такие значения, как если бы ядра были неподвижными. Значения энергии, соответствующие этим волновым функциям, зависят, однако, от координат ядер и изменяются

при движении последних. Изменение энергии электронов, обусловленное движением ядер, может рассматриваться как эффективная потенциальная энергия ядер и в качестве таковой может быть введено в уравнение Шредингера, содержащее координаты ядер. Можно представить себе, что электроны движутся при этом столь быстро, что на ядра действует только некоторая средняя сила, создаваемая электронами; именно такому предположению и отвечает рассмотрение изменения энергии электронов как некоторой эффективной потенциальной энергии ядер¹⁾.

Рассмотрим кристалл, содержащий N_s атомов, и предположим, что на каждую элементарную ячейку кристаллической решетки приходится по одному атому, т. е. что все атомы обладают одинаковыми массами и эквивалентны во всех остальных отношениях. Предположим, далее, что имеют место периодические граничные условия, так что атомы, расположенные в плоскости $x = A_x - a$, где A_x — период, отсчитываемый в направлении x , и a — постоянная решетки, являются смежными с атомами, расположенными в плоскости $x = 0$. Если кристалл подвергается трансляции, то атомы, проходящие через плоскость $x = A_x$, можно представить себе как проходящие тем самым через плоскость $x = 0$, причем это движение не связано с изменением импульса. Можно, однако, представить себе также, что кристалл как целое перемещается за границы исследуемой области от плоскости $x = 0$ до плоскости $x = A_x$, рассматривая при этом потенциальную энергию взаимодействия любых двух ядер как периодическую функцию их координат x_1, y_1, z_1 и x_2, y_2, z_2 , удовлетворяющую уравнению

$$\begin{aligned} \mathcal{V}(x_1 - x_2, y_1 - y_2, z_1 - z_2) = \\ = \mathcal{V}(x_1 - x_2 + n_x A_x, y_1 - y_2 + n_y A_y, z_1 - z_2 + n_z A_z), \end{aligned} \quad (17.32)$$

где A_x, A_y, A_z — периоды, а n_x, n_y, n_z — целые числа.

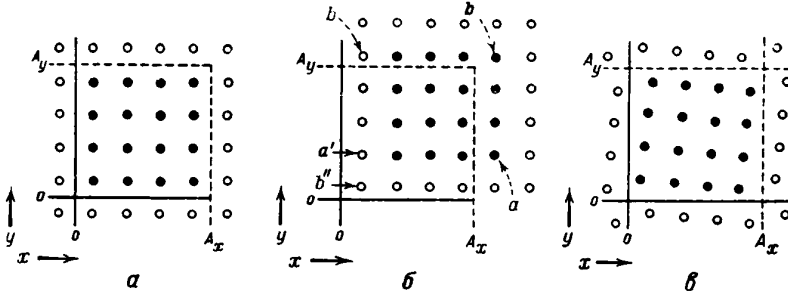
На фиг. 142, б схематически показано расположение атомов кристалла и их периодических „изобразений“. Атомы a и b , находящиеся вне границ рассматриваемой области кристалла, эквивалентны связанным с ними условиями периодичности атомам a' и b'' , расположенным внутри этой области²⁾.

В системе такого рода может возникнуть вращающий момент, так что моменты количества движения не являются постоянными движения. Таким образом, если рассматриваемая область кристалла в целом слегка повернута, как это показано на фиг. 142, в, то „изображения“ атомов, создающиеся вне границ области в силу периодических граничных условий, будут оказывать воздействие на атомы,

¹⁾ Вопрос о разделении координат ядер и электронов рассматривается подробно во многих руководствах. См., например, [7—9]. — *Прим. авт.*

²⁾ Вопрос о роли различных граничных условий рассмотрен в статье Борна [10]. — *Прим. авт.*

находящиеся внутри этой области; возникающие при этом силы будут стремиться повернуть рассматриваемую область кристалла в обратном направлении так, чтобы атомные плоскости снова стали параллельными границам периода. Рассматриваемая модель несколько искусственна; следуя соображениям, приведенным ранее в § 9 гл. 15, можно, однако, представить себе двухмерный кристалл, в котором такие граничные условия будут выполняться. Можно также предполагать, что общий



Ф и г. 142. Влияние периодических граничных условий.

a — нормальное расположение атомов; *b* — расположение атомов после трансляции; *c* — расположение атомов после поворота.

характер результатов, получаемых для такой периодической модели, безусловно сохранился бы и при другой менее искусственной постановке задачи.

Нормальные колебания. Задача настоящего параграфа — получить решение уравнения Шредингера, характеризующее движение атомных ядер. Для описания положений ядер мы будем при этом пользоваться тремя группами векторов: \mathbf{R}_i , \mathbf{R}'_i и $\delta\mathbf{R}_i$. Векторы \mathbf{R} мы определим следующим образом. Для удобства выберем некоторую систему отсчета, в которой ядра расположены идеально правильно, образуя периодическую решетку. Один из атомов этой решетки, называемый нулевым, выберем за начало отсчета. Радиус-векторы, определяющие положения остальных атомов, образуют систему векторов \mathbf{R}_i , так что \mathbf{R}_i есть трансляция решетки. Вектор $\mathbf{i}_x A_x$, длина которого есть целое кратное постоянной решетки, заканчивается в одном из узлов решетки, образуемой векторами \mathbf{R}_i . При наличии возмущения атом i смещается на $\delta\mathbf{R}_i$, так что

$$\delta\mathbf{R}_i = \mathbf{R}'_i - \mathbf{R}_i, \quad (17.33)$$

т. е. смещение атома из положения равновесия равно разности векторов \mathbf{R}'_i и \mathbf{R}_i , определяющих соответственно истинное положение атома и его положение равновесия.

Вводя в рассмотрение радиус-векторы \mathbf{R}'_i и сопряженные им импульсы \mathbf{P}_i , можно записать уравнение Шредингера, описывающее движение ядер, в виде

$$\left[\frac{1}{2M_a} \sum_i \mathbf{P}_i^2 + \mathcal{V}(\mathbf{R}'_1, \dots, \mathbf{R}'_{N_b}) \right] \Phi = \mathcal{E} \Phi, \quad (17.34)$$

где $\mathbf{P}_i = (\hbar/i) \nabla_{\mathbf{R}'_i}$, а индексы указывают номера ядер или атомов. Если амплитуды колебаний малы, то \mathcal{V} может быть разложено в ряд и представлено в квадратичной форме; при этом мы получаем ряд нормальных колебаний, характеризующихся координатами q_i и импульсами p_i . При наличии периодических граничных условий эти колебания могут быть описаны посредством бегущих волн (см. фиг. 55, стр. 171). В трехмерном кристалле каждый из атомов может колебаться в трех направлениях. Для того чтобы найти в этом случае решения, описывающие бегущие волны, исследуем сначала две вспомогательные задачи.

Прежде всего рассмотрим классическую задачу об отыскании нормальных колебаний. Для ее решения удобно воспользоваться рядом величин $\delta \mathbf{R}_i$, представляющих собой смещения атомов из равновесных положений в кристалле, характеризующихся векторами \mathbf{R}_i . Классические уравнения движения, получаемые из уравнений Гамильтона или же из уравнения $\mathbf{F} = m\mathbf{a}$, имеют при этом вид

$$M_a \frac{d^2}{dt^2} \delta \mathbf{R}_i = - \nabla_i \mathcal{V}(\delta \mathbf{R}_1, \delta \mathbf{R}_2 \dots), \quad (17.35)$$

где ∇_i — градиент в пространстве $\delta \mathbf{R}_i$, а \mathcal{V} берется приближенно в виде квадратичной функции переменных $\delta \mathbf{R}_i$. Нормальные колебания для такой системы с учетом влияния периодичности структуры кристалла на форму функции $\mathcal{V}(\delta \mathbf{R}_1, \delta \mathbf{R}_2 \dots)$ исследуются подробно во многих руководствах. Методы решения этой задачи аналогичны методам, применяемым нами в других разделах настоящей главы¹⁾. Нормальные колебания этого типа характеризуются тем, что все атомы совершают простые гармонические колебания с одинаковыми частотами и фазами (изменяется только направление движения, что соответствует разности фаз в 180° , как это показано на фиг. 52, $\kappa = 0$, стр. 164).

Амплитуды нормальных колебаний мы будем обозначать через q с соответствующими индексами. Смещения атомов, если их выразить через нормальные координаты, имеют вид

$$\delta \mathbf{R}_i = \sum_{k,\alpha} \mathbf{G}_{k,\alpha} (q_{k\alpha c} \cos \mathbf{k} \cdot \mathbf{R}_i + q_{k\alpha s} \sin \mathbf{k} \cdot \mathbf{R}_i); \quad (17.36)$$

каждая нормальная координата изменяется при этом согласно простому гармоническому закону, независимо от остальных координат.

¹⁾ Аналогичные задачи рассматриваются в § 6 гл. 14 и § 3 гл. 15. См. также монографии Зейтца [7], Бриллюэна [11] и Вильсона [12]. — *Прим. авт.*

Волновой вектор \mathbf{k} должен удовлетворять периодическим граничным условиям, так что

$$k_x A_x = 2\pi n_x, \quad k_y A_y = 2\pi n_y, \quad k_z A_z = 2\pi n_z, \quad (17.37)$$

где n_x, n_y, n_z — целые числа. Каждому вектору \mathbf{k} отвечают три направления поляризации, для которых индекс α равен соответственно 1, 2, 3. Эти три направления поляризации определяются единичными векторами поляризации $\mathbf{G}_{k\alpha}$. В упруго изотропном кристалле для каждого данного \mathbf{k} один из трех векторов $\mathbf{G}_{k\alpha}$ параллелен \mathbf{k} и описывает продольное колебание, тогда как два других вектора $\mathbf{G}_{k\alpha}$ взаимно перпендикулярны и перпендикулярны вектору \mathbf{k} . Для кубического кристалла, как это будет показано при решении задачи 1, векторы $\mathbf{G}_{k\alpha}$ взаимно перпендикулярны, но уже не характеризуют чисто поперечное и чисто продольное колебания. В этом случае каждой паре значений \mathbf{k} и α отвечают две составляющие колебания; одна из них, обладающая нормальной координатой $q_{k\alpha c}$, изменяется как $\cos \mathbf{k} \cdot \mathbf{r}$, другая, с нормальной координатой $q_{k\alpha s}$ — как $\sin \mathbf{k} \cdot \mathbf{r}$.

Векторы \mathbf{k} связаны с векторами \mathbf{P} , образующими зону Бриллюэна [см. (5.20) — (5.22)] простым соотношением

$$\hbar \mathbf{k} = \mathbf{P}. \quad (17.38)$$

Векторы \mathbf{k} также определяют зону Бриллюэна; ряд значений \mathbf{k} , определяющих зону, мы будем называть *основным рядом векторов* \mathbf{k} . Соотношение между \mathbf{k} и \mathbf{P} будет играть существенную роль в дальнейшем в связи с исследованием электронных переходов, при которых изменения \mathbf{P} сопровождаются возбуждением колебаний решетки. Значения \mathbf{k} для этих колебаний определяются изменениями соответствующих значений \mathbf{P}/\hbar .

Поскольку речь идет о стоячих волнах, описываемых функциями $\cos \mathbf{k} \cdot \mathbf{R}_i$ и $\sin \mathbf{k} \cdot \mathbf{R}_i$, очевидно, что векторы \mathbf{k} и $-\mathbf{k}$ отвечают одним и тем же смещениям. Поэтому при суммировании по возможным значениям \mathbf{k} мы не будем рассматривать векторы, обладающие взаимно противоположными направлениями. (С этой целью можно исключить из рассмотрения все отрицательные значения n_x ; далее, при $n_x = 0$ все отрицательные значения n_y , а при $n_x = n_y = 0$ отрицательные значения n_z . См. задачу 1.) Очевидно также, что на поверхности зоны Бриллюэна значения \mathbf{k} подчиняются соотношениям такого же типа, что и значения \mathbf{P} (см. фиг. 60, а, стр. 186). Для простого кубического кристалла границам зоны Бриллюэна, образуемой векторами \mathbf{k} , соответствуют значения $|k_x|$, $|k_y|$ и $|k_z|$, равные π/a , где a — постоянная решетки. Каждому данному вектору \mathbf{k} отвечают три возможных направления единичного вектора $\mathbf{G}_{k\alpha}$, для которых α принимает значения 1, 2, 3. Частоты нормальных колебаний определяются величиной \mathbf{k} и направлением поляризации; мы будем обозначать их соответственно через $\omega_{k\alpha}$. Колебаниям по закону синуса и косинуса отвечает одно и то же значение частоты $\omega_{k\alpha}$. Методы получения этих, а также

других результатов решения задачи о нормальных колебаниях будут рассмотрены в конце данного параграфа.

Особый интерес представляют три нормальных колебания, соответствующие условию $k=0$ и $\omega_{0,\alpha}=0$. Для одного из них \mathbf{Q}_{01} можно положить равным \mathbf{i}_x — единичному вектору направления x . При $k=0$ член, содержащий синус, обращается в нуль, а член, содержащий косинус, дает смещение, равное $\delta\mathbf{R}_i = \mathbf{i}_x q_{01}$ и соответствующее трансляции кристалла. Эти три нормальных колебания описывают движение центра тяжести кристалла и могут быть представлены в форме

$$\mathbf{R}_c = \mathbf{i}_x q_{01} + \mathbf{i}_y q_{02} + \mathbf{i}_z q_{03}, \quad (17.39)$$

где

$$\mathbf{R}_c = \frac{\sum \delta\mathbf{R}_i}{N_g},$$

причем суммирование производится здесь по всем ядрам, так как смещения, обусловленные каждым из колебаний, для которых $k \neq 0$, удовлетворяют условию $\sum \delta\mathbf{R}_i = 0$.

Полная система нормальных колебаний соответствует преобразованию координат вида

$$x_i = \sum a_{ij} q_j,$$

где x_i — координата ядра, например y -компонента вектора $\delta\mathbf{R}_i$, а q_j — координата, характеризующая одно из нормальных колебаний. Функция Гамильтона в нормальных координатах имеет вид

$$\mathcal{H} = \frac{1}{2M_c} \mathbf{P}_c^2 + \sum \frac{1}{2M} (p_i^2 + M^2 \omega_i^2 q_i^2), \quad (17.40)$$

где

$$M_c = N_g M_a \text{ — масса кристалла,} \quad (17.41)$$

$$M = \frac{M_a}{2}, \quad (17.42)$$

$$\mathbf{P}_c = \sum \mathbf{P}_i. \quad (17.43)$$

К массе каждого из ядер мы добавили массу заполненной электронной оболочки и валентных электронов, следующих за ядром при его движении. Таким образом, M_c в действительности есть масса всего кристалла. Вектор \mathbf{P}_c представляет собой импульс кристалла в целом и является сопряженным \mathbf{R}_c . Суммирование остальных членов производится по всем значениям $q_{k\alpha c}$ и $q_{k\alpha g}$, для которых $k \neq 0$.

В дальнейшем мы будем записывать функцию Гамильтона в следующей сокращенной форме:

$$\mathcal{H} = \frac{1}{2M_c} \mathbf{P}_c^2 + \sum \mathcal{H}_i, \quad (17.44)$$

где i — совокупность значений ka для стоячей волны, так что

$$\mathcal{H}_i = \frac{1}{2M} [p_{ic}^2 + p_{is}^2 + M^2 \omega_i^2 (q_{ic}^2 + q_{is}^2)]. \quad (17.45)$$

О координатах q мы будем говорить при этом, как о нормальных координатах для стоячих волн (сокращенно q_{std})¹⁾.

Преобразование к координатам q и R_c представляет собой каноническое преобразование как в классическом, так и в квантовомеханическом смысле (см. задачи 2 и 3), так что задача об отыскании собственных значений и волновой функции эквивалентна задаче о системе $3(N_s - 1)$ независимых гармонических осцилляторов с центром тяжести R_c .

Следующие три примера дополняют полученные выше результаты.

ЗАДАЧИ

1. Задача заключается в доказательстве того, что функции (17.36) описывают нормальные колебания атомов простой кубической решетки, где в недеформированном состоянии каждый из атомов находится в центре симметрии кристалла. Исследование этого упрощенного случая даст возможность выяснить некоторые общие особенности задачи о нормальных колебаниях; эта задача рассмотрена в более общем виде в работах [7—9].

Предположим, что положением равновесия одного из атомов служит начало координат, а равновесные положения остальных атомов определяются векторами решетки R_i . Смещение атома i из его положения равновесия R_i равно

$$(x_i, y_i, z_i) = \delta R_i.$$

Атом, положение равновесия которого есть $R_i + R_s$, смещается на

$$(x_{i+s}, y_{i+s}, z_{i+s}) = \delta R_{i+s}.$$

При малых смещениях сила F_i , действующая на атом i , будет являться линейной функцией смещения. Показать, что x -компонента этой силы F_{iwx} имеет вид

$$\begin{aligned} -F_{iwx} &= \sum_s \left[x_{i+s} \frac{\partial^2 V}{\partial x_i \partial x_{i+s}} + y_{i+s} \frac{\partial^2 V}{\partial x_i \partial y_{i+s}} + z_{i+s} \frac{\partial^2 V}{\partial x_i \partial z_{i+s}} \right] = \\ &= \sum_s \left[(x_{i+s} - x_i) \frac{\partial^2 V}{\partial x_i \partial x_{i+s}} + (y_{i+s} - y_i) \frac{\partial^2 V}{\partial x_i \partial y_{i+s}} + \right. \\ &\quad \left. + (z_{i+s} - z_i) \frac{\partial^2 V}{\partial x_i \partial z_{i+s}} \right]. \end{aligned}$$

¹⁾ Индекс „std“ представляет собой сокращение английского слова „standing“, что значит „стоячая“. — Прим. ред.

Второе из этих равенств следует из уравнения

$$x_i \sum_s \frac{\partial^2 V}{\partial x_i \partial x_{i+s}} + y_i \sum_s \frac{\partial^2 V}{\partial x_i \partial y_{i+s}} + z_i \sum_s \frac{\partial^2 V}{\partial x_i \partial z_{i+s}} = 0,$$

показывающего, что трансляция кристалла как целого не сопровождается возникновением сил, стремящихся возвратить атомы в их положения равновесия.

Показать, что функция вида

$$\delta \mathbf{R}_i = \mathbf{G} e^{i\mathbf{k} \cdot \mathbf{R}_i + i\omega t} \equiv \mathbf{G} e^{i\theta_i}$$

удовлетворяет уравнениям движения, если

$$\begin{aligned} -M_a \ddot{x}_i &= \omega^2 M_a G_x e^{i\theta_i} = -F_{i\omega} = e^{i\theta_i} G_x \sum_s (e^{i\mathbf{k} \cdot \mathbf{R}_s} - 1) \frac{\partial^2 V}{\partial x_i \partial x_{i+s}} + \\ &+ e^{i\theta_i} G_y \sum_s (e^{i\mathbf{k} \cdot \mathbf{R}_s} - 1) \frac{\partial^2 V}{\partial x_i \partial y_{i+s}} + e^{i\theta_i} G_z \sum_s (e^{i\mathbf{k} \cdot \mathbf{R}_s} - 1) \frac{\partial^2 V}{\partial x_i \partial z_{i+s}} = \\ &= e^{i\theta_i} [M_{xx} G_x + M_{xy} G_y + M_{xz} G_z] \end{aligned}$$

и если аналогичные уравнения имеют место также и для G_y и G_z . Исходя из условия периодичности кристаллической решетки и предположения о том, что каждый из атомов находится в центре симметрии решетки, показать, что

$$\frac{\partial^2 V}{\partial x_i \partial y_{i+s}} = \frac{\partial^2 V}{\partial x_i \partial y_{i-s}} = \frac{\partial^2 V}{\partial y_i \partial x_{i+s}} \text{ и т. д.}$$

и, следовательно, что величины

$$M_{xy} = M_{yx}, \quad M_{xz} = M_{zx}, \quad M_{yz} = M_{zy}$$

являются вещественными функциями \mathbf{k} . Уравнения движения приобретают, таким образом, вид

$$\begin{aligned} \omega^2 M_a G_x &= M_{xx} G_x + M_{xy} G_y + M_{xz} G_z, \\ \omega^2 M_a G_y &= M_{yx} G_x + M_{yy} G_y + M_{yz} G_z, \\ \omega^2 M_a G_z &= M_{zx} G_x + M_{zy} G_y + M_{zz} G_z. \end{aligned}$$

Эти уравнения эквивалентны уравнению

$$\omega^2 M_a \mathbf{G} = \mathbf{M} \mathbf{G},$$

где \mathbf{M} — матрица, компонентами которой служат величины M_{xx} , M_{xy} и т. д. Матрица \mathbf{M} — эрмитова, так как $M_{xy} = M_{yx}$ и $M_{yx} = M_{xy}^*$, поскольку все эти величины вещественны. Все собственные значения $\omega^2 M_a$ положительны, как это можно показать, вычислив потенциальную энергию. Рассмотрите смещение

$$\delta \mathbf{R}_i = \text{Re}(q \mathbf{G} e^{i\mathbf{k} \cdot \mathbf{R}_i}). \quad (\text{A})$$

При таком смещении на атом i действует сила F_i , x -компонента которой имеет вид

$$F_{ix} = -(\operatorname{Re} e^{i\mathbf{k} \cdot \mathbf{R}_i}) [M_{xx}G_x + M_{xy}G_y + M_{xz}G_z] = \\ = -q \cos \mathbf{k} \cdot \mathbf{R}_i [M_{xx}G_x + M_{xy}G_y + M_{xz}G_z].$$

Энергия, необходимая для увеличения q на dq , равна

$$-\sum F_i \cdot d(\delta \mathbf{R}_i) = (\sum \cos^2 \mathbf{k} \cdot \mathbf{R}_i) q dq \cdot \{M_{xx}G_x^2 + M_{xy}G_xG_y + \\ + M_{xz}G_xG_z + M_{yx}G_yG_x + M_{yy}G_y^2 + M_{yz}G_yG_z + \\ + M_{zx}G_zG_x + M_{zy}G_zG_y + M_{zz}G_z^2\},$$

так что энергия, накапливаемая в результате смещения (A), за исключением тех случаев, когда $\mathbf{k} = 0$ и $\mathbf{k} = (\pi/a, \pi/a, \pi/a)$, равна

$$q^2 \frac{1}{4} N_s (\mathbf{G} | M | \mathbf{G}),$$

где $(\mathbf{G} | M | \mathbf{G})$ — квадратичная форма компонент \mathbf{G} . Получить также соотношения

$$\sum_i \cos(\mathbf{k} \cdot \mathbf{R}_i) \cos(\mathbf{k}' \cdot \mathbf{R}_i) = \sum_i \sin(\mathbf{k} \cdot \mathbf{R}_i) \sin(\mathbf{k}' \cdot \mathbf{R}_i) = \frac{N_s}{2} \delta_{\mathbf{k}, \mathbf{k}'},$$

которые понадобятся в дальнейшем и справедливы, если \mathbf{k} отлично от нуля и не равно $(\pi/a, \pi/a, \pi/a)$. При $\mathbf{k} = 0$ и $\mathbf{k} = (\pi/a, \pi/a, \pi/a)$ первая из этих сумм равна $N_s \delta_{\mathbf{k}, \mathbf{k}'}$, тогда как вторая обращается в нуль. Показать также, что все смешанные члены, содержащие произведение синуса на косинус, равны нулю.

Поскольку предполагается, что положению равновесия (все q равны нулю) отвечает минимальное значение энергии, очевидно, что $(\mathbf{G} | M | \mathbf{G})$ — положительная, определенная, квадратичная форма, так что собственные значения ω^2 положительны. Так как коэффициенты при M вещественны, собственные векторы (G_x, G_y, G_z) могут быть выбраны вещественными.

Исходя из свойств эрмитовой формы, доказать, что три собственных вектора \mathbf{G} , которые мы будем обозначать через \mathbf{G}_α или через $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$, взаимно перпендикулярны. Доказать также, что в случае смещения

$$\delta \mathbf{R}_i \equiv \mathbf{G}_1 (q_{1c} \cos \mathbf{k} \cdot \mathbf{R}_i + q_{1s} \sin \mathbf{k} \cdot \mathbf{R}_i) + \mathbf{G}_2 (q_{2c} \cos \mathbf{k} \cdot \mathbf{R}_i + q_{2s} \sin \mathbf{k} \cdot \mathbf{R}_i) + \\ + \mathbf{G}_3 (q_{3c} \cos \mathbf{k} \cdot \mathbf{R}_i + q_{3s} \sin \mathbf{k} \cdot \mathbf{R}_i)$$

потенциальная энергия равна

$$\frac{1}{4} N_s M_\alpha [\omega_1^2 (q_{1c}^2 + q_{1s}^2) + \omega_2^2 (q_{2c}^2 + q_{2s}^2) + \omega_3^2 (q_{3c}^2 + q_{3s}^2)],$$

а кинетическая энергия равна

$$\frac{1}{4} N_s M_a [\dot{q}_{1c}^2 + \dot{q}_{1s}^2 + \dot{q}_{2c}^2 + \dot{q}_{2s}^2 + \dot{q}_{3c}^2 + \dot{q}_{3s}^2].$$

Показать, что при $\mathbf{k} = 0$ синусоидальные члены должны быть опущены и $\omega = 0$, так что в качестве координат при этом могут быть выбраны q_{01} , q_{02} , q_{03} . Это дает

$$\delta R_i = i_x q_{01} + i_y q_{02} + i_z q_{03},$$

где i_x , i_y , i_z — единичные векторы, отложенные вдоль осей x , y , z .

Подсчитать общее число нормальных колебаний и показать, что при учете движения центра тяжести кристалла оно равно $3N_s$, что соответствует трем степеням свободы на атом. Выяснить, чем отличаются друг от друга те случаи, когда число атомов, расположенных вдоль граней кристалла, является соответственно четным и нечетным, и показать, что все состояния, отвечающие поверхности зоны Бриллюэна, эквивалентны; см. фиг. 60, а и 55, z (стр. 186 и 171).

Отсюда следует, что $3N_s$ координат ядер могут быть выражены через $3N_s$ нормальных координат, причем формулы, определяющие кинетическую и потенциальную энергию, приобретают вид

$$E_{\text{кин.}} = \frac{M_a}{2} (\dot{q}_{01}^2 + \dot{q}_{02}^2 + \dot{q}_{03}^2) + \frac{M}{2} \sum_{\mathbf{k} \neq 0} (\text{все } \dot{q}^2),$$

$$E_{\text{пот.}} = \frac{M}{2} \sum_{\mathbf{k} \neq 0} \sum_{\alpha} \omega_{\mathbf{k}\alpha}^2 (q_{\mathbf{k}\alpha s}^2 + q_{\mathbf{k}\alpha c}^2),$$

где $M_a = N_s M_a$ — масса кристалла, а $M = M_a/2$.

2. Показать, что уравнение Шредингера будет записано в нормальных координатах, если заменить в нем \dot{q} на p/M и воспользоваться тем, что $p = (\hbar/i) (\partial/\partial q)$.

Основным при этом является, конечно, предположение о том, что для ядер $\mathbf{P}_i = (\hbar/i) \nabla_{\mathbf{R}_i}$. Следует доказать, что преобразование от координат \mathbf{R}_i к координатам q приводит к правильному определению оператора кинетической энергии, выраженного через переменные q . В этом можно убедиться, показав, что

$$\frac{1}{M_a} \sum_i \nabla_{\mathbf{R}_i}^2 = \frac{1}{M_c} \left[\frac{\partial^2}{\partial q_{01}^2} + \frac{\partial^2}{\partial q_{02}^2} + \frac{\partial^2}{\partial q_{03}^2} \right] + \frac{1}{M} \sum_{\mathbf{k} \neq 0} \frac{\partial^2}{\partial q^2}.$$

Доказать это в более общей форме, показав, что если вещественное линейное преобразование A от одной группы координат q_i к другой группе координат Q_j имеет вид

$$Q_i = \sum_j a_{ij} q_j,$$

причем диагональная форма выражения, определяющего кинетическую энергию, сохраняется:

$$\sum_i M_i \dot{Q}_i^2 = \sum_j m_j \dot{q}_j^2,$$

то преобразование B , определяемое соотношениями

$$b_{ij} = \sqrt{\frac{M_i}{m_j}} a_{ij},$$

обладает тем свойством, что

$$\sum_i b_{ik} b_{ij} = \delta_{kj},$$

так что, поскольку коэффициенты b_{ij} вещественны, B унитарно (см. задачу 3). Из унитарности преобразования B следует, что

$$\sum_i b_{ik} b_{ji} = \delta_{ij}.$$

Принимая это во внимание, доказать, что

$$\sum_i \frac{1}{M_i} \frac{\partial^2}{\partial Q_i^2} = \sum_j \frac{1}{m_j} \frac{\partial^2}{\partial q_j^2},$$

откуда следует, что в нормальных координатах уравнение Шредингера сохраняет обычную форму.

Доказать, что классические импульсы $P_i = M_i \dot{Q}_i$ и $p_j = m_j \dot{q}_j$ преобразуются согласно уравнениям

$$P_i = \sum_j c_{ij} p_j, \quad \text{где } c_{ij} = \frac{M_i}{m_j} a_{ij} = \sqrt{\frac{M_i}{m_j}} b_{ij},$$

и что это преобразование является контактным, так как

$$\sum_i P_i dQ_i = \sum_j p_j dq_j.$$

Показать, что операторы $(\hbar/i) \partial/\partial Q_i$ и $(\hbar/i) \partial/\partial q_i$ преобразуются точно таким же путем, т. е.

$$\frac{\hbar}{i} \frac{\partial}{\partial Q_i} = \sum_j c_{ij} \frac{\hbar}{i} \frac{\partial}{\partial q_j}.$$

Показать, что если операторы p_i , q_i , удовлетворяющие соотношениям коммутативности

$$(p_i q_j - q_j p_i) = \frac{i}{\hbar} \delta_{ij},$$

преобразуются согласно уравнениям

$$P_i = \sum_j c_{ij} p_j \quad \text{и} \quad Q_i = \sum_j a_{ij} q_j,$$

то операторы P_i и Q_i удовлетворяют аналогичным соотношениям коммутативности.

3. Установить свойства унитарного преобразования от n переменных x_j к n переменным y_i . Пусть преобразование A имеет вид

$$y_i = \sum_j a_{ij} x_j, \quad \text{или сокращенно } y = Ax, \quad (\text{A})$$

причем все величины здесь могут быть комплексными. Если преобразование удовлетворяет требованию

$$\sum_i y_i^* y_i = \sum_j x_j^* x_j,$$

то говорят, что оно унитарно. Исходя из этого, показать, что

$$\sum_i a_{ij}^* a_{ik} = \delta_{jk},$$

так что матрица обратного преобразования A^{-1} имеет вид

$$(A^{-1})_{ji} = a_{ij}^*, \quad \text{где } A^{-1}y = x.$$

Показать, что если преобразование A унитарно, то этим свойством обладает также и A^{-1} и, следовательно, коэффициенты a_{ij} также удовлетворяют условию

$$\sum_i a_{ji}^* a_{ki} = \delta_{jk}.$$

§ 4. ВЗАИМОДЕЙСТВИЕ АТОМНЫХ КОЛЕБАНИЙ С ИЗБЫТОЧНЫМ ЭЛЕКТРОНОМ

В этом параграфе будут рассмотрены явления, связанные с введением в кристалл одного лишнего электрона сверх тех, которые осуществляют валентные связи. Потребуем, чтобы волновая функция этого электрона удовлетворяла периодическим граничным условиям, характеризующимся теми же периодами A_x, A_y, A_z , что и граничные условия для атомных ядер. Избыточный электрон взаимодействует с ядрами и с валентными электронами. Как и в предыдущем параграфе, предположим, что состояния валентных электронов определяются пространственным расположением ядер. В таком случае энергия взаимодействия может быть представлена в форме $\mathcal{U}(\mathbf{r}_e, \delta\mathbf{R}_1, \dots)$, где \mathbf{r}_e — радиус-вектор электрона, а $\delta\mathbf{R}_i$ — координаты ядер [см. (17.33)]. Поскольку кристалл обладает свойством периодичности, \mathcal{U} должно удовлетворять уравнению

$$\mathcal{U}(\mathbf{r}_e + \mathbf{A}, \delta\mathbf{R}_1, \dots, \delta\mathbf{R}_{N_g}) = \mathcal{U}(\mathbf{r}_e, \delta\mathbf{R}_1, \dots, \delta\mathbf{R}_{N_g}), \quad (17.46)$$

где \mathbf{A} — любой вектор, компоненты которого являются целыми кратными периодам A_x, A_y, A_z . Вводя в рассмотрение нормальные координаты и координату центра тяжести кристалла \mathbf{R}_c , можно записать

$$\begin{aligned} \mathcal{U} &= \mathcal{U}(\mathbf{r}_e - \mathbf{R}_c, q_1, q_2, \dots) = \mathcal{U}_0(\mathbf{r}_e - \mathbf{R}_c) + \\ &+ \mathcal{U}_1(\mathbf{r}_e - \mathbf{R}_c, q_1, q_2, \dots). \end{aligned} \quad (17.47)$$

В этом можно убедиться следующим образом. Если q_1, q_2, \dots принимают значения, отвечающие такой деформации решетки, при которой центр тяжести кристалла остается неподвижным, так что $\mathbf{R}_c = 0$, то потенциальная энергия будет иметь вид: $\mathcal{U}(\mathbf{r}_e, q_1, q_2, \dots)$. Если $\mathbf{R}_c \neq 0$, то все распределение потенциала смещается на \mathbf{R}_c ; это можно учесть путем замены \mathbf{r}_e на $(\mathbf{r}_e - \mathbf{R}_c)$. Член \mathcal{U}_0 соответствует идеальной решетке и не должен зависеть от координат q_1, q_2, \dots . Член \mathcal{U}_1 обращается в нуль, когда все q_1, q_2, \dots равны нулю, и характеризует влияние деформаций; именно этот член определяет вероятности переходов системы из одних состояний в другие. При вычислениях мы будем рассматривать только те слагаемые \mathcal{U}_1 , которые линейны относительно q_1, q_2, \dots . (Поскольку электрон, находящийся в точке \mathbf{r}_e , повидимому, не обладает минимумом потенциальной энергии, разложение функции \mathcal{U}_1 в ряд будет содержать также и линейные члены.)

При таком выборе энергии взаимодействия электрона с решеткой оператор \mathcal{H} приобретает вид

$$\mathcal{H} = \frac{1}{2M_c} \mathbf{P}_c^2 + \sum_i \mathcal{H}_i + \frac{1}{2m_e} \mathbf{p}_e^2 + \mathcal{U}(\mathbf{r}_e - \mathbf{R}_c, q_1, q_2, \dots), \quad (17.48)$$

где m_e — истинная, а не эффективная масса электрона. Так как \mathcal{H} содержит теперь \mathbf{R}_c , то \mathbf{P}_c уже не коммутирует с \mathcal{H} и не представляет постоянную движения. Однако поскольку \mathbf{r}_e и \mathbf{R}_c входят в выражение для \mathcal{H} только в виде разности $\mathbf{r} = \mathbf{r}_e - \mathbf{R}_c$, то вместо \mathbf{r}_e и \mathbf{R}_c можно ввести в рассмотрение две новые координаты, а вместо \mathbf{P}_c и \mathbf{p}_e — новые импульсы, произведя с этой целью следующие преобразования:

$$\mathbf{p}_T = \mathbf{p}_e + \mathbf{P}_c, \quad \mathbf{p} = \frac{M_c}{M_T} \mathbf{p}_e - \frac{m_e}{M_T} \mathbf{P}_c, \quad (17.49)$$

$$\mathbf{r}_T = \frac{m_e}{M_T} \mathbf{r}_e + \frac{M_c}{M_T} \mathbf{R}_c, \quad \mathbf{r} = \mathbf{r}_e - \mathbf{R}_c, \quad (17.50)$$

$$M_T = M_c + m_e. \quad (17.51)$$

Здесь \mathbf{p}_T и \mathbf{r}_T , очевидно, импульс и координата центра тяжести системы в целом, т. е. кристалла, содержащего избыточный электрон. Как легко убедиться, при таком преобразовании \mathcal{H} приобретает вид

$$\mathcal{H} = \frac{1}{2M_T} \mathbf{p}_T^2 + \sum_i \mathcal{H}_i + \frac{1}{2m} \mathbf{p}^2 + \mathcal{U}(\mathbf{r}, q_1, q_2, \dots), \quad (17.52)$$

где m — приведенная масса:

$$m = \frac{m_e M_c}{m_e + M_c} = \frac{m_e M_c}{M_T}. \quad (17.53)$$

Обратное преобразование определяется соотношениями

$$\mathbf{p}_e = \frac{m_e}{M_T} \mathbf{p}_T + \mathbf{p}, \quad \mathbf{P}_e = \frac{M_e}{M_T} \mathbf{p}_T - \mathbf{p}, \quad (17.54)$$

$$\mathbf{r}_e = \mathbf{r}_T + \frac{M_e}{M_T} \mathbf{r}, \quad \mathbf{R}_e = \mathbf{r}_T - \frac{m_e}{M_T} \mathbf{r}. \quad (17.55)$$

Читатель может убедиться в том, что, например, для x -составляющих этих векторов удовлетворяются условия задачи 2, рассмотренной в § 3, так что \mathbf{p} и \mathbf{p}_T могут быть заменены соответственно операторами $(\hbar/i)\nabla_{\mathbf{r}}$ и $(\hbar/i)\nabla_{\mathbf{r}_T}$.

Выражение, определяющее \mathcal{H} , удобно записать в виде суммы четырех членов

$$\mathcal{H} = \mathcal{H}_{\mathbf{r}_T} + \mathcal{H}_{\mathbf{r}} + \sum_i \mathcal{H}_i + \mathcal{U}_1(\mathbf{r}, q_1, q_2, \dots), \quad (17.56)$$

где слагаемые \mathcal{H}_i отвечают нормальным колебаниям [см. (17.45)]; новые операторы $\mathcal{H}_{\mathbf{r}_T}$ и $\mathcal{H}_{\mathbf{r}}$ имеют вид

$$\mathcal{H}_{\mathbf{r}_T} = \frac{1}{2M_T} \mathbf{p}_T^2, \quad (17.57)$$

$$\mathcal{H}_{\mathbf{r}} = \frac{1}{2m} \mathbf{p}^2 + \mathcal{U}_0(\mathbf{r}). \quad (17.58)$$

За исключением члена \mathcal{U}_1 , каждый член выражения (17.56) зависит только от одной пары взаимно сопряженных координат и импульсов. В таких случаях говорят, что переменные разделяются; решения уравнения

$$(\mathcal{H} - \mathcal{U}_1)\Phi = \mathcal{E}\Phi$$

при этом могут быть получены в виде произведений собственных функций операторов $\mathcal{H}_{\mathbf{r}_T}$, $\mathcal{H}_{\mathbf{r}}$ и операторов \mathcal{H}_i . Этот прием лежит в основе вычисления матричных элементов переходов, осуществляемого в § 6. В данном параграфе мы будем иметь дело с волновыми функциями, зависящими от координат \mathbf{r}_T и \mathbf{r} ; в § 5 будут исследованы волновые функции, зависящие от координат q_1, q_2, \dots , а в § 6 — волновая функция Φ системы в целом.

Рассмотрим, таким образом, волновую функцию вида $\chi(\mathbf{r}_T)\psi(\mathbf{r})$, удовлетворяющую уравнению

$$(\mathcal{H}_{\mathbf{r}_T} + \mathcal{H}_{\mathbf{r}})\chi(\mathbf{r}_T)\psi(\mathbf{r}) = \mathcal{E}\chi(\mathbf{r}_T)\psi(\mathbf{r})$$

и соответствующим граничным условиям.

Следуя обычному методу разделения переменных, разделим это уравнение на $\chi\psi$; принимая во внимание, что оператор $\mathcal{H}_{\mathbf{r}_T}$ не действует на переменную \mathbf{r} , от которой зависит функция ψ , а

оператор \mathcal{H}_T не действует на координату \mathbf{r}_T , функцией которой является χ , получаем далее

$$\frac{\mathcal{H}_{\mathbf{r}_T}\chi}{\chi} + \frac{\mathcal{H}_T\psi}{\psi} = \mathcal{E}.$$

Отсюда следует, что оба члена левой части этого равенства остаются постоянными (не зависят от \mathbf{r} и \mathbf{r}_T), и мы имеем

$$\mathcal{H}_{\mathbf{r}_T}\chi = \mathcal{E}'\chi, \quad \mathcal{H}_T\psi = \mathcal{E}''\psi$$

и

$$\mathcal{E}' + \mathcal{E}'' = \mathcal{E}.$$

Очевидно, что \mathbf{p}_T коммутирует с $\mathcal{H}_{\mathbf{r}_T}$, так что \mathbf{p}_T — постоянная движения. Это приводит к следующим соотношениям:

$$\chi = V^{-1/2} e^{i\mathbf{p}_T \cdot \mathbf{r}_T / \hbar}, \quad \mathcal{E}' = \frac{\mathbf{p}_T^2}{2M_T}, \quad (14.59)$$

где $V = A_x A_y A_z$ — нормирующий множитель, так что интеграл от $\chi^* \chi dV_{\mathbf{r}_T}$, взятый по всему объему, равен единице.

Уравнение, которому удовлетворяет функция ψ , имеет точно такой же вид, как и уравнение для электрона, движущегося в заданном поле $\mathcal{U}_0(\mathbf{r})$, с той лишь разницей, что вместо m_e оно содержит m . Поскольку

$$m = \frac{m_e M_e}{M_e + m_e} = \frac{m_e}{[1 + (m_e/M_e)]},$$

различие между m и m_e ничтожно. Решения этого уравнения могут быть, следовательно, представлены в виде обычных функций Блоха $\psi_{\mathbf{p}}(\mathbf{r})$, где

$$\mathcal{H}_T \psi_{\mathbf{p}}(\mathbf{r}) = \mathcal{E}(\mathbf{p}) \psi_{\mathbf{p}}(\mathbf{r}), \quad (17.60)$$

$$\psi_{\mathbf{p}}(\mathbf{r}) = V^{-1/2} e^{i\mathbf{p} \cdot \mathbf{r} / \hbar} u_{\mathbf{p}}(\mathbf{r}), \quad (17.61)$$

причем множитель $u_{\mathbf{p}}(\mathbf{r})$ нормирован к единице объема и обладает периодичностью неискаженной решетки.

Граничные условия для переменных \mathbf{r}_T и \mathbf{r} могут быть получены из граничных условий, которым удовлетворяют избыточный электрон и ядра. При перемещении любой из этих частиц на расстояние, равное одному из периодов A_x , A_y , A_z , волновая функция должна оставаться неизменной, так как новая конфигурация частиц при этом эквивалентна старой.

Граничные условия для функции $\chi(\mathbf{r}_T)$ могут быть найдены следующим образом. Если все частицы одновременно перемещаются на расстояние, равное одному из трех периодов A_x , A_y , A_z , то при этом аналогичным образом изменяется также и \mathbf{r}_T , тогда как значения \mathbf{r} и q_1, q_2, \dots остаются неизменными. Очевидно поэтому, что

функция $\chi(\mathbf{r}_T)$ должна удовлетворять периодическим граничным условиям и, следовательно, векторы \mathbf{P}_T должны иметь вид

$$\mathbf{P}_T = \hbar \left[i_x \frac{2\pi n_x}{A_x} + i_y \frac{2\pi n_y}{A_y} + i_z \frac{2\pi n_z}{A_z} \right], \quad (17.62)$$

где n_x, n_y, n_z — целые числа. Эти векторы определяют зону Бриллюэна. Они не заканчиваются, однако, на поверхности зоны, и энергия $\mathbf{P}_T^2/2M_T$ неограниченно возрастает с увеличением $|\mathbf{P}_T|$.

В дальнейшем мы будем иметь дело также и с другими векторами типа (17.62), длина которых не ограничена размерами зоны Бриллюэна. Мы будем называть их *векторами неограниченной Р-решетки*, а векторы, длина которых в $1/\hbar$ раз больше, — соответственно, *векторами неограниченной к-решетки*.

Чтобы выяснить, к каким изменениям приводит увеличение \mathbf{r} на величину, равную одному из периодов A_x, A_y, A_z , предположим, что на расстояние $i_x A_x$ перемещается только избыточный электрон. Это не должно было бы повлиять на волновую функцию. Изменение \mathbf{r}_e на $i_x A_x$ при данном постоянном значении \mathbf{r}_e эквивалентно, однако [см. (17.50)], изменению \mathbf{r}_T на $i_x A_x \frac{m_e}{M_T}$ и \mathbf{r} на $i_x A_x$. В результате функция $\chi\psi$ умножается на дополнительный множитель, равный

$$\exp \left[i \frac{\mathbf{P}_T \cdot i_x A_x (m_e/M_T) + \mathbf{P} \cdot i_x A_x}{\hbar} \right] = \exp \left[i \left(\frac{m_e}{M_T} \mathbf{P}_T + \mathbf{P} \right) \frac{i_x A_x}{\hbar} \right].$$

Отсюда, а также из аналогичных выражений, получаемых в тех случаях, когда \mathbf{r}_e изменяется на $i_y A_y$ и на $i_z A_z$, следует, что вектор, заключенный в круглые скобки, должен быть вектором неограниченной Р-решетки, т. е. вектором вида (17.62). Таким образом, можно написать

$$\mathbf{P} = \mathbf{P}_e - \frac{m_e}{M_T} \mathbf{P}_T, \quad (17.63)$$

где \mathbf{P}_e — вектор неограниченной решетки. При последующих вычислениях будет существенным то обстоятельство, что при постоянном значении \mathbf{P}_T разность векторов \mathbf{P} представляет собой вектор неограниченной решетки, несмотря на то, что сами векторы \mathbf{P} не обладают этим свойством. (Член, содержащий \mathbf{P}_T , соответствует учету движения центра тяжести кристалла, см. задачу 1.)

В § 6 мы покажем, что для переходов, обусловленных возмущением \mathcal{U}_1 , соблюдаются правила отбора, весьма сходные с условием сохранения импульса, согласно которому изменение импульса электрона \mathbf{P} должно компенсироваться изменением аналогичной величины для колебаний решетки. В действительности, однако, вопрос о сохранении импульса при этом не рассматривается вовсе, так как исследуемые переходы не связаны с изменением \mathbf{r}_T или же с изменением

величины P_T , которая представляет собой полный импульс системы в целом. Законы сохранения фундаментальным образом связаны с существованием ряда операторов $\mathcal{L}(R_j)$, которые будут рассмотрены в следующих параграфах. Оператор $\mathcal{L}(R_j)$ коммутирует с \mathcal{H} и, следовательно, его собственное значение есть постоянная движения. Именно для собственных значений операторов $\mathcal{L}(R_j)$ и справедливы законы сохранения.

Оператор смещения ¹⁾. Рассмотрим группу операторов $\mathcal{L}(R_j)$, которые соответствуют смещениям не на расстояния, равные периодам A_x, A_y, A_z , а на расстояния, определяемые векторами R_j кристаллической решетки. В результате применения оператора $\mathcal{L}(R_j)$ волновая функция системы изменяется следующим образом. Функция $\mathcal{L}(R_j)\Phi$ принимает то же значение, что и функция $\Phi(R_c, r_e, q_1, q_2, \dots)$, если координата электрона равна $r_e + R_j$, а координаты ядер отвечают такой же деформации, что и в случае функции Φ , с той, однако, разницей, что картина деформации (но не сами ядра) смещена на R_j . Из периодичности \mathcal{V} и \mathcal{U} следует, что если Φ есть собственная функция оператора \mathcal{H} , то $\mathcal{L}(R_i)\Phi$ также есть собственная функция \mathcal{H} .

Можно доказать, что оператор $\mathcal{L}(R_j)$ коммутирует с \mathcal{H} , а также определить собственные значения \mathcal{L} . С этой целью выразим Φ через „координаты деформации“ S_i , где

$$S_i = \delta R_i - R_c = R'_i - R_i - R_c.$$

Координаты S_i характеризуют, таким образом, ту деформацию, которая имеет место в дополнение к простой трансляции кристалла. Волновая функция системы может быть при этом записана в форме $\Phi(R_c, S_i, r_e)$, где $N_s + 1$ переменных R_c и S_i — линейные функции N_s векторов R'_i , определяющих положения ядер. Применение оператора $\mathcal{L}(R_j)$ к функции Φ приводит, очевидно, к замене переменной S_i на переменную S_i , причем

$$R_i = R_i + R_j,$$

а также к замене r_e на $r_e - R_j$. Если картина деформации и избыточный электрон переместились на R_j , то S_i примет при этом значение, которое имел раньше вектор S_i , а $r_e - R_j$ — свое прежнее значение r_e , так что

$$\begin{aligned} \mathcal{L}(R_j)\Phi \text{ для смещенной картины деформации} &= \\ &= \Phi \text{ для несмещенной картины деформации.} \end{aligned}$$

¹⁾ Введение оператора смещения дает возможность более строгого обоснования правил отбора, формулируемых в § 6. При выводе формул, определяющих среднюю длину свободного пробега, этот оператор существенной роли, однако, не играет. — *Прим. авт.*

Это равенство означает, что при смещении деформации решетки и избыточного электрона на R_j волновая функция $\mathcal{L}(R_j)\Phi$ принимает те же значения, что и волновая функция Φ .

Оператор кинетической энергии коммутирует с $\mathcal{L}(R_j)$, так как он представляет собой операцию ∇^2 , применяемую ко всем координатам, определяющим функцию Φ , и совершенно безразлично, применяется ли эта операция до или после применения оператора $\mathcal{L}(R_j)$. Поскольку значения потенциальной энергии \mathcal{V} и \mathcal{U} не изменяются при изменении координат, вызванном применением оператора $\mathcal{L}(R_j)$, последний коммутирует также с \mathcal{V} и \mathcal{U} и, следовательно, с оператором \mathcal{H} системы в целом.

Операторы $\mathcal{L}(R_i)$, очевидно, удовлетворяют уравнениям

$$\mathcal{L}(R_i)\mathcal{L}(R_j) = \mathcal{L}(R_j)\mathcal{L}(R_i) = \mathcal{L}(R_i + R_j)$$

и

$$\mathcal{L}(i_x A_x) = \mathcal{L}(i_y A_y) = \mathcal{L}(i_z A_z) = 1,$$

так как последние операции соответствуют смещению избыточного электрона на расстояние, равное одному из периодов. Следовательно, если собственные значения операторов $\mathcal{L}(i_x a)$, $\mathcal{L}(i_y a)$ и $\mathcal{L}(i_z a)$, где a — постоянная решетки для кубической структуры, равны

$$e^{-iP_x a/\hbar}, \quad e^{-iP_y a/\hbar}, \quad e^{-iP_z a/\hbar},$$

то собственное значение оператора $\mathcal{L}(R_i)$ должно равняться

$$e^{-iP_x R_i/\hbar}$$

и вектор P_x должен быть вектором неограниченной решетки типа (17.62). Однако поскольку в этом случае речь идет только о трансляции на векторы решетки, возможные значения P_x заключены внутри зоны Бриллюэна.

В следующем параграфе мы установим связь между векторами P_x и бегущими волнами, характеризующими колебания ядер, а в § 6 покажем, каким образом условие сохранения P_x приводит к правилам отбора для переходов из одних состояний в другие.

ЗАДАЧИ

1. Показать, что если пренебречь членом \mathcal{U}_1 , то волновая функция может быть представлена в форме

$$e^{i(P_{c0} \cdot R_c + P_e \cdot r_e)/\hbar} \Phi(r, q_1, q_2, \dots),$$

где P_{c0} и P_e — векторы решетки $\hbar k$, а функция $\Phi(r, q_1, q_2, \dots)$ обладает периодичностью решетки по отношению к переменной r . Каково соотношение между P_{c0} и P_e и векторами P_T и P , рассмотренными выше?

2. Показать, что если одно из ядер есть ядро тяжелого изотопа, то оператор $\mathcal{L}(\mathbf{R}_i)$ не коммутирует с кинетической энергией ядер. (То обстоятельство, что \mathcal{L} не коммутирует с \mathcal{H} , не означает, однако, что правила отбора, рассматриваемые в § 6, несправедливы, так как последние связаны с линейным приближением от функции \mathcal{U}_1 , пригодным при низких температурах независимо от содержания изотопов в кристалле.)

3. Показать, что оператор $\mathcal{L}(\mathbf{R}_i)$ существует также и для кристалла типа NaCl, в каждой элементарной ячейке которого содержатся два различных атома.

4. Исследовать свойства операторов $\mathcal{L}(\mathbf{R}_i)$ с точки зрения теории групп. Почему неприводимые представления этих операторов одномерны?

§ 5. ПРЕОБРАЗОВАНИЕ К БЕГУЩИМ ВОЛНАМ

Координаты q , введенные в рассмотрение в § 3, соответствуют нормальным колебаниям и стоячим волнам. Путем контактного преобразования можно получить новую группу координат, соответствующих бегущим волнам. Необходимое для этого преобразование обладает тем свойством, что оно смешивает значения p и q , относящиеся к данному колебанию, т. е. значения p и q , отвечающие данному волновому вектору \mathbf{k} и определенному направлению поляризации α . Функция Гамильтона \mathcal{H}_i для этих двух степеней свободы, которые мы обозначим через q_c и q_s (опуская индексы \mathbf{k} и α), имеет вид

$$\mathcal{H}_i = \frac{[p_c^2 + p_s^2 + M^2 \omega^2 (q_c^2 + q_s^2)]}{2M}. \quad (17.64)$$

Преобразование от старых координат q и p , описывающих стоячие волны (будем обозначать их через $q_{\text{ст}i}$), к координатам, характеризующим бегущие волны ($q_{\text{бег}i}$)¹⁾, определяется соотношениями

$$\begin{aligned} q_c &= (2M\omega)^{-1/2} (q_- + p_+), & q_s &= (2M\omega)^{-1/2} (q_+ + p_-), \\ p_c &= \left(\frac{M\omega}{2}\right)^{1/2} (p_- - q_+), & p_s &= \left(\frac{M\omega}{2}\right)^{1/2} (p_+ - q_-). \end{aligned} \quad (17.65)$$

При этом \mathcal{H}_i преобразуется к виду

$$\mathcal{H}_i = \frac{1}{2} \omega (p_+^2 + q_+^2 + p_-^2 + q_-^2). \quad (17.66)$$

¹⁾ Индекс „бег“ представляет собой сокращение английского слова „travelling“, что значит „бегущая“. — Прим. ред.

С помощью соотношений (17.65) или соотношений, определяющих обратное преобразование

$$q_+ = \left(\frac{M\omega}{2}\right)^{1/2} q_s - (2M\omega)^{-1/2} p_c, \quad q_- = \left(\frac{M\omega}{2}\right)^{1/2} q_c - (2M\omega)^{-1/2} p_s,$$

$$p_+ = \left(\frac{M\omega}{2}\right)^{1/2} q_c + (2M\omega)^{-1/2} p_s, \quad p_- = \left(\frac{M\omega}{2}\right)^{1/2} q_s + (2M\omega)^{-1/2} p_c, \quad (17.67)$$

можно показать, что новые переменные p и $q_{\text{гн}}$, если рассматривать их как операторы, должны удовлетворять квантовомеханическим правилам коммутативности

$$p_+ q_+ - q_+ p_+ = \frac{\hbar}{i}, \quad p_+ q_- - q_- p_+ = 0, \quad (17.68)$$

поскольку старые переменные p и $q_{\text{ст}}$ обладают этим свойством (см. результаты, полученные в § 3 при решении задачи 2). Рассматриваемое преобразование является контактным также и в классическом смысле, так как

$$p_+ dq_+ + p_- dq_- = p_c dq_c + p_s dq_s + dS, \quad (17.69)$$

где dS — полный дифференциал.

Покажем теперь, что, согласно классической теории, движение, описываемое переменными p_+, q_+ , представляет собой бегущую волну, распространяющуюся в направлении $+\mathbf{k}$, тогда как переменные p_-, q_- описывают волну, распространяющуюся в направлении $-\mathbf{k}$. В соответствии с этим новые переменные, получаемые путем преобразования любой пары значений $q_{k\alpha}$ и $q_{k\alpha}$, естественно обозначать не через $q_{k\alpha+}$ и $q_{k\alpha-}$, а через $q_{k\alpha}$ и $q_{-k\alpha}$, включив в рассмотрение отрицательные значения \mathbf{k} (напомним, что, пользуясь при построении зон Бриллюэна координатами $q_{\text{ст}}$, мы учитывали только половину возможных значений \mathbf{k} , удовлетворяющих периодическим граничным условиям, так как волновой вектор $-\mathbf{k}$ отвечает той же стоячей волне, что и волновой вектор $+\mathbf{k}$). Таким образом, в системе координат $q_{\text{гн}}$ при суммировании по \mathbf{k} следует учитывать все значения \mathbf{k} , заключенные в зоне Бриллюэна; каждой паре значений $k\alpha$ при этом соответствует только одна координата.

Классическую картину движения, описываемого переменными p_+ и q_+ , получим, решив уравнения Гамильтона

$$\dot{q}_+ = \frac{\partial \mathcal{H}}{\partial p_+} = \omega p_+ \quad \text{и} \quad \dot{p}_+ = -\frac{\partial \mathcal{H}}{\partial q_+} = -\omega q_+, \quad (17.70)$$

что дает

$$q_+ = A_+ \cos(\omega t + \theta_+), \quad (17.71a)$$

$$p_+ = -A_+ \sin(\omega t + \theta_+). \quad (17.71b)$$

Подставляя эти значения q_+ и p_+ в выражение, определяющее δR_i , и опуская индексы \mathbf{k} и α , получаем

$$\delta R_i = G [\cos(\mathbf{k} \cdot \mathbf{R}_i) q_c + \sin(\mathbf{k} \cdot \mathbf{R}_i) q_s], \quad (17.72a)$$

$$\delta R_i = A_+ (2M\omega)^{1/2} G \sin[\mathbf{k} \cdot \mathbf{R}_i - \omega t - \theta_+], \quad (17.72b)$$

что соответствует волне, распространяющейся в направлении $+\mathbf{k}$ с фазовой скоростью $\omega/|\mathbf{k}|$.

Аналогично можно показать, что решение $p_- q_-$ отвечает волне, распространяющейся в направлении $-\mathbf{k}$.

Покажем теперь, каким образом координаты q_+ и q_- используются для описания бегущих волн в квантовомеханической теории. Как было установлено в § 4, энергия избыточного электрона в деформированной решетке является функцией положения электрона \mathbf{r} , отсчитываемого от центра тяжести кристалла, и деформации решетки, выраженной через переменные q_{std} . Разложим эту энергию в ряд по степеням q и предположим, что деформация настолько мала, что можно ограничиться рассмотрением одних лишь линейных членов разложения. Это приводит к уравнению

$$\mathcal{U}(\mathbf{r}, q_{std}) = \mathcal{U}_0(\mathbf{r}) + \mathcal{U}_1(\mathbf{r}, q_{std}), \quad (17.73)$$

где возмущающий член \mathcal{U}_1 равен

$$\mathcal{U}_1(\mathbf{r}, q_{std}) = \sum_{ka} \left[q_{kas} \left(\frac{\partial \mathcal{U}}{\partial q_{kas}} \right) + q_{kas} \left(\frac{\partial \mathcal{U}}{\partial q_{kas}} \right) \right] \quad (17.74)$$

(частные производные взяты при условии отсутствия деформации). При заданных \mathbf{k} и α значения q_c и q_s определяют синусоидальное возмущение, причем δR_i определяется формулой (17.72a). Методы вычисления энергии взаимодействия электрона с возмущением такого типа будут рассмотрены в следующем параграфе. Здесь мы исследуем некоторые общие вопросы, связанные с определением матричных элементов p , q и δR_i .

Для описания бегущих волн, соответствующих колебаниям ядер, удобно воспользоваться квантовыми числами. Ту часть волновой функции, которая зависит от q_s и q_c , запишем в виде $\varphi(n_+, n_-; q_c, q_s)$ и потребуем, чтобы она была собственной функцией как операторов \mathcal{H}_+ и \mathcal{H}_- в отдельности [с собственными значениями $\hbar\omega(n_+ + 1/2)$ и $\hbar\omega(n_- + 1/2)$], так и оператора $\mathcal{H}_i = \mathcal{H}_+ + \mathcal{H}_-$, представляющего полную энергию системы. Выразив q_+ и q_- через q_{std} , получаем

$$\mathcal{H}_+ = \frac{1}{2} \omega (p_+^2 + q_+^2) = \frac{1}{2} \mathcal{H}_i + \frac{1}{2} \omega (p_s q_c - p_c q_s) \quad (17.75)$$

и

$$\mathcal{H}_- = \frac{1}{2} \omega (p_-^2 + q_-^2) = \frac{1}{2} \mathcal{H}_i - \frac{1}{2} \omega (p_s q_c - p_c q_s). \quad (17.76)$$

Поскольку \mathcal{H}_+ и \mathcal{H}_- — операторы Гамильтона для гармонических осцилляторов, имеем

$$\begin{aligned}\mathcal{H}_+\varphi(n_+, n_-; q_c, q_s) &= \hbar\omega\left(n_+ + \frac{1}{2}\right)\varphi(n_+, n_-; q_c, q_s), \\ \mathcal{H}_-\varphi(n_+, n_-; q_c, q_s) &= \hbar\omega\left(n_- + \frac{1}{2}\right)\varphi(n_+, n_-; q_c, q_s),\end{aligned}\quad (17.77)$$

так что собственное значение \mathcal{H}_i равно

$$\mathcal{E}_i = \hbar\omega(n_+ + n_- + 1) = \hbar\omega(n_i + 1). \quad (17.78)$$

При данном значении энергии \mathcal{E}_i существует $(n_i + 1)$ способов ее подразделения между слагаемыми, содержащими квантовые числа n_+ и n_- .

Для переходов, связанных с изменением квантовых чисел n_+ и n_- , матричный элемент δR_i линеен относительно q_c и q_s и, следовательно, линеен относительно p_+ , q_+ , p_- и q_- . Если n_- изменяется, то матричные элементы p_+ и q_+ равны нулю, так как p_+ и q_+ коммутируют с \mathcal{H}_- . Если n_- остается постоянным, то матричные элементы p_+ и q_+ отличны от нуля только в том случае, когда n_+ изменяется на ± 1 . Они определяются при этом обычными формулами¹⁾:

$$\begin{aligned}(n_+, n_- | p_+ \text{ или } q_+ | n'_+, n_-) &= \\ &= \int \int \varphi^*(n_+, n_-; q_c, q_s) [p_+ \text{ или } q_+] \varphi(n'_+, n_-; q_c, q_s) dq_c dq_s,\end{aligned}\quad (17.79)$$

$$(n_+, n_- | p_+ | n_+ \pm 1, n_-) = \begin{cases} \sqrt{\hbar(n_+ + 1)/2} \\ \sqrt{\hbar n_+ / 2} \end{cases}, \quad (17.80)$$

$$(n_+, n_- | q_+ | n_+ \pm 1, n_-) = \begin{cases} i\sqrt{\hbar(n_+ + 1)/2} \\ i\sqrt{\hbar n_+ / 2} \end{cases}, \quad (17.81)$$

Формулы, определяющие матричные элементы p_- и q_- , имеют аналогичный вид и могут быть получены из формул (17.79) — (17.81) в результате перестановки индексов (разность численных значений индексов при этом должна оставаться равной ± 1).

Матричный элемент δR_i найдем, выразив матричные элементы q_c и q_s через матричные элементы (17.80) и (17.81) и подставив результаты в (17.72а). Это дает

$$(n_+, n_- | \delta R_i | n_+ \pm 1, n_-) = G \sqrt{\frac{\hbar}{4M\omega}} \begin{cases} \sqrt{n_+ + 1} \exp(i\mathbf{k} \cdot \mathbf{R}_i) \\ \sqrt{n_+} \exp(-i\mathbf{k} \cdot \mathbf{R}_i) \end{cases}, \quad (17.82)$$

Напомним, что M есть половина массы кристалла [см. (17.40)]. Аналогично для переходов, связанных с изменением квантового числа n_- ,

¹⁾ Задача о гармоническом осцилляторе рассматривается, например, в монографии Дирака [13]. — *Прим. авт.*

матричный элемент δR_i равен

$$(n_+, n_- | \delta R_i | n_+, n_- \pm 1) = \\ = G \sqrt{\frac{\hbar}{4M\omega}} \begin{cases} i \sqrt{n_- + 1} \exp(-ik \cdot R_i), \\ -i \sqrt{n_-} \exp(ik \cdot R_i). \end{cases} \quad (17.83)$$

Вещественная и мнимая части этих выражений соответствуют переменным q_c и q_s в отдельности.

Поскольку матричные элементы δR_i зависят только от квантового числа, изменяющегося при переходах, можно представить их в виде функций этого числа, обозначив последнее через n_{ka} в соответствии с обозначением координат q_{ka} . Оба матричных элемента δR_i приобретают при этом вид

$$(n_{ka} | \delta R_i | n_{ka} + \delta n_{ka}) = G_{ka} \left[\hbar \frac{2n_{ka} + 1 + \delta n_{ka}}{8M\omega_{ka}} \right]^{1/2} \exp(i\delta n_{ka} k \cdot R_i); \quad (17.84)$$

мы опустили здесь множитель $\pm i$, поскольку для вычисления вероятностей переходов нужно знать только абсолютное значение выражения (17.84). Матричный элемент δR_i отличен от нуля лишь при условии $|\delta n_{ka}| = 1$.

Волновая функция системы, описывающая колебания ядер и являющаяся собственной функцией всех операторов $\mathcal{H}_{\text{яд.}}$, может быть записана в виде произведения

$$\varphi(n_{k_1\alpha_1}, n_{k_2\alpha_2}, \dots) = \prod_{k_j\alpha_j} \varphi(n_{k_j\alpha_j}, n_{-k_j\alpha_j}; q_{k_j\alpha_j c}, q_{k_j\alpha_j s}) = \prod_j \varphi_j; \quad (17.85)$$

число множителей равно здесь числу различных значений пар индексов $k_j\alpha_j$. Матричные элементы q_{std} и, следовательно, матричные элементы δR_i отличны от нуля лишь для таких переходов, при которых одно и только одно из квантовых чисел n_k изменяется на ± 1 .

Применение операторов $\mathcal{S}(R_s)$ к функции Φ^1). Основная причина, по которой собственные значения n_{ka} могут быть использованы для описания бегущих волн, заключается в том, что если волновую функцию системы Φ выразить через n_{ka} и функции χ_{pT} и ψ_p , то она станет при этом собственной функцией оператора $\mathcal{S}(R_s)$. Это означает следующее. Волновая функция Φ' , полученная в результате применения оператора $\mathcal{S}(R_s)$ к функции Φ , равна просто $L_s\Phi$, где L_s — собственное значение оператора $\mathcal{S}(R_s)$ (как будет показано ниже, L_s — число, абсолютное значение которого равно единице). Волновая функция $\mathcal{S}(R_s)\Phi \equiv \Phi'$ определяется следующим образом (см. § 4). Рассмотрим функцию Φ для некоторого ряда значений координат системы в целом,

¹⁾ Как было упомянуто в § 4, при вычислении средней длины свободного пробега использование этого метода не является необходимым. — *Прим. авт.*

который обозначим сокращенно через Q ; затем сместим картину деформации ядер и положение избыточного электрона на \mathbf{R}_s и обозначим новый ряд значений координат через Q' . Очевидно, Q' есть функция Q и \mathbf{R}_s , а Q зависит точно таким же образом от Q' и $-\mathbf{R}_s$; эти соотношения могут быть записаны в виде

$$Q' = \lambda(Q, \mathbf{R}_s) \text{ и } Q = \lambda(Q', -\mathbf{R}_s). \quad (17.86)$$

Определение функции $\mathcal{L}(\mathbf{R}_s)\Phi$ имеет при этом вид

$$[\mathcal{L}(\mathbf{R}_s)\Phi](Q') \equiv \Phi'(Q') = \Phi(Q), \quad (17.87a)$$

так что

$$\Phi'(Q') = \Phi[\lambda(Q', -\mathbf{R}_s)]. \quad (17.87b)$$

Доказательство того, что Φ есть собственная функция оператора $\mathcal{L}(\mathbf{R}_s)$, сводится к доказательству того, что

$$\Phi'(Q') = \Phi[\lambda(Q', -\mathbf{R}_s)] = L_s\Phi(Q'), \quad (17.87b)$$

т. е. того, что для любого ряда значений координат Q' функция Φ' равна произведению числа L_s на функцию Φ , отвечающую этому ряду значений Q' .

Чтобы получить этот результат, следует определить форму функции λ для координат \mathbf{r}_T , \mathbf{r} и q_1, q_2, \dots . Для координат \mathbf{r} и \mathbf{r}_T имеем

$$\begin{aligned} \mathbf{r}' &= \lambda_{\mathbf{r}}(\mathbf{r}, \mathbf{R}_s) = \mathbf{r} + \mathbf{R}_s, \\ \lambda_{\mathbf{r}}(\mathbf{r}', -\mathbf{R}_s) &= \mathbf{r}' - \mathbf{R}_s, \end{aligned} \quad (17.88a)$$

$$\begin{aligned} \mathbf{r}'_T &= \lambda_{\mathbf{r}_T}(\mathbf{r}_T, \mathbf{R}_s) = \mathbf{r}_T + \frac{m_e}{M_T} \mathbf{R}_s, \\ \lambda_{\mathbf{r}_T}(\mathbf{r}'_T, -\mathbf{R}_s) &= \mathbf{r}'_T - \frac{m_e}{M_T} \mathbf{R}_s, \end{aligned} \quad (17.88b)$$

как это следует из (17.50) и того обстоятельства, что смещение картины деформации не должно изменять величины $\sum_i \delta \mathbf{R}_i$, так что значение \mathbf{R}_s остается неизменным.

Для исследования влияния смещения \mathbf{R}_s на значения координат q рассмотрим две такие координаты, соответствующие данным \mathbf{k} и α ; эти координаты обозначим через q_c и q_s , опуская для краткости индексы $\mathbf{k}\alpha$. Введем новые координаты

$$\rho = (q_c^2 + q_s^2)^{1/2}, \quad \theta = \text{tg}^{-1} \frac{q_s}{q_c}; \quad (17.89a)$$

$$q_c = \rho \cos \theta, \quad q_s = \rho \sin \theta. \quad (17.89b)$$

Как мы сейчас покажем, изменение этих координат, обусловленное смещением деформации на \mathbf{R}_s , определяется соотношениями (17.95). Введение новых координат дает возможность рассматривать два гар-

монических осциллятора q_c и q_s как один двухмерный изотропный осциллятор. Для такого осциллятора оператор $(\hbar/i) \partial/\partial\theta$, представляющий момент количества движения, коммутирует с \mathcal{H} , причем собственные значения этого оператора изменяются от $\hbar n$ до $-\hbar n$, отличаясь друг от друга на $2\hbar$, а квантовое число n определяет полную энергию, равную $\hbar\omega(n+1)$. Независимо от этого результата имеем чисто формально

$$\begin{aligned} \frac{\hbar}{i} \cdot \frac{\partial}{\partial\theta} &= \frac{\hbar}{i} \left[\frac{\partial q_c}{\partial\theta} \frac{\partial}{\partial q_c} + \frac{\partial q_s}{\partial\theta} \frac{\partial}{\partial q_s} \right] = \\ &= \frac{\hbar}{i} \left[-q_s \frac{\partial}{\partial q_c} + q_c \frac{\partial}{\partial q_s} \right] = q_c p_s - q_s p_c. \end{aligned} \quad (17.90)$$

Выражение (17.75) приобретает при этом вид

$$\mathcal{H}_+ = \frac{1}{2} \mathcal{H}_i + \frac{\hbar\omega}{2i} \frac{\partial}{\partial\theta}. \quad (17.91)$$

Отсюда следует, что волновая функция $\varphi(n_+, n_-; q_c, q_s)$ есть собственная функция оператора $\partial/\partial\theta$, собственное значение которого равно $i(n_+ - n_-)$. Поэтому можно написать

$$\frac{\partial}{\partial\theta} \varphi(n_+, n_-; q_c, q_s) = i(n_+ - n_-) \varphi(n_+, n_-; q_c, q_s), \quad (17.92)$$

так что

$$\varphi(n_+, n_-; q_c, q_s) = e^{i(n_+ - n_-)\theta} f(\rho). \quad (17.93)$$

Выразив $\delta\mathbf{R}_i$ через координаты ρ и θ , получаем

$$\begin{aligned} \delta\mathbf{R}_i &= \mathbf{G}_{k\alpha} [\cos \mathbf{k} \cdot \mathbf{R}_{i\rho} \cos \theta + \sin \mathbf{k} \cdot \mathbf{R}_{i\rho} \sin \theta] = \\ &= \mathbf{G}_{k\alpha} \rho \cos [\mathbf{k} \cdot \mathbf{R}_i - \theta]. \end{aligned} \quad (17.94)$$

Таким образом, в результате смещения на \mathbf{R}_s координата θ становится равной $\theta' = \theta + \mathbf{k} \cdot \mathbf{R}_s$, так что то значение $\delta\mathbf{R}$, которое отвечало раньше точке \mathbf{R}_i , отвечает теперь точке $\mathbf{R}_i + \mathbf{R}_s$. Соответственно можно записать:

$$\theta' = \lambda_\theta(\theta, \mathbf{R}_s) = \theta + \mathbf{k} \cdot \mathbf{R}_s, \quad \lambda_\theta(\theta', -\mathbf{R}_s) = \theta' - \mathbf{k} \cdot \mathbf{R}_s; \quad (17.95a)$$

$$\rho' = \lambda_\rho(\rho, \mathbf{R}_s) = \rho, \quad \lambda_\rho(\rho', -\mathbf{R}_s) = \rho'. \quad (17.95b)$$

Аналогичные соотношения будут иметь место при всех значениях \mathbf{k} и α .

Воспользовавшись свойствами функций $\chi_{\mathbf{p}_T}(\mathbf{r}_T)$ и $\psi_{\mathbf{p}}(\mathbf{r})$, определяемыми формулами (17.59) и (17.61), а также свойствами φ [см. (17.93)] и следуя соотношению (17.87в), легко вычислить значение функции

$$\Phi = \chi_{\mathbf{p}_T} \psi_{\mathbf{p}} \varphi(M_{k\alpha}) \quad (17.96)$$

для ряда значений координат $\lambda(Q', -\mathbf{R}_s)$.

Каждый из множителей, определяющих функцию Φ , дает при этом выражение вида $\exp(-i\mathbf{P} \cdot \mathbf{R}_s/\hbar)$ или $\exp[-i(n_+ - n_-)\mathbf{k} \cdot \mathbf{R}_s]$. Двучлен $(n_+ - n_-)\mathbf{k}$ может быть записан в форме $n_{\mathbf{k}}\mathbf{k} + n_{-\mathbf{k}}(-\mathbf{k})$, так что собственное значение L_s , определенное в § 3 как $\exp(-i\mathbf{P}_s \cdot \mathbf{R}_s/\hbar)$, приобретает вид

$$L_s = \exp^{-i\mathbf{P}_s \cdot \mathbf{R}_s/\hbar} = \exp -i \left[\mathbf{P} + \frac{m_0}{M_T} \mathbf{P}_T + \hbar \sum n_{\mathbf{k}\alpha} \mathbf{k} \right] \frac{\mathbf{R}_s}{\hbar}; \quad (17.97)$$

суммирование производится здесь по всем значениям пар индексов $\mathbf{k}\alpha$, включая как положительные, так и отрицательные значения \mathbf{k} .

Вектор $\mathbf{P} + (m_0/M_T)\mathbf{P}_T$ есть вектор неограниченной \mathbf{P} -решетки, как это было выяснено при анализе выражения (17.63). Согласно (17.38), этим свойством обладает и сумма векторов $\hbar\mathbf{k}$. Следовательно, \mathbf{P}_s есть вектор неограниченной \mathbf{P} -решетки (в согласии с заключением, к которому мы пришли в § 4).

Знак минус в выражении (17.97), определяющем L_s , связан с тем обстоятельством, что рассматриваемая задача интересует нас с точки зрения смещения значений волновой функции, а не с точки зрения изменения от Q до $\lambda(Q, \mathbf{R}_s)$ того ряда значений координат, которые определяют данную волновую функцию. Последнее, как легко показать, характеризуется соотношением вида

$$\Phi[\lambda(Q, \mathbf{R}_s)] = e^{i\mathbf{P}_s \cdot \mathbf{R}_s/\hbar} \Phi(Q). \quad (17.98)$$

Таким образом, величина \mathbf{P}_s играет здесь роль, аналогичную \mathbf{P} , \mathbf{P}_T и $\hbar n_{\mathbf{k}\alpha}\mathbf{k}$, поскольку речь идет о ее знаке.

Как \mathbf{P}_s , так и \mathbf{P}_T являются постоянными движения для оператора полной энергии \mathcal{H} . Поэтому *при переходах, обусловленных возмущениями \mathcal{U}_1 , вектор $\mathbf{P} + \hbar \sum n_{\mathbf{k}\alpha}\mathbf{k}$ должен оставаться неизменным* даже при учете членов, нелинейных относительно координат q_1, q_2, \dots . Именно этот закон сохранения \mathbf{P}_s и лежит в основе тех правил отбора, которые будут рассмотрены в следующем параграфе.

§ 6. ВЫЧИСЛЕНИЕ МАТРИЧНОГО ЭЛЕМЕНТА

Как уже было указано в § 2, вероятности переходов определяются матричными элементами оператора \mathcal{H} , взятыми по отношению к различным состояниям рассматриваемой системы. Волновые функции, описывающие эти состояния, приближенно представляют собой собственные функции оператора \mathcal{H} и являются точными собственными функциями оператора \mathcal{H}_0 , который лишь незначительно отличается от \mathcal{H} . В том случае, который будет рассмотрен в настоящем параграфе, волновые функции являются собственными функциями оператора

$$\mathcal{H}_0 = \mathcal{H} - \mathcal{U}_1, \quad (17.99)$$

где \mathcal{U}_1 — дополнительная потенциальная энергия избыточного электрона, обусловленная деформацией решетки, описываемой координатами q [см. (17.47)].

Собственные функции Φ оператора \mathcal{H}_0 могут быть записаны в виде произведения уже рассмотренных выше волновых функций $\chi_{\mathbf{r}_T}$ и $\psi_{\mathbf{r}}$, зависящих от переменных \mathbf{r}_T и \mathbf{r} , и волновых функций φ , зависящих от координат q_e и q_a , соответствующих различным значениям параметров \mathbf{k} и α . Волновые функции $\chi_{\mathbf{r}_T}$ и $\psi_{\mathbf{r}}$ [см. (17.59) и (17.61)] имеют вид

$$\chi_{\mathbf{r}_T}(\mathbf{r}_T) = V^{-1/2} e^{i(\mathbf{P}_T \cdot \mathbf{r}_T / \hbar)}, \quad (17.100)$$

$$\psi_{\mathbf{r}}(\mathbf{r}) = V^{-1/2} e^{i(\mathbf{P} \cdot \mathbf{r} / \hbar)} u_{\mathbf{P}}(\mathbf{r}). \quad (17.101)$$

Множитель $V^{-1/2}$, где $V = A_x A_y A_z$, нормирует волновую функцию по рассматриваемой области, причем предполагается, что $u_{\mathbf{P}}(\mathbf{r})$ нормировано к единице объема. Нормированные волновые функции, зависящие от координат q , записываются в виде

$$\varphi(n_{\mathbf{k}\alpha}, n_{-\mathbf{k}\alpha}; q_{\mathbf{k}\alpha c}, q_{\mathbf{k}\alpha s}) \equiv \varphi_j, \quad (17.102)$$

где φ_j — сокращенное обозначение, которым мы будем пользоваться в дальнейшем, а индекс j заменяет индексы $\mathbf{k}\alpha$, относящиеся к стоячим волнам. Волновая функция системы может быть, таким образом, записана в виде

$$\Phi(\mathbf{P}_T, \mathbf{P}, n) = \chi_{\mathbf{r}_T} \psi_{\mathbf{r}} \prod \varphi_i. \quad (17.103)$$

В обозначениях, принятых в § 4 [см. (17.56) — (17.58)], оператор \mathcal{H}_0 имеет вид

$$\begin{aligned} \mathcal{H}_0 &= \frac{1}{2M_T} p_T^2 + \frac{1}{2m} p^2 + \mathcal{U}_0(\mathbf{r}) + \sum_i \mathcal{H}_i = \\ &= \mathcal{H}_{\mathbf{r}_T} + \mathcal{H}_{\mathbf{r}} + \sum_i \mathcal{H}_i. \end{aligned} \quad (17.104)$$

Каждое из слагаемых, определяющих \mathcal{H}_0 , действует только на один из множителей функции Φ , так что энергия системы \mathcal{E} равна

$$\mathcal{E} = \frac{1}{2M_T} \mathbf{P}_T^2 + \mathcal{E}(\mathbf{P}) + \sum_i \hbar \omega_i [n_{+i} + n_{-i} + 1]. \quad (17.105)$$

Отдельные члены этого выражения были рассмотрены выше и соответствуют операторам $\mathcal{H}_{\mathbf{r}_T}$ [см. (17.59)], $\mathcal{H}_{\mathbf{r}}$ [см. (17.60)] и \mathcal{H}_i [см. (17.78)].

Волновые функции Φ , описывающие состояние системы в целом, образуют ортонормированную группу функций; исследованию подлежат

только вопрос об ортогональности функций $\psi_{\mathbf{P}}$. Нетрудно убедиться в том, что при смещении возможных значений \mathbf{P} из зоны Бриллюэна на вектор \mathbf{P}_T [см. (17.63)] соотношения ортогональности между функциями $\psi_{\mathbf{P}}$ не нарушаются.

Матричные элементы оператора \mathcal{H}_0 , взятые по отношению к ортогональным собственным функциям этого оператора Φ , равны нулю, так что все переходы обусловлены наличием члена $\mathcal{U}_1(\mathbf{r}, q_1, q_2, \dots)$. Предположим, что амплитуды колебаний решетки столь малы, что \mathcal{U}_1 может быть представлено в виде суммы членов, содержащих координаты q в первой степени:

$$\begin{aligned} \mathcal{U}_1(\mathbf{r}, q_{i_c}, q_{i_s}) &= \sum_i q_{i_c} \frac{\partial \mathcal{U}}{\partial q_{i_c}} + \sum_i q_{i_s} \frac{\partial \mathcal{U}}{\partial q_{i_s}} = \\ &= \sum_i q_{i_c} \mathcal{U}_{i_c}(\mathbf{r}) + \sum_i q_{i_s} \mathcal{U}_{i_s}(\mathbf{r}). \end{aligned} \quad (17.106)$$

Последняя форма записи этого выражения связана с тем обстоятельством, что частные производные здесь вычисляются при условии равенства всех q нулю, так что эти производные зависят только от \mathbf{r} .

Матричные элементы двух членов ряда (17.106), соответствующих данному значению индексов $k\alpha$ (которое мы сокращенно обозначили через i), взятые по отношению к двум функциям Φ , отличны от нуля только в том случае, если

- а) все значения $n_{+j}n_{-j}$ при $j \neq i$ остаются неизменными;
- б) только одно из значений n_{+i} либо только одно из значений n_{-i} изменяется на ± 1 .

Если условие а) не имеет места, то матричный элемент обращается в нуль в силу взаимной ортогональности двух функций φ_j , стоящих под знаком интеграла. Матричные элементы q_{i_c} и q_{i_s} отличны от нуля только при условии б); их значения определяются при этом формулами (17.80) и (17.81). Поскольку \mathcal{U}_1 не зависит от \mathbf{r}_T , то \mathbf{P}_T есть постоянная величина, как это было выяснено в § 4; при любых переходах значение \mathbf{P}_T остается неизменным.

Нужно, таким образом, исследовать лишь те случаи, когда изменяются импульс \mathbf{P} и только одно из квантовых чисел n . Исходное и конечное значения числа n обозначим через n_i и n'_i , где индекс i заменяет совокупность двух индексов $k\alpha$. Матричный элемент \mathcal{U}_1 при этом равен

$$\begin{aligned} (\mathbf{P}n_i | \mathcal{U}_1 | \mathbf{P}'n'_i) &= \int \Phi^* \mathcal{U}_1 \Phi d' \quad (\text{все координаты}) = \\ &= \frac{1}{V} \int \psi_{\mathbf{P}}^*(\mathbf{r}) \varphi^*(n_{i_+}, n_{i_-}; q_{i_c}, q_{i_s}) [q_{i_c} \mathcal{U}_{i_c} + q_{i_s} \mathcal{U}_{i_s}] \times \\ &\quad \times \psi_{\mathbf{P}'}(\mathbf{r}) \varphi(n'_{i_+}, n'_{i_-}; q_{i_c}, q_{i_s}) dV_{\mathbf{r}} dq_{i_c} dq_{i_s}, \end{aligned} \quad (17.107)$$

где скобка $[n_{i+}, n_{i-}]'$ обозначает, что только одно из двух квантовых чисел n изменилось на ± 1 , а $dV_{\mathbf{r}}$ — элемент объема \mathbf{r} -пространства.

В этом же параграфе, несколько дальше, мы будем иметь дело непосредственно с выражениями, содержащими под знаком интеграла \mathcal{U}_{i_c} и \mathcal{U}_{i_b} . Имеет, однако, смысл исследовать сначала выражение (17.107) с помощью метода, при применении которого воздействие каждой из q -волн рассматривается как суммарный результат смещений отдельных ядер. Введем в рассмотрение понятие „деформации, отвечающей отдельному ядру“, дав ему следующее определение: представим себе, что в деформированном кристалле при данном значении \mathbf{r} избыточный электрон и все ядра, за исключением ядра i , остаются неподвижными, тогда как ядро i смещается на $\delta\mathbf{R}_i$. При такой деформации изменение $\delta\mathcal{U}_i$ потенциальной энергии \mathcal{U}_i будет функцией положения электрона \mathbf{r} по отношению к центру тяжести деформированного кристалла, а также функцией смещения $\delta\mathbf{R}_i$. При малых деформациях $\delta\mathcal{U}_i$ будет линейной функцией трех компонент $\delta\mathbf{R}_i$. Соответственно можно записать

$$\delta\mathcal{U}_i = \delta\mathbf{R}_i \cdot \mathbf{W}_{\mathbf{R}_i}(\mathbf{r}).$$

Отметим еще раз, что $\delta\mathcal{U}_i$ есть изменение энергии, обусловленное смещением ядра i на $\delta\mathbf{R}_i$, при условии, что положения избыточного электрона и всех остальных ядер остаются неизменными. Суммарный результат всех таких „деформаций, отвечающих отдельным ядрам“, эквивалентен деформации, вызванной q_c -волной¹⁾. Таким образом, изменение энергии \mathcal{U}_i , обусловленное q_c -волной, может быть определено двумя эквивалентными способами:

$$\sum_i \delta\mathcal{U}_i = \sum_i \delta\mathbf{R}_i \cdot \mathbf{W}_{\mathbf{R}_i}(\mathbf{r}) = q_c \mathcal{U}_c(\mathbf{r}). \quad (17.108)$$

Недеформированная решетка периодична; поэтому очевидно, что

$$\mathbf{W}_{\mathbf{R}_i}(\mathbf{r}) = \mathbf{W}_0(\mathbf{r} - \mathbf{R}_i). \quad (17.109)$$

Отсюда следует, что смещение любого из ядер приводит к точно такому же изменению потенциальной энергии, что и смещение ядра с нулевым номером, с той, однако, разницей, что оно сопровождается трансляцией на вектор решетки \mathbf{R}_i . Функция $\mathbf{W}_0(\mathbf{r} - \mathbf{R}_i)$ входит под знак интеграла, определяющего матричный элемент \mathcal{U}_i . Воспользовавшись определением функций $\psi_{\mathbf{r}}$ и заменяя переменные интегрирования,

¹⁾ Следует отметить, что суммарное воздействие всех смещений $\delta\mathbf{R}_i$, вызванных q_c -волной, не должно приводить к изменению положения центра тяжести кристалла [см. (17.39)]. Хотя каждая „деформация, отвечающая отдельному ядру“, обуславливает некоторое изменение \mathbf{R}_c и, следовательно, изменение \mathbf{r} при данном положении электрона, суммарное влияние всех этих деформаций таково, что \mathbf{r} остается неизменным. — *Прим. авт.*

можно преобразовать интеграл от функции $W_0(\mathbf{r} - \mathbf{R}_j)$ к более простому виду:

$$\begin{aligned} I &\equiv \int \psi_{\mathbf{P}}^*(\mathbf{r}) W_0(\mathbf{r} - \mathbf{R}_j) \psi'_{\mathbf{P}}(\mathbf{r}) dV_{\mathbf{r}} = \\ &= \int \psi_{\mathbf{P}}^*(\mathbf{r}' + \mathbf{R}_j) W_0(\mathbf{r}') \psi'_{\mathbf{P}}(\mathbf{r}' + \mathbf{R}_j) dV_{\mathbf{r}'} = \\ &= e^{i(\mathbf{P}' - \mathbf{P}) \cdot \mathbf{R}_j / \hbar} \int \psi_{\mathbf{P}}^*(\mathbf{r}') W_0(\mathbf{r}') \psi'_{\mathbf{P}}(\mathbf{r}') dV_{\mathbf{r}'} = \\ &= e^{i(\mathbf{P}' - \mathbf{P}) \cdot \mathbf{R}_j / \hbar} I_0, \quad (17.110) \end{aligned}$$

так как интеграл по \mathbf{r}' эквивалентен интегралу по \mathbf{r}^1 .

Для перехода, при котором $n_{k\alpha} \rightarrow n_{k\alpha} + \delta n_{k\alpha}$, выражая матричный элемент \mathcal{U}_1 через интеграл I_0 и матричный элемент смещения $\delta \mathbf{R}_j$ [см. (17.84)], получаем

$$\begin{aligned} (\mathbf{P}, n_{k\alpha} | \mathcal{U}_1 | \mathbf{P}', n_{k\alpha} + \delta n_{k\alpha}) &= \sum_j \int \Phi^* \delta \mathbf{R}_j \cdot \mathbf{W}_{\mathbf{R}_j} \Phi' d(\text{все координаты}) = \\ &= \sum (n_{k\alpha} | \delta \mathbf{R}_j | n_{k\alpha} + \delta n_{k\alpha}) I_j = \\ &= G_{k\alpha} \cdot I_0 \left[\hbar \frac{2n_{k\alpha} + 1 + \delta n_{k\alpha}}{8M\omega_{k\alpha}} \right]^{1/2} \sum_j e^{i(\hbar \delta n_{k\alpha} \mathbf{k} + \mathbf{P}' - \mathbf{P}) \cdot \mathbf{R}_j / \hbar}. \quad (17.111) \end{aligned}$$

Сумма экспоненциальных членов отлична здесь от нуля лишь при условии ²⁾

$$\mathbf{P}' + \hbar \delta n_{k\alpha} \mathbf{k} = \mathbf{P}, \quad (17.112)$$

так что матричные элементы отличны от нуля только для переходов между такими состояниями, которым соответствует одно и то же значение величины

$$\mathbf{P} + \hbar \sum n_{k\alpha} \mathbf{k} \quad (17.113)$$

так, как если бы импульс фонона $\hbar \sum n_{k\alpha} \mathbf{k}$ совместно с импульсом электрона \mathbf{P} подчинялся закону сохранения. Мы знаем, однако, что импульс системы в целом есть \mathbf{P}_T , так что этот закон сохранения имеет иной смысл, как об этом уже было сказано в конце § 4 и 5.

¹⁾ Это заключение основано на том обстоятельстве, что хотя в общем случае ни $\psi_{\mathbf{P}}$, ни $\psi'_{\mathbf{P}}$ не подчиняются периодическим граничным условиям, произведение $\psi_{\mathbf{P}}^* \psi'_{\mathbf{P}}$ этим условиям удовлетворяет. — *Прим. авт.*

²⁾ В теории электропроводности металлов предполагаются возможными такие случаи, когда $\mathbf{P}' \pm \hbar \mathbf{k} - \mathbf{P}$ есть вектор, компоненты которого являются целыми кратными граней зоны Бриллюэна, так что этот вектор обладает периодичностью, т. е. соответствует вектору обратной решетки, умноженному на 2π . Такие переходы были названы Пайерлсом „процессами переброса“ [14]. — *Прим. авт.*

Подводя итоги, можно сказать следующее. Матричные элементы \mathcal{U}_1 отличны от нуля только при таких изменениях функции Φ , при которых

- а) не изменяется \mathbf{P}_T ;
- б) изменяется только одно из квантовых чисел n_k , причем оно становится равным $n_{k\alpha} \pm 1 \equiv n_{k\alpha} + \delta n_{k\alpha}$;
- в) \mathbf{P} переходит в \mathbf{P}' так, что

$$\mathbf{P} = \mathbf{P}' + \hbar \delta n_{k\alpha} \mathbf{k}.$$

При любых значениях \mathbf{P} и \mathbf{k} всегда существует такое возможное значение \mathbf{P}' , которое будет удовлетворять условию в), так как подобно $\hbar \mathbf{k}$ разности векторов \mathbf{P} и \mathbf{P}' являются векторами неограниченной решетки.

Если эти условия удовлетворяются, то все члены, стоящие под знаком суммы по \mathbf{R}_j , одинаковы и добавляются к N_s , так что (не учитывая фазового множителя, абсолютное значение которого равно единице) имеем

$$\begin{aligned} (\mathbf{P} n_{k\alpha} | \mathcal{U}_1 | \mathbf{P}' n_{k\alpha} + \delta n_{k\alpha}) &= \\ &= \mathbf{G}_{k\alpha} \cdot \mathbf{I}_0(\mathbf{P}, \mathbf{P}') N_s \left[\hbar \frac{2n_{k\alpha} + 1 + \delta n_{k\alpha}}{8M\omega_{k\alpha}} \right]^{1/2}. \end{aligned} \quad (17.114)$$

Напомним, что N_s есть число ячеек в кристалле и $M = N_s (M_a/2) = M_c/2$ (где M_a — масса атома); \mathbf{I}_0 зависит от \mathbf{P} и \mathbf{P}' .

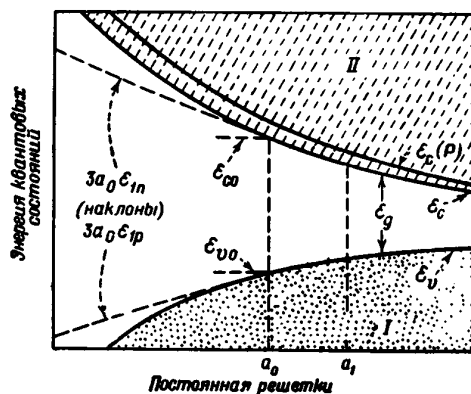
Для вычисления матричного элемента $\mathbf{G}_{k\alpha} \cdot \mathbf{I}_0$ разработан ряд методов, основанных на приближенном определении изменения потенциальной энергии \mathcal{U}_1 при смещении ядер. Мы рассмотрим здесь один из таких методов, предложенный недавно Бардином и Шокли¹⁾. Этот метод дает возможность вычислить матричный элемент \mathcal{U}_1 , исходя из некоторых общих свойств энергетических зон, а также сопоставить результаты теории с экспериментальными данными.

Метод потенциалов деформации. Сущность излагаемого здесь метода основана на предположении о том, что деформация, вызванная колебаниями решетки, обуславливает локальные изменения размеров и формы элементарной ячейки кристалла. Фиг. 143 иллюстрирует изменения зонной структуры энергетического спектра, которых можно ожидать при изменении постоянной решетки в случае таких кристаллов, как кремний или германий. Если бы деформация приводила к простому изменению постоянной решетки, не сопровождающемуся изменением симметрии кристалла, то энергия электрона на дне зоны проводимости определялась бы кривой \mathcal{E}_c . Энергия электрона, находящегося в некотором состоянии вблизи дна зоны, определялась бы при этом кривой $\mathcal{E}_c(\mathbf{P})$. Фиг. 143 иллюстрирует влияние чистого

¹⁾ См. [15] и [16]. В последней работе показано, что к потенциалам деформации может быть применен метод «эффективной массы». — *Прим. авт.*

сжатия или же расширения. Ниже мы исследуем случай деформации более общего типа.

Вычисление изменения энергии электрона в различных участках деформированного кристалла по наклону кривых, приведенных на фиг. 143, к сожалению, сопряжено со следующей трудностью. В случае однородно деформированного кристалла определение потенциальной энергии $\mathcal{U}_0(\mathbf{r})$, выбираемой за нуль отсчета, не однозначно. Численные расчеты по кривым, изображенным на фиг. 143, связаны поэтому с известной степенью произвола. Например, при построении



Фиг. 143. Зависимость энергии от постоянной решетки.

I — зона валентных связей; II — зона проводимости.

таких кривых для алмаза (см. фиг. 54, стр. 168) Кимбалл предположил, что вблизи каждого из атомов кристалла потенциальная энергия принимает одно и то же численное значение при любых значениях постоянной решетки. Как указывает сам автор, это предположение совершенно произвольно.

Кривые, изображенные на фиг. 143, можно, однако, интерпретировать следующим образом. Предположим, что постоянная решетки недеформированного кристалла равна a_0 . Представим себе, что после деформации некоторый участок кристалла остается неизменным, но что в результате постепенного изменения постоянной решетки возникает другая большая область, в которой кристалл однородно расширен так, что постоянная решетки приобрела значение a_1 . Можно думать, что при этих условиях соответствующие значения энергии электрона на дне зоны проводимости $\epsilon_c(a_0)$ и $\epsilon_c(a_1)$ будут определены вполне однозначно. В общем случае это, однако, не так, поскольку деформация обычно сопровождается возникновением электрического поля. Для любого данного вида деформации величина $\epsilon_c(a_1)$ может быть однозначно связана с величиной $\epsilon_c(a_0)$, но какая-либо другая

деформация, обуславливающая точно такое же значение постоянной решетки a_1 , может привести к появлению иных электрических полей, к иному среднему значению потенциала в области расширения кристалла и, следовательно, к другим значениям энергии электрона $\mathcal{E}_c(a_1)$. Учет наличия этих электрических полей связан со значительными трудностями; возникновение таких полей не исключено даже в случае неполярных кристаллов типа алмаза. Экспериментальные данные о температурной зависимости подвижности в кремнии и германии свидетельствуют, однако, о том, что в этих кристаллах при обычных условиях исследования такие эффекты существенной роли не играют. В конце настоящего параграфа будет показано, что при деформации, вызываемой продольными волнами, в кристалле возникает электростатический потенциал, изменяющийся как Δ/k^2 (Δ — объемное расширение, вызываемое q -волной), но что в кристаллах с ковалентным типом связи этот эффект мал по сравнению с эффектом смещения зон (см. задачу 1).

В отличие от значений энергий \mathcal{E}_c и \mathcal{E}_v ширина запрещенной зоны \mathcal{E}_g является однозначной функцией деформации решетки, так как любое изменение начала отсчета потенциальной энергии приводит к одинаковому смещению всех энергетических уровней. В гл. 12 были указаны методы экспериментального определения зависимости ширины запрещенной зоны от постоянной решетки.

Поскольку эффекты, связанные с возникновением в кристалле макроскопических электрических полей, при наличии которых потенциальная энергия электрона зависит от характера распределения деформации, предполагаются ничтожными, мы будем рассматривать \mathcal{E}_c и \mathcal{E}_v как однозначные функции локальной деформации. Если ограничиться рассмотрением членов, линейных относительно деформации, то для кубического кристалла

$$\mathcal{E}_c = \mathcal{E}_{c0} + \mathcal{E}_{1n}\Delta; \quad (17.115a)$$

сокращенно во всех уравнениях вплоть до (17.133) будем полагать

$$\mathcal{E} = \mathcal{E}_0 + \mathcal{E}_1\Delta, \quad (17.115a')$$

$$\mathcal{E}_v = \mathcal{E}_{v0} + \mathcal{E}_{1p}\Delta, \quad (17.115b)$$

$$\mathcal{E}_G = \mathcal{E}_{G0} + \mathcal{E}_{1G}\Delta, \quad (17.115b)$$

$$\mathcal{E}_{1G} = \mathcal{E}_{1n} - \mathcal{E}_{1p}. \quad (17.115\gamma)$$

Здесь Δ — объемное расширение:

$$\Delta = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} = \nabla_{R_i} \cdot \delta R_i,$$

а \mathcal{E}_{1n} и \mathcal{E}_{1p} — постоянные величины; индексы p и n введены здесь в связи с тем, что в дальнейшем нами будет установлено соотношение между этими величинами и другими параметрами, относящимися

соответственно к дыркам и электронам. Остальные компоненты деформации равны нулю в силу симметрии в условиях линейного приближения. Это можно проиллюстрировать на примере двух деформаций чистого сдвига, для одной из которых ϵ'_{xy} положительно и равно значению $-\epsilon'_{xy}$ для второй деформации. В плоскости $z=0$ такие деформации вызывают изменение угла между осями x и y соответственно на $\pm \delta\theta$; вследствие симметрии кубического кристалла энергия \mathcal{E}_0 в обоих случаях изменяется при этом одинаково. Следовательно, \mathcal{E}_c должно быть четной функцией ϵ_{xy} и член, линейный относительно ϵ_{xy} , должен обращаться в нуль.

В дальнейшем в данном параграфе речь будет идти о переходах электрона из одних состояний в другие, и мы будем пользоваться сокращенной формой записи (17.115а'), причем будем пренебрегать членами, учитывающими изменение эффективной массы при деформации. Анализируя ход кривой $\mathcal{E}_c(\mathbf{P})$, приведенной на фиг. 143, можно убедиться в том, что эти члены — порядка произведения \mathbf{P}^2 на некоторую функцию деформации. Действительно, минимум кривой \mathcal{E}_c , отвечающей зоне проводимости, соответствует условию $\mathbf{P}=0$ (или же углу зоны Бриллюэна). Кривая $\mathcal{E}_c(\mathbf{P})$ лежит выше кривой \mathcal{E}_0 на $m_n(\mathbf{P}^2/2)$. Если кривизна кривой, характеризующей зависимость \mathcal{E} от \mathbf{P}^2 , меняется с деформацией, то это эквивалентно изменению величины m_n ; при малых деформациях последнее должно быть, однако, линейной функцией деформации. Поэтому можно записать:

$$\mathcal{E}(\mathbf{P}, \epsilon) = \mathcal{E}_0(\mathbf{P}) + \mathcal{E}_1\Delta + \text{члены порядка произведения } \mathbf{P}^2 \text{ на линейную функцию деформации.} \quad (17.116)$$

Член $\mathcal{E}_1\Delta$ мы будем называть *потенциалом деформации*.

(В кристаллах со структурой алмаза имеют также место „оптические колебания“, при которых два атома, находящиеся в данной элементарной ячейке, движутся в противоположных направлениях. Если рассматриваемая энергетическая зона не вырождена, так что \mathcal{E} является однозначной функцией деформации, то эти оптические колебания не будут давать членов первого порядка. Это следует из того, что в случае длинных волн, интересующих нас при исследовании полупроводников, оптические колебания возникают в результате смещения одной кубической гранецентрированной решетки по отношению к другой, и в силу симметрии при этом не существует однозначной функции, линейной по отношению к компонентам смещения. Если зона вырождена, могут иметься также члены, линейные относительно отдельных компонент смещения¹⁾; среднее значение энергии взаимо-

¹⁾ Бардин предположил, что „оптические колебания“ могут играть существенную роль при поглощении света в инфракрасной области электронами проводимости или дырками [17]; Райдер и Шокли пришли к заключению, что оптические колебания существенны также при рассмотрении вопроса о потере энергии электронами в сильных электрических полях. — *Прим. авт.*

действия, однако, и в этом случае не будет зависеть от оптических колебаний. Исследование этих вопросов может быть осуществлено наилучшим образом с помощью теории представлений групп симметрии¹⁾; оно выходит за рамки настоящей книги.)

Докажем теперь теорему, лежащую в основе метода Бардина — Шокли: *Матричный элемент \mathcal{U}_1 равен матричному элементу потенциала деформации при условии, что речь идет о переходах электронов между состояниями, характеризующимися малыми значениями \mathbf{P} и \mathbf{P}'* . Интеграл, который следует вычислить, чтобы доказать эту теорему, получается из выражения (17.107) в результате интегрирования по координатам q :

$$\begin{aligned} (\mathbf{P}', n | \mathcal{U}_1 | \mathbf{P}, n \pm 1) = & (n | q_c | n \pm 1) \int \psi_{\mathbf{P}'}^*(\mathbf{r}) \mathcal{U}_c(\mathbf{r}) \psi_{\mathbf{P}}(\mathbf{r}) dV_{\mathbf{r}} + \\ & + (n | q_s | n \pm 1) \int \psi_{\mathbf{P}'}^*(\mathbf{r}) \mathcal{U}_s(\mathbf{r}) \psi_{\mathbf{P}}(\mathbf{r}) dV_{\mathbf{r}}. \quad (17.117) \end{aligned}$$

С целью упрощения последующих вычислений здесь опущен индекс i , причем \mathbf{P} заменено на \mathbf{P}' , а \mathbf{P}' — на \mathbf{P} .

Два интеграла, входящие в выражение (17.117), весьма сходны друг с другом; рассмотрим первый из них. В этом интеграле \mathcal{U}_c соответствует энергии возмущения, обусловленного деформацией вида $\delta R_i = \mathbf{G}\eta \cos \mathbf{k} \cdot \mathbf{R}_i$, где η — некоторый параметр, который мы будем стремиться к нулю, чтобы исследовать роль членов, содержащих производные $\partial/\partial q_c$. Предположим, таким образом, что η — малая величина, и в последующих выражениях будем пренебрегать членами, содержащими η^2 . При $\eta \rightarrow 0$ очевидно, что $\mathcal{U}_c = \partial \mathcal{U}_1 / \partial \eta$. Деформация решетки, определенная выше через смещения ядер, в случае относительно малых значений \mathbf{k} и длинных волн, интересующих нас в исследуемой задаче о движении избыточного электрона, медленно изменяется при переходе от одной из элементарных ячеек к следующей. Поэтому можно ввести в рассмотрение объемное расширение, определив его как

$$\Delta(\mathbf{r}) = \nabla \cdot \mathbf{G}\eta \cos \mathbf{k} \cdot \mathbf{r} = -\eta \mathbf{G} \cdot \mathbf{k} \sin \mathbf{k} \cdot \mathbf{r}.$$

Вычислим записанный выше интеграл в предположении, что такая деформация действительно имеет место в кристалле, а η столь мало, что членами, нелинейными относительно η , можно пренебречь. Пользуясь приближенными методами, докажем, что

$$\begin{aligned} \int \psi_{\mathbf{P}'}^*(\mathbf{r}) \eta \mathcal{U}_c(\mathbf{r}) \psi_{\mathbf{P}}(\mathbf{r}) dV_{\mathbf{r}} = & \frac{\eta}{V} \int \exp \left[i(\mathbf{P} - \mathbf{P}') \cdot \frac{\mathbf{r}}{\hbar} \right] \mathcal{E}_1 \Delta(\mathbf{r}) dV_{\mathbf{r}} + \\ & + \text{члены порядка } \eta \frac{(\mathbf{P}^2 \text{ или } \mathbf{P}'^2)}{m}, \quad (17.118) \end{aligned}$$

¹⁾ См., например, [5]. — *Прим. авт.*

т. е., что при малых значениях P и P' матричный элемент может быть выражен через \mathcal{E}_1 так, как если бы энергия возмущения равнялась

$$\delta\mathcal{U}(\mathbf{r}) = \mathcal{E}_1\Delta(\mathbf{r}).$$

Для доказательства этой теоремы необходимо ввести в рассмотрение точные значения волновых функций, описывающих движение электрона в *однородно деформированном* кристалле.

Состояние деформации в любой точке кристалла, деформированного η -волной, мы будем характеризовать заданием функции $\eta_\varepsilon(\mathbf{r})$, где ε — сокращенное обозначение e_{xx} , e_{yy} , e_{zz} и т. д., а также любой трансляции или поворота. Элементарная ячейка, расположенная в точке \mathbf{r}_a , деформирована при этом на $\eta_\varepsilon(\mathbf{r}_a)$. Определим теперь волновую функцию, отвечающую этой ячейке, представив себе, что последняя периодически повторяется во всех направлениях так, что мы имеем однородно деформированную решетку, элементарная ячейка которой есть ячейка, расположенная в точке \mathbf{r}_a . Волновая функция, описывающая движение электрона в такой решетке, имеет вид

$$\psi_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)] = e^{i\mathbf{P}\cdot\mathbf{r}/\hbar} u_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)]; \quad (17.119)$$

в случае малых деформаций значения функции $u_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)]$ в ячейке \mathbf{r}_a будут лишь очень незначительно отличаться от значений функции $u_{\mathbf{P}}(\mathbf{r} - \delta\mathbf{R}_a)$. Потенциальная энергия электрона в ячейке \mathbf{r}_a равна $\mathcal{U}_0(\mathbf{r}) + \eta\mathcal{U}_\varepsilon(\mathbf{r})$, а волновая функция $\psi[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)]$ в области этой ячейки удовлетворяет волновому уравнению

$$\begin{aligned} \left[\frac{1}{2m} p^2 + \mathcal{U}_0(\mathbf{r}) + \eta\mathcal{U}_\varepsilon(\mathbf{r}) \right] \psi_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)] &= \\ &= [\mathcal{H}_r + \eta\mathcal{U}_\varepsilon(\mathbf{r})] \psi_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)] = \\ &= [\mathcal{E}_0(\mathbf{P}) + \mathcal{E}_1\eta\Delta(\mathbf{r}_a)] \psi_{\mathbf{P}}[\mathbf{r}, \eta_\varepsilon(\mathbf{r}_a)]. \end{aligned} \quad (17.120)$$

Мы пренебрегли здесь членами порядка $\hbar P^2/m_n$ [см. (17.116)], учитывая зависимость эффективной массы m_n от деформации. Следует отметить, что оператор p^2 в этом уравнении действует только на \mathbf{r} и не действует на \mathbf{r}_a .

Вывод уравнения (17.120) — один из наиболее существенных этапов нашего исследования. Отметим, что $\mathcal{U}_\varepsilon(\mathbf{r})$ есть вещественная функция положения и что при анализе значений $\mathcal{E}_0 + \mathcal{E}_1\Delta$ мы имеем дело только с вещественными функциями. Позднее мы объединим члены, содержащие \mathcal{U}_ε и \mathcal{U}_δ , и получим выражение, в которое входит комплексный множитель $\exp(i\mathbf{k}\cdot\mathbf{r})$; это будет сделано, однако, лишь после исследования интегралов, содержащих функции \mathcal{U}_ε и \mathcal{U}_δ в отдельности.

Определим волновую функцию так, чтобы она учитывала изменение $\varepsilon(\mathbf{r})$. Такая функция имеет вид

$$\psi_{\mathbf{P}}[\mathbf{r}, \eta\varepsilon(\mathbf{r})] = e^{i\mathbf{P}\cdot\mathbf{r}/\hbar} u_{\mathbf{P}}[\mathbf{r}, \eta\varepsilon(\mathbf{r})] = \psi_{\mathbf{P}}(\mathbf{r}) + \delta\psi_{\mathbf{P}}(\mathbf{r}). \quad (17.121)$$

Форма этой функции изменяется при переходе от ячейки к ячейке в соответствии с периодичностью решетки, образованной повторением элементарной ячейки. В действительности форма функции $\psi_{\mathbf{P}}[\mathbf{r}, \eta\varepsilon(\mathbf{r})]$ слегка меняется даже в пределах каждой элементарной ячейки, поскольку на протяжении любой ячейки $\varepsilon(\mathbf{r})$ несколько изменяется. При малых η член $\delta\psi$, очевидно, линеен относительно η и поэтому можно пренебречь его произведением на другие члены, пропорциональные η .

Докажем теперь теорему, выражаемую равенством (17.118), рассматрив с этой целью интеграл

$$I = \int \psi_{\mathbf{P}}^*(\mathbf{r}) [\mathcal{H}_{\mathbf{r}} + \eta \mathcal{U}_{\mathbf{c}}(\mathbf{r})] \psi_{\mathbf{P}}[\mathbf{r}, \eta\varepsilon(\mathbf{r})] dV_{\mathbf{r}}. \quad (17.122a)$$

Применяя оператор, заключенный в квадратные скобки, к функции $\psi_{\mathbf{P}}$, получаем

$$I = \int \psi_{\mathbf{P}}^* [\mathcal{E}_0(\mathbf{P}) + \mathcal{E}_1 \eta \Delta(\mathbf{r})] (\psi_{\mathbf{P}} + \delta\psi_{\mathbf{P}}) dV_{\mathbf{r}} + \\ + \text{члены, содержащие } \frac{\partial u_{\mathbf{P}}[\mathbf{r}, \eta\varepsilon(\mathbf{r})]}{\partial \varepsilon} \eta \nabla \varepsilon(\mathbf{r}). \quad (17.122b)$$

Можно показать, что последние члены этого выражения — порядка $\eta P^2/m$ (см. задачу 2). Учитывая, что функции ψ ортогональны, а членом $\eta \delta\psi_{\mathbf{P}}$ (пропорциональным η^2) можно пренебречь, находим

$$I \approx \eta \mathcal{E}_1 \int \psi_{\mathbf{P}}^* \Delta \psi_{\mathbf{P}} dV_{\mathbf{r}} + \mathcal{E}_0(\mathbf{P}) \int \psi_{\mathbf{P}}^* \delta\psi_{\mathbf{P}} dV_{\mathbf{r}}. \quad (17.123)$$

Вычислим теперь интеграл I другим методом, разложив в ряд последние два множителя подинтегральной функции и используя то обстоятельство, что $\psi_{\mathbf{P}}$ и $\psi_{\mathbf{P}'}$ — ортогональные собственные функции оператора $\mathcal{H}_{\mathbf{r}}$. Ограничивая разложение членами порядка η^2 , имеем

$$I = \int \psi_{\mathbf{P}}^* \eta \mathcal{U}_{\mathbf{c}} \psi_{\mathbf{P}} dV_{\mathbf{r}} + \int \psi_{\mathbf{P}}^* \mathcal{H}_{\mathbf{r}} \delta\psi_{\mathbf{P}} dV_{\mathbf{r}} = \\ = \eta \int \psi_{\mathbf{P}}^* \mathcal{U}_{\mathbf{c}} \psi_{\mathbf{P}} dV_{\mathbf{r}} + \mathcal{E}_0(\mathbf{P}') \int \psi_{\mathbf{P}}^* \delta\psi_{\mathbf{P}} dV_{\mathbf{r}}; \quad (17.124)$$

вторая скобка этого выражения получена при учете эрмитовых свойств оператора $\mathcal{H}_{\mathbf{r}}$ (см. задачу 3). Совместное решение уравнений (17.123) и (17.124) приводит к следующему значению интеграла:

$$\eta \int \psi_{\mathbf{P}}^* \mathcal{U}_{\mathbf{c}} \psi_{\mathbf{P}} dV_{\mathbf{r}} = \eta \mathcal{E}_1 \int \psi_{\mathbf{P}}^* \psi_{\mathbf{P}} \Delta dV_{\mathbf{r}} + \\ + [\mathcal{E}_0(\mathbf{P}) - \mathcal{E}_0(\mathbf{P}')] \int \psi_{\mathbf{P}}^* \delta\psi_{\mathbf{P}} dV_{\mathbf{r}}. \quad (17.125a)$$

Последний член здесь порядка $\eta P^2/2m_n$, и им можно пренебречь. Искомый интеграл от \mathcal{U}_c получим, разделив (17.125а) на η :

$$\int \Psi_{\mathbf{P}}^* \mathcal{U}_c \Psi_{\mathbf{P}} dV_{\mathbf{r}} = \mathcal{E}_1 \int \Psi_{\mathbf{P}'}^* \Psi_{\mathbf{P}} \Delta dV_{\mathbf{r}}. \quad (17.125б)$$

Этот результат показывает, что „интеграл взаимодействия“ $\int \Psi_{\mathbf{P}}^* \mathcal{U}_c \Psi_{\mathbf{P}} dV_{\mathbf{r}}$ имеет в точности такое же значение, которое мы получили бы, введя в рассмотрение потенциал деформации $\mathcal{E}_1 \Delta$ (см. гл. 11).

Преобразуем теперь (17.117) так, чтобы получить выражение того же вида, что и формула (17.65), определяющая значение δR_i . При этом заменим в наших уравнениях \mathbf{P} на \mathbf{P}' и \mathbf{P}' на \mathbf{P} для того, чтобы получить согласие с прежними обозначениями. Равенство (17.117) в таком случае может быть переписано в виде

$$\begin{aligned} (\mathbf{P}, n_{k\alpha} | \mathcal{U}_1 | \mathbf{P}', n_{k\alpha} + \delta n_{k\alpha}) &= \\ &= \mathcal{E}_1 \int \Psi_{\mathbf{P}}^* \Psi_{\mathbf{P}'} (n_{k\alpha} | \nabla \cdot \mathbf{G}_{k\alpha} (q_c \cos \mathbf{k} \cdot \mathbf{r} + q_s \sin \mathbf{k} \cdot \mathbf{r}) | n_{k\alpha} + \delta n_{k\alpha}) dV_{\mathbf{r}} = \\ &= \mathcal{E}_1 \mathbf{G}_{k\alpha} \cdot \mathbf{k} \int \Psi_{\mathbf{P}}^* \Psi_{\mathbf{P}'} \exp(i \delta n_{k\alpha} \mathbf{k} \cdot \mathbf{r}) dV_{\mathbf{r}} \cdot \left[\hbar \frac{2n_{k\alpha} + 1 + \delta n_{k\alpha}}{8M\omega_{k\alpha}} \right]^{1/2}. \end{aligned} \quad (17.126)$$

Первое из этих выражений получаем, выразив объемное расширение, вызываемое q_c - и q_s -волнами, через дивергенцию, а второе — воспользовавшись формулой для матричного элемента $q_c \cos \mathbf{k} \cdot \mathbf{r} + q_s \sin \mathbf{k} \cdot \mathbf{r}$, получаемой из формулы (17.83) в результате замены \mathbf{R}_i на \mathbf{r} . Здесь не учтены некоторые множители, абсолютное значение которых равно единице, например i и i^2 ; это обстоятельство, однако, несущественно, поскольку вероятность перехода определяется квадратом модуля выражения (17.126).

В согласии с формулами (17.112) и (17.113), интеграл, входящий в (17.126), отличен от нуля только при условии

$$\mathbf{P} = \mathbf{P}' + \hbar \mathbf{k} \delta n_{k\alpha}.$$

Если это условие выполняется, он равен

$$\frac{1}{V} \int u_{\mathbf{P}}^*(\mathbf{r}) u_{\mathbf{P}'}(\mathbf{r}) dV_{\mathbf{r}}; \quad (17.127)$$

напомним, что функция $u_{\mathbf{P}}$ нормирована к единице объема. В случае невырожденной зоны и малых значений \mathbf{P} можно записать $u_{\mathbf{P}}(\mathbf{r})$ в форме

$$u_{\mathbf{P}}(\mathbf{r}) = u_0(\mathbf{r}) + P_x u_x(\mathbf{r}) + P_y u_y(\mathbf{r}) + P_z u_z(\mathbf{r}),$$

причем функции u могут быть выбраны так, чтобы они удовлетворяли условиям ортогональности. Воспользовавшись этим равенством, можно доказать, что с точностью до членов порядка \mathbf{P}^2 интеграл

(17.127) равен единице. Таким образом, для разрешенных переходов имеем окончательно

$$\begin{aligned} (P, n_{k\alpha} | \mathcal{U}_1 | P', n_{k\alpha} + \delta n_{k\alpha}) = \\ = \mathcal{E}_1 \mathbf{G}_{k\alpha} \cdot \mathbf{k} \left[\hbar \frac{2n_{k\alpha} + 1 + \delta n_{k\alpha}}{8M\omega_{k\alpha}} \right]^{1/2}. \end{aligned} \quad (17.128)$$

Возможные значения $\delta n_{k\alpha}$ равны здесь ± 1 [см. (17.79)—(17.84)], а M составляет половину массы кристалла [см. (17.42)]. Формула (17.128) эквивалентна аналогичной формуле, полученной ранее другими авторами, отличаясь от нее иной интерпретацией величины \mathcal{E}_1 ¹⁾.

Связь матричного элемента с упругими постоянными. Для тех колебаний решетки, которые представляют интерес при исследовании свойств полупроводников, значения $\hbar k$ малы по сравнению с величиной $\hbar/2a$, так как изменения P при этом незначительны и, следовательно, длины k -волн велики по сравнению с постоянной решетки a . По отношению к таким колебаниям твердое тело может рассматриваться как континуум, так что для их описания можно воспользоваться макроскопической теорией упругости.

При вычислении матричного элемента стало очевидным, что существенную роль играют те колебания решетки, которые обуславливают расширение кристалла и обладают большой составляющей единичного вектора $\mathbf{G}_{k\alpha}$ в направлении \mathbf{k} . В случае изотропного твердого тела мы имеем одно чисто продольное и два чисто поперечных длинноволновых колебания; в кубическом кристалле аналогичная картина имеет место для волн, распространяющихся в направлениях [100], [110] и [111]. Для других направлений мы имеем смесь продольной и поперечной поляризации. Значения упругих постоянных для кристаллов германия (в *дин/см²*)²⁾:

$$c_{11} = 1,292 \cdot 10^{12}; \quad c_{12} = 0,479 \cdot 10^{12}; \quad c_{44} = 0,670 \cdot 10^{12}.$$

В табл. 14 приведены значения скоростей распространения упругих колебаний для германия (в *см/сек*).

Приведенные здесь данные заимствованы из работы Бонда, Мэзона и др. [19].

Для других направлений продольная скорость распространения лежит в интервале между своими экстремальными значениями, соответствующими направлениям [100] и [111]. Хотя в этом случае

¹⁾ \mathcal{E}_1 равно постоянной C , фигурирующей у Вильсона [12], и составляет $2/3$ значения постоянной C , вводимой Зоммерфельдом и Бете [18], а также Зейтцем [5]. — *Прим. авт.*

²⁾ Новые данные об упругих постоянных и их температурной зависимости для монокристаллов кремния приведены в заметке Мак-Скинмина и др. [22]. — *Прим. ред.*

Таблица 14

Направление поляризации	Направление распространения (параллельно \mathbf{k})		
	[100]	[110]	[111]
Продольное	$4,92 \cdot 10^5$ (а)	$5,39 \cdot 10^5$ (б)	$5,54 \cdot 10^5$ (в)
Поперечное [001] . . .	$3,54 \cdot 10^6$ (г)	$3,54 \cdot 10^6$ (д)	—
„ [110]	—	$2,75 \cdot 10^6$ (е)	$3,03 \cdot 10^5$ (ж)

Формулы, определяющие значение ρc^2 , имеют вид:

$$(а) c_{11}, \quad (б) \frac{c_{11} + c_{12} + 2c_{44}}{2}, \quad (в) \frac{c_{11} + 2c_{12} + 4c_{44}}{3}, \quad (г) c_{44}$$

$$(д) c_{44}, \quad (е) \frac{c_{11} - c_{12}}{2}, \quad (ж) \frac{c_{11} - c_{12} + c_{44}}{3}.$$

По данным о постоянной решетки, $\rho = 5,85 \text{ г/см}^3$.

колебания не будут чисто продольными, значение величины $(\mathbf{G}_{k\alpha} \cdot \mathbf{k})^2$, входящей в определение вероятности перехода, почти равно k^2 , и, поскольку

$$(\mathbf{G}_{k_1} \cdot \mathbf{k})^2 + (\mathbf{G}_{k_2} \cdot \mathbf{k})^2 + (\mathbf{G}_{k_3} \cdot \mathbf{k})^2 = k^2,$$

поперечные колебания играют при этом весьма малую роль. Мы видим, таким образом, что обычное приближение, в котором кубический кристалл трактуется как изотропный и при исследовании процессов рассеяния учитываются только продольные колебания, приводит к относительно малым ошибкам.

Поскольку значения продольных скоростей для различных направлений близки друг к другу, мы введем в рассмотрение среднее значение упругой постоянной для продольного направления c_{11} . В качестве c_{11} произвольно выберем значение упругой постоянной для продольной волны, распространяющейся в направлении [110]; оно является промежуточным между экстремальными значениями, отвечающими направлениям [100] и [111]. Введем также в рассмотрение среднюю скорость распространения, определив ее как

$$c = \left(\frac{c_{11}}{\rho} \right)^{1/2}.$$

Пренебрегая учетом поперечных колебаний, мы видим, что для каждого перехода из состояния P в состояние P' или же из состояния P_0 в состояние P_j (обозначения, принятые в § 2) существуют только две возможности:

$$\hbar \mathbf{k} = P' - P \quad \text{при} \quad \delta n_k = +1$$

либо

$$\hbar \mathbf{k} = P - P' \quad \text{при} \quad \delta n_k = -1.$$

Таким образом, если не считать этой неопределенности, то конечное состояние однозначно характеризуется значением импульса P' .

Покажем теперь, каким образом оба матричных элемента можно трактовать как один матричный элемент $|\mathcal{U}_{0j}|^2$, рассмотренный в § 2.

Воспользовавшись определением c_{II} , можно представить матричный элемент (17.128) в форме, наиболее удобной для запоминания. С этой целью введем в рассмотрение среднее значение энергии $\langle \mathcal{E}_{k\alpha} \rangle$ нормального колебания до и после столкновения, где

$$\langle \mathcal{E}_{k\alpha} \rangle = \frac{\hbar \omega_{k\alpha} (2n_{k\alpha} + 1 + \delta n_{k\alpha})}{2} = \frac{\hbar \omega_{k\alpha} \left(n_{k\alpha} + \frac{1}{2} + n'_{k\alpha} + \frac{1}{2} \right)}{2}.$$

Отметим, кроме того, что

$$(I) \quad M = \frac{M_0}{2} = \frac{1}{2} \rho V,$$

где ρ — плотность кристалла, а V — его объем, и

$$(II) \quad \frac{\omega_{k\alpha}}{k} = c_{k\alpha} \approx \sqrt{\frac{c_{II}}{\rho}},$$

где $c_{k\alpha}$ — скорость распространения волны, характеризуемой параметрами $k\alpha$. Воспользовавшись этими соотношениями, преобразуем выражение (17.128) к виду

$$\begin{aligned} |(P n_{k\alpha} | \mathcal{U}_i | P' n_{k\alpha} + \delta n_{k\alpha})|^2 &= \frac{\mathcal{E}_1^2 (2n_{k\alpha} + 1 + \delta n_{k\alpha})}{4 V c_{II}} = \\ &= \mathcal{E}_1^2 \frac{\langle \mathcal{E}_{k\alpha} \rangle}{2 V c_{II}} = \mathcal{E}_1^2 \frac{\langle \Delta^2 \rangle}{2}. \end{aligned} \quad (17.129)$$

Здесь $\langle \Delta^2 \rangle$ — среднее значение квадрата объемного расширения, соответствующее средней энергии $\langle \mathcal{E}_{k\alpha} \rangle$ и определяемое следующим образом. В случае упругих колебаний энергия в среднем подразделяется поровну на потенциальную и кинетическую. Плотность потенциальной энергии при этом равна $1/2 c_{II} (\epsilon_{kk})^2$, где ϵ_{kk} — деформация в продольном направлении, равная объемному расширению Δ , вызванному продольной волной. Полагая среднее значение потенциальной энергии равным половине полной энергии, получаем

$$\frac{1}{2} V c_{II} \langle \Delta^2 \rangle = \frac{1}{2} \langle \mathcal{E}_{k\alpha} \rangle,$$

что приводит к последнему из записанных выше уравнений.

Этот результат может быть сформулирован словесно следующим образом: *квадрат матричного элемента можно определить, вычислив среднее значение квадрата объемного расширения, соответствующее энергии бегущей волны, усредненной до и после столкновения. Это среднее значение квадрата объемного расширения $\langle \Delta^2 \rangle$ соответствует среднему значению квадрата потенциала деформации $\mathcal{E}_1^2 \langle \Delta^2 \rangle$, а матричный элемент равен половине численной*

величины последнего. В такой формулировке этот результат может быть применен также и в случае волн, поляризация которых лишь частично является продольной.

В § 2 мы приближенно предполагали, что при переходе $\mathbf{P} \rightarrow \mathbf{P}'$ изменение энергии системы полностью обусловлено изменением энергии одного лишь электрона. Можно показать, что в интересующих нас условиях это предположение хорошо оправдывается. Рассмотрим сперва требование сохранения энергии при переходе, согласно которому изменение $\mathcal{E}(\mathbf{P})$ равно $\pm \hbar\omega_{\mathbf{k}}$. В приближенном случае сферической симметрии, когда $\mathcal{E}(\mathbf{P}) = \mathbf{P}^2/2m$, изменение энергии можно выразить через $\delta|\mathbf{P}|$:

$$\delta\mathcal{E}|\mathbf{P}| = \frac{|\mathbf{P}|\delta|\mathbf{P}|}{m} = |\mathbf{v}|\delta|\mathbf{P}|;$$

с другой стороны, изменение энергии колебаний решетки равно

$$\hbar\omega = \frac{\hbar 2\pi c}{\lambda} = \hbar c |\mathbf{k}| = c |\mathbf{P}' - \mathbf{P}|.$$

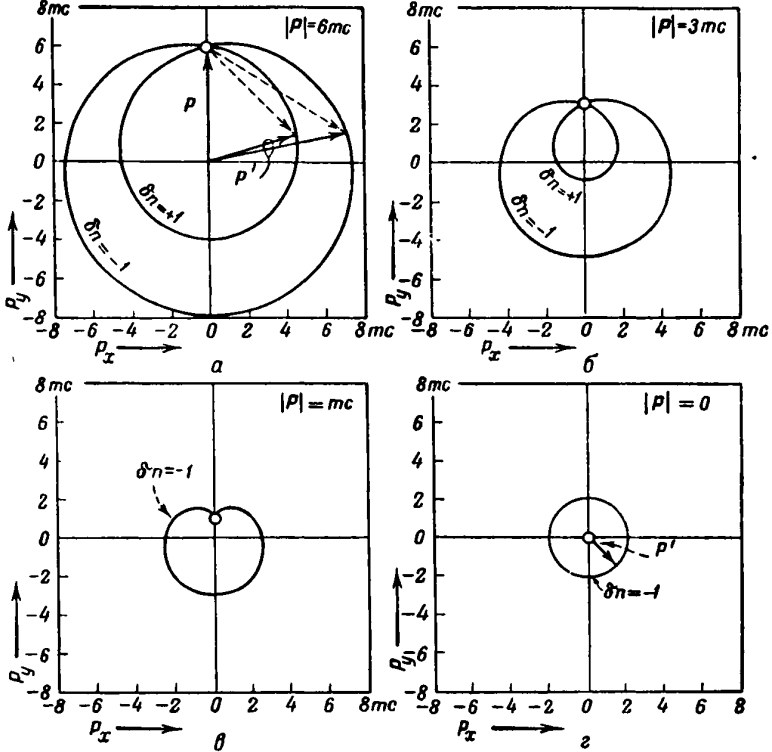
Приравнивая эти два выражения, получаем

$$\delta|\mathbf{P}| = \frac{c}{|\mathbf{v}|} |\mathbf{P}' - \mathbf{P}|. \quad (17.130)$$

При комнатной температуре $|\mathbf{v}|$ — порядка 10^7 см/сек, тогда как c примерно в 20 раз меньше. Таким образом, изменения $|\mathbf{P}|$ составляют только несколько процентов от $|\mathbf{P}|$. На фиг. 144 изображены конечные состояния \mathbf{P}' , в которые происходят переходы из состояний \mathbf{P} . При больших значениях разности $|\mathbf{P} - \mathbf{P}'|$ значение \mathbf{k} велико и расстояние между двумя поверхностями таково, что $\delta\mathcal{E}(\mathbf{P}) = 2\hbar\omega_{\mathbf{k}}$. Если \mathbf{P} принимает значение, отличное от значений, отвечающих одной из этих поверхностей, то благодаря наличию члена $\mathcal{E}(\mathbf{P})$ полная энергия будет при этом изменяться как $\mathbf{v} \cdot \delta\mathbf{P}$, а благодаря наличию члена $\hbar\omega_{\mathbf{k}}$ — как произведение c на компоненту $\delta\mathbf{P}$, параллельную вектору \mathbf{k} . Отсюда следует, что изменения энергии в значительной степени обусловлены изменениями величины $|\mathbf{P}|$. Поскольку расстояние между поверхностями составляет всего лишь несколько процентов от $|\mathbf{P}|$, мы можем рассматривать совокупность обеих поверхностей как одну сферическую поверхность. При очень малых значениях энергии электронов $|\mathbf{v}|$, однако, сравнимо с c . В этих условиях поверхности приобретают форму, изображенную на фиг. 144, б. При еще более низких значениях энергии они приобретают форму, показанную на фиг. 144, в; в этом случае переходы, при которых электроны теряют свою энергию, оказываются невозможными. Однако эти интересные случаи играют существенную роль только при очень низких температурах и лежат вне области рассматриваемых здесь явлений ¹⁾.

¹⁾ По этому вопросу см., например, [5]. — *Прим. авт.*

Соответственно мы будем аппроксимировать конечные состояния P' с помощью сферической поверхности. Переходы от P к P' определяются двумя независимыми процессами, отвечающими условиям $\delta n = +1$ и $\delta n = -1$. Вероятности этих процессов следовало бы вычислить



Фиг. 144. Начальные и конечные значения P для переходов, не сопровождающихся изменениями энергии.

a — два почти сферических слоя, соответствующих испусканию и поглощению фотона, $|P_0| = 6 m c$; $б$ — то же для электрона с $|P_0| = 3 m c$; $в$ — случай электрона с $|P_0| = m c$, потеря энергии электроном невозможна; $г$ — случай $P_0 = 0$.

в отдельности и затем сложить их. Однако при этом мы получили бы точно такой же результат, как если бы просто сложили квадраты абсолютных значений двух матричных элементов. Эффективное значение $|\mathcal{U}_{j0}|^2$, приведенное в § 2, где P и P' заменены на P_0 и P_j , равно

$$|\mathcal{U}_{j0}|^2 = |(P_0 n_{k\alpha} |\mathcal{U}_1| P_j n_{k\alpha} + 1)|^2 + |(P_0 n_{-k\alpha} |\mathcal{U}_1| P_j n_{-k\alpha} - 1)|^2 = \\ = \mathcal{E}_1^2 \frac{\hbar \omega_{k\alpha} (n_{k\alpha} + 1) + \hbar \omega_{-k\alpha} n_{-k\alpha}}{2Vc_{\mu}}. \quad (17.131)$$

Так как колебания, характеризующиеся параметрами $k\alpha$ и $-\mathbf{k}\alpha$, обладают одной и той же частотой, оба квантовых числа n здесь одинаковы. Далее, для рассматриваемых переходов энергия фононов будет порядка

$$\hbar\omega = \frac{\hbar c 2\pi}{\lambda} = \hbar c |\mathbf{k}| \approx c |\mathbf{P}|;$$

поскольку эта величина значительно меньше $|\mathbf{v}| |\mathbf{P}| \approx kT$, интересующие нас колебания решетки соответствуют области энергий, лежащей выше характеристической температуры, и, следовательно,

$$\hbar\omega_{k\alpha} \left(n_{k\alpha} + \frac{1}{2} \right) = kT.$$

Воспользовавшись этим соотношением для вычисления $(2n_{k\alpha} + 1)$, получаем выражение, приведенное в § 2:

$$|\mathcal{U}_{j0}|^2 = \mathcal{E}_2^2 \frac{kT}{Vc_{II}} = \mathcal{E}_1^2 \langle \Delta^2 \rangle = \frac{AT}{V}; \quad A = \frac{\mathcal{E}_1^2 k}{Vc_{II}}, \quad (17.132)$$

где A не зависит от температуры. Значение $\langle \Delta^2 \rangle$ соответствует упругой энергии $c_{II} \epsilon_{kk}^2/2$, равной $kT/2$; введение в формулу одной и той же величины c_{II} для всех направлений распространения связано здесь с приближенным предположением об изотропности среды. Численное значение выражения (17.132) вдвое больше численного значения выражения (17.129), так как оно учитывает переходы, сопровождающиеся как поглощением, так и испусканием фононов.

Подстановка (17.132) в формулу для средней длины свободного пробега, выведенную в § 2, дает

$$\frac{1}{l(T)} = \frac{ATm_n^2}{\pi\hbar^4} = \frac{\mathcal{E}_{1n}^2 kTm_n^2}{\pi\hbar^4 c_{II}}. \quad (17.133)$$

Мы заменили здесь \mathcal{E}_1 на \mathcal{E}_{1n} , чтобы напомнить о том, что эта формула выведена для случая избыточного электрона. В следующем параграфе будет показано, что соответствующая формула, определяющая среднюю длину свободного пробега дырки, отличается только тем, что \mathcal{E}_{1n} заменено в ней на \mathcal{E}_{1p} , а m_n — на m_p .

ЗАДАЧИ

1. Вычислить изменения электростатического потенциала в кристалле, вызванные стоячей q -волной большой длины.

Рассмотрим выражение $\mathbf{W}_0(\mathbf{r}) \cdot \delta\mathbf{R}_0$. Оно представляет собой изменение потенциальной энергии, обусловленное смещением нулевого ядра на $\delta\mathbf{R}_0$. Конечно, \mathbf{W}_0 имеет наибольшее значение в нулевой ячейке, где происходит смещение. На больших расстояниях от рассматриваемого атома смещение одного лишь ядра эквивалентно созданию электрического диполя $+Ze\delta\mathbf{R}_0$. Этот диполь в значительной степени

нейтрализован в результате компенсирующего смещения электронов атомного остатка, число которых равно $Z - 4$, и валентных электронов, число которых равно 4. Электроны атомного остатка будут при этом смещаться почти на точно такое же расстояние, что и ядро, и электрический момент поэтому уменьшится до $4e\delta R_0$. Если бы между четырьмя ковалентными связями, окружающими ядро, не было взаимодействия, то валентные электроны в точности компенсировали бы момент $4e\delta R_0$. Это следует из того обстоятельства, что центр тяжести валентных электронов, связывающих нулевой атом и соседний с ним атом № 1, расположен на середине расстояния между этими двумя атомами; валентные электроны будут, таким образом, смещаться на расстояние, равное половине смещения нулевого атома, и создавать дипольный момент $2(-e)^{1/2}\delta R_0 = -e\delta R_0$. Следовательно, четыре ковалентные связи должны были бы в точности компенсировать дипольный момент $4e\delta R_0$, обусловленный смещением электронов атомного остатка. Между ковалентными связями существует, однако, взаимодействие, так что если нулевой атом (например, атом, расположенный в центре фиг. 4) смещается непосредственно в сторону одного из своих соседей (например, к атому, ближайшему к читателю на фиг. 4), то углы между ковалентными связями, которые он образует с тремя другими атомами, при этом будут уменьшаться; такое взаимное сближение связей вследствие наличия обменного эффекта в свою очередь будет вызывать смещение распределения электронной плотности, отвечающего трем рассматриваемым связям, на расстояние, меньшее $1/2\delta R_0$. Следовательно, движение ядер будет обуславливать появление очень небольшого результирующего дипольного момента, который мы обозначим через $e'\delta R_0$, где $e' \ll e$.

Предположим поэтому, что на больших расстояниях от точки R_0 $\delta R_0 \cdot W_0(\mathbf{r})$ есть поле диполя $e'\delta R_0$, создаваемое последним в среде, диэлектрическая постоянная которой равна χ . В таком случае при распространении волны $\delta R_i = Gq_c \cos \mathbf{k} \cdot \mathbf{R}_i$ в кристалле возникнет поляризационная волна. Показать, что при такой поляризации создается электростатический потенциал

$$\varphi(\mathbf{r}) = \frac{4\pi e' N_e}{\chi V k^2} (\mathbf{k} \cdot \mathbf{G}) q_c \sin \mathbf{k} \cdot \mathbf{r} = - \frac{4\pi e' N_e}{\chi V k^2} \Delta(\mathbf{r}),$$

где N_e/V — число ядер, приходящееся на единицу объема, а $\Delta(\mathbf{r})$ — объемное расширение, вызываемое прохождением волны. Показать, что если значение \mathbf{k} , характеризующее волну, отвечает электрону с энергией $\hbar T$ при комнатной температуре в германии, то

$$\varphi = 5,1 \cdot 10^{-2} \left(\frac{e'}{e}\right) \Delta \text{ абс. единиц} = 15 \left(\frac{e'}{e}\right) \text{ в.}$$

Так как e' составляет, вероятно, только несколько процентов от e , то этот член ничтожно мал по сравнению с другими рассмотренными

нами членами, определяющими изменение структуры энергетических зон. Однако k^2 изменяется прямо пропорционально абсолютной температуре T , так что в области температуры жидкого водорода или при более низких температурах этот член может оказаться существенным.

2. Показать, что в формуле (17.122б) члены, содержащие $du_p(\mathbf{r}, \epsilon)/d\epsilon$, — порядка $\eta P^2/m$.

Прежде всего отметим, что члены, содержащие производную $d/d\epsilon$, умножены на $\nabla\epsilon$ и что $\nabla\epsilon \sim k$, а также, что $\hbar k$ и P — одного и того же порядка величины. Отметим далее, что единственные члены, которые не сводятся непосредственно к членам порядка ηP^2 , содержат множители вида

$$\frac{\partial^2}{\partial x \partial \epsilon} u_p(x, y, z, \epsilon).$$

Эти члены приводят к интегралам, сходным с интегралами, определяющими \bar{p} в выражении $\mathbf{v} = \nabla_P \mathcal{E}(P)$; следовательно, их численное значение не превышает $\eta P^2/m_n$.

3. Доказать, что

$$\int \psi_p^* \mathcal{H}_r \delta\psi_p dV_r = \int (\mathcal{H}_r \psi_p)^* \delta\psi_p dV_r.$$

Для этого при интегрировании по частям выражения

$$\int \psi_p^* \frac{\partial^2}{\partial x^2} \delta\psi_p dV_r$$

по объему кристалла следует учесть, что $\delta\psi_p$ обладает свойством периодичности.

4. При анализе уравнения (17.26) влияние состояний, в которые возможны переходы из состояния b_j , было учтено путем введения в уравнение члена $-i\hbar\eta_j b_j$. Это привело к экспоненциальному выражению вида $\exp(-\eta_j t)$ для коэффициента b_j . Хотя этот метод, как это следует из (17.11) и (17.17), точен, когда речь идет о переходах из данного исходного состояния в некоторую группу конечных состояний, он требует, однако, тщательной проверки при исследовании вопроса о переходах из некоторой группы исходных состояний в ту же группу конечных состояний. В дальнейшем мы будем относить коэффициенты b к тем состояниям, в которые возможны переходы из состояния a_0 , а коэффициенты c — к тем состояниям, в которые возможны переходы из состояний b . Очевидно, что существует три класса c -состояний: 1) состояния, отличающиеся от состояния a_0 наличием двух лишних фононов (что соответствует двум переходам, сопровождающимся испусканием фонона); 2) состояния, кото-

рые характеризуются наличием лишнего фона одного типа и недостатком одного из фононов другого типа; 3) состояния, для которых число фононов на два меньше, чем в состоянии a_0 .

Рассмотрим состояние типа 1, предположив, что оно отличается от состояния a_0 наличием одного лишнего фона с волновым числом k и одного лишнего фона с волновым числом k' ; соответствующий коэффициент обозначим через $c(k, k')$. Такое состояние, как это следует из приводимой ниже схемы, может возникнуть одним из двух способов:

a_0	$b(k)$	$b(k')$	$c(k, k')$
P	$P - \hbar k$	$P - \hbar k'$	$P - \hbar(k + k')$
n_k	$n_k + 1$	n_k	$n_k + 1$
$n_{k'}$	$n_{k'}$	$n_{k'} + 1$	$n_{k'} + 1$

Каждый столбец этой схемы соответствует некоторому определенному состоянию; состояние $c(k, k')$ может возникнуть в результате перехода

$$a_0 \rightarrow b(k) \rightarrow c(k, k')$$

либо в результате перехода

$$a_0 \rightarrow b(k') \rightarrow c(k, k').$$

Это показывает, что на переход из любого состояния $b(k)$ в группу состояний $c(k, k')$ может повлиять наличие другого состояния $b(k')$, из которого также возможны переходы в эту же группу состояний. Если этот эффект выражен достаточно ясно, то это значит, что использованное при выводе уравнения (17.26) предположение о том, что переход из каждого из b -состояний характеризуется своим собственным значением параметра η , может привести к ошибочным результатам.

Предлагается следующая задача. Существенную роль играют те b - и c -состояния, для которых энергия системы имеет то же значение, что и для состояния a_0 ; все остальные коэффициенты b и c относительно малы. Можно сказать, что каждое состояние $b(k)$ возбуждает свою собственную группу состояний $c(k, k')$, в которую из него происходят переходы. Показать, что любое c -состояние, для которого коэффициент $c(k, k')$ имеет значительную величину, возникает в основном только из одного из состояний $b(k)$ и что роль другого возможного состояния $b(k')$ при этом мала. Показать, следовательно, что c -состояния, в которые происходят переходы из любого из состояний $b(k)$, возникают, главным образом, именно из этого состояния $b(k)$, так что влияние состояния $b(k')$ ничтожно мало. Показать также, что аналогичные условия имеют место и для c -состояний типа 2 и 3,

Эти рассуждения свидетельствуют о том, что переход из любого из b -состояний может быть охарактеризован путем введения в (17.26) члена вида $i\hbar\eta b$ и, следовательно, что аналогия между классической и квантовой теориями столкновений может быть продолжена сколь угодно далеко.

Б. Рассмотрим возможные переходы, иллюстрируемые фиг. 144, для того случая, когда $|\mathbf{P}|$ сравнимо с $m\epsilon$. Показать, что если \mathbf{P}_0 первоначально параллельно оси P_y и столкновения происходят в плоскости $P_x P_y$, то форма кривых, приведенных на фиг. 144, определяется решениями уравнения

$$\pm c |\mathbf{P}' - \mathbf{P}_0| = \frac{1}{2m} (\mathbf{P}'^2 - \mathbf{P}_0^2),$$

где $\delta n = \mp 1$, и что эти кривые представляют собой геометрическое место точек пересечения параболы вращения $\mathcal{E} = (P_x^2 - P_y^2)/2m$ с круговым конусом, вершина которого лежит на параболе при $P_y = |\mathbf{P}_0|$, $P_x = 0$, $\mathcal{E} = \mathbf{P}_0^2/2m$. Показать, что если \mathbf{P}' параллельно оси P_y , то существуют два решения, причем одно из них всегда соответствует поглощению фонона при $\delta n = -1$, а второе — переходу от испускания фонона к поглощению фонона при $|\mathbf{P}_0| = m\epsilon$, что отвечает касанию образующей конуса с поверхностью параболы вращения. Основываясь на этих результатах, убедитесь в том, что кривые, приведенные на фиг. 144, дают качественно правильную картину рассматриваемых переходов.

§ 7. ВЫЧИСЛЕНИЕ МАТРИЧНЫХ ЭЛЕМЕНТОВ ПРИ УЧЕТЕ ПРИНЦИПА ПАУЛИ

В этом параграфе будут исследованы те изменения, которые должны быть внесены в теорию рассеяния при наличии большого числа электронов. Сначала рассмотрим тот простейший случай, когда в кристалле присутствует только одна дырка, и покажем, что он по существу эквивалентен уже исследованному случаю одного избыточного электрона. Вычисления, связанные с определением вероятностей переходов для отдельной дырки, будут затем обобщены на случай частично заполненной зоны (см. стр. 627); при этом будут получены результаты, лежащие в основе теории, изложенной в § 2 гл. 11.

Случай одной дырки. Чтобы исследовать тот случай, когда в кристалле присутствует одна дырка, рассмотрим $2N$ волновых функций, соответствующих зоне валентных связей. Эти функции определим следующим образом:

$$\chi(a; \mathbf{q}_c) = \psi(\mathbf{P}_a; \mathbf{r}_c) \gamma_a(s_c), \quad (17.134)$$

где $\mathbf{q}_c = (\mathbf{r}_c, s_c)$ — совокупность значений радиус-вектора \mathbf{r}_c , определяющего положение электрона c по отношению к центру тяжести кристалла (без учета массы рассматриваемых электронов), и спина s_c . Индекс a здесь дважды проходит все значения, соответствующие Р-решетке: один раз при положительном спине, когда $\gamma_a = \alpha$, второй раз при отрицательном спине, когда $\gamma_a = \beta$, где α и β — спиновые функции, определенные в § 7 гл. 15. В тех случаях, когда нас будет интересовать только форма функции χ , будем писать сокращенно $\chi(a)$. (Индексы a, b, c в дальнейшем относятся к электронам, а индексы i и j — к атомным колебаниям.) Определение функции

$$\psi(\mathbf{P}_a; \mathbf{r}) \equiv \psi_{\mathbf{P}_a}(\mathbf{r}),$$

введенное в § 6 [см. (17.101)], дает условие нормировки функции $\chi(a; \mathbf{q})$. Как и в § 7 гл. 15, под $\int d\mathbf{q}$ будет подразумеваться интегрирование по объему и суммирование по возможным значениям спина.

Волновая функция, описывающая состояние системы электронов, должна быть антисимметричной; поскольку предполагается, что в кристалле присутствует только одна дырка, ее наличие может быть при этом учтено путем исключения из рассмотрения волновой функции, соответствующей отсутствующему электрону (см. § 8 гл. 15). Такую волновую функцию запишем следующим образом:

$$A_a(\text{все } \mathbf{q}_c) = \frac{1}{\sqrt{(2N-1)!}} \sum \{\Delta\} \varepsilon(\Delta) \prod_{c=1}^{2N-1} \chi(e_c; \Lambda \mathbf{q}_c); \quad (17.135)$$

здесь e_c принимает значения, отвечающие всем возможным функциям χ , за исключением функции $\chi(a)$, а $\sum \{\Delta\}$ означает суммирование по перестановкам $(2N-1)$ координат в выражении, определяющем волновую функцию в форме определителя [см. (15.115) и (15.116)].

Волновая функция, описывающая нормальные колебания решетки, может быть записана в той же форме, что и в предыдущем параграфе [см. (17.102) и (17.103), а также (17.85)]:

$$\varphi(\text{все } n) = \prod_{i,j} \varphi(n_{k_i a_j}, n_{-k_i a_j}; q_{k_i a_j c}, q_{k_i a_j s}) = \prod_i \varphi_i. \quad (17.136)$$

Каждый элементарный процесс при этом попрежнему связан с изменением только одного из квантовых чисел n .

Волновая функция системы в целом (без учета движения центра тяжести) имеет, таким образом, вид

$$\Phi(\mathbf{P}_a, \text{все } n) = A_a \prod_i \varphi_i = A_a \varphi(\text{все } n). \quad (17.137)$$

Она является собственной функцией оператора Гамильтона

$$\mathcal{H}_0 = \sum_c \mathcal{H}_{r_c} + \sum_i \mathcal{H}_i, \quad (17.138)$$

где \mathcal{H}_i — операторы Гамильтона для гармонических осцилляторов, а

$$\mathcal{H}_{r_c} = \frac{1}{2m} p_c^2 + \mathcal{U}_0(r_c). \quad (17.139)$$

В случае почти заполненной зоны каждый из членов \mathcal{U}_0 должен учитывать усредненное влияние всех остальных электронов данной зоны и содержать потенциальную энергию, обусловленную наличием ядер, электронов атомных остатков, а также валентных электронов, находящихся в других зонах. При определении оператора \mathcal{H}_0 нами опущен член, связанный с отклонениями энергии $\mathcal{U}_0(r_c)$ от ее среднего значения, вызванными флюктуациями положений остальных электронов. Этот член обозначим через \mathcal{U}_2 . Точное значение потенциальной энергии всех электронов может быть записано в виде

$$\begin{aligned} \mathcal{U}(\text{все } r_c, \text{ все } q_i) &= \sum_c \mathcal{U}_0(r_c) + \sum_c \mathcal{U}_1(r_c, \text{ все } q_i) + \mathcal{U}_2(\text{все } r_c) = \\ &= \sum_c \mathcal{U}_0(r_c) + \mathcal{U}_{1T}(\text{все } r_c, \text{ все } q_i), \end{aligned} \quad (17.140)$$

где \mathcal{U}_{1T} — полная энергия возмущения.

Сумма членов $\mathcal{U}_0(r_c) + \mathcal{U}_1(r_c, \text{ все } q_i)$ представляет собой потенциальную энергию электрона c , обусловленную его взаимодействием с ядрами, а также со всеми остальными электронами, без учета флюктуаций положений последних. Члены \mathcal{U}_0 и \mathcal{U}_1 имеют здесь, таким образом, тот же смысл, что и в § 4, с той, однако, разницей, что мы будем теперь предполагать, что некоторые из электронов, поведение которых в среднем описывается этими членами, также могут совершать переходы. Член \mathcal{U}_2 характеризует влияние флюктуаций мгновенных положений электронов, образующих зону валентных связей. При рассмотрении изменения энергии \mathcal{U}_1 , вызванного смещением ядер, учитываются также изменения средней плотности заряда этих электронов, поскольку их волновые функции изменяются вследствие колебаний решетки (см. задачу 1 в § 6). Члены \mathcal{U}_1 имеют весьма существенное значение; что же касается возмущения \mathcal{U}_2 , то мы покажем, что в том случае, когда в кристалле присутствует только одна дырка, оно не может служить причиной переходов.

(Другой член, опущенный нами при определении \mathcal{H}_0 , связан с учетом инерционного взаимодействия между электронами при вычислении приведенной массы системы. Этот член ничтожно мал, см. задачу 1.)

Вычислим теперь матричный элемент энергии возмущения для состояний

$$\Phi = A_a \prod_i \varphi_i \quad \text{и} \quad \Phi' = A_b \prod_i \varphi'_i. \quad (17.141)$$

Энергия возмущения имеет вид

$$\mathcal{U}_{1T} = \sum_{\sigma} \mathcal{U}_1(\mathbf{r}_{\sigma}, \text{ все } q_i) + \mathcal{U}_2(\text{ все } \mathbf{r}_{\sigma}), \quad (17.142)$$

и интересующий нас матричный элемент равен

$$(\mathbf{P}_a \text{ все } n | \mathcal{U}_{1T} | \mathbf{P}_b \text{ все } n') = \int \Phi^* \mathcal{U}_{1T} \Phi' d(\text{ все } \mathbf{q} \text{ и } q), \quad (17.143a)$$

$$(\mathbf{P}_a \text{ все } n | \mathcal{U}_{1T} | \mathbf{P}_b \text{ все } n') = \\ = \int A_a^* \varphi^* (\text{ все } n) \mathcal{U}_{1T} A_b \varphi (\text{ все } n') d(\text{ все } \mathbf{q} \text{ и } q), \quad (17.143б)$$

$$(\mathbf{P}_a \text{ все } n | \mathcal{U}_{1T} | \mathbf{P}_b \text{ все } n') = \\ = \frac{1}{(2N-1)!} \int \left[\sum_{\Delta} \{\Delta\} \varepsilon(\Delta) \prod_c \chi(e_c^*; \Delta \mathbf{q}_c) \right] \prod_i \varphi_i^* \times \\ \times \left[\sum_{\mathfrak{S}} \{\mathfrak{S}\} \varepsilon(\mathfrak{S}) \prod_d \chi(f_d; \mathfrak{S} \mathbf{q}_d) \right] \prod_j \varphi'_j \times \\ \times \left[\sum_{\sigma} \mathcal{U}_1(\mathbf{r}_{\sigma}, \text{ все } q_i) + \mathcal{U}_2(\text{ все } \mathbf{r}_{\sigma}) \right] d(\text{ все } \mathbf{q} \text{ и } q); \quad (17.143в)$$

здесь в первой квадратной скобке из рассмотрения исключено значение e_c , равное a , а во второй скобке — значение f_d , равное b . Почти все интегралы, входящие в это громоздкое выражение, равны либо нулю, либо единице.

Прежде всего рассмотрим член, содержащий энергию \mathcal{U}_2 ; при изменении любого из квантовых чисел n этот член, очевидно, обращается в нуль. Он равен, кроме того, нулю в тех случаях, когда при суммировании по всем состояниям, соответствующим функциям A_a и A_b , значения сумм $\sum \mathbf{P}$ оказываются различными.

Действительно, если значения $\sum \mathbf{P}$ различны, то при изменении координат, вызванном смещением каждого из электронов на расстояние \mathbf{R}_i , волновая функция A умножается на дополнительный множитель $\exp[i(\sum \mathbf{P}) \cdot \mathbf{R}_i/\hbar]$; значение \mathcal{U}_2 при этом, однако, не изменяется. В силу периодических граничных условий такое смещение не должно вызывать изменения численного значения рассматриваемого интеграла; следовательно, если смещение на расстояние \mathbf{R}_i приводит к изменению значения $A_a^* A_b$, интеграл должен равняться нулю. Для волновых функций, соответствующих случаю одной дырки, сумма $\sum \mathbf{P}$ равна $-\mathbf{P}_a$ для

функции A_a и равна $-P_b$ для функции A_b , так что член, содержащий \mathcal{U}_2 , отличен от нуля только при условии $P_a = P_b$. Если, однако, $P_a = P_b$, то переходы невозможны, так как квантовые числа n не могут при этом изменяться. Таким образом, наличие члена \mathcal{U}_2 влияет только на среднюю энергию данного состояния, но этот эффект, по видимому, уже учитывается членами $\mathcal{U}_0(r_c)$. В том случае, когда в кристалле присутствует только одна дырка, член \mathcal{U}_2 можно поэтому вообще не учитывать.

Покажем теперь, что все члены, содержащие $\mathcal{U}_1(r_c, \text{ все } q_i)$, имеют одно и то же значение. Это очевидно прежде всего из соображений симметрии, так как при построении волновых функций A_a и A_b все электроны предполагаются совершенно одинаковыми. Этот результат может быть также доказан более строго путем исследования любого из членов выражения (17.143б), например члена, содержащего $\mathcal{U}_1(r_f, \text{ все } q_i)$. Покажем, что этот член имеет точно такое же значение, как и член, содержащий $\mathcal{U}_1(r_g, \text{ все } q_i)$. Заменим у переменных $q_f = (r_f, s_f)$ и $q_g = (r_g, s_g)$ индекс f на индекс g , и наоборот. Это не должно повлиять на численное значение интеграла, так как во всех случаях интегрирование по r_f и r_g осуществляется в пределах объема кристалла $V = A_x A_y A_z$, а спины принимают значения $\pm 1/2$.

Интеграл, который мы получим из интеграла от $\mathcal{U}_1(r_f, \text{ все } q_i)$ в результате замены f на g и g на f , отличается, однако, от интеграла от $\mathcal{U}_1(r_g, \text{ все } q_i)$ тем, что в выражениях, определяющих функции A_a^* и A_b , переменные q_f и q_g обменялись местами. Поскольку в результате этих перестановок интеграл оказывается умноженным на -1 , интеграл от $\mathcal{U}_1(r_f, \text{ все } q_i)$ будет иметь точно такое же значение, что и интеграл от $\mathcal{U}_1(r_g, \text{ все } q_i)$, если в выражениях, определяющих A_a^* и A_b , мы заменим q_f на q_g и q_g на q_f . Аналогичным образом можно показать, что все остальные члены, содержащие \mathcal{U}_1 , имеют одно и то же значение.

В дальнейшем можно поэтому ограничиться рассмотрением какого-либо одного из членов ряда $\mathcal{U}_1(r_c, \text{ все } q_i)$, например члена $\mathcal{U}_1(r_1, \text{ все } q_i)$, и умножить затем полученный результат на $(2N-1)$. Первая из скобок выражения (17.143в) содержит под знаком суммы $(2N-1)!$ членов, столько же членов содержит и вторая скобка, так что $\mathcal{U}_1(r_1, \text{ все } q_i)$ в общей сложности входит в $[(2N-1)!]^2$ членов. Поскольку функции χ ортонормированы, для каждого из этих членов интеграл по $q_2 \dots q_{2N-1}$ будет обращаться в нуль, за исключением того случая, когда функции χ^* и χ (число таких функций равно $2N-2$) в (17.143в) одинаковы и расположены в одной и той же последовательности. Когда эти функции расположены в одинаковой последовательности, интеграл по q равен единице. Ряд функций χ , общий для функций A_a и A_b , может быть получен только путем выделения функций $\chi(a)$ и $\chi(b)$ из общего ряда $2N$ функций χ .

Следовательно, (17.143в) отлично от нуля только тогда, когда первая из скобок этого выражения содержит функцию $\chi^*(b; \mathbf{q}_1)$, а вторая — функцию $\chi(a; \mathbf{q}_1)$. Оно отлично от нуля, кроме того, только в том случае, если перестановки Δ и \mathcal{S} при этом таковы, что $(2N-2)$ остальных значений q правильно сочетаются попарно в группах функций χ , являющихся общими для A_a и A_b . Имеется $(2N-2)!$ способов выбрать одновременно Δ и \mathcal{S} так, чтобы они удовлетворяли этому требованию. В каждом случае $\varepsilon(\Delta)\varepsilon(\mathcal{S})$ имеет при этом одно и то же значение, равное $+1$ либо -1 , в зависимости от числа перестановок, необходимых для того, чтобы ввести переменную \mathbf{q}_1 в выражение, определяющее функцию $\chi^*(b)$, числа перестановок, необходимых для введения \mathbf{q}_1 в выражение для $\chi(a)$, и, наконец, числа перестановок, нужных для того, чтобы расположить остальные значения \mathbf{q} , определяющие функции χ в формуле (17.143в), так, чтобы они сочетались попарно с соответствующими значениями \mathbf{q} , определяющими функции χ^* . Поскольку при вычислении вероятности перехода знак матричного элемента не играет никакой роли, нет необходимости устанавливать здесь правила его определения.

Запишем теперь матричный элемент (17.143), опуская член, содержащий \mathcal{U}_2 , вычисляя только член, содержащий $\mathcal{U}_1(\mathbf{q}_1, \text{все } q_i)$, и умножая получающийся при этом результат на $(2N-1)!$ для того, чтобы учесть остальные слагаемые \mathcal{U}_1 , а также на $(2N-2)!$ с тем, чтобы учесть различные способы попарного выбора значений \mathbf{q} . В результате мы получим множитель $(2N-1)(2N-2)! = (2N-1)!$, который в точности сократится с коэффициентом выражения (17.143в). Матричный элемент приобретает, таким образом, следующий вид:

$$\begin{aligned} & (\mathbf{P}_a, \text{ все } n | \mathcal{U}_{1T} | \mathbf{P}_b, \text{ все } n') = \\ & = \pm \int \chi^*(b; \mathbf{r}_1, s_1) \prod_i \varphi_i^* \mathcal{U}_1(\mathbf{r}_1, \text{ все } q_i) \chi(a; \mathbf{r}_1, s_1) \times \\ & \quad \times \prod_j \varphi_j' d\mathbf{q}_j d(\text{все } q). \end{aligned} \quad (17.144)$$

Наличие знака \pm , обязанное своим происхождением множителю $\varepsilon(\Delta)\varepsilon(\mathcal{S})$, как уже было сказано выше, на значение вероятности перехода влияния не оказывает. Отметим, что при интегрировании функция A_a сводится к $\chi^*(b)$, а функция A_b — к $\chi(a)$. Эта взаимная перестановка параметров a и b объясняется тем, что вероятность перехода дырки из состояния a в состояние b равна вероятности перехода электрона из состояния b в состояние a . Дальнейшее исследование выражения (17.144) показывает, что оно имеет точно такой же вид, как и выражение, полученное нами для матричного элемента в § 6 настоящей главы; если \mathcal{U}_1 предполагается линейной функцией переменных q_i , то мы получим при этом те же правила отбора, что и выше, и (17.144) сведется к формуле (17.107), согласно которой изменяется только одно из квантовых чисел n .

Правила отбора существуют также и для спина. Поскольку \mathcal{U}_1 от спина не зависит, интеграл (17.144) отличен от нуля только в том случае, когда как $\chi(a)$, так и $\chi(b)$ содержат спиновую функцию α или же спиновую функцию β . При рассматриваемых переходах спин сохраняется, поэтому после суммирования по спиновым координатам спиновые функции в окончательный результат не входят и функции χ сводятся к соответствующим функциям $\psi(P) \equiv \psi_P$.

Очевидно, таким образом, что матричный элемент, определяющий вероятность перехода дырки из состояния P_a в состояние P_b , равен матричному элементу, определяющему вероятность перехода электрона из состояния P_b в состояние P_a . Последний может быть вычислен с помощью методов, описанных в § 6; единственное различие заключается в том, что вместо волновых функций, отвечающих состояниям, близким к дну зоны проводимости, в этом случае следует пользоваться волновыми функциями, отвечающими состояниям, расположенным вблизи верхнего края зоны валентных связей.

Поскольку в случае дырки матричный элемент энергии возмущения имеет точно такую же форму, как и в случае электрона, очевидно, что методы, изложенные в § 2, могут быть применены также и для вычисления вероятностей переходов дырок из одних состояний в другие. В результате для средней длины свободного пробега дырки мы получим выражение вида (17.133) с той, однако, разницей, что вместо скорости изменения величины \mathcal{E}_e при объемном расширении кристалла [см. (17.115) и (17.125)] оно будет содержать соответственно скорость изменения величины \mathcal{E}_v . Формулы, определяющие среднюю длину свободного пробега электрона и дырки, имеют, таким образом, вид:

$$\frac{1}{l_n(T)} = \frac{\mathcal{E}_{1n}^2 k T m_n^2}{\pi \hbar^4 c_{11}}, \quad (17.145)$$

$$\frac{1}{l_p(T)} = \frac{\mathcal{E}_{1p}^2 k T m_p^2}{\pi \hbar^4 c_{11}}. \quad (17.146)$$

Следует упомянуть об одном существенном различии между переходами дырок и переходами электронов. При электронных переходах сохраняется величина

$$P + \hbar \sum n_{ka} k, \quad (17.147)$$

тогда как в случае дырок закон сохранения относится к величине

$$-P + \hbar \sum n_{ka} k. \quad (17.148)$$

Как было отмечено в § 8 гл. 15, дырке естественно, однако, приписать импульс $P_B = -P$, где P — импульс, соответствующий той волновой функции, которую мы исключили из рассмотрения, а P_B — импульс остальных электронов зоны. Такое определение P_B находится в согласии с нашими представлениями о положительном заряде и

положительной массе дырки. Очевидно, что оно согласуется также и с условием сохранения величины

$$P_B + \hbar \sum n_{k\alpha k}. \quad (17.149)$$

Это показывает, что формальная аналогия поведения дырок и электронов станет полной, поскольку речь идет о законах их рассеяния и ускорения, если ввести в рассмотрение величину P_B . Исследование этой формальной аналогии не приводит, однако, к более глубокому пониманию механизма изучаемых явлений, и мы не будем поэтому останавливаться на нем более подробно.

Переходя к рассмотрению свойств системы, содержащей большое число электронов или же большое число дырок, следует указать, что возможность сведения формулы для многоэлектронного матричного элемента (17.143) к одноэлектронной формуле (17.144) не связана с какими-либо специальными предположениями о форме функции \mathcal{U}_1 , помимо предположения о том, что отдельные члены \mathcal{U}_1 зависят от координат отдельных электронов. То обстоятельство, что \mathcal{U}_1 является линейной функцией нормальных координат, было использовано нами только для того, чтобы показать, каким образом соотношение (17.144) или же аналогичное соотношение, полученное в § 6, приводит к формулировке правил отбора для P и квантовых чисел n_i . Следовательно, (17.143) может быть записано в форме (17.144) также и в тех случаях, когда рассеяние обусловлено ионами примесей, искажениями решетки кристалла, например дислокациями, или же любыми другими причинами, вызывающими отклонение от идеального периодического силового поля. Отсюда мы приходим к заключению, что рассеяние дырок искажениями решетки должно происходить точно так же, как и рассеяние электронов, с той разницей, что для описания процесса рассеяния следует пользоваться не волновыми функциями, отвечающими зоне проводимости, но волновыми функциями, отвечающими зоне валентных связей. Это заключение в несколько более простой форме уже было использовано в гл. 11 при исследовании вопроса о рассеянии дырок на примесях.

Промежуточный случай: большое число электронов и дырок. Если зона заполнена частично, так что значительная часть квантовых состояний занята, то принцип Паули, согласно которому переходы электронов в занятые состояния невозможны, приобретает существенное значение. Мы покажем здесь, что влияние принципа Паули на значение матричных элементов может быть исследовано на основе теории, изложенной в гл. 11, где то обстоятельство, что рассматриваемые квантовые состояния частично заняты, было учтено путем уменьшения скорости электронных переходов. С этой целью в выражение для скорости перехода был введен множитель $(1 - f)$, определяющий долю конечных состояний, в которые возможны переходы. В § 2 настоящей главы в весьма общей форме была продемонстрирована эквивалентность

результатов квантовой теории и теории, основанной на классических модельных представлениях, для чего было проведено сопоставление временной зависимости распределения конечных состояний для обоих случаев. В случае системы, содержащей большое число дырок и электронов, подобное сопоставление не представляется столь ценным, как при построении описательной теории, и мы ограничимся поэтому вычислением матричных элементов. Определив последние, мы найдем затем вероятности переходов, подставив значения матричных элементов, относящиеся к исследуемому случаю, в полученную ранее формулу, определяющую вероятность перехода отдельного электрона.

Соответственно речь будет идти об M электронах, которые могут занять часть $2N$ квантовых состояний, отвечающих некоторой энергетической зоне или же нескольким перекрывающимся зонам. Антисимметричные волновые функции, описывающие состояния такой системы электронов, будут иметь вид

$$A = A(a_1, \dots, a_M; \mathbf{q}_1, \dots, \mathbf{q}_M), \quad (17.150)$$

где параметры a относятся к M волновым функциям вида $\chi(a)$ [см. (17.134)]. Порядок расположения этих параметров в выражении (17.150) не играет никакой роли, поскольку изменение этого порядка может лишь привести к изменению знака перед определителем, в форме которого задана функция A [см. (15.116)].

Волновые функции, описывающие нормальные колебания атомов решетки, могут быть выражены через квантовые числа, соответствующие бегущим волнам [см. (17.85), (17.102), (17.103) и (17.136)]:

$$\varphi(\text{все } n) = \prod_{j\ell} \varphi(n_{k_j a_j}, n_{-k_j a_j}; q_{k_j a_j \ell}, q_{-k_j a_j \ell}) = \prod_{\ell} \varphi_{\ell}. \quad (17.151)$$

Волновая функция, характеризующая исходное состояние системы, принимает, таким образом, вид

$$\Phi = A(a_1, \dots, a_M; \mathbf{q}_1, \dots, \mathbf{q}_M) \varphi(\text{все } n) = A\varphi. \quad (17.152)$$

Энергия возмущения имеет то же значение, что и прежде, и может быть записана в виде

$$\mathcal{U}_1 T = \sum_{c=1}^M \mathcal{U}_1(\mathbf{r}_c, \text{все } q_1) + \mathcal{U}_2(\mathbf{r}_1, \dots, \mathbf{r}_M); \quad (17.153)$$

отдельные члены этого выражения сохраняют тот же смысл, что и в формуле (17.142).

Интересующие нас матричные элементы связывают две такие функции Φ , которые отличаются друг от друга одним или несколькими значениями параметра a , а также тем, что одно из квантовых чисел n для одной из этих функций отличается на ± 1 от одного из квантовых чисел n , определяющих вторую функцию. Это правило отбора является, очевидно, непосредственным следствием линейной зависимости

энергии $\mathcal{U}_1(\mathbf{r}_0, \text{ все } q_i)$ от переменных q . В случае одной дырки, как мы видели выше, членом \mathcal{U}_2 можно пренебречь. В теории электропроводности этим членом можно пренебречь также и в том случае, когда зона почти заполнена или же почти пуста, а поверхности постоянной энергии приближенно предполагаются сферическими. Как уже отмечалось выше, при переходах, обусловленных возмущением \mathcal{U}_2 , сохраняется величина $\sum_i P_{a_i}$. Если поверхности постоянной энергии обладают сферической формой, то P_a и v_a связаны друг с другом линейным соотношением, так что

$$\sum v_a = \frac{1}{m_n} \sum P_a \quad \text{или} \quad \sum v_a = -\frac{1}{m_p} \sum P_a. \quad (17.154)$$

Следовательно, взаимодействие электронов с электронами (или дырок с дырками), вызванное наличием возмущения \mathcal{U}_2 , не изменяет характера движения электронов, т. е. не изменяет величины тока и поэтому не оказывает влияния на процесс релаксации последнего. Если, однако, поверхности постоянной энергии не обладают сферической формой или же если зоны вырождены так, что каждому значению P отвечает несколько значений $\mathcal{E}(P)$ и $\mathbf{v}(P)$, то член \mathcal{U}_2 , вообще говоря, мог бы иметь существенное значение. Наблюдаемый на опыте характер зависимости подвижности от температуры указывает, однако, на то, что в случае германия и кремния, а также для обычных металлов эти эффекты не играют существенной роли, поэтому в дальнейшем мы будем ими пренебрегать¹⁾. Если членом \mathcal{U}_2 можно пренебречь, то остаются в силе все те соображения, которые привели нас выше от формулы (17.143) к формуле (17.144). Отсюда следует, что интересующий нас матричный элемент отличен от нуля лишь в том случае, когда изменяется только один из параметров α , так как при изменении двух или большего числа параметров функции A^* и A' , входящие под знак интеграла, определяющего матричный элемент, не будут содержать общих групп $(M-1)$ функций χ . Следовательно, матричный элемент отличен от нуля лишь для таких переходов, при которых изменяется только одно из значений a и только одно из квантовых чисел n . В этих случаях он определяется, очевидно, формулой (17.144), которая, в силу существования правил отбора для квантового числа n , эквивалентна формуле (17.107) и содержит волновые функции, описывающие исходное и конечное состояния электронов, а также осцилляторные волновые функции, отвечающие тому из нормальных колебаний, которое участвует в данном переходе.

Искомые матричные элементы эквивалентны, таким образом, матричным элементам, получаемым при рассмотрении одноэлектронных

¹⁾ Вопрос о рассеянии электронов в металлах рассматривается, в частности, в монографии Мотта и Джонса [20]. См., кроме того, статью Бардина [21], в которой также затронут вопрос о роли взаимодействия электронов с электронами. — *Прим. авт.*

переходов из исходных занятых состояний $a_1, \dots, a_r, \dots, a_M$ в новую группу состояний $a'_1, \dots, a'_r, \dots, a'_M$, отличающуюся от исходной только значением параметра a_r , причем переход происходит в неза-
 ятое состояние с тем же значением спина.

Как это следует из формулы (17.23), вероятность перехода из данного исходного состояния в некоторую группу конечных состояний пропорциональна плотности конечных состояний и квадрату абсолютного значения матричного элемента. Применяя эту формулу к случаю системы, содержащей большое число электронов, мы приходим к заключению, что для любого из электронных переходов ($a_r \rightarrow a'_r$) плотность конечных состояний уменьшилась, так как часть f состояний, которые могли бы быть выбраны в качестве конечного состояния a'_r , уже занята. Таким образом, вероятность перехода имеет в этом случае точно такое же значение, как если бы речь шла о переходах отдельного электрона из исходного состояния a_r , с той, однако, разницей, что выражение, определяющее эту вероятность, содержит теперь дополнительный множитель $(1 - f)$. С физической точки зрения этот результат эквивалентен соображениям, изложенным в § 2, согласно которым распределение электронов при переходах таково, как если бы каждый из них совершал эти переходы независимо от остальных электронов, при условии, что переходы в занятые состояния запрещены.

Использование классической модели, рассмотренной в § 2 гл. 11, тем самым до некоторой степени оправдано. Отметим еще раз, что мы показали аналитически, что матричный элемент, связывающий два состояния системы в целом, различающиеся друг от друга тем, что одно из электронных состояний (a_r) заменено другим состоянием (a'_r), равен (с точностью до множителя ± 1) матричному элементу для системы, содержащей только один избыточный электрон, который переходит из состояния, описываемого волновой функцией $\chi(a_r)$, в состояние, описываемое волновой функцией $\chi(a'_r)$. (При этом делается весьма искусственное предположение о том, что потенциальная энергия электрона в обоих случаях имеет одно и то же значение.) Мы показали, кроме того, что нельзя построить такие волновые функции, которые отвечали бы тому случаю, когда два электрона находятся в одном и том же конечном состоянии; отсюда следует, что матричные элементы, соответствующие переходам в уже занятые состояния, должны равняться нулю. Было также отмечено, что поскольку формула для вероятности переходов содержит плотность конечных состояний, в нее должен быть введен множитель $(1 - f)$, чтобы учесть уменьшение числа тех конечных состояний, в которые возможны переходы.

Напомним, что формула (17.23), определяющая вероятности переходов, получила подтверждение, поскольку мы показали, что классическая и квантовомеханическая модели рассматриваемых явлений

обладают эквивалентными свойствами. Эта эквивалентность классической и квантовой теорий была доказана лишь для того случая, когда в кристалле присутствует только один избыточный электрон (или дырка). Чтобы установить ее для системы, содержащей большое число электронов, необходимо было бы обобщить метод Вайскопфа — Вигнера на случай такой системы. Как уже было указано выше, повышение строгости изложения при этом не окупило бы, однако, излишнего усложнения теории.

ЗАДАЧА

При определении приведенной массы системы, состоящей из большого числа электронов, взаимодействующих с кристаллом, мы получаем оператор Гамильтона, содержащий члены, недиагональные по отношению к импульсам электронов. Обозначим координаты электронов, отсчитываемые от некоторого заданного начала координат, через r_{e1} , r_{e2} и т. д. а радиус-вектор, определяющий положение кристалла, — через R_c . В таком случае $r_a = r_{ea} - R_c$ есть координата, определяющая положение электрона по отношению к кристаллу; именно такой смысл и имеет величина r_a , упоминаемая в данном параграфе. Показать, что уравнения, которыми в рассматриваемом случае следовало бы заменить уравнения (17.49) — (17.52), имеют вид

$$\begin{aligned} p_T &= p_c + \sum_a p_{ea}, \\ p_a &= p_{ea} - \frac{m_e}{M_T} p_T, \\ r_T &= \frac{M_c}{M_T} R_c + \frac{m_e}{M_T} \sum r_{ea}, \\ r_a &= r_{ea} - R_c \end{aligned}$$

и что оператор Гамильтона приобретает при этом вид

$$\mathcal{H} = \frac{1}{2M_T} p_T^2 + \frac{1}{2m} \sum p_a^2 + \frac{1}{2M_c} \sum_a \sum_{b \neq a} p_a \cdot p_b + \mathcal{U}(\text{все } r_a, \text{ все } q_i),$$

где, как и прежде, $m = m_e M_c / (m_e + M_c)$. Эффективная масса, входящая в знаменатель третьего члена последнего выражения, столь велика, что этот член представляет собой лишь незначительную поправку.

ЛИТЕРАТУРА

1. Weisskopf V., Wigner E., Zs. f. Phys., **63**, 54 (1930); **65**, 18 (1930)
2. Heitler W., The Quantum Theory of Radiation, Oxford, 1936 (см. перевод: Гейтлер В., Квантовая теория излучения, М.—Л., 1936).
3. Schiff L. I., Quantum Mechanics, New York, 1949.
4. Pierce B. O., A Short Table of Integrals, New York, 1929.
5. Seitz F., Phys. Rev., **73**, 549 (1948).

6. Whittaker, Watson, *Modern Analysis*, Cambridge, 1927 (см. перевод: Уиттекер, Ватсон, Курс современного анализа, М.—Л., 1934).
7. Seitz F., *The Modern Theory of Solids*, New York, 1940 (см. перевод: Зейтц Ф., Современная теория твердых тел, М.—Л., 1949).
8. Eyring H., Walter J., Kimball G., *Quantum Chemistry*, New York, 1944 (см. перевод: Эйринг Г., Уолтер Д., Кимбалл Д., Квантовая химия, ИЛ, 1948).
9. Feynman R. P., *Phys. Rev.*, **56**, 340 (1939).
10. Born M., *Proc. Phys. Soc.*, **54**, 362 (1942).
11. Brillouin L., *Wave Propagation in Periodic Structures*, New York, 1946.
12. Wilson A. H., *The Theory of Metals*, Cambridge, 1936 (см. перевод: Вильсон А., Квантовая теория металлов, М.—Л., 1941).
13. Dirac P. A. M., *The Principles of Quantum Mechanics*, Oxford, 1930 (см. перевод: Дирак П., Квантовая механика, М.—Л., 1937).
14. Peierls R., *Ann. d. Phys.*, **12**, 154 (1932).
15. Shockley W., Bardeen J., *Phys. Rev.*, **77**, 407 (1950).
16. Bardeen J., Shockley W., *Phys. Rev.*, **80**, 72 (1950).
17. Bardeen J., *Phys. Rev.*, **79**, 216 (1950).
18. Sommerfeld A., Bethe H., *Handb. d. Phys.*, Vol. XXIV (см. перевод: Зоммерфельд А., Бете Г., Электронная теория металлов, М.—Л., 1938).
19. Bond W., Mason W., McSkimin H., Olsen K., Teal G., *Phys. Rev.*, **78**, 176 (1950).
20. Mott N., Jones H., *The Properties of Metals and Alloys*, Oxford, 1936.
21. Bardeen J., *Phys. Rev.*, **52**, 688 (1937).
- 22*. McSkimin H. J., Bond W. L., Buchler E., Teal G. K., *Phys. Rev.*, **83**, 1080 (1951).

БИБЛИОГРАФИЯ ¹⁾

Часть I

ВВЕДЕНИЕ В ЭЛЕКТРОНИКУ ТРАНЗИСТОРОВ

1948 г.

- Bardeen J., Brattain W. H., Транзистор, полупроводниковый триод, Phys. Rev., 74, 230.
- Brattain W. H., Bardeen J., Природа прямого тока в точечных контактах с германием, Phys. Rev., 74, 231.
- Shockley W., Pearson G. L., Модуляция проводимости тонких слоев полупроводника поверхностными зарядами, Phys. Rev., 74, 232.
- * Wells W., Опытный германиевый кристаллический усилитель, Audio Engineering, 32, 6, 11, July.
- * White S. Y., Опытный германиевый кристаллический усилитель, Audio Engineering, 32, 28, 39, August.
- * White S. Y., Конструкция кристаллических усилителей, Audio Engineering, 32, 26, September.
- * White S. Y., Экспериментальные данные о германиевых кристаллических усилителях, Audio Engineering, 32, 32, October.
- * White S. Y., Характеристики кристаллических усилителей, Audio Engineering, 32, 18, November.
- * White S. Y., Анализ характеристик германиевых триодов, Audio Engineering, 32, 19, 40, December.

1949 г.

- Aisberg E., Транзистрон-транзистор, Toute la Radio, 137, 218.
- Angello S. J., Полупроводниковые выпрямители, Electr. Eng., 68, 865.
- Bardeen J., Brattain W. H., Физические принципы действия транзистора, Phys. Rev., 75, 1208.
- Becker J. A., Фотоэффект в полупроводниках, Electr. Eng., 68, 937.
- Becker J. A., Shive J. N., Транзистор — новый полупроводниковый усилитель, Electr. Eng., 68, 215.
- Becker M., Fan H. Y., Фотоэдс-эффект на $p-n$ -границах, созданных в германии бомбардировкой α -частицами и дейтронами, Phys. Rev., 75, 1631.
- Benzer S., Германиевые выпрямители высокого обратного напряжения, Journ. Appl. Phys., 20, 804.
- * Brown J. B., Транзистор с магнитным смещением, Phys. Rev., 76, 1736.
- Dunlap W. C., Германий — важный новый полупроводник, G. E. Rev., 52, 15.
- * Eberhard E., Endres R., Moore R., Счетные схемы с транзисторами, R. C. A. Rev., 10, 459.
- * Fan H., Gray R., Барьерный слой в германии p -типа, Phys. Rev., 76, 458.
- Johnson W. E., Lark-Horovitz K., Полупроводники, облученные нейтронами, Phys. Rev., 76, 442.

¹⁾ Звездочками отмечена литература, добавленная редактором перевода.

- * Красилов А. В., Параметры кристаллических триодов и новые схемы применения их, Вестник информации, № 17 (61), 1.
- * Lark-Horowitz K., Электропроводность полупроводников, *Electr. Eng.*, **68**, 1047.
- Lehans F. W., Генератор на транзисторе для телеметрии, *Electronics*, **22**, 90.
- Lehovec K., Испытание транзисторов, *Electronics*, **22**, 88.
- * Lehovec K., О дырочном токе в германиевом транзисторе, *Phys. Rev.*, **75**, 1100.
- Reich H. J., Ungvary R. L., Триггерная схема с транзистором, *Rev. Sci. Instr.*, **20**, 586.
- * Rowland W., Haegeler, Кристаллический тетрод-смеситель, *Electronics*, **22**, 80.
- Shive J. N., Новое фотосопротивление из германия, *Phys. Rev.*, **76**, 375.
- Webster W. M., Eberhard E., Barton L. E., Несколько новых схем для трехконтактных полупроводниковых усилителей, *R. C. A. Rev.*, **10**, 5.

1950 г.

- Aigrain P., Dugas C., Характеристики транзисторов, *Compt. Rend.*, **230**, 377.
- * Те же, Испытание кристаллических триодов, *Electronics*, **20**, 118.
- * Bardeen J., Теория зависимости характеристик точечных контактов с германием от концентрации дырок, *Bell Syst. Techn. Journ.*, **29**, 469.
- Becker J. A., Транзисторы, *Electr. Eng.*, **69**, 58.
- Busch G., Электрическая проводимость неметаллов, *Zs. f. Ang. Math. und Phys.*, **1**, 3, 81 [см. перевод: Успехи физич. наук, **47**, 258 (1952)].
- * Brown C. B., Работа транзисторов в области высоких частот, *Electronics*, **20**, 81.
- Fricke H., Полупроводниковые триоды и тетроды как усилители и смесители, *E. T. Z.*, **71**, 133.
- Gebble H. A., Vanbury P. C., Hogarth C. A., Кристаллические диоды и триоды из сернистого свинца, *Proc. Phys. Soc.*, **B63**, 371.
- * Hall R., Dunlap W. C., Переходы типа $p-n$, образованные при диффузии примеси, *Phys. Rev.*, **80**, 467.
- Heina H., Кристаллический германиевый триод, *Sylvania Techn.*, **3**, 13.
- Hunter L., Физическая интерпретация характеристик транзистора типа А, *Phys. Rev.*, **77**, 558.
- * Hunter L. P., Brown R. E., Прибор для производственного испытания транзисторов, *Electronics*, **23**, 96.
- Meacham L. A., Michaels S. E., Наблюдения быстрого удаления запасенных дырок из германиевых транзисторов и варисторов, *Phys. Rev.*, **78**, 175.
- * Pfann W. G., Транзистор как обратимый усилитель, *Proc. Inst. Rad. Eng.*, **38**, 1222.
- * Pfann W. G., Scaff J. H., Транзистор из германия p -типа, *Proc. Inst. Rad. Eng.*, **38**, 1151.
- * Shive I. N., Фототранзистор, *Bell Lab. Rec.*, **28**, 337.
- * Shockley W., Теория высоких значений альфа (коэффициента усиления по току) для коллекторных контактов с германием, *Phys. Rev.*, **78**, 294.
- * Slade B. N., Высокоэффективный транзистор с большим расстоянием между контактами, *R. C. A. Rev.*, **11**, 517.
- Turner R. P., Кристаллический приемник с транзисторным усилителем, *Radio and Telev. News*, **43**, 38, 153.

1951 г.

- * Беляев А. Ф., Логинов В. Н., Кристаллические детекторы и усилители, М.—Л.

- * Bowers W. B., Схема умножения частоты с транзистором, *Electronics*, **24**, 140, March.
- * D. G. F. and R. K. J., Транзистор без точечных контактов, *Electronics*, **24**, 82.
- * Goncher F. S., Измерение диффузии дырок в германии *n*-типа, *Phys. Rev.*, **81**, 475.
- * Goncher F. S., Pearson G. L., Sparks M., Опыты с германием, обладающим внутренней *p-n*-границей и теория явлений, *Phys. Rev.*, **81**, 637.
- * Haegele R. W., Визуальный метод испытания транзисторов и применение его к контролю формовки коллектора, *Sylvania Techn.*, **4**, 61.
- * Hall H. H., Vardeen J., Pearson G. L., Влияние давления и температуры на сопротивление *p-n*-переходов в германии, *Phys. Rev.*, **84**, 129, October.
- * Haynes J. R., Shockley W., Подвижность и время жизни введенных дырок и электронов в германии, *Phys. Rev.*, **81**, 835.
- * Hogarth C. A., Кристаллический диод и триод из селенистого свинца, *Proc. Phys. Soc.*, **64B**, 822.
- * Калашников С. Г., Эрастов Л. Н., Образование запиорных слоев в некоторых сплавах, *ЖТФ*, **21**, 129.
- * Kanai Y., Усиление тока на *p-n*-переходе в кремнии, *Journ. Phys. Soc. (Japan)*, **6**, 211.
- * McKay K. G., Переход типа *n-p-n* как модель для явления вторичной фотопроводимости, *Phys. Rev.*, **84**, 833.
- * Kogos L. L., Schwartz R. F., Схема частотной модуляции с транзистором, *Electronics*, **24**, 140, March.
- * Longini R. L., Электрическая формовка транзисторов из германия *n*-типа с применением контактов из сплавов, *Phys. Rev.*, **84**, 1254.
- * Matare H. F., Явления в граничном слое и статистические колебания тока в трехэлектродном кристалле, *Zs. f. Phys.*, **131**, 82.
- * Pfann W. G., Влияние состава точечного электрода на свойства выпрямляющих переходов в германии, *Phys. Rev.*, **81**, 882, March.
- * Pietsenpol W. J., Выпрямитель и фотоэлемент с *p-n*-переходами, *Phys. Rev.*, **82**, 120.
- * Prim R. C., Некоторые дифференциальные уравнения в частных производных, описывающие поток дырок и электронов в полупроводниках, *Bell Syst. Techn. Journ.*, **30**, 1174, October.
- * Salow H., Германиевые триоды (обзор), *Zs. angew. Phys.*, **3**, 231.
- * Schultheiss P. M., Reich H., Некоторые триггерные схемы с транзистором, *Proc. Inst. Rad. Eng.*, **39**, 627, June.
- * Scott T. R., Кристаллические триоды, *Electr. Commun.*, **28**, 195.
- * Skalnick J. G., Reich H. J., Gibson J. E., Flynn T., Изменение характеристик транзисторов под действием вспомогательного тока, *Electronics*, **24**, 142, 228, 232, September.
- * Stade B. W., Метод улучшения электрической и механической стабильности транзисторов с точечными контактами, *R. C. A. Rev.*, **12**, 651.
- * Stahl F. A., Влияние металла зонда на положение *p-n*-границы в германии, *Sylvania Techn.*, **4**, 61, July.
- * Stelmak J. P., Электрическая формовка транзисторов *n*-типа с контактами из фосфорных сплавов, *Phys. Rev.*, **83**, 165.
- * Teal G. K., Sparks M., Buchler E., Выращивание монокристаллов германия с *p-n*-переходами, *Phys. Rev.*, **81**, 637, February.
- * Ugilette G. A., Применение германиевых транзисторов, *Antenna*, **23**, 161.
- * Wallace R. T., Jr., Pietsenpol W. J., Некоторые особенности схем и применения *n-p-n*-транзисторов, *Proc. Inst. Rad. Eng.*, **39**, 753; *Bell Syst. Techn. Journ.*, **30**, 530.

1952 г.

- * Aigrain P., Сообщение об исследовании явлений выпрямления и усиления в германии, *Ann. d. Phys.*, **7**, 140.
- * Vanbury P. C., Транзистор из сернистого свинца с двойной поверхностью, *Proc. Phys. Soc.*, **65B**, 236.
- * Billig E., Физика транзисторов (обзор), *Brit. Journ. of Appl. Phys.*, **3**, 241.
- * Chaplin G. B. V., Снятие характеристик транзистора при помощи катодного осциллографа, *Journ. Sci. Instr.*, **29**, 142.
- * Epstein G. S., Bush J. A., Shellhorn B., Применение транзисторов в технике связи, *Electronics*, **25**, 98.
- * Maisch J., Транзисторы. Физика и устройство транзисторов. *AEU*, **6**, 73.
- * Moline J., Магнитный и кристаллический усилители, *Radio* (франц.), No 4, 1.
- * Pearson G. L., Read W. T., Shockley W., Зондирование слоя объемного заряда в $p-n$ -переходе, *Phys. Rev.*, **85**, 1055.
- * Roessler E., Основные сведения о транзисторах (обзор), *E. T. Z.*, **73**, 338.
- * Schultz O., Германиевый триод и его применение (обзор), *Fernm. Praxis*, **29**, 33.
- * Sittner W. R., Усиление тока в транзисторе типа А, *Proc. Ind. Rad. Eng.*, **40**, 448.
- * Топпезен Тор Н., Наблюдение транзисторного действия на новых материалах, *Proc. Phys. Soc.*, **65B**, 737.
- * Valdes L. V., Электрическая формовка коллекторного электрода транзистора из германия n -типа, *Proc. Ind. Rad. Eng.*, **40**, 445.
- * Вавилов В. С., Полупроводниковые триоды без точечных контактов, *Успехи физич. наук*, **46**, 96.

Часть II

ОПИСАТЕЛЬНАЯ ТЕОРИЯ ПОЛУПРОВОДНИКОВ

Книги по зонной теории твердых тел

1931 г.

Hume-Rothery W., *Metallic State, Electrical Properties and Theories*, Oxford.

1933 г.

Sommerfeld A., Bethc H., *Elektronentheorie der Metalle*, *Handb. d. Phys.*, **24**, pt. 2, Berlin (см. перевод: Бете Г., Зоммерфельд А., *Электронная теория металлов*, М.—Л., 1938).

1936 г.

- Ewald P. P., Röschl Th., Prandtl L., *The Physics of Solids and Fluids*, London, 2ed. (см. перевод: Эвальд П., Пошль Ф., Прандтль Л., *Физика твердых и жидких тел*, М.—Л., 1939).
- Fröhlich H., *Elektronentheorie der Metalle*, Berlin.
- Mott N. F., Jones H., *The Theory of the Properties of Metals and Alloys*, Oxford.
- Wilson A. N., *Theory of Metals*, Cambridge University Press (см. перевод: Вильсон А. Г., *Теория металлов*, М.—Л., 1941).

1939 г.

- Slater J. C., Introduction to Chemical Physics, New York.
Wilson A. H., Semiconductors and Metals; an Introduction to the Electron Theory of Metals, Cambridge University Press.

1940 г.

- Mott N. F., Gurney R. F., Electronic Processes in Ionic Crystals, Oxford (см. перевод: Мотт Н., Гурни Р., Электронные процессы в ионных кристаллах, ИЛ, 1950).
Seitz F., The Modern Theory of Solids, New York (см. перевод: Зейтц Ф., Современная теория твердого тела, М.—Л., 1949).

1943 г.

- Seitz F., The Physics of Metals, New York (см. перевод: Зейтц Ф., Физика металлов, М.—Л., 1947).

1947 г.

- Hume-Rothery W., Atomic Theory for Students of Metallurgy, London.
* Волькенштейн Ф. Ф., Электропроводность полупроводников, М.—Л.

1948 г.

- Hume-Rothery W., Electrons, Atoms, Metals and Alloys, London (см. перевод: Юм-Розери В., Электроны и металлы, М., 1950).

1949 г.

- Rice F. O., Teller E., Structure of Matter, New York.

1950 г.

- * Добрецов Л. Н., Электронная и ионная эмиссия, М.—Л.
* Френкель Я. И., Введение в физику металлов, 2-е изд., М.—Л.

1951 г.

- * Адирович Э. И., Некоторые вопросы теории люминесценции кристаллов, М.—Л., гл. III.
* Пекар С. И., Исследования по электронной теории кристаллов, М.—Л.

Обзорные работы

1933 г.

- * Иоффе А. Ф., Электропроводность полупроводников, М.—Л.

1934 г.

- Slater J. C., Электронная структура металлов, Rev. Mod. Phys., 6, 209.

1935 г.

- * Нордгейм Л., Теория металлического состояния, Успехи физич. наук, 15, 571, 675, 779, 939.

1937 г.

Seitz F., Johnson R. P., Modern Theory of Solids, Journ. Appl. Phys., 8, 84, 186, 246 [см. перевод: Зейтц Ф., Джонсон Р., Теория металлов, Успехи физич. наук, 23, 89, 293 (1940)].

1940 г.

* Давыдов Б. И., Шмушкевич И. М., Теория электронных полупроводников, Успехи физич. наук, 24, 21.

1941 г.

* Бардин И. П., Электропроводность металлов, Успехи физич. наук, 25, 19.

Часть III

НЕКОТОРЫЕ ВОПРОСЫ КВАНТОМЕХАНИЧЕСКОЙ ТЕОРИИ

Книги по квантовой механике

1932 г.

* Фок В. А., Начала квантовой механики, Л.

1934 г.

* Френкель Я. И., Волновая механика, М.—Л., т. I.
Gurphey R. W., Elementary Quantum Mechanics, Cambridge.

1935 г.

* Френкель Я. И., Волновая механика, М.—Л., т. II.
Pauling L., Wilson E. B., Introduction to Quantum Mechanics, New York.

1937 г.

Kemble E. C., Fundamental Principles of Quantum Mechanics, New York.

1938 г.

Dushman S., Elements of Quantum Mechanics, New York.
Rojansky V., Introductory Quantum Mechanics, New York.

1945 г.

Heitler W., Elementary Wave Mechanics, Oxford (см. перевод: Гайтлер В., Элементарная квантовая механика, ИЛ, 1948).

1946 г.

Eyring H., Walter J., Kimball G. E., Quantum Chemistry, New York (см. перевод: Эйринг Г., Уолтер Д., Кимбалл Д., Квантовая химия, ИЛ, 1948).

1947 г.

Dirac P. A. M., *The Principles of Quantum Mechanics*, Oxford, 3rd ed. (см. перевод: Дирак П. А. М., *Основы квантовой механики*, 2-е изд., М.—Л., 1937).

1948 г.

- * Ландау Л. Д., Лифшиц Е. М., *Квантовая механика*, М.—Л.
Mott N. F., Sneddon I. N., *Wave Mechanics and its Applications*, Oxford.

1949 г.

- * Блохинцев Д. И., *Основы квантовой механики*, М.—Л.
Schiff L. I., *Quantum Mechanics*, New York.

1952 г.

- * Шпольский Э. В., *Атомная физика*, ч. I и II, 3-е изд., М.—Л.

Книги по статистической физике

1930 г.

Brillouin L., *Statistiques quantiques et leurs applications*, Paris, 2 v. (см. перевод: Бриллюэн Л., *Квантовая статистика*, Харьков — Киев, 1934).

1936 г.

Fowler R. H., *Statistical Mechanics; the Theory of the Properties of Matter in Equilibrium*, Cambridge.

1938 г.

Tolman R. C., *Principles of Statistical Mechanics*, Oxford.

1939 г.

Fowler R. H., Guggenheim E. A., *Statistical Thermodynamics*, Cambridge (см. перевод: Фаулер Р., Гуггенгейм Э., *Статистическая термодинамика*, ИЛ, 1949).
Slater J. C., *Introduction to Chemical Physics*, New York.

1940 г.

- Mayer J. E., Mayer M. G., *Statistical Mechanics*, New York (см. перевод: Майер и Гепперт-Майер, *Статистическая механика*, ИЛ, 1952).
* Гуревич Л. Э., *Основы физической кинетики*, М.—Л.

1944 г.

- * Леонтович М. А., *Статистическая физика*, М.—Л.

1946 г.

Schroedinger E., *Statistical Thermodynamics*, Cambridge (см. перевод: Шредингер Э., *Статистическая термодинамика*, ИЛ, 1948).

1948 г.

- * Френкель Я. И., Статистическая физика, Изд. АН СССР.

1949 г.

- Gurney R. W., Introduction to Statistical Mechanics, New York.
Rushbrooke G. S., Introduction to Statistical Mechanics, Oxford.

1950 г.

- * Левич В. Г., Введение в статистическую физику, М.—Л.

1951 г.

- * Ландау Л. Д., Лифшиц Е. М., Статистическая физика, М.—Л.
* Леонтович М. А., Введение в термодинамику, М.—Л.

ПРИЛОЖЕНИЕ

В. ШОКЛИ, М. СПАРКС и Г. ТИЛ

ТРАНЗИСТОРЫ С $p-n$ -ПЕРЕХОДАМИ

[Phys. Rev., 83, № 1, 151 (1951)]

§ 1. ВВЕДЕНИЕ

В этой статье рассмотрены явления, происходящие при приложении напряжения к полупроводникам, состоящим из нескольких областей с проводимостью различного типа. Такие полупроводники, состоящие, в частности, из двух областей с проводимостью одного типа, между которыми находится третья область с проводимостью противоположного типа, имеют большое практическое значение в транзисторной электронике. Они могут быть использованы также для более наглядной иллюстрации механизма диффузии электрона и дырок. В частности, с помощью таких полупроводников хорошо можно продемонстрировать явление *внутренних контактных потенциалов*.

Транзисторы, в которых нелинейные эффекты обусловлены взаимодействием областей p - и n -типа внутри германия, называются транзисторами с $p-n$ -переходами, в отличие от точечно-контактных приборов, в которых существенную роль играет контакт металл—полупроводник. Имеется ряд возможных транзисторов с $p-n$ -переходами. Транзистор с $p-n-p$ -переходами уже рассматривался с теоретической точки зрения [1]¹⁾. В этой статье мы будем рассматривать главным образом транзисторы с $n-p-n$ -переходами [2], фототранзистор n -типа с $p-n$ -ловушкой у коллектора и $p-n-p-n$ -транзистор с эмиттером p -типа и с $p-n$ -ловушкой у коллектора.

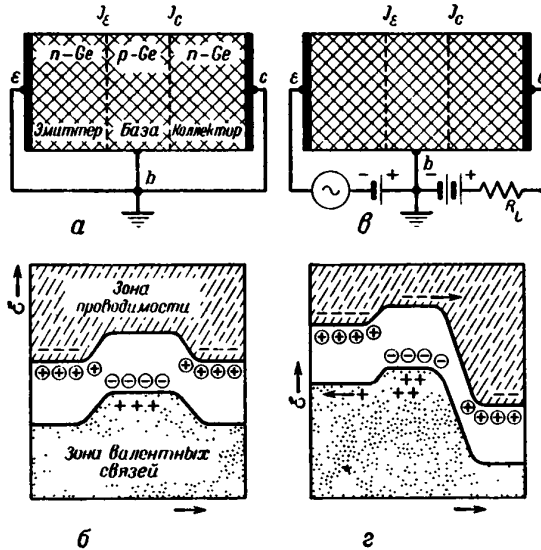
В следующих параграфах сначала будут описаны основные явления, связанные с работой указанных выше приборов. Далее будет дано описание устройства и электрических характеристик транзисторов с $n-p-n$ -переходами. Наконец, будет дана теория работы приборов, на основе которой будут получены вольтамперные характеристики для некоторых типов этих приборов и проведено сравнение теории с экспериментом.

§ 2. РАБОТА ПОЛУПРОВОДНИКА С $n-p-n$ -ПЕРЕХОДАМИ В КАЧЕСТВЕ ТРАНЗИСТОРА И В КАЧЕСТВЕ УМНОЖИТЕЛЯ С ЛОВУШКОЙ

На фиг. 1 показан полупроводник германий с $n-p-n$ -переходами и показано, как он может быть использован в качестве транзистора. Подобно транзисторам типа А между эмиттером и основным

¹⁾ См. также основной текст настоящей книги. — *Прим. ред.*

электродом и между ним и коллектором имеются выпрямляющие переходы. Однако в отличие от транзисторов типа А выпрямление происходит внутри германия, а не на контактах между металлическими электродами и германием, которые являются по существу омическими. Имеются еще и другие важные различия между $n-p-n$ -транзисторами и транзисторами типа А. Движение введенных носителей тока в $n-p-n$ -транзисторах обуславливается в основном диффузией,



Фиг. 1. Структура $n-p-n$ -транзистора и схема энергетических уровней.

а, б — в условиях термического равновесия; в, г — подано постоянное смещение.

а не дрейфом в электрическом поле. Умножение тока в области коллектора, обуславливающее неустойчивость транзисторов типа А при положительной обратной связи, отсутствует у транзисторов с $n-p-n$ -переходами.

В этом параграфе кратко описывается теория действия транзистора. Дополнительные детали читатель найдет в цитируемой литературе. В § 5 и следующих разделах некоторые аспекты теории даны в аналитической форме. Анализ значительно упрощен благодаря тому, что сделаны следующие предположения (которые рассматриваются также в работе [1]¹).

¹) См. также основной текст книги. — Прим. ред.

1. Доноры и акцепторы полностью ионизированы (это предположение хорошо выполняется для германия при комнатной температуре).

2. В каждой области концентрация носителей тока одного знака (большинства носителей) значительно больше, чем другого (меньшинства носителей).

3. В любой области общая скорость рекомбинации линейно связана с отклонением концентрации носителей тока, присутствующих в меньшем числе, от ее значения при тепловом равновесии¹⁾.

4. Объемный заряд всюду в полупроводнике (кроме областей объемного заряда в $p-n$ -переходах) незначителен.

На фиг. 1 показаны энергетические зоны рассматриваемого транзистора для случая нулевых смещений и для смещений, при которых транзистор может работать в качестве усилителя. В последнем случае напряжение смещения прикладывается к барьеру J_c (правая часть фигуры) в обратном направлении. При этом электроны не могут переходить из области коллектора в основной электрод, так как их потенциальная энергия недостаточна для преодоления потенциального барьера. По этим же самым причинам дырки не могут переходить из основного электрода в область коллектора. Наоборот, из области эмиттера электроны могут переходить в область основного электрода, так как высота потенциального барьера между этими областями невелика. Из области основного электрода некоторые из этих электронов, благодаря диффузии, могут переходить через барьер коллектора. Прохождение электронов через барьер эмиттера зависит от высоты этого барьера, которая, в свою очередь, определяется величиной приложенного к барьеру напряжения (изменением величины этого напряжения можно менять высоту барьера).

Если слой основного электрода очень тонок, то число электронов, рекомбинирующих в нем с дырками, будет незначительным и через этот слой будет проходить электронный ток. Если, кроме того, электропроводность в области эмиттера значительно больше, чем в области основного электрода, то число электронов, проходящих через потенциальный барьер из области эмиттера в область основного электрода, будет во много раз больше числа дырок, проходящих в обратном направлении. Таким образом, электрический ток через потенциальный барьер эмиттера будет определяться в основном потоком электронов из области эмиттера в область основного электрода.

При таких условиях работа этого прибора имеет очень близкую аналогию с работой вакуумных электронных ламп. Область эмиттера соответствует при этом катоду, основного электрода — сеточному пространству, а область коллектора — аноду. В хорошо рассчитанных приборах управляемый электронный ток, проходящий через основной электрод, может быть значительно больше, чем проходящий через ту же область ток управления. Транзистор имеет, таким образом, очень

¹⁾ Предположения 2 и 3 позволяют использовать линейные уравнения при вычислении тока, обусловленного введенными носителями.—Прим. авт.

высокий коэффициент усиления по току. Транзистор может, следовательно, работать как триод в схеме с заземленным катодом, если заземлить область эмиттера и сигнал подавать на основной электрод. Из приведенных в § 4 вольтамперных характеристик транзисторов видно, что пропускание тока является почти совершенным.

Интересно отметить, что как в электронных лампах, так и в транзисторах управление тока осуществляется взаимодействием двух электронных потоков. В электронных вакуумных лампах сеточный ток управляет термоэлектронным током в сеточном пространстве. В транзисторах поток дырок в основном электроде изменяет напряжение на барьере эмиттера и управляет потоком электронов через слой основного электрода.

При тех же самых действующих смещениях устройство, показанное на фиг. 1, может быть использовано в качестве коллектора транзистора с большим коэффициентом усиления [3]. Оно может быть использовано также и как фототранзистор [4]. Первое из этих применений будет описано позднее, а сейчас рассмотрим, каким образом такое устройство обеспечивает усиление фототоков. Прежде всего отметим, что электрод, соединенный с основным электродом, в этом случае не нужен. Если германий освещается вблизи барьеров, то возникающие при этом пары электрон — дырка будут разделяться полем барьеров и, следовательно, дырки будут двигаться в область основного электрода. Накапливаясь здесь, они будут заряжать этот слой положительно и, тем самым, понижать высоту потенциального барьера. В результате этого большее число электронов будет преодолевать потенциальный барьер и двигаться к коллектору. Когда свет выключается, влияние добавочных дырок прекращается, так как часть дырок диффундирует из основного электрода в область эмиттера и соединяется там с электронами, тогда как другая их часть рекомбинирует в области основного электрода с электронами, которые перешли сюда благодаря диффузии из области эмиттера. Если, однако, слой основного электрода очень тонок и концентрация электронов в области эмиттера очень велика, то число электронов, проходящих через барьер из этой области в основной электрод, будет значительно больше числа дырок, проникающих в слой эмиттера и рекомбинирующих там. В § 7 будет показано, что получаемое в этом случае усиление тока прямо пропорционально отношению электропроводностей этих двух слоев и обратно пропорционально толщине слоя основного электрода.

Интересным следствием диффузии электронов через слой основного электрода является возникновение *внутренних контактных потенциалов* [5]. Для иллюстрации этого мы предположим, что слой основного электрода заземлен и что потенциал прикладывается к эмиттеру. Если, кроме того, имеется добавочный омический контакт в области основного электрода, то он будет находиться, конечно, при потенциале земли. Однако если контакт является выпрямляющим и, особенно, если проходящий через него ток определяется преимущественно потоком

электронов, то его потенциал скорее будет определяться концентрацией электронов в слое основного электрода, чем потенциалом заземленного контакта. Область коллектора является примером такого контакта. Установлено как теоретически, так и экспериментально, что если слой основного электрода заземляется, то потенциал, приложенный к области эмиттера, передается через него (несмотря на то, что его электростатический потенциал практически не изменяется) и проявляется в области коллектора, которая стремится принять потенциал, приблизительно равный потенциалу области эмиттера, по крайней мере для некоторой определенной области напряжений. Опытные данные и теория этого явления рассматриваются в § 7 и 8.

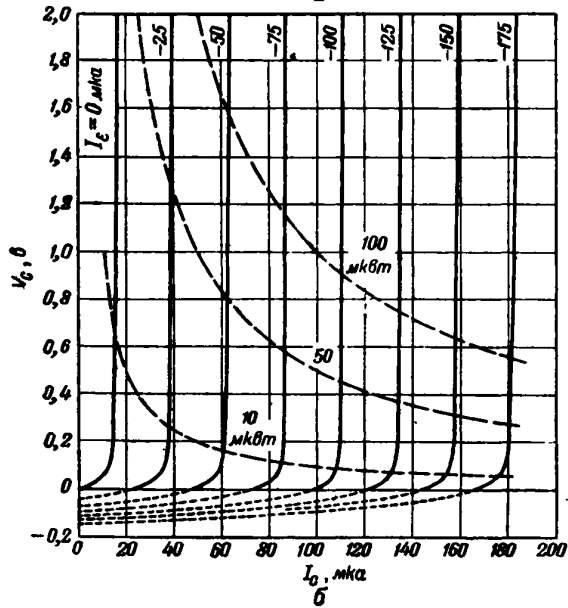
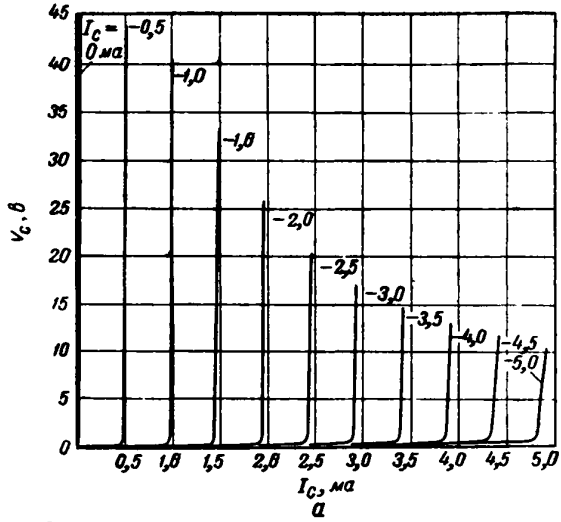
§ 3. ОПИСАНИЕ ОПЫТНЫХ МОДЕЛЕЙ

Опытные модели были изготовлены из монокристалла германия, в котором тонкий слой p -типа находился между двумя областями n -типа. Сделанные модели обладали различной электропроводностью. Обычными значениями электропроводности различных областей $n-p-n$ -структуры, показанной на фиг. 1, были следующие: эмиттер — $100 \text{ ом}^{-1} \cdot \text{см}^{-1}$, основной электрод — $1 \text{ ом}^{-1} \cdot \text{см}^{-1}$ и коллектор — $0,1 \text{ ом}^{-1} \cdot \text{см}^{-1}$. Типичные значения времени жизни меньшинства носителей тока в области коллектора составляли 300—400 мксек. Опыт показывает, что времена жизни носителей тока в других областях, в которых непосредственное их измерение труднее, не очень сильно отличаются от этих значений. Электроды всех трех областей представляют собой обычные механически прочные омические контакты. Работа таких структур обуславливается процессами, происходящими внутри монокристалла, а не на контактах между электродами и германием.

§ 4. РАБОЧИЕ ХАРАКТЕРИСТИКИ

Рабочие характеристики $n-p-n$ -транзисторов будут рассмотрены здесь очень кратко. Работа этих приборов определяется процессами, происходящими в $p-n$ -переходах, величина активной площади которых в противоположность усилителям с точечными контактами может быть легко изменена. Это дает возможность увеличивать выходную мощность прибора, не изменяя плотности тока. Один из наибольших $n-p-n$ -транзисторов, использованный в качестве усилителя, имел площадь перехода, равную $0,3 \text{ см}^2$, при толщине слоя основного электрода около $0,07 \text{ см}$ и обеспечивал 2 вт неискаженной выходной мощности при работе в режиме усиления по классу А. Его предельная частота была около $10\,000 \text{ гц}$, причем это значение, как показано в § 9, находится в согласии с результатами, полученными на основе расчета скорости диффузии носителей тока через слой основного электрода с электропроводностью p -типа.

Возможности малого расхода мощности при работе этих моделей изучены более подробно. Типичная величина площади их $p-n$ -переходов



Фиг. 2. Вольтамперные характеристики $n-p-n$ -транзистора.

составляет примерно $0,01 \text{ см}^2$, а толщина слоя основного электрода — около $1,5 \cdot 10^{-3} \text{ см}$. Получаемое с ними усиление равно 50 дб , а их шумовые числа при 1000 гц составляют около $10-15 \text{ дб}$. Каждая из этих величин по своему значению на несколько порядков лучше соответствующей величины для кристаллических усилителей с точечными контактами. При работе с сигналами малой интенсивности они дают, по существу, уже полное усиление начиная с $0,1 \text{ в}$ напряжения на коллекторе и, следовательно, являются исключительно хорошими усилителями очень малых мощностей. При несколько меньшем усилении их к. п. д. может доходить до $48-49\%$, в то время как его теоретически возможный максимум для усилителей класса A равен 50% . Уоллэс и Томас построили генератор, который работает при входной мощности, равной всего лишь $0,6 \text{ мвт}$. Такие маленькие транзисторы могут работать также в качестве усилителей при максимальной выходной мощности, равной нескольким сотням мвт .

Вольтамперные характеристики $n-p-n$ -транзисторов показаны на фиг. 2. Предельная частота этих усилителей при работе их в схемах с высоким усилением определяется главным образом емкостью $p-n$ -перехода коллектора и лежит значительно ниже границы, определяемой скоростью диффузии через слой основного электрода. Более подробное рассмотрение емкостных эффектов можно найти в статье Уоллэса и Питенполя.

§ 5. ОБЩИЕ ТЕОРЕТИЧЕСКИЕ ПОЛОЖЕНИЯ И ГРАНИЧНЫЕ УСЛОВИЯ

В этом и следующих параграфах основные теоретические представления, изложенные в § 2, даются в аналитической форме. Перечень главных обозначений приводится ниже. В этом и в ближайших параграфах будет рассмотрен $n-p-n$ -полупроводник с общей точки зрения. Поэтому вместо терминов эмиттер и коллектор удобно использовать индексы l и r для обозначения двух областей с электропроводностью n -типа (соответственно левой и правой). Это особенно удобно при описании тех случаев, когда на любую из этих двух областей может быть подано смещение как на коллектор или как на эмиттер.

Сначала установим граничные условия на $p-n$ -переходах при приложенных к ним напряжениях. Нулевое значение электростатического потенциала ψ внутри полупроводника может быть выбрано произвольно. В нашем случае это значение выбрано так, что $-q\psi$ равно приблизительно энергии электрона, находящегося на энергетическом уровне, проходящем через середину запретной зоны. При условиях равновесия точное значение ψ удовлетворяет уравнениям

$$n = n_i e^{q(\psi - \varphi)/kT}, \quad (1)$$

$$p = n_i e^{q(\varphi - \psi)/kT}. \quad (2)$$

При неравновесных условиях подобные уравнения определяют квази-уровни Ферми φ_n и φ_p в зависимости от ψ , n и p

$$n = n_i e^{q(\psi - \varphi_n)/kT}, \quad (3)$$

$$p = n_i e^{q(\varphi_p - \psi)/kT}. \quad (4)$$

ОБОЗНАЧЕНИЯ

- A — площадь поперечного сечения прибора;
- $b = \mu_n/\mu_p = 2,1$;
- $B_i = kT/q [1 - e^{-q\varphi_i/kT}]$;
- $B_r = kT/q [1 - e^{-q\varphi_r/kT}]$;
- D_n и D_p — коэффициенты диффузии электронов и дырок;
- G_{lr} , G_{lr} и пр. — активные проводимости при нулевом напряжении;
- n — концентрация электронов;
- n_b — концентрация электронов в слое основного электрода;
- n_i — концентрация электронов внутри образца;
- n_1 — отклонение концентрации электронов от ее значения при тепловом равновесии;
- N_d и N_a — концентрации доноров и акцепторов;
- p — концентрация дырок;
- q — заряд дырки; заряд электрона;
- V и v — компоненты постоянного и переменного напряжения;
- W — толщина слоя основного электрода;
- μ_n и μ_p — подвижности электронов и дырок, равные соответственно 3600 и 1700 $\text{см}^2/\text{в} \cdot \text{сек}$;
- σ_i — проводимость германия;
- τ_{pl} , τ_{pr} и τ_{nb} — времена жизни меньшинства носителей;
- φ — (уровень Ферми)/ $(-q)$;
- φ_n — квази-уровень Ферми электронов;
- φ_p — квази-уровень Ферми дырок;
- φ_l , φ_b и φ_r — напряжения, приложенные к трем областям;
- ψ — электростатический потенциал.

Квази-уровни Ферми вводятся для удобства при обсуждении граничных условий на $p-n$ -переходах и значения приложенных напряжений. Для зависимости плотности токов от φ_n и φ_p принимается очень простая форма:

$$I_n = qD_n \nabla n + q\mu_n nE = -q\mu_n n \nabla \varphi_n \quad (5)$$

$$I_p = -qD_p \nabla p + q\mu_p pE = -q\mu_p p \nabla \varphi_p. \quad (6)$$

Из этих уравнений с очевидностью следует, что данный электронный ток будет вызывать значительно более сильные изменения φ_n и φ_p при прохождении через область с проводимостью p -типа, где электронная концентрация мала, чем при прохождении его через область с проводимостью n -типа. В действительности, отношение электронных проводимостей этих двух областей так велико, что в области n -типа величину φ_n можно считать, по существу, постоянной. В согласии с предположением, что концентрация носителей тока одного знака мала по сравнению с концентрацией носителей тока другого знака и что объемный заряд незначителен, следует, что потенциал ψ внутри каждой области, по существу, также постоянен. Если электроды областей l и r являются омическими и если они расположены на таком большом расстоянии от $p-n$ -переходов, что вводимые носители тока не достигают их, то значения φ_n для этих областей просто равны приложенным к ним напряжениям. Из предположения, что ток к основному электроду является в основном дырочным, следует, что φ_p для этой области равно φ_b .

Воспользуемся теперь этими заключениями для выяснения граничных условий на $p-n$ -переходе J_l . Предположим для простоты, что основной электрод заземляется, или, другими словами, будем рассматривать только случаи, когда

$$\varphi_l \neq 0, \quad \varphi_b = 0, \quad \varphi_r \neq 0. \quad (7)$$

На основании вышеизложенного можно заключить, что значение квази-уровня Ферми φ_n в области перехода остается постоянным и максимальный градиент этой величины имеет место только в области слоя b . Следовательно, в области основного электрода вблизи J_l , φ_n по существу можно считать равным φ_l . Концентрация электронов в этой области вблизи J_l определяется, таким образом, формулой

$$n \text{ (в слое } b, \text{ вблизи } J_l) = n_b e^{-q\varphi_l/kT}, \quad (8)$$

где n_b — концентрация электронов в соответствующей точке при тепловом равновесии. [Это выражение непосредственно следует из (3), если принять во внимание, что ψ и φ_p в слое b не зависят от приложенных напряжений (7).]

Мы будем иметь дело главным образом с отклонением концентраций от их равновесных значений (неравновесное значение будем отмечать в дальнейшем индексом 1). Отклонение, соответствующее соотношению (8), равно

$$n_1 = n_b [e^{-q\varphi_l/kT} - 1] \equiv -n_b q \frac{B_l}{kT}. \quad (9)$$

Введенная в это выражение величина B_l определяется как

$$B_l \equiv \frac{kT}{q} [1 - e^{-q\varphi_l/kT}]. \quad (10)$$

Эта величина введена потому, что все токи, с которыми мы будем иметь дело, также зависят от напряжения. Коэффициент B_i имеет размерность напряжения и при малых значениях φ_i приблизительно равен φ_i .

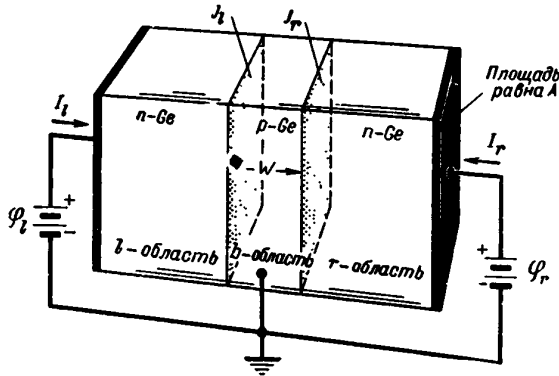
Аналогичные рассуждения приводят к подобным же соотношениям для концентрации дырок в области l :

$$p \text{ (в области } l, \text{ вблизи } J_l) = p_l e^{-q\varphi_l/kT}, \quad (11)$$

$$p_l = p_l [e^{-q\varphi_l/kT} - 1] \equiv -p_l q \frac{B_l}{kT}. \quad (12)$$

§ 6. ВОЛЬТАМПЕРНЫЕ ХАРАКТЕРИСТИКИ

Анализ, проведенный в предыдущем разделе, показывает, что вблизи J_l отклонения концентраций как дырок, так и электронов от их равновесных значений пропорциональны B_i . Эти отклонения концентрации вызывают диффузионные токи, которые должны исчезать при тепловом равновесии, когда $B_i = 0$. В линейном приближении



Фиг. 3. Схема $n-p-n$ транзистора.

(см. § 2) эти токи будут пропорциональны B_i . Выбранные обозначения для токов разного направления показаны на фиг. 3. В соответствии с этими символами, ток в область b , обусловленный потоком дырок через J_l , будем обозначать

$$I_{lp} = G_{lp} B_i. \quad (13)$$

Если площадь поперечного сечения A , проводимость и время жизни τ_{pl} сохраняют постоянное значение, то для коэффициента G_{lp} легко можно получить следующее выражение [1, 2]:

$$G_{lp} = \frac{q^2 p_{nl} A}{L_{pl}} = \frac{\sigma_i^2 b A}{(1+b)^2 \sigma_l L_{pl}}, \quad (14)$$

в котором диффузионная длина L_{pi} определяется уравнением

$$L_{pi} = (D_p \tau_{pi})^{1/2}. \quad (15)$$

Уравнение (14) выражает активную проводимость G_{ip} в зависимости от собственной электропроводности σ_i и от действительной проводимости области n -типа. Величина b равна отношению подвижностей

$$b = \frac{\mu_n}{\mu_p}. \quad (16)$$

Для случая $V_r = 0$ можно также непосредственно оценить значение электронного тока, проходящего через J_i . Некоторые из электронов, вошедших в слой b через J_i , будут проникать в область r даже при нулевом напряжении на переходе J_r . Это является следствием того, что отклонение концентрации n_1 в области перехода J_r должно быть равно нулю при $\varphi_r = 0$. Расчеты показывают, что электронные токи через $p-n$ -переходы равны [см. (6)]¹⁾

$$I_{in} = \left[\frac{q \mu_n n_b A}{L_n} \operatorname{cth} \frac{W}{L_n} \right] B_i \equiv G_{in} B_i, \quad (17)$$

$$I_{rn} = - \left[\frac{q \mu_n n_b A}{L_n} \operatorname{csch} \frac{W}{L_n} \right] B_i \equiv G_{rn} B_i. \quad (18)$$

Активную проводимость G_{in} можно выразить в зависимости от величин, характеризующих свойства слоя основного электрода:

$$G_{in} = \frac{\sigma_i^2 b A}{(1+b^2) \sigma_b L_n} \operatorname{cth} \frac{W}{L_n}. \quad (19)$$

Подобные рассуждения приводят к соответствующей системе уравнений для J_r . Воспользовавшись величинами G и B , зависимость тока от напряжения можно записать в следующем виде:

$$I_i = G_{ii} B_i + G_{ir} B_r, \quad (20)$$

$$I_r = G_{ri} B_i + G_{rr} B_r, \quad (21)$$

$$G_{ii} = G_{in} + G_{ip}, \quad G_{ir} = G_{in}, \quad (22)$$

$$G_{ri} = G_{rin}, \quad G_{rr} = G_{rrn} + G_{rrp}. \quad (23)$$

Для малых напряжений, как уже отмечалось раньше, можно написать

$$B_i \approx \varphi_i, \quad B_r \approx \varphi_r, \quad \text{при } |\varphi_{i,r}| \ll \frac{kT}{q}. \quad (24)$$

Таким образом, при малых напряжениях коэффициенты в уравнениях (20), (21) приблизительно равны компонентам активной проводимости.

¹⁾ См. также основной текст книги. — Прим. ред.

Если напряжение складывается из напряжения смещения и небольшого переменного напряжения

$$\varphi_i = V_i + v_i, \quad \varphi_r = V_r + v_r, \quad (25)$$

то уравнения для малых сигналов принимают вид

$$i_i = g_{ii}v_i + g_{ir}v_r, \quad (26)$$

$$i_r = g_{ri}v_i + g_{rr}v_r. \quad (27)$$

При этом G и g связаны соотношением

$$\frac{g_{ii}}{G_{ii}} = \frac{g_{ri}}{G_{ri}} = \exp\left(-\frac{qV_i}{kT}\right). \quad (28)$$

Подобные соотношения имеют место и для двух других коэффициентов.

Воспользовавшись этими соотношениями, можно также показать, что каждый коэффициент g пропорционален отклонению соответствующего члена GB от его значения при насыщении; в этом случае $B = kT/q$. Символически это можно записать так:

$$g = (\text{Отклонению } GB \text{ от его значения при насыщении}) \cdot \frac{q}{kT}. \quad (29)$$

Очевидно, что для модели, к которой применимы уравнения (17) и (18) вследствие симметрии, можно написать

$$G_{ir} = G_{ri}. \quad (30)$$

Если в среднем слое проводимость изменяется асимметрично или если время жизни носителей тока на одной стороне слоя больше, чем в другой, то последнее равенство не имеет силы. Однако можно показать, что если справедливы описанные в § 2 линейные приближения, то это равенство выполняется при любой геометрии среднего слоя. Доказательство этого утверждения дано в Дополнении (см. [3]). Это приводит нас к заключению, что режим работы идеализированной модели транзистора, в которой все токи являются линейными функциями концентрации носителей, описываются тремя параметрами, по крайней мере при работе в области низких частот. Этими параметрами являются коэффициенты уравнений (20) и (21). Всего их 4, но из них 2 связаны соотношением (30).

С помощью этих коэффициентов можно выразить усиление по току модели при работе любого из двух $p-n$ -переходов в качестве эмиттера. Коэффициенты усиления по току в этих двух случаях, вообще говоря, не равны между собой. При использовании левого $p-n$ -перехода в качестве эмиттера, усиление по току определяется соотношением

$$\begin{aligned} \alpha_i &\equiv -\left.\frac{\partial I_r}{\partial I_i}\right|_{\varphi_r = \text{const}} = -\frac{G_{ri}}{G_{ii}} = -\frac{G_{rin}}{G_{iin} + G_{ip}} = \\ &= -\frac{G_{rin}}{G_{iin}} \frac{G_{iin}}{G_{iin} + G_{ip}} = \gamma_i \beta_i, \quad (31) \end{aligned}$$

где γ — та часть тока, текущего через эмиттерный переход (под влиянием приложенного к эмиттеру напряжения), которая переносится носителями, находящимися в области основного электрода в меньшем числе. Очевидно, что

$$\gamma_i = \frac{G_{in}}{G_{in} + G_{np}}. \quad (32)$$

Та часть этих вошедших в основной электрод носителей, которая доходит до коллектора, определяется соотношением

$$\beta_i = -\frac{G_{im}}{G_{in}} = \operatorname{sch} \frac{W}{L_n}. \quad (33)$$

„Собственный“ коэффициент усиления по току («intrinsic α » или α^*) коллекторного перехода определяется как отношение изменения полного тока к изменению тока, который обуславливается входящим до него потоком носителей заряда с меньшей концентрацией. Коэффициент α^* простого коллектора с $p-n$ -переходом равен 1. В коллекторе с ловушкой, который мы рассмотрим в следующем параграфе, доходящий до коллектора ток вызывает поток электронов и α^* может доходить до 100 и даже больше.

Если средний слой настолько тонок, что гиперболические функции (17) и (18) можно аппроксимировать первыми членами их разложения в ряд, то

$$\beta_i \approx 1 - \frac{1}{2} \left(\frac{W}{L_n} \right)^2 \approx 1, \quad (34)$$

$$\gamma_i \approx \frac{1}{1 + (\sigma_b W / \sigma_i L_{pi})}, \quad (35)$$

$$\alpha_i \approx \frac{1}{1 + (\sigma_b W / \sigma_i L_{pi})}, \quad (36)$$

$$G_{im} \approx \frac{\sigma_i^2 b A}{(1 + b)^2 \sigma_b W}. \quad (37)$$

Заменяя l на r , подобные выражения можно получить также и для J_r .

§ 7. СПЕЦИАЛЬНЫЕ РЕЖИМЫ РАБОТЫ

Используем выведенные в предыдущем параграфе уравнения для анализа работы транзисторов в некоторых специальных случаях их применения.

А. Работа транзистора в качестве усилителя. Для работы в области линейного режима к коллектору необходимо приложить достаточно большое напряжение, чтобы его ток достиг насыщения. Это условие соответствует прямолинейным участкам характеристик, показанных на фиг. 2. Необходимое для этого напряжение должно превы-

шать $4kT/q$. При выполнении этого условия значения токов эмиттера и коллектора приближенно могут быть выражены уравнениями

$$I_i = G_{ii}B_i - \frac{\alpha_i G_{ii} kT}{q}, \quad (38)$$

$$I_r = -\alpha_i G_{ii}B_i + \frac{G_{rr} kT}{q} = -\alpha_i I_i + (G_{rr} + \alpha_i^2 G_{ii}) \frac{kT}{q}. \quad (39)$$

Уравнение (39) объясняет параллельность линейных характеристик (см. фиг. 2, а), соответствующих различным значениям тока I_i .

Для расчета схем важно знать полную проводимость эмиттера g_{ii} . В принятом в этой статье приближении, в котором не учитываются также последовательные омические сопротивления, эта полная проводимость равна $1/r_e$ (r_e — одно из сопротивлений эквивалентного контура). Наиболее интересная область обычно соответствует значениям $-\varphi_i > 4kT/q$, для которой на основании (29) можно написать

$$\frac{1}{r_e} = g_{ii} = \frac{\partial I_i}{\partial \varphi_i} = G_{ii} e^{-q\varphi_i/kT} \approx \frac{I_i q}{kT} \approx 40 I_i \text{ ом}^{-1}. \quad (40)$$

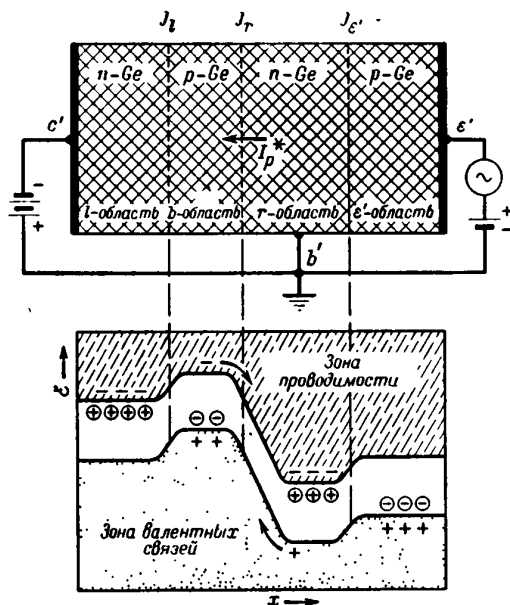
Если транзистор работает с заземленным эмиттером и значение тока коллектора соответствует области насыщения, то полная проводимость входа равна

$$\frac{dI_b}{d\varphi_b} = \frac{d(-I_i - I_r)}{d(-\varphi_b)} = (1 - \alpha_i) G_{ii} e^{-q\varphi_i/kT}, \quad (41)$$

в то время как крутизна характеристики в $1/(1 - \alpha_i)$ раз больше. Таким образом, если $\alpha_i = 0,99$, то при работе в области насыщения (тока коллектора) усиление по току будет достигать 100. Согласно (39) идеальному насыщению соответствует бесконечно большое полное сопротивление коллектора и теоретическое усиление по мощности, таким образом, становится в этом случае также бесконечно большим. На самом деле, сопротивление коллектора составляет в настоящее время $10^7 - 10^8 \text{ ом}$. Для токов, больших по сравнению с током насыщения, соотношение (41) при $I_r = 10^{-4} a$ (типичное значение для характеристик, соответствующих высокому коэффициенту усиления) приводит к следующему значению для полной проводимости $(1 - \alpha_i) I_i q/kT = 10^{-2} \cdot 10^{-4} \cdot 40 = 4 \cdot 10^{-6} \text{ ом}^{-1}$. Усиление по мощности при согласованной нагрузке примерно равно отношению полных сопротивлений выхода и входа, умноженному на квадрат усиления по току, т. е. $10^7 \cdot 4 \cdot 10^{-6} \cdot 10^4 \approx 66 \text{ дб}$. Как указано в § 4, усиление по мощности в практически применяемых схемах доходит уже до 50 дб.

Б. Коллектор с ловушкой в $p-n-p$ -транзисторе. В точечно-контактных транзисторах часто наблюдается усиление по току в том смысле, что при постоянном напряжении смещения на коллекторе изменение тока эмиттера приводит к значительно боль-

шему (в несколько раз) изменению тока коллектора. Это, возможно, объясняется тем, что около коллекторного перехода образуется „ $p-n$ -ловушка“, что приводит к высоким значениям „собственного“ коэффициента усиления по току коллектора. Мы будем теоретически иллюстрировать эту теорию, воспользовавшись моделью транзистора n -типа, в котором дырки вводятся эмиттером в виде $p-n$ -перехода и коллектор которого состоит из слоев с электропроводностью n - и p -типа. Эта модель показана на фиг. 4. Смещения в этом случае такие же,



Фиг. 4. Коллектор с $p-n$ -ловушкой в $p-n-p-n$ -транзисторе.

как и у транзистора, показанного на фиг. 1, если не считать дополнительной области эмиттера и отсутствия электрода у слоя b . Здесь e' , b' и c' — соответственно эмиттер, основной электрод и коллектор. Действующие смещения соответствуют обратному напряжению на переходе J_r и прямому на переходе J_i . Слой основного электрода b будет находиться под таким потенциалом, что результирующий ток через него будет равен нулю:

$$I_r + I_i = G_{ii}(1 - \alpha_i) V_i + G_{rr}(1 - \alpha_r) V_r = 0. \quad (42)$$

При больших обратных смещениях значение V_r будет положительно, V_i — отрицательно, а

$$\varphi_i - \varphi_r = V_{c'}, \quad (43)$$

где $V_{c'}$ — напряжение коллектора. При больших $V_{c'}$ значение B_r будет соответствовать области насыщения и φ_i будет определяться соотношением

$$B_i = - \frac{G_{rr} (1 - \alpha_r) \frac{kT}{q}}{G_{ii} (1 - \alpha_i)}, \quad (44)$$

так что ток насыщения будет равен

$$I_r (\text{насыщения}) = \frac{(1 - \alpha_r \alpha_i) G_{rr} kT}{q (1 - \alpha_i)}. \quad (45)$$

Если дырочный ток I_p^* , обусловленный выходящими из v' дырками, достигает J_p , то условие

$$I_r + I_i + I_p^* = 0 \quad (46)$$

приведет к изменению

$$\Delta B_i = - \frac{I_p^*}{G_{ii} (1 - \alpha_i)}, \quad (47)$$

благодаря которому электронный ток через J_r увеличивается на

$$\Delta I_{nr} = G_{ri} \Delta B_i = \frac{\alpha_i I_p^*}{1 - \alpha_i}, \quad (48)$$

причем α^* этой сложной модели определяется соотношением

$$\alpha^* = \frac{\Delta I_{nr} + I_p^*}{I_p^*} = \frac{1}{1 - \alpha_i}. \quad (49)$$

В указанном выше приближении для тонкого слоя

$$\alpha^* \approx 1 + \frac{\alpha_i l_p l_i}{\sigma_b W}. \quad (50)$$

Таким образом, при малой толщине слоев и при большой проводимости l областей можно получить очень большие значения α^* .

Нижняя часть фиг. 4 объясняет название „ p — n -ловушка“. Высокий потенциал для электронов в слое b является низким потенциалом для дырок. Дырки, выходящие из области v' , захватываются этой ловушкой, в результате чего возникает прямое смещающее напряжение на J_i , что вызывает усиление электронного тока.

В. Фототранзистор. Модель, показанная на фиг. 4, может работать также и в качестве фототранзистора, если вместо введения дырок из области эмиттера создавать пары электрон — дырка с помощью света. В этом случае необходимо иметь только два электрода (b' и c'), причем область v' отсутствует. Если пары создаются около J_r , то под действием поля электроны и дырки будут двигаться в противо-

положных направлениях. Число дырок, проходящих при этом через J_r , будет равно числу разделенных пар. Усиление дырочного тока, проходящего через переход J_r , будет происходить при этих условиях точно так же, как в случае $p-n-p-n$ -транзистора. Таким образом, кажущийся квантовый выход для электронно-дырочных пар, создаваемых у J_r [на основании тех же самых причин, которые привели к уравнению (50)], равен

$$\alpha^* \approx 1 + \frac{\sigma_i L_{pi}}{\sigma_b W}. \quad (51)$$

В следующем параграфе мы приведем некоторые экспериментальные значения α^* для фототранзисторов.

Г. Внутренние контактные потенциалы. Для рассмотрения вопроса о внутренних контактных потенциалах мы вернемся к модели, показанной на фиг. 3. Если область b заземляется, а область r может принимать свой собственный потенциал, то изменение потенциала области l будет вызывать изменение потенциала области r , несмотря на то, что потенциал в области b , измеренный с помощью омического контакта, был бы всюду равен нулю. Свободный потенциал незаземленной области r определяется из условия, что полный ток должен быть равен нулю, т. е.

$$I_r = G_{ri} B_i + G_{rr} B_r = 0, \quad (52)$$

$$B_r = -\frac{G_{ri}}{G_{rr}} B_i = -\frac{G_{ir}}{G_{rr}} B_i = \alpha_r B_i, \quad (53)$$

$$e^{-q\varphi_r/kT} = 1 - \alpha_r + \alpha_r e^{-q\varphi_i/kT}. \quad (54)$$

Уравнение (54) выражает φ_r в зависимости от φ_i и α_r . Оно принимает простую форму для предельных условий смещения.

Смещение в прямом направлении: $-q\varphi_i/kT \gg 1$

$$\varphi_r = \varphi_i - \frac{kT}{q} \ln \alpha_r, \quad (55)$$

$$\alpha_r = e^{q(\varphi_i - \varphi_r)/kT}. \quad (56)$$

Нулевое смещение: $|q\varphi_i/kT| \ll 1$

$$\varphi_r = \alpha_r \varphi_i. \quad (57)$$

Смещение в обратном направлении: $q\varphi_i/kT \gg 1$

$$\varphi_r = -\frac{kT}{q} \ln(1 - \alpha_r), \quad (58)$$

$$\alpha_r = 1 - e^{-q\varphi_r/kT}. \quad (59)$$

В следующем параграфе мы покажем, что эти соотношения приблизительно выполняются.

§ 8. СРАВНЕНИЕ С ЭКСПЕРИМЕНТОМ

В этом параграфе будут рассмотрены экспериментальные данные, полученные главным образом с $n-p-n$ -транзисторами и будет показано, что их можно объяснить на основе изложенной выше теории. Эти данные были получены при двух различных условиях; в первом случае напряжения были малы по сравнению с kT/q . Эти данные используются для определения величин G на основе приближения (24). Вторая группа данных была получена для широкой области напряжений. При всех условиях, однако, один из двух коэффициентов B рассматривается как независимая величина, а другой считается либо постоянным, либо пропорциональным первому. Следовательно, любой ток можно выразить в форме

$$I = c + mB = I_S + m \left[B - \frac{kT}{q} \right] = I_S - I_V e^{-qV/kT}, \quad (60)$$

где V — напряжение, от которого зависит B ,

$$I_S = c + \frac{mkT}{q} \quad (61)$$

значение тока насыщения при больших положительных значениях V и

$$I_V = \frac{mkT}{q}. \quad (62)$$

Как I_S , так и I_V легко вычисляются в зависимости от величин G . Вычисленные таким образом значения I_S и I_V сравниваются здесь с их экспериментальными значениями, а также исследуется зависимость I от V .

Необходимо отметить, что значения величин G сильно зависят от температуры. Рассмотрим, например,

$$G_{rl} = \frac{b\sigma_i^2 A}{(1+b)^2 \sigma_p W}. \quad (63)$$

В этом выражении ¹⁾

$$\sigma_i^2 \sim e^{-\epsilon_G/kT}, \quad (64)$$

где ϵ_G — ширина запрещенной зоны и

$$\sigma_p \sim T^{-3/2}. \quad (65)$$

Следовательно, значение G_{rl} при $T_0 + \Delta T = 300^\circ \text{K} + \Delta T$ приблизительно равно

$$G_{rl}(T_0 + \Delta T) \approx G_{rl}(T_0) \exp \left[\left(\frac{\epsilon_G}{kT_0^2} \right) + \left(\frac{3}{2T_0} \right) \right] \Delta T = G_{rl}(T_0) e^{(0,095 + 0,005) \Delta T}, \quad (66)$$

¹⁾ n_i^2 изменяется как $T^3 \exp(-\epsilon_G/kT)$, а $\mu_n \mu_p$ — как T^{-3} [2, стр. 475 и 287]. — Прим. авт.

т. е. при повышении температуры на 1°C G_{ji} возрастает приблизительно на $10^0/0$. Это возрастание объясняется главным образом изменением σ_j^3 и приблизительно одинаково для всех величин G .

Большой температурный коэффициент величин G объясняет очень большую зависимость токов от температуры при постоянном напряжении. Однако это не означает, что в правильно рассчитанном контуре транзистор будет очень чувствительным к изменениям температуры. Например, значение α_1 , как видно из соотношения (36), зависит только от величины $\sigma_b W / \sigma_e L_{pe}$, имеющей небольшой температурный коэффициент. При постоянном токе эмиттера его сопротивление пропорционально абсолютной температуре. Таким образом, наиболее важные с точки зрения работы схем величины имеют небольшие температурные коэффициенты.

Так как в этом параграфе рассматриваются транзисторы, имеющие по одному контакту в качестве эмиттера и по одному в качестве коллектора, то мы откажемся от терминологии „левый“ и „правый“ и используем более привычные для транзисторов индексы: 1 — для эмиттера и 2 — для коллектора. Тогда уравнения для токов примут следующий вид:

для больших сигналов

$$I_e = G_{11}B_1 + G_{12}B_2, \quad (67)$$

$$I_c = G_{21}B_1 + G_{22}B_2; \quad (68)$$

для небольших сигналов

$$I_e = g_{11}v_e + g_{12}v_c, \quad (69)$$

$$i_c = g_{21}v_e + g_{22}v_c, \quad (70)$$

где каждый коэффициент g зависит от соответствующего ему напряжения

$$g_{ij} = G_{ij}e^{-qV_j/kT}. \quad (71)$$

В области низких напряжений активные проводимости были измерены при напряжении 2 мв. Для этого небольшого напряжения

$$B(+2 \text{ мв}) = 1,95 \text{ мв}, \quad (72)$$

$$B(-2 \text{ мв}) = -2,05 \text{ мв}, \quad (73)$$

так что токи в этой области почти линейны и даже это незначительное отклонение их от линейности практически устраняется усреднением по двум полярностям. Величина I_e измерялась с заземленным коллектором, G_{11} было вычислено из отношения I_e/V_e , а I_e и V_c были измерены с разомкнутой цепью коллектора. В этом случае $I_e/V_e = G_{11}(1 - \alpha_1\alpha_2)$ и $V_c/V_e = -G_{21}/G_{22} = \alpha_2$. Подобные данные были получены также с приложенным к коллектору напряжением. Наконец, коллектор и эмиттер были соединены вместе и измерен суммарный ток. В этом

случае

$$\frac{I_e + I_c}{V_e} = G_{11} + G_{22} - G_{12} - G_{21}. \quad (74)$$

Для величин G были выбраны следующие значения (активные проводимости в мком^{-1} при 22°C):

$$G_{11} = 8,8; \quad G_{22} = 33,3; \quad G_{12} = G_{21} = -7,8. \quad (75)$$

Как показано в табл. 1, эти значения в пределах точности эксперимента соответствуют результатам измерений. Возможность получения такого соответствия с помощью трех постоянных, конечно, нельзя еще рассматривать как окончательное подтверждение теории, изложенной в § 6. Однако любой пассивный трехклеммный прибор в отсутствие магнитного поля должен удовлетворять условию обратимости и описываться тремя постоянными. Табл. 1 показывает по существу согласованность измерений.

Таблица 1

АКТИВНЫЕ ПРОВОДИМОСТИ $n-p-n$ -ТРАНЗИСТОРА ПРИ НУЛЕВОМ СМЕЩЕНИИ
(в мком^{-1} при $T = 22^\circ\text{C}$)

	Измеренные	Вычисленные
G_{11}	$8,8 \pm 0,5$	8,8
G_{22}	$33,3 \pm 0,5$	33,3
$G_{11}(1 - \alpha_1\alpha_2)$	$6,9 \pm 0,5$	6,9
$G_{22}(1 - \alpha_1\alpha_2)$	$26,5 \pm 0,5$	26,4
$\alpha_2 = V_c/V_e$ при $I_c = 0$	$0,86 \pm 0,02$	0,89
$\alpha_1 = V_c/V_c$ при $I_c = 0$	$0,22 \pm 0,02$	0,23
$G_{11} + G_{12} - 2G_{21}$	$26,2 \pm 0,5$	26,5

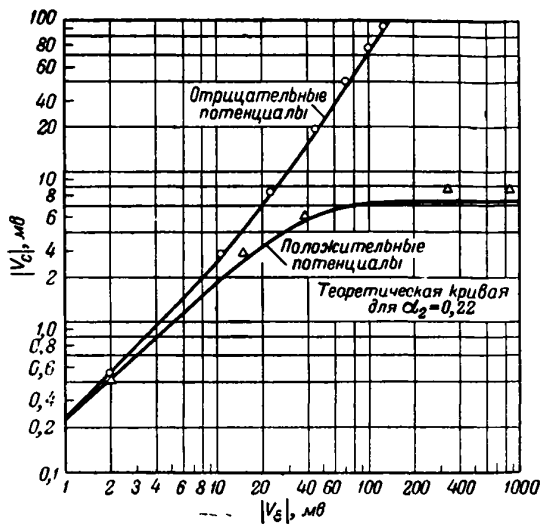
Точных значений постоянных для слоя основного электрода изучаемого транзистора получить не удалось. Однако порядок величин G находится в разумном согласии с их оценками, полученными на основании изучения свойств исходных материалов и способа изготовления усилителей. Мы не рассчитываем на точное совпадение экспериментальных и теоретических данных¹⁾, но для примера рассмотрим транзистор с $A = 0,003 \text{ см}^2$, $W = 2,5 \cdot 10^{-3} \text{ см}$, $\sigma_a = 20$, $\sigma_b = 10$, $\sigma_c = 0,1$, $\sigma_i = 1/60 \text{ ом}^{-1} \cdot \text{см}^{-1}$, $\tau_{pc} = 40$ и $\tau_{pa} = 10$ мксек. При этих условиях уравнения (13), (34) и (37) дают следующие значения величин G (в мком^{-1}):

$$G_{11n} = 7,3, \quad (76)$$

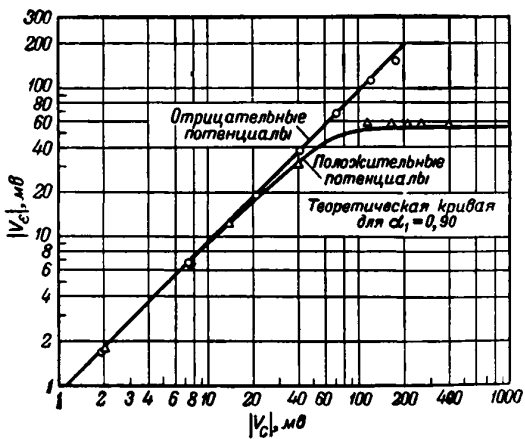
$$G_{11p} = 0,5, \quad (77)$$

$$G_{22p} = 45. \quad (78)$$

¹⁾ Для наиболее простого случая $p-n$ -перехода расхождение экспериментальных и теоретических данных составляет примерно 20%. — *Прим. авт.*



Ф и г. 5. Внутренний контактный потенциал на разомкнутом коллекторе.



Ф и г. 6. Внутренний контактный потенциал на разомкнутом эмиттере.

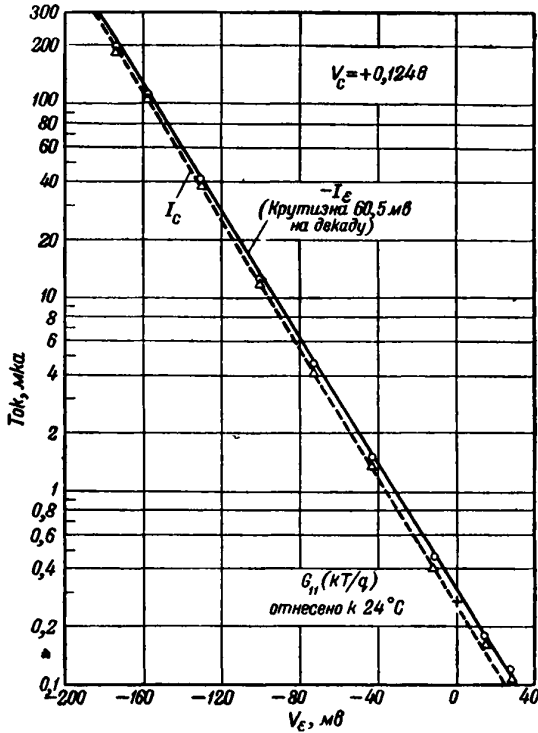
При $\tau_n = 40$ мксек. значение β больше 0,99 и, следовательно, может быть принято при вычислении величин G равным 1. Теперь можно вычислить

$$G_{11} = G_{11n} + G_{11p} = 7,8, \quad (79)$$

$$-G_{12} = -G_{12n} = \beta_1 G_{11n} = \beta_2 G_{22n} = 7,3, \quad (80)$$

$$G_{22} = G_{22n} + G_{22p} = 7,3 + 45 = 52,3. \quad (81)$$

Согласно изложенной выше теории, отличие значения α_1 от единицы обуславливается главным образом потоком дырок через J_c . Подобное



Фиг. 7. Зависимость токов эмиттера и коллектора от напряжения на эмиттере (на коллектор подано смещение до насыщения).

заключение остается верным и для α_2 . Ток основного электрода почти полностью определяется этими потоками дырок, причем степень их рекомбинации в области основного электрода незначительна.

Влияние внутренних контактных потенциалов показано на фиг. 5 и 6.

На фиг. 5 через V_c обозначено приложенное к эмиттеру напряжение, а через V_c — измеренный потенциал коллектора при отсутствии

тока. Полученные данные находятся в общем согласии с теорией и значением α_2 (α_2 соответствует переходу коллектора), полученным при измерениях в области низких напряжений и использованным при составлении табл. 1.

На фиг. 6 показаны результаты, полученные при подаче напряжения на коллектор. Видно, что эти данные незначительно отличаются от теоретических кривых для случая обратных смещений. Подстановка полученных результатов в уравнение (59) приводит к следующим значениям α :

$$\alpha_2 = 0,226 \quad \text{и} \quad \alpha_2 = 0,894, \quad (82)$$

которые показывают, что степень согласия теоретических и экспериментальных данных определяется небольшими изменениями α . Проверку уравнения (56), относящегося к случаю прямого смещения, нельзя считать удовлетворительной, так как при этих условиях токи относительно велики и падением напряжения на последовательно включенных сопротивлениях нельзя пренебрегать. Полученные значения, однако, приблизительно совпадают с приведенными в табл. 1.

На фиг. 7 даны зависимости I_c и I_e от V_e при подаче на коллектор смещения, соответствующего области насыщения. Формулы для этого случая имеют вид

$$I_c = G_{11}B_1 + G_{12} \frac{kT}{q} = G_{11}(1 - \alpha_1) \frac{kT}{q} - G_{11} \frac{kT}{q} e^{-qV_e/kT}, \quad (83)$$

$$I_e = G_{12}B_1 + G_{22} \frac{kT}{q} = G_{22}(1 - \alpha_2) \frac{kT}{q} + \alpha_1 G_{11} \frac{kT}{q} e^{-qV_e/kT}. \quad (84)$$

Видно, что графики, приведенные на фиг. 7, соответствуют экспоненциальной зависимости между током и напряжением на эмиттере, а наклон их находится в хорошем согласии с теорией, согласно которой при изменении тока в 10 раз изменение напряжения должно быть равно (все значения соответствуют $T = 297^\circ \text{K}$)

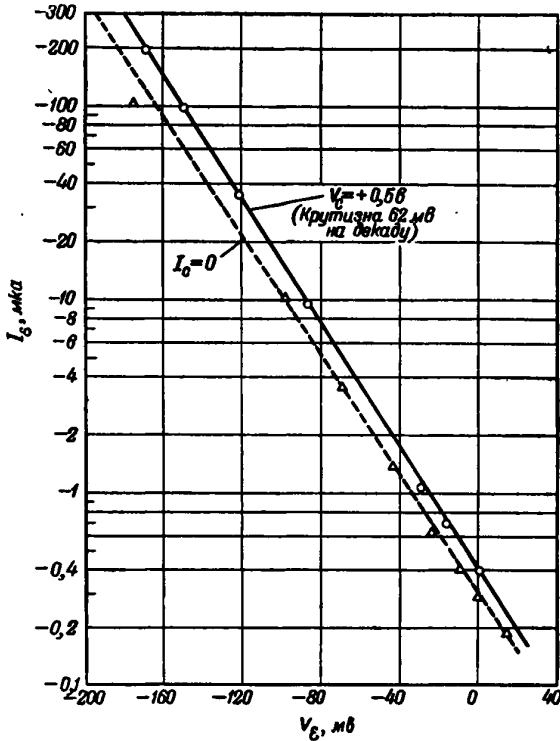
$$\Delta V = 2,30 \frac{kT}{q} = \frac{2,30}{39,4} = 59,0 \text{ мв}. \quad (85)$$

Данные табл. 1 (данные исправлены для $\Delta T = 2^\circ \text{C}$) и фиг. 7 приводят к следующим значениям I_S и I_V (табл. 2):

Таблица 2

	Данные фиг. 7	Данные табл. 1
$G_{11}(1 - \alpha_1) kT/q$	0,021 мка	0,025 мка
$G_{11} (kT/q)$	0,30 "	0,27 "
$G_{22}(1 - \alpha_2) kT/q$	0,88 "	0,78 "
$\alpha_1 G_{11} kT/q$	0,27 "	0,24 "

Значения, соответствующие области насыщения, были получены непосредственно из опытных данных. В использованном интервале напряжений насыщение в области коллектора не было полным и значения, соответствующие насыщению, были исправлены введением члена, учитывающего сопротивление утечки порядка 1 мгом



Фиг. 8. Зависимость тока эмиттера от напряжения эмиттера (для двух режимов коллектора).

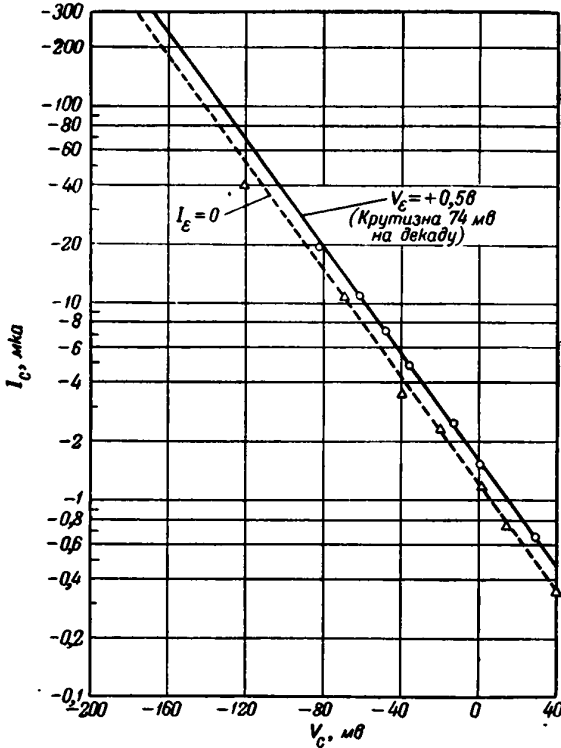
(происхождение этого эффекта утечки остается неясным, и ток коллектора стремится к насыщению при более высоких обратных смещениях).

На фиг. 8 снова показана зависимость между током и напряжением эмиттера. Транзистор находился в этом опыте при несколько более высокой температуре. Измерения производились при более высоком напряжении на коллекторе ($V_C = +0,5 \text{ в}$) в одном случае и со свободным коллектором ($I_C = 0$) в другом. Отношение двух

членов должно было бы быть равно

$$\frac{G_{11}}{G_{11}(1 - \alpha_1 \alpha_2)} = 1,25. \quad (86)$$

В удовлетворительном согласии с теорией экспериментальное значение этого отношения равно 1,30.



Фиг. 9. Зависимость тока коллектора от напряжения коллектора (для двух режимов эмиттера).

Полученная при аналогичных условиях зависимость тока коллектора от напряжения коллектора показана на фиг. 9. В хорошем согласии с теорией отношение двух членов в этом случае также равно 1,30. Необходимо отметить, однако, что крутизна при этом (при изменении напряжения на 74 мВ ток изменяется в 10 раз) соответствует слишком высокой температуре $+103^\circ \text{C}$. Такая крутизна имеет место при таких

малых значениях тока, что объяснить это побочными эффектами, вносимыми последовательно включенными сопротивлениями, становится уже трудно.

Имеется существенная разница между природой токов, зависимость которых от напряжения показана на фиг. 7, и тока коллектора, зависимость которого от напряжения на коллекторе показана на фиг. 9. Первые переносятся главным образом электронами, диффундирующими через слой основного электрода и попадающими в область коллектора. Основанием для этого заключения является то, что в этом случае значение α_1 равно 0,89, а это означает, что только 11% тока обуславливается потоком электронов, рекомбинирующих в слое основного электрода, или потоком дырок, диффундирующих в область эмиттера и рекомбинирующих там (или обратными процессами, в зависимости от полярности). С другой стороны, большая часть тока коллектора обуславливается потоком дырок, рекомбинирующих в области коллектора [это заключение находится в согласии с объяснением значений величин G в уравнениях (79)—(81)]. С точки зрения теории нужно ожидать, что если процесс рекомбинации заключается в захвате носителей тока центрами рекомбинации, ловушками [2, стр. 342], то в этом случае скорость рекомбинации с увеличением концентрации введенных носителей тока будет возрастать медленнее, чем по линейному закону, так как с увеличением концентрации носителей тока ловушки будут стремиться к насыщению. Эта точка зрения находит себе некоторое экспериментальное подтверждение в работе Гаучера и Хайнса, которые нашли, что с увеличением концентрации носителей тока время их жизни увеличивается. Может быть, этот механизм объясняет также более медленное, чем это должно быть, возрастание тока с увеличением напряжения на коллекторе (см. фиг. 9).

Нелинейность процесса рекомбинации дырок в области эмиттера подтверждается также зависимостью α_1 от тока эмиттера. Это можно видеть на фиг. 7. С увеличением тока эмиттера отношение тока коллектора к току эмиттера (сверх насыщения) возрастает от 0,9 примерно до 0,95. Это обусловлено более медленным ростом дырочного тока при увеличении напряжения, чем это имеет место для тока эмиттера.

Таким образом, тенденция α к увеличению с возрастанием тока эмиттера является, по всей вероятности, общим свойством $n-p-n$ -транзисторов. Уоллс и Питенполь [3] показали, что значение α может доходить до 0,995.

Очень большую чувствительность имеют $n-p-n$ -фототранзисторы, и их квантовый выход составляет, по крайней мере, несколько сотен. Согласно изложенному в § 7, такая большая величина квантового выхода привела бы к таким значениям α_1 ($\alpha_1 = 1 - 1/\alpha^*$), которые по своей величине сравнимы или даже больше, чем самые большие их значения для $n-p-n$ -транзисторов.

§ 9. НЕКОТОРЫЕ РАСЧЕТНЫЕ СООБРАЖЕНИЯ

Основная цель предыдущих параграфов заключалась в рассмотрении следствий диффузионной теории и в сравнении их с экспериментом. Для этой цели исследовалась работа приборов в наиболее простых условиях — при малых токах и нулевой частоте. Однако с точки зрения практического применения приборов не менее важное значение имеет их работа при больших токах и высоких частотах. В этом параграфе мы кратко рассмотрим наиболее важные из факторов, которые необходимо учитывать при расчете приборов.

Большое значение имеет предельная рабочая частота. Она может определяться характером внешней цепи или некоторыми внутренними свойствами транзистора. Наиболее важным из этих последних является время диффузии через слой основного электрода. Это время равно

$$\tau_D = \frac{W^2}{D_n}. \quad (87)$$

При круговой частоте ω решение для n_1 в слое основного электрода имеет вид

$$\exp[i\omega t \pm x(1 + i\omega\tau_n)^{1/2}/L_n], \quad (88)$$

и это дает следующее значение для β :

$$\beta = \frac{\operatorname{sch}(1 + i\omega\tau_n)^{1/2} W}{L_n}. \quad (89)$$

Для частот, удовлетворяющих условию $\omega\tau_n \gg 1$, это приводит к

$$\beta = \frac{2}{\exp(1+i)\left(\frac{\omega\tau_D}{2}\right)^{1/2} + \exp(-1-i)\left(\frac{\omega\tau_D}{2}\right)^{1/2}}. \quad (90)$$

Из этого уравнения видно, что при $\omega\tau_D \gg 2$ имеется отставание фазы на $(\omega\tau_D/2)^{1/2}$ радианов и равнос ему по величине затухание в неперах. Усиление мощности, которое пропорционально во многих случаях β^2 , уменьшается приблизительно на $3 \delta\theta$ при $\omega\tau_D = 2$ или $f = D_n/\pi W^2 \approx 30/W^2$. При $W = 2,5 \cdot 10^{-3}$ см это составляет примерно $5 \cdot 10^6$ гц.

В дополнение к этому основному ограничению могут быть еще ограничения, обусловленные емкостью и омическими сопротивлениями. Для иллюстрации этого рассмотрим схему с транзистором, эмиттер которого заземлен. Такая схема аналогична ламповой схеме с заземленным катодом. В этом случае сигналы переменного тока, подводимые к основному электроду, создают также некоторое переменное падение напряжения на коллекторе. Это напряжение заряжает емкость $p-n$ -перехода коллектора, причем заряжающий ток должен обуславливаться потоком дырок в слой основного электрода. Для слоя

с $\sigma_b = 10 \text{ ом}^{-1} \cdot \text{см}^{-1}$ и $W = 2,5 \cdot 10^{-3} \text{ см}$ сопротивление единичного квадрата от одного его края до противоположного равно 40 ом. Если емкость равна 2000 пф/см^2 , а усиление по напряжению равно 10, то входной сигнал будет влиять на потенциал слоя основного электрода только при ширине транзистора меньшей, чем

$$(10 \cdot \omega \cdot 2 \cdot 10^{-9} \cdot 40)^{-1/2} = (\omega \cdot 8 \cdot 10^{-7})^{-1/2}.$$

При частоте $5 \cdot 10^6 \text{ гц}$ это соответствует ширине, равной 2 мм. Ток, протекающий через слой основного электрода, должен заряжать также емкость эмиттера, которая может быть значительно больше емкости коллектора, обусловленной диффузионными эффектами.

Если транзистор слишком широк, то сигнал, приложенный к основному электроду, будет ослабляться (различная плотность тока по ширине основного электрода) и при работе будет использоваться поэтому только часть слоя основного электрода. В оставшейся части слоя основного электрода будут преобладать емкостные напряжения. Обусловленные этими напряжениями токи будут создавать, в свою очередь, большие активные емкости между эмиттером и коллектором.

Очевидно, что рассмотренные здесь эффекты требуют детальных вычислений при расчете любого частного случая. Однако общие физические принципы, на которых может быть основан расчет $n-p-n$ -транзисторов, остаются такими же, как и в случае $p-n$ -переходов и $n-p-n$ -транзисторов, работающих при малых смещениях и низких частотах.

Небольшая величина шумовых чисел этих транзисторов не может быть объяснена до создания хорошо разработанной теории генерации шумов. Однако их величина находится в грубом согласии с теорией, основанной на представлении о модуляции шумами механизма рекомбинации [2]. Согласно этой теории, каждый элемент объема является источником шумов, значение квадрата тока помех которого пропорционально квадрату отклонения концентрации меньшинства носителей тока от ее нормального значения. Применение этой теории к $n-p-n$ -модели и к транзисторам типа А показывает, что наблюдаемая разница между их шумовыми числами, равная 40 дБ и больше, может быть объяснена различными плотностями токов и геометрией прибора.

ДОПОЛНЕНИЕ

Доказательство равенства $G_{ir} = G_r$

Мы докажем это равенство, допустив, что концентрация электронов в слое основного электрода мала по сравнению с концентрацией дырок при тепловом равновесии. В этом случае влиянием введенных электронов на распределение потенциала можно пренебречь и отклонение концентрации электронов от ее равновесного значения выразить линейным уравнением (фиг. 10). $I_{ni}(B_i, B_r)$ обозначает электронный

ток, проходящий в основной электрод через J_l . Равенство $G_{lr} = G_{rl}$ будет доказано, если показать, что

$$I_{nl}(0, B) = I_{nr}(B, 0). \quad (\text{Д. 1})$$

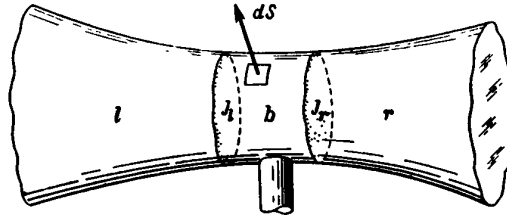
Мы предполагаем, что в слое основного электрода электростатический потенциал ψ и время жизни τ являются произвольными функциями положения. В качестве граничного условия на внешних поверхностях и на металлическом контакте принимается, что

$$I_n dS = -qn_1 s |dS|, \quad (\text{Д. 2})$$

где dS — внешняя нормаль, s — скорость поверхностной рекомбинации, а

$$n_1 = n - n_p \quad (\text{Д. 3})$$

есть отклонение концентрации n от ее значения при тепловом равновесии.



Фиг. 10. Область вблизи основного электрода, по поверхности которой производится интегрирование.

Приняты следующие обозначения решений, соответствующих напряжениям, приложенным к двум $p-n$ -переходам:

$$V_l = B, \quad V_r = 0, \quad n_1, \quad I_n, \quad (\text{Д. 4})$$

$$V_l = 0, \quad V_r = B, \quad n'_1, \quad I'_n. \quad (\text{Д. 5})$$

Выражения для токов (Д. 1) принимают вид

$$I_{nl}(0, B) = - \int_{J_l} I'_n dS, \quad (\text{Д. 6})$$

$$I_{nr}(B, 0) = - \int_{J_r} I_n dS. \quad (\text{Д. 7})$$

Для доказательства равенства (Д. 1) введем вектор

$$\mathbf{A} = \frac{n_1}{n_b} I'_n - \frac{n'_1}{n_b} I_n. \quad (\text{Д. 8})$$

Так как на внешних поверхностях $I_n dS$ пропорционально n_1 , то $A ds = 0$ на этих поверхностях. Следовательно, $\int A dS$ по поверхности слоя основного электрода равен

$$\int A dS = \frac{qB}{kT} [-I_{ni}(0, B) + I_{nr}(B, 0)], \quad (\text{Д. 9})$$

так как мы имеем на переходах

$$\frac{n_1}{n_b} = \begin{cases} \frac{qB}{kT} & \text{на } J_l, \\ 0 & \text{на } J_r, \end{cases} \quad (\text{Д. 10})$$

$$\frac{n'_1}{n_b} = \begin{cases} 0 & \text{на } J_l, \\ \frac{qB}{kT} & \text{на } J_r. \end{cases} \quad (\text{Д. 11})$$

Кроме того, можно показать, что в области основного электрода $\nabla \cdot \mathbf{A} = 0$ и, следовательно, по теореме Гаусса $\int \mathbf{A} dS = 0$.

Таким образом, равенство (Д. 1) доказано. Справедливость равенства $\nabla \cdot \mathbf{A} = 0$ вытекает из симметричности выражения $\nabla \cdot (n_1/n_b) I'_n$ относительно n_1 и n'_1

$$\begin{aligned} \nabla \cdot \left(\frac{n_1}{n_b} I'_n \right) &= \frac{1}{n_b} (\nabla n_1 - n_1 \nabla \ln n_b) \cdot I'_n + \frac{n_1}{n_b} \nabla \cdot I'_n = \\ &= \frac{1}{n_b} \left[(\nabla n_1 + n_1 q \frac{E}{kT}) (q v_n n'_1 E + q D_n \nabla n'_1) - q \frac{n_1 n'_1}{\tau} \right]. \quad (\text{Д. 12}) \end{aligned}$$

Благодаря этой симметричности дивергенции два члена в соотношении (Д. 8) сокращаются.

ЛИТЕРАТУРА

1. Shockley W., Bell Syst. Techn. Journ., 28, 435 (1949).
2. Wallace R. L., Piefenpol W. J., Bell Syst. Techn. Journ., July, 1951.
3. Shockley W., Phys. Rev., 78, 294 (1950).
4. Shive J. N., Phys. Rev., 76, 575 (1949).
5. Bardeen J., Bell Syst. Techn. Journ., 29, 469 (1950).

Дж. МОРТОН

СОВРЕМЕННОЕ СОСТОЯНИЕ ТРАНЗИСТОРНОЙ ЭЛЕКТРОНИКИ ¹⁾

{Bell Syst. Techn. Journ., 31, № 3, 411 (1952)}

§ 1. ВВЕДЕНИЕ

Когда почти четыре года тому назад был изобретен транзистор, стало очевидным, что перед техникой связи открылось новое направление развития. Этот относительно простой прибор смог выполнять самые разнообразные функции в усилительных схемах, где ранее почти монопольно господствовали электронные лампы. Прибор оказался небольших размеров, не требовал расхода энергии на подогрев, и были все основания ожидать, что механически он будет очень прочным. Более того, он был изготовлен из материалов, свойства которых, повидимому, сохраняются неограниченно долго, и был достаточно прост в изготовлении.

Однако, как и следовало ожидать, практическое использование возможностей этого нового электронного прибора встретило на своем пути целый ряд трудностей.

Хотя принцип действия прибора в общем был ясен, поведение его во время работы иногда оказывалось совершенно непонятным, что, естественно, ограничивало возможности его применения, например:

а) При совершенно одинаковой технологии изготовления, отдельные приборы значительно отличались друг от друга — *воспроизводимость была плохой*.

б) Свойства значительной части приборов при их испытании внезапно и по непонятным причинам изменялись с течением времени, а также при изменении температуры, тогда как некоторые из них имели очень устойчивые характеристики — *надежность приборов в работе была также плохой*.

в) Ограниченность теории и несовершенство существующей в то время технологии являлись серьезным тормозом в работе по улучшению качества приборов и их проектированию. Еще нельзя было ставить вопрос об изготовлении транзисторов с заранее заданными характеристиками, которые определялись назначением прибора. Рабочие характеристики были ограничены в отношении усиления шумового числа, диапазона частот и мощности — *возможности проектирования приборов были крайне ограничены*.

Для успешного применения транзисторов в качестве действующих элементов рабочих схем нужно было преодолеть все эти трудности.

¹⁾ Сокращенный перевод. — Прим. ред.

Прежде всего необходимо было понять причины указанных недостатков, полнее разработать теорию действия прибора и усовершенствовать технологию его изготовления в направлении, обеспечивающем возможность получения заданных характеристик.

За последние два года достигнуты значительные успехи в деле преодоления этих трудностей, хотя и сейчас еще мы не можем говорить о полной ликвидации перечисленных выше ограничений. Эти успехи достигнуты благодаря дальнейшему развитию теории, улучшению технологии и улучшению качества исходного материала. Следует указать основные из них:

а) Сделаны первые шаги в применении теории к объяснению и предсказанию электрических характеристик, качество которых, как было установлено, определяется физической структурой и свойствами исходных материалов.

б) Найденны причины и установлен физический смысл некоторых ранее непонятных явлений, связанных с изменением свойств приборов при их испытании. Это позволило улучшить технологические методы изготовления приборов и сделало возможным их контроль. Установлено, что неоднородность исходного материала маскирует наличие зависимости электрических свойств транзисторов даже от его объемных характеристик (например, от удельного сопротивления германия) и тем более от явлений, связанных с процессом их изготовления.

в) В результате получена возможность изготавливать в лабораторных условиях транзисторы с некоторыми заранее заданными характеристиками, лежащими в пределах, определяемых допусками, и с удовлетворительным выходом.

г) Современные транзисторы обладают значительно большей прочностью и более продолжительным сроком службы, нежели старые типы. Получено также некоторое снижение температурной зависимости их электрических характеристик. Однако это не означает, что в настоящее время уже устранены все трудности, связанные с проблемой надежности приборов в работе.

д) Современное состояние транзисторной электроники позволило экспериментально проверить некоторые выводы теории, что естественно помогло установить новые пределы рабочих характеристик точечно-контактных приборов. Еще более важен тот факт, что в настоящее время в лабораторных условиях уже получены новые приборы с $p-n$ -переходами, характеристики которых могут меняться в более широких пределах.

е) Наличие достаточно надежных в работе современных приборов с более широкими пределами их рабочих характеристик позволяет продолжать изучение возможностей их применения в радиотехнических системах и отдельных схемах на более реальной основе. Такое изучение, в свою очередь, будет стимулировать дальнейшую разработку приборов, обладающих новыми, более совершенными характеристиками.

Целью настоящей статьи является краткое описание успехов, достигнутых в последние годы в работе по преодолению упомянутых трудностей, связанных с воспроизводимостью, надежностью в работе и качеством электрических характеристик транзисторов. В связи с тем, что целый ряд различных типов транзисторов находится в стадии усовершенствования, здесь будут описаны электрические характеристики лишь некоторых из них. Это, конечно, не означает, что один какой-нибудь тип транзистора сочетает в себе все свойства этого прибора. Объем статьи не позволяет подробно описать все или даже многие типы транзисторов. Наша задача заключается в кратком обзоре современного состояния транзисторной электроники.

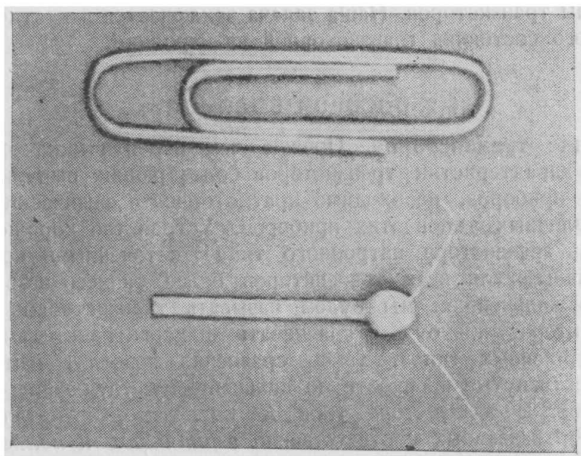
§ 2. ВОСПРОИЗВОДИМОСТЬ

Описание транзисторов. Прежде чем перейти к качественному сравнению характеристик транзисторов более ранних выпусков и современных приборов, необходимо кратко описать физические основы работы различных типов этих приборов. Устройство хорошо известного теперь транзистора патронного типа¹⁾ с точечными контактами было типичным для всех транзисторов более раннего изготовления, которые в дальнейшем мы будем называть транзисторами типа А, и их характеристики будем сравнивать с характеристиками новых, более совершенных типов. Такое сравнение позволит нам оценить успехи, достигнутые в работе по дальнейшему улучшению качества транзисторов. В транзисторе типа А имеются два металлических электрода²⁾, названных соответственно *эмиттером* и *коллектором*; они образуют на верхней поверхности небольшого образца германия *n*-типа два выпрямляющих точечных контакта. Третий электрод находится на нижней поверхности образца и представляет собой обычный омический контакт большой площади. Выпрямляющие свойства точечных контактов обусловлены наличием потенциальных *p* — *n*-барьеров между основной массой образца германия *n*-типа и небольшими включениями *p*-типа под коллектором и эмиттером. Когда к контакту коллектора в обратном направлении (минус на коллектор, плюс на нижний омический электрод) прикладывается относительно большое напряжение смещения, обратный ток или поток электронов из коллектора в образец германия — благодаря относительно большому значению полного обратного сопротивления барьера коллектора — будет незначительным. Если же к контакту эмиттера приложить небольшое напряжение смещения (десять вольта) в прямом направлении (плюс на эмиттер, минус на нижний омический электрод), то будет иметь место поток дырок из области включений *p*-типа под эмиттером в образец германия *n*-типа. Эти дырки под действием поля, обусло-

1) См. фиг. 19 основного текста, стр. 55. — *Прим. ред.*

2) См. фиг. 20 основного текста, стр. 56. — *Прим. ред.*

вленного начальным электронным током коллектора, будут двигаться в направлении коллектора и увеличивать тем самым значение тока в коллекторной цепи. Вследствие своего положительного заряда дырки могут понижать потенциальный барьер для потока электронов от коллектора к основному электроду. Это приведет к увеличению тока в цепи коллектора. На каждую дырку, проникающую через потенциальный барьер, будет приходиться несколько электронов, протекающих в цепи коллектора. Таким образом, изменение тока коллектора



Ф и г. 1. „Бусинковый“ транзистор (тип М 1689).

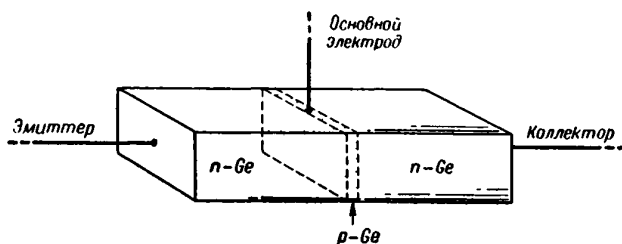
будет отличаться от изменения тока эмиттера. Отношение изменения тока коллектора к изменению тока эмиттера (при постоянном напряжении коллектора) называется *усилением по току*. В транзисторах с точечными контактами коэффициент усиления по току может быть больше единицы. Так как ток коллектора проходит через большое сопротивление, а ток эмиттера через малое, получается кроме того и усиление по напряжению.

Некоторые из новых транзисторов по своей физической схеме очень похожи на транзисторы типа А. Однако электрические характеристики новых транзисторов, как будет показано ниже, значительно улучшены не только в отношении воспроизводимости и надежности в работе, но и в отношении их пределов.

Для использования в миниатюрных пакетных схемах транзистор с точечными контактами был значительно уменьшен за счет того, что в нем были оставлены только самые необходимые части. На фиг. 1 показана фотография так называемого „бусинкового“ транзистора (для сравнения размеров на фигуре показана также фотография

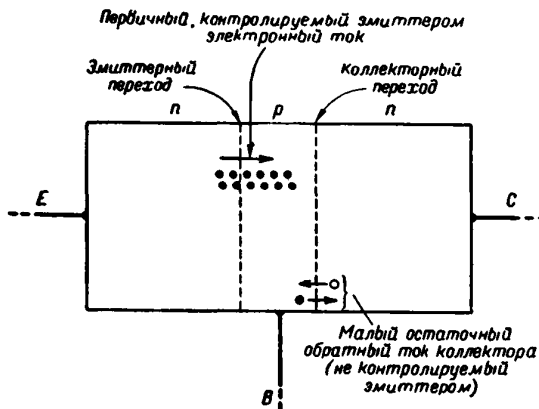
канцелярской скрепки), в форме которого сейчас изготавливаются некоторые типы улучшенных транзисторов.

Недавно Шокли, Спаркс, Тил¹⁾, а также Уоллэс и Питенполь [2] описали новый тип транзистора с $n-p-n$ -переходами. Схема такого



Фиг. 2. Транзистор с $n-p-n$ -переходами.

транзистора показана на фиг. 2. В средней части монокристалла германия n -типа имеется тонкий искусственно созданный слой германия p -типа (две области n -типа и одна область p -типа являются частями одного монокристалла). Все три электрода (эмиттер, основной



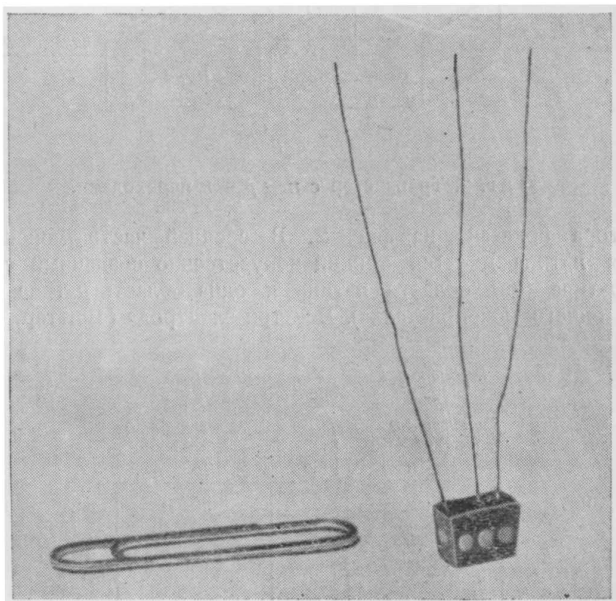
Фиг. 3. Схема транзистора с $n-p-n$ -переходами.

электрод, коллектор) являются в этом случае обычными невыпрямляющими омическими контактами. Если не считать различия в порядке изменения типа проводимости (в случае точечно-контактных транзисторов это

¹⁾ См. статью этих авторов, помещенную в настоящей книге, стр. 643.—
Прим. ред.

проводимость $p-n-p$ -типа, а в нашем случае $n-p-n$ -типа), принцип действия этих двух приборов по существу одинаков.

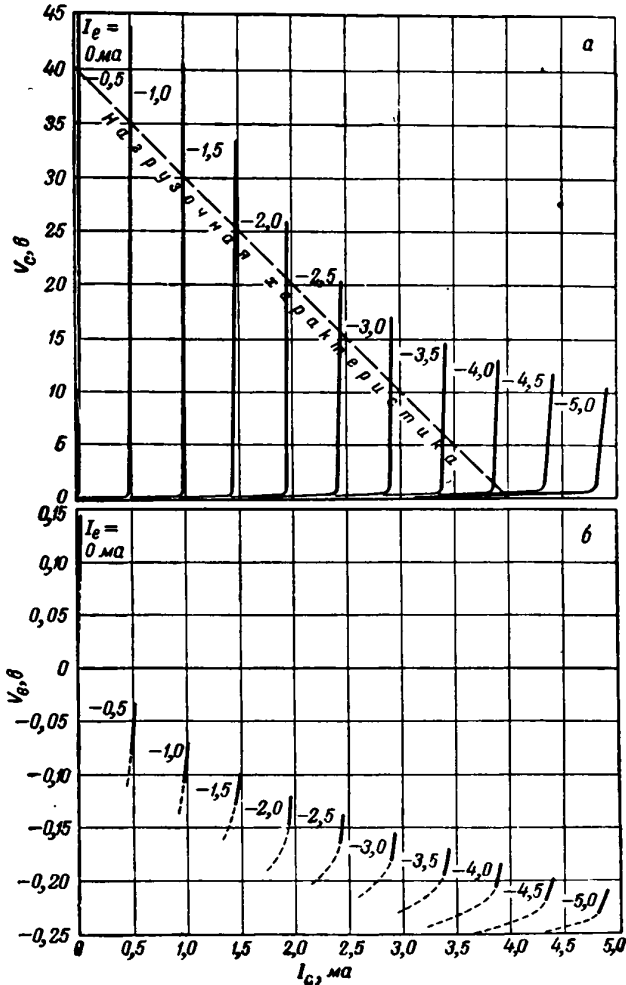
Когда к контакту коллектора (фиг. 3) прикладывается напряжение смещения в обратном направлении, т. е. электрод C заряжается положительно относительно электрода B , то лишь небольшое количество электронов и дырок диффундирует сквозь барьер коллектора и обратный ток незначителен. Так как полное обратное сопротивление



Фиг. 4. Внешний вид улучшенной модели транзистора с $n-p-n$ -переходами (тип М 1752).

потенциального барьера такого контакта во много раз больше, чем в случае точечных контактов, величина этого тока значительно меньше величины обратного тока точечно-контактных приборов и в меньшей степени зависит от напряжения коллектора. Если теперь к контакту эмиттера приложить небольшое напряжение смещения (несколько десятых вольта) в прямом направлении (минус на эмиттер E и плюс на основной электрод B), то сравнительно большое число электронов будет диффундировать из богатой электронами области (n -типа) эмиттера в область (p -типа) основного электрода. Если при этом область основного электрода настолько тонка, что попадающие в нее электроны не рекомбинируют (в объеме или на поверхности) с дырками, то практически

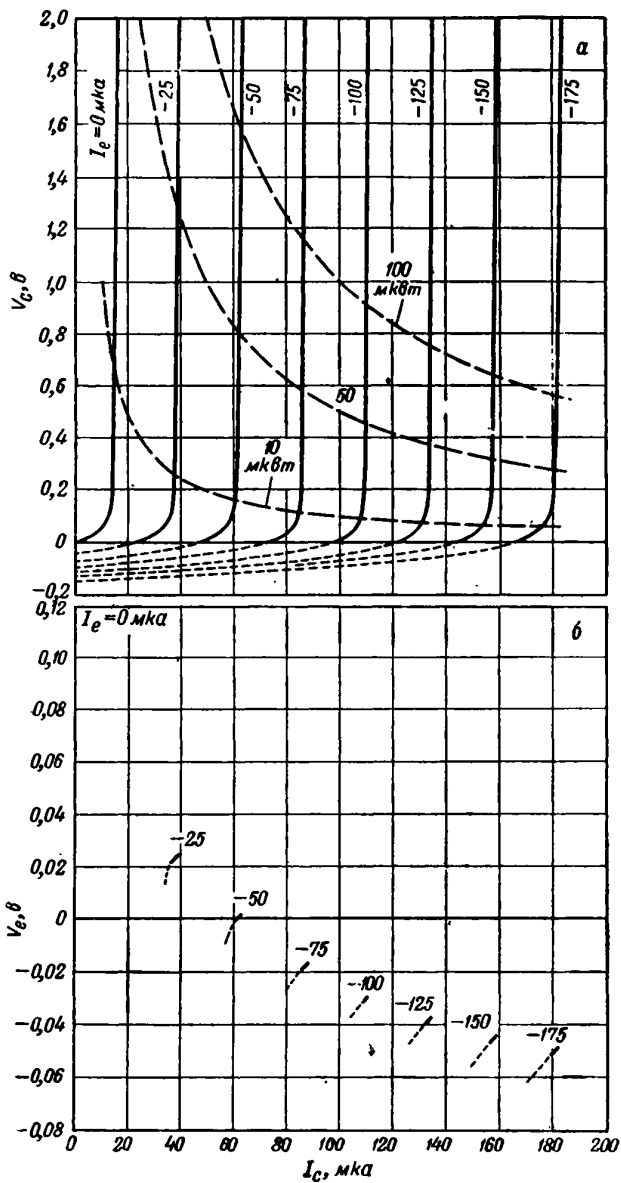
все эти электроны будут диффундировать через потенциальный барьер коллектора. Далее под действием поля они будут двигаться через коллекторный контакт и увеличивать тем самым ток коллектора.



Ф и г. 5. Статические характеристики транзистора М 1752 с $n-p-n$ -переходами.

a — для коллектора; b — для эмиттера.

Так же как и раньше, ввиду того что из области эмиттера в область основного электрода электроны проходят через малое сопро-



Ф и г. 6. Статические характеристики транзистора М 1752 для области более низких напряжений.

а — для коллектора; б — для эмиттера.

тивление, а из основного электрода в область коллектора — через очень большое сопротивление объемных $p-n$ -переходов, будет иметь место очень высокое усиление по напряжению. Следует отметить, что транзисторы такого типа не дают большого усиления по току, максимальное его значение не превышает единицы. Однако, в связи с тем что контакты этого типа выпрямляют значительно лучше, чем точечные, очень большая величина отношения полного обратного сопротивления коллектора к полному прямому сопротивлению эмиттера с избытком компенсирует это более низкое усиление по току по сравнению с точечно-контактными приборами. Таким образом, по мощности такие транзисторы могут иметь значительно большее усиление на каскад, чем точечно-контактные [1, 2]. На фиг. 4 показана фотография улучшенной модели транзистора этого типа, названной М 1752.

На фиг. 5,а приведено семейство статических характеристик коллектора этой модели. Сравнение их с характеристиками точечно-контактных приборов показывает, что транзисторы М 1752 имеют значительно более высокое полное обратное сопротивление барьера коллектора (которое, как видно из графика, почти не зависит от напряжения коллектора) и соответственно меньший обратный ток в коллекторной цепи при токе эмиттера, равном нулю.

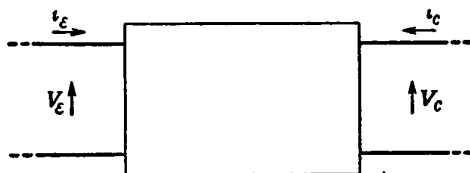
На фиг. 6,а приводятся те же характеристики для области более низких напряжений. Они, как видно из графика, сохраняют почти идеальный прямолинейный характер и одинаковое расстояние между собой вплоть до напряжений порядка 0,1 в и токов в несколько *мкА*. Таким образом, линейное усиление в режиме класса А возможно и при малых мощностях на коллекторе — порядка нескольких *мкВт*. Для справки на том же графике (фиг. 6,а) пунктирными линиями показаны кривые, соответствующие рассеянию постоянной мощности на коллекторе, равной 10, 50 и 100 *мкВт*.

Воспроизводимость линейных характеристик. Успехи на пути повышения воспроизводимости тех свойств транзисторов, которые существенны для усиления малых сигналов, можно характеризовать средней величиной линейных импедансов разомкнутого транзистора, а также величиной отклонения от этого среднего значения. Выражения для этих импедансов получены Райдером и Кирхером [2]. Подобная характеристика воспроизводимости предполагает, естественно, существование налаженного статистического контроля процессов, ведущих к улучшению транзисторов со сравнительно нормальным распределением свойств, для которого могут быть даны средние значения и величина максимально допускаемых отклонений от них. Можно отметить, что в настоящее время такая возможность существует для большинства транзисторов, находящихся в стадии дальнейшего усовершенствования.

При разработке старых приборов типа А такой контроль не применялся. Однако в дальнейшем будет дана оценка воспроизводимости

их характеристик. Для ознакомления с обозначениями и терминологией рассмотрим кратко метод определения линейных характеристик транзисторов.

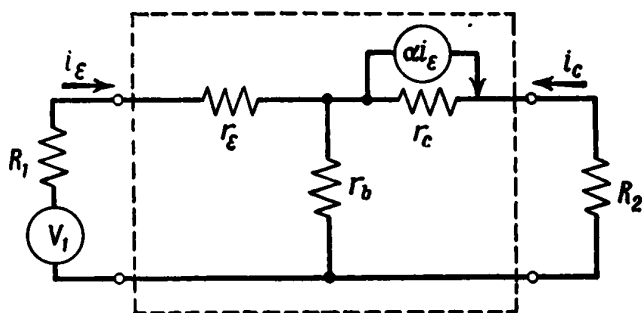
На фиг. 7 показана общепринятая схема четырехполюсника, изображающая транзистор; входные клеммы четырехполюсника соответствуют эмиттеру и основному электроду транзистора, а выходные клеммы — коллектору и основному электроду.



Ф и г. 7. Схема четырехполюсника.

$$V_a = i_a z_{11} + i_c z_{12}; \quad V_c = i_b z_{21} + i_c z_{22}.$$

Линейная связь между дифференциальными токами и напряжениями коллектора и эмиттера на достаточно малых участках статических характеристик может быть представлена двумя линейными уравнениями [2]. Коэффициенты этих уравнений являются обобщенными



Ф и г. 8. Эквивалентный контур.

$$\begin{aligned} \bar{r}_a &= 250 \text{ ом} & \bar{r}_c &= 20\,000 \text{ ом} \\ \bar{r}_b &= 250 \text{ ом} & a &= 2 \end{aligned}$$

полными сопротивлениями транзистора или наклонами соответствующих статических характеристик при фиксированных значениях постоянного тока. Эти уравнения можно написать для любой из большого числа эквивалентных схем, одна из которых, показанная на фиг. 8, повидимому, наиболее употребительна,

В этой схеме r_e по своей величине почти совпадает с прямым полным сопротивлением эмиттера, а r_c — с обратным полным сопротивлением коллектора; r_b — полное сопротивление обратной связи образца германия, α — коэффициент усиления по току, характеризующий процессы как накопления, так и умножения носителей заряда, если таковые имеются.

Численное значение α почти совпадает с величиной отмеченного выше коэффициента усиления по току на коллекторе. Средние значения этих параметров эквивалентного контура для транзистора типа А приведены на рисунке. В табл. 1 даны пределы их изменения по состоянию на сентябрь 1949 г. Там же приведены пределы отклонений [4] параметров и для новых точечно-контактных транзисторов, находящихся в настоящее время в стадии дальнейшего совершенствования.

Пределы изменения параметров транзисторов типа А даны по отношению к их средним значениям (см. фиг. 8). Приведенные же пределы отклонений параметров от их средних значений относятся к большинству современных типов точечно-контактных приборов, причем сами эти средние значения зависят от типа рассматриваемого транзистора. В табл. 2 даны средние значения параметров транзистора типа М 1729, применяемого в виде усилителя, который имеет очень большое сходство с более старыми приборами типа А. Для сравнения в этой же таблице приводятся некоторые типичные значения параметров транзистора типа М1752 с $n-p-n$ -переходами, который находится сейчас в стадии совершенствования и поэтому не имеет еще общепринятого конструктивного оформления.

Райдер и Кирхер [2] показали, что транзисторы с заземленным основанием могут быть неустойчивыми по отношению к короткому замыканию, если $\alpha > 1$ и r_b слишком велико, так как в этом случае r_b является параметром положительной обратной связи.

На фиг. 9 приведена кривая устойчивости схемы для номинальных значений r_e и r_c , равных 700 и 20 000 Ω соответственно. Если значения

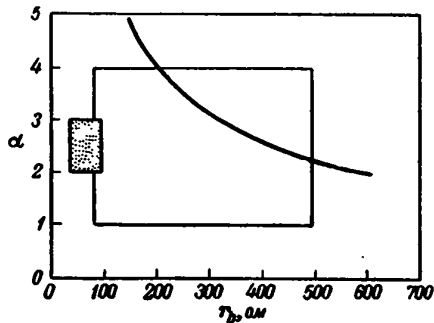
Таблица 1
ПРЕДЕЛЫ ИЗМЕНЕНИЯ ПАРАМЕТРОВ
ЭКВИВАЛЕНТНОГО КОНТУРА ДЛЯ
ТРАНЗИСТОРА ТИПА А

Параметр	Пределы изменения параметров	
	по состоянию на сентябрь 1949 г.	по состоянию на январь 1952 г., %
α	4:1	± 20
r_c	7:1	± 30
r_e	3:1	± 20
r_b	7:1	± 25

Таблица 2
СРАВНЕНИЕ ХАРАКТЕРИСТИК
ТРАНЗИСТОРОВ М1729 И М1752

Параметры	М1729	М1752
r_e	120	25
r_b	75	250
r_c	15 000	$5 \cdot 10^6$
α	2,5	0,95

α и r_b транзисторов соответствуют точкам, расположенным выше этой кривой, то схема с такими транзисторами будет неустойчивой, т. е. величина параметров α и r_b будет изменяться при коротком замыкании. Наоборот, при значениях α и r_b , соответствующих точкам, находящимся ниже этой кривой, схема будет устойчивой при любой оконечной нагрузке. Незаштрихованный прямоугольник на этой фигуре ограничивает значения α и r_b , характерные для транзисторов типа А по состоянию на сентябрь 1949 г. Расположение этого прямоугольника показывает, что из большого числа таких приборов приблизительно

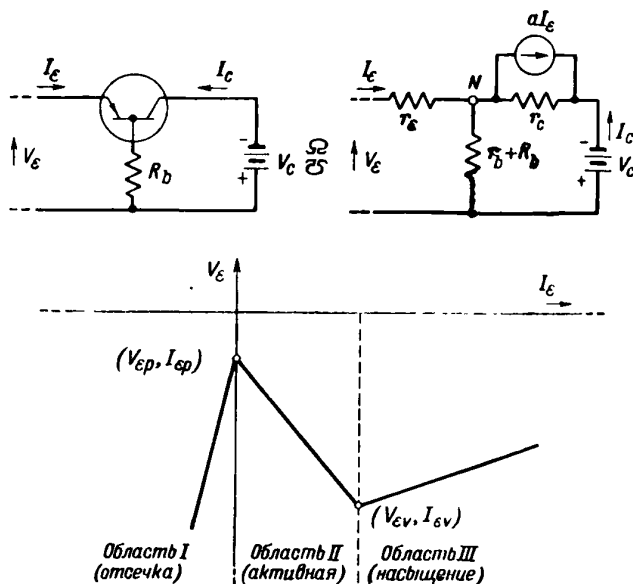


Фиг. 9. Кривая устойчивости контура и диапазон изменений α и r_b .

только половина могла бы обеспечить устойчивость контура при коротком его замыкании. Маленький заштрихованный прямоугольник ограничивает значения α и r_b , которые в настоящее время имеет транзистор М1729, находящийся еще в стадии усовершенствования. Из расположения этого прямоугольника видно, что эти приборы обладают значительно меньшим разбросом значений параметров α и r_b и что каждый из них при коротком замыкании обеспечивает устойчивость контура.

Интересно отметить, что примерно таким же разбросом значений параметров, порядка $\pm 20 - \pm 25\%$, обладают и современные электронные лампы, производство которых хорошо налажено. Приведенные здесь данные по воспроизводимости значений линейных параметров эквивалентного контура имеют силу практически для всех классов точечно-контактных приборов, которые сейчас конструируются для использования в установках по передаче незатухающих волн. Пока еще слишком рано утверждать, что такое же положение имеет место и в случае транзисторов с $n-p-n$ -переходами. Однако есть основания полагать, что после ряда усовершенствований подобные результаты будут получены и с приборами этого типа.

Воспроизводимость характеристик для сигналов большой мощности, применяемых в импульсных схемах. Хорошо известно, что электронные приборы, применяемые в установках, работающих на сигналах большой мощности, особенно в коммутирующих и счетно-решающих устройствах, должны обладать характеристиками, управляемыми в очень широкой области — от запирания до насыщения. Транзисторы с такими характеристиками могут найти себе самое широкое



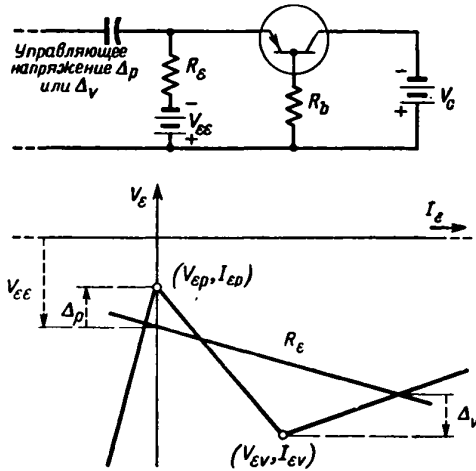
Фиг. 10. Транзистор в схеме с заземленным эмиттером; схема включения, эквивалентная схема и характеристики.

применение в различных импульсных схемах, например таких, как двоичные счетные устройства, накопительные счетчики, импульсные усилители с обратной связью, усилители задержанных импульсов, усилители строб-импульсов и импульсные генераторы. Объем этой статьи не позволяет подробно описать все технические особенности этой проблемы. Такое описание, несомненно, найдет себе место в ряде будущих работ, посвященных более узким вопросам. Однако краткое описание некоторых наиболее важных импульсных характеристик, конечно, необходимо.

Общей особенностью практически всех исследованных до настоящего времени импульсных схем является то, что благодаря наличию усиления по току транзистор может давать различные типы двузначных

характеристик отрицательного сопротивления для любой пары или для всех пар клемм.

Типичная простая схема и соответствующая характеристика для клемм эмиттер—земля показаны на фиг. 10. Чтобы сделать схему нестабильной, последовательно с основанием включается достаточно большое сопротивление. В области I, где эмиттер заряжен отрицательно, входное сопротивление по существу является обратной характеристикой эмиттера, как простого диода. В области II, когда эмиттер



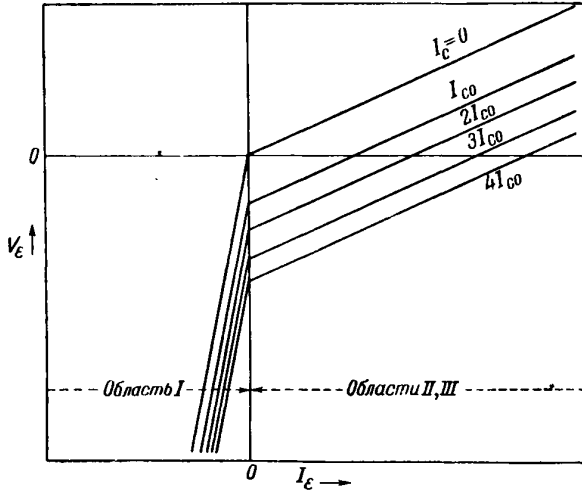
Фиг. 11. Схема удвоенной стабильности и характеристика, указывающие на величину необходимого управляющего напряжения.

становится положительным, коэффициент усиления по току быстро возрастает до значения, большего единицы. Если R_b достаточно велико и коэффициент усиления по току больше единицы, то напряжение эмиттер—земля начнет уменьшаться вследствие большого увеличения тока коллектора, доводя напряжение узла N до отрицательного значения много быстрее, чем это достигается уменьшением тока эмиттера, протекающего через сопротивление r_e . Эта точка перехода называется *пиковой точкой*.

Далее, если значение $\alpha(r_b + R_b)$ достаточно велико в указанном выше смысле, то входное сопротивление в области II может быть отрицательным. Когда напряжение внутреннего узла приближается к значению напряжения на клемме коллектора, достигается „точка минимума“. В этой точке благодаря притоку дырок полное сопротивление коллектора уменьшается до минимального значения, и за этой точкой оно по существу равно нулю; транзистор, как говорят, нахо-

дится в *состоянии насыщения*. От этой точки входное полное сопротивление снова становится положительным и определяется почти всецело сопротивлениями основного электрода и эмиттера.

Присоединяя клеммы эмиттер—основной электрод различными способами к цепям смещения, состоящим из сопротивлений и емкостей, можно заставить такой четырехполюсник выполнять разнообразные



Фиг. 12. Идеализированные характеристики эмиттера.

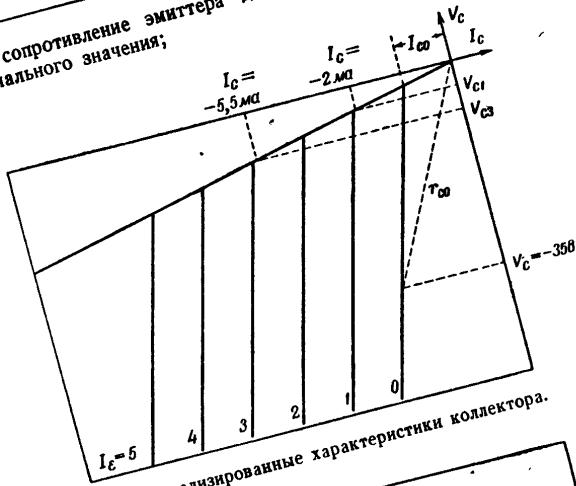
функции, требующие моноустойчивых, неустойчивых и биустойчивых характеристик. При таких условиях ток эмиттера и соответственно ток коллектора все время будут изменяться между значениями, соответствующими запиранию и насыщению.

Для примера на фиг. 11 указаны значения смещения эмиттера и сопротивления нагрузки, при которых имеются три возможных равновесных значения тока и напряжения эмиттера.

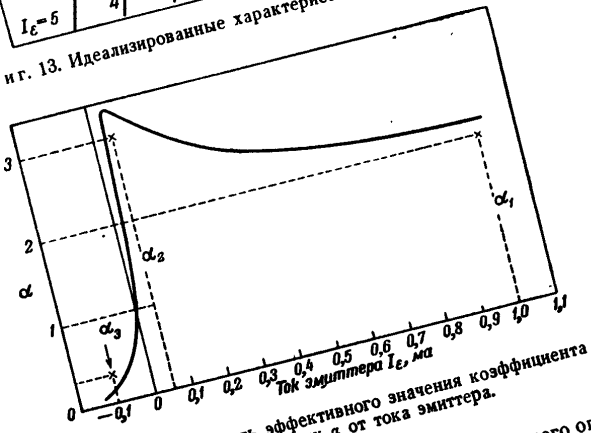
Можно показать, что две точки пересечения в областях I и III соответствуют устойчивым состояниям, тогда как две точки пересечения в области II — неустойчивы. Следовательно, если устойчивое равновесие вначале имеет место в области I, достаточно небольшого положительного импульса Δ_p , приложенного к эмиттеру, чтобы перейти от стабильной точки I к стабильной точке II, и наоборот, малый отрицательный импульс — Δ_n будет соответствовать при этом переходу от большого тока к меньшему.

Конструктор, имеющий дело с транзисторными схемами, прежде всего заинтересован в воспроизводимости (с различными транзисторами данного типа) следующих параметров характеристики:

а) полное сопротивление эмиттера должно быть больше определенного минимального значения;



Ф и г. 13. Идеализированные характеристики коллектора.



Ф и г. 14. Зависимость эффективного значения коэффициента усиления по току α от тока эмиттера.

б) пиковая точка V_{ep} должна лежать ниже некоторого определенного максимума;
 в) значение отрицательного сопротивления должно быть больше определенного минимума;

г) точка минимума $V_{\nu\beta}$, $I_{\epsilon\beta}$, должна лежать выше определенного минимума;

д) крутизна в области III должна быть меньше определенного максимума, чтобы можно было управлять ее величиной внешними способами.

Можно показать, что для целого ряда схем эти условия могут быть выполнены путем точного установления определенных максимальных и минимальных границ на статических характеристиках.

На фиг. 12 показана идеализированная серия входных или эмиттерных характеристик, из которой видно, что точным определением минимального значения обратного сопротивления в области I выполняется условие а). Точным определением максимальной крутизны в областях II и III выполняется условие д).

Таблица 3

ХАРАКТЕРИСТИКИ „БУСИНКОВОГО“ ТРАНЗИСТОРА M1689

Испытываемый параметр	Условия	Минимальное значение	Максимальное значение
r_{c0} — сопротивление отключенного коллектора при постоянном токе	$V_c = -35$ в $I_e = 0$ ма	17 500 ом	—
V_{c1} — напряжение на коллекторе	$I_c = -2$ ма $I_e = 1$ ма	—	-3 в
V_{c3} — напряжение на коллекторе	$I_c = -5,5$ ма $I_e = 3$ ма	—	-4 в
Сопротивление выключенного эмиттера	$V_e = -10$ в	50 000 ом	—
R_{11} — сопротивление включенного эмиттера	$V_c = -10$ в $I_e = 1$ ма	—	800 ом
α_1	$V_c = -30$ в $I_e = 1,0$ ма	1,5	—
α_2	$V_c = -30$ в $I_e = +0,05$ ма	2,0	—
α_3	$V_c = -30$ в $I_e = -0,1$ ма	—	0,3
R_{12} — сопротивление обратной связи для разомкнутой цепи	$V_c = -10$ в $I_e = +1$ ма	—	500 ом
R_{21} — сопротивление разомкнутой цепи (в прямом направлении)	$V_c = -10$ в $I_e = +1$ ма	15 000 ом	—
R_{22} — выходное сопротивление разомкнутой цепи	$V_c = -10$ в $I_e = +1$ ма	10 000 ом	—

Обратимся теперь к идеализированному семейству характеристик коллектора на фиг. 13. Точным определением максимального значения $V_{св}$ можно обеспечить выполнение условия г), а точным определением минимального значения $r_{с0}$ — условия б).

Наконец, требуя, чтобы коэффициент усиления по току при возрастании I_c (фиг. 14) изменялся от незначительной величины при малых отрицательных значениях I_c до значения много большего единицы при небольших положительных величинах I_c и оставался значительно большим единицы при больших значениях I_c , можно удовлетворить условия б) и в).

В табл. 3 даны максимальные и минимальные значения некоторых параметров коммутирующего „бусинкового“ транзистора М 1689, находящегося сейчас в стадии дальнейшего усовершенствования. Применяя в настоящее время эти приборы в различных более сложных импульсных схемах, получают схемы с воспроизводимостью порядка ± 2 дб.

§ 3. НАДЕЖНОСТЬ

Срок службы. Достаточно точного критерия оценки надежности в работе для электронных ламп еще не существует. То же можно сказать и в отношении транзисторов. Однако в той мере, в какой эти величины теперь могут быть определены, в табл. 4 дано представление о состоянии вопроса теперь и в сентябре 1949 г.

Статистическая оценка половины срока службы приборов в лучшем случае произвольна, и необходима экстраполяция кривых срока службы в предположении, что определенный закон старения будет иметь силу в течение более долгого времени ¹⁾.

В сентябре 1949 г. время испытания приборов типа А на срок службы равнялось приблизительно 4 000 часов. Предполагая, что усиление по току падает во времени по экспоненциальному закону, нельзя было на основании такого испытания достаточно точно оценить половину срока службы.

Сейчас время испытания этих же самых приборов типа А составляет уже приблизительно 20 000 часов. Это позволяет более надежно оценить срок их службы; в настоящее время половину срока службы можно оценить несколько больше, чем 70 000 часов. Необходимо подчеркнуть, однако, что эти приборы типа А были сделаны более чем два года тому назад — по менее совершенной технологии и из худших материалов. Надо считать, что эти транзисторы, значительно улучшенные в настоящее время и изготавливаемые из более качественных материалов и по более совершенной технологии, окажутся

¹⁾ Оценка срока службы зависит, конечно, от определения момента выхода из строя. Во время проводившихся опытов транзисторы работали как усилители класса А. Было принято, что транзистор выбывает из строя, когда его усиление в режиме класса А уменьшается больше чем на 3 дб по сравнению с его начальным значением. — *Прим. авт.*

Таблица 4

ДЛИТЕЛЬНОСТЬ РАБОТЫ И МЕХАНИЧЕСКАЯ ПРОЧНОСТЬ ТРАНЗИСТОРОВ

Характеристика надежности	По состоянию на сентябрь 1949 г.	По состоянию на январь 1952 г.
Средний срок службы	до 10 000 часов	> 70 000 часов
Эквивалентный температурный коэффициент сопротивления коллектора r_c	$-1\%/^{\circ}\text{C}$	$-1/4\%/^{\circ}\text{C}$
Удар	?	> 20 000 г
Вибрации	?	При 20—5000 гц можно не принимать в расчет при ускорениях до 100 г

приборами еще более высокого качества. Хотя эти новые данные перспективны, тем не менее слишком рано экстраполировать их очень далеко.

Температурная зависимость. Транзисторы, подобно другим полупроводниковым приборам, более чувствительны к температурным изменениям, чем электронные лампы. Из всех параметров линейного эквивалентного контура полное сопротивление коллектора r_c и усиление по току α являются самыми чувствительными. В интервале температур от -40 до $+80^{\circ}\text{C}$ другие параметры значительно менее чувствительны. Для транзисторов типа А температурная зависимость r_c и α показана на фиг. 15.

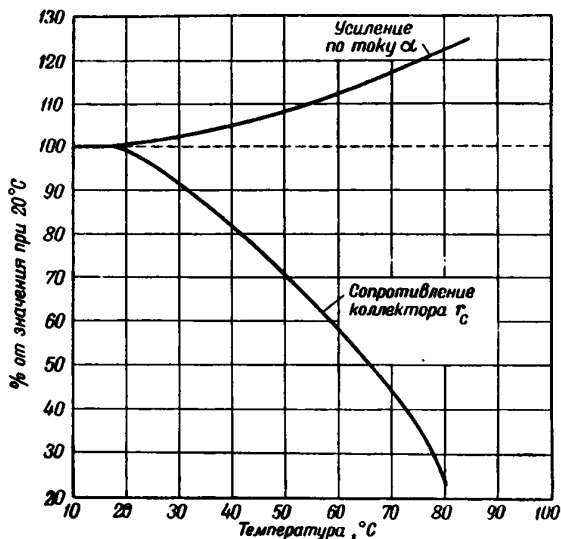
Для сравнения (см. табл. 4) по этой зависимости, несмотря на ее криволинейный характер, был вычислен средний температурный коэффициент r_c . Он оказался приблизительно равным 1% на 1°C .

Таким образом, при повышении температуры от комнатной до 80°C , r_c первых транзисторов типа А уменьшается приблизительно на 20—30%; в то же самое время α увеличивается на 20—30%. В настоящее время для большинства типов точечно-контактных приборов эта температурная зависимость r_c уменьшена приблизительно в 4 раза, причем температурная зависимость коэффициента усиления по току осталась примерно такой же.

На фиг. 16 показана температурная зависимость r_c и α для разрабатываемого сейчас транзистора M1729. Как и раньше для вычисления среднего температурного коэффициента, эта зависимость считалась приблизительно прямолинейной. В линейных схемах, например таких, как усилитель с заземленным основным электродом, усиление мощности по классу А приблизительно пропорционально $\alpha^2 r_c$; следовательно, усиление такого усилителя будет оставаться по существу постоянным в пределах 1—2 дб, в интервале температур от -40 до $+80^{\circ}\text{C}$. Для применения в импульсных схемах и для смещений постоянного тока в точечно-контактных транзисторах имеет значение тот факт, что постоянный ток коллектора (при фиксированных значениях тока

эмиттера и напряжения коллектора) изменяется приблизительно с такой же скоростью, как r_c (полное сопротивление коллектора при небольших сигналах). Подобное уменьшение температурной зависимости получено также и для коммутирующих транзисторов.

На фиг. 17 показана температурная зависимость некоторых параметров (см. табл. 3) коммутирующего бусинкового транзистора М1689.



Фиг. 15. Зависимость сопротивления коллектора и коэффициента усиления по току α от температуры.

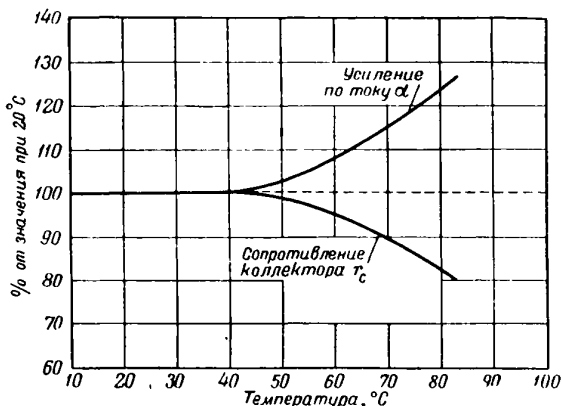
Результаты изучения работы коммутирующих транзисторов дают возможность полагать, что в большинстве случаев эти приборы могут надежно работать при температуре $+70^\circ\text{C}$, а в некоторых случаях и при $+80^\circ\text{C}$.

Температурная зависимость транзисторов с $n-p-n$ -переходами еще мало изучена, так как прибор находится в значительно более ранней стадии своего развития. Предварительные данные указывают на меньшую температурную зависимость таких параметров как α и r_c для небольших сигналов. С другой стороны, изменение постоянного тока, особенно I_{c0} , от температуры во много раз больше, порядка 10% на 1°C ¹⁾. Утешительно здесь лишь то, что обычно I_{c0} значительно меньше действительного действующего значения I_c .

В заключение можно сказать, что, несмотря на достигнутые значительные успехи в изучении и уменьшении температурной зависи-

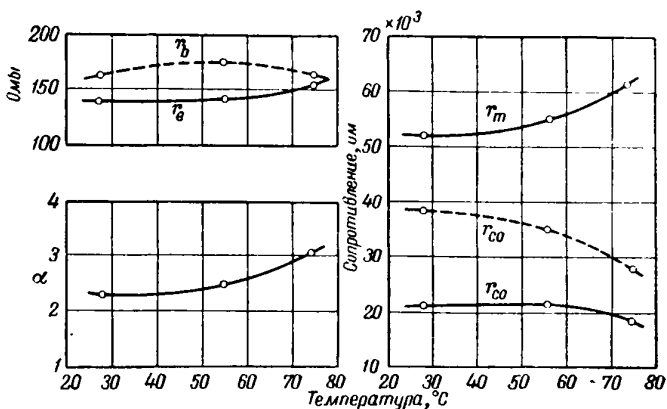
¹⁾ I_{c0} — ток коллектора при токе эмиттера, равном нулю. — Прим. авт.

мости параметров транзисторов, все же до сих пор нельзя считать ограничения, накладываемые температурной зависимостью, полностью



Ф и г. 16. Зависимость сопротивления коллектора и коэффициента усиления по току α от температуры для транзистора типа М 1729.

преодоленными. Для лучшей работы прибора решение этой проблемы должно быть доведено до конца.



Ф и г. 17. Зависимость параметров транзистора М 1689 от температуры.

Удар и вибрации. В отношении механической прочности точечно-контактные транзисторы были испытаны на удар вплоть до 20 000 g без заметного изменения их электрических характеристик. Испытание

точечно-контактных транзисторов и транзисторов с $n-p-n$ -переходами на вибрацию с частотой от 20 до 5 000 $гц$ при ускорениях до 100 g не привело к заметным изменениям каких-нибудь электрических характеристик транзистора, т. е. если такие изменения и существуют, то они находятся ниже естественного уровня шумов. Вибрационные испытания вплоть до ускорений в 1 000 g на нескольких частотах в звуковой области также не обнаружили заметных изменений характеристик транзистора.

§ 4. РАЗМЕРЫ И МОЩНОСТЬ

Размеры. Благодаря своим малым размерам транзисторы находят в настоящее время все новые области применения, заведомо недоступные другим электронным приборам. Патронный транзистор типа А (табл. 5) имеет объем $\sim 0,3 \text{ см}^3$, в то время как объем сверхминиатюрной лампы равен $\sim 2 \text{ см}^3$, а миниатюрной $\sim 15 \text{ см}^3$.

Таблица 5

СРАВНЕНИЕ ХАРАКТЕРИСТИК ТРАНЗИСТОРА ТИПА А (1949 г.)
И УСОВЕРШЕНСТВОВАННЫХ ТИПОВ (1952 г.)

Характеристика	Тип А (по состоянию на сентябрь 1949 г.)	Усовершенствованные типы (по состоянию на январь 1952 г.)
Объем	$\sim 0,3 \text{ см}^3$	$\sim 8 \text{ мм}^3$ М1689, точно-контактный триод
		$\sim 30 \text{ мм}^3$ М1752, с $n-p-n$ -переходами
Минимальное напряжение на коллекторе при работе в режиме класса А	30 в	2 в М1768, М1734, точно-контактные триоды
		0,2 в М1752, с $n-p-n$ -переходами
Минимальная мощность, рассеиваемая на коллекторе при работе в режиме класса А	50 мвт	2 мвт М1768, точно-контактный триод
		10 мквт М1752, с $n-p-n$ -переходами
К. п. д. при работе в режиме класса А	20%	35% М1768, М1729, точно-контактные триоды
		49% М1752, с $n-p-n$ -переходами

Современный точечно-контактный бусинковый транзистор М1689, имеющий по существу такие же электрические характеристики, как и коммутирующий транзистор патронного типа М1698¹⁾, занимает объем, равный только 8 мм³. Бусинковый транзистор с $n-p-n$ -переходами имеет объем, приблизительно равный 30 мм³, который в случае необходимости можно еще уменьшить.

Дальнейшее сокращение размеров приборов будет происходить за счет уменьшения объема их второстепенных деталей. Нужно отметить, что дальнейшее сокращение объема этих деталей может быть обеспечено за счет понижения рабочего напряжения, мощности и температуры.

Рабочая мощность. Большим преимуществом транзистора является отсутствие необходимости в подогреве катода. Кроме того, за последние два года были сделаны значительные успехи в понижении необходимых для практической работы значений напряжения коллектора и мощности.

Рассмотрим минимальное напряжение коллектора, при котором усиление небольших сигналов в режиме класса А находится еще в пределах от 3 до 6 дБ от его полного значения.

В сентябре 1949 г. транзистор типа А обеспечивал необходимое усиление при напряжении коллектора, равном 30 в. В настоящее время некоторые точечно-контактные приборы (М1768 и М1734) хорошо работают при напряжениях коллектора от 2 до 6 в, даже в случаях достаточно высоких частот. Транзистор М1752 с $n-p-n$ -переходами дает нужное усиление при напряжениях коллектора от 0,2 до 1 в. При этих условиях минимальная мощность, рассеиваемая на коллекторе, необходимая для получения нужного усиления, может составлять только 2—10 мвт для точечно-контактных приборов и от 10 до 100 мквт для транзисторов с $n-p-n$ -переходами²⁾. В настоящее время к. п. д. в режиме усиления класса А уже повышен до 35% для точечно-контактных приборов и до 49% (максимально возможное значение к. п. д. в этом случае равно 50%) для транзисторов с $n-p-n$ -переходами; к. п. д. транзисторов, работающих в режимах В и С, близки к их теоретическим предельным значениям.

Характеристики транзисторов. Точные технические условия на электрические характеристики транзисторов, естественно, зависят от типа транзистора и его назначения. Следует отметить, что в настоящее время эти приборы так многочисленны, что уже необходима их классификация по типам.

¹⁾ Транзистор М1698 патронного типа с точечными контактами, годный по своим электрическим характеристикам для применения в импульсных и коммутирующих схемах. В настоящее время прибор находит себе применение в лабораториях, разрабатывающих новые схемы и в тех случаях, когда нет необходимости в особо маленьких размерах. — *Прим. авт.*

²⁾ В некоторых специальных случаях практическая работа прибора может быть обеспечена при мощностях от 0,1 до 1 мквт. — *Прим. авт.*

В этой статье мы можем отметить только наиболее важные характеристики некоторых из них. Выше было указано, что ни один тип транзистора не объединяет в себе всех хороших качеств этого нового прибора в большей степени, чем электронные лампы.

Табл. 6 дает возможность судить об успехах, сделанных за последние два года в работе по улучшению отдельных, наиболее важных характеристик некоторых точечно-контактных приборов и транзисторов с $n-p-n$ -переходами. Здесь, как и раньше, характеристики транзисторов типа А даны по состоянию на сентябрь 1949 г.

Для некоторых применений, связанных с коммутацией и передачей сигнала, нужны транзисторы, имеющие большой коэффициент усиления по току. Лабораторные модели приборов, занимающие по своей конструкции среднее положение между точечно-контактными транзисторами и транзисторами с $n-p-n$ -переходами, дают усиление по току до 50 *дб*.

Для применений, связанных с передачей сигнала, усиление по току на каскад у некоторых типов точечно-контактных приборов (M1768, M1729) увеличено уже до 20—24 *дб*, а у транзистора M1752 с $n-p-n$ -переходами — до 45—50 *дб*.

Для применений, требующих малого уровня шумов и высокой чувствительности, шумовое число точечно-контактных приборов уменьшено до 40—45 *дб*, а транзистора M1752 с $n-p-n$ -переходами — до 10—20 *дб*. При этом необходимо помнить, что шумовые числа определяются при 1000 *гц* и что их величина изменяется обратно пропорционально частоте (при изменении частоты в 10 раз шумовое число изменяется приблизительно на 11 *дб*).

Для применений, связанных с видеосигналами, промежуточной частотой и переключениями большой скорости, значительно улучшена частотная характеристика. Точечно-контактный транзистор M1729, применяемый в видеоусилителях, вплоть до 7 *мгц*, дает усиление около 18—20 *дб* на каскад. Точечно-контактный прибор M1734 находится сейчас в стадии дальнейшей разработки для применения в высоко-частотных генераторах и для коммутации микросекундных импульсов. Созданные в лабораторных условиях на базе транзисторов M1734 предварительные модели усилителей промежуточной частоты (24 *мгц*) дают усиление около 18 *дб* на каскад и имеют полосу пропускания в несколько *мгц*. Однако необходимо еще снизить сопротивление обратной связи этого транзистора. В импульснопередающих установках транзисторы M1734 хорошо работают в качестве импульсных генераторов и усилителей импульсов длительностью в 0,5 *мксек.*, требуя при этом только 6—8 *в* коллекторного напряжения и 12—20 *вт* коллекторной мощности на каскад. Амплитуды усиленных импульсов могут достигать при этом до 4—5 *в* (из 6 *в* полного напряжения на коллекторе), а время нарастания импульса составляет всего 0,01—0,02 *мксек.*

Улучшением теплоотвода выходная мощность в режиме усиления класса А лабораторных моделей транзисторов с $n-p-n$ -переходами

повышена в настоящее время до 2 *вт*. Эта величина, однако, не является верхним пределом, а скорее представляет собой необходимое значение для некоторых частных применений.

Таблица 6

Рабочие характеристики	Тип А (по состоянию на сентябрь 1949 г.)	Усовершенствованные типы (по состоянию на январь 1952 г.)
Усиление по току α	$5 \times$	$50 \times$ Триод с $n-p-n$ -переходами
Усиление на каскад при работе в режиме класса А	18 дб	22 дб М1729, М1768, точечно-контактный триод
		45 дб М1752, триод с $n-p-n$ -переходами
Шумовое число при 1 000 гц	60 дб	45 дб М1768, точечно-контактный триод
		10 дб М1752, триод с $n-p-n$ -переходами
Частотная характеристика f_c	5 мггц	7—10 мггц М1729, точечно-контактный триод
		20—50 мггц М1734, точечно-контактный триод
Выходная мощность при работе в режиме класса А	0,5 <i>вт</i>	2 <i>вт</i> Триод с $n-p-n$ -переходами
Характеристики переключения	Нет	Хорошие М1698, М1689, М1734, точечно-контактный триод
Сопrotивление обратной связи r_b	250 ом	70 ом М1729, точечно-контактный триод
Отношение фототока к тепловому току	2:1	20:1 М1740, триод с $n-p-n$ -переходами

Характеристиками, необходимыми для работы в переключающих устройствах, обладают уже сейчас описанные ранее точечно-контактные приборы типов М1698, М1689 и М1734. Уже сейчас с помощью точечно-контактных транзисторов М1689 и М1698 можно управлять телефонными реле, работающими при относительно больших токах — от 50 до 100 *ма*.

Новые фототранзисторы [4] с $n-p-n$ -переходами являются более совершенными приборами, чем точечно-контактные приборы более раннего изготовления [5]. Несмотря на то, что их квантовый к. п. д. меньше, чем к. п. д. точечно-контактных приборов, они привлекают к себе большое внимание. Отношение их светового тока к темновому сейчас значительно улучшено, а полное сопротивление коллектора увеличено в 10—100 раз, что дает возможность получать значительно большее выходное напряжение при том же самом световом потоке.

§ 5. НЕКОТОРЫЕ ПРИМЕНЕНИЯ ТРАНЗИСТОРОВ

Пакетные схемы для передачи данных. Для определения возможностей применения транзисторов в миниатюрных пакетных схемах были изучены наиболее важные функции импульсной кодовой системы передачи данных. Это изучение должно было привести к улучшению составных пакетных схем в отношении размеров, надежности в работе и потребляемой мощности. В частности, было улучшено выполнение кодирующих, передающих, считающих, регистрирующих функций и функций последовательного сложения. В результате этой работы получены диод М1728 с $n-p-n$ -переходами, фотоэлемент М1740 с $n-p-n$ -переходами и коммутирующий бусиновый транзистор М1689, которые применяются в пакетных схемах.

Транзисторы (с хорошей взаимозаменяемостью) успешно выполняют основные функции: 1) в 4-значном обратимом двоичном счетном устройстве; 2) в 6-значном угловом кодировщике положения; 3) в 6-значном двоичном трансляционном устройстве; 4) в 5-значном переводном регистраторе; 5) в счетных машинах последовательного действия.

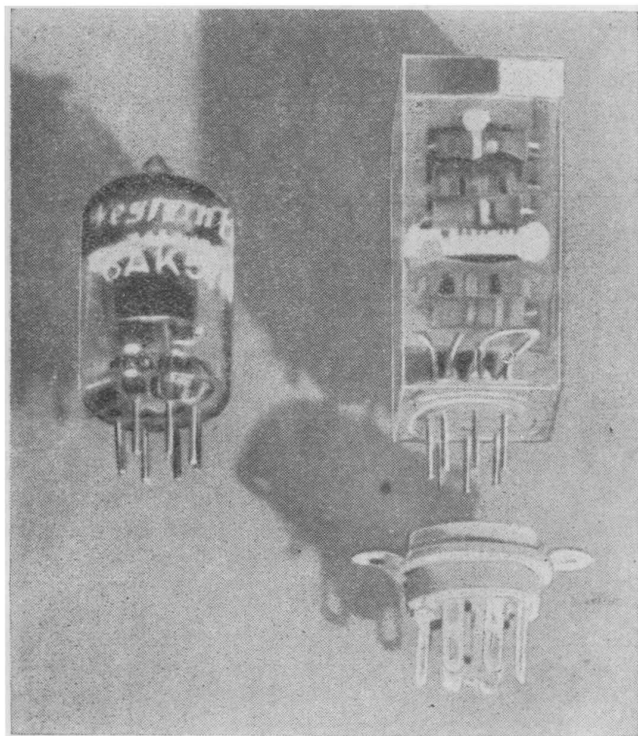
Для выполнения этих основных функций было построено 7 различных типов малых пакетных схем: регенеративный усилитель строб-импульсов, импульсный усилитель, диодный усилитель строб-импульсов, двоичное счетное устройство, усилитель задержанных импульсов и др.

Конечным результатом этой работы явилось наглядное доказательство возможности создания более или менее надежных систем для передачи данных из воспроизводимых и уже изготовляемых транзисторных пакетных схем. Дальнейшая работа по улучшению должна будет заключаться в том, чтобы сделать такую систему пакетных схем годной для эксплуатационного использования. По сравнению с ламповыми системами современная лабораторная модель такой транзисторной системы занимает в 10 раз меньший объем и потребляет в 10 раз меньшую мощность при выполнении одной и той же работы. На фиг. 18 дана фотография пакетной схемы транзисторного счетчика, а на фиг. 19 — фотография, на которой показаны обе стороны различных типов пакетных схем, применяемых в настоящее время¹⁾.

¹⁾ При изготовлении этих пакетных схем был использован способ авто- сборки частей. — *Прим. авт.*

Корпус этих систем изготовлен из прозрачной пластмассы, которая в будущем, очевидно, для этой цели применяться не будет.

На фиг. 20 показано несколько пакетных схем; для увеличения прочности и теплопроводности в пластмассу в качестве наполнителя введен кремнезем. Лабораторная модель счетчика, показанная на фиг. 20,

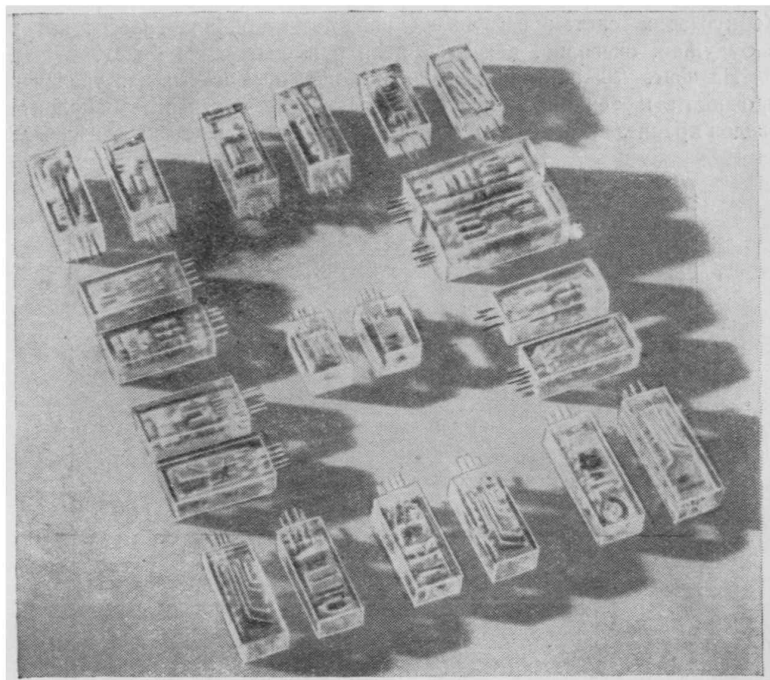


Фиг. 18. Пакетная схема транзисторного счетчика.

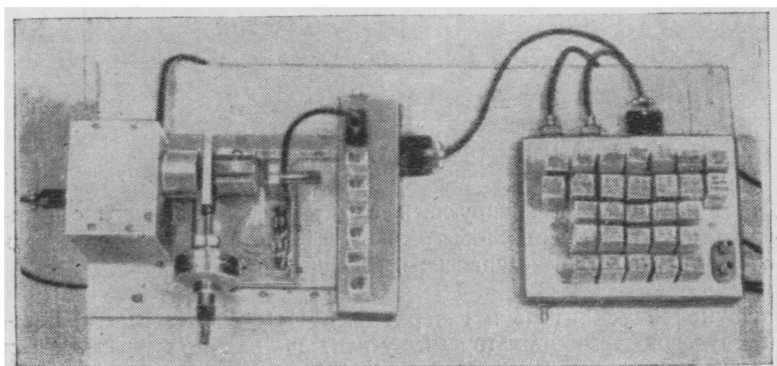
состоит из 6-значной кодирующей схемы (слева), связанной с 6 импульсными усилителями с обратной связью, которые в свою очередь питают 6-значный трансляционно-переводный счетчик.

Звуковой усилитель и генератор¹⁾. На фиг. 21 (справа) показан звуковой трансформаторный усилитель на двух транзисторах

¹⁾ Этот раздел написан по материалам работы Уоллеса и Питенполя [1]. —
Прим. авт.

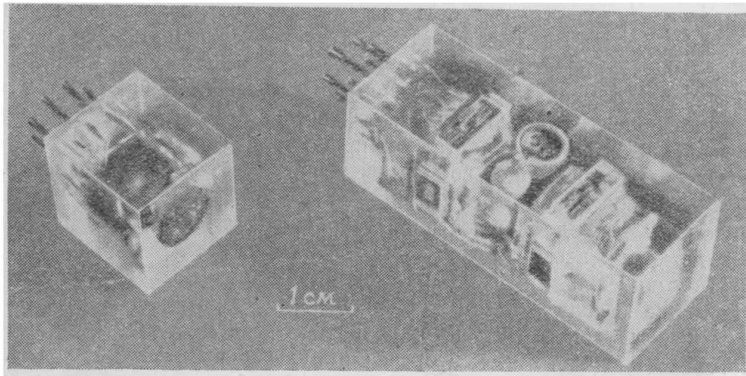


Ф и г. 19. Фотография ряда пакетных схем.



Ф и г. 20. Лабораторная модель счетчика.

М1752 с $n-p-n$ -переходами. Этот усилитель имеет полосу пропускания 100—20 000 гц и дает усиление по мощности до 90 дб.



Фиг. 21. Звуковой трансформаторный усилитель с двумя транзисторами М1752 с $n-p-n$ -переходами и транзисторный звуковой генератор.

В пределах от 1 до 20 в его усиление практически не зависит от напряжения коллектора, увеличивается с ростом напряжения только неискаженная мощность выхода при согласованной нагрузке. При напряжении коллектора 1,5 в ток коллектора приблизительно равен 0,5 ма (на один элемент) при полном расходе мощности в 1,5 мвт. Таким образом, выходная мощность усилителя класса А каждого транзистора составляет около 0,7 мвт. Шумовое число такого усилителя в полосе 1000 гц составляет от 10 до 15 дб в зависимости от действующих смещений.

На фиг. 21 (слева) показан небольшой транзисторный звуковой генератор, имеющий один транзистор М1752, трансформатор и один конденсатор. Для определения минимальной мощности, необходимой для устойчивой генерации, генератор испытывался при постепенном уменьшении напряжения коллектора. Испытания показали, что устойчивая генерация имеет место до напряжения, равного 55 мв. Ток коллектора и полный расход мощности были равны при этом соответственно 1,5 мка и 0,09 мквт.

§ 6. ЗАКЛЮЧЕНИЕ

В отношении воспроизводимости и возможностей применения транзисторы, находящиеся сейчас в стадии усовершенствования, не уступают электронным лампам.

Транзисторы, несомненно, имеют ббльший срок службы и ббльшую механическую прочность (выдерживают более сильные удары и вибрации), чем большинство электронных ламп. Что касается влияния температуры, то транзисторы значительно уступают электронным лампам, так как в большинстве случаев верхний предел их работы равен приблизительно только 70—80°С. Значение этого ограничения часто уменьшается меньшим потреблением мощности и, стало быть, меньшей степенью самонагрева прибора. Тем не менее, это ограничение является недостатком прибора.

Очевидно преимущество транзисторов в отношении размеров, а также в отношении потребляемой мощности. Прибор может работать при общем расходе мощности в несколько *мвт*, а в некоторых случаях даже при нескольких *мквт*.

В отношении пределов рабочих характеристик можно сделать следующие выводы: особое внимание нужно обратить на сравнительную оценку работы транзисторов и электронных ламп в импульсных системах (при частотах повторения импульсов вплоть до 1—2 *мгц*), так как в этом случае они, по существу, обеспечивают одинаковые рабочие характеристики, но транзисторы имеют значительное превосходство в отношении размеров и потребляемой мощности. Необходимо помнить, что по некоторым параметрам транзисторы являются приборами очень высокого качества, но в отношении температурной зависимости они хуже электронных ламп.

Такие же выводы, по существу, применимы и к транзисторам (особенно к транзисторам с *n—p—n*-переходами), работающим в низкочастотных (< 1 *мгц*) установках по передаче незатухающих волн. В области частот от 1 до 100 *мгц* все характеристики электронных ламп (исключая, может быть, только шумы и полосу пропускания) являются очень хорошими, так что в этом случае транзисторы имеют преимущество только в отношении размеров и прочности.

Таким образом, можно сказать, что несмотря на необходимость решить еще целый ряд проблем, связанных с физической сущностью работы самого транзистора, инженерам нужно серьезно исследовать возможности применения транзисторов не только с целью создания лучших систем, но также для того, чтобы оказывать влияние на работу по улучшению этих приборов для применения их в тех системах, которым они больше всего соответствуют.

ЛИТЕРАТУРА

1. Wallace R. L., Pietenpol W. Y., Bell Syst. Techn. Journ., 30, 530 (1951).
2. Ryder R. M., Kircher R. J., Bell Syst. Techn. Journ., 28, 367 (1949).
3. A. S. T. M. Manual, Quality Control of Materials, January 1951, part III p. 55.
4. Pietenpol W. Y., Phys. rev., 82, 121 (1951).
5. Shive Y. N., Bell Lab. Rec., 28, 337 (1950).

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Акцепторные уровни 273—277
 Акцепторы 29, 30, 41, 191
 Алмаз 18—26, 37—39, 297, 298
 — подвижность носителей тока 33, 393
 — проводимость 21—26, 38, 39, 393
 — распределение квантовых состояний 269, 270
 — структура 18—22, 187, 393
 — энергетические зоны 168, 187, 189—191, 213, 269, 270, 297, 298, 393, 394
 Аномальная дырка и аномальный электрон 401
 Ансамбль канонический 551—553
 — микроканонический 527, 551—553
 Атомный остаток 18, 20

 Бегущие волны 169—174, 591—598

 Валентная зона (зона валентной связи), см. Энергетическая зона
 Валентные кристаллы 18, 168
 Введение носителей тока, см. Носители тока
 Вероятности плотность 193—195, 427, 429, 432, 483, 508
 — ток 480—483
 Вероятность перехода 231—234, 236, 302, 305, 306, 563, 564, 569, 571, 572, 598, 626, 630
 Внутренние контактные потенциалы 415, 643, 646, 647, 659, 663, 664.
 Водород, атом 17—19, 157—159, 163, 272, 273
 — молекула 17—19
 Волновая функция 18, 39, 156—163, 429—443
 — — антисимметричная 163, 506—509
 — — комплексно сопряженная 430
 — — нормированная 159, 428, 506, 507
 — — ортогональность 442, 443
 — — разложение по собственным функциям 443
 — — связь с вероятностью 159, 160, 430

 Волновое уравнение 156—161
 — — Шредингера 154, 156—161, 206, 427—431, 434, 439, 441, 481, 488, 489, 510, 566, 576, 582, 583
 — — — обратимость во времени 485
 — — — решение для одномерной модели атома 444—450
 — — — — — кристалла 450—460
 — — — — — трехмерной модели кристалла 460—468
 — — — — — электрона в периодическом поле 169—173
 Волновой пакет дырочный 510—516
 — — электрона 490—504, 521, 522
 Волновые пакеты 197—201, 210, 223, 475—477
 — — общие свойства 483—490
 Волновые функции Блоха 155, 156, 172—179, 193—204, 206, 207, 317
 — — — для одномерного кристалла 172—179, 198—200, 450—460
 — — — для трехмерного кристалла 182, 183, 193, 194, 459—468, 477, 492
 Время жизни носителей тока, см. Носители тока
 — — — в нитевидном образце 376—387
 Время пролета 84, 89, 90
 — — в нитевидном транзисторе 93, 95, 100
 — — в транзисторе типа А 137—139
 Выпрямленс, см. Переходы p - n -типа и Точечный контакт
 Вырождения температура 291, 292
 Вырожденные полупроводники 291, 334—340
 Вырожденный электронный газ 39, 291, 292, 336—340

 Гармонические осцилляторы 158, 165, 166, 169—171, 186—189, 440, 597
 — — нормальные колебания 158, 161, 166, 170—172, 177, 191, 192, 579
 — — распределение энергии по Планку 523—533, 553

- Германий 16—20, 22, 25—27, 30—33, 37, 38, 41—43, 114, 115, 126, 127, 140, 141, 215, 256, 312, 333, 336, 337, 340, 342, 343, 393—397, 408, 409, 419
 — введение носителей тока 78—97
 — монокристаллы 33, 397, 409
 — подвижность 33, 296, 396—398
 — при комнатной температуре 296—298
 — проводимость 30—32, 41—43, 190, 191, 243, 296, 340, 397
 — сопротивление растекания 130, 131
 — упругие свойства 611, 612
 — уровень Ферми 293, 294, 297
 — формовка 140—143
 — энергетические зоны 190, 191, 271, 275
 Групповая скорость 198—203, 223, 475—480, 514
- Детального равновесия принцип 305, 309, 356, 404, 569
 Детерминант цепи 65, 73
 Дифракция электронов 427—429, 431—435
 Диффузионная длина 357, 371, 653
 Диффузия носителей тока, см. Носители тока
 Донорные уровни 271—277, 294
 Доноры 29, 30, 41, 191
 Дырки 23—30, 218—223
 — аномальные 401
 — введение 79—97, 126, 128, 354
 — влияние на характеристики точечных контактов 90—92
 — время жизни 87, 93, 94, 97, 99, 113, 354
 — движение в зоне валентных связей 218—228
 — концентрация 37, 79—92, 354
 — подвижность 81, 84, 86, 93, 96, 146
 — свойства 510—516, 626, 627
 — спонтанное возникновение 101, 354
 Дырочный ток 81—99, 129, 131, 134, 143—145
 — — в переходах p - n -типа 117—125, 651—655
 — — измерения 84—90, 92
 — — импульс 93—97, 131, 391, 392, 403, 412, 414
- Запорный слой 129, 131, 132, 137, 141
 Запрещенная зона, см. Энергетическая зона
- Зоны Бриллюэна 154, 155, 166—204, 206—226, 239, 250, 251, 421, 461, 468, 469, 577
 — — для кристалла одномерного 180, 181, 203
 — — — с простой кубической структурой 182—187, 577
 — — — трехмерного 181—191, 193, 203, 204, 213, 214, 461, 468, 469, 577
 — — при воздействии электрического и магнитного полей 206—226, 239, 250, 251
- Идеальный кристалл, см. Кристалл идеальный
 Избыточный электрон 23—37, 213—218, 226, 271, 286, 409, 490—504, 543, 560, 584—591, 593
 — — взаимодействие с атомными колебаниями 584—591
- Квази-импульс 154, 178—189, 203—223, 240
 Квази-уровни Ферми 364, 365, 367, 650, 651
 Квантованная система 524, 551—554
 Квантовая механика элементарная 424—474
 — — задачи 469—474
 Квантовые состояния 154, 162—193, 208, 461, 524—527, 551, 552
 — — априорная вероятность 524, 526, 551
 — — распределение по энергиям 267—277, 298, 525—527, 540
 — — — — роль примесей 271—277
 — — сохранение числа состояний 167, 180, 185, 208, 267
 — — эффективная плотность в зоне проводимости и валентной зоне 290, 295, 539, 540
 Квантовые числа 158, 159, 162, 163, 169, 170
 Квантовый выход 147, 410, 659, 668
 Ковалентная связь 17—21, 617
 Ковалентные кристаллы 17
 Коллектор 55, 56, 90—92, 643—670, 675—694
 Контакт металл — полупроводник, см. Точечный контакт
 Конуэлла — Вайскопфа формула 316, 333, 336, 340, 341
 Концентрация носителей тока, см. Носители тока

- Оператор импульса 433, 434, 437, 440, 495
 — кинетической энергии 435, 438, 518, 590
 — скорости 493—495
 — смещения 589—591, 595—598
 Операторы, общие свойства 434—440, 583
 — собственные значения 439, 442, 443
 — — — вырожденные 443
 — собственные функции 439—443, 598
 — — — ортогональность 442, 443
 — — — ортонормированная форма 444, 511, 512
 «Оптические колебания» 606, 607
 Основной электрод 57, 643—672, 677—694
 Отрицательное сопротивление 48, 49, 109, 137, 688
- Пакетные схемы 698—701
 Паули принцип 162, 163, 232, 309, 506, 507, 509, 627
 Переходы p — n -типа 104, 114—125, 147, 148, 358—365, 409, 643, 653
 — — — полная проводимость 374, 375
 — — — сопротивление 368, 373, 394, 395
 — — — теория 365—376
 Периодические граничные условия 183, 194, 197, 206, 207, 421, 439, 463, 510, 516—522, 574, 575
 Периодическое силовое поле 169—179, 206, 211, 317, 627
 Планка формула 532
 Плотность заряда 287, 363, 507—509, 512—514
 Плотность тока 354, 365, 372, 512—514
 p — n -ловушка 145, 146, 655, 656—658
 Поверхностные заряды 49—52
 — состояния 52—55, 128, 141, 298, 451
 Поверхность постоянной энергии 186, 187, 203, 213, 214, 399—402
 Подвижность носителей тока, см. Носители тока
 Полезная мощность 64, 65
 Постоянная спада 376, 377, 381, 386, 387
 Потенциальная яма 444
- Потенциалы деформации 319—324, 392—394, 402, 603—616
 Примесная проводимость и примесные полупроводники 27—43, 261, 274, 275, 340
 Примесные центры 41, 276, 277, 302
 — — — рассеяние носителей тока 310—317, 332, 333, 340—343, 559
 Проводимость дырочная (p -типа) 17, 23—32, 42, 81, 87—90, 103, 115, 121, 372
 — примесная 27—43, 261, 274, 275, 340
 — собственная 31, 33, 36, 40, 333, 334, 360, 373, 394, 652, 653
 — электронная (n -типа) 17, 23—32, 83—90, 115, 119, 121
 Проводники кожные 17
 — электронные 17
 Пуассона уравнение 362, 363
- Развертывающее поле 86
 Разрешенная зона, см. Энергетическая зона
 Распределение 524—539, 551—554
 — задачи 551—554
 — Планка 532
 — Ферми—Дирака 536, 537
 — — — функция распределения 278—283, 537
 Рассеяние носителей тока, см. Носители тока
 Рекомбинационно-эмиссионный центр 405—407
 Рекомбинация носителей тока, см. Носители тока
 Релаксация время 82, 244, 306—310, 316, 324, 330—337, 341
- Свободная энергия 538, 539, 556
 Свободный пробег, время 236—238, 248
 — — — среднее время 229, 231, 234—238, 244, 248, 259—261, 263, 264, 329—337, 341, 348, 349, 571
 — — — задачи на вычисление 263, 264, 348, 349
 — — — средняя длина 244, 259, 316, 322, 323, 332, 344—346, 571, 572, 616, 626
 Связь ковалентная 17—21, 617
 — металлическая 20, 21
 Скорость групповая 198—203, 223, 475—480, 514
 — фазовая 198, 199, 202, 475
 — дрейфа 244—248, 382

- Случайные процессы 229—233
 Собственная проводимость 31, 33, 36, 40, 333, 334, 360, 373, 394
 Собственные значения 158
 — — оператора 439, 442, 443
 — — вырожденные 443
 — функции 158, 159, 170, 439
 — — оператора 439—443, 598
 — — ортогональность 442, 443
 — — ортонормированная форма 444, 511, 512
 — — энергии 440—443, 463
 Сопrotивление, см. Электропроводность
 — растекания 129—131, 136, 402
 Спин 161—163, 504, 505
 Статистическая механика, приложение к полупроводникам 524—556
 — — связь с термодинамикой 533—536, 551—556
 — — сумма 552—554
 Стационарные состояния 442
 Суля эффект 97—101, 376, 384—387, 401, 413
 — — аномальный 401
 Счетчики транзисторные 22, 147, 699—701

 Теллур 393, 394
 Температура 528, 535
 Теория переходов 560—639
 — — задачи 618, 628
 — — квантовая 564—571
 — — эквивалентность классической и квантовой моделей 561—564, 630, 631
 Тепловые колебания решетки 25, 26, 28, 29, 31, 36, 154, 603, 611
 — — переходы ими обусловленные 233, 308, 310, 317—324, 560—573
 Термистор 48, 49
 Точечный контакт 48, 87, 90—92, 114, 125—132
 — — влияние концентрации дырок 90—92
 — — емкость 131, 132
 — — проводимость 91, 92, 98, 100, 101
 — — теория формовки 139—143
 Транзистор нитевидный 15, 16, 84—87, 90, 92, 147, 148, 265, 390, 407, 408
 — — время жизни дырок 113, 376—387, 404
 — — — пролета дырок 111—114
 Транзистор нитевидный, коэффициент усиления по току 112—114, 143
 — — — положительная обратная связь 108—110
 — — — теория 104—114, 376—487
 — — — шумы 402—407
 — — — эквивалентная схема 105—108, 110, 112, 113
 — — — эффект Суля 97—99, 384—387
 Транзистор типа А 47, 54—76, 78, 90, 108, 114, 125, 132—139, 643, 644, 676, 694—697
 — — — в качестве усилителя 62—76
 — — — время пролета 137—139
 — — — задачи на расчеты 72—76
 — — — коэффициент усиления по току 61, 62, 133, 134, 143—146, 691, 692, 697
 — — — положительная обратная связь 61, 65—67, 137, 644
 — — — срок службы 690, 691
 — — — теория 132—139
 — — — характеристики 56—60, 62, 694—697
 — — — воспроизводимость и устойчивость 681—684, 691, 692
 — — — шумы 68—72, 75, 76, 670, 697
 — — — эквивалентная схема 60—68, 134—137, 682, 683
 Транзисторная электроника 15, 17
 — — задачи на расчеты 411—415
 — — современное состояние 407—411, 673—702
 Транзисторы в качестве генератора 701, 702
 — — — счетчика 22, 147, 699—701
 — — — усилителя 48—51, 62—76, 104, 105, 645—649, 654—658, 661, 663—670, 695—701
 — — — элемента схемы 47—76, 682, 683
 — — задачи 72—76, 148, 149
 — — надежность работы 690—694
 — — размеры 694, 695
 — — срок службы 690, 691
 — — температурная чувствительность 691—693, 702
 — — физическая теория 83, 104—149
 — — с p - n -переходами 120—125, 643—672
 — — — аналогия с вакуумным триодом 122—125, 645, 646
 — — — n - p - n -транзистор 643—656, 659—672, 677—681, 683, 684, 692, 694—698, 701, 702

- Транзисторы с $p-n$ -переходами в качестве усилителя 645—649, 654—658, 661, 663—668, 696, 698
 — — — — — граничные условия 649—652
 — — — — — проводимость 651—656, 659—668, 670—672
 — — — — — характеристики 647—649, 679—681, 683, 694—698, 702
 — — — — — экспериментальные данные 660—668
 — — — — — $p-n-p$ -транзистор 120—125, 146
 — — — — — $p-n-p-n$ -транзистор 656—658
 — с точечными контактами 643, 656, 676—678, 690—698
 — — — — — бусинковый 676, 689, 690, 693—698
 — — — — — рабочие характеристики 696—698
 Трансляции 461, 464, 473
 Углерод 18—20, 167
 Умова — Пойнтинга вектор 195—197
 Унитарное преобразование 584
 Уровень Ферми 279, 282—285, 290, 291, 293, 295, 360—364, 537—539
 — — — — — зависимость от температуры 540—549
 Усиление в рабочем режиме 65, 66
 — динамический коэффициент 70, 71
 — полезное 66, 67, 70
 — — — — — максимальное 66, 67
 Усиление по мощности 50, 51, 64, 65, 83, 125, 656, 681
 — по напряжению 65, 677, 681
 — по току 51, 61, 62, 67, 92, 106—108, 112—114, 125, 129, 133, 134, 140, 143—146, 415, 645, 646, 654—659, 676, 681, 688—693, 696, 697
 Фазовая скорость 198, 199, 202, 475
 Фактор шумов, см. Шумы
 Ферми — Дирака статистика 278—300
 — — — — — для полупроводников 286—300
 — — — — — случай комнатной температуры 296—298
 — — — — — для простой модели 283—285
 — — — — — задачи 299, 300
 — — — — — классическое приближение 281
 — — — — — функция распределения 278—283, 305, 537
 Ферми квази-уровни, см. Квази-уровни Ферми
 — уровни, см. Уровень Ферми
 Фока уравнение 509, 510
 Фонон 94, 304, 348, 565, 602, 615, 616, 618—620
 Формовка 139—143
 Фосфор 32
 — в примеси к кремнию 32—40, 42, 345
 Фотопроводимость 22—25, 43, 114, 146, 147, 390, 410, 646
 Фототранзисторы 146, 147, 646, 658, 659, 668, 698
 Химический потенциал 537, 538
 Холла подвижность 84, 254, 256, 329, 331—334, 396—399
 — — элементарная теория 250—256
 — постоянная 254, 257—263, 331—334, 336
 — угол 250, 254, 257—262, 357
 — эффект 32, 84, 97, 248—265, 324—344, 357, 384, 396—402
 — — задачи на вычисления 264, 265
 — — измерения 32, 256—258
 — — нормальный и аномальный 401
 — — теория 248—256, 324—344, 396—402
 — — формулы 260—263, 331—337
 Шумы 670, 696, 697
 — в транзисторе типа А 68—72, 75, 76, 670, 697
 — теория 402—407
 — фактор 71, 72
 — эффект передачи 404, 405
 Эйнштейна соотношение для подвижности 155, 352, 355, 356
 Экситон 94
 Электрические фильтры 178
 Электрическое и магнитное поля, действие на электроны и дырки 205—228, 231, 235, 248—256, 324—349, 355—358, 384—387, 396—402, 490—504, 510—522
 — — — — — задачи 227, 228, 346—349
 Электроны в кристалле аномальные 401
 — — — — — обобщение второго закона Ньютона 205—207, 214
 — — — — — одномерный случай 194—197, 202
 — — — — — отрицательная масса 210, 221, 228

- Электроны в кристалле, плотность тока 194—197, 229, 230
 — — — под действием электрического и магнитного полей 206—228, 231, 235, 248—256, 324—349, 355—358, 396—402, 490—504
 — — — скорости и токи 79—90, 124—126, 129, 134, 143—145, 193—204, 217, 222, 229, 239—248, 480, **514**
 — — — средняя скорость 193—198, 217, 240
 — — — трехмерный случай 203
 — — — энергия 201—203, 217, 514
 Электропроводность, теория 229—248, 301—350, 651—656
 — формулы 31, 32, 241, 334—336, 373—375
 Электростатический потенциал 53, 54, 358—365, 605, 616, 617
 — — связь с уровнями энергии 359
 Элементарная ячейка 183, 185, 187, 462, 481
 — — деформация при тепловых колебаниях 603, 604, 608
 Элементарные векторы 461—464, 473
 Эмиттер 55, 56, 86, 90, 120, 126, 128, 143, 643—670, 675—694
 — линии и трубки тока 138, 139
 Энергетическая зона 17, 22, 26, 117—122, 165—192, 202, 203, 319, 358—365, 450—460, 469, 470
 — — валентных электронов (валентная зона) 26, 30, 118, 119, 359, 540
 Энергетическая зона запрещенная 27, 118, 119, 167, 190, 191, 361, 450, 451, 460
 — — — проводимости (свободная зона) 26, 30, 118, 119, 190, 191, 514, 541
 — — — эффективное число состояний 290, 295, 539, 540
 — — — разрешенная 167, 450, 451, 460
 — — — ширина зоны 27, 166, 267, 550
 Энергетические зоны металлов, изоляторов и полупроводников 189—191
 — уровни примесные 271—277
 — — — акцепторные 273—277
 — — — донорные 271—277
 — — связь с электростатическим потенциалом 359
 — — чистых кристаллов 267—271
 Энергия активации 36—42, 294, 546—550
 — закон сохранения 208—211, 222, 490
 — ионизации примесных центров 39—41, 276, 277
 — свободная 538, 539, 556
 — связи 39, 40, 167, 168, 272, 294, 542
 — сравнительная шкала 280
 Энтропия 528, 535, 556
 Эрмитовы операторы 436—440
 Эффективная масса носителей тока, см. Носители тока

ОГЛАВЛЕНИЕ

Предисловие редактора перевода	3
Из предисловия автора	7
Обозначения	9

Часть I

ВВЕДЕНИЕ В ЭЛЕКТРОНИКУ ТРАНЗИСТОРОВ

(Перевод Б. И. Болтакса)

Глава 1. Объемные свойства полупроводников	15
§ 1. Избыточные электроны и дырки — носители электрического тока	17
§ 2. Примесные полупроводники. Доноры и акцепторы	27
§ 3. Интерпретация данных по проводимости и эффекту Холла	31
Задачи	43
Литература	43
Приложение к гл. 1	46
Глава 2. Транзистор как элемент схемы	47
§ 1. Состояние вопроса о полупроводниковых усилительных устройствах до появления транзистора	47
§ 2. Транзистор типа А	54
Задачи	72
Литература	77
Глава 3. Количественное исследование процесса введения дырок и электронов в германий	78
§ 1. Введение носителей тока в германий	78
§ 2. Влияние магнитного поля на концентрацию дырок и электронов. Эффект Суля	97
Задачи	102
Литература	103
Глава 4. Физическая теория транзисторов	104
§ 1. Теория нитевидных транзисторов	104
§ 2. Переходы типа $p-n$ и транзисторы с $p-n$ -переходами	114
§ 3. О природе контакта между металлом и полупроводником	125
§ 4. Теория транзистора типа А	132

§ 5. О теории „формовки“ и высоких значениях α	139
§ 6. Фототранзисторы и счетчики	146
Задачи	147
Литература	149

Часть II

ОПИСАТЕЛЬНАЯ ТЕОРИЯ ПОЛУПРОВОДНИКОВ

(Перевод А. Н. Арсеньевой)

Глава 5. Квантовые состояния, энергетические зоны и зоны Бриллюэна	153
§ 1. Введение	153
§ 2. Квантовые состояния, спин и принцип Паули	156
§ 3. Квантовые состояния и энергетические зоны в кристаллах	163
§ 4. Волновые функции Блоха в периодическом поле	169
§ 5. Зоны Бриллюэна и квази-импульс P как квантовое число	178
§ 6. Энергетические зоны металлов, изоляторов и полупроводников	189
Задачи	191
Литература	192
Глава 6. Скорости и токи электронов в кристаллах	193
§ 1. Понятия скорости и тока	193
§ 2. Вектор Умова — Пойнтинга и метод плотности потока	195
§ 3. Волновые пакеты и групповая скорость	197
§ 4. Скорости, соответствующие квантовым состояниям в зоне Бриллюэна	200
Глава 7. Электроны и дырки в электрических и магнитных полях	205
§ 1. Действие силы $\dot{P} = F$	206
§ 2. Теорема сохранения квантовых состояний	208
§ 3. Принцип сохранения энергии при наличии электрического поля	208
§ 4. Действие магнитного поля	211
§ 5. Поведение избыточного электрона в кристалле; эффективная масса и перестройка зоны Бриллюэна	213
§ 6. Движение дырки	218
§ 7. Выводы	223
§ 8. Связь между квантовой и классической механикой	226
Задачи	227
Литература	228
Глава 8. Введение в теорию электропроводности и эффект Холла	229
§ 1. Введение	229
§ 2. Случайные процессы и вероятности переходов	231
§ 3. Среднее время свободного пробега	234
§ 4. Рассмотрение средней скорости, подвижности и проводимости при помощи метода зон Бриллюэна	239
§ 5. Рассмотрение скорости дрейфа, подвижности и проводимости на основе корпускулярных представлений	244
§ 6. Рассмотрение эффекта Холла с помощью корпускулярных представлений	248

§ 7. Рассмотрение эффекта Холла с помощью теории зон Бриллюэна. Подвижность, определенная из эффекта Холла, подвижность, определенная по скорости дрейфа и микроскопическая подвижность	250
§ 8. Измерение эффекта Холла. Практические единицы и единицы MKS	256
§ 9. Различные виды формул для эффекта Холла	260
Задачи	263
Литература	265
Глава 9. Распределение квантовых состояний по энергиям	267
§ 1. Схема энергетических уровней для чистых кристаллов	267
§ 2. Уровни энергии, обусловленные примесями, — донорные и акцепторные уровни	271
Задачи	277
Литература	277
Глава 10. Статистика Ферми — Дирака в применении к полупроводникам	278
§ 1. Функция распределения Ферми — Дирака	278
§ 2. Статистика Ферми — Дирака в применении к простой модели	283
§ 3. Статистика Ферми — Дирака в применении к полупроводникам	286
§ 4. Случай комнатной температуры	296
Задачи	299
Литература	300
Глава 11. Математическая теория проводимости и эффекта Холла	301
§ 1. Введение	301
§ 2. Время релаксации	302
§ 3. Переходы, обусловленные тепловым возбуждением и рассеянием на примесях	310
§ 4. Средние значения подвижностей, среднее время свободного пробега и постоянные Холла	324
§ 5. Сравнение с опытом	337
Задачи	346
Литература	350
Глава 12. Приложение теории к электронике транзисторов	351
§ 1. Введение	351
§ 2. Принятые обозначения; уравнение непрерывности	352
§ 3. Соотношение Эйнштейна, связывающее подвижность с коэффициентом диффузии	355
§ 4. Энергетические зоны, электростатический потенциал и квазиуровни Ферми	358
§ 5. Теория $p-n$ -перехода	365
§ 6. Время жизни носителей заряда в нитевидных образцах германия и эффект Суля	376
§ 7. Решение уравнения непрерывности для случая больших плотностей введенных носителей тока (решение Херинга)	387
§ 8. Зависимость энергий \mathcal{E}_{1p} , \mathcal{E}_{1n} и \mathcal{E}_{1G} от объема	392

§ 9. Подвижность, определяемая по скорости дрейфа (подвижность дрейфа), и подвижность, определяемая по эффекту Холла (подвижность Холла)	396
§ 10. К теории шумов	402
§ 11. Заключительные замечания	407
Задачи	411
Литература	415

Часть III

НЕКОТОРЫЕ ВОПРОСЫ КВАНТОМЕХАНИЧЕСКОЙ ТЕОРИИ

(Перевод Т. А. Конторовой)

Глава 13. Введение	419
Глава 14. Элементарная квантовая механика	424
§ 1. Математическая теория линии передачи без потерь	424
§ 2. Уравнение Шредингера	427
§ 3. Операторные методы	436
§ 4. Собственные функции оператора энергии	440
§ 5. Решения для простого „атома“	444
§ 6. Решения для простого „кристалла“	450
§ 7. Функции Блоха для трех измерений	460
Задачи	469
Литература	474
Глава 15. Скорости и ускорения электронов и дырок; ток (теория)	475
§ 1. Групповая скорость	475
§ 2. Случай электронных волн	476
§ 3. Аналогия из теории электрических цепей	477
§ 4. Ток вероятности	480
§ 5. Некоторые общие свойства квантовомеханических волновых пакетов	483
§ 6. Ускорение электронного волнового пакета электрическими и магнитными полями	490
§ 7. Принцип Паули и антисимметрия волновых функций	504
§ 8. Дырочный волновой пакет	510
§ 9. Периодические граничные условия и электрические поля	516
Литература	522
Глава 16. Статистическая механика в приложении к полупроводникам	524
§ 1. Вывод законов распределения для простых квантованных систем	524
§ 2. Определение эффективного числа квантовых состояний в энергетической зоне	539
§ 3. Температурная зависимость уровня Ферми	540
§ 4. Энергия активации	546
Задачи	551
Литература	557

Глава 17. Теория переходов для дырок и электронов	558
§ 1. Введение	558
§ 2. Теория переходов; связь средней длины свободного пробега с матричными элементами	560
Задачи	572
§ 3. Нормальные колебания атомов кристалла	573
Задачи	579
§ 4. Взаимодействие атомных колебаний с избыточным электроном	584
Задачи	590
§ 5. Преобразование к бегущим волнам	591
§ 6. Вычисление матричного элемента	598
Задачи	616
§ 7. Вычисление матричных элементов при учете принципа Паули	620
Задача	631
Литература	631
Библиография	633

ПРИЛОЖЕНИЕ

(Перевод П. В. Гулятьева)

В. Шокли, М. Спаркс и Г. Тил. Транзисторы с $p-n$-переходами	643
§ 1. Введение	643
§ 2. Работа полупроводника с $n-p-n$ -переходами в качестве транзистора и в качестве умножителя с ловушкой	643
§ 3. Описание опытных моделей	647
§ 4. Рабочие характеристики	647
§ 5. Общие теоретические положения и граничные условия	649
§ 6. Вольтамперные характеристики	652
§ 7. Специальные режимы работы	655
§ 8. Сравнение с экспериментом	660
§ 9. Некоторые расчетные соображения	669
Дополнение. Доказательство равенства $G_{lr} = G_{rl}$	670
Литература	672
Дж. Мортон. Современное состояние транзисторной электроники	673
§ 1. Введение	673
§ 2. Воспроизводимость	675
§ 3. Надежность	690
§ 4. Размеры и мощность	694
§ 5. Некоторые применения транзисторов	698
§ 6. Заключение	701
Литература	702
Предметный указатель	703

Редактор *Л. В. Гессен*
Технические редакторы *Б. И. Корни-*
лов, А. Н. Никифорова

Сдано в производство 11/XII 1952 г.
Подписано к печати 16/VII 1953 г.
А-04037. Бумага $60 \times 92\frac{1}{16} = 22,4$ бум. л.
44,8 печ. л.

Уч.-издат. л. 49,3. Изд. № 2/1638.
Цена 36 р. 50 к. Зак. 4042.

4-я типография им. Евг. Соколовой
Союзполиграфпрома Главиздата
Министерства культуры СССР.
Ленинград, Измайловский пр., 29.

ОПЕЧАТКА

<i>Стр.</i>	<i>Строка</i>	<i>Напечатано</i>	<i>Следует читать</i>
112	11 стр.	т _р	т _г

Зак. 4042.