



**R&C**  
*Dynamics*

# Quantum Transport: Atom to Transistor

---

Supriyo Datta

Purdue University



CAMBRIDGE  
UNIVERSITY PRESS

С. Датта

---

# Квантовый транспорт от атома к транзистору

Перевод с английского к. ф.-м. н. Д. В. Хомицкого  
Под редакцией проф. В. Я. Демиховского



Москва ♦ Ижевск

2009

УДК 530.1  
ББК 22.314  
Д 209



Издание осуществлено при финансовой поддержке  
Российского фонда фундаментальных исследований  
по проекту №08-02-07047.

### Датта С.

Квантовый транспорт: от атома к транзистору. — М.-Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2009. — 532 с.

Книга С. Датта, автора известных работ в области нанофизики и наноэлектроники, посвящена проблемам электронного транспорта в низкоразмерных полупроводниковых структурах. Последовательно рассматриваются различные квантовые системы, от атома водорода до нанотранзистора. Излагаются наиболее общие понятия и методы неравновесной статистической механики и кинетики, и при этом не предполагается первоначальное знакомство читателя с квантовой механикой. В книге приведено большое количество численных примеров, а также программ в пакете MATLAB для проведения расчетов квантовых состояний и транспорта в объемных полупроводниках и низкоразмерных структурах. Дополнения к данному изданию в форме видеолекций, посвященных ключевым результатам книги, содержатся на Интернет-сайте автора.

Книга рассчитана на научных работников, инженеров, студентов старших курсов и аспирантов, специализирующихся в области физики конденсированного состояния и наноэлектроники.

**ISBN 978-5-93972-744-0**

**ББК 22.314**

© S. Datta 2005

This work was originally published in English by Cambridge University Press under the title *Quantum Transport: Atom to Transistor*. This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

© Перевод на русский язык:

НИЦ «Регулярная и хаотическая динамика», 2009

<http://shop.rcd.ru>

<http://ics.org.ru>

# Оглавление

Предисловие редактора перевода . . . . .	9
Предисловие автора к русскому изданию . . . . .	11
Предисловие . . . . .	17
Благодарности . . . . .	21
Список обозначений . . . . .	22
ГЛАВА 1* Пролог. Электрическое сопротивление (квантовый подход) . . . . . 25	
1.1. Диаграмма энергетических уровней . . . . .	28
1.2. Как возникает электрический ток? . . . . .	33
1.3. Квант кондактанса . . . . .	38
1.4. Профиль распределения потенциала . . . . .	42
1.5. Кулоновская блокада . . . . .	47
1.6. Переход к закону Ома . . . . .	51
ГЛАВА 2* Уравнение Шредингера . . . . . 67	
2.1. Атом водорода . . . . .	67
2.2. Метод конечных разностей . . . . .	73
2.3. Примеры . . . . .	76
2.3.1. Частица в потенциальной яме . . . . .	77
2.3.2. Частица в трехмерной потенциальной яме . . . . .	83
ГЛАВА 3* Приближение самосогласованного поля . . . . . 90	
3.1. Метод самосогласованного поля (ССП) . . . . .	90
3.2. Связь с многоэлектронной картиной . . . . .	96
3.3. Типы связи . . . . .	106
3.3.1. Валентные электроны . . . . .	106
3.3.2. Ионная связь . . . . .	106
3.3.3. Ковалентная связь . . . . .	110
3.4. Дополнение: многоэлектронная картина . . . . .	115

6 ◇ Оглавление

<b>ГЛАВА 4. Базисные функции</b>	127
*4.1. Базисные функции как инструмент численных методов	128
*4.2. Базисные функции как методология	135
4.3. Равновесная матрица плотности	142
4.4. Дополнение	147
4.4.1. Матрица плотности	147
4.4.2. Теория возмущений	149
<b>ГЛАВА 5. Зонная структура</b>	156
*5.1. Элементарные примеры	157
*5.2. Основной результат	163
*5.3. Обычные полупроводники	171
5.4. Влияние спин-орбитального взаимодействия	176
5.5. Дополнение: уравнение Дирака	183
<b>ГЛАВА 6. Подзоны</b>	187
*6.1. Квантовые ямы, проволоки, точки и «нанотрубки»	188
*6.2. Плотность состояний	199
*6.3. Минимальное сопротивление проволоки	208
6.4. Скорость электрона в (под)зоне	214
<b>ГЛАВА 7. Емкость</b>	220
*7.1. Модельный гамильтониан	222
7.2. Электронная плотность. Матрица плотности	229
*7.3. Квантовая и электростатическая емкость	239
7.4. Дополнение: многозонный гамильтониан в приближении эффективной массы	246
<b>ГЛАВА 8*. Уширение уровней</b>	255
8.1. Открытые системы	258
8.2. Локальная плотность состояний	265
8.3. Время жизни	276
8.4. Что образует контакт (резервуар)?	284
<b>ГЛАВА 9. Когерентный транспорт</b>	297
*9.1. Предварительный обзор	297
9.2. Матрица плотности	304
9.3. Входящий и исходящий токи	312
*9.4. Функция пропускания	315
*9.5. Примеры	325

9.5.1. Аналитический пример . . . . .	325
9.5.2. Численный пример . . . . .	328
<b>ГЛАВА 10. Некогерентный транспорт . . . . .</b>	<b>339</b>
10.1. Почему атом излучает свет? . . . . .	342
10.2. Примеры . . . . .	352
10.2.1. Атомные переходы . . . . .	352
10.2.2. Межзонные переходы в полупроводниках . . . . .	356
10.2.3. Внутризонные переходы в полупроводниках . . . . .	359
10.3. Входящий и выходящий токи . . . . .	362
10.4. Дополнение: фононы . . . . .	367
<b>ГЛАВА 11. От атома к транзистору . . . . .</b>	<b>380</b>
11.1. Уравнения квантового транспорта . . . . .	381
*11.2. Закон Ома . . . . .	387
11.2.1. Классический транспорт . . . . .	387
11.2.2. Когерентный транспорт (одна подзона) . . . . .	390
11.2.3. Когерентный транспорт (много подзон) . . . . .	391
11.2.4. Квантовый транспорт с дефазировкой . . . . .	395
*11.3. Где происходит выделение тепла? . . . . .	396
11.3.1. Эффект «Пельтье» . . . . .	398
*11.4. Где происходит падение напряжения? . . . . .	400
<b>ГЛАВА 12. Эпилог . . . . .</b>	<b>415</b>
<b>Приложение: формализм вторичного квантования . . . . .</b>	<b>425</b>
П.1. Корреляционные функции . . . . .	426
П.2. Неравновесная матрица плотности . . . . .	432
П.3. Входящий и исходящий потоки . . . . .	437
П.4. Электронный ток и неупругие процессы . . . . .	441
П.5. Кулоновская блокада и резонанс Кондо . . . . .	447
<b>Программы на языке MATLAB для построения иллюстраций . .</b>	<b>455</b>
<b>Дополнительная литература . . . . .</b>	<b>516</b>
1. Пролог . . . . .	516
2. Уравнение Шредингера . . . . .	517
3. Самосогласованное поле . . . . .	518
4. Базисные функции . . . . .	519
5. Зонная структура . . . . .	519
6. Подзоны . . . . .	520
7. Емкость . . . . .	520

8 ◇ **Оглавление**

8.	Уширение уровней . . . . .	520
9.	Когерентный транспорт . . . . .	521
10.	Некогерентный транспорт . . . . .	521
11.	От атома к транзистору . . . . .	522
12.	Приложение: формализм вторичного квантования . . . . .	522
<b>Литература . . . . .</b>		524
<b>Предметный указатель . . . . .</b>		528

# **Предисловие редактора перевода**

В книге известного американского физика, специалиста в области теории конденсированного состояния, нанофизики и спинtronики профессора С. Датты рассматриваются проблемы, связанные с протеканием электрического тока в низкоразмерных структурах и приборах наноэлектроники.

При анализе различных кинетических процессов вnanoобъектах, описываемых уравнением Шредингера и связанных контактами с макроскопическим резервуаром, автор использует метод неравновесных функций Грина. Этот метод был разработан Л. В. Келдышем в 1964 году и в настоящее время широко используется для решения задач физической кинетики. В книге основные уравнения метода неравновесных функций Грина вводятся на базе сравнительно простых соображений, доступных широкому кругу научных работников и инженеров, занимающихся как решением фундаментальных проблем, так и разработкой приборов наноэлектроники. Эти уравнения служат мостом между «атомом и транзистором», т. е. между миром, где действуют обратимые законы, описываемые с помощью гамильтониана, и миром необратимым с его диссипативными процессами.

Следует отметить еще одну примечательную особенность книги: результаты развивающейся теории в ней, как правило, иллюстрируются численными примерами и задачами. С этой целью приведено большое число программ, предназначенных для выполнения расчетов квантовых состояний в низкоразмерных структурах, определения электронных спектров в стандартных полупроводниках, а также для моделирования процессов протекания тока в приборах наноэлектроники. При этом для расчетов используется пакет «MATLAB».

Книга рассчитана на широкий круг читателей, и в том числе на тех, кто впервые знакомится с проблемами нанофизики, — студентов и аспирантов. Она также совершенно необходима любому теоретику и экспериментатору, соприкасающемуся с проблемами физики наноструктур и наноприборов. Можно согласиться с автором книги в том, что многие вопросы, затронутые в ней, должны войти в программы вузовских курсов.

Пользуясь случаем, хочу поблагодарить профессора Суприйо Датта, который написал предисловие к русскому изданию, привел список технических опечаток, допущенных в предыдущем издании книги, и дал ряд ссылок на разработанные им новые информационные ресурсы — лекции и задачи, — к которым открыт свободный доступ в сети Интернет.

*В. Демиховский*

# **Предисловие автора к русскому изданию**

Мне приятно, что моя книга станет доступной моим коллегам и студентам в России, и я благодарен профессору В. Я. Демиховскому и доктору Д. В. Хомицкому, а также Российскому фонду фундаментальных исследований, благодаря которым это стало возможным. Я высоко оцениваю приложенные ими усилия по подготовке русского перевода книги и хотел бы использовать появившуюся возможность для того, чтобы сказать здесь несколько слов о ее содержании, т. е. о квантовом транспорте.

Обычно для описания процесса протекания тока, а также других неравновесных явлений используются уравнения гидродинамического типа, в которых свойства образца характеризуются параметрами объемной структуры, такими как проводимость, диэлектрическая и магнитная проницаемости и др. При этом в решении задачи принимают участие две категории исследователей: физики и материаловеды. Первые занимаются поиском новых эффектов и новых материалов, позволяющих улучшить характеристики новых приборов, а вторые разрабатывают эти приборы. При решении задач эти две группы исследователей, как правило, пользуются различными подходами и в значительной степени различным инструментарием. Так, физики для изучения равновесных или квазиравновесных свойств используют атомистические модели, в то время как инженеры-материаловеды при исследовании сильно неравновесных систем используют в основном макроскопические подходы. Несмотря на то, что со дня изобретения транзистора эти обоснованные подходы успешно работают вот уже на протяжении шестидесяти лет, нам представляется, что подобная ситуация не вполне оправдана в условиях, когда на первое место выходят проблемы, связанные с развитием нанотехнологий.

Рассмотрим *простой пример*, который иллюстрирует рис. 1. Пусть имеется молекула с двумя уровнями энергии, отстоящими друг от друга, к примеру, на расстоянии 1 эВ, и способная поглощать фотоны с такой энергией. Чтобы использовать эту структуру в качестве «ячейки солнечной батареи», преобразующей фотоны с энергией 1 эВ в электрическую энергию, необходимо изготовить «контакты», через которые электроны будут

входить в структуру и выходить из нее. Если контакты соединены со структурой так, что на каждый уровень приходит и уходит одинаковое число электронов (см. рис. 1 *a*), то в силу симметрии системы не следует ожидать появления какого-либо тока во внешней цепи. Действительно, почему электроны должны преимущественно перемещаться слева направо, а не справа налево? Но если изменить схему так, чтобы левый контакт вводил и выводил электроны лишь с нижнего уровня, а правый взаимодействовал только с верхним уровнем (см. рис. 1 *б*), то во внешней цепи возникнет электрический ток, направление которого указано стрелками. Таким образом, в схеме с одной молекулой электрический ток может возникать или отсутствовать в зависимости от способа соединения с контактами.

Сам по себе этот принцип, иллюстрируемый рисунком, не является новым. Он используется в обычных фотогальванических приборах, содержащих *p-n* переходы. Новым является использование подобных принципов в приборах, имеющих масштаб отдельных молекул. Стандартный подход заключался бы в применении для расчета такого макроскопического параметра, как коэффициент поглощения атомных моделей, и далее в использовании гидродинамических уравнений, позволяющих вычислить ток. Я хочу подчеркнуть, что при таком подходе различия в схемах, приведенных на рис. 1 *а* и рис. 1 *б*, не могут быть установлены. *Тем не менее, различие здесь является принципиальным*, так как электрический ток возникает только во второй схеме.

Может показаться, что структура, показанная на рисунке, дает чисто академический пример электронного прибора, поскольку реально работающая ячейка фотоэлемента состоит из огромного числа молекул, и присоединение такого же числа индивидуальных контактов практически неосуществимо. Тем не менее, мы полагаем, что в будущем при исследовании новых физических эффектов, а также в приложениях появятся именно такие схемы, в которых будут присутствовать отдельные наноразмерные контакты в разных точках наноразмерной структуры. Подобным образом работают и существующие в природе многочисленные биологические «приборы». Этот принцип будет использован и в создаваемых человеком электронных приборах завтрашнего дня, что позволит увеличить эффективность современных полуклассических устройств типа солнечных батарей и изучать такие «экзотерические» эффекты как «перепутывание» (entanglement) квантовых состояний.

По мере того, как прогресс в области нанотехнологий открывает двери в новую эру прикладной квантовой механики, становятся все более необходимыми подходы, не требующие обращения к макроскопическим гидродинамическим уравнениям. Нам необходимы современные методы, позволя-

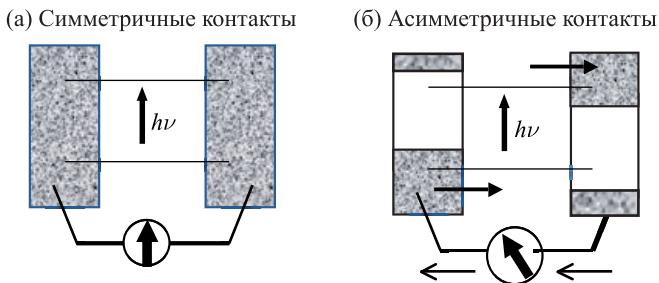


Рис. 1. Простой пример, иллюстрирующий ключевую роль, которую играют «контакты» в наноразмерных электронных приборах. Описать работу подобной структуры, имеющей масштаб порядка нанометра, с помощью макроскопических гидродинамических уравнений невозможно

ющие описывать структуры, взаимодействующие с контактами непосредственно на масштабах порядка нанометра.

*Метод неравновесных функций Грина (НРФГ)–Ландауэра*, разработанный группой Университета Пюрдью, представляет собой один из подходов к этой задаче. Он широко используется исследователями, занимающимися наноэлектроникой, и может применяться к широкому классу нанообъектов, начиная от молекул и заканчивая структурами с баллистическим или диффузионным транспортом. Основные идеи этого подхода я старался изложить в данной книге. В моей более ранней книге «Электронный транспорт в мезоскопических системах» (ЭТМС) (Издательство Университета Кембриджа, 1995) акцент был сделан на использовании метода Ландауэра, который сопоставлялся с методом НРФГ лишь в последней главе. В настоящей книге главным является именно полное описание метода НРФГ–Ландауэра. Кроме этого книга содержит главы (со второй по шестую), посвященные традиционным, более привычным для подобного курса темам, таким как зонная структура полупроводников и других материалов. Эти главы будут полезны и студентам младших курсов.

Хотя, начиная с классических работ Швингера, Каданова, Бейма, Келдыша и других исследователей, выполненных в 1960-х годах, формализм метода НРФГ широко используется в физике конденсированного состояния, я хотел бы отметить, что разработанный мною метод изложения носит несколько отличный характер. Он не имеет аналогов в существующей учебной литературе. Основные уравнения (11.1.1) и (11.1.2), используемые

в этой книге, являются стационарными вариантами полных, зависящих от времени уравнений, приведенных в Приложении, (П.3.6.):

$$G = [EI - H - \Sigma]^{-1} \quad (1)$$

и

$$G^n = G\Sigma^{in}G^+, \quad (2)$$

где в стационарном состоянии  $G \equiv G^R$ ,  $\Sigma \equiv \Sigma^R$  и  $\Sigma^{in} \equiv I\Sigma^<$ .

Данные уравнения фактически совпадают с *уравнениями (75)–(77)* из работы Л. В. Келдыша «Диаграммная техника для неравновесных процессов», опубликованной в журнале ЖЭТФ (т. 47, с. 1515 (1964)). Они были получены с помощью теории возмущений, развитой для системы многих частиц. Построенная Л. В. Келдышем диаграммная техника предназначена для расчета неравновесных процессов; она позволяет находить собственно-энергетические функции  $\Sigma^R$  и  $\Sigma^<$ . Авторы большинства работ, посвященных методу НРФГ, следуют именно этому подходу.

В данной книге основные уравнения (1) и (2) получены из простых соображений, в которых используется одиночестичное уравнение Шредингера. Для более продвинутого читателя в книгу включено Приложение, в котором используется техника операторов вторичного квантования и приведен более строгий вывод, не требующий, однако, использования полного набора методов теории возмущений теории систем многих частиц. Что касается собственно-энергетических функций, то здесь наше изложение не выходит за рамки самосогласованного борновского приближения. Такое приближение является достаточным для адекватного описания многих задач, интересующих как физиков, так и инженеров, по крайней мере, пока не затрагиваются сильные корреляционные эффекты, рассмотрение которых, так или иначе, требует применения методов, выходящих за рамки теории возмущений. Независимо от способа получения величин  $\Sigma$  и  $\Sigma^{in}$  мы рассматриваем уравнения (1) и (2) как соотношения, отражающие основную идею метода НРФГ, подобно тому, как основной смысл приближения Больцмана содержится в уравнении

$$(\vec{\nu} \cdot \vec{\nabla} f) + (\vec{F} \cdot \vec{\nabla}_k f) = S_{op}f \quad (3)$$

независимо от того, каким способом получено выражение для оператора рассеяния (интеграла столкновений)  $S_{op}$ . Метод НРФГ учит нас тому, каким образом следует сочетать описание квантовой динамики с наличием «контактов», подобно тому, как приближение Больцмана учит нас сочетанию принципов классической динамики с наличием тех же «контактов», т. е. всевозможных необратимых процессов.

Так или иначе, но, по нашему мнению, метод НРФГ должен более широко применяться в учебных курсах по физике конденсированного состояния и не выглядеть неким эзотерическим инструментом, доступным только узким специалистам. Надеюсь, что наш оригинальный подход поможет сделать метод НРФГ частью стандартного учебного плана как для физиков, так и для инженеров, работающих сnanoструктурами. Наша цель состоит в том, чтобы эти специалисты смогли успешно применять этот метод для решения широкого класса фундаментальных и прикладных задач, в которых речь идет о «присоединении контактов к системе, описываемой уравнением Шредингера» так, как это иллюстрирует рис. 1.

Имея в виду эту стратегию, мы разработали два дополнительных курса: «Основы наноэлектроники» и «Квантовый транспорт», первый из которых рассчитан на студентов младших, а второй — на студентов старших курсов, причем у слушателей не предполагается владения какими-либо методами, выходящими за рамки линейной алгебры. Тематика этих курсов частично перекрывается, но в целом они разработаны как дополняющие друг друга и могут изучаться в любом порядке. Лекции по этим курсам записаны на видео в рамках ресурсов Системы Вычислительной Нанотехнологии (СВН) и в настоящее время являются доступными для пользователей Интернета через доступ к ресурсу nanoХАБ этой системы. Интересующийся читатель может отыскать данные лекции вместе с соответствующими домашними заданиями и упражнениями, обратившись по следующим Интернет-адресам:

<http://www.nanohub.org/resources/5346>

<http://www.nanohub.org/resources/6218>

Задача о протекании тока затрагивает один из наиболее фундаментальных вопросов физики, относящийся к природе «трения» в микроскопических системах, а также к появлению необратимости из изначально обратимых во времени законов. В этой связи мы выражаем уверенность в том, что метод НРФГ–Ландауэра представляет интерес и для значительно более широкого круга читателей, помимо тех, кто специализируется в области наноэлектроники. Имея это в виду, я пытался донести основные идеи книги до возможно более широкой аудитории, предполагая наличие у нее лишь основных базовых знаний. В дополнение я хочу упомянуть также свои лекции, прочитанные на пятидневной летней школе «Наноэлектроника и значение понятия “сопротивления”», прошедшей в июле 2008 года. Читатель может найти их в Интернете по адресу

<http://www.nanohub.org/resources/5279>

Я постоянно работаю над расширением и дополнением этих лекций, в частности, разработкой их более подробного текстового сопровождения. Данный проект составляет часть новой образовательной программы «Электроника: подход от малого к большому», поддерживаемой Национальным научным фондом, Фондом корпорации «Интел». В реализации этого проекта мне помогают мои коллеги Марк Лундстрем и Эшраф Элам.

С наилучшими пожеланиями всем моим читателям и с благодарностью за их интерес и внимание,

Суприйо Датта,  
Вэст Лафайет, январь 2009 г.

# Предисловие

В этой книге рассматриваются основные представления и методы микроскопической теории вещества, причем, основное внимание уделено изучению электронных состояний, а также проблемам, связанным с протеканием тока. Еще сто лет тому назад атомистическая концепция казалась отнюдь не очевидной и многие известные в те годы ученые сомневались в реальной пользе принципов, постулирующих существование атомов, которые нельзя увидеть<sup>1</sup>. Никто, однако, не предполагал, что к концу двадцатого века ученые смогут «увидеть» атомы, сфотографировать их и даже построить «nanoструктуры» с размерами порядка нанометра<sup>2</sup>. Свойства таких nanoструктур нельзя объяснить на языке макроскопической физики, использующей понятия подвижности или диффузии. Здесь необходимы атомистические, или микроскопические, подходы.

Микроскопическая теория материи начала развиваться в двадцатом веке вместе с квантовой механикой и постепенно приобрела черты инженерной дисциплины. Эта теория используется, например, при изучении структуры новых материалов, их электрических и механических свойств, а также скоростей химических реакций. В этой книге, в основном, рассматриваются задачи, связанные с протеканием тока через nanoструктуры, к которым приложена разность потенциалов. По мере уменьшения размеров электронных приборов до атомных масштабов эти задачи приобретают все большую практическую значимость. Данная область науки и техники быстро развивается, и поэтому конкретные примеры, рассмотренные здесь, могут сохранить, а могут и потерять свою актуальность по истечении, скажем, двадцати лет. Тем не менее, обсуждаемые в этой книге вопросы и, в том числе, задача о протекании тока имеют непосредственное отношение к фундаментальным понятиям квантовой механики и неравновесной статистической физики. Они будут определять принцип работы наноразмерных приборов в течение многих последующих лет. Настоящая книга, в основном, построена как учебник, в котором обсуждаются проблемы, иллюстрирующие основные физические принципы, а не как адресованная исследователям монография, в которой анализируются реальные задачи.

---

<sup>1</sup>Интересный исторический обзор этой проблемы можно найти в [32].

<sup>2</sup>Расстояние между атомами составляет приблизительно 0.25 нм.

В теории электрического транспорта используются не только идеи и методы квантовой механики, но в равной мере и представления неравновесной статистической механики. Эти разделы традиционно излагаются в нескольких физических и химических учебных курсах, которые читаются в течение нескольких семестров. При работе над этой книгой я стремился сконцентрировать основные положения теории электрического транспорта в объеме одного учебника, который студенты старших курсов могут усвоить в течение одного семестра. При этом часть параграфов книги, отмеченных знаком (\*), адресована студентам средних курсов. При подборе материала предполагалось, что читатель знаком с уравнениями в частных производных и матричной алгеброй, а также имеет навыки работы в среде MATLAB или эквивалентной интерактивной системе компьютерных вычислений.

Первая глава и Приложение отличаются от основного материала книги; они должны объединить отдельные вопросы, рассматриваемые в книге, в единую физическую проблему. В первой главе рассмотрены основные идеи и подходы, используемые для описания квантового транспорта. Этот материал доступен начинающим, тем, кто не имеет предварительной подготовки в области квантовой механики. Приложение, с другой стороны, предназначено более продвинутому читателю и посвящено описанию математического аппарата, на современном языке «вторичного квантования». Обе эти главы построены на материале обзорной статьи [15].

Закончив свою предыдущую книгу «Электронный транспорт в мезоскопических системах» (ЭТМС) [13], я не думал, что когда-либо буду писать еще. Но книга ЭТМС была написана в начале 1990-х годов, когда квантовый транспорт был интересен, в основном, физикам-исследователям. С тех пор масштабы электронных устройств непрерывно уменьшались, их размеры стали измеряться в нанометрах, и квантовый транспорт в таких структурах стал темой, актуальной также для инженеров-электронщиков. Определить основную ориентацию учебника мне помог многолетний друг и коллега Марк Лундстрем, который убедил меня взяться за книгу, адресованную инженерной аудитории. Именно такое изменение предполагаемой аудитории (хотя я надеюсь, что и физики найдут эту книгу полезной) определило, скажем, изменение обозначения электрического заряда с « $e$ » на « $q$ », хотя я не менял обозначение « $i$ » на « $j$ », поскольку уравнение Шредингера со слагаемым вида « $-j\partial\psi/\partial t$ » выглядит просто неправильным!

В любом случае, данная книга значительно отличается от книги ЭТМС. Последняя начиналась с обсуждения приближения эффективной массы, которое считалось уже известным читателю из курса физики твердого тела. В этой книге, в главах от второй до седьмой, я последовательно продви-

гаюсь от описания атома водорода к зонным спектрам  $E(k)$  и приближению эффективной массы. Важно еще и то, что книга ЭТМС посвящена изучению проводимости при малых напряжениях (или «теории линейного отклика»), а также физической интерпретации этих процессов в проводниках малых размеров. Акцент был сделан на методе коэффициентов пропускания. В настоящей книге (глава 1 и главы 8–11) я обращал внимание на анализ полных вольт-амперных характеристик, а также на проведении самосогласованных расчетов. При этом в соответствующих местах делаются ссылки на метод коэффициентов пропускания: обсуждается формула Ландауэра и зонды Бюттикера. Но основное место в книге отведено методу неравновесных функций Грина, который, как я убежден, дает наиболее строгое описание транспорта в квантовых структурах и позволяет оценить адекватность других упрощенных методов. Этот подход соединяет модель полностью когерентного транспорта, развитую в мезоскопической физике с моделью некогерентного транспорта в электронных приборах, основанной на уравнении Больцмана.

Метод неравновесных функций Грина обычно излагается на весьма абстрактном языке теории систем многих частиц, однако в этой книге я попытался изложить его на языке, более доступном для широкой аудитории. В своем наиболее простом варианте этот метод сводится к уравнениям баланса в одноуровневой системе, который может быть рассказан даже студентам младших курсов. Именно поэтому я начинаю изложение в главе 1 с «вводной» версии теории, которая затем развивается в полноценное матричное описание и иллюстрируется примерами в главе 11. Обоснование этого аппарата на языке вторичного квантования дано в Приложениях. Можно сказать, что стиль этого учебника существенно отличается от стиля книги ЭТМС, которая базируется, в основном, на методе коэффициентов пропускания и содержит лишь краткое упоминание в последней главе о методе неравновесных функций Грина.

Эта книга отличается от предыдущих еще и тем, что в ней широко использована среда для вычислений MATLAB. Чтобы проиллюстрировать результаты теории, я привожу многочисленные численные примеры и соответствующие программы в пакете MATLAB. Все программы собраны в заключительном разделе книги, но их можно загрузить и с моего Интернет-сайта<sup>3</sup>. Настоятельно рекомендую читателям разрабатывать свои собственные компьютерные программы и воспроизвести представленные в книге результаты. Подобный практический опыт поможет за короткое время овладеть описанными в книге разнообразными и важными понятиями и методами.

---

<sup>3</sup><http://dynamo.ecn.purdue.edu/~datta>

Дополнительные задачи, облегчающие усвоение теоретического материала, размещены на моем Интернет-сайте, и я буду рад поделиться своими решениями с заинтересованными читателями. Планирую пополнять этот список новыми задачами и приглашаю читателей поделиться своими задачами с остальными. Был бы рад ускорить этот процесс, добавляя ссылки на соответствующие сайты.

Эта книга возникла в результате чтения на протяжении ряда лет лекций для студентов старших, а с недавнего времени младших и средних курсов. Читателю будет полезно ознакомиться и с видеолекциями (презентациями), посвященными главным разделам данной книги. Благодаря поддержке Электронного центра Университета Пюрдью, Сети вычислительной нанотехнологии Национального научного фонда, а также Института наноэлектроники и компьютерных вычислений НАСА эти видеолекции открыты для свободного доступа в сети Интернет.

## **Благодарности**

Я хотел бы выразить свою признательность очень широкому кругу людей. Не буду перечислять их всех поименно, замечу лишь, что на протяжении ряда лет я имел удовольствие общаться со многими блестящими студентами и выдающимися коллегами в Университете Пюрдью, а также в других местах. Им, семье, поддерживающей меня, и всем тем, кто помогал мне учиться, я посвящаю эту книгу.

# Список обозначений

## Фундаментальные константы

$q$	заряд электрона	$1.602 \times 10^{-19}$ Кл
$h$	постоянная Планка	$6.626 \times 10^{-34}$ Дж·сек
$\hbar$	$h/2pi$	$1.055 \times 10^{-34}$ Дж·сек
$m$	масса свободного электрона	$9.11 \times 10^{-31}$ кг
$\epsilon_0$	диэлектрическая постоянная вакуума	$8.854 \times 10^{-12}$ Ф/м
$a_0$	боровский радиус, $4\pi\epsilon_0\hbar^2/mq^2$	0.0529 нм
$G_0 = q^2/h$	квант проводимости	$38.7 \times 10^{-6} S$ ( $S = 1/\Omega = A/V$ ) = $= 1/(25.8 \times 10^3 \Omega)$

В книге мы будем использовать рационализированную систему МКС, где энергия будет измеряться в электрон-вольтах:  $1 \text{ эВ} = 1.602 \times 10^{-19} \text{ Дж}$ .

Энергия  $E_0 = q^2/8\pi\epsilon_0 a_0$  в этой системе равна 13.6 эВ.

## Некоторые другие используемые обозначения

$I$	ток (внешний)	Амперы (A)
$J$	ток (внутренний)	Амперы (A)
$V$	напряжение	Вольты (B)
$R$	сопротивление	Омы ( $\Omega = B/A$ )
$G$	проводимость	Сименс ( $S = A/B$ )
$a$	постоянная решетки	метры (m)
$t_0$	$\hbar^2/sm^*a^2$	электрон-вольты (эВ)
$t$	время	секунды (c)

$m^*$	эффективная масса	килограммы (кг)
$m_c$	эффективная масса в зоне проводимости	килограммы (кг)
$\gamma_{1,2,3}$	параметры Латинжера	безразмерные
$\vec{F}$	электрическое поле	В/м
$L$	длина канала	м
$S$	площадь поперечного сечения	м <sup>2</sup>
$\vec{k}$	волновой вектор	/м
$\vec{v}$	скорость	м/сек
$n_s$	электронная плотность на единицу площади	/м <sup>2</sup>
$n_L$	электронная плотность на единицу длины	/м
$N$	число электронов или число фотонов	безразмерное
$\rho$	матрица плотности	безразмерная
$\varepsilon$	энергия уровня	эВ
$H$	гамильтониан	эВ
$U$	самосогласованный потенциал	эВ
$E$	энергия	эВ
$\mu$	электрохимический потенциал	эВ
$f(E)$	распределение Ферми	/эВ
$n(E)$	электронная плотность на единицу энергии	/эВ
$D(E)$	плотность состояний	/эВ
$A(E)$	спектральная функция	/эВ
$G^n(E)$	(то же, что и $-iG^<$ ) корреляционная функция	/эВ
$G^p(E)$	(то же, что и $+iG^<$ ) корреляционная функция	/эВ
$G(E)$	функция Грина (запаздывающая)	/эВ
$\bar{T}(E)$	функция пропускания	безразмерная
$T(E)$	вероятность пропускания (< 1)	безразмерная
$\gamma, \Gamma(E)$	уширение	эВ

24 ◇ Список обозначений

$\Sigma(E)$	собственная энергетическая функция эВ (запаздывающая)	
$\Sigma^{\text{in}}(E)$	(то же, что $-i\Sigma^<$ ) функция для рассеяния внутрь структуры	эВ
$\vartheta(E)$	единичная степ-функция	$\begin{cases} = 1, & E > 0 \\ = 0, & E < 0 \end{cases}$ безразмерная
$\delta(E)$	дельта-функция Дирака	/эВ
$\delta_{nm}$	дельта-символ Кронекера	$\begin{cases} = 1, & n = m \\ = 0, & n \neq m \end{cases}$ безразмерный
+	операция, обозначающая транспонирование и сопряжение (эрмитово сопряжение)	
T	операция, обозначающая транспонирование	

## Глава 1

# Пролог. Электрическое сопротивление (квантовый подход)

Перед читателем не вполне обычный пролог к книге. По размерам его можно сравнить с отдельной главой и связано это с тем, что автор стремился обсудить все основные положения современной теории, которые используются при описании электрического транспорта в малых объектах. Фактически речь идет о структурах, в которых электронный спектр состоит лишь из одного или нескольких уровней, лежащих в интересующем нас интервале энергий.

Замечательно то, что такое обсуждение можно провести, не привлекая каких-либо фундаментальных сведений из квантовой механики. Знание квантовой механики в полном объеме нам будет необходимо для понимания того, каким образом возникают энергетические уровни, а также для описания массивных проводников, содержащих большое количество энергетических уровней. Прежде чем мы с головой погрузимся в эти детали (а для этого у нас есть целая книга!), предварительно будет полезно понять смысл тех факторов, которые влияют на вольт-амперные характеристики действительно малых объектов.

При изучении электрического сопротивления малых объектов развивающийся в книге подход «от малого к большому» отличается от стандартного подхода «от большого к малому». На младших курсах мы начинаем изучение этой проблемы с утверждения о том, что кондактанс  $G$  (величина, обратная сопротивлению) большого макроскопического проводника прямо пропорционален площади его поперечного сечения  $A$  и обратно пропорционален длине  $L$ :

$$G = \sigma A / L \quad (\text{закон Ома}),$$

где удельная проводимость  $\sigma$  является характеристикой материала проводника. Позднее, на старших курсах, мы узнаем о факторах, определяющих проводимость, и лишь после этого, в конце концов, задумываемся о том, что происходит в том случае, когда проводник уменьшается настолько, что

становится невозможным определить его проводимость. Я уверен, что подобный подход «от большого к малому» сложился исторически. Действительно, до недавнего времени никто не имел четкого представления о том, как описать кондактанс очень маленького объекта, если вообще можно говорить о кондактансе микроскопически малых объектов. Чтобы провести измерение кондактанса малого объекта, к нему необходимо присоединить два больших контакта, которые соединяют его с батареей. До конца прошлого века мы не знали, как присоединить контакты к столь малому объекту как, например, молекула, и не было ясно, чему равен кондактанс такого объекта. Сегодня ответы на подобные вопросы кажутся достаточно простыми, за исключением, может быть, некоторых сравнительно редких явлений, таких как, например, эффект Кондо, которые наблюдаются лишь в определенном диапазоне параметров. Весьма вероятно, что в будущих экспериментах с микроскопическими проводниками будет обнаружено еще много новых эффектов и развивающиеся в книге представления, безусловно, не являются последним словом. Но, я думаю, они должны стать «первым словом», поскольку традиционный подход «от большого к малому» скрывает многое из простой физики очень маленьких проводников.

Типичная структура, к обсуждению которой мы будем часто обращаться в этой книге, — это простой вариант «нанотранзистора», состоящего из полупроводникового канала, отделенного от металлического затвора слоем изолятора (обычно двуокиси кремния) (рис. 1.1). Области, обозначенные на рисунке как «исток» и «сток», представляют собой две контактные площадки, которые обладают высокой проводимостью. Сопротивление канала определяет ток, протекающий от истока к стоку, когда между ними приложено электрическое напряжение  $V_D$ . Напряжение на затворе  $V_G$  используется для управления концентрацией электронов в канале и, следовательно, его сопротивлением. Подобный механизм лежит в основе работы любого полевого транзистора (FET/ПТ), хотя в разных версиях детали могут отличаться. В последние годы длина канала  $L$  в выпускаемых полевых транзисторах постоянно уменьшалась от значения 10 мкм в 1960 г. до 0,1 мкм в 2000 г. Это позволило увеличить за указанный период плотность монтажа транзисторов в  $(100)^2 = 10\,000$  раз и в итоге получить увеличение вычислительной мощности процессоров. Такое увеличение плотности упаковки является главной движущей силой компьютерной революции. Как долго еще будет продолжаться это уменьшение размеров? На самом деле никто не знает. Но несомненно одно: независимо от того, что будут представлять собой будущие электронные приборы, мы должны будем научиться описывать их электронные характеристики и моделировать их работу. Примеры, которые я буду использовать в этой книге, лет через двадцать

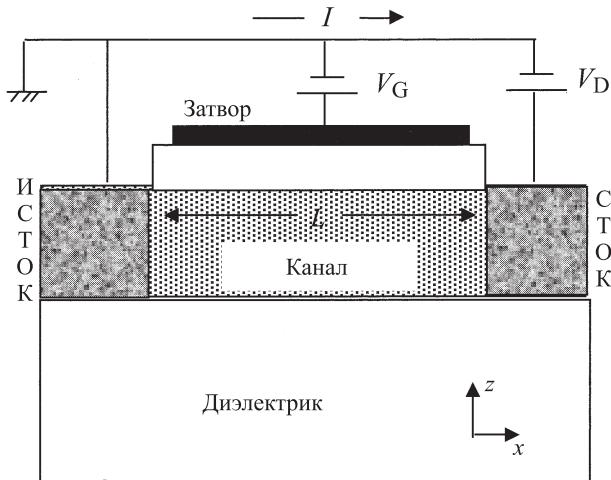


Рис. 1.1. Схема наноразмерного полевого транзистора. Слой изолятора должен быть достаточно толстым для того, чтобы предотвратить протекания тока в контакт затвора, но при этом быть достаточно тонким для того, чтобы, изменения напряжение на затворе, можно было бы эффективно управлять электронной концентрацией в канале

могут приобрести большое значение, а может быть и нет. Однако проблема протекания электрического тока затрагивает некоторые наиболее общие вопросы физики, имеющие отношение к природе необратимости в микроскопическом масштабе. Понятия, которые мы будем обсуждать в этой книге, являются ключевыми и фундаментальными понятиями как квантовой механики, так и неравновесной статистической механики. Вполне вероятно, что многие годы эти понятия будут иметь прямое отношение к анализу и конструированию наноприборов.

**Обзор содержания.** Чтобы смоделировать протекание электрического тока, необходимо, прежде всего, рассмотреть равновесную диаграмму энергетических уровней и определить положение электрохимического потенциала или уровня Ферми  $\mu$ . При этом можно на время забыть о контактах — истоке и стоке (параграф 1.1). Ток возникает, если некоторое внешнее устройство (батарея) поддерживает два этих контакта при разных значениях электрохимических потенциалов  $\mu_1$  и  $\mu_2$ , что приводит канал в неравновесное состояние (параграф 1.2). Сила тока, протекающего через микроскопическое устройство лишь с одним энергетическим уровнем в интересующем

нас интервале, вычисляется очень просто и, как можно было ожидать, зависит от качества контактов. Но совершенно неочевидно то, что для канала с одним уровнем (в интересующем нас энергетическом интервале) существует максимальный кондактанс, который является фундаментальной постоянной, зависящей от заряда электрона и постоянной Планка:

$$G_0 \equiv q^2/h = 38.7\mu S = (25.8k\Omega)^{-1}. \quad (1.1)$$

Заметим, что это обстоятельство не принималось во внимание вплоть до конца 1980-х гг.

Фактически, микроскопические каналы обычно содержат два вырожденных уровня с одинаковой энергией, один из которых отвечает проекции спина «вверх», а второй отвечает проекции спина «вниз». В этой ситуации максимальный кондактанс становится равным  $2G_0$ . Если контакты плохие, мы всегда будем измерять меньший кондактанс. Но суть в том, что существует верхний предел кондактанса, т. е. максимум того, что можно измерить, располагая даже самыми превосходными контактами (параграф 1.3). В параграфе 1.4 я объясню важность той роли, которую играют кулоновское отталкивание и законы электростатики в целом в определении формы вольт-амперных ( $I - V$ ) характеристик, а также как эти аспекты связаны с уравнениями квантового транспорта. Приняв во внимание все перечисленные факторы, мы сможем описать процесс протекания тока в одноуровневом канале, сильно взаимодействующем с контактами. Но если канал связан слабо, появляются некоторые дополнительные физические аспекты, которые мы обсудим в параграфе 1.5. Наконец, в параграфе 1.6 я объясню, каким образом одноуровневое описание распространяется на более массивные устройства с большим числом энергетических уровней, и в какой момент начинает работать закон Ома. Это тот самый переход к более крупным устройствам, который требует использования современных концепций квантовой статистической механики, составляющих предмет обсуждения и в остальных главах данной книги.

## 1.1. Диаграмма энергетических уровней

На рисунке 1.1.1 приведены типичные вольт-амперные характеристики полевого транзистора, схема которого представлена на рис. 1.1. Ширину транзистора в направлении оси  $y$ , перпендикулярной плоскости листа, будем считать равной 1 мкм. При малых напряжениях на затворе транзистор находится в запертом состоянии и через него при приложенном напряжении на стоке  $V_D$  протекает очень слабый ток. Если напряжение на затворе

превысит т. н. пороговое значение  $V_T$ , транзистор включается и сила тока в состоянии «включено» растет вместе с напряжением на затворе  $V_G$ . При фиксированном напряжении на затворе сила тока  $I$  сначала возрастает вместе с напряжением стока, но затем стремится к насыщению. Попытаемся понять, почему сила тока начинает возрастать в тот момент, когда напряжение на затворе превысит  $V_T$  (рис. 1.1.1 а).

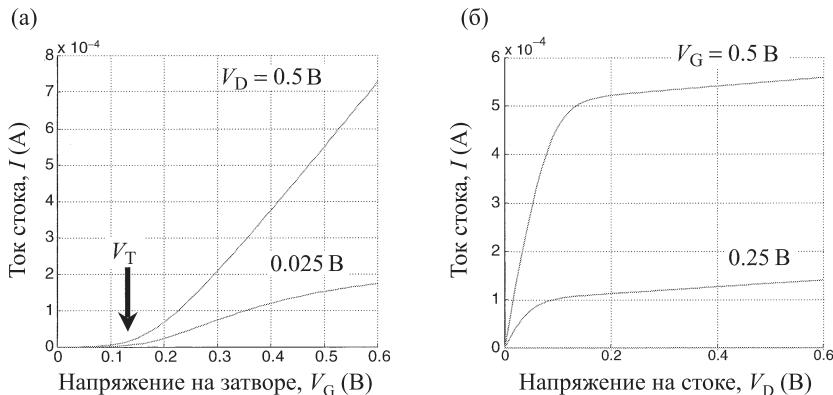
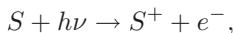


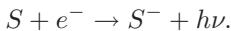
Рис. 1.1.1. (а) Ток стока  $I$ , как функция напряжения на затворе  $V_G$ , при разных величинах напряжения стока  $V_D$ . (б) Ток стока, построенный как функция напряжения стока, при разных величинах напряжения на затворе

При анализе работы любой неоднородной структуры или прибора (подобной структуре, изображенной на рис. 1.1) необходимо, прежде всего, рассмотреть соответствующую равновесную диаграмму энергетических уровней, предполагая равенство нулю напряжения между истоком и стоком. Электроны в полупроводнике заполняют энергетические уровни, которые формируют энергетические зоны, как это показано на рис. 1.1.2. Заметим, что в существовании энергетических зон можно убедиться с помощью экспериментов, позволяющих определять положения заполненных уровней из измерений минимальной энергии фотона, необходимой для того, чтобы выбить электрон в вакуум (эксперименты по фотоэмиссии (РЕ)). Схематически мы можем описать этот процесс следующим образом:



где « $S$ » представляет полупроводниковое устройство (или, в сущности, любой материал!).

Конечно, подобным образом невозможно определить положение пустых уровней, поскольку они не содержат электронов, которые можно было бы выбить. Для этого необходим эксперимент по обратной фотоэмиссии, в котором падающие электроны поглощаются с излучением фотонов:



Существуют и другие эксперименты, и в том числе основанные на измерении поглощения света, которые дают информацию о расположении энергетических уровней. Все подобные эксперименты давали бы эквивалентную информацию, если бы электроны не взаимодействовали друг с другом, и мы могли бы выбивать один электрон, не оказывая никакого влияния на его окружение. Однако в реальном мире необходимы весьма нетривиальные рассуждения, позволяющие связать измеряемые энергии с величинами, используемыми в тех или иных теоретических моделях. В главе 3 мы обсудим некоторые из этих проблем.



Рис. 1.1.2. Разрешенные уровни энергии, которые могут быть заняты электронами, в активной части прибора, как, например, канал на рис. 1.1. Положительное напряжение на затворе  $V_G$  смещает энергетические уровни вниз, в то время как значение электрохимического потенциала  $\mu$  зафиксировано контактами истока и стока, находящимися в равновесии друг с другом ( $V_D = 0$ )

Предположим, что в больших контактных областях (обозначенных на рис. 1.1 как исток и сток) энергетические уровни квантовых состояний имеют непрерывное распределение. Это будет справедливо, если контакты металлические, но не всегда оказывается правильным, если мы имеем дело с полупроводниковыми контактами. В них могут возникать различные интересные эффекты (см. упражнение У.1.4), и в том числе такие как снижение силы тока при повышении напряжения (или «отрицательное дифференциальное сопротивлением» (ОДС/NDR)); однако в нашем обсуждении мы не будем рассматривать такие возможности. Разрешенные состояния могут быть заполнены вплоть до некоторой энергии  $\mu$  (называемой «электрохимическим потенциалом»). Последняя также может быть определена с помощью фотоэмиссионных измерений, в которых определяется работа выхода, т. е. минимальная энергия фотона, необходимая для того, чтобы выбрать фотозелектрон из металла. Работа выхода определяет, насколько электрохимический потенциал  $\mu$  находится ниже уровня вакуума.

**Функция Ферми.** Если исток и сток соединены с каналом при  $V_D$ , равном нулю, то электроны будут втекать в устройство и вытекать из него, приводя все области в равновесие с общим электрохимическим потенциалом  $\mu$ , подобно тому как два материала в равновесии достигают общей температуры  $T$ . В таком равновесном состоянии (по прошествии длительного времени) среднее число электронов на любом энергетическом уровне обычно не является целым числом, а определяется функцией Ферми:

$$f_0(E - \mu) = \frac{1}{1 + \exp[(E - \mu)/k_B T]}. \quad (1.1.1)$$

Энергетические уровни, расположенные много ниже  $\mu$ , всегда заполнены, так что  $f_0 = 1$ , тогда как энергетические уровни, расположенные много выше  $\mu$ , всегда не заняты и  $f_0 = 0$ . Та часть энергетических уровней, расположенных в пределах нескольких  $k_B T$  в окрестности  $\mu$ , заполнена с некоторой конечной вероятностью, а другая — пуста, так что среднее число электронов лежит между 0 и 1:  $0 \leq f_0 \leq 1$  (рис. 1.1.3). Обратите внимание на то, что это число не может превышать единицу, поскольку принцип Паули запрещает заполнение уровня более чем одним электроном.

**Проводимость  $n$ -типа.** Приложенное к затвору положительное напряжение  $V_G$  понижает энергетические уровни в канале. Однако при этом положение уровней энергии в истоке и стоке не меняется и, следовательно, электрохимический потенциал  $\mu$  (который должен быть всюду одинаков) остается неподвижным и переводит транзистор в открытое состояние. Дей-

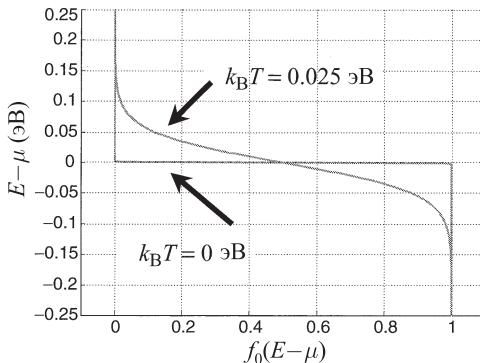


Рис. 1.1.3. Функция Ферми (уравнение (1.1.1)), среднее число электронов, заполняющих состояние с энергией  $E$ , если оно находится в равновесии с большим контактом («резервуаром»), имеющим электрохимический потенциал  $\mu$

ствительно, как мы увидим в следующем разделе, электрический ток, возникающий под влиянием внешнего напряжения, зависит от числа энергетических уровней, находящихся в окрестности  $E = \mu$ . Таким образом, пороговое напряжение на затворе  $V_T$ , необходимое для перевода транзистора в открытое состояние, определяется разностью энергий между равновесным электрохимическим потенциалом  $\mu$  и самым нижним доступным пустым состоянием или нижней границей зоны проводимости (рис. 1.1.2).

**Проводимость  $p$ -типа.** Обратите внимание на то, что количество электронов в канале не является фактором, однозначно определяющим протекание электрического тока. В частности, отрицательное напряжение на затворе ( $V_G < 0$ ) снижает количество электронов в канале. Тем не менее, коль скоро электрохимический потенциал переместится в заполненную зону, как это показано на рис. 1.1.4, канал станет более проводящим, благодаря наличию состояний (заполненных или нет) в окрестности  $E = \mu$ . Это пример проводимости  $p$ -типа или «дырочной» проводимости в отличие от проводимости  $n$ -типа или электронной проводимости, механизм которой иллюстрирует рис. 1.1.2. Итак, для появления электрического тока необходимо наличие состояний вблизи  $E = \mu$ , но они не обязательно должны быть пустыми. Заполненные состояния также хороши, и определить с помощью этого эксперимента, является ли проводимость  $n$ -типа (рис. 1.1.2) или  $p$ -типа (рис. 1.1.4), невозможно. К этому вопросу мы еще раз вернемся в параграфе 1.2, где будем обсуждать, как в ответ на приложенное к контактам истока и стока напряжение возникает электрический ток.

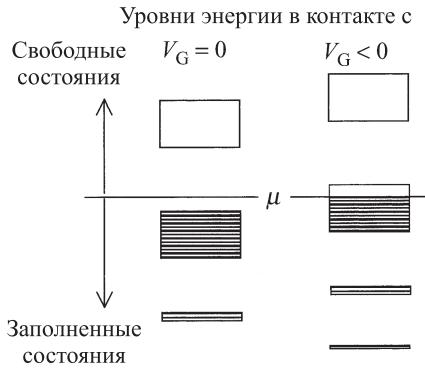


Рис. 1.1.4. Механизм проводимости *p*-типа или дырочной проводимости. Отрицательное напряжение на затворе ( $V_G < 0$ ) уменьшает количество электронов в канале. Тем не менее, канал становится проводящим, как только электрохимический потенциал  $\mu$  окажется в заполненной зоне, поскольку проводимость зависит от наличия состояний вблизи  $E = \mu$ , а не от общего числа электронов

Рисунки 1.1.2 и 1.1.4 наводят на мысль, что с одним и тем же устройством можно работать как с прибором *n*- или *p*-типа, просто изменения полярность напряжения на затворе. Это справедливо для приборов с коротким каналом, если, как мы предполагали, состояния в контактах имеют непрерывное распределение. Но в общем случае это не обязательно так: например, в устройствах с длинным каналом вблизи контактов могут возникать «обедненные области», формы которых могут быть различны в приборах *n*- и *p*-типа.

## 1.2. Как возникает электрический ток?

Ранее мы установили, что проводимость связана с теми электронными состояниями, у которых энергетические уровни лежат в окрестности  $E = \mu$ ; причем, не важно, заполненные это уровни или свободные. Чтобы понять причину этого, рассмотрим механизм переноса электронов от истока к стоку. Пусть батарея понижает уровни энергии электронов в области стока относительно их положения на истоке (это имеет место при условии, что  $V_D$  положительно). Тогда в стационарном состоянии значения электрохимического потенциала на стоке и истоке будут отличаться на  $qV_D$ :

$$\mu_1 - \mu_2 = qV_D, \quad (1.2.1)$$

и соответствующие функции Ферми на разных электродах будут иметь вид:

$$f_1(E) = \frac{1}{1 + \exp[(E - \mu_1)/k_B T]} = f_0(E - \mu_1), \quad (1.2.2a)$$

$$f_2(E) = \frac{1}{1 + \exp[(E - \mu_2)/k_B T]} = f_0(E - \mu_2). \quad (1.2.2b)$$

Каждый контакт «стремится» привести канал в равновесие с собой. При этом электроны истока переходят в канал, но равновесие не может быть реализовано, пока сток продолжает отбирать электроны из канала. Канал, таким образом, участвует в процессе установления баланса между двумя резервуарами с разными значениями химического потенциала и остается в неравновесном состоянии (см. рис. 1.2.1).

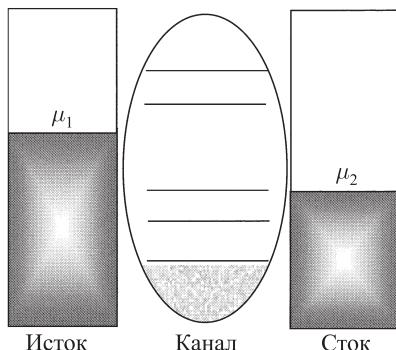


Рис. 1.2.1. Положительное напряжение  $V_D$ , приложенное к стоку, понижает его электрохимический потенциал:  $\mu_1 - \mu_2 = qV_D$ . Контакты истока и стока стремятся установить разные распределения Ферми в канале, и при этом он остается в промежуточном состоянии

**Уравнения баланса для одноуровневой модели.** Процесс установления баланса легко описать, рассматривая простую одноуровневую систему, смешенную таким образом, что ее уровень энергии  $\varepsilon$  находится между электрохимическими потенциалами двух контактов (рис 1.2.2). Равновесию с контактом 1 соответствует функция  $f_1(\varepsilon)$ , тогда как контакту 2 —  $f_2(\varepsilon)$ , где  $f_1$  и  $f_2$  — функции Ферми истока и стока, определяемые уравнениями (1.2.2). Ток  $I_1$  через левый контакт пропорционален ( $f_1 - N$ ) и имеет вид

$$I_1 = \frac{q\gamma_1}{\hbar} (f_1 - N). \quad (1.2.3a)$$

Здесь  $-q$  — заряд электрона и аргумент  $\varepsilon$  опущен. Аналогично, результирующий ток  $I_2$  через правый контакт пропорционален  $(f_2 - N)$  и может быть записан как

$$I_2 = \frac{q\gamma_2}{\hbar}(f_2 - N). \quad (1.2.3b)$$

Константы  $\gamma_1/\hbar$  и  $\gamma_2/\hbar$  определяют скорости ухода электрона, первоначально заполнившего уровень  $\varepsilon$  на контакты истока и стока соответственно. Эти величины имеют размерность обратного времени, так что  $\gamma_1$  и  $\gamma_2$  имеют размерность энергии. В принципе, мы могли бы экспериментально измерить эти величины, которые имеют размерность частоты, так что  $\gamma_1$  и  $\gamma_2$  имеют размерность энергии. В конце этого раздела я добавлю еще несколько слов о физике, выходящей за пределы этих уравнений. Теперь же мы рассмотрим некоторые вытекающие из них следствия.

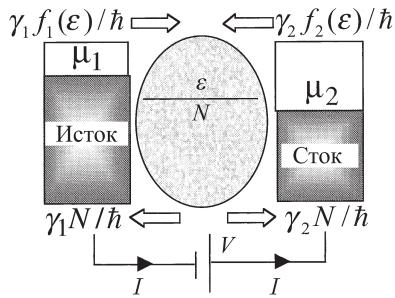


Рис. 1.2.2. Входящие и выходящие токи электронов на границах между одноуровневым каналом и истоком и стоком: простая иллюстрация уравнения баланса

**Ток в одноуровневой модели.** В стационарном состоянии результирующий поток, направленный внутрь канала или из него, равен нулю:  $I_1 + I_2 = 0$ , так что из уравнений (1.2.3a) и (1.2.3b) мы получаем естественный результат:

$$N = \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2}, \quad (1.2.4)$$

т. е. количество электронов  $N$  является средним по контактам 1 и 2 соответственно. Подставляя этот результат в (1.2.3a) или (1.2.3b), мы получаем выражение для тока стационарного состояния:

$$I = I_1 = -I_2 = \frac{q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(\varepsilon) - f_2(\varepsilon)]. \quad (1.2.5)$$

Это ток для одной проекции спина. При наличии двух спиновых состояний с одинаковой энергией данную величину тока следует увеличить вдвое.

Полученный выше простой результат позволяет проиллюстрировать некоторые, важные моменты, касающиеся процесса протекания электрического тока. Во-первых, ток не возникает, если  $f_1(\varepsilon) = f_2(\varepsilon)$ . На уровне, лежащем гораздо ниже обоих электрохимических потенциалов  $\mu_1$  и  $\mu_2$ , имеет место равенство  $f_1(\varepsilon) = f_2(\varepsilon) = 1$ , поэтому этот уровень не дает вклада в электрический ток. Точно так же не вносит вклад в ток уровень, который лежит гораздо выше обоих потенциалов  $\mu_1$  и  $\mu_2$  и имеет  $f_1(\varepsilon) = f_2(\varepsilon) = 0$ . Электрический ток возникает только в том случае, когда уровень находится в пределах нескольких  $k_B T$  в окрестности  $\mu_1$  и  $\mu_2$ , мы имеем  $f_1(\varepsilon) \neq f_2(\varepsilon)$ . Таким образом, протекание электрического тока связано с различными условиями на двух контактах. Контакт 1 продолжает закачивать электроны в канал, стремясь увеличить их число от  $N$  до  $f_1$ , тогда как контакт 2 продолжает их извлекать, пытаясь уменьшить их число до  $f_2$ . Результирующим эффектом является непрерывный перенос электронов от контакта 1 к контакту 2, соответствующий силе тока  $I$  во внешнем контуре (рис. 1.2.2). Обратите внимание на то, что направление тока обратно потоку электронов, поскольку электроны имеют отрицательный заряд.

Теперь становится понятно, почему процесс проводимости требует наличия состояний в окрестности  $E = \mu$ . Свободны (*n*-тип, рис. 1.1.2) или заполнены (*p*-тип, рис. 1.1.4) эти состояния в равновесии, не имеет значения до тех пор, пока не приложено напряжение стока. В случае пустых состояний электроны сначала инжектируются с отрицательного контакта, а затем собираются положительным контактом. В случае заполненных состояний электроны сначала собираются положительным контактом и впоследствии добавляются из отрицательного контакта. В любом случае, мы получаем ток, идущий по внешнему контуру в одном и том же направлении.

**Входящий и исходящий потоки.** Уравнения (1.2.3a) и (1.2.3b) выглядят элементарными и редко вызывают вопросы. Но они скрывают несколько тонких проблем, которые могут озадачить более продвинутого читателя, и поэтому я считаю себя обязанным вкратце остановиться на этих проблемах. Я понимаю, что иду на определенный риск смутить читателей, которые считают, что уже все поняли и спешат перейти к следующей главе, опустив окончание этого раздела.

Правые части уравнений (1.2.3a) и (1.2.3b) можно интерпретировать как разности между входящим и выходящим токами из истока и стока соответственно (см. рис. 1.2.2). Рассмотрим, к примеру, электронные потоки истока. Выражение для выходящего тока  $\gamma_1 N / \hbar$  обосновать легко, посколь-

ку  $\gamma_1/\hbar$  определяет интенсивность (или скорость), с которой электрон, изначально занимавший уровень  $\varepsilon$ , вылетает в контакт истока. Сложнее обосновать выражение для входящего тока  $\gamma_1 f_1/\hbar$ , поскольку в контактах имеется множество электронов в различных состояниях, каждый из которых стремится заполнить одно состояние внутри канала, и совершенно неочевидно, каким образом суммировать входящий ток всех этих состояний. Для этого обычно используют следующие термодинамические аргументы. Если бы канал находился в равновесии с истоком, то результирующий поток был бы равен нулю, т. е. входящий ток был бы равен выходящему току. Но в состоянии равновесия последний равен  $\gamma_1 f_1/\hbar$ , так как среднее число электронов  $N$  равно  $f_1$ . В отсутствии равновесия  $N$  отличается от  $f_1$ , но входящий ток остается неизменным, поскольку он зависит только от условий внутри контактов, которые также остаются неизменными (обратите внимание на то, что выходящий ток изменяется, образуя полный ток, выражение для которого получено выше).

**«Принцип запрета Паули»?** Вдумчивый читатель может не согласиться с тем, что сделанным мной утверждением, а именно, с тем, что входящий ток «зависит только от условий в контактах». Возникает вопрос: не может ли наличие других электронов в канале привести к уменьшению протекающего тока в соответствии с принципом запрета Паули? Можно привести аргумент, заключающийся в том, что входящий и выходящий токи истока определяются соответственно как

$$\gamma_1 f_1(1 - N) \quad \text{и} \quad \gamma_1 N(1 - f_1)$$

вместо величин

$$\gamma_1 f_1 \quad \text{и} \quad \gamma_1 N,$$

указанных на рис. 1.1.2. Нетрудно видеть, что величина полного тока, обусловленного разницей между входящим и выходящим токами, в обоих случаях одна и та же, так что эти рассуждения могут показаться «академическими». Однако совсем не академический интерес представляет эффект уширения уровня, возникающий при взаимодействии с контактами. Именно учет этого уширения, как мы увидим в следующем разделе, необходим для получения количественно правильных результатов. Я предполагаю определять входящий и выходящий токи таким образом, чтобы отнесенный к одному электрону выходящий поток ( $\gamma_1 = \gamma_1 N/N$ ) был равен величине этого уширения. При этом будем иметь в виду, что разность этих потоков равна полному току. Будет ли эта величина, обусловленная связью с истоком, равна  $\gamma_1$  или  $\gamma_1(1 - f_1)$ , зависит от конкретной системы. Можно показать, что до тех пор пока в формирование входящего и выходящего токов не вовлекается процесс релаксации энергии или неупругие взаимодействия, уширение

равно  $\gamma_1$ , независимо от фактора заполнения  $f_1$  в контакте. Мы обсудим этот вопрос в последующих главах 9 и 10, однако соответствующее рассмотрение требует использования описанного в приложении современного формализма.

### 1.3. Квант кондактанса

Рассмотрим устройство, к которому приложено небольшое напряжение, вызывающее расщепление электрохимических потенциалов истока и стока (рис. 1.3.1 *a*). Мы можем записать выражение для силы тока, проходящего через это устройство (1.2.5), и упростить его, предположив, что  $\mu_1 > \varepsilon > \mu_2$ , а температура настолько низка, что  $f_1(\varepsilon) \equiv f_0(E - \mu_1) \approx 1$  и  $f_2(\varepsilon) \equiv f_0(E - \mu_2) \approx 0$ :

$$I = \frac{q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} = \frac{q \gamma_1}{2 \hbar}, \quad \text{если } \gamma_2 = \gamma_1. \quad (1.3.1a)$$

Полученный результат может навести на мысль о том, что можно неограниченно увеличивать силу тока, проходящего через этот одноуровневый прибор, увеличивая величину  $\gamma_1 (= \gamma_2)$ , т. е. усиливая связь прибора с контактами. Однако, согласно одному из главных результатов мезоскопической физики, максимальный кондактанс одноуровневого устройства равен  $G_0$  (см. уравнение (1.1)). Для того чтобы избежать подобного противоречия, следует учесть эффект уширения уровня, которое неизбежно сопровождает любое взаимодействие с ним. Учет уширения должен привести к выходу части уровней за пределы интервала энергии, границами которого являются  $\mu_1$  и  $\mu_2$ , и в котором происходит протекание тока. В результате сила тока уменьшается по сравнению с величиной, определяемой уравнением (1.3.1a), на коэффициент  $(\mu_1 - \mu_2)/C\gamma_1$ , который равен доле уровня, попадающей в «окно» между  $\mu_1$  и  $\mu_2$ . Здесь  $C\gamma_1$  — эффективная ширина уровня, а  $C$  — численная константа. Поскольку  $\mu_1 - \mu_2 = qV_D$ , из уравнения (1.3.1b) следует

$$I = \frac{q\gamma_1}{2\hbar} \frac{qV_D}{C\gamma_1} \rightarrow G = \frac{I}{V_D} = \frac{q^2}{2C\hbar}, \quad (1.3.1b)$$

т. е. кондактанс действительно достигает постоянной величины, независимой от силы взаимодействия канала с контактами ( $\gamma_1 = \gamma_2$ ). Теперь для того, чтобы получить оценку для  $C$ , мы проведем более строгое вычисление.

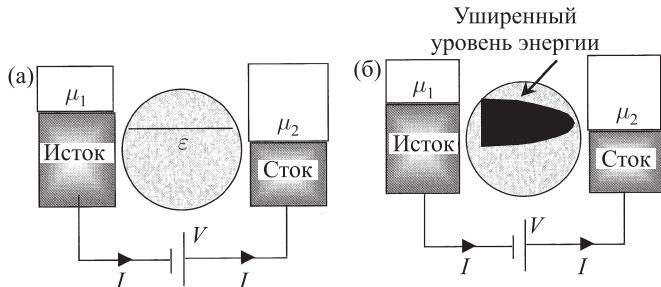


Рис. 1.3.1. (а) Канал с приложенным к нему небольшим напряжением, вызывающим расщепление электрохимических потенциалов истока и стока  $\mu_1 > \varepsilon > \mu_2$ . (б) Взаимодействие канала с контактами неизбежно вызывает уширение уровня и связанное с этим распространение части энергетического уровня за пределы интервала между  $\mu_1$  и  $\mu_2$ , где протекает ток

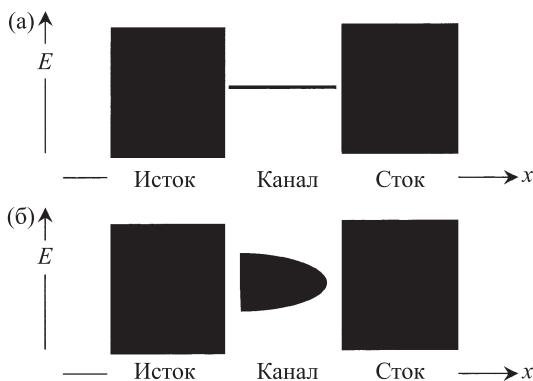


Рис. 1.3.2

Причину этого уширения можно понять, заметив, что до установления связи канала с истоком и стоком плотность состояний (ПС)  $D(E)$  выглядела примерно так, как показано на рис. 1.3.2 а (черный цвет означает высокую плотность состояний). В этом случае мы имеем один дискретный уровень в канале и непрерывное распределение состояний в контактах истока и стока. При включении взаимодействия эти состояния частично «перетекают» в канал и одновременно состояния канала распространяются в контакты. Поскольку потеря происходит при фиксированной энергии, а приобретенные состояния распределяются в некотором интервале энергий, суммарным

эффектом является уширение плотности состояний канала или переход от исходного дельтаобразного распределения к диффузному распределению (как показано на рис. 1.3.2 б). В главе 8 мы убедимся в том, что существует «правило сумм», требующее сохранения полного числа состояний. Поэтому проинтегрированный по всем энергиям уровень по-прежнему соответствует только одному электрону. В принципе, размытая плотность состояний могла бы иметь любую форму, однако в простейшем случае она изображается центрированной при  $E = \varepsilon$  лоренцевской функцией, интеграл от которой по всем энергиям равен единице:

$$D_\varepsilon(E) = \frac{\gamma/2\pi}{(E - \varepsilon)^2 + (\gamma/2)^2}. \quad (1.3.2)$$

Исходную дельтаобразную плотность состояний можно представить как предел выражения  $D_\varepsilon(E)$ , при стремлении ширины уровня  $\gamma \rightarrow 0$  к нулю (рис. 1.3.3), причем ширина уровня  $\gamma$ , как и следовало ожидать, пропорциональна интенсивности связи. Действительно, оказывается, что  $\gamma = \gamma_1 + \gamma_2$ , где  $\gamma_1/\hbar$  и  $\gamma_2/\hbar$  — это введенные в параграфе 1.2 скорости ухода. Это следует из проведенного в главе 8 квантово-механического анализа. На данном этапе мы можем объяснить это как результат действия «принципа неопределенности» для энергии и времени, согласно которому произведение времени жизни ( $= \hbar/\gamma$ ) квантового состояния и неопределенности его энергии ( $\gamma$ ) равно постоянной Планка  $\hbar$ . Следует заметить, что в общем случае форма линии не обязательно является лоренцевской, что обычно отражается в энергетической зависимости функции  $\gamma(E)$ .

В любом случае, основной вывод состоит в том, что взаимодействие с контактами ведет к уширению одного дискретного энергетического уровня, что и приводит к непрерывной плотности состояний, задаваемой уравнением (1.3.2). Мы можем учесть этот эффект, модифицировав наше выражение для силы тока (уравнение (1.2.5))

$$I = \frac{q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(\varepsilon) - f_2(\varepsilon)],$$

для того, чтобы проинтегрировать (или просуммировать) его по распределенным состояниям,  $D_\varepsilon(E)dE$ :

$$I = \frac{q}{\hbar} \int_{-\infty}^{+\infty} dE D_\varepsilon(E) \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(E) - f_2(E)]. \quad (1.3.3)$$

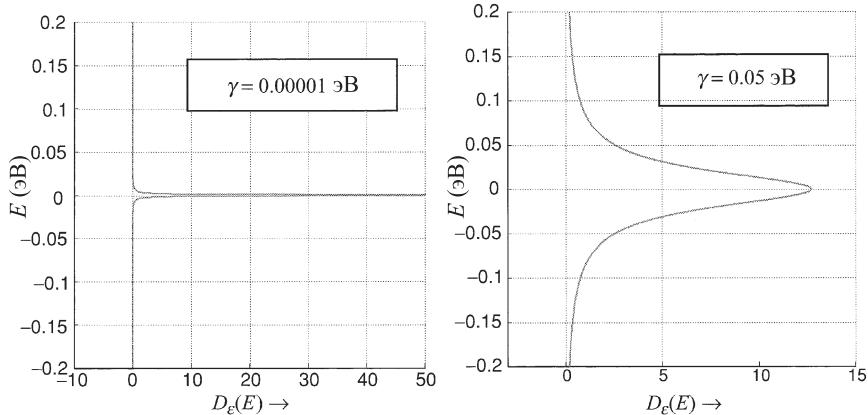


Рис. 1.3.3. При взаимодействии с контактами энергетический уровень при  $E = \varepsilon$  испытывает уширение, которое можно описать, вводя непрерывную плотность состояний  $D_\varepsilon(E)$ . Показаны центрированные при  $\varepsilon = 0$  эВ кривые  $D_\varepsilon(E)$  для двух разных значений константы связи  $\gamma$

При низких температурах можно записать:

$$f_1(E) - f_2(E) = 1, \quad \text{если} \quad \mu_1 > E > \mu_2, \quad = 0 \quad \text{в другом случае},$$

так что величина тока будет определяться следующим уравнением:

$$I = \frac{q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} \int_{\mu_2}^{\mu_1} dE D_\varepsilon(E).$$

Если приложенное напряжение настолько мало, что плотность состояний можно считать постоянной в интервале  $\mu_1 > E > \mu_2$ , мы можем использовать уравнение (1.3.2), чтобы записать

$$I = \frac{q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} (\mu_1 - \mu_2) \frac{(\gamma_1 + \gamma_2)/2\pi}{(\mu - \varepsilon)^2 + (\gamma_1 + \gamma_2)^2}.$$

Максимальная величина тока получается, если энергетический уровень  $\varepsilon$  совпадает с  $\mu$ , равным среднему значению между  $\mu_1$  и  $\mu_2$ . Учитывая,

что  $\mu_1 - \mu_2 = qV_D$ , можно записать максимальный кондактанс как

$$G \equiv \frac{I}{V_D} = \frac{q^2}{h} \frac{4\gamma_1\gamma_2}{\gamma_1 + \gamma_2} = \frac{q^2}{h}, \quad \text{если } \gamma_1 = \gamma_2.$$

Уравнение (1.3.3) для тока представляет собой обобщение нашего предыдущего результата (1.2.5), учитывающего влияние уширения. Подобным образом можно получить обобщение выражения для количества электронов  $N$  (см. уравнение (1.2.4))

$$N = \frac{\gamma_1 f_1(\varepsilon) + \gamma_2 f_2(\varepsilon)}{\gamma_1 + \gamma_2},$$

учитывающее уширение плотности состояний:

$$N = \int_{-\infty}^{+\infty} dE D_\varepsilon(E) \frac{\gamma_1 f_1(E) + \gamma_2 f_2(E)}{\gamma_1 + \gamma_2}. \quad (1.3.4)$$

## 1.4. Профиль распределения потенциала

Физики часто фокусируют внимание на проводимости при малом напряжении, которое определяется исключительно свойствами энергетических уровней в окрестности равновесного электрохимического потенциала  $\mu$ . Это т. н. «линейный отклик». Такой подход не позволяет, однако описать полные вольт-амперные характеристики. Для полного описания необходимо знать потенциал внутри канала, возникающий в ответ на приложенные к внешним электродам — истоку, стоку и затвору — напряжения. Чтобы убедиться в этом, рассмотрим одноуровневый канал с равновесным электрохимическим потенциалом  $\mu$ , расположенным чуть выше энергетического уровня  $\varepsilon$ , как это показано на рис. 1.4.1. Когда мы прикладываем напряжение между истоком и стоком, электрохимические потенциалы расходятся на  $qV$ :  $\mu_1 - \mu_2 = qV$ . Мы знаем, что при низкой температуре ток возникает только в том случае, если уровень  $\varepsilon$  находится между  $\mu_1$  и  $\mu_2$ . В зависимости от того, какое влияние на энергетический уровень  $\varepsilon$  оказывает приложенное напряжение, могут реализоваться различные варианты. Если пренебречь влиянием затвора, то можно ожидать, что внутри канала потенциал будет лежать ровно посередине между его значениями в истоке и стоке,  $\varepsilon \rightarrow \varepsilon - (V/2)$ , как это изображено на рис. 1.4.2 для приложенных положительном и отрицательном напряжениях (обратите внимание на то,

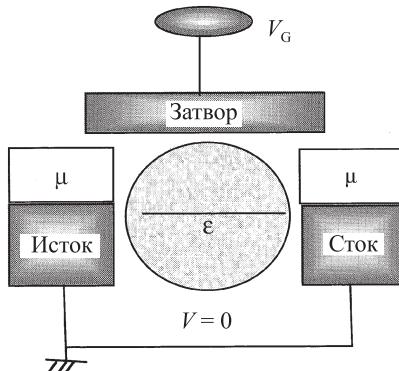
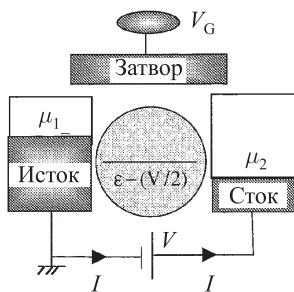
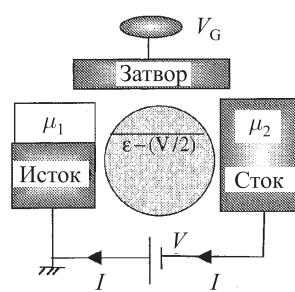


Рис. 1.4.1

что потенциал источника остается постоянным). Очевидно, что при любой полярности смещения ( $V > 0$  или  $V < 0$ ) энергетический уровень лежит посередине между  $\mu_1$  и  $\mu_2$ , что приводит к равным величинам силы тока для  $+V$  и  $-V$ .

(а)  $V > 0$ (б)  $V < 0$ Рис. 1.4.2. Положение энергетических уровней (а) при прямом смещении ( $V > 0$ ) и (б) при обратном смещении ( $V < 0$ ), при условии, что потенциал канала лежит ровно посередине между его значениями в истоке и стоке

Совершенно иная картина возникает, если допустить, что затвор настолько сильно связан с каналом, что энергетический уровень следует за потенциалом затвора и не подвергается влиянию напряжения стока, другими словами,  $\varepsilon$  остается неподвижным относительно электрохимического по-

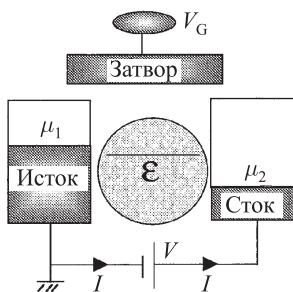
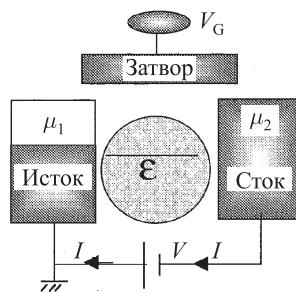
(a)  $V > 0$ (б)  $V < 0$ 

Рис. 1.4.3. Положение энергетических уровней (а) при положительном напряжении ( $V > 0$ ) и (б) при отрицательном напряжении ( $V < 0$ ), при условии, что потенциал канала остается неподвижным относительно электрохимического потенциала истока

тенциала истока (рис. 1.4.3). Тогда при положительном смещении ( $V > 0$ ) энергетический уровень находится между  $\mu_1$  и  $\mu_2$ , а при отрицательном ( $V < 0$ ) — за пределами этого интервала. В этой ситуации вольт-амперная характеристика может стать *ассиметричной* по  $V$ . Очевидно, что профиль распределения потенциала оказывает сильное влияние на форму вольт-амперной характеристики и поэтому даже простейшая модель должна учитывать это обстоятельство.

Итак, каким же образом следует проводить расчет потенциала внутри канала? Если бы канал был непроводящим, проблема свелась бы к решению уравнения Лапласа ( $\varepsilon_r$  — относительная диэлектрическая проницаемость, которая может быть пространственно зависимой)

$$\vec{\nabla} \cdot (\varepsilon_r \vec{\nabla} V) = 0$$

с граничными условиями  $V = 0$  (электрод истока),  $V = V_G$  (электрод затвора) и  $V = V_D$  (электрод стока). Рассматривая канал как точечный объект и пренебрегая любыми пространственными изменениями потенциала внутри него, мы могли бы наглядно представить решение этого уравнения эквивалентной схемы, содержащей емкости, как показано на рис. 1.4.4.

Потенциальная энергия в канале получается умножением электростатического потенциала  $V$  на электронный заряд  $-q$ :

$$U_L = \frac{C_G}{C_E}(-qV_G) + \frac{C_D}{C_E}(-qV_D). \quad (1.4.1a)$$

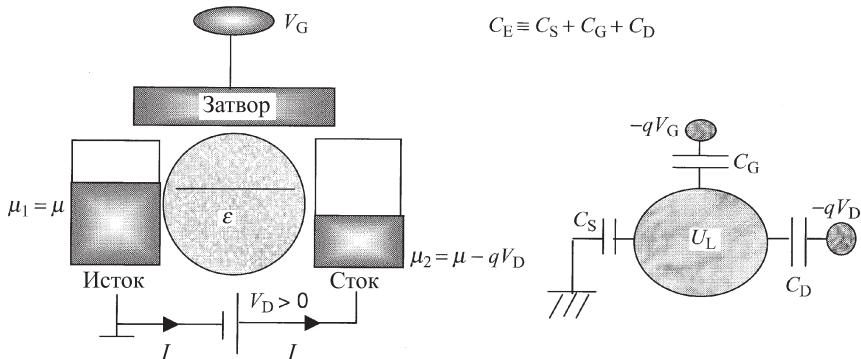


Рис. 1.4.4. Простейшая емкостная эквивалентная схема, используемая для расчета «лапласовского» потенциала активной области  $U_L$ , возникающего в ответ на внешние напряжения затвора и стока  $V_G$  и  $V_D$ . Символ  $C_E$  обозначает полную электростатическую (индекс  $E$ ) емкость. Действительный потенциал  $U$  может отличаться от  $U_L$ , если в окрестности энергий  $\mu_1$  и  $\mu_2$  существует заметная плотность электронных состояний

Здесь мы снабдили потенциальную энергию нижним индексом  $L$  как напоминание о том, что она получена из уравнения Лапласа, т. е. в пренебрежении любым изменением электронного заряда. Ясно, что такое предположение обоснованно при очень небольшом количестве электронных состояний в окрестности энергий  $\mu_1$  и  $\mu_2$ . Если же в канале происходит изменение электронной плотности  $\Delta\rho$ , то нам необходимо решать уравнение Пуассона для потенциала

$$\vec{\nabla} \cdot (\varepsilon_r \vec{\nabla} V) = -\Delta\rho/\varepsilon_0.$$

На языке нашей емкостной эквивалентной схемы это означает, что можно записать изменение заряда как сумму зарядов трех конденсаторов:

$$-q\Delta N = C_S V + C_G(V - V_G) + C_D(V - V_D),$$

так что потенциальная энергия  $U = -qV$  определяется суммой лапласовского потенциала и дополнительного слагаемого, пропорционального изменению числа электронов:

$$U = U_L + \frac{q^2}{C_E} \Delta N. \quad (1.4.1b)$$

Константа  $q^2/C_E \equiv U_0$  определяет изменения потенциальной энергии вследствие появления *одного* дополнительного электрона. Она называется одноэлектронной энергией электростатического отталкивания. Значение этой энергии мы обсудим в следующем разделе. Изменение  $\Delta N$  количества электронов вычисляется относительно того первоначального числа электронов в канале  $N_0$ , которое соответствует предполагаемому положению энергетического уровня  $\varepsilon$ .

**Самосогласованное решение. Итерационный метод.** В структурах, имеющих малые размеры, влияние потенциала  $U$  сводится к сдвигу кривой плотности состояний вдоль оси энергий. Этот сдвиг можно непосредственно включить в выражение для числа электронов  $N$  (уравнение (1.3.4)) и силы тока  $I$  (уравнение (1.3.3)):

$$N = \int_{-\infty}^{+\infty} dE D_\varepsilon(E - U) \frac{\gamma_1 f_1(E) + \gamma_2 f_2(E)}{\gamma_1 + \gamma_2}, \quad (1.4.2)$$

$$I = \frac{q}{\hbar} \int_{-\infty}^{+\infty} dE D_\varepsilon(E - U) \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(E) - f_2(E)]. \quad (1.4.3)$$

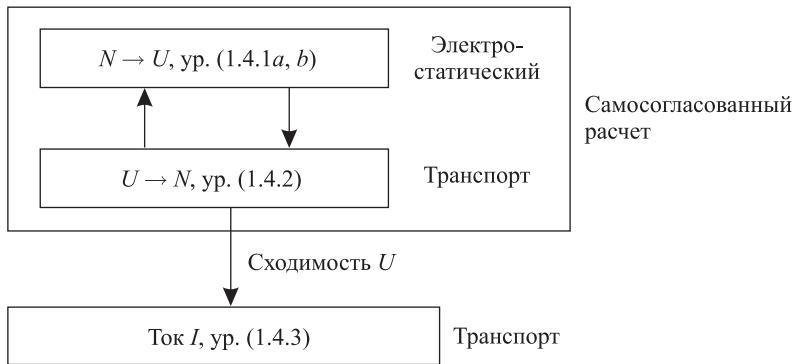


Рис. 1.4.5. Итерационная процедура, используемая для самосогласованного вычисления  $N$  и  $U$

В правую часть уравнения (1.4.2) входит величина  $U$ , которая, в свою очередь, зависит от  $N$  через соотношение (1.4.1b). Значит эти два урав-

нения должны решаться одновременно, т. е. «самосогласованно». Такое решение обычно выполняется с помощью итерационной процедуры, схема которой изображена на рис. 1.4.5.

Схема вычислений такова. Отталкиваясь от первоначальной оценки для  $U$ , мы вычисляем  $N$  из уравнения (1.4.2) с учетом определенной в (1.3.2) плотности состояний  $D_\varepsilon(E)$ . Затем с помощью уравнения (1.4.1b) с учетом определенной в (1.4.1a) «лапласовской» энергии  $U_L$  вычисляем соответствующий потенциал  $U$  и сравниваем результат с его исходной оценкой. Если полученное значение  $U$  не достаточно близко к нашей первоначальной оценке, мы пересматриваем ее, используя подходящий алгоритм, например, полагая:

$$U_{\text{new}} = U_{\text{old}} + \alpha(U_{\text{calc}} - U_{\text{old}}), \quad (1.4.4)$$

где  $\alpha$  — некоторое положительное число (обычно  $< 1$ ), которое должно выбираться максимально большим, но не приводящим к расходимости всей процедуры решения, которая может проявиться в росте  $U_{\text{calc}} - U_{\text{old}}$  от одной итерации к другой. Итерационный процесс должен повторяться до тех пор, пока мы не найдем потенциал  $U$  и не вычислим число электронов  $N$ , при котором новое значение  $U$  будет достаточно близко к предыдущему (скажем, в пределах долей  $k_B T$ ). Коль скоро сходящееся решение  $U$  найдено, величина тока может быть вычислена из уравнения (1.4.3).

На рис. 1.4.6 представлены найденные с помощью процедуры самосогласования сила тока  $I$  и количество электронов  $N$ , как функции приложенного напряжения стока (см. схему на рис. 1.4.5).

## 1.5. Кулоновская блокада

Электростатическая модель, в которой используется уравнение Пуасона, является хорошим нулевым приближением в задаче об электрон-электронном взаимодействии. Такой подход иногда называют приближением Хартри. Считается, однако, что эта модель приводит к завышению эффектов взаимодействия и поэтому нуждается в корректировке. Для адекватного описания электрон-электронного взаимодействия необходимо учитывать обменные и корреляционные эффекты. Поиск соответствующей функции  $U(N)$  (если таковая существует!), которая заменила бы наш простой результат (сравните с уравнением (1.4.1b))

$$U(N) = q^2(N - N_0)/C_E,$$

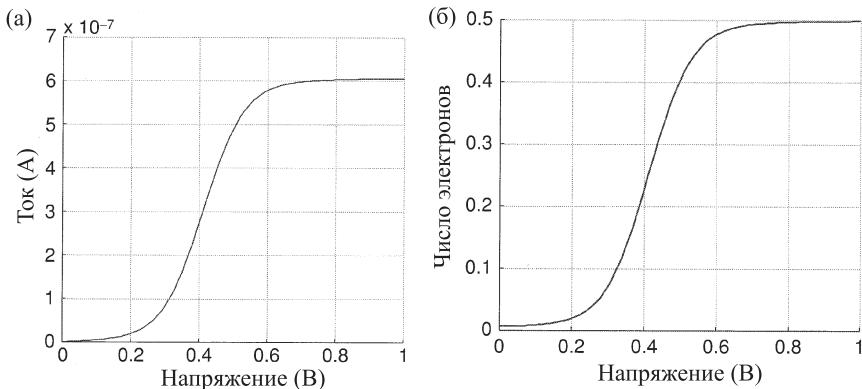


Рис. 1.4.6. (а) Величина тока как функция напряжения стока, вычисленная ССП-методом (рис. 1.4.5) при  $\mu = 0$ ,  $\varepsilon = 0.2$  эВ,  $V_G = 0$ ,  $k_B T = 0.025$  эВ,  $U_0 = 0.025$  эВ,  $C_D/C_E = 0.5$  и  $\gamma_1 = \gamma_2 = 0.005$  эВ. (б) Количество электронов как функция напряжения стока, вычисленная ССП-методом (рис. 1.4.5) при тех же значениях параметров, что и на рис. (а)

является, вероятно, одной из центральных проблем физики многоэлектронных систем, и он может привести к необходимости рассмотрения явлений, выходящих за рамки этой книги (к таким явлениям относится, например, электронный магнетизм). Однако есть один очень важный эффект, о котором здесь необходимо упомянуть. Речь идет об эффекте кулоновской блокады или одноэлектронной зарядке. Этот эффект радикально влияет на протекание электрического тока даже в простом одноуровневом устройстве. Позвольте мне объяснить, как это происходит.

Обычно энергетические уровни являются двукратно вырожденными по спину, т. е. на них могут находиться два электрона, один из которых имеет проекцию спина, направленную вверх, а второй проекцию спина вниз. Обычно это означает лишь то, что все наши результаты должны умножаться на два. Даже самое маленькое устройство имеет скорее два уровня, чем один, и его максимальный кондактанс будет равен обсуждавшемуся выше удвоенному кванту кондактанса  $G_0 \equiv q^2/h$ . Все выражения для количества электронов и для силы тока должны быть умножены на два. Однако существует и менее тривиальное следствие, которое я бы хотел обсудить.

Рассмотрим канал с двумя вырожденными по спину уровнями (рис. 1.5.1), содержащий в нейтральном зарядовом состоянии только один

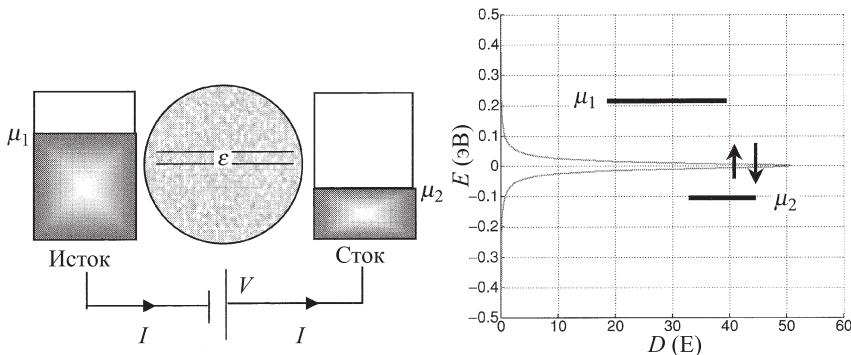


Рис. 1.5.1

электрон ( $N_0 = 1$ ). Мы ожидаем, что уширенная плотность состояний будет вдвое больше нашего предыдущего результата (см. уравнение (1.3.2))

$$D_\varepsilon = 2(\text{спин}) \times \frac{\gamma/2\pi}{(E - \varepsilon)^2 + (\gamma/2)^2}, \quad (1.5.1)$$

где полная ширина есть сумма вкладов, возникающих благодаря индивидуальной связи канала с каждым из двух контактов:  $\gamma = \gamma_1 + \gamma_2$ , как и раньше. Поскольку имеющиеся в наличии состояния нейтрального канала заполнены лишь наполовину, электрохимический потенциал будет находиться точно посередине уширенной плотности состояний (ПС). Поэтому мы могли бы ожидать появления большого тока под влиянием смещения, приводящего к расщеплению электрохимических потенциалов истока и стока.

Однако при определенных условиях плотность состояний выглядит как одна из функций, показанных на рис. 1.5.2, а именно плотность состояний расщепляется на две части, соответствующие спину вниз и спину вверх и разделенные одноэлектронной энергией электростатического отталкивания

$$U_0 \equiv q^2/C_E. \quad (1.5.2)$$

Когда приложено небольшое смещение, возникает очень слабый ток, поскольку в интервале между  $\mu_1$  и  $\mu_2$  электронные состояния отсутствуют. Такая «кулоновская блокада» экспериментально наблюдается в системах с энергией электростатического отталкивания  $U_0$ , превышающей ширину уровня  $\gamma$ .

Почему должны возникать два пика плотности состояний? Это трудно объяснить, опираясь на простую картину ССП. Два пика с одинаковой

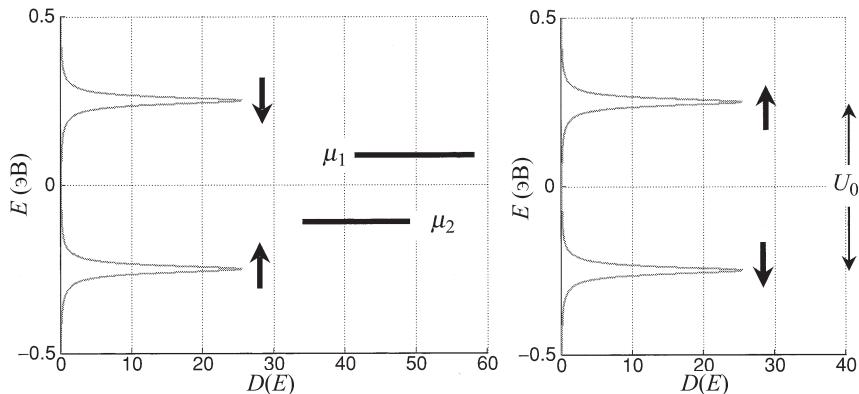


Рис. 1.5.2. При определенных условиях вместо одного вырожденного пика, показанного на рис. 1.5.1, плотность состояний расщепляется на две части, соответствующие спину вниз и спину вверх, разделенные одноэлектронной энергией электростатического отталкивания  $U_0$  ( $\gamma = 0.05$  эВ,  $U_0 = 0.25$  эВ)

энергией («вырожденные») всегда должны оставаться вырожденными, пока они испытывают действие одного и того же самосогласованного потенциала  $U$ . Дело в том, что ни один электрон не чувствует создаваемого им самим потенциала. Предположим, что уровень, отвечающий спину вверх, заполняется первым, вследствие чего уровень со спином вниз поднимается на  $U_0$ . Но, поскольку электрон не чувствует никакого самодействия, уровень со спином вверх остается неподвижным, что приводит к изображенной в левой части рис. 1.5.2 картине. Конечно, также вероятно, что первым заполнится уровень со спином вниз, что, в свою очередь, приведет к результату, изображеному в правой части рис. 1.5.2. В обоих случаях плотность состояний в окрестности  $\mu$  подавляется по сравнению с величиной, рассчитанной в рамках приближения самосогласованного поля (рис. 1.5.1).

Описание протекания электрического тока в таком режиме кулоновской блокады требует совершенно иного подхода, который в этой книге мы обсуждать не будем, за исключением краткого обзора, приведенного в параграфе 3.4. Когда же нам действительно следует позаботиться об учете эффектов кулоновской блокады? Ответ ясен: только в том случае, если  $U_0$  превышает  $k_B T$  и  $\gamma$  ( $= \gamma_1 + \gamma_2$ ). В других случаях, ССП-метод дает результаты, во многом похожие на те, что следуют из корректного рассмотрения (см. рис. 3.4.3).

Что же определяет величину  $U_0$ ? Ответ очевиден — это пространственная протяженность (локализация) электронной волновой функции. Если мы размажем один электрон по поверхности сферы радиуса  $R$ , то, как известно из курса общей физики, ее потенциал будет равен  $q/4\pi\varepsilon_r\varepsilon_0 R$ . Это означает, что энергия, необходимая для размещения на сфере другого электрона, будет равна  $q^2/4\pi\varepsilon_r\varepsilon_0 R \cong U_0$ , что составляет  $\sim 0.025$  эВ, если  $R = 5$  нм, а  $\varepsilon_r = 10$ . Уровням со слабо локализованными волновыми функциями (большой радиус  $R$ ) соответствуют маленькие энергии зарядки  $U_0$ , и метод самосогласованного поля обеспечивает приемлемое описание даже при самых низких температурах. Но если радиус  $R$  достаточно мал, энергия электростатического отталкивания  $U_0$  может превысить  $k_B T$ , и мы можем оказаться в том режиме, который плохо описывается методом самосогласованного поля, и когда эффекты одноэлектронной зарядки играют основную роль.

## 1.6. Переход к закону Ома

После обсуждения основных физических явлений, сопровождающих протекание электрического тока через малые проводники, нам следует перейти к обсуждению новых факторов, появляющихся при рассмотрении аналогичных проблем в массивных проводниках. При описании электронной проводимости малых проводников можно выделить три следующих режима.

- *Режим самосогласованного поля (CCP).* Если  $k_B T$  и/или  $\gamma$  соизмеримы с  $U_0$ , можно использовать метод CCP, который описан в параграфе 1.4.
- *Режим кулоновской блокады (КБ).* В случае, когда  $U_0$  существенно превышает как  $k_B T$ , так и  $\gamma$ , применение метода CCP становится неадекватным. Более корректным является использование (при определенных условиях) многоэлектронного уравнения эволюции, которое будет обсуждаться в параграфе 3.4.
- *Промежуточный режим.* В условиях, когда  $U_0$  соизмеримо с наибольшим по значению параметром  $k_B T$  и/или  $\gamma$ , упрощенный подход не имеет места: метод CCP не описывает корректно распределение зарядов, в то время как метод уравнения эволюции не учитывает правильно эффекты уширения.

В промежуточном режиме, как известно, даже при рассмотрении малых проводников могут возникать новые физические явления, для объяснения которых необходимо привлекать новые современные физические концепции. Так, например, в этом режиме экспериментально наблюдался эффект

Кондо, связанный с появлением дополнительного пика плотности состояний возле  $E = \mu$ . Этот пик существует одновременно с двумя пиками, показанными на рис. 1.5.2 и отстоящими друг от друга на  $U_0$ .

При рассмотрении больших проводников можно также говорить о существовании трех режимов транспорта, берущих начало от режимов, рассмотренных выше. Мы можем рассматривать большой проводник как массив доменов, как это показано на рис. 1.6.1. Энергия связи  $t$  между элементами массива приводит к эффекту, напоминающему (но не вполне точно соответствующему) уширение  $\gamma$ , которое мы связывали с наличием контактов. Если  $t \geq U_0$ , поведение полной проводимости будет относиться к режиму ССП и может быть описано в рамках метода ССП, изложенного в параграфе 1.4. В случае, когда  $t \ll U_0$ , относиться к режиму КБ и в принципе может рассматриваться в рамках многоэлектронного уравнения эволюции (которое будет обсуждаться в параграфе 3.4), если только можно пренебречь квантовой интерференцией между уровнями энергии. Здесь все зависит от конкретного расстояния между уровнями, а также от интенсивности процессов фазовой десинхронизации, которые будут обсуждаться в главе 10. С другой стороны, описание электронного транспорта в относящихся к промежуточному режиму больших проводниках, для которых  $\gamma_s \ll t \leq U_0$ , требует специального подхода. В самом деле, принято считать, что высокотемпературные сверхпроводники (чья микроскопическая теория все еще остается противоречивой) состоят из доменов, связь между которыми находится на грани между режимами ССП и КБ.

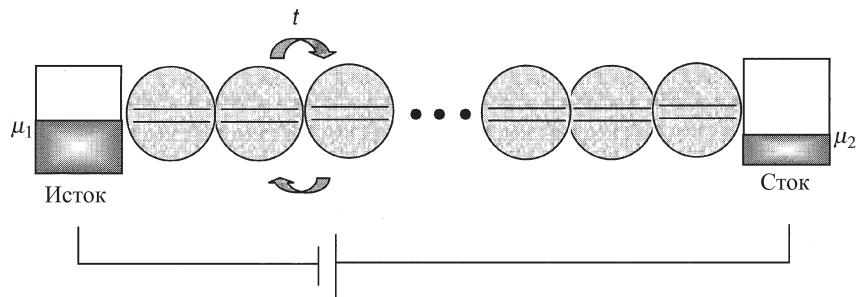


Рис. 1.6.1. Большой проводник, рассматриваемый как массив доменов. Если проводник на рисунке продолжить в поперечном направлении к плоскости рисунка, то каждый домен будет представлять массив доменов в направлении, перпендикулярном плоскости рисунка

Чем более делокализованными являются электронные волновые функции (большие  $t$ ), тем более точным является приближение ССП, которое,

в основном, будет обсуждаться в этой книге. Я постараюсь объяснить, как простое одноуровневое описание, использованное в параграфе 1.4, распространяется на случай более массивных проводников, и в первую очередь тех нанотранзисторов, в которых приближение ССП учитывает электронно-электронное взаимодействие через усредненный потенциал  $U(r)$ .

**Сводка результатов, относящихся к одноуровневым проводникам.** Выше рассматривался механизм протекания тока через устройство, содержащее лишь один квантовый уровень. Стартовав в параграфе 1.2 со случая простого дискретного уровня ( $\varepsilon$ ), мы перешли в параграфе 1.3 к обсуждению более общей модели уширенного уровня, описываемого лоренцевской плотностью состояний

$$D_\varepsilon(E) = 2 \times \frac{\gamma/2\pi}{(E - \varepsilon)^2 + (\gamma/2)^2}, \quad \gamma \equiv \gamma_1 + \gamma_2, \quad (1.6.1)$$

и самосогласованным потенциалом (см. параграф 1.4)

$$\begin{aligned} U &= U_L + U_0(N - N_0), \\ U_L &= \frac{C_G}{C_E}(-qV_G) + \frac{C_D}{C_E}(-qV_D), \\ U_0 &= \frac{q^2}{C_E}, \quad C_E = C_G + C_S + C_D. \end{aligned} \quad (1.6.2)$$

Число электронов в канале  $N$  в рамках ограниченного метода ССП определяется как

$$N = \int_{-\infty}^{+\infty} dE n(E),$$

где

$$n(E) = D(E - U) \left( \frac{\gamma_1}{\gamma} f_1(E) + \frac{\gamma_2}{\gamma} f_2(E) \right), \quad (1.6.4)$$

в то время как токи в двух контактах описываются выражениями

$$I_1 = \frac{q}{\hbar} \int_{-\infty}^{+\infty} dE \gamma_1 [D(E - U) f_1(E) - n(E)], \quad (1.6.5a)$$

$$I_2 = \frac{q}{\hbar} \int_{-\infty}^{+\infty} dE \gamma_2 [D(E - U) f_2(E) - n(E)]. \quad (1.6.5b)$$

В стационарном состоянии сумма двух токов должна быть равна нулю, что позволяет исключить величину  $n(E)$ :

$$I = \frac{q}{h} \int_{-\infty}^{+\infty} dE \bar{T}(E) [f_1(E) - f_2(E)],$$

где величина

$$\bar{T}(E) = D(E - U) 2\pi\gamma_1\gamma_2/\gamma \quad (1.6.6)$$

называется *коэффициентом пропускания*. Эта величина играет центральную роль в мезоскопической физике (см. параграф 9.4). Обратите внимание, что функции Ферми  $f_1$  и  $f_2$  даются выражениями

$$f_1(E) = f_0(E - \mu_1), \quad (1.6.7)$$

$$f_2(E) = f_0(E - \mu_2), \quad (1.6.8)$$

где  $f_0 \equiv [1 + \exp(E/k_B T)]^{-1}$ , а электрохимические потенциалы в контактах источника и стока описываются выражениями

$$\mu_1 = \mu, \quad (1.6.9)$$

$$\mu_2 = \mu - qV_D, \quad (1.6.10)$$

где  $\mu$  есть равновесное значение химического потенциала.

Следует отметить, что в уравнениях (1.6.4)–(1.6.6) я использовал функцию  $D(E)$  вместо  $D_\varepsilon(E)$  для определения плотности состояний. Позвольте объяснить, почему я это сделал.

**Массивные проводники — эвристический подход.** Функция  $D_\varepsilon(E)$  (см. уравнение (1.6.1)) определяет плотность состояний, полученную с учетом уширения дискретного уровня  $\varepsilon$ , в то время как  $D(E)$  определяет плотность состояний в общем случае для многоуровневого проводника с большим числом энергетических уровней (рис. 1.6.2).

Если мы сделаем довольно смелое предположение о том, что все уровни работают независимо, то можно будет использовать те же уравнения, что и для одноуровневой структуры, заменив одноуровневую плотность состояний  $D_\varepsilon(E)$  в уравнении (1.6.1) полной плотностью состояний  $D(E)$ . Имея это в виду, я буду при описании тока в канале обращаться к уравнениям (1.6.4)–(1.6.6) как к модели независимых уровней.

**Нанотранзистор — простая модель.** В качестве примера системы с независимыми уровнями построим модель нанотранзистора, показанного

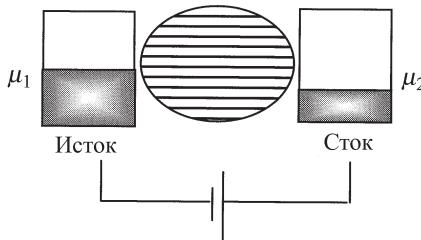


Рис. 1.6.2

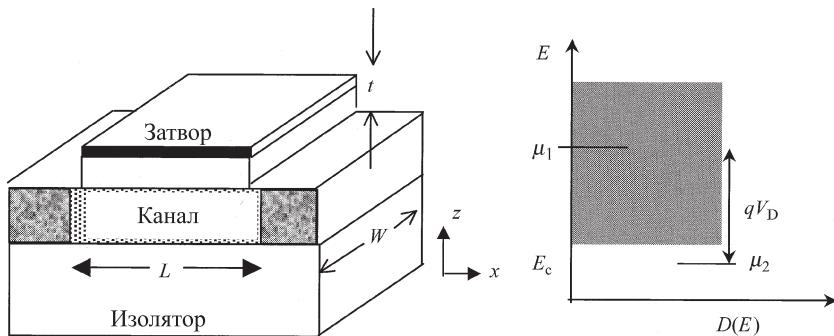


Рис. 1.6.3

на рис. 1.1, записав плотность состояний в виде (см. рис. 1.6.3, где  $W$  есть ширина в направлении  $y$ )

$$D(E) = m_c WL/\pi\hbar^2 \vartheta(E - E_c). \quad (1.6.11)$$

Здесь мы воспользуемся результатом, который будет обсуждаться ниже в главе 6. Этот результат заключается в том, что плотность состояний на единицу площади в массивном двумерном (2D) проводнике, с эффективной массой  $m_c$ , для энергий, превосходящих энергию края зоны проводимости  $E_c$ , равна  $m_c/\pi\hbar^2$ . Скорости эмиссии могут быть записаны в предположении, что электроны удаляются из контакта со скоростью  $\nu_R$

$$\gamma_1 = \gamma_2 = \hbar\nu_R/L. \quad (1.6.12)$$

Вольт-амперные характеристики, показанные на рис. 1.1.1, были получены с использованием следующих параметров модели:  $m_c = 0.25m$ ,

$C_G = 2\epsilon_r \epsilon_0 WL/t$ ,  $C_S = C_D = 0.05C_G$ ,  $W = 1$  мкм,  $L = 10$  нм, толщина изолятора  $t = 1.5$  нм, значение  $\nu_R = 10^7$  см/с. При высоких значениях потенциала стока  $V_D$  ток достигает насыщения, когда  $\mu_2$  падает ниже значения  $E_c$ , поскольку нет дополнительных состояний, которые могли бы внести вклад в ток. Следует обратить внимание на то, что емкость затвора  $C_G$  намного превосходит другие емкости, что помогает удерживать потенциал канала фиксированным относительно источника, по мере того как потенциал стока возрастает (см. уравнение (1.6.3)). Другими словами, при приложении напряжения к стоку положение «дна» плотности состояний  $E_c$  в канале останется неизменным по отношению к  $\mu_1$  и поэтому ток не будет стремиться к насыщению. Важным свойством правильно спроектированного транзистора является то, что затвор расположен намного ближе к каналу по сравнению с величиной  $L$ , что позволяет сохранять потенциал канала постоянным, несмотря на изменение напряжения стока  $V_D$ .

Я хотел бы заметить, что модель, использованная в наших рассуждениях, не учитывает конкретный профиль потенциала вдоль длины канала, рассматривая последний как маленький ящик с постоянным потенциалом  $U$ , приведенным в уравнении (1.6.3). Тем не менее, результаты (рис. 1.1.1) удивительно близки к экспериментальным наблюдениям и другим более реалистичным моделям. Дело в том, что ток в правильно сконструированныхnanoструктурах контролируется из малой области канала, расположенной вблизи истока, чья длина может составлять лишь малую долю от фактической длины  $L$ . К счастью, нам не нужно точно знать длину этой области, поскольку данная модель приводит к той же зависимости тока от параметра  $L$ .

**Закон Ома?** Приведет ли модель независимых уровней к выводу закона Ома, если мы будем рассматривать проводимость при малых напряжениях в массивном проводнике длины  $L$  и поперечным сечением  $S$ ? Поскольку ток пропорционален плотности состояний  $D(E)$  (см. уравнение (1.6.5)), которая, в свою очередь, пропорциональна объему  $SL$  проводника, то взаимодействие с контактами уменьшается обратно пропорционально длине  $L$  проводника, т. е. чем длиннее проводник, тем слабее взаимодействие состояний канала с контактами. В то время как плотность состояний растет с увеличением объема, взаимодействие с контактами падает как  $1/L$ , поэтому проводимость ведет себя как

$$G \sim SL/L = S.$$

Закон Ома, однако, говорит нам о том, что проводимость должна быть пропорциональна отношению  $S/L$ ; мы же предсказываем, что она пропор-

циональна  $S$ . Причина этого несоответствия состоит в том, что мы моделируем сейчас случай т. н. *баллистического* проводника, в котором электроны распространяются свободно и где сопротивление возникает лишь в контактах. Проводимость такого проводника, в самом деле, не зависит от его длины. Зависимость сопротивления от длины в законе Ома появляется вследствие процессов рассеяния в проводнике, которые до сих пор еще не были включены в наши рассуждения.

Без учета рассеяния мы можем записать параметр взаимодействия в виде  $\gamma = \hbar\nu/L$ , где  $\nu$  есть скорость электрона в процессе его движения с «отражением» от контактов. В присутствии рассеивателей — дефектов, неоднородностей и т. д., электрон не будет двигаться баллистически со скоростью  $\nu$ . Можно ожидать, что в этом случае электрон скорее будет двигаться по случайной траектории с характерным временем  $\tau \sim L^2/D$ , затрачиваемым на преодоление пути от одного контакта до другого. Заметим, что здесь  $D$  — это коэффициент диффузии, а не плотность состояний. Как следствие, уширение уровней будет иметь вид  $\gamma = \hbar/\tau \sim \hbar D/L^2$  вместо  $\hbar\nu/L$ . Поскольку плотность состояний пропорциональна произведению  $SL$ , это приводит к пропорциональности  $\sim S/L$  в соответствии с законом Ома. Подобные рассуждения выглядят очень простыми, но важно отметить, что диффузионное поведение не следует автоматически из квантовомеханического описания. Даже если бы мы уточнили нашу модель, вводя «когерентные» рассеиватели, это не привело бы к моментальному установлению закона Ома.

Полное описание требует учета процессов рассеяния с нарушением фазы, приводящих к изменению состояния внешнего объекта. Так, если электрон рассеивается на абсолютно твердом, т. е. не изменяющемся дефекте кристаллической решетки, то это рассеяние называется когерентным. Если же электрон отдает часть своей энергии решетке, что приводит к возникновению колебаний, то процесс рассеяния происходит с нарушением фазы. Это некогерентное рассеяние.

Существование подобных некогерентных рассеивателей необходимо также для того, чтобы забрать энергию у электронов и вызвать диссипацию. В данном разделе мы построили простую модель, позволяющую рассчитать сопротивление  $R$ , но при этом величина джоулевых потерь  $I^2R$  нигде не фигурировала, поскольку вся диссипация происходит только в контактах. Последнее становится очевидным, если мы взглянем на то, что происходит при перемещении электрона от источника к стоку (рис. 1.6.4). Вакантное состояние, или дырка, остается в источнике при значении энергии ниже, чем  $\mu_1$ , в то время как электрон появляется в стоке при энергии большей, чем  $\mu_2$ . Процессы диссипации энергии быстро уменьшают энергию элек-

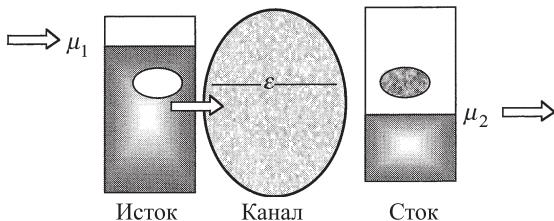


Рис. 1.6.4. Когда электрон перемещается от источника к стоку, в источнике остается незаполненное место, или дырка, в то время как электрон появляется в контакте стока. Процессы диссипации энергии быстро уменьшают энергию электрона до значения  $\mu_2$  внутри стока, а энергию дырки — увеличивают до значения  $\mu_1$  в источнике. При обсуждении нашей модели мы не рассматриваем явно эти процессы; просто постулируется, что контакты поддерживаются в равновесии с фиксированными значениями электрохимических потенциалов

трана до значения энергии  $\mu_2$  внутри стока, а энергию дырки — поднимают до значения энергии  $\mu_1$  в источнике. В конечном итоге эффект заключается в перемещении электрона из состояния с энергией  $\mu_1$  в источнике в состояние  $\mu_2$  в стоке, и в нашей модели энергия ( $\mu_1 - \mu_2$ ) диссирирует частично в источнике, а частично в стоке, но не в канале. Известно, что существует экспериментальное подтверждение того, что в наноразмерных проводниках большая часть тепла выделяется в контактах, т. е. вне канала, и что это позволяет экспериментаторам пропускать большие токи через малый проводник, не опасаясь сжечь его. Следует отметить, что длинные проводники имеют существенно некогерентный характер рассеяния внутри канала, что необходимо учесть в нашей модели.

Следует подчеркнуть, что переход от модели баллистического проводника к закону Ома требует разработки гораздо более глубокой модели протекания тока нежели модель независимых уровней (уравнения (1.6.4)–(1.6.6)), хотя последняя, зачастую, обеспечивает адекватное описание проводников малой длины. Я позволю себе кратко описать сейчас суть этой «продвинутой модели», которую мы будем развивать в данной книге и проиллюстрируем примерами в главе 11.

**Многоуровневые проводники — от чисел к матрицам.** Модель независимых уровней, которую мы разработали в данной главе, позволяет сформулировать основные принципы теории протекания тока через проводник. В этой модели рассматривается расположение равновесного электрохимического потенциала  $\mu$  в том интервале энергий, где определена плотность

состояний  $D(E)$ , рассмотрено уширение уровней  $\gamma_{1,2}$  вследствие взаимодействия с контактами 1 и 2, определен самосогласованный потенциал  $U$ , описывающий влияние внешних электродов, и изменение числа электронов  $N$ . В общем случае при построении модели многоуровневого проводника с  $n$  энергетическими уровнями каждую из этих величин следует заменить на соответствующую матрицу размера ( $n \times n$ ):

$\varepsilon \rightarrow [H]$	Гамильтониан
$\gamma_{1,2} \rightarrow [\Gamma_{1,2}(E)]$	Матрица уширения
$2\pi D(E) \rightarrow [A(E)]$	Спектральная функция
$2\pi n(E) \rightarrow [G^n(E)]$	Корреляционная функция
$U \rightarrow [U]$	Матрица самосогласованного потенциала
$N \rightarrow [\rho] = \int (dE/2\pi)[G^n(E)]$	Матрица плотности.

В частности, взаимодействие с контактами описывается матрицей «собственно энергетической функции»  $[\Sigma_{1,2}(E)]$ , чья антиэрмитова часть есть матрица уширения:  $\Gamma_{1,2} = i[\Sigma_{1,2} - \Sigma_{1,2}^\dagger]$ . Все представляющие интерес величины могут быть рассчитаны с помощью этих матриц. К примеру, в параграфе 1.2 мы обсуждали входящий и выходящий токи электронов в одноуровневой структуре. Рис. 1.1.1 иллюстрирует, как эти концепции обобщаются в терминах указанных выше матриц. Я хотел бы отметить заранее, что с целью подчеркнуть сходство с понятием электронной плотности я использовал в книге обозначение  $G^n(E)$  вместо обычно используемого в литературе обозначения неравновесных функций Грина (НРФГ)  $-iG^<(E)$ , введенных в работах [37, 24] и [25].

Обратите внимание на то, что в матричной модели (рис. 1.6.5б) я добавил третий «контакт», обозначенный здесь как « $s$ -контакт». С помощью этого понятия мы будем описывать процессы рассеяния, без учета которых нельзя, в частности, совершить переход к закону Ома. Необходимость учета контактов ( $\Gamma_1$  и  $\Gamma_2$ ) была осознана в процессе развития мезоскопической физики в 1980-х годах в связи с интерпретацией экспериментов. До тех пор считалось, что контакты приводят лишь к малым отклонениям от идеальной модели. Понять физику проводимости в терминах  $s$ -контакта, насколько мне известно, никто не пытался до тех пор, пока Бюттикер в середине 1980-х не ввел в рассмотрение данную идею феноменологически (см. [5, 13]). Впоследствии в моей работе [11] в рамках микроскопической модели было показано, что процессы некогерентного рассеяния действуют подобно некоторому фиктивному «контакту», распределенному вдоль канала, который забирает и вновь инжектирует электроны. Как и для случая

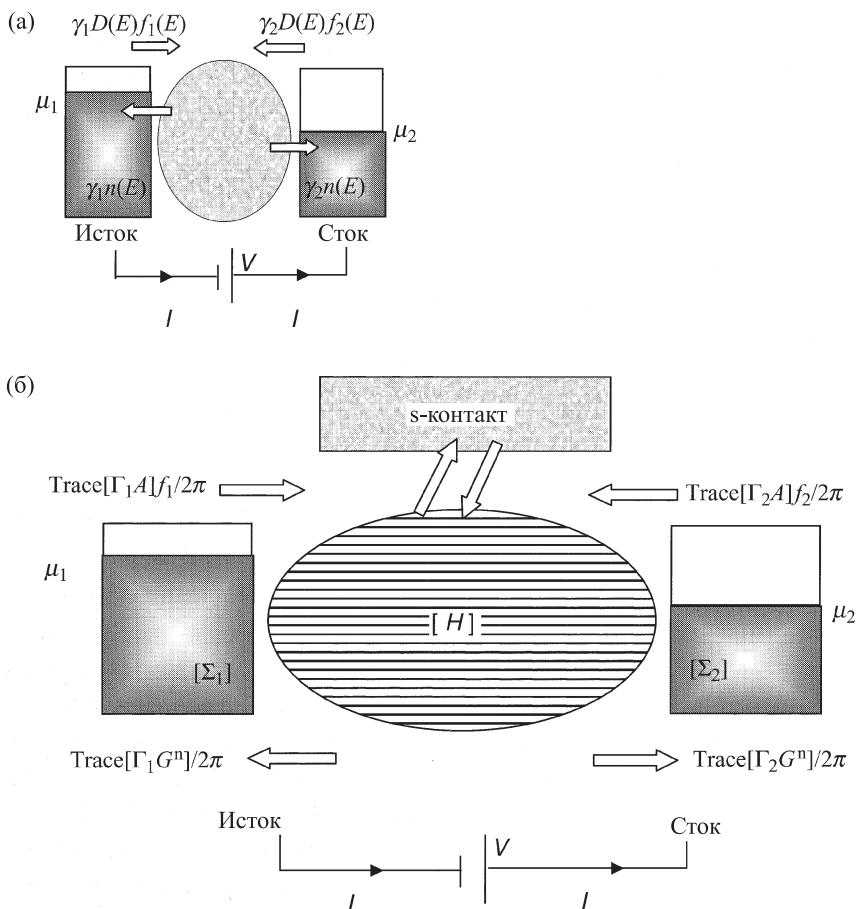


Рис. 1.6.5. От чисел к матрицам: входящий и выходящий токи электронов в контактах источника и стока данной структуры. (а) Простой результат модели независимых уровней, см. уравнения (1.6.4)–(1.6.6). (б) Общая матричная модель, которая будет развита в данной книге. Без «*s*-контакта» эта модель эквивалентна уравнению (6) в работе [39]. Данный «*s*-контакт», распределенный вдоль канала, описывает некогерентные процессы рассеяния [11]. В общем случае этот «контакт» не может быть описан функцией Ферми, в отличие от настоящих контактов

реальных контактов, взаимодействие с этим «контактом» также может быть описано с помощью матрицы уширения  $\Gamma_s$ . Однако в отличие от реальных контактов рассеивающий контакт в общем случае не может быть описан функцией Ферми. Поэтому, хотя выходящий ток и определяется выражением  $\text{Trace}[\Gamma_s G^n / 2\pi]$ , расчет входящего тока требует отдельного рассмотрения, которое мы проведем в главе 10. Полный набор уравнений будет обсуждаться в главе 11.

Читатель может задать естественный вопрос. Почему числа превратились в матрицы, а не просто в векторы-столбцы? Пусть, к примеру, единичный домен проводника характеризуется уровнем энергии  $\varepsilon$ . Тогда имеет смысл полагать, что в случае многих единичных доменов нам следует говорить об уровне энергии  $\varepsilon(n)$  в каждой  $n$ -й ячейке. Зачем нам нужна матрица  $H(m, n)$ ? Этот вопрос отсылает нас к основам аппарата квантовой механики, где все физические величины задаются матрицами. Мы можем отыскать такое представление, в котором матрица  $[H]$  будет диагональной, и в том же представлении можно записать собственные значения энергии в виде вектора-столбца  $\varepsilon(n)$ . Если в этом представлении и все прочие матрицы имеют приблизительно диагональный вид, мы и в самом деле могли бы работать с векторами-столбцами из  $n$  элементов, а не с матрицами из  $n^2$  элементов. Так поступают обычно при «квазиклассическом» рассмотрении. Однако в общем случае не существует одного представления, в котором бы все матрицы были диагональными, и поэтому требуется полноценное квантовое рассмотрение.

Схема расчета, представленная на рис. 1.6.5 б без  $s$ -контакта, часто используется для анализа небольших устройств. В этой форме она идентична результату, полученному в работе [39] (уравнение (6)), где использовался основанный на формализме НРФГ метод работы [8]. Чтобы сделать этот подход доступным читателю, не знакомому с наиболее сложными разделами физики многих частиц, я получу эти результаты, используя элементарные рассуждения. Результаты, обсуждаемые в данном разделе (рис. 1.6.5 а), можно рассматривать как частный случай этого общего подхода, для которого все матрицы имеют размер  $(1 \times 1)$ , т. е. являются числами. В самом деле, если существует представление, диагонализующее все эти матрицы, то матричная модель без  $s$ -контакта должна естественным образом следовать из рис. 1.6.5 а. Тогда в этом специальном представлении мы могли бы записать отдельные уравнения для тока, протекающего через каждый диагональный элемент (или уровень), сложить их и записать сумму, т. е. след. Получающиеся в результате уравнения должны быть справедливы в любом представлении, поскольку след является инвариантом по отношению к замене базиса. В общем случае, однако, матричная модель не может быть

получена таким простым путем, поскольку не существует одного представления, диагонализующего все матрицы. В главах 8–10 я вывел полную матричную модель (рис. 1.6.5б), используя элементарную квантовую механику. В Приложение я поместил краткий вывод тех же результатов, изложенный на языке вторичного квантования, но и там я попытался сохранить несколько упрощенный уровень изложения.

Я должен отметить, что картина, приведенная на рис. 1.6.5, является недостаточной для вычисления тока: необходимы дополнительные уравнения для определения «плотности состояний»  $A[E]$  и «электронной плотности»  $G^n[E]$ . В нашей элементарной модели (рис. 1.6.5а) мы записали плотность состояний с помощью «подстановки» (см. уравнение (1.6.1)), но отдельного уравнения для электронной плотности не потребовалось. Эта величина была получена из условия равенства токов (см. вывод уравнения (1.2.3) для случая дискретного уровня и уравнения (1.6.4) для уширенного уровня). Точно также и в матричной модели (рис. 1.6.5б) (без  $s$ -контакта), построенной в работе [39], было указано, что величину  $[G^n(E)]$  можно исключить таким же образом, но лишь только в том случае, если  $[\Gamma_1]$  равна постоянной, умноженной на  $[\Gamma_2]$ . Однако такое рассмотрение может оказаться справедливым лишь для очень коротких каналов. В противном случае контакт источника удален от контакта стока, что приводит к существенному отличию связанных с различными контактами матриц  $[\Gamma_1]$  от  $[\Gamma_2]$ . В этом случае нам потребуются дополнительные уравнения для того, чтобы определить как  $[A(E)]$ , так и  $[G^n(E)]$ .

Эти матрицы (как диагональные, так и недиагональные) скрывают в себе информацию об описываемых ими физических явлениях: следующие пять глав посвящены обсуждению структуры гамильтоновской матрицы  $[H]$ , глава 7 посвящена рассмотрению свойств матрицы плотности  $[\rho]$ , глава 8 — матрице  $[\Sigma]$ , в главе 9 обсуждаются все эти понятия с целью получения диаграммы входящего/выходящего тока, показанной на рис. 1.6.5. В главе 10 вводится матрица  $\Gamma_s$ , описывающая рассеяние, что завершает построение модели диссипативного квантового транспорта. Наконец, в главе 11 мы проиллюстрируем все развитые положения с помощью ряда примеров, описывающих переход от баллистического транспорта к закону Ома, или, другими словами, от атома к транзистору.

После столь длинного введения мы готовы обратиться к «деталям». В следующей главе начнем с обсуждения вопроса о том, как записать гамильтониан  $[H]$  для данной структуры. Собственные значения этого гамильтониана расскажут нам, каковы уровни энергии системы. Мы пройдем путь от атома водорода в главе 2 к проблемам, связанным с твердыми телами в главе 5, а затем спустимся «вниз» к наноструктурам в главе

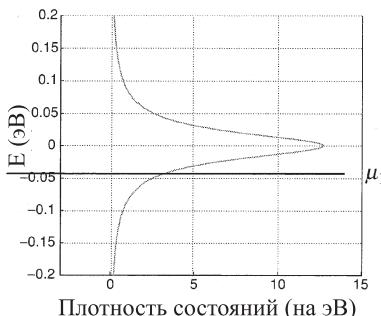
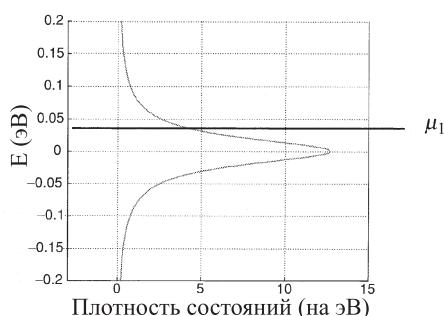
6. В следующей главе будет проведено обсуждение проблемы, с которой непосредственно связано зарождение квантовой механики, а именно, с атомом водорода.

## Упражнения

**У.1.1.** Рассмотрите канал с одним вырожденным по спину уровнем, предполагая следующие значения параметров:  $\mu = 0$ ,  $\varepsilon = 0.2$  эВ,  $k_B T = 0.025$  эВ,  $\gamma_1 = \gamma_2 = 0.005$  эВ. Вычислите, используя метод самосогласованного поля (ССП), зависимость тока от напряжения стока  $V_D$ , предполагая, что  $V_G = 0$ , параметр  $U_L = -qV_D/2$  и  $U_0 = 0.1$  эВ, 0.25 эВ. Сравните результаты с рис. 1.4.6.

**У.1.2.** Вычислите зависимость тока от напряжений затвора и стока для нанотранзистора в системе координат, показанной на рис. 1.1.1, используя уравнения ССП (1.6.2)–(1.6.8) с функцией  $D(E) = m_c WL/\pi\hbar^2$  и  $\gamma_1 = \gamma_2 = \hbar\nu_R/L$ . Используйте следующие значения параметров:  $m_c = 0.25m$ ,  $C_G = 2\varepsilon_r\varepsilon_0 WL/t$ ,  $C_S = C_D = 0.05C_G$ ,  $W = 1\mu m$ ,  $L = 10$  нм, толщина изолятора  $t = 1.5$  нм, скорость  $\nu_R = 10^7$  см/с.

**У.1.3.** Термоэлектрический эффект: в данной главе мы обсудили протекание тока, возникающее при наличии напряжения между двумя контактами. В данном случае ток зависит от плотности состояний вблизи энергии Ферми, и не важно, расположена ли равновесная энергия Ферми  $\mu_1$  (а) на нижней или (б) на верхней границе плотности состояний.

(а)  $n$ -типа(б)  $p$ -типа

Однако, если мы просто нагреем один контакт по отношению к другому таким образом, что  $T_1 > T_2$  (без приложения напряжения), то возник-

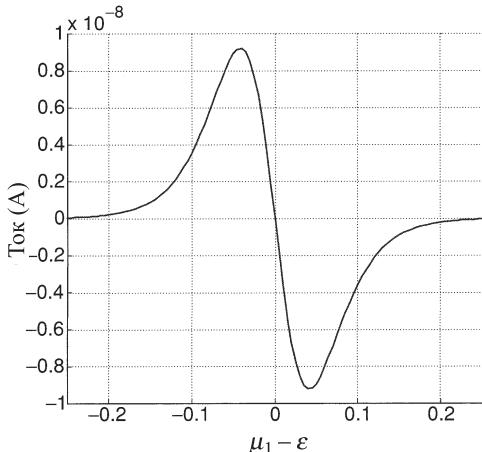


Рис. У.1.3. У читателя должен получиться график, похожий на этот, на котором видно изменение направления тока на противоположное при переходе от структуры *n*-типа ( $\mu_1 < \varepsilon$ ) к структуре *p*-типа ( $\mu_1 > \varepsilon$ )

нет электрический ток, который будет протекать в различном направлении в случаях (а) и (б).

Чтобы обнаружить этот эффект теоретически, вычислите ток из уравнения (1.6.6) при значении  $U = 0$  (учитывая, что нет необходимости решать задачу самосогласованно),  $V_D = 0$  и  $V_G = 0$ , и при двух значениях  $k_B T_1 = 0.026$  эВ и  $k_B T_2 = 0.025$  эВ:

$$f_1(E) \equiv \left[ 1 + \exp\left(\frac{E - \mu_1}{k_B T_1}\right) \right]^{-1} \text{ и } f_2(E) \equiv \left[ 1 + \exp\left(\frac{E - \mu_1}{k_B T_2}\right) \right]^{-1},$$

Постройте ток как функцию  $\mu_1 - \varepsilon$  в интервале изменения этой величины от -0.25 до 0.25 эВ, предполагая, что  $\gamma_1 = \gamma_2 = 0.05$  эВ (рис. У.1.3). Впервые на данную задачу обратили внимание в работе [42].

**У.1.4.** Отрицательное дифференциальное сопротивление. На рис. 1.4.6 *a* показаны зависимости  $I$  от  $V_D$ , вычисленные с помощью самосогласованного решения уравнений (1.6.2)–(1.6.5*b*) в предположении, что

$$\begin{aligned} \varepsilon &= 0.2 \text{ эВ}, & k_B T &= 0.025 \text{ эВ}, & U_0 &= 0.025 \text{ эВ}, & V_G &= 0, \\ \mu_1 &= 0, & \mu_2 &= \mu_1 - qV_D, & U_L &= -qV_D/2. \end{aligned}$$

Параметры уширения  $\gamma_1$  и  $\gamma_2$ , возникающего при учете двух контактов, предполагаются равными 0.005 эВ.

Предположим теперь, что параметр  $\gamma_1$  равен 0.005 эВ для значений  $E > 0$  и равен нулю для  $E < 0$  (причем  $\gamma_2$  предполагается все еще независимым от энергии и равным 0.005 эВ). Покажите, что вольт-амперные характеристики продемонстрируют в этом случае эффект отрицательного дифференциального сопротивления (ОДС), который заключается в падении тока с ростом напряжения. Отрицательное дифференциальное сопротивление возникает лишь при одном знаке приложенного напряжения, а при другом знаке напряжения этот эффект отсутствует, как показано на рис. У.1.4. На эту задачу впервые указали авторы работы [43].

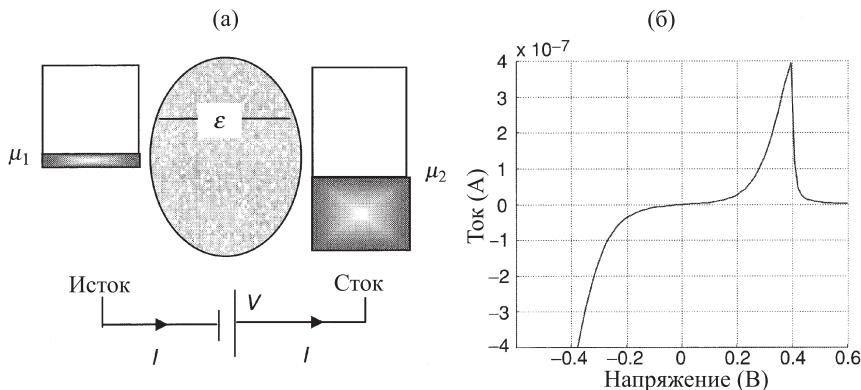


Рис. У.1.4

**У.1.5 (Закон Ома).** Подход «от малого к большому», предложенный в данной главе, приводит к следующему выражению для тока (из уравнения (1.6.6), где  $\gamma_1 = \gamma_2$  и  $U = 0$ ):

$$I = (q\gamma_1/2\hbar) \int_{-\infty}^{+\infty} dE D(E) (f_1(E) - f_2(E)). \quad (A)$$

Обычный подход «от большого к малому» приводит к выражению для тока, пропорциональному площади поперечного сечения  $S$ , обратной длине  $1/L$ , электронной плотности  $n$  и подвижности  $\bar{\mu}$  (черта добавлена,

чтобы отличить ее от электрохимического потенциала  $\mu$ ):

$$I = (S/L)n\bar{\mu}(\mu_1 - \mu_2). \quad (B)$$

Как перейти от (A) к (B)? Подсказки состоят в следующем: (1) связите параметр уширения с диффузионной постоянной  $\bar{D}$  через соотношение  $\gamma_1 = 2\hbar\bar{D}/L^2$ , как было упомянуто в параграфе 1.6; (2) предположите, что электронная система является «невырожденной», так что функции Ферми аппроксимируются функциями Больцмана, и предположите также, что разность потенциалов намного меньше  $k_B T$ , в силу чего  $f_1(E) - f_2(E) \approx e^{-(E-\mu_1)/k_B T}(\mu_1 - \mu_2)/k_B T$ ; наконец, (3) используйте формулу Эйнштейна  $\bar{D}/\bar{\mu} = k_B T/q$  и соотношение  $nSL = \int_{-\infty}^{+\infty} dE D(E) e^{-(E-\mu)/k_B T}$ .

## Глава 2

# Уравнение Шредингера

На протяжении нескольких следующих глав мы должны научиться тому, каким образом следует записывать гамильтониан  $[H]$  заданной структуры (см. рис. 1.6.5). В данной главе (1) мы начнем рассмотрение с задачи об атоме водорода (параграф 2.1) и обсуждения тех фактов, которые привели к открытию уравнения Шредингера, (2) далее рассмотрим простой численный метод, называемый методом конечных разностей (параграф 2.2), который может быть использован для преобразования этого дифференциального уравнения в матричное уравнение, и, наконец, (3) рассмотрим несколько численных примеров (параграф 2.3), показывающих, как производится диагонализация гамильтоновской матрицы и находятся энергетические уровни и волновые функции.

### 2.1. Атом водорода

В начале двадцатого века ученые пытались построить модель атомов, о которых было известно лишь то, что они состоят из отрицательно заряженных частиц, окружающих положительно зараженное ядро. Простая модель представляет электрон (с массой  $m$  и зарядом  $-q$ ) движущимся по орбите вокруг ядра (с зарядом  $Zq$ ) на расстоянии радиуса  $r$ , причем электрон притягивается к ядру силами электростатического притяжения, во многом напоминающего гравитационное притяжение, которое удерживает планеты на орбитах вокруг Солнца.

$$\frac{Zq^2}{4\pi\varepsilon_0 r^2} = \frac{mv^2}{r} \Rightarrow v = \sqrt{\frac{Zq^2}{4\pi\varepsilon_0 mr}} \quad (2.1.1)$$

электростатическая сила = центростремительной силе.

Более быстро движущийся электрон движется по орбите, имеющей меньший радиус. Полная энергия связана с радиусом его орбиты соотно-

шением

$$E = -\frac{Zq^2}{4\pi\varepsilon_0 r} + \frac{mv^2}{2} = -\frac{Zq^2}{8\pi\varepsilon_0 r} \quad (2.1.2)$$

потенциальная + кинетическая  
энергия + энергия = полной энергии.

Однако вскоре стало ясно, что эта упрощенная точка зрения была неадекватной, поскольку в соответствии с классической электродинамикой электрон, движущийся по орбите, должен излучать электромагнитные волны, что привело бы к непрерывной потере энергии и падению (по спирали) электрона на ядро. В рамках классического подхода невозможно построить стабильную структуру подобной системы. Исключением является лишь случай, когда электрон находится непосредственно на самом ядре, что, разумеется, противоречит эксперименту. Стало очевидным, что при описании атомной системы необходимо отказаться от классического подхода.

Нильс Бор постулировал, что в стационарных состояниях электрон движется вокруг ядра по орбитам, расположенным на определенных расстояниях от него, и при этом имеет фиксированные значения углового момента. Позднее стало ясно, что размеры орбит могут быть найдены из предположения о том, что электроны описываются волновыми функциями, характеризующимися определенной длиной волны де Броиля  $h/mv$ , где  $h$  есть постоянная Планка. При движении по стационарной орбите периметр этой орбиты должен быть равным целому числу длин волн:

$$2\pi r = n(h/mv). \quad (2.1.3)$$

Используя уравнения (2.1.3), (2.1.1) и (2.1.2), можно найти радиус и энергию электрона, движущегося по стационарной орбите:

$$r_n = (n^2/Z)a_0 \quad \text{боровский радиус}, \quad (2.1.4)$$

где

$$a_0 = 4\pi\varepsilon_0\hbar^2/mq = 0.0529 \text{ нм}, \quad (2.1.5)$$

$$E_n = -(Z^2/n^2)E_0, \quad (2.1.6a)$$

$$\text{где } E_0 = q^2/8\pi\varepsilon_0a_0 = 13.6 \text{ эВ} \quad (1 \text{ Ридберг}). \quad (2.1.6b)$$

Если электрон находится на орбите с наименьшей энергией ( $n = 1$ ), он не может далее терять свою энергию, поскольку орбит с более низкой энергией просто не существует (рис. 2.1.2 a). Если мы сообщим атому дополнительную энергию, то электрон перейдет в возбужденное состояние,

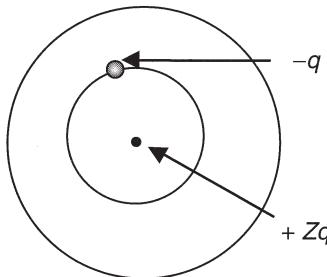


Рис. 2.1.1. В стационарных состояниях длина окружности, по которой движется электрон, равна целому числу длин волн де Броиля

т. е. на более высокую стационарную орбиту (рис. 2.1.2 б). По мере того, как он шаг за шагом будет опускаться на нижележащие энергетические состояния, будут излучаться фотоны, чья энергия  $h\nu$  равна разности энергий на орbitах  $m$  и  $n$ :

$$h\nu = E_m - E_n = E_0 Z^2 \left( \frac{1}{n^2} - \frac{1}{m^2} \right). \quad (2.1.7)$$

Было экспериментально установлено, что спектр света, излучаемого атомом водорода, действительно состоит из дискретных частот, описываемых этим соотношением с целыми значениями  $m$  и  $n$ . Такое удивительно точное соответствие с экспериментом привело к выводу о том, что в этой простой модели, обычно называемой моделью Бора, скрывается фундаментальная идея, правильно описывающая главные свойства данного явления.

*Уравнение Шредингера*, составляющее основу квантовой механики, позволяет рассчитать энергетические уровни в поле произвольного потенциала  $U(\vec{r})$ .

$$i\hbar \frac{\partial \Psi}{\partial t} = \left( -\frac{\hbar^2}{2m} \nabla^2 + U(\vec{r}) \right) \Psi. \quad (2.1.8)$$

Каким же образом из этого уравнения следует существование дискретных энергетических уровней? Математически можно показать, что если  $U(\vec{r}) = -Zq^2/4\pi\epsilon_0 r$  для потенциала ядра с зарядом  $+Zq$ , то решения этого уравнения характеризуются тремя индексами  $n$ ,  $l$  и  $m$ :

$$\Psi(\vec{r}, t) = \phi_{nlm}(\vec{r}) \exp(-iE_n t/\hbar), \quad (2.1.9)$$

где энергия  $E_n$  зависит только от индекса  $n$  и дается выражением  $E_n = -(Z^2/n^2)E_0$  в соответствии с полученным ранее эвристическим принци-

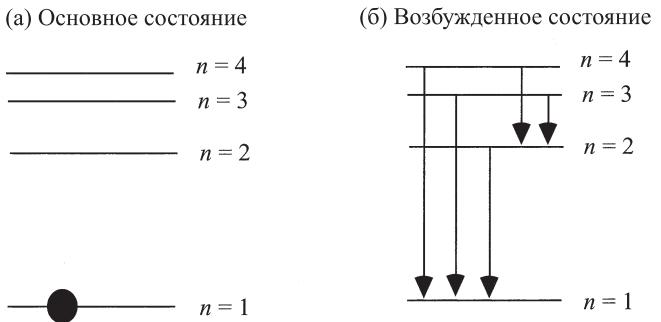


Рис. 2.1.2. (а) Электрон релаксирует в состояния с наименьшей энергией ( $n = 1$ ).  
(б) Переходя на нижние энергетические, он излучает фотоны, чья энергия  $h\nu$  определяется разницей энергий начального и конечного состояний

пом (см. уравнение (2.1.6 a)). Уравнение Шредингера, описывающее электроны, имеет, разумеется, иной смысл нежели, к примеру, уравнение, описывающее акустическую волну в резонаторе. Но энергия электрона  $E$  играет примерно ту же роль, что и частота акустической волны. Подобно волне в акустическом резонаторе, электронная волна в атомном «резонаторе» (параметры которого определяются потенциальной энергией  $U(\vec{r})$ ) имеет свои резонансные частоты. Рассмотрим этот вопрос более подробно.

**Волны в резонаторе.** Давайте, прежде всего, рассмотрим простой пример, а именно колебания одномерной (1D) струны  $u(x, t)$ , которые описываются одномерным волновым уравнением:

$$\frac{\partial^2 u}{\partial t^2} = \mu^2 \frac{\partial^2 u}{\partial x^2}. \quad (2.1.10)$$

Решения этого уравнения могут быть записаны в виде плоских волн, обладающих линейным законом дисперсии  $\omega = \pm \mu k$ :

$$u = u_0 \exp(ikx) \exp(-i\omega t) \Rightarrow \omega^2 = \mu^2 k^2. \quad (2.1.11)$$

Что произойдет, если мы закрепим два конца струны, и их смещение в этих точках будет равно нулю (рис. 2.1.3)? В этом случае разрешенные значения  $k$  будут «квантоваться», что приведет к появлению дискретного набора резонансных частот:

$$u = u_0 \exp(ikx) \exp(-i\omega t) \Rightarrow k = n\pi/L \Rightarrow \omega = n\pi\nu/L. \quad (2.1.12)$$



Рис. 2.1.3. Стоячие волны. (а) Акустические волны в «гитарной» струне; отклонение от равновесия на обоих ее концах равно нулю. (б) Электронные волны в одномерной потенциальной яме с бесконечно высокими стенками, и волновая функция, обращающаяся в нуль на границах ямы

Подобным образом ведет себя и решение уравнения Шредингера. Если потенциал, ограничивающий движение частицы, отсутствует ( $U = 0$ ), решения уравнения Шредингера можно записать как

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2}, \quad (2.1.13)$$

т. е. в виде плоских волн, которым отвечает параболический закон дисперсии  $E = \hbar^2 k^2 / 2m$ :

$$\Psi = \Psi_0 \exp(ikx) \exp(-iEt/\hbar) \Rightarrow E = \hbar^2 k^2 / 2m. \quad (2.1.14)$$

Если мы закрепим концы струны, то получим стоячие волны с квантованными (точнее дискретными) значениями  $k$  и резонансными частотами:

$$\begin{aligned} \Psi &= \Psi_0 \exp(ikx) \exp(-iEt/\hbar) \Rightarrow \\ &\Rightarrow k = n\pi/L \Rightarrow E = \hbar^2 \pi^2 n^2 / 2mL^2. \end{aligned} \quad (2.1.15)$$

Атомные «резонаторы», разумеется, определяются потенциалами  $U(\vec{r})$  более сложного вида, нежели простой одномерный потенциал, показанный на рис. 2.1.2 б. Тем не менее, основной принцип остается неизменным: как только мы ограничиваем волну областью резонатора, ее частота или энергия принимают дискретные значения.

**«Периодический» резонатор.** Другой тип резонатора, который мы будем часто рассматривать, представляет собой кольцо (рис. 2.1.4), где конечная точка при  $x = L$  вновь соединена с начальной точкой  $x = 0$  и, таким образом, конец и начало отсутствуют. Реально встречающиеся резонаторы редко имеют подобную форму, но подобная идеализация используется довольно часто, поскольку она упрощает математические расчеты. Эта модель

будет обоснована, если нас интересуют решения, относящиеся к внутренним областям резонатора, на которые не влияют условия на его границах (или краях), и при этом наши расчеты становятся проще. Однако подобные рассуждения могут быть не справедливыми в случае «nanoструктур», где реальные условия на поверхности могут существенно влиять на результаты измерений экспериментатора.

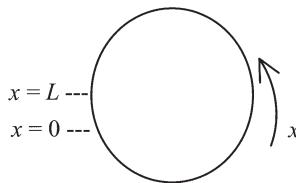


Рис. 2.1.4. Стоячие волны в кольце

Так или иначе, для периодического резонатора собственные функции даются выражениями (сравните с уравнением (2.1.15))

$$\begin{aligned} \Psi &= \Psi_0 \sin(kx) \exp(-iEt/\hbar) \\ \text{и } \Psi &= \Psi_0 \cos(kx) \exp(-iEt/\hbar), \end{aligned} \quad (2.1.16)$$

где  $k = 2n\pi/L \Rightarrow E = 2\hbar^2\pi^2/mL^2$ .

Здесь значения  $k$  разделены интервалом  $2\pi/L$  вместо  $\pi/L$ , поэтому их число вдвое меньше числа разрешенных значений. При этом каждому значению  $k$  отвечают функции косинус и синус, которые теперь появляются парами.

Важно отметить, что, как только появляются вырожденные собственные состояния, т. е. две или более собственных функций отвечают одинаковым собственным значениям, любая их линейная комбинация также будет собственной функцией, отвечающей тому же собственному значению. Следовательно, можно записывать собственные состояния в виде

$$\begin{aligned} \Psi &= \Psi_0 \exp(+ikx) \exp(-iEt/\hbar) \\ \text{и } \Psi &= \Psi_0 \exp(-ikx) \exp(-iEt/\hbar), \end{aligned} \quad (2.1.17)$$

где  $k = 2n\pi/L \Rightarrow E = 2\hbar^2\pi^2/mL^2$ .

Подобная запись часто используется в аналитических расчетах, причем, первая из функций рассматривается как состояние с волновым числом  $+k$ , распространяющееся в положительном направлении оси  $x$ , в то время

как вторая — как состояние с волновым числом  $-k$ , распространяющееся в отрицательном направлении оси  $x$ .

**Электронная плотность и плотность потока вероятности.** Электрон, описываемый волновой функцией  $\Psi(x, t)$ , может быть обнаружен в объеме  $dV$  с вероятностью  $\Psi^* \Psi dV$ . Если имеется несколько электронов, мы можем просуммировать величину  $\Psi^* \Psi$  для всех электронов и получить среднюю электронную плотность  $n(x, t)$ . Какую величину следует суммировать, для того чтобы получить плотность потока вероятности  $J(x, t)$ ?

Соответствующее выражение для плотности потока вероятности

$$J = \frac{i\hbar}{2m} \left( \Psi \frac{\partial \Psi^*}{\partial x} - \Psi^* \frac{\partial \Psi}{\partial x} \right) \quad (2.1.18)$$

можно обосновывать, исходя из следующих соображений: поскольку волновая функция  $\Psi(x, t)$  подчиняется уравнению Шредингера, можно показать, что должно выполняться соотношение

$$\frac{\partial J}{\partial x} + \frac{\partial n}{\partial t} = 0, \quad (2.1.19)$$

если  $J$  определяется формулой (2.1.18) и  $n = \Psi^* \Psi$ . Это соотношение обеспечивает выполнение условия непрерывности независимо от типа волновой функции. Плотность электрического тока получается из  $J$  умножением на заряд электрона  $-q$ .

Можно проверить непосредственно, что состояния  $+k$  и  $-k$  в уравнении (2.1.17) переносят равные по величине и противоположные по направлению ненулевые токи, пропорциональные электронной плотности:

$$J = (\hbar k/m) \Psi \Psi^*, \quad (2.1.20)$$

что подразумевает отождествление величины  $(\hbar k/m)$  со скоростью электрона  $\mu$  (поскольку мы предполагаем равенство величин  $J$  и  $n\mu$ ). Этот вывод, однако, является справедливым лишь для волновых функций, имеющих вид плоских волн, как в уравнении (2.1.17). К примеру, состояния, описывающиеся волновыми функциями в виде синуса или косинуса в уравнении (2.1.16), несут нулевой ток. Действительно, из уравнения (2.1.18) следует равенство нулю тока для любой вещественной волновой функции.

## 2.2. Метод конечных разностей

Уравнение Шредингера для атома водорода может быть решено аналитически, однако большинство других встречающихся на практике задач

требуют численного решения. В этом параграфе я расскажу об одном простом способе получения численного решения уравнения Шредингера. Большинство численных методов имеют одну общую черту — они используют один прием для замены волновой функции  $\Psi(\vec{r}, t)$  на вектор-столбец  $\psi(t)$ , а дифференциального оператора  $H_{op}$  на матрицу  $[H]$ , вследствие чего уравнение Шредингера превращается из дифференциального уравнения в частных производных  $i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}, t) = H_{op} \Psi(\vec{r}, t)$  в матричное уравнение (или уравнение в конечных разностях)  $i\hbar \frac{d}{dt} \{\psi(t)\} = [H]\{\psi(t)\}$ .

Это преобразование можно осуществить многими способами, но наиболее простой состоит в записи уравнения на дискретной решетке. Чтобы увидеть, как это делается, рассмотрим для простоты случай только одного измерения и проведем дискретизацию переменной  $x$ , описывающей положение частицы, на решетке, как это показано на рис. 2.2.1:  $x_n = na$ .

Можно представить волновую функцию  $\Psi(x, t)$  как вектор-столбец  $\{\psi_1(t) \psi_2(t) \dots \dots\}^T$  (символ  $T$  обозначает транспонирование), который содержит значения волновой функции только в точках решетки в момент времени  $t$ . Опуская для простоты переменную  $t$ , можно записать

$$\{\psi_1 \psi_2 \dots \dots\} = \{\Psi(x_1) \Psi(x_2) \dots \dots\}.$$

Это представление является точным лишь в пределе  $a \rightarrow 0$ , до тех пор, пока величина  $a$  остается малой по сравнению с пространственным масштабом, на котором изменяется  $\Psi$ .

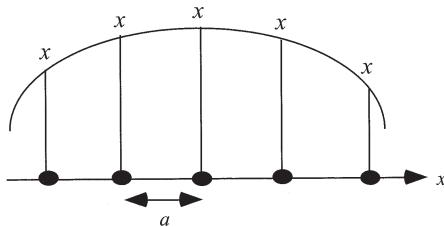


Рис. 2.2.1. Непрерывная функция может быть представлена своими значениями в совокупности точек на дискретной решетке

Следующим шагом является получение матричного представления оператора Гамильтона

$$H_{op} \equiv -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x).$$

В целом то, что мы делаем, можно назвать переходом от *дифференциального* уравнения к *разностному* уравнению. Для этой цели используется стандартная процедура, называемая методом конечных разностей:

$$\left( \frac{\partial^2 \Psi}{\partial x^2} \right) \rightarrow \frac{1}{a^2} [\Psi(x_{n+1}) - 2\Psi(x_n) + \Psi(x_{n-1})]$$

и

$$U(x)\Psi(x) \rightarrow U(x_n)\Psi(x_n).$$

Это позволяет нам записать уравнение в форме ( обратите внимание на обозначения  $t_0 \equiv \hbar^2/2ma^2$  и  $U_n \equiv U(x_n)$ )

$$\begin{aligned} i\hbar \frac{d\psi_n}{dt} &= [H_{op}\psi]_{x=x_n} = (U_0 + 2t_0)\psi_n - t_0\psi_{n-1} - t_0\psi_{n+1} = \\ &= \sum_m [(U_0 + 2t_0)\delta_{n,m} - t_0\delta_{n,m+1} - t_0\delta_{n,m-1}] \psi_m, \end{aligned} \quad (2.2.1)$$

где  $\delta_{n,m}$  есть символ Кронекера, равный единице при  $n = m$  и равный нулю при  $n \neq m$ . Уравнение (2.2.1) может быть записано в матричном виде:

$$i\hbar \frac{d}{dt} \{\psi(t)\} = [H]\{\psi(t)\}. \quad (2.2.2)$$

Элементы матрицы, описывающей гамильтониан, определяются выражением

$$H_{n,m} = [U + 2t_0]\delta_{n,m} - t_0\delta_{n,m+1} - t_0\delta_{n,m-1}, \quad (2.2.3)$$

где  $t_0 \equiv \hbar^2/2ma^2$  и  $U_n \equiv U(x_n)$ . Такая запись означает, что матрица, осуществляющая представление оператора  $\hat{H}$ , имеет вид

$$\begin{pmatrix} & 1 & 2 & \dots N-1 & N \\ 1 & 2t_0 + U_1 & -t_0 & 0 & 0 \\ 2 & -t_0 & 2t_0 + U_2 & 0 & 0 \\ & & \dots & \dots & \dots \\ N-1 & 0 & 0 & 2t_0 + U_{N-1} & -t_0 \\ N & 0 & 0 & -t_0 & 2t_0 + U_N \end{pmatrix}. \quad (2.2.4)$$

Для заданного потенциала  $U(x)$  данная матрица может быть записана явно, как только мы выберем соответствующую постоянную решетки  $a$ .

**Собственные значения и собственные вектора.** Теперь, когда мы представили уравнение Шредингера в матричной форме (уравнение (2.2.2))

$$i\hbar \frac{d}{dt} \{\psi(t)\} = [H] \{\psi(t)\},$$

возникает вопрос о том, каким образом можно вычислить  $\{\psi(t)\}$ , если задано некоторое начальное состояние  $\{\psi(0)\}$ ? Сначала следует воспользоваться стандартной процедурой и найти собственные значения  $E_\alpha$  и собственные векторы  $\{\alpha\}$  матрицы  $[H]$ :

$$[H]\{\alpha\} = E_\alpha\{\alpha\}. \quad (2.2.5)$$

Используя уравнение (2.2.5), легко убедиться в том, что волновая функция  $\{\psi(t)\} = e^{-iE_\alpha t/\hbar}\{\alpha\}$  удовлетворяет уравнению (2.2.2). Поскольку уравнение (2.2.2) является линейным, любая суперпозиция таких решений

$$\{\psi(t)\} = \sum_{\alpha} C_{\alpha} e^{-iE_{\alpha}t/\hbar}\{\alpha\} \quad (2.2.6)$$

также будет решением. Можно показать, что выражение (2.2.6) является достаточно «общим» в том смысле, что любое решение уравнения (2.2.2) может быть записано в таком виде. Для заданного начального состояния мы можем найти коэффициенты  $C_\alpha$ , так что волновая функция в последующие моменты времени  $t$  в этом случае дается выражением (2.2.6). Позднее мы обсудим, как вычисляются коэффициенты  $C_\alpha$ , но сейчас только сделаем утверждение о том, что динамику системы можно описать в терминах собственных значений и соответствующих собственных векторов, связанных с этими уровнями матрицы  $[H]$ . Вот почему первым шагом при рассмотрении любой системы является запись матрицы  $[H]$  и поиск ее собственных значений и собственных векторов. Это несложно сделать, используя стандартные математические пакеты, такие как MATLAB, и мы продемонстрируем это в следующем параграфе при обсуждении конкретных примеров.

## 2.3. Примеры

Для того чтобы закрепить навыки расчета собственных значений и собственных векторов в рамках метода конечных разностей, мы рассмотрим в этом параграфе несколько примеров. Будут рассматриваться простые задачи, допускающие аналитическое решение, что позволит использовать аналитические результаты для оценки точности наших численных решений.

Преимущество численного подхода состоит, разумеется, в том, что он позволяет столь же легко решать и более сложные задачи, для которых аналитическое решение неизвестно.

### 2.3.1. Частица в потенциальной яме

Давайте теперь рассмотрим задачу о квантовой частице в «потенциальной яме». Мы уже упоминали об этой задаче в параграфе 2.1. Пусть потенциальная энергия является постоянной и конечной внутри ямы, а на границах при  $x = 0$  и  $x = L$  (рис. 2.3.1) движение ограничено бесконечно высокими стенками. Собственные функции в этой задаче  $\phi_\alpha(x)$  описываются выражениями

$$\phi_\alpha(x) \sim \sin(k_\alpha x), \quad \text{где } k_\alpha = \alpha\pi/L, \alpha = 1, 2, \dots,$$

а соответствующие энергии равны  $E_\alpha = \hbar^2 k_\alpha^2 / 2m$ .

Данная задача может быть решена также и численно. Если волновая функция задана на дискретной решетке со 100 точками, то матрица гамильтониана  $[H]$ , которую можно записать, используя уравнение (2.2.4), где все  $U_n = 0$ , будет иметь размер  $100 \times 100$ :

$$\begin{pmatrix} 1 & 2 & \dots & 99 & 100 \\ 2t_0 & -t_0 & 0 & 0 & 0 \\ 2 & -t_0 & 2t_0 & 0 & 0 \\ & \dots & \dots & \dots & \dots \\ 99 & 0 & 0 & 2t_0 & -t_0 \\ 100 & 0 & 0 & -t_0 & 2t_0 \end{pmatrix}. \quad (2.3.1)$$

Записав элементы данной матрицы, можно обратиться к стандартному математическому пакету, например, такому как MATLAB, и найти собственные значения и соответствующие собственные вектора. Мы нашли 100 собственных значений, показанных на рис. 2.3.2 a. Их распределение с достаточно хорошей точностью соответствует аналитическому результату  $E_\alpha = \hbar^2 \pi^2 \alpha^2 / 2mL^2$ , где  $L = 101a$ , только в области малых энергий, но расходится при больших энергиях вследствие быстрых осцилляций волновой функции. Использованное при решении задачи представление оператора второй производной в конечных разностях, где  $t_0 \equiv \hbar^2 / 2ma^2$

$$-\frac{\hbar^2}{2m} \left( \frac{\partial^2 \Psi}{\partial x^2} \right)_{x=x_n} \rightarrow -t_0 [\Psi(x_{n+1}) - 2\Psi(x_n) + \Psi(x_{n-1})],$$

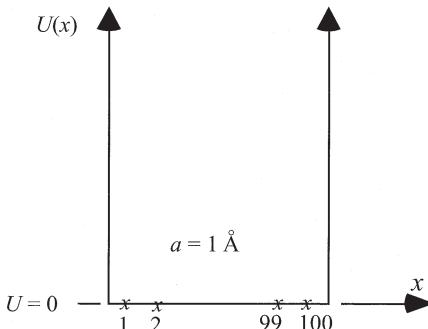


Рис. 2.3.1. Энергетические уровни «частицы в яме», найденные с использованием дискретной решетки из 100 точек с пространственным шагом  $a = 1 \text{ \AA}$

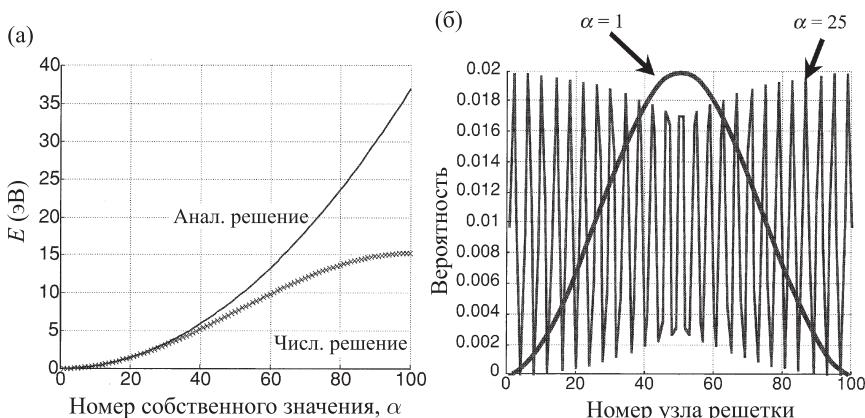


Рис. 2.3.2. (а) Результаты численного расчета (см. рис. 2.3.1) 100 собственных значений согласуются с результатами аналитического расчета в области низких энергий, но отличаются от них в силу быстрой осцилляции волновых функций в области высоких энергий. (б) Распределение плотности вероятности (квадрата волновой функции) для собственных значений с  $\alpha = 1$  и  $\alpha = 25$

является достаточно точным лишь при условии, что  $\Psi$  меняется достаточно медленно на масштабе длины  $a$ . В самом деле, если мы положим  $\Psi \sim \sin(k_\alpha x)$ , то можно непосредственно показать, что

$$-\frac{\hbar^2}{2m} \left( \frac{\partial^2 \Psi}{\partial x^2} \right)_{x=x_n} = t_0(k_\alpha a)^2 \Psi(x_n),$$

в то время как

$$-t_0[\Psi(x_{n+1}) - 2\Psi(x_n) + \Psi(x_{n-1})] = -2t_0(1 - \cos k_\alpha)\Psi(x_n).$$

Поскольку  $k_\alpha = \alpha\pi/L$ , собственные значения, найденные аналитически, лежат на параболической кривой, в то время как распределение собственных значений, найденных численно, ведет себя как косинус:

$$\text{Аналитический результат } E_\alpha = t_0(\pi\alpha a/L)^2;$$

$$\text{Численный результат } E_\alpha = 2t_0[1 - \cos(\alpha\pi a/L)].$$

Эти результаты эквивалентны лишь при  $k_\alpha a = \alpha\pi a/L \gg 1$ , когда  $\cos(k_\alpha a) \approx 1 - (k_\alpha^2 a^2/2)^2$ .

**Нормировка.** На рис. 2.3.2б показан квадрат собственной функции, отвечающий собственным значениям  $\alpha = 1$  и  $\alpha = 25$ . В аналитических расчетах функции обычно нормируются так, чтобы выполнялось условие

$$\int_{-\infty}^{+\infty} dx |\psi_\alpha(x)|^2 = 1.$$

При численных расчетах нормированный собственный вектор удовлетворяет условию

$$\sum_{n=1}^N |\phi_\alpha(x_n)|^2 = 1.$$

Следовательно, при сравнении численных и аналитических результатов следует ожидать выполнения равенства

$$\begin{array}{ccc} |\phi_\alpha(x_n)|^2 & = & |\phi_\alpha(x)|^2, \\ \text{численно} & & \text{аналитически} \end{array} \quad (2.3.2)$$

где  $a$  есть постоянная решетки (см. рис. 2.3.1). Так, в данном случае

$$\begin{array}{ccc} |\phi_\alpha(x)|^2 & = & (2/L) \sin^2(k_\alpha x) \rightarrow |\phi_\alpha(x_n)|^2 = (2a/L) \sin^2(k_\alpha x_n), \\ & & \text{аналитически} & \text{численно} \end{array}$$

Поскольку мы используем значение  $a = 1 \text{ \AA}$  и  $L = 101 \text{ \AA}$ , полученное численно распределение вероятности должно иметь максимальное значение  $2a/L \approx 0.02$ , как показано на рис. 2.3.2б.

**Границные условия.** Необходимо рассмотреть еще один вопрос: строго говоря, матрица  $[H]$  является бесконечно большой, но на практике мы всегда ограничиваем ее каким-либо конечным числом точек  $N$ . Последнее означает, что мы заменяем два граничных элемента (см. (2.2.1)):

$$-t_0 + (2t_0 + U_1)\psi_1 - t_0\psi_2 \text{ на } (2t_0 + U_1)\psi_1 - t_0\psi_2$$

и

$$-t_0\psi_{N-1} + (2t_0 + U_N)\psi_N - t_0\psi_{N+1} \text{ на } -t_0\psi_{N-1} + (2t_0 + U_N)\psi_N.$$

Фактически это означает, что мы приравниваем  $\psi_0$  и  $\psi_{N+1}$  к нулю. Данное граничное условие справедливо, если потенциал является бесконечно большим в точках 0 и  $N + 1$ , как показано на рис. 2.3.3. Истинное значение потенциала в граничных точках не будет влиять на результаты до тех пор, пока волновые функции остаются фактически равными нулю в этих точках.

Еще один широко применяемый тип граничных условий это *периодические граничные условия*, когда предполагается, что последняя точка соединена с первой точкой и, таким образом, границ у системы нет вовсе (см. рис. 2.3.4). Ранее мы уже отметили (рис. 2.1.4), что такое предположение допустимо, если мы интересуемся только свойствами внутренней области структуры. Тогда мы можем принять любую их форму граничных условий, если она упрощает наши расчеты.

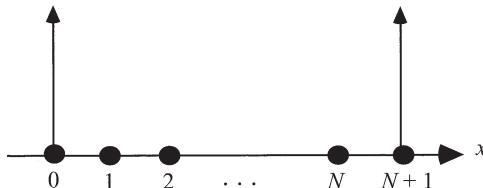


Рис. 2.3.3. Граничные условия, при которых  $\psi_0 = 0$  и  $\psi_{N+1} = 0$ , могут быть использованы, если мы предполагаем бесконечно большое значение потенциала в точках 0 и  $N + 1$

Математически введение периодических граничных условий приводит к следующему изменению гамильтоновской матрицы

$$H = \begin{matrix} & 1 & 2 & \dots & 99 & 100 \\ 1 & 2t_0 & -t_0 & 0 & -t_0 \\ 2 & -t_0 & 2t_0 & 0 & 0 \\ & \dots & \dots & \dots & \dots \\ 99 & 0 & 0 & 2t_0 & -t_0 \\ 100 & -t_0 & 0 & -t_0 & 2t_0 \end{matrix} \quad (2.3.3)$$

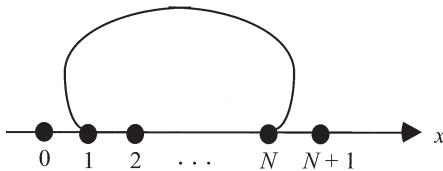


Рис. 2.3.4. Периодические граничные условия предполагают, что у системы нет граничных точек. Точка  $N$  вновь соединяется с точкой 1, как если бы структура имела форму кольца, где точка  $(N + 1)$  эквивалентна точке 1

Обратите внимание на то, что единственное изменение в матрице по сравнению с граничными условиями на бесконечно высокой стенке (см. уравнение (2.3.1)) касается только элементов  $H(1, 100)$  и  $H(100, 1)$ . Эта операция меняет собственные значения и собственные вектора, но изменения незаметны, если число точек на решетке велико. Собственные функции теперь записываются как

$$\phi_\alpha(x) \sim \sin(k_\alpha x) \quad \text{и} \quad \cos(k_\alpha x),$$

где  $k_\alpha = \alpha 2\pi/L$ ,  $\alpha = 1, 2, \dots$ , вместо

$$\phi_\alpha(x) \sim \sin(k_\alpha x),$$

где  $k_\alpha = \alpha \pi/L$ ,  $\alpha = 1, 2, \dots$

Значения  $k\alpha$  отстоят друг от друга на расстоянии  $2\pi/L$  вместо  $\pi/L$ , так что число разрешенных значений уменьшается наполовину, но собственные значения теперь появляются парами, как это следует из рис. 2.3.5. Каждому собственному значению  $k_\alpha$  при этом отвечают функции синус и косинус.

Как отмечалось ранее, вместо записи собственных состояний в форме

$$\cos(k_\alpha x) \quad \text{и} \quad \sin(k_\alpha x)$$

мы с тем же успехом можем их записать как

$$e^{ik_\alpha x} = \cos(k_\alpha x) + i \sin(k_\alpha x) \quad \text{и} \quad e^{-ik_\alpha x} = \cos(k_\alpha x) - i \sin(k_\alpha x).$$

Такая запись часто используется в аналитических расчетах, но численные расчеты в типичных случаях будут выдавать собственные вектора в форме  $\cos(k_\alpha x)$  и  $\sin(k_\alpha x)$ . Обе формы одинаково правильны, и лишь для конкретных расчетов одна из них может оказаться предпочтительнее другой.

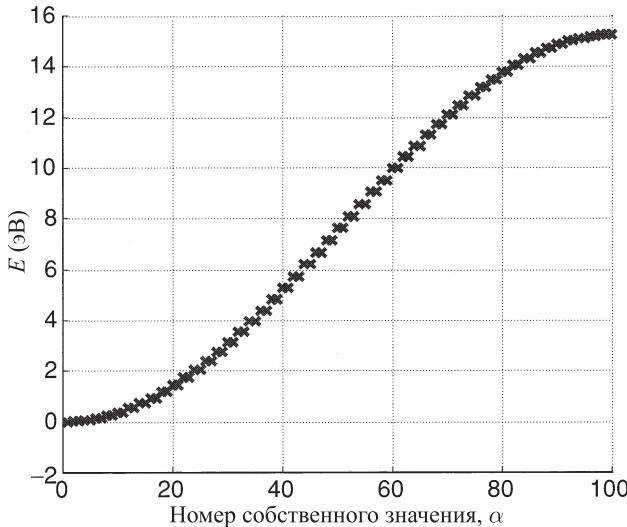


Рис. 2.3.5. Собственные значения энергии в яме длиной 101 Å с периодическими граничными условиями (те же, что и на рис. 2.3.1): теперь они объединены в пары

**Число собственных значений.** При проведении численных расчетов мы всегда получаем конечное число собственных значений, в то время как при аналитическом решении это число собственных значений бесконечно. Этого следовало ожидать, поскольку матрица конечных размеров может иметь лишь конечное число собственных значений. Не следует удивляться тому, что мы не получаем бесконечного числа значений  $E_\alpha$ , отвечающих бесконечной последовательности  $k_\alpha a = \alpha 2\pi a/L$ , как это имеет место при аналитическом решении. Причина состоит в том, что для дискретной решетки волновые функции

$$\sin(k_\alpha x) \quad \text{и} \quad \sin([k_\alpha + (2\pi/a)]x)$$

представляют собой одно и то же состояние, поскольку для любого узла решетки  $x_n = na$  имеет место

$$\sin(k_\alpha x_n) = \sin([k_\alpha + (2\pi/a)]x_n).$$

Эти функции не равны друг другу в промежутке между двумя узлами решетки и поэтому представляют различные состояния в непрерывной модели. Но при переходе к описанию на дискретной решетке значения  $k_\alpha$ ,

отличающиеся на  $2\pi/a$ , будут представлять идентичные состояния, и лишь значения  $k_\alpha a$ , лежащие в интервале с размерами  $2\pi$ , будут соответствовать одинаковым решениям. Поскольку  $k_\alpha a = \alpha\pi a/L = \alpha\pi/N$ , это означает, что необходимо рассматривать лишь  $N$  различных значений  $\alpha$ . Обычно в качестве интервала изменения  $k_\alpha a$  выбирается первая зона Бриллюэна

$$-\pi < k_\alpha a < +\pi \quad \text{для периодических граничных условий}$$

и

$$0 < k_\alpha a < +\pi \quad \text{для граничных условий с бесконечно высокой стенкой.}$$

### 2.3.2. Частица в трехмерной потенциальной яме

Выше мы обсудили метод конечных разностей на примере одномерной задачи. Однако основная идея этого метода может быть распространена на случай двух и трех измерений. Диагональные элементы матрицы  $[H]$  тогда будут равны величине  $t_0$ , умноженной на число ближайших соседей (два в случае одного измерения, четыре в случае двух измерений и шесть в случае трех измерений) плюс слагаемое  $U(\vec{r})$ , т. е. потенциальная энергия в узле решетки. Недиагональные элементы для ближайших узлов решетки будут равны  $-t_0$ . Следовательно, имеем

$$\begin{aligned} H_{nm} &= \nu t_0 & n = m, \\ H_{nm} &= -t_0 & n \text{ и } m \text{ есть ближайшие соседи}, \\ H_{nm} &= 0 & \text{в других случаях}. \end{aligned} \tag{2.3.4}$$

где  $\nu$  есть число ближайших соседей.

Однако в случае двух или трех измерений мы сталкиваемся с практическими трудностями. Если узлы решетки отстоят друг от друга на расстояние в  $1 \text{ \AA}$ , то одномерная задача с  $L = 101 \text{ \AA}$  требует матрицы  $[H]$  размером  $100 \times 100$ . Однако в трехмерном случае потребуется уже матрица размером  $10^6 \times 10^6$ . Это означает, что практически мы ограничены задачами очень малого размера. Однако если координатные переменные разделяются, то мы можем работать с тремя отдельными одномерными задачами вместо одной гигантской трехмерной задачи. Такая ситуация возможна, если потенциал может быть разделен на части, зависящие от  $x$ ,  $y$  и  $z$ :

$$U(\vec{r}) = U_x(x) + U_y(y) + U_z(z). \tag{2.3.5}$$

Волновая функция в этом случае может быть записана в виде произведения:

$$\Psi(\vec{r}) = X(x)Y(y)Z(z),$$

где каждая из функций  $X(x)$ ,  $Y(y)$  и  $Z(z)$  находится как решение отдельного одномерного уравнения Шредингера:

$$\begin{aligned} E_x X(x) &= \left( -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U_x(x) \right), \\ E_y Y(y) &= \left( -\frac{\hbar^2}{2m} \frac{d^2}{dy^2} + U_y(y) \right), \\ E_z Z(z) &= \left( -\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + U_z(z) \right). \end{aligned} \quad (2.3.6)$$

Полная энергия  $E$  равна сумме энергий, связанных с каждым измерением:  $E = E_x + E_y + E_z$ .

**Сферически симметричный потенциал.** Некоторые задачи могут не допускать разделения переменных в декартовых координатах, но позволяют осуществить его в цилиндрических или сферических координатах. Например, в задаче об атоме водорода с потенциалом  $U(\vec{r}) = -q^2/4\pi\varepsilon_0 r$  переменные не разделяются в декартовых координатах  $(x, y, z)$ . Однако они могут быть разделены в сферических координатах  $(r, \theta, \phi)$ , где волновая функция может быть записана в виде

$$\Psi(r, \theta, \phi) = [f(r)/r]Y_l^m(\theta, \phi). \quad (2.3.7)$$

Здесь радиальная часть волновой функции  $f(r)$  может быть найдена как решение радиального волнового уравнения:

$$Ef(r) = \left( -\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + U(r) + \frac{l(l+1)\hbar^2}{2mr^2} \right) f(r), \quad (2.3.8)$$

где  $l = 0$  для  $s$ -уровней,  $l = 1$  для  $p$ -уровней и т. д. Функции  $Y_l^m(\theta, \phi)$  представляют собой сферические гармоники, определяемые выражениями

$$Y_0^0(\theta, \phi) = \sqrt{1/4\pi},$$

$$Y_1^0(\theta, \phi) = \sqrt{3/4\pi} \cos \theta,$$

$$Y_0^{\pm 1}(\theta, \phi) = \sqrt{3/8\pi} \sin \theta e^{\pm i\phi}$$

и т. д. Уравнение (2.3.8) может быть решено численно с использованием метода конечных разностей, который мы описали выше.

**Нормировка.** Следует обратить внимание на то, что полные волновые функции должны быть нормированы так, чтобы выполнялось условие

$$\int_0^\infty dr r^2 \int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi |\Psi|^2 = 1.$$

Поскольку согласно (2.3.7)

$$\Psi(r, \theta, \phi) = [f(r)/r] Y_l^m(\theta, \phi),$$

и сферические гармоники нормированы так, что

$$\int_0^\pi d\theta \sin \theta \int_0^{2\pi} d\phi |Y_l^m|^2 = 1,$$

нетрудно увидеть, что радиальная часть волновой функции  $f(r)$  удовлетворяет условию нормировки

$$\int_0^\infty dr |f(r)|^2 = 1. \quad (2.3.9)$$

Отсюда следует, что функция  $|f(r)|^2$  определяет радиальное распределение плотности вероятности, так что величина  $|f(r)|^2 \Delta r$  равна вероятности обнаружить электрон в сферическом слое между значениями радиуса  $r$  и  $(r + \Delta r)$ . Результаты численных расчетов с постоянной решеткой  $a$  со-поставлены со значениями величины  $|f(r)|^2 a$ , найденными аналитически. К примеру, для уровней  $1s$  и  $2s$  имеем

$$|f_{1s}|^2 a = (4ar^2/a_0^3)e^{-2r/a_0}, \quad (2.3.10a)$$

$$|f_{2s}|^2 a = (ar^2/8a_0^3) \left(2 - \frac{r}{a_0}\right) e^{-2r/a_0}. \quad (2.3.10b)$$

**Численные результаты.** Если решение проведено на решетке со 100 узлами и постоянная решетки есть  $a = 0.05$  Å, то два наименьших собственных значения с  $l = 0$  (что отвечает уровням  $1s$  и  $2s$ ) равны

$$E_{1s} = -13.56 \text{ эВ} \text{ и } E_{2s} = -2.96 \text{ эВ},$$

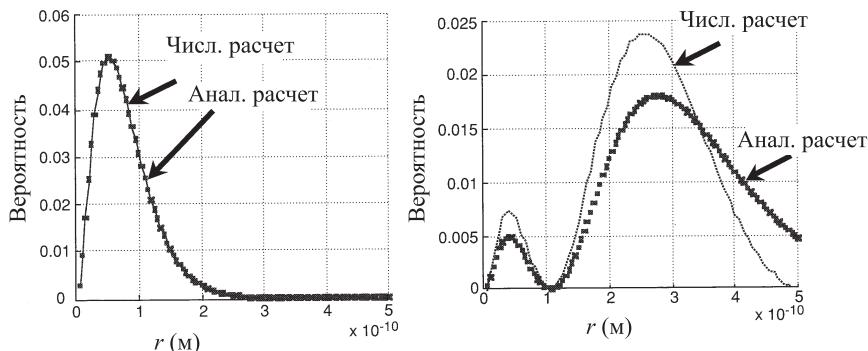


Рис. 2.3.6. Распределение радиальной плотности вероятности  $|f(r)|^2$ , отвечающее двум низшим собственным значениям  $E_{1s} = -13.56$  эВ и  $E_{2s} = -2.96$  эВ для  $l = 0$  (что отвечает уровням  $1s$  и  $2s$ ). Точки показаны решения уравнений (2.3.10 а, б), найденные аналитически, сплошными линиями показаны результаты численного расчета, полученного в модели решетки со 100 узлами и постоянной решетки  $a = 0.05$  Å

что сопоставимо с аналитическими результатами (см. уравнение (2.2.6))  $E_{1s} = -13.59$  эВ и  $E_{2s} = -3.4$  эВ. Видно, что значения энергии уровня  $1s$  хорошо согласуются, в то время как для уровня  $2s$  расхождение становится существенным. Причину этого легко объяснить, построив графики соответствующих функций  $|f(r)|^2$ , и сравнить их с аналитическими результатами. Из рис. 2.3.6 следует, что имеется хорошее соответствие между численным и аналитическим решениями для волновой функции уровня  $1s$ , в то время как между функции уровня  $2s$  имеется значительное расхождение. Этую проблему можно разрешить, выбрав большие значения постоянной решетки, а именно,  $a = 0.1$  Å. Рисунок 2.3.7 показывает, что теперь рассчитанные волновые функции достаточно хорошо соответствуют аналитическим результатам, и рассчитанное значение энергии уровня  $2s$ , равное  $-3.39$  эВ, находится в хорошем согласии с аналитическим результатом. В то же время значение энергии состояния  $1s$  несколько уменьшается до  $-13.47$  эВ, поскольку теперь волновая функция представлена не достаточно частым набором точек. Мы могли бы улучшить соответствия как для  $1s$ , так и для  $2s$  состояний, использовав представление волновой функции на решетке из 200 точек с постоянной решеткой  $a = 0.05$  Å. Это позволило бы сохранить как достаточно подробное описание, так и достаточно большой простран-

ственный промежуток. В этом случае вычисления заняли бы больше времени, поскольку нам пришлось бы вычислить собственные значения матрицы размером  $(200 \times 200)$ , а не  $(100 \times 100)$ .

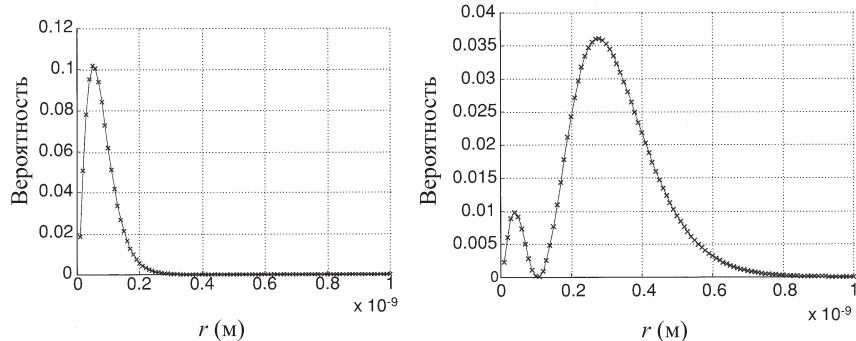


Рис. 2.3.7. Распределение радиальной плотности вероятности  $|f(r)|^2$ , отвечающее двум низшим собственным значениям  $E_{1s} = -13.47$  эВ и  $E_{2s} = -3.39$  эВ для  $l = 0$  (что отвечает уровням  $1s$  и  $2s$ ). Сплошной линией представлен результат аналитического решения уравнений (2.3.10 *a,b*)), крестики « $\times$ » показывают результат численного расчета, полученного в модели решетки со 100 узлами и постоянной решетки  $a = 0.1$  Å

Этот простой пример иллюстрирует важные моменты, которые необходимо иметь в виду при выборе решетки для проведения численных расчетов. Постоянная решетки  $a$ , с одной стороны, должна быть достаточно малой для того, чтобы волновая функция была представлена адекватно, и в то же время период решетки должен быть достаточно большим для того, чтобы была представлена вся интересующая нас область изменения волновой функции в пространстве. Если потребовалось бы точное описание всех собственных состояний, наша задача стала бы безнадежной. Однако нам, как правило, необходима информация о собственных состояниях, расположенных в определенном интервале энергий, и в соответствии с этим мы должны выбирать матрицу  $[H]$  таким образом, чтобы обеспечить требуемую точность в этом интервале энергий.

## Упражнения

**У2.1.(а)** Используйте дискретную решетку со 100 узлами и шагом решетки 1 Å для расчета энергий частицы в потенциальной яме с бесконечно

высокими стенками и сравните со значением энергии  $E_\alpha = \hbar^2 \pi^2 \alpha^2 / 2mL^2$  (ср. с рис. 2.3.2 а). Постройте на рисунке распределение плотности вероятности (квадрата волновой функции) для собственных значений  $\alpha = 1$  и  $\alpha = 50$  (ср. с рис. 2.3.2 б). (б) Найдите собственные значения, используя периодические граничные условия, и сравните с рис. 2.3.5.

**У.2.2.** (а) Получите радиальное уравнение, указанное в формуле (2.3.8), при помощи следующей схемы: (1) Запишите оператор  $\nabla^2$  из уравнения Шредингера в сферических координатах:

$$\nabla^2 \equiv \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right);$$

(2) Обратите внимание на то, что сферические гармоники  $Y_m^l(\theta, \phi)$  являются собственными функциями угловой части оператора Лапласа:

$$\left( \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right) Y_l^m = -l(l+1) Y_l^m;$$

(3) Запишите волновую функцию в форме  $\Psi(r, \theta, \phi) = \psi(r) Y_l^m(\theta, \phi)$  и обратите внимание на то, что

$$\nabla^2 \psi = \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{l(l+1)}{r^2} \right) \psi,$$

(4) что позволяет упростить уравнение Шредингера и записать для радиальной части выражение

$$E\psi = \left( -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) + U(r) + \frac{\hbar^2 l(l+1)}{2mr^2} \right) \psi,$$

из которого, наконец, (5) записав подстановку  $\psi(r) = f(r)/r$ , получить выражение (2.3.8) для  $f(r)$ .

(б) Используйте дискретную решетку со 100 узлами и шагом  $a$  для решения уравнения (2.3.8)

$$Ef(r) = \left( -\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + U(r) + \frac{l(l+1)\hbar^2}{2mr^2} \right) f(r)$$

для уровней энергии  $1s$  и  $2s$  атома водорода. Постройте соответствующие радиальные плотности распределения вероятности  $|f(r)|^2$  и сравните их с аналитическим результатом для (а)  $a = 0.05 \text{ \AA}$  (ср. с рис. 2.3.6) и (б) с  $a = 0.1 \text{ \AA}$  (ср. с рис. 2.3.7).

Строго говоря, необходимо заменять массу электрона приведенной массой для учета движения ядра, но эта поправка является малой по сравнению с принятым нами сейчас уровнем точности.

**У.2.3.** Используйте выражение (2.1.18) для расчета плотности тока, связанного с движением электрона, имеющего волновую функцию

$$\Psi(x, t) = (e^{+\gamma x} + ae^{-\gamma x})e^{-iEt/\hbar},$$

предполагая, что  $\gamma$  является (а) чисто мнимой ( $= i\beta$ ) и (б)  $\gamma$  является действительной.

## **Глава 3**

# **Приближение самосогласованного поля**

При переходе от атома водорода, в котором только один электрон, к многоэлектронным атомам мы сталкиваемся с серьезной и еще не полностью решенной проблемой электрон-электронных взаимодействий. В этой главе я планирую обсудить следующие вопросы: (1) метод самосогласованного поля (ССП), позволяющий приблизенно включить электрон-электронное взаимодействие в уравнение Шредингера (параграф 3.1); (2) интерпретацию энергетических уровней, полученных из т. н. «одноэлектронного» уравнения Шредингера (параграф 3.2); (3) энергетические соображения, описывающие процесс, в котором атомы «связываются» для образования молекул (параграф 3.3). Наконец, в дополнительном параграфе 3.4 будут развиты некоторые положения параграфа 2.3, которые могут вызвать интерес у заинтересовавшихся читателей.

### **3.1. Метод самосогласованного поля (ССП)**

Сразу после создания теории атома водорода одним из первых успехов квантовой механики стало объяснение структуры периодической таблицы атомов. Это объяснение базировалось как на анализе решений уравнения Шредингера, так и на учете принципа Паули, который требует, чтобы каждое состояние было заполнено не более чем одним электроном. Как известно, собственные значения энергии атома для каждого значения квантового числа  $l$ , начиная с  $l = 0$  (см. уравнение (2.3.8)), нумеруются целыми числами  $n$  начиная от  $n = l + 1$ . Для любых значений  $(n, l)$  имеется  $(2l + 1)$  состояний, описываемых функциями с определенным квантовым числом  $m$ , которые принадлежат одному значению уровня энергии. Для каждой тройки чисел  $(n, l, m)$  имеется уровень с проекцией спина вверх и уровень с проекцией спина вниз, и поэтому для данной пары  $(n, l)$  кратность вырождения уровней равна  $2l(l + 1)$ . Картина энергетических уровней имеет приблизительно такой вид, как показано на рис. 3.1.1.

Элементы периодической таблицы расположены по мере возрастания числа электронов в атомах. Их электронная структура может быть записана следующим образом: водород,  $1s^1$ ; гелий,  $1s^2$ ; литий,  $1s^2 2s^1$ ; бериллий,  $1s^2 2s^2$ ; бор,  $1s^2 2s^2 2p^1$  и т. д., где верхний индекс обозначает число электронов, занимающих соответствующую орбиталь.

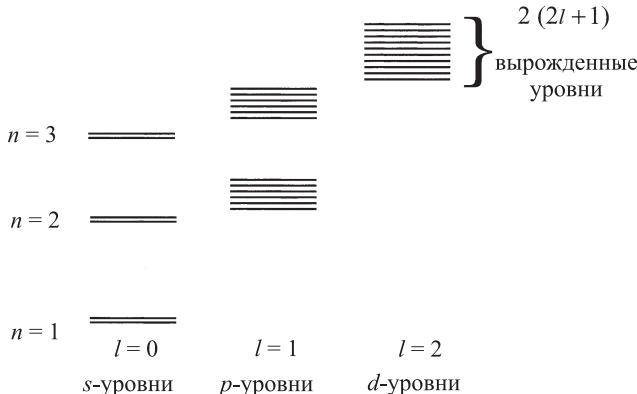


Рис. 3.1.1

Как можно рассчитать уровни энергии многоэлектронного атома? Стационарное уравнение Шредингера

$$E_\alpha \Phi_\alpha(\vec{r}) = H_{\text{оп}} \Phi_\alpha(\vec{r}), \quad \text{где} \quad H_{\text{оп}} \equiv -\frac{\hbar^2}{2m} \nabla^2 + U(\vec{r}),$$

обеспечивает вполне точное описание наблюдаемых спектров всех атомов, а не только атома водорода. Однако описание многоэлектронных атомов включает электрон-электронные взаимодействия, которые учитываются добавлением «самосогласованного поля (ССП)»  $U_{\text{SCF}}(\vec{r})$  к потенциалу ядра  $U_{\text{nuc}}(\vec{r})$ :  $U(\vec{r}) = U_{\text{nuc}}(\vec{r}) + U_{\text{SCF}}(\vec{r})$ , подобно тому, как в параграфе 1.4 мы добавляли дополнительный потенциал к потенциалу Лапласа  $U_L$  (см. уравнение (1.4.1b)). Потенциал ядра  $U_{\text{nuc}}(\vec{r})$ , как и  $U_L$ , является заданным, в то время как  $U_{\text{SCF}}(\vec{r})$  зависит от электронных волновых функций и должен рассчитываться с помощью самосогласованной итерационной процедуры. В этом параграфе мы опишем данную процедуру и обсудим связанные с ней проблемы.

Прежде всего, рассмотрим атом гелия, состоящий из двух электронов, связанных с ядром, несущим два положительных заряда  $+2q$ . Как будут

выглядеть уровни энергии этого атома? Сначала будем действовать, как при решении задачи об атоме водорода, правда, полагая потенциал равным

$$U(\vec{r}) = -2q^2/4\pi\varepsilon_0 r$$

вместо

$$U(\vec{r}) = -q^2/4\pi\varepsilon_0 r.$$

Решая уравнение Шредингера с потенциалом  $U(\vec{r}) = -Zq^2/4\pi\varepsilon_0 r$ , мы найдем уровни энергии, описываемые выражением

$$E_n = -(Z^2/n^2)E_0 = -54.4 \text{ эВ}/n^2 \quad (Z = 2),$$

как это предсказывает простая модель Бора (см. уравнения (2.1.6a, б)). Однако этот результат совсем не согласуется с данными экспериментов. К примеру, потенциал ионизации гелия равен  $\approx 25$  эВ, т. е. для ионизации атома гелия необходимо иметь фотон с энергией как минимум в 25 эВ:



Это означает, что уровень  $1s$  в атоме гелия имеет энергию  $-25$  эВ, а не  $-54.4$  эВ, как предсказывает простая модель. Как же мы смогли ошибиться более чем на 30 эВ? Оказывается, это произошло потому, что мы не учли взаимодействие со вторым электроном атома гелия. Если мы измерим энергию, которая необходима для удаления второго электрона из  $\text{He}^+$



то результат (известный как второй ионизационный потенциал) будет действительно близок к 54.4 эВ. Но при этом первый потенциал ионизации будет приблизительно на 30 эВ меньше. Значит, для удаления электрона из нейтрального атома гелия потребуется энергия, которая на 30 эВ меньше чем энергия, необходимая для удаления электрона из иона гелия ( $\text{He}^+$ ), который уже потерял один электрон. Причина этого в том, что электрон в атоме гелия испытывает отталкивание со стороны другого электрона, и это эффективно увеличивает его энергию на 30 эВ и, соответственно, облегчает процесс ионизации атома (рис. 3.1.2).

В общем случае уровни ионизации многоэлектронных атомов могут быть приближенно рассчитаны из уравнения Шредингера путем добавления к потенциальному ядру  $U_{\text{nuc}}(\vec{r})$  потенциала самосогласованного поля  $U_{\text{SCF}}(\vec{r})$ , который обусловлен влиянием других электронов (рис. 3.1.3):

$$U(\vec{r}) = U_{\text{nuc}}(\vec{r}) + U_{\text{SCF}}(\vec{r}).$$
 (3.1.2)

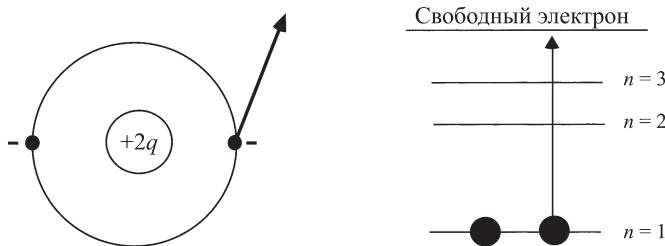


Рис. 3.1.2. Для ионизации нейтрального атома гелия требуется приблизительно 25 эВ энергии, что позволяет определить значение энергии основного состояния  $n = 1$  как равное –25 эВ

Для всех атомов потенциал ядра возникает вследствие наличия заряда ядра  $+Zq$ , расположенного в начале координат и описываемого потенциалом  $U_{\text{nuc}}(\vec{r}) = -Zq^2/4\pi\varepsilon_0 r$ . Самосогласованное поле возникает из-за наличия остальных  $(Z - 1)$  электронов, поскольку электрон не чувствует потенциала, создаваемого им самим. Чтобы рассчитать потенциал  $U_{\text{SCF}}(\vec{r})$ , необходимо знать электронные волновые функции, которые, в свою очередь, должны рассчитываться из уравнения Шредингера, содержащего потенциал  $U_{\text{SCF}}(\vec{r})$ . Это означает, что расчеты должны выполняться самосогласованно, по обозначенной ниже схеме.

*Шаг 1.* Представить приблизительно электронный потенциал  $U_{\text{SCF}}(\vec{r})$ .

*Шаг 2.* Найти собственные функции и собственные значения из уравнения Шредингера.

*Шаг 3.* Рассчитать электронную плотность  $n(\vec{r})$ .

*Шаг 4.* Рассчитать электронный потенциал  $U_{\text{SCF}}(\vec{r})$ .

*Шаг 5.* Если вновь рассчитанный потенциал  $U_{\text{SCF}}(\vec{r})$  существенно отличается от предыдущего приближения, необходимо заменить это приближение на вновь рассчитанное и вернуться к шагу 2. Если же полученный новый потенциал  $U_{\text{SCF}}(\vec{r})$  отличается от предыдущего, скажем, не более чем на 10 мэВ, то можно сделать вывод, что имеет место сходимость, и завершить расчет.

При выполнении *Шага 2* мы фактически можем пользоваться тем же методом, который был использован для атома водорода, хотя аналитическое решение обычно не представляется возможным. Потенциал  $U_{\text{SCF}}(\vec{r})$  в общем случае не является изотропным (т. е. зависит от угловых переменных  $(\theta, \phi)$ ), но для атомов его можно считать изотропным без опасения получить какую-либо существенную ошибку. При этом, зависимость от переменной  $r$  оказывается достаточно сложной, что делает невозможным ана-

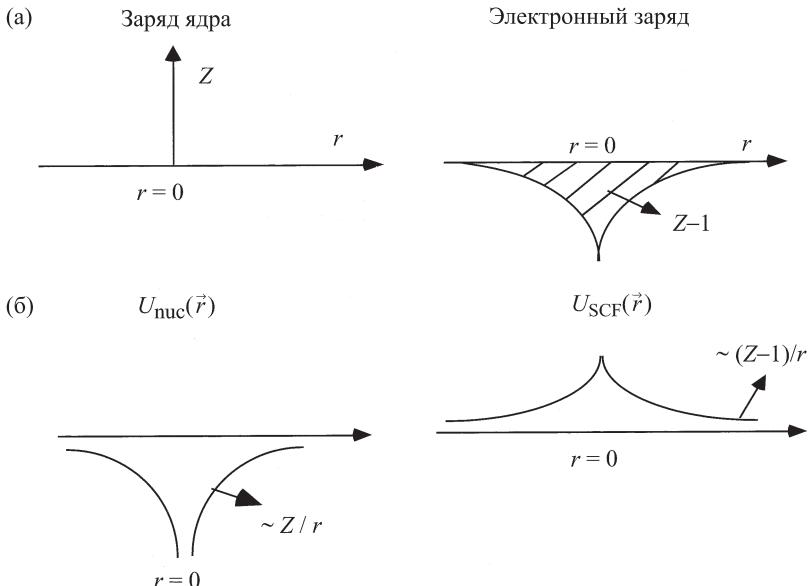


Рис. 3.1.3. Схематическое изображение (а) распределения зарядовой плотности на ядре (слева) и электронной зарядовой плотности (справа); (б) потенциальная энергия ядра  $U_{\text{нuc}}(\vec{r})$ , которую «видит» дополнительный электрон (слева), и потенциал  $U_{\text{SCF}}(\vec{r})$ , обусловленный наличием других электронов (справа). Последний должен рассчитываться самосогласованным образом

литическое решение. Использование численных методов, однако, позволяет решить уравнение Шредингера для любого потенциала  $U(r)$ , что оказывается столь же просто, как решить задачу для атома водорода с  $U(r) \sim 1/r$ .

При выполнении *Шага 3* необходимо просуммировать распределения вероятности по всем занятым собственным состояниям:

$$n(\vec{r}) = \sum_{\text{occ } \alpha} |\Phi_\alpha(\vec{r})|^2 = \sum_{\text{занятые } n,l,m} \left| \frac{f_n(r)}{r} \right|^2 |Y_{lm}(\theta, \phi)|^2. \quad (3.1.3)$$

Если мы предположим, что распределение плотности заряда является изотропным, то сможем написать

$$\sigma(r) \equiv \int r^2 \sin \theta d\theta d\phi n(\vec{r}) = \sum_{\text{занятые } n,l,m} |f_n(r)|^2. \quad (3.1.4)$$

При выполнении *Шага 4* можно непосредственно использовать формулу электростатики, чтобы найти

$$U_{\text{SCF}}(r) = \frac{Z-1}{Z} \left[ \frac{q^2}{4\pi\epsilon_0 r} \int_0^r dr' \sigma(r') + \frac{q^2}{4\pi\epsilon_0} \int_0^\infty dr' \frac{\sigma(r')}{r'} \right]. \quad (3.1.5)$$

Два слагаемых в уравнении (3.1.5) определяют вклады в заряд от сферы радиуса  $r$ , а также от распределения заряда вне этой сферы, как это показано на рис. 3.1.4. Первое слагаемое есть значение потенциала на расстоянии  $r$  от центра сферы вне ее. Можно показать, что его значение совпадает со случаем, когда весь заряд сконцентрирован в центре сферы:

$$\frac{q^2}{4\pi\epsilon_0 r} \int_0^r dr' \sigma(r').$$

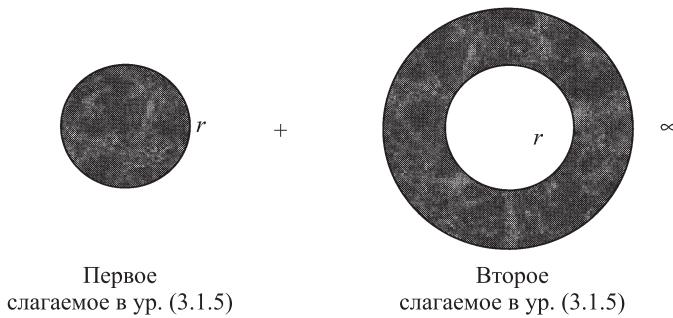


Рис. 3.1.4

Второе слагаемое представляет собой значение потенциала на расстоянии  $r$  от центра заряженной сферы в точке внутри нее. Можно показать, что его значение совпадает со значением потенциала в центре сферы (потенциал является постоянным во всех точках внутри сферы, поскольку электрическое поле там равно нулю):

$$\frac{q^2}{4\pi\epsilon_0} \int_0^\infty dr' \frac{\sigma(r')}{r'}.$$

Полный потенциал найдем, складывая две компоненты.

Чтобы объяснить появление множителя  $(Z - 1)/Z$  в уравнении (3.1.5), заметим, что правильная зарядовая плотность для каждого собственного состояния не должна включать само рассматриваемое состояние, поскольку электрон не чувствует свой собственный потенциал. К примеру, у кремния имеется 14 электронов  $1s^2 2s^2 2p^6 3s^2 3p^2$ , и самосогласованное поле учитывает их все, кроме одного — для уровня  $3p$  мы исключаем  $3p$ -электрон, для уровня  $3s$  исключаются  $3s$ -электрон и т. д. Более удобно, однако, просто рассмотреть полную плотность заряда и умножить ее на  $(Z - 1)/Z$ . Такая операция сохраняет сферическую симметрию в распределении заряда, и возникающее различие, обычно, не бывает существенным. Обратите внимание на то, что полный электронный заряд равен  $Z$ :

$$\int_0^\infty dr \sigma(r) = \sum_{\text{занятые } n,l,m} 1 = Z, \quad (3.1.6)$$

поскольку радиальные волновые функции являются нормированными:  $\int_0^\infty dr |f_n(r)|^2 = 1$ .

**Атом гелия.** На рис. 3.1.5 показан профиль потенциала и распределение плотности вероятности для  $1s$ -состояния гелия, полученный методом ССП, который мы только что обосновали. Для сравнения также показаны результаты для  $1s$ -уровня атома водорода, которые обсуждались в предыдущей главе.

**Атом кремния.** На рис. 3.1.6 приведено распределение плотности вероятности для состояний  $1s$  и  $2p$  в кремнии, полученное с использованием метода ССП. Для сравнения показано также подобное распределение для уровня  $1s$  в атоме водорода. Легко видеть, что волновая функция состояния  $1s$  в кремнии более сильно ограничена в пространстве, нежели в состоянии  $1s$  атома водорода. Такая ситуация типична для глубоких уровней, и это объясняет, почему подобные состояния в твердых телах остаются сильно локализованными, в то время как электроны на внешних оболочках (такие как  $3p$ ) являются делокализованными.

## 3.2. Связь с многоэлектронной картиной

**Многоэлектронное уравнение Шредингера.** Важно отдавать себе отчет в том, что метод ССП является приближенным и широко используется лишь потому, что зачастую не удается применить более точные методы. К примеру, если требуется рассчитать собственные состояния атома гелия

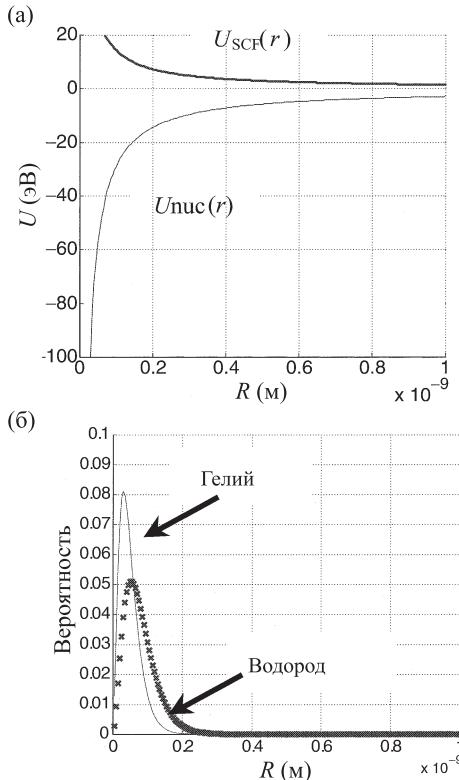


Рис. 3.1.5. Пример использования метода самосогласованного поля для расчета состояний атома гелия. (а) Потенциал ядра  $U_{\text{nuc}}(r)$  и самосогласованный электронный потенциал  $U_{\text{SCF}}(r)$ . (б) Радиальное распределение плотности вероятности для  $1s$ -состояния в атомах гелия и водорода

с двумя электронами, то необходимо решить двухэлектронное уравнение Шредингера вида

$$E\Psi(\vec{r}_1, \vec{r}_2) = \left( -\frac{\hbar^2}{2m}\nabla_1^2 - \frac{\hbar^2}{2m}\nabla_2^2 + U(\vec{r}_1) + U(\vec{r}_2) + U_{ee}(\vec{r}_1, \vec{r}_2) \right) \Psi(\vec{r}_1, \vec{r}_2), \quad (3.2.1)$$

где  $\vec{r}_1$  и  $\vec{r}_2$  есть координаты двух электронов, а  $U_{ee}$  представляет собой потенциальную энергию их взаимного отталкивания:  $U_{ee}(\vec{r}_1, \vec{r}_2) = e^2/4\pi\varepsilon_0|\vec{r}_1 - \vec{r}_2|$ . Такую задачу решить значительно труднее, чем одно-

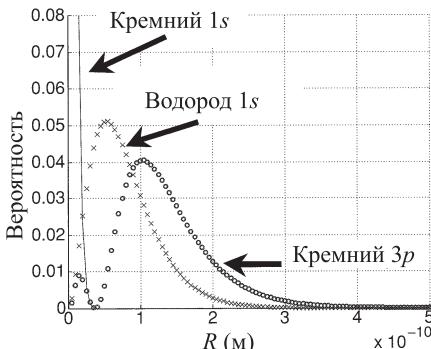


Рис. 3.1.6. Пример использования метода самосогласованного поля для расчета состояний атома кремния. Показаны распределения радиальной плотности вероятности для уровней 1s атома водорода, а также для уровней 1s и 3p атома кремния

электронное уравнение Шредингера, но, в принципе, это возможно. Однако как только мы переходим к многоэлектронным атомам, этот подход становится слишком громоздким даже с вычислительной точки зрения, поэтому достаточно редко применяется напрямую. Тем не менее, предположим, что мы смогли рассчитать энергетические уровни многоэлектронного атома. Как можно использовать наши предыдущие результаты (в принципе, а не на практике), чтобы построить *диаграмму энергетических уровней одного электрона*, подобную той, что мы строили ранее? Ответ зависит от того, какую информацию мы хотим получить, изучая систему одноэлектронных энергетических уровней.

**Энергии ионизации и энергии сродства.** В этой книге нас интересует, в основном, процесс протекания тока, при котором происходит переход в систему нового электрона и последующий его уход из системы, либо наоборот, как это обсуждалось в главе 1. Следовательно, нам хотелось бы, чтобы одноэлектронные уровни определяли энергию, необходимую для удаления электрона из атома (энергию ионизации), либо энергию, которая требуется для введения электрона в атом (энергию сродства) (рис. 3.2.1).

При рассмотрении энергий ионизации одноэлектронные энергии  $\varepsilon_n$  представляют собой разность между энергией основного состояния  $E_G(N)$  нейтрального атома с  $N$  электронами и  $n$ -м энергетическим уровнем  $E_n(N-1)$  положительно заряженного атома с  $(N-1)$  электронами:

$$\varepsilon_n = E_G(N) - E_n(N-1). \quad (3.2.2a)$$

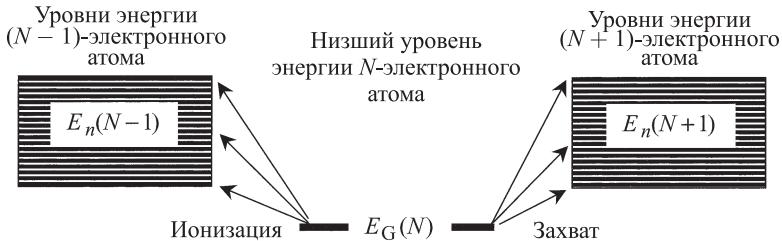
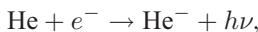


Рис. 3.2.1. Одноэлектронные энергетические уровни описывают разности энергий между уровнями атома с  $N$  электронами и атома с  $(N - 1)$  или  $(N + 1)$  электронами. Первые (называемые энергиями ионизации) представляют собой заполненные состояния, из которых можно удалить электрон, в то время как вторые (энергии сродства) являются незаполненными состояниями, на которые электрон может быть добавлен

Эти уровни энергии измеряются при регистрации энергии фотона, необходимой для ионизации электрона на данный уровень. Подобные эксперименты по фотоэмиссии очень полезны для определения занятых энергетических уровней атомов, молекул, а также твердых тел. Однако они предоставляют информацию лишь о занятых уровнях, таких как уровень  $1s$  в атоме гелия, или о валентной зоне полупроводника. Для определения положения свободных уровней, таких как уровень  $2s$  в атоме гелия или зона проводимости полупроводника, необходим эксперимент по обратной фотоэмиссии (ОФЭ) (см. рис. 3.2.2):



что позволит измерить сродство атома, определяющее способность к захвату дополнительных электронов. Для расчета уровней сродства нам необходимо рассмотреть разность между основным уровнем энергии  $E_G(N)$  и  $n$ -м уровнем  $E_n(N + 1)$  отрицательно ионизованного атома с  $(N + 1)$  электронами:

$$\varepsilon_n = E_n(N + 1) - E_G(N). \quad (3.2.2b)$$

Если мы хотим, чтобы атомные уровни определяли энергии оптических переходов, нам следует рассматривать разность между энергией основного состояния  $E_G(N)$  и  $n$ -м энергетическим уровнем  $E_n(N)$  атома с  $N$  электронами, поскольку видимый свет не меняет полного числа электронов в атоме, а просто переводит их на более высокие энергетические уровни:

$$\varepsilon_n = E_n(N) - E_G(N).$$

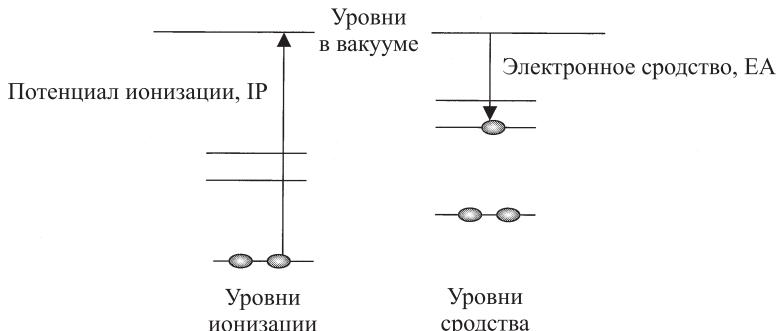


Рис. 3.2.2. Энергии ионизации и сродства зависят от потенциала отталкивания, образованного  $Z - 1$  и  $Z$  электронами соответственно, поэтому последние располагаются выше по энергии на величину  $U_0$ , равную электростатической энергии одного электрона

Не существует априорной причины того, что найденная из этих расчетов величина запрещенной зоны будет соответствовать величине запрещенной зоны, полученной из расчета энергий ионизации либо сродства. В массивных твердых телах в отсутствие экситонных эффектов обычно предполагают, что оптическая ширина запрещенной зоны равна расстоянию между валентной зоной и зоной проводимости, но этот вывод не обязан быть справедливым для небольшихnanoструктур.

**Электростатическая энергия одного электрона<sup>1</sup>.** Как отмечалось выше, непосредственный расчет уровней энергии состоит в расчете энергий  $E_G(N)$  и  $E_n(N \pm 1)$  с помощью  $N$ - и  $(N \pm 1)$ -двухэлектронного электронного уравнения Шредингера (см. уравнение (3.2.1)). Эта процедура, однако, обычно является невыполнимой, и единственным полезным на практике подходом к задачам с многоэлектронными атомами, молекулами или твердыми телами является включение эффективного потенциала  $U_{SCF}(\vec{r})$  в уравнение Шредингера, как это было описано выше.

Как следует выбирать эффективный потенциал? Если мы обозначим через  $U_{ee}(N)$  полную энергию электрон-электронного взаимодействия  $N$ -электронной системы, то правильная функция  $U_{SCF}$ , определяющая энергию ионизации, будет равна изменению энергии взаимодействия при переходе от  $N$ -к  $(N - 1)$ -электронному атому:

$$[U_{SCF}]_{\text{ионизации}} = U_{ee}(N) - U_{ee}(N - 1). \quad (3.2.3a)$$

<sup>1</sup>В оригинале название этого пункта звучит как “Single-electron charging energy”, т. е. энергия, связанная с появлением заряда одного электрона. — Прим. ред. перев.

Подобно этому правильная функция  $U_{SCF}$ , определяющая энергию сродства, равна разности энергии взаимодействия при переходе от  $N$ -к ( $N + 1$ )-электронному атому:

$$[U_{SCF}]_{\text{сродства}} = U_{ee}(N + 1) - U_{ee}(N). \quad (3.2.3b)$$

Энергия электрон-электронного взаимодействия ансамбля  $N$  электронов пропорциональна числу нетождественных пар:

$$U_{ee}(N) = U_0 N(N - 1)/2, \quad (3.2.4)$$

где  $U_0$  есть средняя энергия взаимодействия, приходящаяся на одну пару, и напоминающая электростатическую энергию одного электрона, введенную в параграфе 1.4. Из уравнений (3.2.3a), (3.2.3b) и (3.2.4), как легко видеть, следует, что

$$[U_{SCF}]_{\text{ионизации}} = U_0(N - 1), \quad \text{а} [U_{SCF}]_{\text{сродства}} = U_0 N. \quad (3.2.5)$$

Последнее означает, что для расчета энергий ионизации атома с  $Z$  электронами следует использовать потенциал, образованный ( $Z - 1$ ) электронами (для гелия это потенциал одного электрона), как мы и поступали в предыдущем параграфе. Для расчета энергий сродства следует использовать потенциал, созданный  $Z$  электронами (для гелия это будет потенциал двух электронов). Уровни энергии, полученные первым способом, будут располагаться ниже по энергии, нежели уровни, определенные вторым способом, на величину, равную электростатической энергии одного электрона  $U_0$ .

Как было отмечено в параграфе 1.5, одноэлектронная электростатическая энергия  $U_0$  зависит от степени локализации электронной волновой функции и может составлять величину в несколько электрон-вольт в атомах. Даже вnanoструктурах, имеющих масштаб порядка 10 нм, эта величина может быть достаточно существенной (к примеру, сравнимой с тепловой энергией  $k_B T$ ).

Обычно используется выражение для самосогласованного потенциала

$$U_{SCF} = \partial U_{ee} / \partial N = U_0 N - (U_0/2), \quad (3.2.6)$$

применимое для всех уровней. Поэтому энергии ионизации располагаются на ( $U_0/2$ ) ниже, а энергии сродства — на ( $U_0/2$ ) выше, чем уровни, рассчитываемые нами. Следует иметь в виду, что даже в том случае, если вычисления в рамках приближения ССП дают энергетические уровни, расстояние

между которыми мало по сравнению с величиной  $k_B T$  (см. рис. 3.2.3 а), структура может не быть хорошим проводником, поскольку одноэлектронные электростатические эффекты приводят к образованию «кулоновской щели» между занятymi и свободными уровнями (рис. 3.2.3 б). Разумеется, этот эффект является существенным лишь в том случае, когда одноэлектронная электростатическая энергия превосходит величину  $k_B T$ .

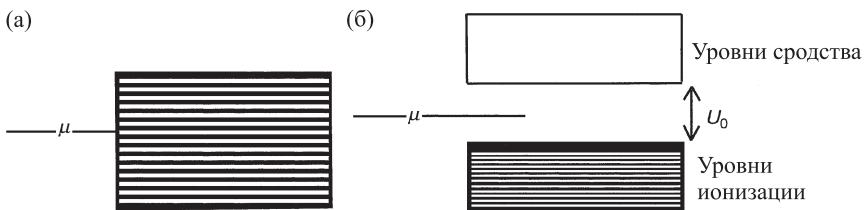


Рис. 3.2.3

**Приближение Хартри.** В массивных проводниках (с большими  $R$ ) величина  $U_0$  пренебрежимо мала, и различием между  $Z$  и  $Z - 1$  можно пренебречь. Самосогласованный потенциал для уровней ионизации и сродства фактически является одинаковым, и выражение

$$U_{SCF} = \partial U_{ee} / \partial N \quad (3.2.7)$$

можно обобщить для получения стандартного выражения, используемого в теории функционала плотности (ТФП):

$$U_{SCF}(\vec{r}) = \frac{\partial U_{ee}}{\partial n(\vec{r})}, \quad (3.2.8)$$

которое означает, что самосогласованный потенциал в точке  $\vec{r}$  равен изменению энергии электрон-электронного взаимодействия при бесконечно малом изменении числа электронов в этой же точке. Если воспользоваться стандартным выражением для  $U_{ee}$  из классической электростатики, то будем иметь

$$U_{ee} = \frac{1}{2} \int d\vec{r} \int d\vec{r}' \frac{q^2 n(\vec{r}) n(\vec{r}')}{4\pi\varepsilon |\vec{r} - \vec{r}'|}. \quad (3.2.9)$$

Уравнение (3.2.8) приводит к приближению Хартри  $U_H(\vec{r})$  для само-

согласованного потенциала:

$$U_H(\vec{r}) = \int d\vec{r}' \frac{q^2 n(\vec{r}')}{4\pi\varepsilon|\vec{r} - \vec{r}'|}, \quad (3.2.10)$$

которое является решением уравнения Пуассона  $-\nabla^2 U_H = -q^2 n/\varepsilon$  в однородной среде. Расчеты характеристик реальных структур зачастую требуют использования более сложных граничных условий, в том числе на границах, где изменяются параметры различных материалов (например, диэлектрические постоянные). В связи с этим более удобным представляется решение модифицированного уравнения Пуассона, в котором учитывается изменение в пространстве относительной диэлектрической проницаемости:

$$-\vec{\nabla} \cdot (\varepsilon \nabla U_H) = q^2 n/\varepsilon_0. \quad (3.2.11)$$

В то же время отметим, что для атомов нет той сложной неоднородности, которую надо было бы учитывать, и, следовательно, более удобной является работа с уравнением (3.2.10).

**Энергия корреляции.** Реальная энергия электрон-электронного взаимодействия оказывается меньше значения, предсказываемого уравнением (3.2.9), что связано с корреляциями электронов, точнее с их стремлением избегать друг друга. Такие корреляции должны быть включены в многоэлектронную картину, но при одноэлектронном описании она не учитывается. Один из способов ее учета состоит в представлении энергии электрон-электронного взаимодействия в виде

$$U_{ee} = \frac{1}{2} \int d\vec{r} \int d\vec{r}' \frac{e^2 n(\vec{r}) n(\vec{r}') [1 - g(\vec{r}, \vec{r}')] }{4\pi\varepsilon|\vec{r} - \vec{r}'|},$$

где  $g$  есть корреляционная функция, отражающая тот факт, что вероятность обнаружить два электрона одновременно в точках  $\vec{r}$  и  $\vec{r}'$  не является просто пропорциональной произведению  $n(\vec{r})n(\vec{r}')$ . Такая энергия имеет меньшую величину, поскольку электроны стараются «избегать» друг друга. Следует отметить, что введенный корреляционный фактор фактически зависит от спина, но пока мы не учитываем подобные детали. Соответствующее значение самосогласованного потенциала также уменьшается и определяется выражением

$$U_{\text{SCF}} = \int d\vec{r}' \frac{e^2 n(\vec{r}') [1 - g(\vec{r}, \vec{r}')] }{4\pi\varepsilon|\vec{r} - \vec{r}'|}. \quad (3.2.12)$$

Для того определения функции  $g(\vec{r}, \vec{r}')$  было выполнено большое число работ, обычно связываемых с т. н. задачей об обменно-корреляционной «дырке».

Основной эффект, связанный с энергией корреляции, заключается в добавлении *отрицательного* слагаемого  $U_{xc}(\vec{r})$  к слагаемому Хартри  $U_H(\vec{r})$ , которое обсуждалось выше (см. уравнение (3.2.10)):

$$U_{xc}(\vec{r}) = -\frac{q^2}{4\pi\varepsilon_0} C[n(\vec{r})]^{1/3}. \quad (3.2.13)$$

В этом выражении  $C$  представляет собой константу порядка единицы. Физической основой данного приближения является следующее: отдельный электрон, введенный в среду с фоновой электронной плотностью  $n(\vec{r})$ , будет отталкивать электроны в окрестности своего местонахождения, образуя положительно заряженную корреляционную «дырку» вокруг себя. Если мы будем моделировать такую дырку как положительно заряженную сферу радиуса  $r_0$ , то сможем оценить величину  $r_0$ , потребовав, чтобы полный заряд внутри сферы был равен величине заряда электрона:

$$n(r)4\pi r_0^3/3 = 1 \rightarrow r_0 = \frac{1}{C}[n(r)]^{-1/3},$$

где  $C$  есть постоянная порядка единицы. Потенциал в уравнении (3.2.13) может рассматриваться как потенциал в центре этого положительного заряда, заключенного в объеме сферы радиуса  $r_0$ :

$$U_{xc}(\vec{r}) = -\frac{q^2}{4\pi\varepsilon_0 r_0}.$$

С развитием теории ССП был выполнен большой объем исследований, в результате которых появились различные усложненные варианты записи уравнения (3.2.13). Поистине удивительно то, что одноэлектронное ССП приближение зачастую позволяет дать весьма точное описание многоэлектронных систем. То, что данное приближение очень хорошо работает, мы не можем в настоящее время объяснить математически, и наша уверенность в методе ССП проистекает из превосходного согласия с экспериментальными результатами, полученными практически для всех атомов из периодической таблицы (см. рис. 3.2.4). Почти все работы, относящиеся к теории электронной структуры атомов, молекул и твердых тел, основаны на этом методе, и мы будем его использовать в дальнейшем.

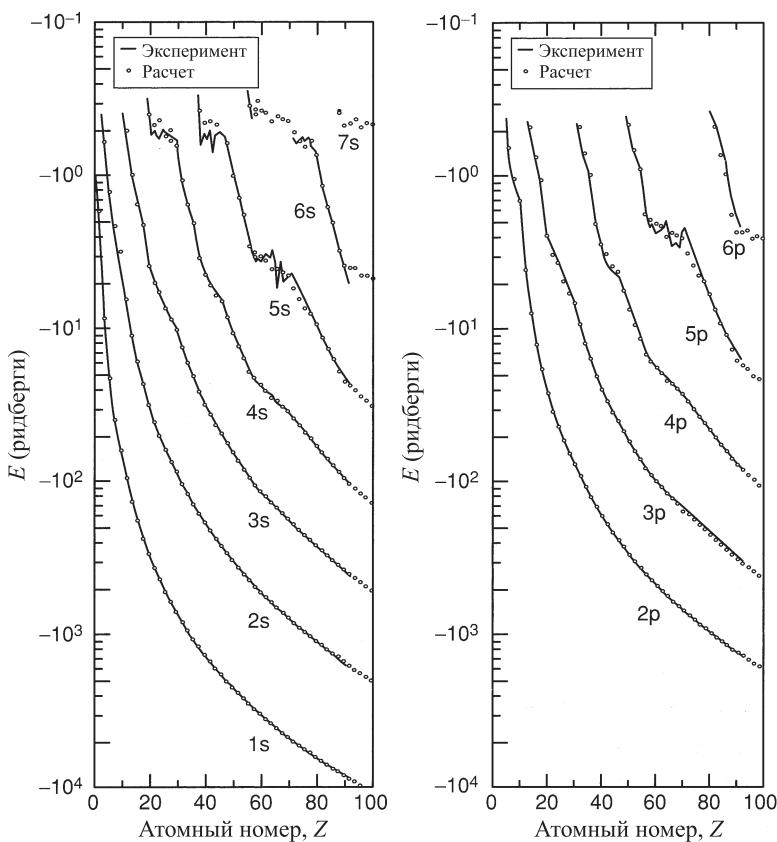


Рис. 3.2.4. Уровни энергии (как функция атомного номера), рассчитанные теоретически с использованием метода самосогласованного поля. Результаты находятся в очень хорошем согласии с экспериментом (взято из [23]). Для атома водорода  $s$ - и  $p$ -уровни являются вырожденными (т. е. имеют одну и ту же энергию). Это является следствием зависимости типа  $\sim 1/r$  для потенциала ядра. Такая зависимость, однако, не имеет места для потенциала самосогласованного поля, образованного электронами, поэтому в многоэлектронных состояниях  $s$ -уровень имеет более низкую энергию по сравнению с  $p$ -уровнем

### 3.3. Типы связи

Объяснение структуры периодической таблицы химических элементов в свое время стало одним из главных достижений квантовой механики. Для такого объяснения было необходимо принять в расчет как структуру энергетического спектра, полученную из уравнения Шредингера, так и принцип запрета Паули, согласно которому каждый уровень должен быть занят не более чем одним электроном. В разделе 3.3.1 мы обсудим общие принципы и в том числе периодический характер зависимости энергетических уровней от атомного номера. В дальнейшем мы обсудим два механизма химической связи: ионный (раздел 3.3.2) и ковалентный (раздел 3.3.3). Общей чертой любого типа связи является наличие механизма, при котором пара атомов  $A$  и  $B$  может понизить свою полную энергию путем формирования молекулы  $AB$ :  $E(AB) < E(A) + E(B)$ .

#### 3.3.1. Валентные электроны

Известно, что лишь электроны, находящиеся на внешней оболочке, т. е. валентные электроны, участвуют в процессе образования связи. Энергии этих валентных электронов изменяются периодически, как это показано на рис. 3.3.1 для первых 86 атомов периодической таблицы, от водорода (атомный номер  $Z = 1$ ) до радона ( $Z = 86$ ). Сюда не включены переходные металлы с  $d$ - и  $f$ -оболочками (см. табл. 3.3.1). Необходимо отметить, что энергии связи имеют тенденцию к уменьшению, по мере того как мы движемся в пределах одного периода периодической таблицы, например, от лития (Li) к неону (Ne). Далее они резко возрастают при переходе к новому периоду и затем уменьшаются при движении от натрия (Na) к аргону (Ar). Эта тенденция проявляется как для  $s$ - так и для  $p$ -уровней и продолжается в более высоких периодах. Можно сказать, что такое периодическое изменение энергии уровней выражает основной принцип периодической таблицы элементов.

#### 3.3.2. Ионная связь

Ионные связи обычно формируются между атомами, находящимися в левой части периодической таблицы (например, натрием Na), и атомом из правой части этой таблицы (например, хлором Cl). Энергетические уровни Na и Cl имеют приблизительно такой вид, как показанный на рис. 3.3.2. Представляется естественным, что при перекрытии волновой функции  $3s$ -состояния атома Na с волновой функцией  $3p$  атома Cl должно произойти понижение полной энергии системы. В самом

Таблица 3.3.1. Первые 86 атомов периодической таблицы, начиная с водорода (атомный номер 1) до радона ( $Z = 86$ ), за исключением переходных металлов с  $d$ - и  $f$ -оболочками ( $Z = 21–28, 39–46$  и  $57–78$ )

H ( $Z = 1$ )			B ( $Z = 5$ )	C ( $Z = 6$ )	N ( $Z = 7$ )	O ( $Z = 8$ )	F ( $Z = 9$ )	He ( $Z = 2$ )
Li ( $Z = 3$ )	Be ( $Z = 4$ )		Al ( $Z = 13$ )	Si ( $Z = 14$ )	P ( $Z = 15$ )	S ( $Z = 16$ )	Cl ( $Z = 17$ )	Ne ( $Z = 10$ )
Na ( $Z = 11$ )	Mg ( $Z = 12$ )		Zn ( $Z = 30$ )	Ga ( $Z = 31$ )	Ge ( $Z = 32$ )	As ( $Z = 33$ )	Se ( $Z = 34$ )	Ar ( $Z = 18$ )
K ( $Z = 19$ )	Ca ( $Z = 20$ )	...	Cu ( $Z = 29$ )	...	Ge ( $Z = 31$ )	As ( $Z = 32$ )	Se ( $Z = 33$ )	Kr ( $Z = 36$ )
Rb ( $Z = 37$ )	Sr ( $Z = 38$ )	...	Ag ( $Z = 47$ )	Cd ( $Z = 48$ )	In ( $Z = 49$ )	Sn ( $Z = 50$ )	Te ( $Z = 51$ )	Xe ( $Z = 54$ )
Cs ( $Z = 55$ )	Ba ( $Z = 56$ )	...	Au ( $Z = 79$ )	Hg ( $Z = 80$ )	Tl ( $Z = 81$ )	Pb ( $Z = 82$ )	Po ( $Z = 83$ )	Rn ( $Z = 86$ )

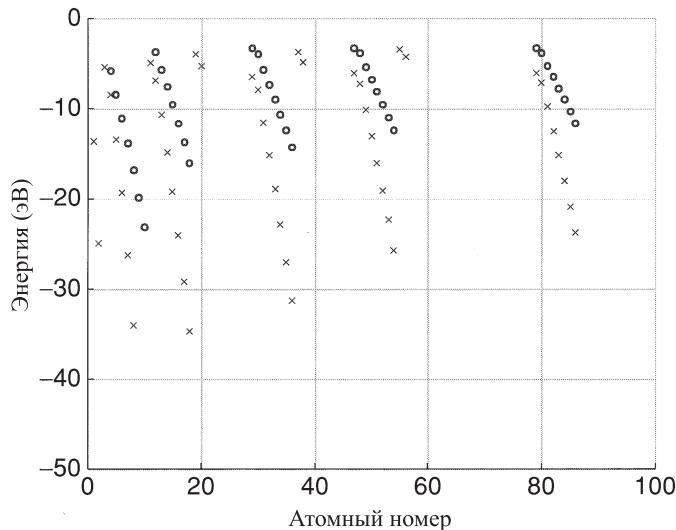


Рис. 3.3.1. Энергия внешних s( $\times$ )- и p( $\circ$ )-уровней первых 86 элементов периодической таблицы за исключением  $d$ - и  $f$ -переходных металлов ( $Z = 21\text{--}28, 39\text{--}46$ , и  $57\text{--}58$ ). Численные данные взяты из [22] и [36]

деле, представляется «очевидным», что энергия связи  $E_{\text{bin}}$  для NaCl будет равна

$$E_{\text{bin}} = E(\text{Na}) + E(\text{Cl}) - E(\text{Na}^+ \text{Cl}^-) = 12.3 - 5.1 = 7.2 \text{ эВ.}$$

Но подобный аргумент представляется не до конца убедительным, поскольку необходимо принимать во внимание и изменение электростатической энергии вследствие образования связи. Правильное значение энергии связи является близким к 4 эВ. Дело в том, что все изображенные здесь энергетические уровни являются *ионизационными* уровнями. Энергия, необходимая для образования иона натрия, определяется его потенциалом ионизации (ПИ)

$$E(\text{Na}^+) - E(\text{Na}) = \text{ПИ}(\text{Na}) = 5 \text{ эВ.} \quad (3.3.1a)$$

При этом энергия, необходимая для образования иона хлора, определяется *электронным средством* (ЭС) хлора, и потому включает дополнительную электростатическую энергию  $U_0$ :

$$E(\text{Cl}) - E(\text{Cl}^-) = \text{ЭС}(\text{Cl}) - U_0 = 13.8 \text{ эВ} - U_0. \quad (3.3.1b)$$

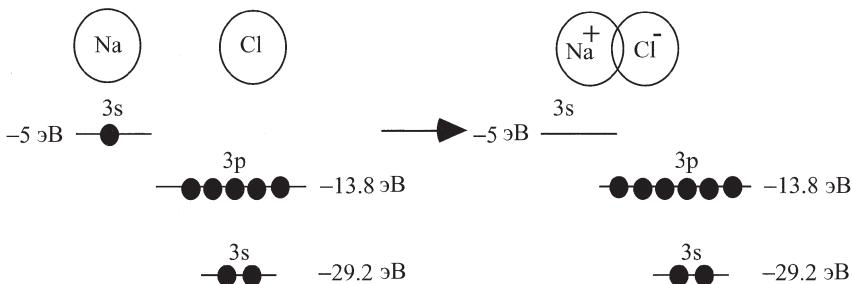


Рис. 3.3.2. Формирование молекулы  $\text{Na}^+\text{Cl}^-$  из отдельных атомов  $\text{Na}$  и  $\text{Cl}$ , когда волновая функция  $3s$ -электрона натрия перекрывается с волновой функцией  $3p$ -электрона хлора, что приводит к снижению полной энергии. Это рассмотрение носит лишь частичный характер, поскольку общий баланс энергии включает также электростатическую энергию, находящуюся в микроскопическом конденсаторе, образованном двумя ионами, как это поясняется в тексте

Комбинируя уравнения (3.3.1a) и (3.3.1b), мы получаем

$$E(\text{Na}) + E(\text{Cl}) - E(\text{Na}^+) - E(\text{Cl}^-) = 8.8 \text{ эВ} - U_0. \quad (3.3.2)$$

Однако эта величина не равна энергии связи в молекуле  $\text{NaCl}$ . Она определяет энергию, полученную при переходе от нейтрального атома натрия и нейтрального атома хлора к ионам  $\text{Na}^+$  и  $\text{Cl}^-$ , полностью пространственно разделенных друг с другом. Если мы позволим ионам  $\text{Na}^+$  и  $\text{Cl}^-$ , находящимся на бесконечно большом расстоянии друг от друга, приблизиться и сформировать молекулу хлорида натрия  $\text{Na}^+\text{Cl}^-$ , то выделившаяся в этой реакции энергия будет равна  $U'_0$ :

$$E(\text{Na}^+) + E(\text{Cl}^-) - E(\text{Na}^+\text{Cl}^-) = U'_0,$$

поэтому энергия связи дается выражением

$$E_{\text{bin}} = E(\text{Na}) + E(\text{Cl}) - E(\text{Na}^+\text{Cl}^-) = 8.8 \text{ эВ} - U_0 + U'_0, \quad (3.3.3)$$

где величина  $U_0 - U'_0$  приблизительно равна 5 эВ, что дает значение энергии связи около 4 эВ. Нас не интересуют детали вычислений (не вполне точных) в этой конкретной задаче. Моя основная идея состоит в том, что, хотя процесс образования связи путем перемещения электрона может выглядеть просто как «уход» электроном с одного атома и переход на другой атом на более низкий энергетический уровень, детальное описание баланса энергии

требует более подробного обсуждения. Можно сказать, что использование схем одноэлектронных энергетических уровней для описания электронных переходов в атомных системах требует известной осторожности.

### 3.3.3. Ковалентная связь

Мы видели, как происходит уменьшение энергии системы при соединении атома, находящегося в левой части периодической таблицы (такого как натрий) с атомом из правой части периодической таблицы (такого как хлор). Атомы из правой части периодической таблицы имеют более низко расположенные электронные уровни, и о них говорят как о более электроотрицательных атомах, по сравнению с атомами, расположенными в левой части таблицы. При сближении двух атомов можно ожидать перехода электронов с высоколежащих энергетических уровней атомов из левой части на низколежащие уровни атомов из правой части, что и приводит к образованию ионной связи.

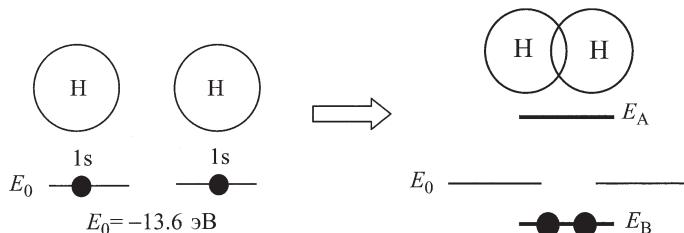


Рис. 3.3.3. Образование молекулы  $\text{H}_2$  из отдельных атомов водорода с формированием нижнего энергетического уровня  $E_B$  и верхнего энергетического уровня  $E_A$

Это рассуждение, однако, не объясняет образования ковалентных связей, в которых участвуют атомы с приблизительно одинаковой электроотрицательностью. Подобный процесс можно считать более тонким. К примеру, трудно сразу увидеть, почему два одинаковых атома водорода стремятся образовать молекулу  $\text{H}_2$ , когда перенос электрона от одного атома к другому не приводит к понижению энергии. На самом деле происходит следующее: когда два атома сближаются, получающиеся энергетические уровни представляют собой исходный уровень, расщепленный на нижний ( $E_B$ ) и верхний ( $E_A$ ) уровни, как показано на рис. 3.3.3. В такой ситуации оба электрона занимают нижний уровень, энергия которого меньше, чем у изолированного атома водорода:  $E_B < E_0$ .

Каким образом мы можем вычислить величину  $E_B$ ? Для этого необходимо решить уравнение Шредингера

$$E_\alpha \Phi_\alpha(\vec{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_N(\vec{r}) + U_{N'}(\vec{r}) + U_{SCF}(\vec{r}) \right) \Phi_\alpha(\vec{r}), \quad (3.3.4)$$

где  $U_N(\vec{r})$  и  $U_{N'}(\vec{r})$  есть потенциалы левого и правого ядер соответственно, а  $U_{SCF}(\vec{r})$  представляет собой потенциал, который действует на один электрон благодаря присутствию другого электрона. Для простоты мы пренебрегаем сейчас потенциалом  $U_{SCF}(\vec{r})$  и рассчитываем энергетические уровни, обусловленные лишь наличием потенциала ядер:

$$E_{\alpha 0} \Phi_{\alpha 0}(\vec{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_N(\vec{r}) + U_{N'}(\vec{r}) \right) \Phi_{\alpha 0}(\vec{r}). \quad (3.3.5)$$

Решение уравнения (3.3.5) с наименьшей энергией может быть приближенно записано как

$$E_{B0} = E_0 + \frac{a + b}{1 + s}, \quad (3.3.6)$$

где (см. параграф 4.1)

$$a = -2E_0 \frac{1 - (1 + \bar{R})e^{-2\bar{R}}}{\bar{R}},$$

$$b = -2E_0(1 + \bar{R})e^{-\bar{R}},$$

$$s = e^{-\bar{R}}[1 + \bar{R} + (\bar{R}^2/3)],$$

$$\bar{R} \equiv R/a_0,$$

а величина  $R$  есть расстояние между центрами атомов водорода.

Сейчас мы попробуем разобраться в природе сил, приводящих к образованию ковалентной связи. Пунктирная линия на рис. 3.3.4 показывает зависимость величины  $E_{B0} - E_0$  от длины связи  $R$ , описываемую уравнением (3.3.6). Экспериментально значение длины связи  $R$  в молекуле водорода равно 0.074 нм, что означает наличие минимума у полной энергии при данном значении  $R$ . Поскольку энергия продолжает уменьшаться по мере уменьшения  $R$ , может возникнуть предположение о том, что два атома водорода просто располагаются в одной точке пространства ( $R = 0$ ). Для ответа на этот вопрос следует рассчитать полную энергию, которая включала бы межэлектронное отталкивание (обратите внимание на то, что функция

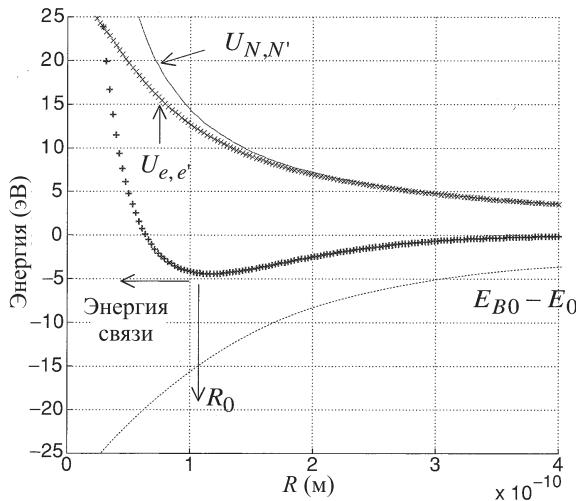


Рис. 3.3.4. Различные энергии как функции расстояния между ядрами  $R$ . Значками  $\times \times \times$  показан приближенный вид энергии межэлектронного отталкивания ( $U_{e,e'}$ ). Сплошной линией показана энергия отталкивания ядер ( $U_{N,N'}$ ). Пунктирная кривая изображает величину  $E_{B0} - E_0$  — энергию уровня связи молекулы  $H_2$ , отсчитанную от уровня  $1s$  атома водорода, рассчитанную приближенно с помощью уравнения Шредингера без какого-либо самосогласованного потенциала. Значками  $+++$  показана энергия связи молекулы  $H_2$  относительно энергии двух атомов водорода, оценка которой проведена из выражения  $2(E_{B0} - E_0) + U_{N,N'} + U_{e,e'}$

$U_{SCF}(r)$  не была включена в уравнение (3.3.6)), а также межъядерное отталкивание. Общий баланс энергии найдем как разность между энергией молекулы водорода ( $H_2$ ) и двумя изолированными атомами водорода ( $2H$ ).

Энергия, необходимая для сборки двух атомов водорода из двух протонов ( $N, N'$ ) и двух электронов ( $e, e'$ ) может быть записана в виде

$$E(2H) = U_{e,N} + U_{e',N'} = 2E_0. \quad (3.3.7a)$$

Энергия, необходимая для сборки молекулы  $H_2$  из двух протонов ( $N, N'$ ) и двух электронов ( $e, e'$ , может быть записана в форме

$$E(H_2) = U_{N,N'} + U_{e,e'} + U_{e,N} + U_{e,N'} + U_{e',N} + U_{e',N'} = 2E_0. \quad (3.3.7b)$$

Уравнение (3.3.6) определяет квантовомеханическое значение как величины ( $U_{e,N} + U_{e,N'}$ ), так и величины ( $U_{e',N} + U_{e',N'}$ ), а также  $E_{B0}$ .

Следовательно,

$$E(\text{H}_2) = U_{N,N'} + U_{e,e'} + 2E_{B0}. \quad (3.3.7c)$$

Энергия связи представляет собой энергию, необходимую для диссоциации молекулы, состоящей из двух атомов водорода, которая может быть записана в виде

$$E_{\text{bin}} = E(\text{H}_2) - E(2\text{H}) = 2(E_{B0} - E_0) + U_{N,N'} + U_{e,e'}. \quad (3.3.8)$$

Данная энергия должна принимать минимальное значение в состоянии равновесия, и она состоит из трех отдельных слагаемых. Уравнение (3.3.6) определяет лишь первое слагаемое, а второе может быть легко найдено, поскольку оно равно энергии электростатического взаимодействия между двумя ядрами, т. е. точечными зарядами:

$$U_{N,N'} = q^2/4\pi\varepsilon_0 R. \quad (3.3.9a)$$

Электростатическое взаимодействие между электронами также должно описываться функцией вида  $q^2/4\pi\varepsilon_0 R$  для больших значений  $R$ , но при этом оно должно стремиться к значению  $\sim q^2/4\pi\varepsilon_0 a_0$  на малых расстояниях, поскольку электронные заряды «размазаны» на масштабе  $\sim a_0$ . Примем для этого вида взаимодействия следующее приближение

$$U_{e,e'} \cong q^2/4\pi\varepsilon_0 \sqrt{R^2 + a_0^2}, \quad (3.3.9b)$$

отметив, что это есть лишь чрезвычайно упрощенная аппроксимация решения очень сложной квантовомеханической задачи. Подчеркнем еще раз, что теория электрон-электронного взаимодействия представляет собой одну из центральных проблем квантовой теории состояния вещества.

Сплошными линиями на рис. 3.3.4 показана зависимость потенциала  $U_{N,N'}$  (уравнение (3.3.9a)), в то время как линия из значков  $\times \times \times$  демонстрирует зависимость потенциала  $U_{e,e'}$  (уравнение (3.3.9b)). Линия из значков  $+++$  изображает полную энергию химической связи, оцениваемую из выражения (3.3.8). Она достигает минимума приблизительно при 0.1 нм, что не слишком сильно отличается от экспериментально определенной длины связи 0.074 нм. Кроме того, значение энергии связи в этом минимуме равно  $\sim 4.5$  эВ, что весьма близко к фактическому экспериментальному значению. Несмотря на грубость использованных приближений, этот пример весьма наглядно иллюстрирует природу химической связи.

**Частота молекулярных колебаний.** Форма кривой, описывающей зависимость энергии связи от  $R$ , позволяет рассматривать молекулу водорода как две массы, соединенные пружиной (рис. 3.3.5).



Рис. 3.3.5. Молекула водорода может рассматриваться как две массы, соединенные пружиной

Сжатая идеальная пружина с жесткостью  $K$  имеет потенциальную энергию вида  $U(R) = K(R - R_0)^2/2$ . Энергия связи молекулы водорода (см. рис. 3.3.4) может быть аппроксимирована функцией  $U(R) \cong U(R_0) + + K(R - R_0)^2/2$ , где эффективная жесткость пружины может быть определена как кривизна  $[d^2U/dR^2]_{R=R_0}$ . В самом деле, частоту колебаний структуры H – H можно оценить как резонансную частоту  $\sqrt{2K/M}$  системы, состоящей из масс, равных массе атома водорода, соединенных пружиной.

**Уровни ионизации.** Как уже было отмечено, энергетические уровни многоэлектронной системы обычно определяют уровни ионизации, представляющие собой энергию, которую необходимо затратить, чтобы удалить электрон из данной системы. В данном контексте это означает, что уровень энергии молекулы водорода  $E_B$  можно представлять как величину

$$E_B = E(\text{H}_2) - E(\text{H}_2^+).$$

Поскольку справедливо равенство  $E(\text{H}_2^+) = U_{N,N'} + U_{e',N} + U_{e',N'}$ , то, воспользовавшись уравнением (3.3.7b), мы можем записать

$$E_B = U_{e,e'} + U_{e,N} + U_{e,N'} = U_{e,e'} + E_{B0}. \quad (3.3.10)$$

Нетрудно проверить, что в наших модельных расчетах (см. рис. 3.3.4)  $E_{B0}$  имеет величину приблизительно на 15 эВ ниже  $E_0$ , но при этом значение  $E_B$  лежит лишь примерно на 4 эВ ниже  $E_0$ . Если бы в уравнение Шредингера было включено самосогласованное поле  $U_{SCF}(r)$ , мы получили бы энергию, лежащую выше (не такое большое по модулю отрицательное число) энергии невзаимодействующих электронов  $E_{B0}$  на величину, равную энергии электрон-электронного взаимодействия  $U_{e,e'}$ .

**Энергия связи.** Можно было бы подумать, что значение энергии связи дается выражением

$$E_{\text{bin}} = 2(E_B - E_0) + U_{N,N'},$$

поскольку  $E_B$  включает энергию электрон-электронного взаимодействия  $U_{e,e'}$ . Однако легко увидеть из уравнений (3.3.8) и (3.3.10), что правильным является выражение

$$E_{\text{bin}} = 2(E_B - E_0) + (U_{N,N'} - U_{e,e'}).$$

Дело в том, что если мы включим электрон-электронное взаимодействие в наши расчеты энергии  $E_B$ , то полная энергия двух электронов НЕ будет равна  $2E_B$ , поскольку при этом энергия взаимодействия между двумя электронами будет учтена дважды. Правильное значение энергии получается при вычитании этой дважды учтенной части:  $2E_B - U_{e,e'}$ .

### 3.4. Дополнение: многоэлектронная картина

Как я упоминал в параграфе 3.2, метод ССП используется столь широко лишь потому, что более точные способы решения, основанные на многоэлектронной картине, представляются слишком сложными. Однако многоэлектронную задачу можно решить точно, если мы имеем дело с небольшим каналом, слабо связанным с окружающей средой и подобным одноравневой системе, которая обсуждалась в параграфе 1.4. В связи с этим будет полезно рассмотреть заново одноуровневую задачу в многоэлектронной постановке и сравнить результаты с теми, которые получаются при использовании метода самосогласованного поля.

**Сопоставление картины одноэлектронных и многоэлектронных уровней.** Схема одноэлектронных и многоэлектронных энергетических уровней в системе, содержащей только один вырожденный по спину уровень с энергией  $\epsilon$ , приведена на рис. 3.4.1. Поскольку каждый одноэлектронный энергетический уровень может быть либо свободным (0), либо занятым (1), многоэлектронные состояния могут нумероваться бинарными числами, где число разрядов равно числу одночастичных состояний. Следовательно,  $N$  одноэлектронных состояний приводят к появлению  $2^N$  многоэлектронных состояний. Это число быстро возрастает с  $N$ , что невероятно усложняет расчеты. Вот почему методы ССП используются так широко, несмотря на то, что они имеют приближенный характер.

Рассмотрим систему с двумя вырожденными одноэлектронными состояниями (со спином вверх и вниз), которые могут быть заполненными либо свободными. Предполагается, что все остальные одноэлектронные состояния не изменяют степени своего заполнения: те, что ниже, остаются заполненными, а те, что выше, — свободными. Предположим, что энергия

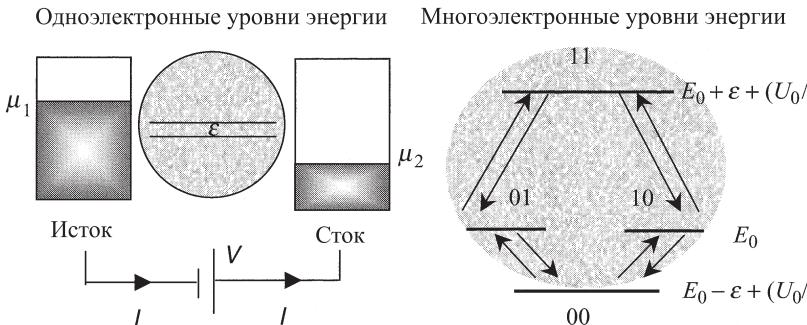


Рис. 3.4.1. Сопоставление одноэлектронных и многоэлектронных уровней в канале с одним вырожденным по спину уровнем с энергией  $\varepsilon$

электрон-электронного взаимодействия дается выражением

$$U_{ee}(N) = (U_0/2)N(N-1) \quad (\text{подобно уравнению (3.2.4)}),$$

что отвечает самосогласованному потенциалу (см. уравнение (3.2.6))

$$\partial U_{ee} / \partial N = U_0 N - (U_0/2).$$

Предположим, что число электронов  $N_0$  в нейтральном состоянии соответствует тому, что одно из этих состояний заполнено. Энергия одноэлектронного уровня энергии  $\varepsilon$  может быть записана как сумма «голых» уровней  $\tilde{\varepsilon}$  (полученных из уравнения Шредингера с одним лишь ядерным потенциалом  $U_N$ ) плюс вклад от самосогласованного потенциала  $[\partial U_{ee} / \partial N]_{N=N_0}$ :

$$\varepsilon = \tilde{\varepsilon} + [\partial U_{ee} / \partial N]_{N=N_0} = \tilde{\varepsilon} + U_0 N_0 - (U_0/2).$$

Рассмотрим теперь многоэлектронную картину. У нас имеется четыре многоэлектронных состояния, которые мы обозначим как 00, 01, 10 и 11. В нейтральном состоянии система находится либо в состоянии (10), либо в состоянии (01), чью полную энергию мы обозначим как

$$E(10) = E(01) \equiv E_0.$$

Мы можем записать энергию других многоэлектронных состояний как

$$\begin{aligned} E(11) &= E_0 + \tilde{\varepsilon} + U_{ee}(N_0 + 1) - U_{ee}(N_0) \\ &= E_0 + \tilde{\varepsilon} + U_0 N_0 = E_0 + \varepsilon + (U_0/2) \end{aligned}$$

и

$$\begin{aligned} E(00) &= E_0 - \tilde{\varepsilon} - U_{ee}(N_0) + U_{ee}(N_0 - 1) \\ &= E_0 - \tilde{\varepsilon} - U_0(N_0 - 1) = E_0 - \varepsilon + (U_0/2). \end{aligned}$$

**Уравнение эволюции.** В многоэлектронной задаче полная система описывается набором вероятностей  $P_\alpha$  найти ее в одном из  $2^N$  возможных состояний  $\alpha$ . Сумма всех таких вероятности должна равняться единице:

$$\sum_{\alpha} P_{\alpha} = 1 \rightarrow P_{00} + P_{01} + P_{10} + P_{11} = 1. \quad (3.4.1)$$

Мы можем рассчитать индивидуальные вероятности, полагая, что система непрерывно переходит между состояниями этого набора, и в стационарных условиях полного потока вероятности, увеличивающего или уменьшающего полную вероятность заполнения каждого состояния, нет:

$$\sum_{\beta} R(\alpha \rightarrow \beta)P_{\alpha} = \sum_{\beta} R(\beta \rightarrow \alpha)P_{\beta}. \quad (3.4.2)$$

Зная величину скоростей эволюции, мы можем рассчитать вероятности различных состояний, решая уравнение (3.4.2). Подобные уравнения называют уравнениями эволюции, и, следовательно, мы можем назвать уравнение (3.4.2) многоэлектронным уравнением эволюции.

В свою очередь, константы скоростей эволюции  $R(\alpha \rightarrow \beta)$  могут быть записаны с учетом конкретной модели взаимодействия системы с ее окружением. К примеру, если предположить, что взаимодействие включает лишь приход и уход отдельных электронов из контактов источника и стока, то константы скорости эволюции для состояний 00 и 01 определяются выражением:

$$\frac{\gamma_1}{\hbar}f'_1 + \frac{\gamma_2}{\hbar}f'_2 \uparrow \left[ \begin{pmatrix} 01 \\ 00 & E_0 \\ 00 & E_0 - \varepsilon + (U_0/2) \end{pmatrix} \right] \downarrow \frac{\gamma_1}{\hbar}(1 - f'_1) + \frac{\gamma_2}{\hbar}(1 - f'_2),$$

где функции

$$f'_1 \equiv f_0(\varepsilon_1 - \mu_1) \quad \text{и} \quad f'_2 \equiv f_0(\varepsilon_1 - \mu_2)$$

определяют число электронов с энергией  $\varepsilon_1 = \varepsilon - (U_0/2)$  в контактах истока и стока соответственно. Скорость попадания в данное состояние пропорциональна числу имеющихся электронов, в то время как скорость ухода из этого состояния пропорциональна числу имеющихся пустых состояний.

Такая же картина имеет место при описании переходов между состояниями 00 и 10, где предполагается, что состояния со спином вверх и вниз описываются одной и той же функцией Ферми в контактах, чего и следовало ожидать, коль скоро каждый контакт локально находится в состоянии равновесия.

Аналогично, мы можем записать константы скорости переноса между состояниями 01 и 11:

$$\frac{\gamma_1}{\hbar} f_1'' + \frac{\gamma_2}{\hbar} f_2'' \uparrow \left[ \begin{pmatrix} \overline{11} & E_0 + \varepsilon + (U_0/2) \\ \overline{01} & E_0 \end{pmatrix} \right] \downarrow \frac{\gamma_1}{\hbar} (1 - f_1'') + \frac{\gamma_2}{\hbar} (1 - f_2''),$$

где функции

$$f_1'' \equiv f_0(\varepsilon_2 - \mu_1) \quad \text{и} \quad f_2'' \equiv f_0(\varepsilon_2 - \mu_2)$$

описывают число электронов с энергией  $\varepsilon_2 = \varepsilon + (U_0/2)$  в контактах истока и стока, соответствующей разности энергии между состояниями 01 и 11. Эта энергия больше, чем энергия разности  $\varepsilon$  между состояниями 00 и 01, поскольку в силу наличия энергии взаимодействия  $U_0$  требуется больше энергии для добавления электрона туда, где уже есть один электрон.

Используя эти константы скоростей эволюции, из уравнения (3.4.2) можно непосредственно показать, что

$$\frac{P_{10}}{P_{00}} = \frac{P_{01}}{P_{00}} = \frac{\gamma_1 f_1' + \gamma_2 f_2'}{\gamma_1 (1 - f_1') + \gamma_2 (1 - f_2')} \quad (3.4.3a)$$

и

$$\frac{P_{11}}{P_{10}} = \frac{P_{11}}{P_{01}} = \frac{\gamma_1 f_1'' + \gamma_2 f_2''}{\gamma_1 (1 - f_1'') + \gamma_2 (1 - f_2'')}. \quad (3.4.3b)$$

Совместно с уравнением (3.4.1) эти выражения определяют все индивидуальные вероятности. На рисунке 3.4.2 показана эволюция этих вероятностей по мере возрастания напряжения  $V_G$  на затворе и постоянном нулевом напряжении стока  $V_D$ . Напряжение на затворе смешает значение энергии одноэлектронного уровня как  $\varepsilon \rightarrow \varepsilon + U_L$  (мы предположили, что  $U_L = -qV_G$ ), а вероятности рассчитываются из уравнений (3.4.3a), (3.4.3b) и (3.4.1), где функции Ферми даются выражениями

$$f_1' = f_0(\varepsilon_1 + U_L - \mu_1), \quad f_2' = f_0(\varepsilon_1 + U_L - \mu_2); \quad (3.4.4a)$$

$$f_1'' = f_0(\varepsilon_2 + U_L - \mu_1), \quad f_2'' = f_0(\varepsilon_2 + U_L - \mu_2). \quad (3.4.4b)$$

Система эволюционирует из состояния 00 (при  $P_{00} = 1$ ), переходит в состояния 01 и 10 ( $P_{01} = P_{10} = 0.5$ ), по мере того как  $\varepsilon_1 + U_L$  становится

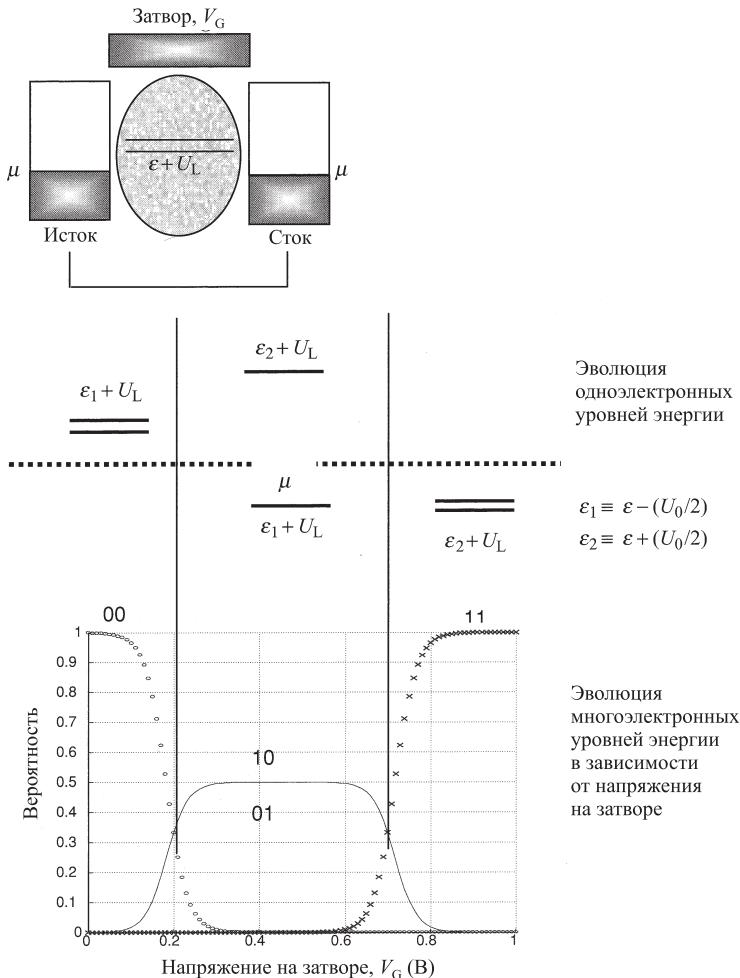


Рис. 3.4.2. Эволюция уровней энергии в канале с одним вырожденным по спину уровнем при возрастании положительного напряжения  $V_G$  на затворе и постоянном нулевом напряжении стока  $V_D$ . Здесь  $\mu = 0$ ,  $\varepsilon = 0.2$  эВ,  $k_B T = 0.025$  эВ,  $U_0 = 0.25$  эВ,  $U_L = -qV_G$ . В нижней части рисунка показаны вероятности нахождения канала в одном из четырех состояний:  $P_{00}(\circ)$ ,  $P_{01} = P_{10}$  (сплошная линия), и  $P_{11}(x)$

меньше  $\mu$ , и, наконец, переходит в состояние 11 ( $P_{11} = 1$ ), когда величина  $\varepsilon_1 + U_L$  становится меньше  $\mu$ .

**Взаимосвязь между многоэлектронной картиной и одноэлектронными уровнями.** В параграфе 3.2 я обращал ваше внимание на то, что одноэлектронные энергетические уровни определяются как *разность* между энергетическими уровнями многоэлектронной задачи, соответствующие состояниям, отличающимся на *один электрон*. Переходы, сопровождающие добавлением одного электрона, соответствуют уровням *средства*, в то время как переходы с удалением одного электрона соответствуют уровням *ионизации*. К примеру (см. рис. 3.4.2), если система находится в состоянии 00, то имеются два вырожденных одноэлектронных уровня  $\varepsilon_1 + U_L$  вида

$$\varepsilon_1 + U_L = E(10) - E(00) = E(01) - E(00) \quad \text{Уровни средства.}$$

Если система находится в состоянии 10, то имеются два одноэлектронных уровня

$$\varepsilon_1 + U_L = E(10) - E(00) \quad \text{Уровень ионизации}$$

$$\text{и } \varepsilon_2 + U_L = E(11) - E(10) \quad \text{Уровень средства.}$$

В состоянии 11 присутствуют два вырожденных одноэлектронных уровня

$$\varepsilon_2 + U_L = E(11) - E(10) = E(11) - E(01) \quad \text{Уровни ионизации.}$$

Уровни средства располагаются выше электрохимического потенциала  $\mu$ , в то время как уровни ионизации находятся под  $\mu$ , как это показано на рис. 3.4.2. В этом смысле чрезвычайно важной концепции, которая используется при интерпретации одноэлектронных энергетических уровней в системе взаимодействующих объектов. Занятые уровни (или уровни ионизации) ответственны за удаление электрона, в то время как свободные уровни (уровни средства) участвуют в процессе добавления электрона в систему. Действительно, именно таким образом все эти уровни определяются экспериментально, причем занятые уровни определяются при помощи фотоэмиссии (ФЭ), а свободные уровни при помощи обратной фотоэмиссии (ОФЭ), как было упомянуто в параграфе 1.1.

**Закон равновесия.** Рисунок 3.4.2 иллюстрирует расчет, выполненный в равновесной системе, когда исток и сток имеют одну и ту же энергию Ферми  $f_1 = f_2$ . Задачи для систем, находящихся в равновесии, фактически не требуют использования уравнения эволюции, подобного уравнению (3.4.2). Достаточно воспользоваться общим принципом равновесной

статистической механики, который утверждает, что вероятность  $P_\alpha$  нахождения системы в многоэлектронном состоянии  $\alpha$  с энергией  $E_\alpha$  и числом электронов  $N_\alpha$  дается выражением

$$P_\alpha = \frac{1}{Z} \exp[-(E_\alpha - \mu N_\alpha)/k_B T], \quad (3.4.5)$$

где постоянная  $Z$  (называемая статистической суммой) определяется из условия, согласно которому вероятности, задаваемые формулой (3.4.5), в сумме дают единицу:

$$Z = \sum_{\alpha} \exp[-(E_\alpha - \mu N_\alpha)/k_B T]. \quad (3.4.6)$$

Данный закон является основным законом равновесной статистической механики. Он применим к любой системе частиц (электронов, фотонов, атомов и т. д.), взаимодействующих или нет (см., к примеру, главу 1 из книги [18]). Функция Ферми представляет собой лишь частный пример этого общего соотношения, которая может быть получена при применении (3.4.6) к системе с одним одноэлектронным энергетическим уровнем, что будет отвечать двум многоэлектронным состояниям:

$$\begin{array}{cccc} \alpha & N_\alpha & E_\alpha & P_\alpha \\ 0 & 0 & 0 & 1/Z \\ 1 & 1 & \varepsilon & 1/Z \exp[(\nu - \varepsilon)/k_B T], \end{array}$$

так что  $Z = 1 + \exp[(\nu - \varepsilon)/k_B T]$ , и можно непосредственно показать, что среднее значение числа электронов равно значению функции Ферми (уравнение (1.1.1)):

$$N = \sum_{\alpha} N_\alpha P_\alpha = P_1 = \frac{\exp[(\nu - \varepsilon)/k_B T]}{1 + \exp[(\nu - \varepsilon)/k_B T]} = f_0[\varepsilon - \nu].$$

При рассмотрении многоэлектронных систем мы можем использовать функцию Ферми лишь в том случае, когда электроны не взаимодействуют. В такой ситуации представляется оправданным выделение одного уровня, который может заселяться независимо от степени заполнения оставшихся уровней. Метод ССП использует функцию Ферми в предположении, что энергия каждого уровня зависит от заполнения остальных уровней. Это, однако, является лишь приближением. При более точном подходе следовало бы отказаться от использования функции Ферми, и вместо того, чтобы

вычислять вероятности различных многочастичных состояний, применить уравнение (3.4.5).

Можно привести пример, иллюстрирующий сказанное. Вероятность заполнения локализованных уровней донора или акцептора (с большими электростатическими энергиями  $U_0$ ) в полупроводниках в состоянии равновесия определяется модифицированной функцией Ферми (в которой  $\nu$  есть степень вырождения уровня)

$$f = \frac{1}{1 + (1/\nu) \exp[(\varepsilon - \mu)/k_B T]}, \quad (3.4.7)$$

а не стандартной функцией Ферми (ср. с уравнением (1.1.1)). Это выражение может быть легко получено для двух вырожденных по спину уровней ( $\nu = 2$ ), если мы предположим, что электростатическая энергия  $U_0$  настолько велика, что состояние 11 имеет нулевую вероятность заполнения. Для оставшихся состояний мы можем записать

$$\begin{array}{cccc} \alpha & N_\alpha & E_\alpha & P_\alpha \\ 00 & 0 & 0 & 1/Z \\ 01 & 1 & \varepsilon & 1/Z \exp[(\nu - \varepsilon)/k_B T] \\ 10 & 1 & \varepsilon & 1/Z \exp[(\nu - \varepsilon)/k_B T], \end{array}$$

так что  $Z = 1 + 2 \exp[(\nu - \varepsilon)/k_B T]$ , и среднее значение числа электроновдается выражением

$$\begin{aligned} N &= \sum_{\alpha} N_{\alpha} P_{\alpha} = P_{01} + P_{10} = \\ &= \frac{2 \exp[(\nu - \varepsilon)/k_B T]}{1 + 2 \exp[(\nu - \varepsilon)/k_B T]} = \frac{1}{1 + 2 \exp[(\nu - \varepsilon)/k_B T]}, \end{aligned}$$

в соответствии с уравнением (3.4.7). Этот результат, известный каждому инженеру, работающему сnanoструктурами, может рассматриваться как частный случай общего результата в уравнении (3.4.5).

Уравнение (3.4.5), однако, может быть использовано лишь для изучения равновесных задач, в то время как наша основная задача — это вычисление тока внеравновесных условиях. Именно поэтому мы обратили внимание на подход, связанный с использованием уравнения эволюции и основанный на уравнении (3.4.2). Для равновесных задач он приводит к тем же результатам, однако помогает также развить новую важную концепцию. Часто приходится слышать утверждения о том, что законы, которые описывают равновесные состояния, носят статистический характер и могут применяться лишь к большим системам. Из уравнения эволюции, однако, очевидно, что закон равновесия (уравнение (3.4.5)) не является исключительным

атрибутом данной системы. Этот закон определяет свойства контактов или «резервуаров». Наше единственное предположение относится к распределению по энергии электронов, которые приходят из контактов. Коль скоро данные «резервуары» являются простыми системами, не имеет значения, насколько сложной или насколько малой является данная «система».

**Вычисление тока.** Решив уравнения эволюции для индивидуальных вероятностей, мы можем возвратиться к неравновесным задачам и рассчитать ток истока:

$$I_1 = -q \sum_{\beta} (\pm) R_1(\alpha \rightarrow \beta) P_{\alpha},$$

$+$ , если в  $\beta$  больше электронов, чем в  $\alpha$   
 $-$ , если в  $\beta$  меньше электронов, чем в  $\alpha$

где величина  $R_1$  описывает часть полного коэффициента перехода  $R$ , связанную с контактом истока. В нашей задаче это сводится к вычислению выражения

$$I_1 = (-q/\hbar)(2\gamma_1 f'_1 P_{00} - \gamma_1(1 - f'_1)(P_{01} + P_{10}) + \gamma_1 f''_1(P_{01} + P_{10}) - 2\gamma_1(1 - f''_1)P_{11}). \quad (3.4.8)$$

На рис. 3.4.3 показана вольт-амперная характеристика, т. е. зависимость тока от напряжения стока ( $I - V_D$ ), рассчитанная в рамках описанного выше подхода. Результат сравнивается с расчетом, основанным на методе ССП, описанном в параграфе 1.4. Вольт-амперные характеристики, полученные методом ССП, отличаются от показанных на рис. 1.4.6 а, поскольку в выражении для самосогласованного потенциала  $U_0(N - N_0)$  имеем значение  $N_0 = 1$ , а не нуль, а также учтено наличие двух спинов. Два упомянутых подхода хорошо согласуются для  $U_0 = 0.025$  эВ, но заметно различаются для  $U_0 = 0.25$  эВ, что демонстрирует влияние кулоновской блокады, связанной с изменением электростатической энергии при добавлении одного электрона (см. упражнение У3.6).

Использование многоэлектронного уравнения эволюции позволяет рассчитать ток в режиме кулоновской блокады, когда одноэлектронная электростатическая энергия  $U_0$  намного превосходит уширение уровней  $\gamma_{1,2}$  и/или тепловую энергию  $k_B T$ . К сожалению, мы не можем использовать этот метод широко в силу двух причин. Во-первых, трудности возрастают экспоненциально с увеличением размеров и становятся непреодолимыми. Во-вторых, не ясно, каким образом можно ввести уширение уровней в данную картину, или учесть квантовую интерференцию между различными энергетическими уровнями. Как следствие, в настоящее время все еще является непростой задачей корректное описание транспорта в промежуточном режиме, когда  $U_0 \sim \gamma_{1,2}, k_B T$ . В этом режиме движение электронов

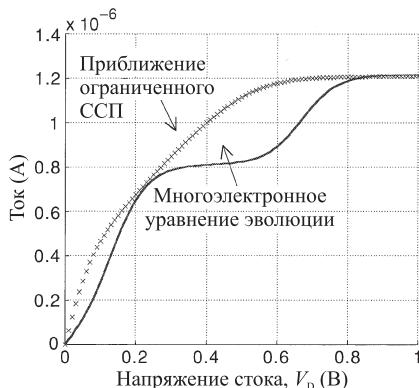
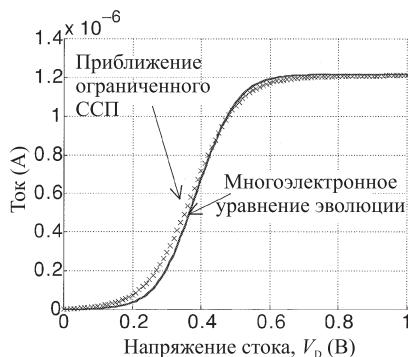
(a)  $U_0 = 0.25 \text{ эВ}$ (б)  $U_0 = 0.025 \text{ эВ}$ 

Рис. 3.4.3. Зависимость тока от напряжения стока  $V_D$ , рассчитанная в предположении  $V_G = 0$  с  $\mu = 0$ ,  $\epsilon = 0.2 \text{ эВ}$ ,  $k_B T = 0.025 \text{ эВ}$ ,  $\gamma_1 = \gamma_2 = 0.005 \text{ эВ}$ ,  $U_L = -qV_D/2$ . Два различных подхода (метод ССП и многоэлектронное уравнение эволюции) хорошо согласуются для  $U_0 = 0.025 \text{ эВ}$ , но приводят к заметно различающимся результатам при  $U_0 = 0.25 \text{ эВ}$ , что свидетельствует о существовании кулоновской блокады, связанной с изменением электростатической энергии при добавлении одного электрона

является «сильно коррелированным», что приводит к заметному отличию двухэлектронных вероятностей вида  $P_{11}$  от произведения одноэлектронных вероятностей вида  $P_{01}P_{10}$ . Было предпринято большое число попыток учесть электронные корреляции в рамках приближения ССП. К примеру, можно предположить, что самосогласованный потенциал  $U_{SCF} = U_0\Delta N$ , который мы использовали, является одинаковым для всех уровней энергии, или орбиталей. Можно было бы использовать «неограниченное» приближение для самосогласованного поля, зависящее от орбиталей так, что потенциал, ощущаемый на уровне  $j$ , исключал бы всякое самовоздействие, обусловленное числом электронов  $n_j$  на этом уровне (см. упражнение У.3.5.(в)):

$$U_{SCF}(j) = U_0(\Delta N - \Delta n_j). \quad (3.4.9)$$

Такие подходы могут приводить к лучшему согласию с результатами многоэлектронного описания, но всякий раз к ним следует относиться с осторожностью, особенно при решении неравновесных задач.

## Упражнения

**У.3.1.** Используйте метод ССП (только слагаемое Хартри) для расчета энергии уровня  $1s$  атома гелия. (а) Постройте график для потенциала ядра  $U_N(r)$  и для самосогласованного электронного потенциала  $U_{SCF}(r)$  (ср. с рис. 3.1.4 *a*). (б) Постройте график волновой функции уровня  $1s$  атома гелия и сравните его с графиком волновой функции уровня  $1s$  в атоме водорода (ср. с рис. 3.1.4 *b*).

**У.3.2.** Используйте метод ССП (только слагаемое Хартри) для расчета энергии уровней  $3s$  и  $3p$  атома кремния. Постройте графики волновых функций состояний  $1s$  и  $3p$  в атоме кремния и сравните их с графиком волновой функции для состояния  $1s$  атома водорода (ср. с рис. 3.1.4 *b*).

**У.3.3.** Постройте приближенно график энергии связи в молекуле водорода как функцию расстояния между атомами водорода. Используя уравнения (3.3.6) и (3.3.9*a*), (3.3.9*b*), сравните результат с рис. 3.3.4.

**У.3.4.** В параграфе 1.2 мы получили следующее выражение для тока через единственный уровень:

$$I = \frac{q}{h} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(\varepsilon) - f_2(\varepsilon)],$$

а также для среднего числа электронов

$$N = \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2},$$

используя систему уравнений эволюции на одиночном одноэлектронном уровне (без спинового вырождения). В многоэлектронной картине у нас есть два уровня «0» и «1», отвечающие одноэлектронному уровню, который является пустым либо заполненным соответственно. Запишите соответствующие уравнения эволюции в данном представлении и выведите заново выражения для величин « $N$ » и « $I$ ».

**У.3.5.** Рассматривая канал с двумя вырожденными по спину уровнями и предполагая следующие значения параметров:  $\mu = 0$ ,  $\varepsilon = 0.2$  эВ,  $k_B T = 0.025$  эВ,  $\gamma_1 = \gamma_2 = 0.005$  эВ, вычислите:

(а) Число электронов как функцию напряжения на затворе  $V_G$  при значении  $V_D = 0$  и  $U_L = -qV_G$ , при этом используйте (1) многоэлектронное уравнение эволюции и (2) ограниченный потенциал ССП, описываемый выражением  $U_{SCF} = U_0(N - N_0)$  с  $N_0 = 1$ . Рассмотрите случаи различных значений  $U_0 = 0.025$  эВ и  $U_0 = 0.25$  эВ.

(б) Зависимость тока от напряжения стока  $V_D$ , предполагая, что  $V_G = 0$ ,  $U_L = -qV_d/2$ , и используя (1) многоэлектронное уравнение эволюции и (2) ограниченный потенциал ССП, указанный в (а). Рассмотрите случаи различных значений  $U_0 = 0.025$  эВ и  $U_0 = 0.25$  эВ и сравните с рис. 3.4.3.

(в) Повторите расчеты, предложенные в пунктах (а) и (б), для случая неограниченного потенциала ССП (уравнение (3.4.9)), в котором исключено самовоздействие:

$$\begin{aligned} U_{SCF}(\uparrow) &= U_0(\Delta N - \Delta n_\uparrow) = U_0(\Delta n_\downarrow) = U_0(n_\downarrow - 0.5), \\ U_{SCF}(\downarrow) &= U_0(n_\uparrow - 0.5). \end{aligned}$$

Примечание: результат может быть различным в зависимости от того, является ли начальное приближение симметричным, т. е.  $U_{SCF}(\uparrow) = U_{SCF}(\downarrow)$ , либо нет, т. е.  $U_{SCF}(\uparrow) \neq U_{SCF}(\downarrow)$ .

**У3.6.** На рис. 3.4.3 *a* ( $U_0 = 0.25$  эВ) многоэлектронный подход порождает наличие двух плато для графика тока: нижнее плато находится в интервале  $\varepsilon_2 + U_L > \mu_1 > \varepsilon_1 + U_L$ , где  $f'_1 \cong 1$ ,  $f''_1 \cong 0$ , а верхнее плато находится в интервале  $\mu_1 > \varepsilon_2 + U_L > \varepsilon_1 + U_L$ , где  $f'_1 \cong 1$ ,  $f''_1 \cong 1$ . В обоих случаях величины  $f'_2 \cong 0$ ,  $f''_2 \cong 0$ . Используя уравнения (3.3.3) и (3.3.8), покажите, что значения тока на этих плато даются выражениями

$$\frac{2\gamma_1\gamma_2}{2\gamma_1 + \gamma_2} \quad \text{и} \quad \frac{2\gamma_1\gamma_2}{\gamma_1 + \gamma_2}$$

соответственно.

## Глава 4

# Базисные функции

В предыдущей главе мы видели, что расчет энергетических уровней в атоме может быть проведен в рамках метода ССП. Это возможно сделать благодаря тому, что сферическая симметрия позволяет свести задачу к одномерной. Молекулы, с другой стороны, не обладают сферической симметрией, и для того, чтобы задача могла быть решена численно, необходим другой, более эффективный подход. Использование разложения по базисным функциям составляет основу другого эффективного метода численного решения уравнения Шредингера (или любого другого дифференциального уравнения подобного типа). В то же время данный аппарат является методически важным инструментом, демонстрирующим фундаментальные положения квантовой механики. В данной главе будут обсуждаться оба этих аспекта.

Основная идея состоит в том, что искомая волновая функция, вообще говоря, может быть представлена в виде разложения по системе базисных функций  $u_m(\vec{r})$ :

$$\Phi(\vec{r}) = \sum_{m=1}^M \phi_m u_m(\vec{r}).$$

После этого мы можем представить волновую функцию как вектор-столбец, состоящий из коэффициентов этого разложения:

$$\Phi(\vec{r}) \rightarrow \{\phi_1 \phi_2 \dots \phi_M\}^T, \quad \text{«}T\text{» означает транспонирование.}$$

Разница, однако, состоит в том, что теперь у нас появляется свобода выбора системы базисных функций  $u_m(\vec{r})$ : если мы выберем функции, которые выглядят весьма похожими на нашу ожидаемую волновую функцию, то можно представить ее достаточно точным разложением всего лишь с небольшим числом членов, что сильно уменьшит размер конечной матрицы  $[H]$ . Это обстоятельство делает данный подход удобным в качестве инструмента численных методов (напоминающего по духу концепцию «интерполяционных функций» в методе конечных элементов (см. [44, 51])), что будет проиллюстрировано нами на простом примере в параграфе 4.1.

Следует отметить, что концепция базисных функций в действительностя является значительно более общей. Эти функции можно рассматривать как координатные оси в гильбертовом пространстве, как это будет описано в параграфе 4.2. Мы продемонстрируем значительные возможности и гибкость этой точки зрения в целом, используя понятие матрицы плотности в параграфе 4.3, а также рассматривая дальнейшие примеры в параграфе 4.4.

## 4.1. Базисные функции как инструмент численных методов

Основная идея может быть сформулирована достаточно простым образом. Мы записываем волновую функцию в виде разложения по любой системе базисных функций  $u_m(\vec{r})$ :

$$\Phi(\vec{r}) = \sum_m \phi_m u_m(\vec{r}) \quad (4.1.1)$$

и подставляем ее в уравнение Шредингера  $E\Phi(\vec{r}) = H_{op}\Phi(\vec{r})$ , получая равенство

$$E \sum_m \phi_m u_m(\vec{r}) = \sum_m \phi_m H_{op} u_m(\vec{r}).$$

Умножая обе части на сопряженную функцию  $u_n^*(\vec{r})$  и интегрируя по всей области определения  $\vec{r}$ , мы получим равенство

$$E \sum_m S_{nm} \phi_m = \sum_m H_{nm} \phi_m,$$

которое можно записать в форме матричного уравнения

$$E[S]\{\phi\} = [H]\{\phi\}, \quad (4.1.2)$$

где коэффициенты

$$S_{nm} = \int d\vec{r} u_n^*(\vec{r}) u_m(\vec{r}), \quad (4.1.3a)$$

$$H_{nm} = \int d\vec{r} u_n^*(\vec{r}) H_{op} u_m(\vec{r}). \quad (4.1.3b)$$

Для продолжения расчета необходимо вычислить эти интегралы, что является наиболее затратным по времени этапом в процессе решения

задачи. Но как только эти матричные элементы найдены, можно будет сразу же найти собственные значения  $E_\alpha$  и собственные вектора  $\{\Phi_\alpha\}$  гамильтоновской матрицы. Собственные функции далее могут быть записаны в «координатном представлении» путем подстановки найденных коэффициентов обратно в исходное разложение (4.1.1):

$$\Phi_\alpha(\vec{r}) = \frac{1}{\sqrt{Z_\alpha}} \sum_m \phi_{m\alpha} u_m(\vec{r}), \quad \Phi_\alpha^*(\vec{r}) = \frac{1}{\sqrt{Z_\alpha}} \sum_n \phi_{n\alpha}^* u_n^*(\vec{r}), \quad (4.1.4)$$

где постоянная  $Z_\alpha$  выбирается исходя из условия правильной нормировки:

$$1 = \int d\vec{r} \Phi_\alpha^*(\vec{r}) \Phi_\alpha(\vec{r}) \rightarrow Z_\alpha = \sum_n \sum_m \phi_{n\alpha}^* \phi_{m\alpha} S_{nm}. \quad (4.1.5)$$

Уравнения (4.1.1)–(4.1.5) составляют основу математических соотношений, возникающих при использовании метода разложения по базисным функциям.

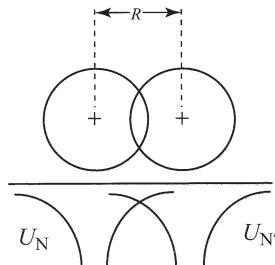


Рис. 4.1.1. Функции  $U_N$  и  $U_{N'}$  являются кулоновскими потенциалами от левого и правого ядер молекулы водорода  $H_2$

**Частный пример.** Рассмотрим следующий частный пример, который поможет нам еще раз обсудить физическое содержание и схему практической реализации рассматриваемого метода. В разделе 3.3.3 мы пришли к выводу о том, что низшее значение энергии находится из уравнения Шредингера, включающего потенциалы двух ядер (рис. 4.1.1), но не учитывающим самосогласованный потенциал,

$$E_{\alpha 0} \Phi_{\alpha 0}(\vec{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_N(\vec{r}) + U_{N'}(\vec{r}) \right) \Phi_{\alpha 0}(\vec{r}), \quad (4.1.6)$$

приближенно можно записать в виде ( $E_1 \equiv -E_0 = -13.6$  эВ)

$$E_{B0} = E_1 + \frac{a+b}{1+s}, \quad (4.1.7)$$

где

$$a = -2E_0 \frac{1 - (1 + R_0)e^{-2R_0}}{R_0}, \quad b = -2E_0(1 + R_0)e^{-R_0},$$

$$s = e^{-R_0}[1 + R_0 + (R_0^2/3)], \quad R_0 \equiv R/a_0,$$

а величина  $R$  есть расстояние между центрами атомов водорода.

Сейчас мы используем метод базисных функций, чтобы показать, как этот результат может быть получен из уравнения (4.1.6).

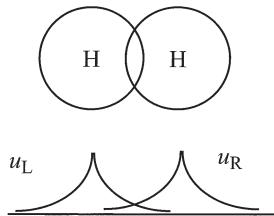


Рис. 4.1.2. Удобный вариант выбора базисных функций для описания молекулы водорода  $H_2$  в виде двух  $1s$ -орбиталей, центры которых расположены вблизи левого и правого ядер соответственно

Обратите внимание на то, что потенциал  $U(\vec{r}) = U_N(\vec{r}) + U_{N'}(\vec{r})$  в уравнении (4.1.6) не является сферически симметричным, в отличие от атомных потенциалов, которые мы обсуждали в главах 2 и 3. Это означает, что мы не можем отдельно решить радиальную часть уравнения Шредингера. Вообще говоря, нам необходимо решать полное трехмерное уравнение Шредингера, что с вычислительной точки зрения является весьма трудной задачей. В то же время часто использование разложения волновой функции по базисным функциям позволяет довести эту задачу до практического решения. В рассматриваемом частном примере мы будем использовать всего лишь две базисные функции

$$\Phi_{\alpha 0}(\vec{r}) = \phi_L(\vec{r})u_L(\vec{r}) + \phi_R(\vec{r})u_R(\vec{r}), \quad (4.1.8)$$

где функции  $u_L(\vec{r})$  и  $u_R(\vec{r})$  представляют собой  $1s$ -орбитали атома водорода, центры которых находятся вблизи левого и правого ядер соответственно

(см. рис. 4.1.2). Это означает, что выполняются соотношения

$$E_1 u_L(\vec{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_N(\vec{r}) \right) u_L(\vec{r}) \quad (4.1.9a)$$

и

$$E_1 u_R(\vec{r}) = \left( -\frac{\hbar^2}{2m} \nabla^2 + U_{N'}(\vec{r}) \right) u_R(\vec{r}). \quad (4.1.9b)$$

Запись волновой функции в виде (4.1.8) мотивирована тем, что это выражение правильно описывает структуру основного состояния, когда атомы достаточно далеко удалены друг от друга в пространстве. В этом случае собственные состояния есть

$$(\phi_L \quad \phi_R) = (1 \quad 0) \quad \text{и} \quad (\phi_L \quad \phi_R) = (0 \quad 1).$$

Следует ожидать, что в том случае, когда длина межатомной связи  $R$  не очень мала (по сравнению с боровским радиусом  $a_0$ ), уравнение (4.1.8) при надлежащем выборе коэффициентов  $\phi_L$  и  $\phi_R$  все еще будет достаточно аккуратно описывать собственные состояния.

Коль скоро мы представили искомую волновую функцию в виде разложения лишь по двум базисным функциям  $u_L$  и  $u_R$ , матрицы  $[S]$  и  $[H]$  в уравнении (4.1.2) будут иметь размер  $(2 \times 2)$ . Элементы этих матриц могут быть найдены из уравнений (4.1.3a), (4.1.3b) с использованием уравнений (4.1.9a), (4.1.9b):

$$S = \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix} \quad \text{и} \quad H = \begin{pmatrix} E_1 + a & E_1 s + b \\ E_1 s + b & E_1 + a \end{pmatrix}, \quad (4.1.10)$$

где

$$s = \int d\vec{r} u_L^*(\vec{r}) u_R(\vec{r}) = \int d\vec{r} u_R^*(\vec{r}) u_L(\vec{r}), \quad (4.1.11a)$$

$$a = \int d\vec{r} u_L^*(\vec{r}) U_{N'}(\vec{r}) u_L(\vec{r}) = \int d\vec{r} u_R^*(\vec{r}) U_N(\vec{r}) u_R(\vec{r}), \quad (4.1.11b)$$

$$\begin{aligned} b &= \int d\vec{r} u_L^*(\vec{r}) U_N(\vec{r}) u_R(\vec{r}) = \int d\vec{r} u_L^*(\vec{r}) U_{N'}(\vec{r}) u_R(\vec{r}) \\ &= \int d\vec{r} u_R^*(\vec{r}) U_N(\vec{r}) u_L(\vec{r}) = \int d\vec{r} u_R^*(\vec{r}) U_{N'}(\vec{r}) u_L(\vec{r}). \end{aligned} \quad (4.1.11c)$$

Отсюда следует, что уравнение (4.1.2) принимает вид

$$E = \begin{pmatrix} \phi_L \\ \phi_R \end{pmatrix} = \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix}^{-1} \begin{pmatrix} E_1 + a & E_1 s + b \\ E_1 s + b & E_1 + a \end{pmatrix} \begin{pmatrix} \phi_L \\ \phi_R \end{pmatrix}, \quad (4.1.12)$$

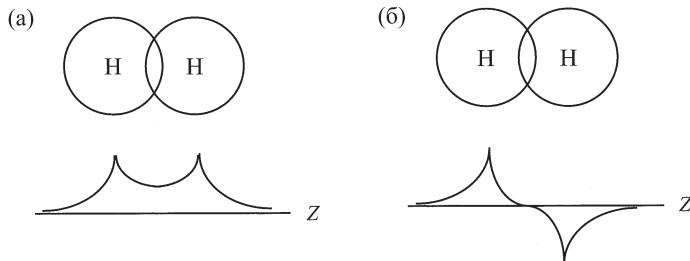


Рис. 4.1.3

из которого нетрудно получить два собственных значения. Нижнее собственное значение отвечает симметричному состоянию (B), а верхнее — антисимметричному состоянию (A):

$$E_B = E_1 + \frac{a+b}{1+s} \quad \text{и} \quad E_A = E_1 + \frac{a-b}{1-s}. \quad (4.1.13)$$

Параметры  $a$ ,  $b$  и  $s$  здесь могут быть рассчитаны путем подстановки известных базисных функций  $u_L(\vec{r})$ ,  $u_R(\vec{r})$  и потенциалов ядер  $U_N(\vec{r})$  и  $U_{N'}(\vec{r})$  в уравнения (4.1.11a), (4.1.11b), (4.1.11c).

Волновые функции, отвечающие симметричному и антисимметричному состояниям, будут определены выражениями

$$(\phi_L \quad \phi_R)_B = (1 \quad 1) \quad \text{и} \quad (\phi_L \quad \phi_R)_A = (1 \quad -1),$$

которые представляют собой симметричную (B) (рис. 4.1.3 а) и антисимметричную (A) (рис. 4.1.3 б) комбинации двух  $1s$ -орбиталей, центры которых расположены вблизи двух ядер. Оба электрона в молекуле водорода  $H_2$  находятся в симметричном состоянии, чья волновая функция есть

$$\Phi_{B0} = \frac{1}{\sqrt{Z}} [u_L(\vec{r}) + u_R(\vec{r})], \quad (4.1.14)$$

где

$$u_L(\vec{r}) = \frac{1}{\sqrt{\pi a_0^3}} \exp \left[ \frac{-|\vec{r} - \vec{r}_L|}{a_0} \right], \quad \vec{r}_L = -(R/2)\vec{z},$$

$$u_R(\vec{r}) = \frac{1}{\sqrt{\pi a_0^3}} \exp \left[ \frac{-|\vec{r} - \vec{r}_R|}{a_0} \right], \quad \vec{r}_L = +(R/2)\vec{z}.$$

Константа  $Z$  находится из условия нормировки волновой функции:

$$1 = \int d\vec{r} \Phi_{B0}^*(\vec{r}) \Phi_{B0}(\vec{r}) = \frac{2(1+s)}{Z} \rightarrow Z = 2(1+s).$$

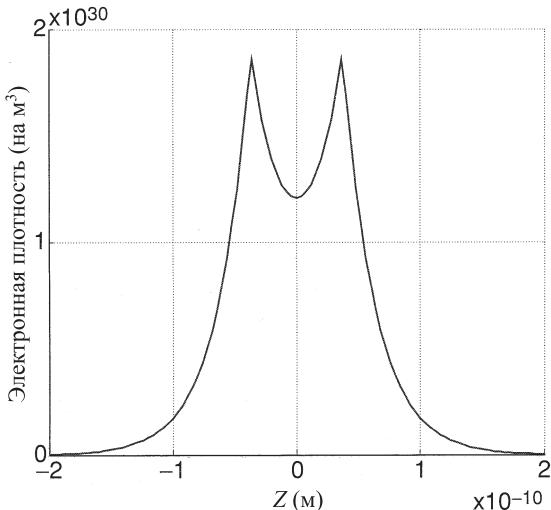


Рис. 4.1.4. Распределения электронной плотности вдоль линии, соединяющей два атома водорода. В равновесии атомы удалены друг от друга на расстояние  $R = 0.074$  нм

На рис. 4.1.4 показано распределение электронной плотности вдоль линии, соединяющей два ядра. Поскольку имеется два электрона — один со спином вверх и другой со спином вниз — распределение электронной плотности  $n(\vec{r})$  в молекуле водорода  $\text{H}_2$  определяется функцией  $|\Phi_{B0}(\vec{r})|^2$ , умноженной на два.

**Как получаются точные результаты при использовании лишь двух базисных функций?** Решая уравнение Шредингера на дискретной решетке, как это делалось в главах 2 и 3, мы фактически используем набор из очень большого числа базисных функций. К примеру, если узлы решетки отстоят друг от друга на  $0.5$  Å, а размер атома составляет  $2.5$  Å, то нам требуется  $5^3 = 125$  узлов решетки (каждый из которых представляет собой базисную функцию), поскольку задача является трехмерной. Что же мы теряем, используя всего одну базисную функцию вместо 125? Ответ

состоит в том, что наши результаты являются достаточно точными лишь в ограниченном интервале энергий.

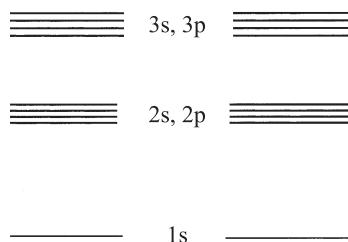


Рис. 4.1.5. Собственные состояния молекулы водорода  $\text{H}_2$ , полученные в предположении, что атомы находятся не слишком близко друг к другу. Эти состояния могут быть использованы как базисные функции при более точном расчете состояний молекулы водорода

Чтобы увидеть это, предположим, что учитывается не только  $1s$  орбиталь, но также и  $2s$ ,  $2p_x$ ,  $2p_y$ ,  $2p_z$ ,  $3s$ ,  $3p_x$ ,  $3p_y$  и  $3p_z$  орбитали (см. рис. 4.1.5). При этом низшие энергетические состояния будут по-прежнему формироваться, в основном, из функций  $1s$  орбиталей с ничтожно малой примесью остальных волновых функций, что позволит получать достаточно точные результаты с использованием лишь одной базисной функции на один атом. Причина состоит в том, что присутствие недиагонального матричного элемента  $M$  существенно меняет собственные значения матрицы

$$\begin{pmatrix} E_1 & M \\ M & E_2 \end{pmatrix}$$

лишь тогда, когда он становится сравнимым с разностью диагональных матричных элементов, т. е. при  $M \geq |E_1 - E_2|$ . Диагональные матричные элементы по величине примерно соответствуют уровням энергии изолированных атомов, поэтому разность  $|E_1 - E_2|$  равна 10 эВ, если, к примеру, иметь в виду  $1s$ - и  $2s$ -уровни атома водорода. Недиагональный матричный элемент  $M$  зависит от степени близости двух атомов, и для типичных молекул с ковалентной связью в твердых телах составляет порядка 2 эВ, что меньше величины  $|E_1 - E_2|$ . В результате симметричное состояние, в основном, формируется из волновых функций типа  $1s$ , и наше рассмотрение, основанное на матрице  $2 \times 2$ , является достаточно точным. При этом следует заметить, что точный расчет вышележащих энергетических уровней потребовал бы учета большего числа базисных функций.

**Расчеты на базе первых принципов (Ab initio) и в полуэмпирических моделях.** Концепция базисных функций широко используется для расчетов на базе первых принципов, когда уравнение Шредингера решается напрямую с учетом самосогласованного поля. Для больших молекул или твердых тел подобные вычисления являются слишком громоздкими ввиду необходимости использовать значительное число базисных функций и, соответственно, интегралов, которые определяют матричные элементы. Интегралы, возникающие при учете самосогласованного поля, требуют особенно много машинного времени. В силу этих обстоятельств широко применяются так называемые полуэмпирические методы, где матричные элементы подбираются исходя из сопоставления теоретических и экспериментальных результатов. Подобные подходы являются весьма полезными в тех случаях, когда мы можем получить значения параметров из эксперимента, а затем использовать их для предсказания результатов в других ситуациях. К примеру, мы можем рассчитать параметры, определяющие уровни энергии бесконечно протяженного твердого тела, а затем использовать эти параметры для расчета энергетических уровней наноструктуры конечных размеров.

## 4.2. Базисные функции как методология

Выше было отмечено, что все практически реализуемые методы расчета энергетических уровней в молекулах и твердых телах подразумевают использование разложения по базисным функциям. Однако концепция базисных функций есть нечто большее, чем просто численный метод. Она развивает нашу интуицию и позволяет дать физическую интерпретацию изучаемым явлениям. В самом деле, концепция записи волновой функции в виде суперпозиции базисных функций является одной из центральных идей квантовой механики, и мы попытаемся показать это на примерах в дальнейшем.

**Векторное пространство и гильбертово пространство.** Полезно сравнить уравнение (4.1.1) с записью обычного трехмерного вектора  $\vec{V}$ , разложенного по трем единичным векторам  $\vec{x}$ ,  $\vec{y}$  и  $\vec{z}$ :

$$\vec{V} = V_x \vec{x} + V_y \vec{y} + V_z \vec{z} \longleftrightarrow \Phi = \phi_1 u_1 + \phi_2 u_2 + \phi_3 u_3 + \dots$$

Волновую функцию можно рассматривать как вектор состояния в  $N$ -мерном пространстве, называемом гильбертовым пространством. Тогда полное число базисных функций  $u_m(\vec{r})$  будет равно  $N$ . Коэффициенты  $\phi_m$  в уравнении (4.1.1) подобны компонентам вектора состояния  $\Phi$ , в то время как функции  $u_m(\vec{r})$  представляют собой единичные векторы вдоль  $N$

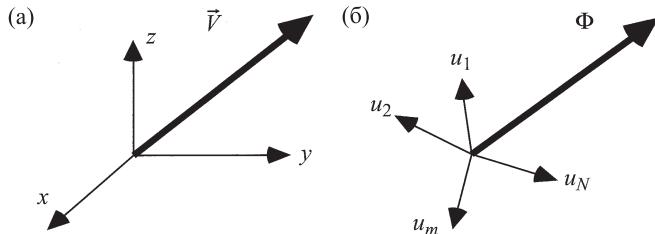


Рис. 4.2.1. (а) Обычный вектор  $\vec{V}$  в трехмерном пространстве может быть выражен через свои компоненты, направленные по осям  $x$ ,  $y$  и  $z$ . (б) Вектор состояния  $\Phi(\vec{r})$  может быть представлен компонентами вдоль «осей», соответствующих базисным функциям  $u_m(\vec{r})$

координатных осей. Выбор другой системы базисных функций  $u_m(\vec{r})$  соответствует выбору другой системы координат: все компоненты (проекции)  $\phi_m$  вдоль различных осей изменяются, при этом оставляя неизменным сам вектор состояния. В принципе, число компонент  $N$  является бесконечным, но на практике зачастую можно получать достаточно точные результаты и с конечным значением  $N$ . Мы попытались продемонстрировать рассматриваемую аналогию на рис. 4.2.1 (заметьте, что не так просто рассуждать об  $N$ -мерном векторе ( $N > 3$ ), используя двумерный лист бумаги). В обозначениях Дирака, которые являются очень удобными и широко используются, вектор состояния, ассоциируемый с волновой функцией  $\Phi(\vec{r})$ , обозначается как  $|\Phi\rangle$  («кет»), а единичные векторы, ассоциируемые с базисными функциями  $u_m(\vec{r})$ , также записываются как кет-векторы  $|m\rangle$ . В этих обозначениях разложение по базисным функциям (см. уравнение (4.1.1)) записывается как

$$\Phi(\vec{r}) = \sum_m \phi_m u_m(\vec{r}) \rightarrow \text{обозначения Дирака} \rightarrow |\Phi\rangle = \sum_m \phi_m |m\rangle. \quad (4.2.1)$$

**Скалярное произведение.** Понятие скалярного произведения является центральным в векторной алгебре:

$$\vec{A} \cdot \vec{B} = A_x B_x + A_y B_y + A_z B_z = \sum_{m=x,y,z} A_m B_m.$$

Соответствующее понятие в гильбертовом пространстве можно определить как интеграл перекрытия между любыми двумя функциями  $f(\vec{r})$  и  $g(\vec{r})$ :

$$\int d\vec{r} f^*(\vec{r}) g(\vec{r}) \rightarrow \text{обозначения Дирака} \rightarrow \langle f | g \rangle.$$

Общие черты интеграла перекрытия, записанного в решеточном представлении, и скалярного произведения можно увидеть из выражения

$$\int d\vec{r} f^*(\vec{r}) g(\vec{r}) \cong a^3 \sum_m f^*(\vec{r}_m) g(\vec{r}_m).$$

В представлении дискретной решетки (см. рис. 2.2.1) «компоненты»  $f_m$  функции  $f(\vec{r})$ , соответствующая базисной функции  $u_m(\vec{r})$ , определяется величиной  $f(\vec{r}_m)$  так же, как число  $A_m$  представляет собой проекцию вектора  $\vec{A}$  на направление базисного вектора  $\vec{m}$ :

$$\int d\vec{r} f^*(\vec{r}) g(\vec{r}) \cong a^3 \sum_m f_m^* g_m \quad \text{ср. с} \quad \vec{A} \cdot \vec{B} = \sum_{m=x,y,z} A_m B_m.$$

Различие здесь заключается в том, что мы берем комплексное сопряжение от одной из функций (что не является принципиальным, если мы имеем дело с вещественными функциями), и это отражается в записи одного вектора как «бра»-вектор  $\langle f |$  в противоположность другому «кет»-вектору  $|g\rangle$ . Скалярное произведение в этих обозначениях представляется объединением «бра» и «кет» в «брaket» (по-английски — «скобка»)  $\langle f | g \rangle$ .

**Ортогональность.** Говорят, что система координат является ортогональной, если выполняются соотношения  $\vec{m} \cdot \vec{n} = \delta_{nm}$ , где индексы  $n$  и  $m$  соответствуют координатам  $x, y, z$ , а символ  $\delta_{nm}$  есть символ Кронекера, определяемый как

$$\begin{aligned} \delta_{nm} &= 1, && \text{если } n = m \\ &= 0, && \text{если } n \neq m. \end{aligned} \tag{4.2.2}$$

Это свойство базисных векторов обычно выполняется (к примеру,  $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{z} = \vec{z} \cdot \vec{x} = 0$ ), но в некоторых случаях приходится иметь дело и с неортогональными координатными системами. По аналогии можно сказать, что базисные функции  $u_m(\vec{r})$  являются ортогональными, если выполняется следующее соотношение:

$$\int d\vec{r} u_n^*(\vec{r}) u_m(\vec{r}) = \delta_{nm} \xrightarrow{\text{Обозначения Дирака}} \langle n | m \rangle = \delta_{nm}. \tag{4.2.3a}$$

Обратите внимание на то, что базисные функции, использованные нами при рассмотрении молекулы водорода (см. рис. 4.1.2), не являются ортогональными, поскольку

$$\int d\vec{r} u_L^*(\vec{r}) u_R(\vec{r}) \equiv s = e^{-R_0} [1 + R_0 + (R_0^2/3)] \neq 0.$$

В общем случае имеет место соотношение

$$\int d\vec{r} \tilde{u}_n^*(\vec{r}) u_m(\vec{r}) = S_{nm} \xrightarrow{\text{Обозначения Дирака}} \langle n | m \rangle = S_{nm}. \quad (4.2.3b)$$

**Ортогонализация.** При работе с неортогональной системой базисных функций  $\{u_n(\vec{r})\}$  мы можем определить новую систему функций

$$\tilde{u}_i(\vec{r}) = \sum_n [S^{-1/2}]_{ni} u_n(\vec{r}), \quad (4.2.4)$$

которые являются ортогональными. Последнее можно продемонстрировать следующим образом:

$$\int \tilde{u}_i(\vec{r}) \tilde{u}_j(\vec{r}) = \sum_n \sum_m [S^{-1/2}]_{in} S_{nm} [S^{-1/2}]_{mj} = [S^{-1/2} S S^{-1/2}]_{ij} = \delta_{ij},$$

где мы воспользовались уравнением (4.2.3b). Это означает, что при использовании в качестве базиса новой системы  $\tilde{u}_i(\vec{r})$  матрица интегралов перекрытия  $[S] = [I]$ , где  $[I]$  превращается в единичную матрицу, т. е. диагональную матрицу с единицами на главной диагонали. Это является свойством базиса, составленного из ортонормированных функций, что делает работу с ним принципиально более удобной.

Даже если мы стартуем с неортогонального базиса, зачастую удобно ортогонализовать его. В процессе ортогонализации можно потерять информацию о локальной природе исходных базисных функций и связанной с этим физикой. К примеру, функции  $\{u_n(\vec{r})\}$ , которые использовались при рассмотрении молекулы водорода, были локализованы соответственно на левом и правом атомах. Но функции ортонормированного базиса  $\{\tilde{u}_i(\vec{r})\}$ , которые являются линейными комбинациями этих двух функций, и будут менее локализованными, чем  $\{u_n(\vec{r})\}$ . Как правило, бывает непросто найти такие функции, которые были бы одновременно и локальными и ортонормированными. Ниже, начиная с этого момента будем предполагать, что базисные функции являются ортонормированными.

**Операторы.** Оператор, подобный  $H_{op}$ , действуя на вектор, изменяет его так, что в результате получается новый вектор, отвечающий другому состоянию. Можно сказать, что в результате такой процедуры происходит поворот вектора в пространстве его компонент. С использованием обычных векторов мы можем представить вращение, введя матрицу поворота

$$\begin{pmatrix} A'_x \\ A'_y \end{pmatrix} = \begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix},$$

где для простоты мы рассмотрели двумерный вектор. Каким же образом нам следует записать матрицу  $[R]$ , отвечающую действию оператора  $R_{op}$ ? Общим принципом здесь является следующее:  $R_{nm} = \vec{n} \cdot (R_{op} \vec{m})$ . К примеру, предположим, что мы рассматриваем оператор, который поворачивает вектор на угол  $\theta$ . В этом случае мы получим следующие выражения:

$$R_{xx} = \vec{x} \cdot (R_{op} \vec{x}) = \vec{x} \cdot (\vec{x} \cos \theta + \vec{y} \sin \theta) = \cos \theta,$$

$$R_{yx} = \vec{y} \cdot (R_{op} \vec{x}) = \vec{y} \cdot (\vec{x} \cos \theta + \vec{y} \sin \theta) = \sin \theta,$$

$$R_{xy} = \vec{x} \cdot (R_{op} \vec{y}) = \vec{x} \cdot (-\vec{x} \sin \theta + \vec{y} \cos \theta) = -\sin \theta,$$

$$R_{yy} = \vec{y} \cdot (R_{op} \vec{y}) = \vec{y} \cdot (-\vec{x} \sin \theta + \vec{y} \cos \theta) = \cos \theta.$$

Матричное представление любого оператора  $A_{op}$  в гильбертовом пространстве записывается с использованием подобных обозначений:

$$[A]_{nm} = \int d\vec{r} u_n^*(\vec{r})(A_{op} u_m(\vec{r})) \xrightarrow{\text{Обозначения Дирака}} [A]_{nm} = \langle n | A_{op} m \rangle. \quad (4.2.5)$$

**Постоянный оператор.** Какая матрица отвечает постоянному оператору, который просто умножает вектор состояния на константу  $C$ ? В общем случае, ответ записывается как матрица

$$[C]_{nm} = C \int d\vec{r} u_n^*(\vec{r}) u_m(\vec{r}) = C [S]_{nm}, \quad (4.2.6)$$

что сводится к произведению  $C[I]$  для ортонормированного базиса.

**Матричное представление уравнения Шредингера.** Матричное представление уравнения Шредингера, полученное в предыдущем разделе,

$$E\Phi(\vec{r}) = H_{op}\Phi(\vec{r}) \rightarrow E[S]\{\phi\} = [H]\{\phi\} \quad (4.2.7)$$

теперь может быть интерпретировано на языке подхода, описанного в данном разделе. Как и оператор вращений в векторном пространстве, любой дифференциальный оператор в гильбертовом пространстве имеет матричное представление. Как только мы выбрали набор базисных функций, оператор  $H_{op}$  становится матрицей  $[H]$ , а постоянная  $E$  становится матрицей вида  $E[S]$ :

$$[S]_{nm} = \langle n | m \rangle \equiv \int d\vec{r} u_n^*(\vec{r}) u_m(\vec{r}), \quad (4.2.8a)$$

$$[H]_{nm} = \langle n | H_{op} m \rangle \equiv \int d\vec{r} u_n^*(\vec{r})(H_{op} u_m(\vec{r})). \quad (4.2.8b)$$

Следуя уравнению (4.2.4), можно ортогонализовать систему базисных функций, так, что в терминах ортогональных базисных функций  $\{\tilde{u}_i(\vec{r})\}$  уравнение Шредингера будет иметь форму стандартного матричного уравнения на собственные значения:

$$E\{\tilde{\phi}\} = [\tilde{H}]\{\tilde{\phi}\},$$

где матричные элементы  $[\tilde{H}]$  определяются выражениями

$$[\tilde{H}]_{ij} = \int d\vec{r} \tilde{u}_i^*(\vec{r})(H_{op} \tilde{u}_j(\vec{r})).$$

**Изменение базиса.** Предположим, что мы выполнили разложение волновой функции в одном базисе и хотим перейти к другому базису:

$$\Phi(\vec{r}) = \sum_m \phi_m u_m(\vec{r}) \rightarrow \Phi(\vec{r}) = \sum_i \phi'_i u'_i(\vec{r}). \quad (4.2.9)$$

Такое преобразование может быть описано путем преобразования матрицы  $[C]$  в другой базис, компоненты которого связаны с компонентами старого базиса соотношением

$$u'_i(\vec{r}) = \sum_m C_{mi} u_m(\vec{r}). \quad (4.2.10)$$

Из уравнений (4.2.9) и (4.2.10) следует, что

$$\phi_m = \sum_m C_{mi} \phi'_i \rightarrow \{\phi\} = [C]\{\phi'\}. \quad (4.2.11a)$$

Подобным образом можно показать, что любая матрица  $[A']$  в новом представлении связана с этой же матрицей  $[A]$  в старом представлении следующим преобразованием:

$$A'_{ji} = \sum_n \sum_m C_{nj}^* A_{nm} C_{mi} \rightarrow [A'] = [C]^\dagger [A] [C]. \quad (4.2.11b)$$

**Унитарное преобразование.** Существует специальный класс преобразований, которые сохраняют «норму» вектора состояния

$$\sum_m \phi_m^* \phi_m = \sum_i \phi'^*_i \phi'_i \rightarrow \{\phi\}^\dagger \{\phi\} = \{\phi'\}^\dagger \{\phi'\}. \quad (4.2.12)$$

Подставляя выражение для  $\{\phi\}$  из уравнения (4.2.11a) в уравнение (4.2.12), получаем

$$\{\phi'\}^\dagger [C]^\dagger [C] \{\phi'\} = \{\phi'\}^\dagger \{\phi'\} \rightarrow [C]^\dagger [C] = [I]. \quad (4.2.13)$$

Матрица  $[C]$ , удовлетворяющая условию (4.2.13), называется унитарной матрицей, а соответствующее преобразование называется унитарным преобразованием.

Обратите внимание на то, что для унитарного преобразования  $[C]^\dagger = [C]^{-1}$ , что позволяет записывать с помощью уравнения (4.2.11b) и обратное преобразование просто как  $[A] = [C][A'][C]^\dagger$ .

**Эрмитовские операторы.** Матрица  $[A]$ , представляющая эрмитов оператор  $A_{op}$ , остается эрмитовой в любом представлении, т. е. она равна матрице  $[A]^\dagger$ , полученной сопряжением и транспонированием исходной:

$$[A] = [A]^\dagger, \quad \text{т. е. } A_{mn} = A_{mn}^*. \quad (4.2.14)$$

Если  $A_{op}$  сводится к умножению на функцию  $U(\vec{r})$ , то нетрудно показать, что этот оператор является эрмитовым, коль скоро он вещественный:

$$[U]_{mn}^* = \left[ \int d\vec{r} u_m^*(\vec{r}) U(\vec{r}) u_n(\vec{r}) \right]^* = [U]_{nm}.$$

Если  $A_{op}$  является дифференциальным оператором вида  $d/dx$  или  $d^2/dx^2$ , то требуется несколько больше усилий, чтобы убедиться в его эрмитовости. Более простым является использование дискретной аппроксимации на решетке, которую мы обсуждали в главе 2. Уравнение (2.3.1) дает матричное представление оператора  $d^2/dx^2$ , откуда ясно видно, что этот оператор является эрмитовым в данном представлении. Можно также показать, что матрица, являющаяся эрмитовой в одном представлении, будет эрмитовой и в любом другом представлении. Оператор Гамильтона является эрмитовым, поскольку он представляет собой сумму эрмитовых операторов вида  $\partial^2/\partial x^2, \partial^2/\partial y^2, \partial^2/\partial z^2$  и  $U(\vec{r})$ . Согласно общему требованию квантовой механики все собственные значения любого оператора  $A_{op}$ , представляющего наблюдаемую физическую величину, должны быть вещественными. Это условие выполняется, если потребовать, чтобы все операторы  $A_{op}$ , а не только один оператор Гамильтона, были эрмитовыми, поскольку собственные значения эрмитовой матрицы являются вещественными.

Эрмитовские матрицы обладают еще одним полезным свойством: если из столбцов, представляющих компоненты ортонормированной системы векторов, сформирована матрица  $[V]$

$$[V] = [\{V_1\} \{V_2\} \cdots],$$

то эта матрица будет унитарной, т. е. будет выполняться равенство  $[V]^\dagger [V] = [I]$ . Данная унитарная матрица может быть использована для трансформации всех векторов-столбцов  $\{\phi\}$  и матриц  $[M]$  в новый базис, в котором найденные собственные векторы используются как базисные:

$$\{\phi\}_{\text{нов}} = [V]^\dagger \{\phi\}_{\text{стар}} \leftrightarrow \{\phi\}_{\text{стар}} = [V] \{\phi\}_{\text{нов}}, \quad (4.2.15)$$

$$[M]_{\text{нов}} = [V]^\dagger [M]_{\text{стар}} [V] \leftrightarrow [M]_{\text{стар}} = [V] [M]_{\text{нов}} [V]^\dagger. \quad (4.2.16)$$

Если  $[V]$  представляет собой матрицу, состоящую из собственных векторов, отвечающих эрмитовой матрице, такой как  $[H]$ , то в новом представлении матрица  $[H]$  будет диагональной, причем на диагонали будут располагаться ее собственные значения:

$$[H'] = [V]^\dagger [H] [V] = \begin{pmatrix} E_1 & 0 & 0 & 0 & \cdots \\ 0 & E_2 & 0 & 0 & \cdots \\ 0 & 0 & E_3 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \quad (4.2.17)$$

По этой причине процесс вычисления собственных функций и собственных значений часто называют диагонализацией.

### 4.3. Равновесная матрица плотности

Понятие матрицы плотности является одним из центральных в статистической механике, и ее подробному обсуждению будет посвящена глава 7. Причина, по которой я ввожу это понятие уже в этой главе, состоит в том, что при работе с данным объектом мы можем продемонстрировать преимущества подхода, связанного с базисными функциями. Мы начнем с краткого объяснения того, что представляет собой матрица плотности. В главе 3 была рассчитана электронная плотность  $n(\vec{r})$  многоэлектронного атома. Для этого было проведено суммирование плотностей вероятности по всем занятым собственным состояниям  $\alpha$ :

$$n(\vec{r}) = \sum_{\text{занятые } \alpha} |\Phi_\alpha(\vec{r})|^2, \quad \text{см. уравнение (3.1.3).}$$

Такое рассмотрение справедливо при низких температурах и в замкнутых системах с фиксированным числом электронов, занимающих низшие доступные энергетические уровни. В общем случае, однако, состояния могут быть заняты частично, и равновесная электронная плотность должна

быть записана в форме

$$n(\vec{r}) = \sum_{\alpha} f_0(\varepsilon_{\alpha} - \mu) |\Phi_{\alpha}(\vec{r})|^2, \quad (4.3.1)$$

где  $f_0(E) \equiv [1 + \exp(E/k_B T)]^{-1}$  есть функция распределения Ферми (рис. 1.1.3), чье значение определяет вероятность заполнения соответствующего состояния. Так, «0» соответствует свободным состояниям, «1» — занятым состояниям, а значения между 1 и 0 определяют среднюю степень заполнения состояния, которое в какие-то моменты времени занято, а в какие-то — свободно.

Можно ли ввести «волновую функцию»  $\Psi(\vec{r})$  такой многоэлектронной системы так, чтобы квадрат ее модуля определял электронную плотность  $n(\vec{r})$ ? Одна из возможностей состоит в том, чтобы записать ее в форме

$$\Psi(\vec{r}) = \sum_{\alpha} C_{\alpha} \Phi_{\alpha}(\vec{r}), \quad (4.3.2)$$

где  $|C_{\alpha}|^2 = f_0(\varepsilon_{\alpha} - \mu)$ . Но такая запись будет не совсем верной. Если мы возведем модуль этой многоэлектронной «волновой функции» в квадрат, то будем иметь

$$n(\vec{r}) = |\Psi(\vec{r})|^2 = \sum_{\alpha} \sum_{\beta} C_{\alpha} C_{\beta}^* \Phi_{\alpha}(\vec{r}) \Phi_{\beta}^*(\vec{r}),$$

что эквивалентно уравнению (4.3.1) в том и только в том случае, если

$$|C_{\alpha}|^2 = f_0(\varepsilon_{\alpha} - \mu) \equiv f_{\alpha}, \quad C_{\alpha} C_{\beta}^* = 0, \quad \alpha \neq \beta. \quad (4.3.3)$$

Выполнить это требование невозможно, если коэффициенты  $C_{\alpha}$  являются обычными числами — в этом случае  $C_{\alpha} C_{\beta}^*$  должно равняться  $\sqrt{f_{\alpha} f_{\beta}}$  и не может быть равным нулю до тех пор, пока оба коэффициента  $C_{\alpha}$  и  $C_{\beta}$  не равны нулю. Если бы мы попытались записать многоэлектронную волновую функцию в форме (4.3.2), то пришлось бы рассматривать коэффициенты  $C_{\alpha}$  как случайные числа, чьи корреляционные коэффициенты определяются выражением (4.3.3).

Поэтому вместо волновой функции многоэлектронных системмы принято вводить матрицы  $\rho(\alpha, \beta)$ , демонстрирующие корреляцию  $C_{\alpha} C_{\beta}^*$  между каждой парой коэффициентов. Такая матрица  $\rho$  называется матрицей плотности, и в представлении собственных векторов ее элементы могут быть записаны как (см. уравнение (4.3.3))

$$\rho(\alpha, \beta) = f_{\alpha} \delta_{\alpha, \beta}, \quad (4.3.4)$$

где  $\delta_{\alpha,\beta}$  есть дельта-символ Кронекера, определенный как

$$\delta_{\alpha,\beta} = \begin{cases} 1, & \text{если } \alpha = \beta, \\ 0, & \text{если } \alpha \neq \beta. \end{cases}$$

Мы можем переписать выражение (4.3.1) для электронной плотности  $n(\vec{r})$  в виде выражения

$$n(\vec{r}) = \sum_{\alpha} \sum_{\beta} \rho(\alpha, \beta) \Phi_{\alpha}(\vec{r}) \Phi_{\beta}^*(\vec{r}), \quad (4.3.5)$$

обобщением которого является определение

$$\tilde{\rho}(\vec{r}, \vec{r}') = \sum_{\alpha} \sum_{\beta} \rho(\alpha, \beta) \Phi_{\alpha}(\vec{r}) \Phi_{\beta}^*(\vec{r}'), \quad (4.3.6)$$

согласно которому электронная плотность  $n(\vec{r})$  равна диагональным элементам данной матрицы:

$$n(\vec{r}) = [\rho(\vec{r}, \vec{r}')]_{\vec{r}'=\vec{r}}. \quad (4.3.7)$$

Теперь можно утверждать, что выражение (4.3.6) описывает *унитарное преобразование из базиса собственных состояний в координатное представление*. Это можно увидеть, обратив внимание на то, что матрица перехода  $[V]$  находится путем записи каждого из собственных состояний (векторов старого базиса) как векторов-столбцов, в координатном представлении:

$$[V]_{\vec{r}, \alpha} = \Phi_{\alpha}(\vec{r}),$$

а также учитывая тот факт, что матрица перехода является унитарной:  $V^{-1} = V^{\dagger}$ :

$$\Rightarrow [V^{-1}]_{\alpha, \vec{r}} = [V^{\dagger}]_{\alpha, \vec{r}} = [V]_{\vec{r}, \alpha}^* = \Phi_{\alpha}^*(\vec{r}),$$

так что выражение (4.3.6) приобретает форму унитарного преобразования:

$$\tilde{\rho}(\vec{r}, \vec{r}') = \sum_{\alpha} \sum_{\beta} V(\vec{r}, \alpha) \rho(\alpha, \beta) V^{\dagger}(\beta, \vec{r}') \Rightarrow \tilde{\rho} = V \rho V^{\dagger}.$$

На основании изложенного выше мы приходим к важному заключению: *матрица плотности  $\rho$  в состоянии равновесия может быть записана как функция Ферми от гамильтоновской матрицы*. При этом  $I$  есть единичная матрица того же размера, как и  $H$ :

$$\rho = f_0(H - \mu I). \quad (4.3.8)$$

Это матричное соотношение носит общий характер и выполняется в любом представлении. К примеру, если мы используем собственные состояния  $\alpha$  гамильтониана  $H$  в качестве базисных, то как матрица  $[H]$ , так и матрица  $\rho$  будут диагональными:

$$[H] = \begin{pmatrix} \varepsilon_1 & 0 & 0 & \cdots \\ 0 & \varepsilon_2 & 0 & \cdots \\ 0 & 0 & \varepsilon_3 & \cdots \\ \cdots & & & \cdots \end{pmatrix},$$

$$\rho = \begin{pmatrix} f_0(\varepsilon_1 - \mu) & 0 & 0 & \cdots \\ 0 & f_0(\varepsilon_2 - \mu) & 0 & \cdots \\ 0 & 0 & f_0(\varepsilon_3 - \mu) & \cdots \\ \cdots & & & \cdots \end{pmatrix}.$$

Именно такой результат содержится в выражении (4.3.4). Утверждение же состоит в том, что соотношение (4.3.8) остается верным не только в представлении собственных векторов гамильтониана, но также и в любом другом представлении. Если известна матрица оператора Гамильтона  $[H]$ , требуется всего лишь три команды на языке MatLab для того, чтобы вычислить матрицу плотности:

```
[V, D] = eig(H)
rho = 1./(1 + exp((diag(D) - mu)./kT))
rho = V * diag(rho) * V'.
```

Первая команда вычисляет диагональную матрицу  $[D]$ , чьи диагональные элементы есть собственные значения  $[H]$ , а также матрицу  $[V]$ , чьи столбцы состоят из соответствующих собственных векторов. Другими словами,  $[D]$  есть гамильтониан  $[H]$ , записанный в базисе собственных векторов, поэтому  $D = V^\dagger H V$ .

Вторая команда рассчитывает матрицу плотности в представлении собственных векторов, что является несложной процедурой, поскольку в этом представлении как  $[H]$ , так и  $[\rho]$  являются диагональными. Наконец, третья команда преобразует  $[\rho]$  обратно к исходному представлению.

На рис. 4.3.1 показана равновесная электронная плотность в одномерной яме, рассчитанная на дискретной решеткой из 100 точек с шагом 2 Å, при значении  $\mu = 0.25$  эВ. Гамильтониан  $[H]$  в этом случае записывается в виде матрицы размера  $(100 \times 100)$ , явный вид которой может быть найден с помощью процедур, описанных в параграфах 2.2 и 2.3. Далее по описанным выше правилам рассчитывается матрица плотности, причем ее

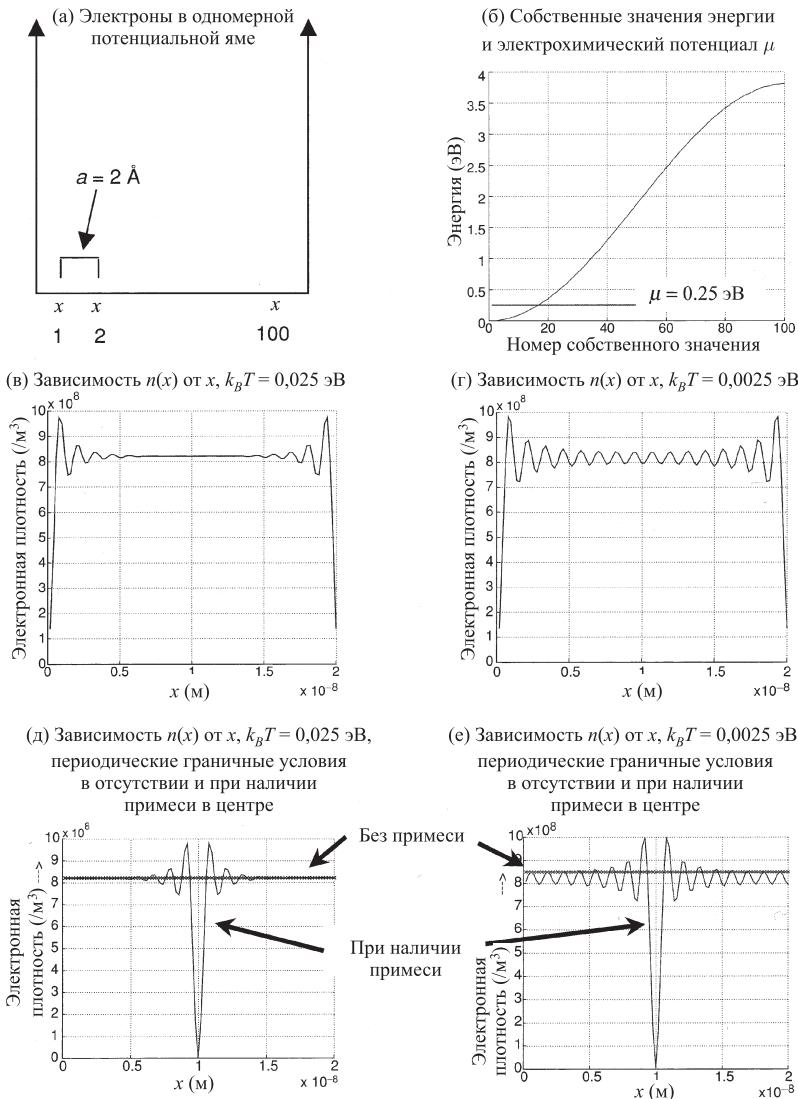


Рис. 4.3.1. Равновесная электронная плотность в одномерной прямоугольной яме, определенная на дискретной решетке из 100 точек с шагом решетки  $2 \text{ \AA}$

диагональные элементы дают значение электронной плотности  $n(x)$  (умноженное на постоянную решетки  $a$ ). Обратите внимание на форму распределения, в виде стоячих волн на рис. 4.3.1  $\delta, \varepsilon, \gamma$ , которые отсутствуют в случае использования периодических граничных условий (рис. 4.3.1  $\alpha, \beta, \zeta$ ). На рисунках 4.3.1  $\delta, \varepsilon$  также показано распределение электронной плотности в форме стоячих волн, возникающих в присутствии сильного отталкивающего потенциала

$$U_0\delta[x - (L/2)], \quad \text{где} \quad U_0 = 2 \text{ эВ} \cdot \text{нм},$$

находящегося в центре потенциальной ямы.

Обратите внимание на то, что матрица плотности может выглядеть различным образом в зависимости от того, какие базисные функции мы используем. В представлении собственных функций она является диагональной, поскольку диагональным является гамильтониан. Однако в координатном представлении решетки у матрицы плотности будут и недиагональные элементы. В любом базисе  $m$  диагональные элементы  $\rho(m, m)$  говорят нам о числе электронов, занимающих состояние  $m$ . В представлении обычного пространства решетки диагональные элементы  $\rho(\vec{r}, \vec{r})$  определяют электронную плотность  $n(\vec{r})$ . След матрицы  $\rho$  (сумма ее диагональных элементов) является инвариантом во всех представлениях и определяет полное число электронов  $N$ :

$$N = \text{Tr}(\rho). \quad (4.3.9)$$

Если нас интересует лишь электронная плотность, то необходимо знать лишь диагональные элементы матрицы плотности. Но недиагональные элементы могут также потребоваться при переходе в другое представление. Они могут также понадобиться для вычисления других физических характеристик (см. раздел 4.4.1).

## 4.4. Дополнение

### 4.4.1. Матрица плотности

В квантовой механике принято отождествлять каждую наблюдаемую величину  $A$  с оператором  $A_{op}$ , для которого в любом базисе может быть найдено матричное представление. Среднее значение  $\langle A \rangle$  данной наблюдаемой величины (то значение, которое мы получим как среднее из серии измерений) определяется величиной

$$\langle A \rangle = \int d\vec{r} \Psi^*(\vec{r}) A_{op} \Psi(\vec{r}).$$

Подставив сюда волновую функцию, записанную в виде разложения по базису в форме уравнения (4.3.2), можно видеть, что

$$\langle A \rangle = \sum_{\alpha} \sum_{\beta} C_{\alpha} C_{\beta}^* \int d\vec{r} \Phi_{\beta}^*(\vec{r}) A_{op} \Psi_{\alpha}(\vec{r}),$$

поэтому

$$\langle A \rangle = \sum_{\alpha} \sum_{\beta} \rho_{\alpha\beta} A_{\alpha\beta} = \text{Tr}[\rho A].$$

Если известна матрица плотности, то полученный результат можно использовать для вычисления среднего значения любой величины, даже если система не находится в состоянии равновесия. Здесь мы ограничились обсуждением лишь равновесной матрицы плотности, поскольку расчет неравновесной матрицы плотности является значительно более сложным. Мы убедимся в этом в последующих главах данной книги.

**Представление плоских волн и синус-косинус представление.** Рассмотрим теперь проводник длины  $L$ , в котором есть только два состояния в виде плоских волн (ПВ):

$$\Psi_+(x) = \frac{1}{\sqrt{L}} e^{+ikx} \quad \text{и} \quad \Psi_-(x) = \frac{1}{\sqrt{L}} e^{-ikx}.$$

Оператор тока в этом базисе имеет вид

$$[J_{op}]_{\pm} = \frac{-q}{L} \begin{pmatrix} (+) & (-) \\ \hbar k/m & 0 \\ 0 & -\hbar k/m \end{pmatrix},$$

и мы можем записать матрицу плотности в виде

$$[\rho]_{\pm} = \begin{pmatrix} (+) & (-) \\ f_+ & 0 \\ 0 & f_- \end{pmatrix},$$

где  $f_+$  и  $f_-$  есть вероятности заполнения двух данных состояний. Мы хотели бы трансформировать обе эти матрицы из базиса «±» в базис «cs», где используются функции типа синус и косинус:

$$\Psi_c = \sqrt{\frac{2}{L}} \cos kx \quad \text{и} \quad \Psi_s = \sqrt{\frac{2}{L}} \sin kx.$$

Для этого следует записать матрицу перехода  $[V]$ , чьи столбцы представляют координаты векторов старого базиса  $(+, -)$  в новом базисе  $(c, s)$ :

$$[V] = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ +i & -i \end{pmatrix},$$

так что в представлении «cs» мы получим

$$[\rho]_{cs} = [V][\rho]_{\pm}[V]^+ = \frac{1}{2} \begin{pmatrix} (c) & (s) \\ f_+ + f_- & -i(f_+ - f_-) \\ -i(f_+ - f_-) & f_+ + f_- \end{pmatrix}$$

и

$$[J_{op}]_{cs} = [V][J_{op}]_{\pm}[V]^+ = \begin{pmatrix} (c) & (s) \\ 0 & -i\hbar k/mL \\ i\hbar k/mL & 0 \end{pmatrix}.$$

Нетрудно проверить, что среднее значение тока  $\langle J \rangle = \text{Tr}[\rho J_{op}]$  является одинаковым в обоих представлениях:

$$\langle J \rangle = (-q/L)(\hbar k/m)[f_+ - f_-].$$

Этот результат следовало ожидать, поскольку след матрицы является инвариантом унитарного преобразования, т. е. остается неизменным в любом представлении. Необходимо отметить, что ток в представлении функций синус-косинус появляется благодаря *недиагональным* элементам оператора тока и матрицы плотности, а не благодаря *диагональным* элементам. Заметим, что недиагональные матричные элементы не имеют такой наглядной интуитивной физической интерпретации, как диагональные элементы. Если за перенос тока отвечают диагональные элементы, то для интерпретации результатов можно использовать квазиклассическое описание в терминах вероятностей заполнения. Но если «физический процесс» описывается недиагональными элементами, нам требуется более общий квантовый подход (я признателен А. У. Оверхаузеру за его предложение данного примера).

#### 4.4.2. Теория возмущений

Предположим, что необходимо найти уровни энергии атома водорода в присутствии внешнего электрического поля  $F$ , направленного вдоль оси  $z$ . Будем использовать в качестве базиса состояния  $1s, 2s, 2p_x, 2p_y$

и  $2p_z$  и запишем в этом базисе матрицу гамильтониана. В отсутствие поля такая матрица была бы диагональной:

$$[H_0] = \begin{pmatrix} 1s & 2s & 2p_x & 2p_y & 2p_z \\ E_1 & 0 & 0 & 0 & 0 \\ 0 & E_2 & 0 & 0 & 0 \\ 0 & 0 & E_2 & 0 & 0 \\ 0 & 0 & 0 & E_2 & 0 \\ 0 & 0 & 0 & 0 & E_2 \end{pmatrix},$$

где  $E_0 = 13.6$  эВ,  $E_1 = -E_0$  и  $E_2 = -E_0/4$ . В присутствии электрического поля появляется дополнительная матрица  $[H_F]$ , которую необходимо прибавить к гамильтониану  $[H]$ . Ее элементы определяются выражениями

$$[H_F]_{nm} = qF \int_0^\infty dr r^2 \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\phi u_n^*(\vec{r}) r \cos \theta u_m(\vec{r}).$$

Используя систему базисных функций

$$u_{1s} = \sqrt{1/\pi a_0^3} e^{-r/a_0},$$

$$u_{2s} = \sqrt{1/32\pi a_0^3} \left(2 - \frac{r}{a_0}\right) e^{-r/a_0},$$

$$u_{2p_x} = \sqrt{1/16\pi a_0^3} \left(\frac{r}{a_0}\right) e^{-r/2a_0} \sin \theta \cos \phi,$$

$$u_{2p_y} = \sqrt{1/16\pi a_0^3} \left(\frac{r}{a_0}\right) e^{-r/2a_0} \sin \theta \sin \phi,$$

$$u_{2p_z} = \sqrt{1/16\pi a_0^3} \left(\frac{r}{a_0}\right) e^{-r/2a_0} \cos \theta,$$

можно вычислить соответствующие интегралы непосредственно и показать, что матрица гамильтониана в этом базисе имеет вид

$$[H_F] = \begin{pmatrix} 1s & 2s & 2p_x & 2p_y & 2p_z \\ 0 & 0 & 0 & 0 & A \\ 0 & 0 & 0 & 0 & B \\ 0 & 0 & 0 & 0 & 0 \\ A & B & 0 & 0 & 0 \end{pmatrix},$$

где величины  $A$  и  $B$  являются линейными функциями электрического поля:

$$A = (128\sqrt{2}/243)a_0F, \quad B = -3a_0F.$$

Следовательно,

$$[H_0] + [H_F] = \begin{pmatrix} 1s & 2s & 2p_z & 2p_x & 2p_y \\ E_1 & 0 & A & 0 & 0 \\ 0 & E_2 & B & 0 & 0 \\ A & B & E_2 & 0 & 0 \\ 0 & 0 & 0 & E_2 & 0 \\ 0 & 0 & 0 & 0 & E_2 \end{pmatrix}.$$

Обратите внимание на то, что здесь по-новому обозначены столбцы и строчки матрицы. Это сделано, чтобы подчеркнуть тот факт, что элементы матрицы, отвечающие состояниям  $2p_x$  и  $2p_y$ , не подвергаются влиянию электрического поля, в то время как остальные состояния подвержены действию электрического поля  $1s$ ,  $2s$  и  $2p_z$ .

**Теория возмущений для вырожденных состояний.** В отсутствие внешнего электрического поля собственное значение  $E_1$  было невырожденным, а собственное значение  $E_2$  вырожденным четырехкратно. Как изменятся эта система уровней при включении электрического поля? Как упоминалось ранее, собственные значения в большей или меньшей степени соответствуют диагональным матричным элементам до тех пор, пока недиагональные элементы  $H_{mn}$  не становятся сравнимыми с разностью между соответствующими диагональными элементами ( $H_{mm} - H_{nn}$ ). Поэтому, как только появляются два вырожденных собственных значения (т. е.  $H_{mm} - H_{nn} = 0$ ), даже небольшой по величине недиагональный элемент  $H_{mn}$  оказывает заметное влияние на спектр. Следовательно, можно ожидать, что электрическое поле окажет существенное влияние на уровни  $2s$  и  $2p_z$ , поскольку до включения поля они были вырожденными. В этой задаче можно получить хорошее приближение, рассматривая блок матрицы  $[H]$ , соответствующий только этим двум уровням:

$$[H_0] + [H_F] = \begin{pmatrix} 2s & 2p_z \\ E_2 & B \\ B & E_2 \end{pmatrix}.$$

Легко показать, что собственные значения этой матрицы равны  $E = E_2 \pm B$ , а соответствующие собственные векторы записываются как

$$|2s\rangle - |2p_z\rangle \quad \text{и} \quad |2s\rangle + |2p_z\rangle.$$

Этот приближенный подход (известный как теория возмущений для вырожденных состояний) дает достаточно хорошие результаты (см. рис. 4.4.1), при условии, что недиагональные элементы (такие как  $A$ ), связывающие эти состояния, малы по сравнению с разностью энергий между этими уровнями (такой как величина  $E_2 - E_1$ ).

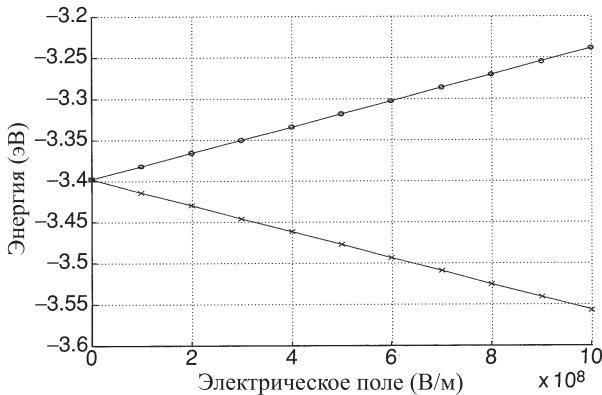


Рис. 4.4.1. Энергия уровней  $2s$  и  $2p_z$  в присутствии электрического поля  $F$ . Сплошные линии показывают результаты непосредственной диагонализации матрицы, в то время как значками  $\circ$  и  $\times$  отмечены результаты теории возмущений для  $E = E_2 \pm B$ .

**Теория возмущений для невырожденных состояний.** Как возмущение действует на уровень  $1s$ ? Поскольку в данном случае нет вырождения, эффект внешнего поля выражен много слабее, и в первом приближении можно просто игнорировать оставшуюся часть матрицы гамильтониана:

$$[H_0] + [H_F] = \begin{pmatrix} 1s \\ E_1 \end{pmatrix},$$

утверждая, тем самым, что собственное значение остается равным  $E_1$ . Мы можем добиться большей точности, если «перенормируем» матрицу следующим образом. Предположим, что мы осуществляем разбиение матрицы  $[H]$  на блоки, записав

$$[H]\{\phi\} = E\{\phi\} \rightarrow [H_0] + [H_F] = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix},$$

где  $H_{11}$  обозначает блок матрицы, который мы хотим сохранить в рассмотрении (в данном случае это блок, отвечающий состоянию  $1s$ ). Нетрудно

далее избавиться от  $\{\phi_2\}$ , чтобы получить

$$[H]' \{\phi_1\} = E \{\phi_1\},$$

где введено обозначение

$$[H]' = [H_{11}] + [H_{12}] [EI - H_{22}]^{-1} H_{21},$$

в котором  $I$  есть единичная матрица того же размера, что  $H_{22}$ . В результате такого преобразования мы получим не слишком много, если бы пришлось искать обратную матрицу  $[EI - H_{22}]^{-1}$  с ее недиагональными элементами. Но в низшем порядке теории возмущений мы можем попросту не учитывать недиагональные элементы  $H_{22}$  и записать обратную матрицу непосредственно. В данном примере это приведет к выражению

$$[H'] \approx E_1 + (0 \quad A) \begin{pmatrix} 1/(E - E_2) & 0 \\ 0 & 1/(E - E_2) \end{pmatrix} \begin{pmatrix} 0 \\ A \end{pmatrix} = E_1 + \frac{A^2}{E - E_2}.$$

В низшем порядке теории возмущений собственное значение  $E$  приблизительно равно  $E_1$ , поэтому

$$[H'] \approx E_1 + (A^2/(E - E_2)),$$

что демонстрирует квадратичный характер поправки к собственному значению энергии невырожденного состояния, в отличие линейной поправки к энергии вырожденных состояний. Такой приближенный подход, известный как теория возмущений для невырожденных состояний, достаточно хорошо описывает поправки к собственным значениям (см. рис. 4.4.2).

## Упражнения

**У.4.1.** Постройте зависимость электронной плотности в молекуле водорода  $n(x)$  от координаты, отложенной вдоль оси, соединяющей два атома водорода, предполагая, что они находятся на расстоянии, равном равновесной длине химической связи  $R = 0.074$  нм. Сравните ваши результаты с результатами на рис. 4.1.4.

**У.4.2.** Рассчитайте равновесную электронную плотность  $n(x)$  в одномерной потенциальной яме, представив волновую функцию на решетке из 100 узлов, расстояние между которыми равно 2 Å. Сравните результат с каждым из графиков на рисунках 4.3.1 *в–е*.

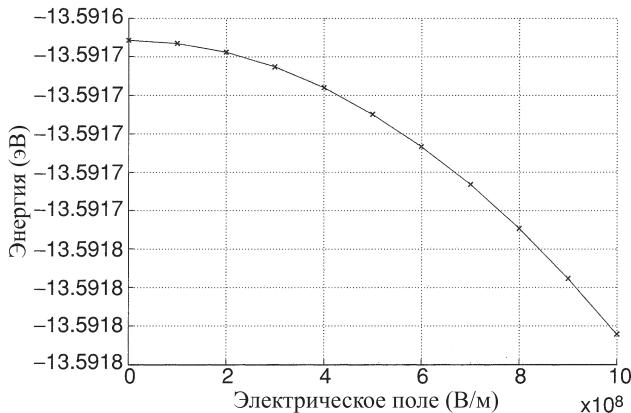


Рис. 4.4.2. Энергия уровня  $1s$  в присутствии электрического поля  $F$ . Сплошная линия показывает результаты непосредственной диагонализации матрицы, а крестиками отмечены результаты, полученные в рамках теории возмущений для энергии уровня  $E = E_1 + (A^2/(E_1 - E_2))$

**У.4.3.** Дано совокупность базисных функций, имеющих треугольный профиль и заданных на одномерной дискретной решетке с постоянной решеткой  $a$  (см. рис. У.4.3).

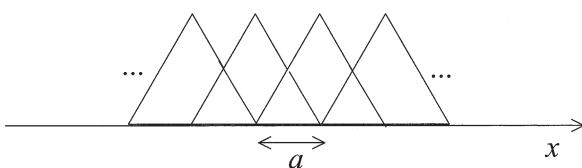


Рис. У.4.3

(а) Вычислите матричные элементы матрицы интегралов перекрытия  $[S]$  из уравнения (4.1.3a).

(б) Вычислите матричные элементы для матрицы гамильтониана  $[H]$  из уравнения (4.1.3b), предполагая, что  $H_{op} = -(\hbar^2/2m) \partial^2/\partial x^2$ .

Эта задача представляет частный случай метода конечных элементов, часто используемого в литературе. Например, можно обратиться к работам [44] и [51]. Как соотносится этот метод с методом конечных разностей, который обсуждался в параграфе 2.2?

**У.4.4.** Найдите собственные значения матрицы

$$\begin{pmatrix} E_1 & 0 & A & 0 & 0 \\ 0 & E_2 & B & 0 & 0 \\ A & B & E_2 & 0 & 0 \\ 0 & 0 & 0 & E_2 & 0 \\ 0 & 0 & 0 & 0 & E_2 \end{pmatrix},$$

где  $E_1 = -E_0$ ,  $E_2 = -E_0/4$  ( $E_0 = 13.6$  эВ) и  $A = (128\sqrt{2}/243)a_0F$ ,  $B = -3a_0F$ , и постройте их как функцию электрического поля  $F$ . Сравните результаты с результатами, полученными в рамках теории возмущений и показанными на рис. 4.4.1 и 4.4.2.

## Глава 5

# Зонная структура

В предыдущей главе мы убедились в том, что волновые функции атомных орбиталей могут быть использованы в качестве базиса при переходе к матричному представлению оператора Гамильтона. После диагонализации соответствующей матрицы находятся собственные значения энергии. В данной главе будет показано, каким образом этот подход может быть использован для расчета собственных значений энергии в бесконечной периодической структуре твердого тела. В первую очередь рассмотрим несколько «элементарных» примеров, которые демонстрируют все основные детали расчета зонной структуры с помощью матричного уравнения

$$E(\phi_0) = [h(\vec{k})](\phi_0),$$

где матрица

$$[h(\vec{k})] = \sum_m [H_{nm}] e^{i\vec{k} \cdot (\vec{d}_m - \vec{d}_n)}.$$

Матрица  $[h(\vec{k})]$  имеет размер  $(b \times b)$ , где  $b$  есть число базисных орбиталей, приходящихся на одну элементарную ячейку. Суммирование по  $m$  происходит по всем соседним ячейкам (включая исходную), с которыми  $n$ -я волновая функция перекрывается (т. е. матричный элемент  $H_{nm}$  отличен от нуля). Сумму можно вычислить, выбирая любую элементарную ячейку  $n$  за исходную. В силу периодичности кристаллической решетки результат будет одинаковым для любого  $n$ . После определения собственных значений матрицы  $[h(\vec{k})]$ , имеющей размеры  $(b \times b)$ , зонную структуру представить графически. Для этого собственные значения следует найти для каждого волнового вектора  $\vec{k}$ , так что соответствующие собственные значения определят  $b$  ветвей энергетического спектра. Это утверждение составляет главный результат, который мы вначале проиллюстрируем на элементарных примерах (параграф 5.1), а затем сформулируем его в общем виде для твердых тел с периодической структурой (параграф 5.2). Далее полученные

результаты будут использованы при обсуждении зонной структуры 3D полупроводников (параграф 5.3). В параграфе 5.4 обсуждается природа спин-орбитального взаимодействия, а также и влияние этого взаимодействия на уровни энергии в полупроводниках. Механизм спин-орбитального взаимодействия имеет релятивистскую природу и может быть описан в рамках уравнения Дирака, которое кратко обсуждается в параграфе 5.5.

## 5.1. Элементарные примеры

Рассмотрим элементарную одномерную модель твердого тела, состоящего из  $N$  атомов (см. рис. 5.1.1). Если выбрана модель с одной орбитальной на атом, то мы сможем записать гамильтоновскую матрицу ( $N \times N$ ) в виде (здесь недиагональные элементы обозначены индексом « $ss$ », хотя рассматриваемые орбитали не обязательно являются орбиталями  $s$ -типа):

$$H = \begin{pmatrix} |1\rangle & |2\rangle & \dots & |N-1\rangle & |N\rangle \\ |1\rangle & E_0 & E_{ss} & & 0 & E_{ss} \\ |2\rangle & E_{ss} & E_0 & & 0 & 0 \\ & & & \dots & & \dots \\ |N-1\rangle & 0 & 0 & & E_0 & E_{ss} \\ |N\rangle & E_{ss} & 0 & & E_{ss} & E_0 \end{pmatrix}. \quad (5.1.1)$$

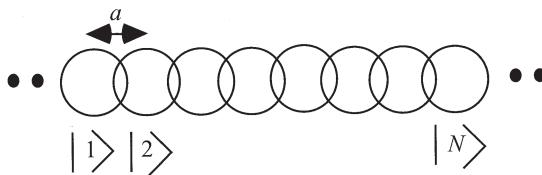


Рис. 5.1.1. Одномерная модель твердого тела

Выше при выводе матрицы мы использовали периодические граничные условия, смысл которых заключается в том, что  $N$ -й атом отождествляется с первым, образуя кольцо. В такой модели появляются ненулевые матричные элементы  $H_{1,N}$  и  $H_{N,1}$ , которые должны обратиться в нуль, если твердое тело резко ограничено. Периодические граничные условия не всегда являются полностью реалистичными, но, поскольку мы обсуждаем объемные свойства массивных твердых тел, выполнение точных граничных условий на их поверхности не имеет большого значения. В такой ситуации

можно выбирать любые граничные условия (в том числе периодические), которые облегчают математические расчеты.

Чему же равны собственные значения матрицы  $[H]$  в выражении (5.1.1)? Эта матрица фактически соответствует матрице, которую мы обсуждали в главе 2 в связи с методом конечных разностей. В рассматриваемой модели собственные значения могут быть записаны в виде (индекс  $\alpha$  есть целое число)

$$E_\alpha = E_0 + 2E_{ss} \cos(k_\alpha a), \quad \text{где } k_\alpha a = \alpha 2\pi/N. \quad (5.1.2)$$

Значения  $k_\alpha a$  определены в интервале от  $-\pi$  до  $\pi$ , где они принимают дискретные значения, разделенные шагом  $2\pi/N$ , как это показано на рис. 5.1.2 а. Если число  $N$  велико, то собственные значения расположены близко друг к другу, а при малых  $N$  они отстоят сравнительно далеко друг от друга (рис. 5.1.2 б).

В чем причина того, что нам так легко удалось найти собственные значения этой матрицы? Причина в ее периодической структуре. Именно поэтому матричное уравнение  $E(\psi) = [H](\psi)$  — это система  $N$  одинаковых по форме уравнений, которые могут быть записаны в форме ( $n = 1, 2, \dots, N$ )

$$E\psi_n = E_0\psi_n + E_{ss}\psi_{n-1} + E_{ss}\psi_{n+1}. \quad (5.1.3)$$

Эта система уравнений может быть решена аналитически путем подстановки

$$\psi_n = \psi_0 e^{ikna}. \quad (5.1.4)$$

Подставляя выражение (5.1.4) в (5.1.3) и сокращая общий множитель  $\exp(ikna)$ , мы получаем, что

$$E\psi_0 = E_0\psi_0 + E_{ss}e^{-ika}\psi_0 + E_{ss}e^{ika}\psi_0,$$

откуда следует, что

$$E = E_0 + 2E_{ss} \cos(ka).$$

Следовательно, решение в форме (5.1.4) удовлетворяет системе уравнений при любом значении  $k$ . Возникает вопрос о том, что будет ограничивать полное число собственных значений (как это и должно быть для матрицы конечных размеров)?

Наличие лишь конечного числа собственных значений является следствием двух факторов. Во-первых, периодические граничные условия требуют, чтобы волновая функция имела период  $Na$ , а из-за конечных размеров решетки разрешенные значения  $k$  могут иметь только дискретные значения  $k_\alpha = \alpha 2\pi/N$  (см. уравнение (5.1.2)). Во-вторых, значения  $ka$ ,

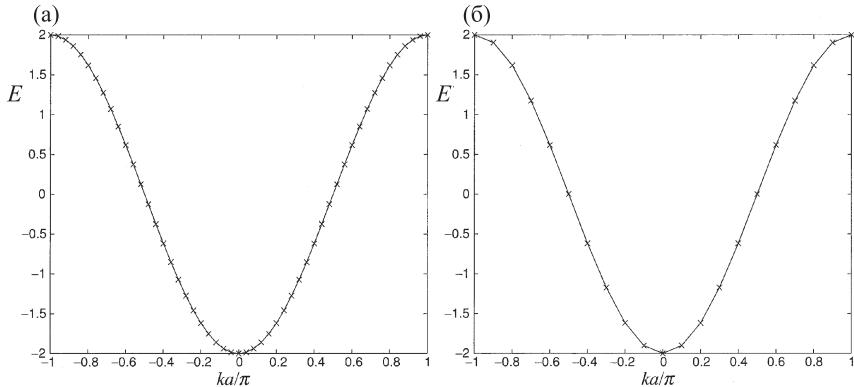


Рис. 5.1.2. Сплошные линии показывают зависимость энергии  $E$  от  $ka/\pi$ , найденную из уравнения (5.1.2) с параметрами  $E_0 = 0$  и  $E_{ss} = -1$ . Крестиками обозначены собственные значения матрицы из выражения (5.1.1), где (а)  $N = 100$  и (б)  $N = 20$

отличающиеся на  $2\pi$ , описывают одно и то же состояние на дискретной решетке, а волновые функции вида

$$\exp(ik_\alpha x) \quad \text{и} \quad \exp(i[k_\alpha + (2\pi/a)]x)$$

принадлежат одному и тому же состоянию, поскольку для любого узла решетки  $x_n = na$  выполняется равенство

$$\exp(ik_\alpha x_n) = \exp(i[k_\alpha + (2\pi/a)]x_n).$$

Данные волновые функции НЕ являются эквивалентными в интервалах между узлами решетки и потому описывают различные состояния в модели непрерывной решетки. Но на дискретной решетке значения  $k_\alpha$ , отличающиеся на  $2\pi/a$ , будут отвечать тождественным состояниям, и лишь значения  $k_\alpha a$ , лежащие в интервале длины  $2\pi$ , будут описывать независимые решения. В принципе, можно рассматривать любой интервал длины  $2\pi$ , но обычно ограничивают разрешенные значения  $k_\alpha a$  областью, называемой первой зоной Бриллюэна:

$$-\pi \leq ka \leq +\pi \quad \text{для периодических граничных условий.} \quad (5.1.5)$$

Интересно отметить, что конечные размеры решетки  $Na$  приводят к дискретному характеру (в единицах  $2\pi/Na$ ) разрешенных значений  $k$ , в то время как дискретность решетки с шагом  $a$  приводит к конечной величине интервала изменения  $k$ , равной  $2\pi/a$ . Число разрешенных значений  $k$ ,

таким образом, составляет

$$(2\pi/a)/(2\pi/Na) = N,$$

что точно равно числу узлов в кристаллической решетке. По этой причине число собственных значений, равное числу разрешенных значений  $k$ , и порядок матрицы  $[H]$ , определяемый числом узлов решетки, совпадают.

**Когда энергия в зоне уменьшается с ростом  $k$ ?** При расчете зависимости энергии от  $k$  (см. рис. 5.1.2) предполагалось что величина  $E_{ss}$  отрицательна, что имело бы место, скажем, при использовании уравнения (4.1.11) для ее расчета (обратите внимание на то, что потенциалы  $U_L$  и  $U_R$  отрицательны), а атомные орбитали являлись бы орбиталями  $s$ -типа. Но если зонные состояния сформированы из атомных орбиталей  $p_x$ -типа, как показано на рисунке 5.1.3 б, то интеграл перекрытия  $E_{ss}$  был бы положительным, и, соответственно, график  $E(k)$  имел бы вид убывающей функции, как показано на рисунке. Подобная ситуация реализуется в валентной зоне обычных полупроводников, формируемой преимущественно из орбиталей  $p$ -типа.

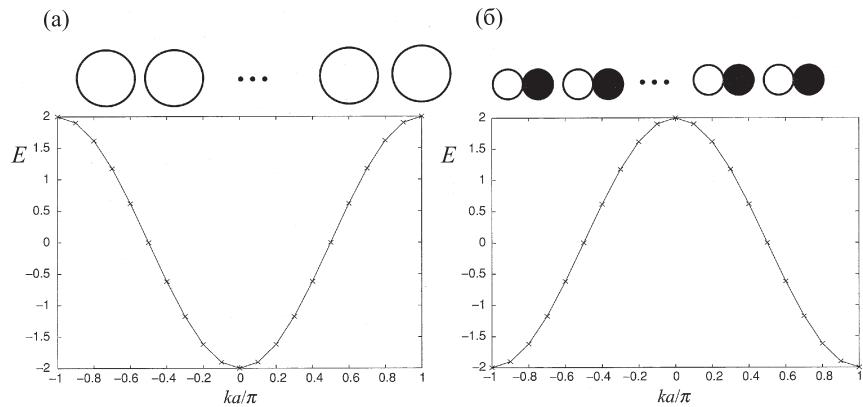


Рис. 5.1.3. То же, что на рис. 5.1.2 б с  $E_0 = 0$  и (а)  $E_{ss} = -1$  и (б)  $E_{ss} = +1$

**Решетка с базисом.** Рассмотрим далее одномерную модель твердо-го тела, чья элементарная ячейка состоит из двух атомов, как показано на рис. 5.1.4. Заметим, что реальные одномерные структуры, подобные изображенной на рис. 5.1.1, имеют тенденцию к спонтанной деформации, при которой происходит удвоение периода, формирование структуры, показанной

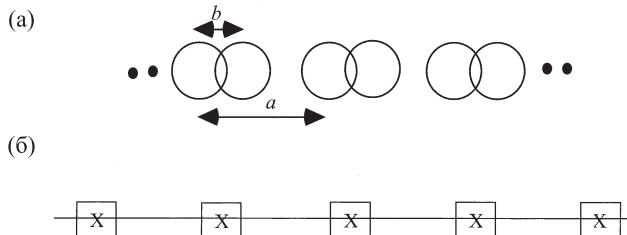


Рис. 5.1.4. (а) Одномерная модель твердого тела, в котором в каждой элементарной ячейке находятся два атома. (б) Базисная решетка, определяющая периодическую структуру твердого тела

на рис. 5.1.4.<sup>1</sup> Энергетические соображения, приводящие к появлению этого структурного перехода, здесь мы обсуждать не будем. Наша цель состоит только в том, чтобы продемонстрировать схему расчета зонной структуры твердого тела, чья элементарная ячейка содержит более чем одну базисную орбиталь. Предположив, что на один атом приходится одна орбиталь, мы можем записать матричное представление гамильтониана  $[H]$  в виде

$$H = \begin{array}{ccccccc} & |1_A\rangle & |1_B\rangle & |2_A\rangle & |2_B\rangle & |3_A\rangle & |3_B\rangle & \dots \\ |1_A\rangle & E_0 & E_{ss} & 0 & 0 & 0 & 0 & \dots \\ |1_B\rangle & E_{ss} & E_0 & E'_{ss} & 0 & 0 & 0 & \dots \\ |2_A\rangle & 0 & E'_{ss} & E_0 & E_{ss} & 0 & 0 & \dots \\ |2_B\rangle & 0 & 0 & E_{ss} & E_0 & E'_{ss} & 0 & \dots \\ |3_A\rangle & 0 & 0 & 0 & E'_{ss} & E_0 & E_{ss} & \dots \\ |3_B\rangle & 0 & 0 & 0 & 0 & E_{ss} & E_0 & \dots \end{array} \quad (5.1.6)$$

В отличие от матрицы, фигурирующей в выражении (5.1.1), здесь присутствуют два различных интеграла перекрытия  $E_{ss}$  и  $E'_{ss}$ , как следствие, и в данной ситуации подстановка (5.1.4) не может быть использована напрямую. Тем не менее, можно сгруппировать элементы матрицы в блоки  $(2 \times 2)$  и записать гамильтониан в виде

$$H = \begin{array}{ccccc} & |1\rangle & |2\rangle & |3\rangle & \dots \\ |1\rangle & H_{11} & H_{12} & 0 & \dots \\ |2\rangle & H_{21} & H_{22} & H_{23} & \dots \\ |3\rangle & 0 & H_{32} & H_{33} & \dots \end{array} \quad (5.1.7)$$

<sup>1</sup>Пайерлсовское удвоение периода. — Прим. ред. перев.

где

$$H_{nn} = \begin{pmatrix} E_0 & E_{ss} \\ E_{ss} & E_0 \end{pmatrix}, \quad H_{n,n+1} = \begin{pmatrix} 0 & 0 \\ E'_{ss} & 0 \end{pmatrix}, \quad H_{n,n-1} = \begin{pmatrix} 0 & E'_{ss} \\ 0 & 0 \end{pmatrix}.$$

Матрица в уравнении (5.1.7) теперь является периодической, и поэтому можно записать матричное уравнение  $E(\psi) = [H](\psi)$  в виде

$$E\phi_n = H_{nn}\phi_n + H_{n,n-1}\phi_{n-1} + H_{n,n+1}\phi_{n+1}, \quad (5.1.8)$$

где  $\phi_n$  представляет собой вектор-столбец  $(2 \times 1)$ , а элемент  $H_{nm}$  является матрицей  $(2 \times 2)$ .

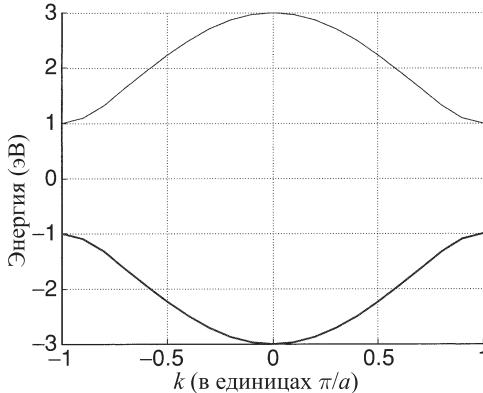


Рис. 5.1.5. Зонная структура «двуахомного» одномерного твердого тела, изображенного на рис. 5.1.4, построенная в соответствии с уравнением (5.1.10) с параметрами  $E_0 = 0$ ,  $E_{ss} = 2$ ,  $E'_{ss} = 1$

Данная система уравнений может быть решена при помощи подстановки

$$\phi_n = \phi_n e^{ikna}. \quad (5.1.9)$$

Подставляя (5.1.9) в систему (5.1.8) и сокращая общий множитель  $\exp[ikna]$ , придем к уравнению

$$E\phi_0 = H_{nn}\phi_0 + H_{n,n-1}e^{-ika}\phi_0 + H_{n,n+1}e^{ika}\phi_0,$$

что приводит к следующему уравнению для двухкомпонентного столбца  $\phi_0$ :

$$E\phi_0 = \begin{pmatrix} E_0 - E & E_{ss} + E'_{ss}e^{-ika} \\ E_{ss} + E'_{ss}e^{ika} & E_0 - E \end{pmatrix} = 0,$$

откуда следует, что

$$E = E_0 \pm (E_{ss}^2 + E_{ss}'^2 + 2E_{ss}E_{ss}' \cos(ka))^{1/2}. \quad (5.1.10)$$

Выражение (5.1.9) описывает дисперсионную зависимость  $E(k)$  или спектр, состоящий из двух ветвей, показанных на рис. 5.1.5.

## 5.2. Основной результат

Описанная выше процедура непосредственно обобщается на общий случай периодической решетки, в которой элементарная ячейка содержит произвольное число базисных функций. Пусть  $n$ -я элементарная ячейка (рис. 5.2.1) взаимодействует с  $m$  соседними ячейками и это взаимодействие описывается матрицей  $[H_{nm}]$ , имеющей размер  $(b \times b)$ , где  $b$  есть число базисных функций в одной элементарной ячейке. Полное матричное уравнение примет вид

$$\sum_m [H_{nm}] \{\phi_m\} = E \{\phi_n\}, \quad (5.2.1)$$

где  $\{\phi_m\}$  есть вектор-столбец  $(b \times 1)$ , обозначающий волновую функцию в  $m$ -й элементарной ячейке.

Для дальнейшего важно, что решение системы уравнений можно найти при помощи подстановки

$$\{\phi_m\} = \{\phi_0\} \exp[i\vec{k} \cdot \vec{d}_m], \quad (5.2.2)$$

благодаря которой уравнение (5.2.1) выглядит одинаково в любой  $n$ -й элементарной ячейке. Это свойство можно назвать дискретным аналогом теоремы Блоха, которая обычно обсуждается в стандартных учебниках, таких, как, например, курс Ашкрофта и Мермина [2]. После подстановки выражения (5.2.2) в (5.2.1) получаем

$$E \{\phi_0\} = [h(\vec{k})] \{\phi_0\}, \quad (5.2.3)$$

где

$$[h(\vec{k})] = \sum_m [H_{nm}] \exp[i\vec{k} \cdot (\vec{d}_m - \vec{d}_n)] \quad (5.2.4)$$

независимо от того, какая именно  $n$ -я элементарная ячейка выбрана для вычисления суммы (5.2.4). Этот вывод составляет основу всех методов расчета зонной структуры твердых тел. Суммирование по  $m$  в уравнении (5.2.4)

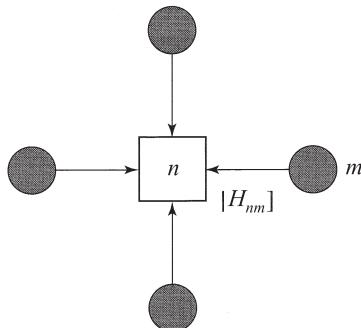


Рис. 5.2.1. Схематическое изображение  $n$ -й элементарной ячейки, взаимодействие которой с соседними  $m$ -ми элементарными ячейками описывается матрицей  $[H_{nm}]$  размера  $(b \times b)$ , где  $b$  есть число базисных функций в одной элементарной ячейке. Расположение ближайших соседей будет отличаться от одного образца твердого тела к другому, но в строго периодическом твердом теле конфигурация будет идентичной независимо от того, какое  $n$  выбрать

включает все соседние элементарные ячейки (в том числе исходную), с которой волновая функция в  $n$ -й ячейке перекрывается, т. е. матричный элемент  $H_{nm}$  отличен от нуля. Размер матрицы  $[h(\vec{k})]$  равен  $(b \times b)$ , где  $b$  есть число базисных функций в элементарной ячейке. Зонная структура может быть определена путем расчета собственных значений матрицы  $[h(\vec{k})]$  размера  $(b \times b)$  для каждого значения  $\vec{k}$ . Она будет состоять из  $b$  ветвей — по одной, отвечающих каждому собственному значению.

**Разрешенные значения  $k$ .** Рассматривая одномерную задачу, мы выяснили, почему  $k$  принимает лишь конечное число разрешенных значений, равное числу элементарных ячеек в твердом теле. Напомним этот основной результат: конечный размер решетки (равный  $Na$ ) приводит к дискретности (в единицах  $2\pi/Na$ ) разрешенных значений  $k$ , в то время как дискретность самой решетки (с шагом  $a$ ) приводит к конечной величине интервала изменения  $k$  (равному  $2\pi/a$ ). Как нам обобщить этот результат на случай более одного числа измерений?

Подобное обобщение выполняется непосредственно, если твердое тело формирует квадратную (или кубическую) решетку, как это показано на рис. 5.2.2. В двумерном случае разрешенные значения  $k$  определяются как

$$[\vec{k}]_{mn} = \vec{x}(m 2\pi/Ma) + \vec{y}(n 2\pi/Nb), \quad (5.2.5)$$

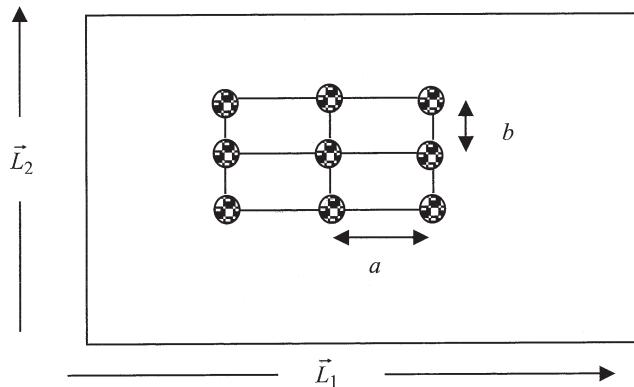


Рис. 5.2.2. Конечная двумерная квадратная решетка с  $M$  элементарными ячейками вдоль направления  $x$  и  $N$  элементарными ячейками вдоль направления  $y$  ( $\vec{L}_1 = \vec{x}Ma$ ,  $\vec{L}_2 = \vec{y}Nb$ )

где  $(m, n)$  есть пара целых чисел, в то время как  $M, N$  представляют собой число элементарных ячеек вдоль направлений  $x$  и  $y$  соответственно. Подобная запись выглядит как разумное обобщение одномерного результата (сравните с выражением (5.1.2):  $k_\alpha = \alpha(2\pi/Na)$ ). Формально мы могли бы вывести уравнение (5.2.5), записав (с учетом того, что  $\vec{L}_1 = \vec{x}Ma$ ,  $\vec{L}_2 = \vec{y}Nb$ )

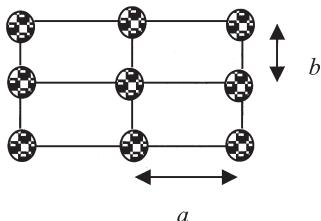
$$\vec{k} \cdot \vec{L}_1 = m 2\pi \rightarrow k_x = m 2\pi/Ma,$$

$$\vec{k} \cdot \vec{L}_2 = n 2\pi \rightarrow k_y = n 2\pi/Nb.$$

**Зона Бриллюэна.** Процедура построения зоны Бриллюэна начинается с построения обратной решетки (см. рис. 5.2.3 б) в  $k$ -пространстве, что может рассматриваться как преобразование Фурье кристаллической решетки, определенной в обычном пространстве. В одномерном случае, как известно, последовательность импульсов с шагом дискретизации  $a$  (см. рис. 5.2.4 а) обладает Фурье-образом, состоящим из набора импульсов, отстоящих друг от друга на расстоянии  $2\pi/a$  (рис. 5.2.4 б). Далее мы можем построить первую зону Бриллюэна с центром в точке  $k = 0$  путем соединения этой точки с соседними узлами обратной решетки и построения перпендикуляров, проходящих через середины отрезков, соединяющих точку  $\vec{k} = (0, 0)$  с соседними узлами обратной решетки.

(а)

Решетка в координатном пространстве



(б)

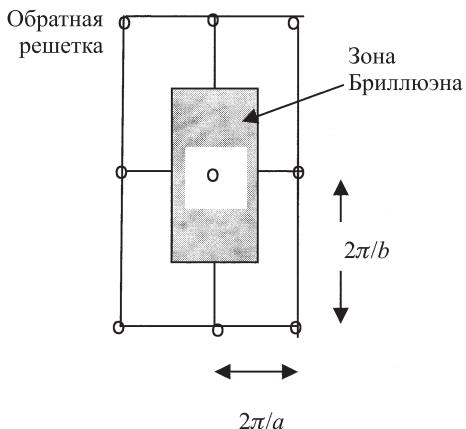


Рис. 5.2.3. (а) Квадратная решетка в координатном пространстве. (б) Соответствующая обратная решетка

Зона Бриллюэна, построенная с помощью данной процедуры, определяет область допустимых значений  $\vec{k}$ :

$$-\pi \leq k_x a \leq +\pi \quad \text{и} \quad -\pi \leq k_y b \leq +\pi, \quad (5.2.6)$$

что согласуется с результатом эвристического обобщения выражения (5.1.5).

**Обратная решетка.** Если кристаллическая решетка в координатном пространстве не является квадратной или кубической, то для построения обратной решетки необходимо разработать более общую процедуру, нежели использованная выше. Отметим в первую очередь, что любой узел кристаллической решетки в трехмерном пространстве задается набором трех целых чисел  $(m, n, p)$ , таких, что

$$\vec{R} = m\vec{a}_1 + n\vec{a}_2 + p\vec{a}_3, \quad (5.2.7a)$$

где векторы  $\vec{a}_1$ ,  $\vec{a}_2$  и  $\vec{a}_3$  называются базисными векторами кристаллической решетки. Узлы же обратной решетки могут быть заданы как

$$\vec{K} = M\vec{A}_1 + N\vec{A}_2 + P\vec{A}_3, \quad (5.2.7b)$$

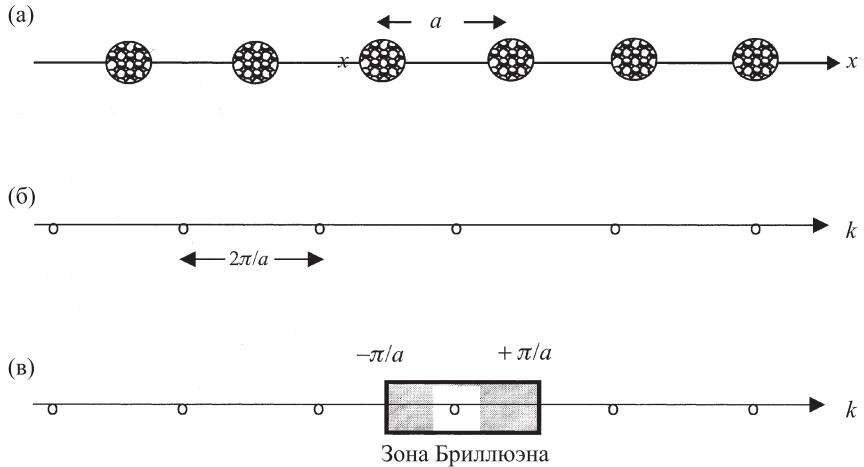


Рис. 5.2.4. (а) Решетка в координатном пространстве. (б) Обратная решетка. (в) Зона Бриллюэна

где  $(M, N, P)$  есть целые числа, а векторы  $\vec{A}_1, \vec{A}_2$  и  $\vec{A}_3$  определены таким образом, чтобы удовлетворять равенствам

$$\vec{A}_j \cdot \vec{a}_i = 2\pi\delta_{ij}, \quad (5.2.8)$$

где  $\delta_{ij}$  есть символ Кронекера (равный единице при  $i = j$  и нулю при  $i \neq j$ ). Равенство (5.2.8) будет выполнено, если записать

$$\vec{A}_1 = \frac{2\pi(\vec{a}_2 \times \vec{a}_3)}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{a}_3)}, \quad \vec{A}_2 = \frac{2\pi(\vec{a}_3 \times \vec{a}_1)}{\vec{a}_2 \cdot (\vec{a}_3 \times \vec{a}_1)}, \quad \vec{A}_3 = \frac{2\pi(\vec{a}_1 \times \vec{a}_2)}{\vec{a}_3 \cdot (\vec{a}_1 \times \vec{a}_2)}. \quad (5.2.9)$$

Нетрудно заметить, что эта формальная процедура приводит к обратной решетке, показанной на рис. 5.2.3 б, если в качестве базисных векторов будут приняты векторы  $\vec{a}_1 = \vec{x}a, \vec{a}_2 = \vec{y}b, \vec{a}_3 = \vec{z}c$ . В этом случае из уравнения (5.2.9) найдем, что

$$\vec{A}_1 = \vec{x}(2\pi/a), \quad \vec{A}_2 = \vec{y}(2\pi/b), \quad \vec{A}_3 = \vec{z}(2\pi/c).$$

Теперь, используя уравнения (5.2.7), можно построить обратную решетку, показанную на рис. 5.2.3 б. Разумеется, в данном случае использование формальной процедуры не является обязательным. Преимущества этого формального подхода можно оценить, работая с неквадратными решетками, в чем мы убедимся ниже, рассматривая одну двумерную решетку.

**Пример двумерной решетки.** Атомы углерода, образующие монослой графита (часто называемого слоем графена), располагаются в узлах гексагональной решетки, показанной на рис. 5.2.5 а. Можно видеть, что данная структура не является строго периодической. В самом деле, соседние атомы углерода находятся не в идентичном окружении. Но если объединить два атома в одну ячейку, то решетка, составленная из таких ячеек, будет периодической, поскольку каждый ее узел будет иметь одинаковое окружение (см. рис. 5.2.5 б).

Положение каждого узла данной периодической решетки может быть описано набором целых чисел  $(m, n, p)$ , где

$$\vec{R} = m\vec{a}_1 + n\vec{a}_2 + p\vec{a}_3, \quad (5.2.10)$$

с базисными векторами

$$\vec{a}_1 = \vec{x}a + \vec{y}b, \quad \vec{a}_2 = \vec{x}a - \vec{y}b, \quad \vec{a}_3 = \vec{z}c,$$

где

$$a \equiv 3a_0/2 \quad \text{и} \quad b \equiv \sqrt{2}a_0/2.$$

Здесь  $c$  есть высота элементарной ячейки в направлении оси  $z$ , которая не играет заметной роли в последующем рассмотрении. Нас интересует распределение электронных состояний в плоскости ( $xy$ ). Плоскости, имеющие различную координату  $z$ , считаются изолированными друг от друга, что не слишком далеко от истинной ситуации в реальном графите. Точки обратной решетки на плоскости  $(k_x, k_y)$  задаются вектором

$$\vec{K} = M\vec{A}_1 + N\vec{A}_2, \quad (5.2.11)$$

где числа  $(M, N)$  являются целыми, а векторы  $\vec{A}_1, \vec{A}_2$  определяются выражениями (5.2.9):

$$\vec{A}_1 = \frac{2\pi(\vec{a}_2 \times \vec{z})}{\vec{a}_1 \cdot (\vec{a}_2 \times \vec{z})} = \vec{x}\frac{\pi}{a} + \vec{y}\frac{\pi}{b},$$

$$\vec{A}_2 = \frac{2\pi(\vec{z} \times \vec{a}_1)}{\vec{a}_2 \cdot (\vec{z} \times \vec{a}_1)} = \vec{x}\frac{\pi}{a} - \vec{y}\frac{\pi}{b}.$$

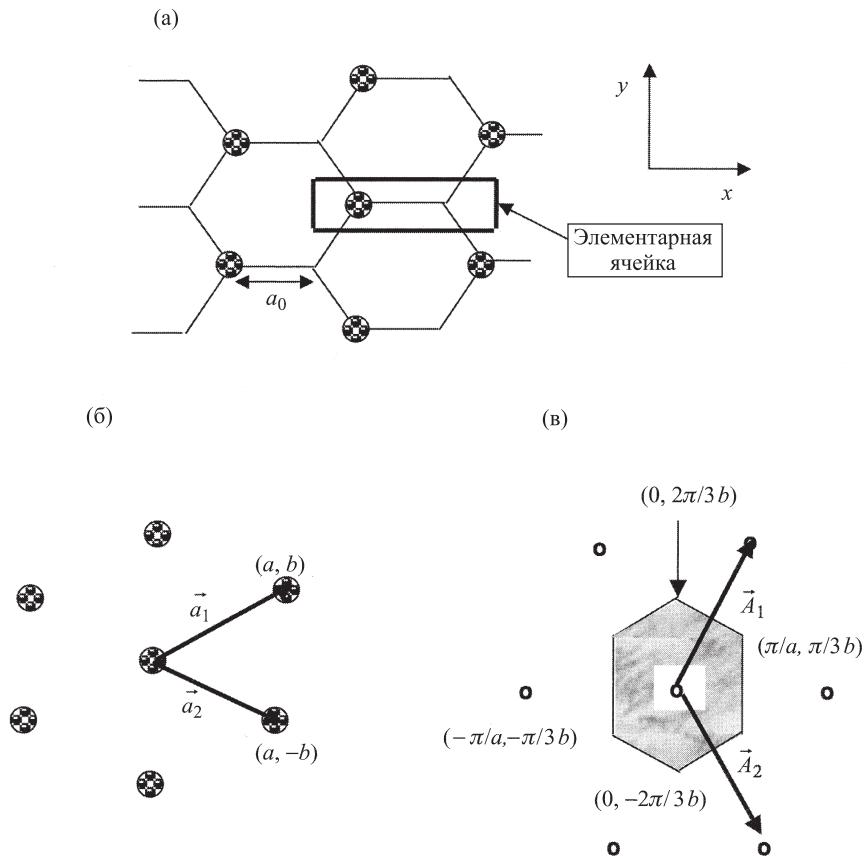


Рис. 5.2.5. (а) Решетка из атомов углерода на поверхности графита имеет элементарную ячейку, состоящую из двух атомов. (б) Решетка в координатном пространстве с периодическим расположением элементарных ячеек и базисными векторами  $\vec{a}_1$  и  $\vec{a}_2$ . (в) Обратная решетка с базисными векторами  $\vec{A}_1$  и  $\vec{A}_2$ , определяемая из соотношений  $\vec{A}_1 \cdot \vec{a}_1 = \vec{A}_2 \cdot \vec{a}_2 = 2\pi$  и  $\vec{A}_1 \cdot \vec{a}_2 = \vec{A}_2 \cdot \vec{a}_1 = 0$ . Показана также зона Бриллюэна (заштрихована), полученная построением перпендикуляров, проходящих через середины отрезков, соединяющих начало координат  $(0, 0)$  с соседними узлами обратной решетки

Используя эти базисные векторы, можно построить обратную решетку, которая показана на рис. 5.2.5 в. Для того чтобы построить зону Бриллюэна, определяющую область разрешенных значений вектора  $\mathbf{k}$ , следует провести перпендикуляры, проходящие через середины отрезков, соединяющих начало координат  $(0, 0)$  с соседними узлами обратной решетки. Область, ограниченная этими перпендикулярами, и есть зона Бриллюэна.

Зона Бриллюэна определяет область изменения  $k$ , в то время как фактические дискретные значения  $k$  являются следствием конечных размеров образца, как это отмечалось при обсуждении выражения (5.2.5). Возникает вопрос: каким образом при данном значении  $k$  находятся соответствующие собственные значения энергии? Ответ ясен: их можно найти, используя выражения (5.2.3) и (5.2.4). Размер матрицы  $[h(\vec{k})]$  зависит от числа базисных функций, приходящихся на одну элементарную ячейку. Если в качестве базисных функций использовать систему четырех валентных орбиталей углерода ( $2s, 2p_x, 2p_y, 2p_z$ ), то будет всего  $4 \times 2 = 8$  базисных функций на одну элементарную ячейку, поскольку она содержит два атома углерода. Тогда для каждого  $k$  будем иметь восемь собственных значений.

Известно, однако, что в графене уровни энергии, относящиеся к  $2s$ -,  $2p_x$ - и  $2p_y$ -орбиталям, значительно отделены по шкале энергий от уровней, относящихся к  $2p_z$ -орбиталям. Другими словами, нет отличных от нуля матричных элементов, связывающих эти два базисных подпространства. Больше того, уровни, относящиеся к  $2s$ -,  $2p_x$ - и  $2p_y$ -орбиталям, расположены либо гораздо выше, либо гораздо ниже уровня Ферми. Поэтому состояния валентной зоны или зоны проводимости вблизи уровня Ферми, которые и отвечают за электрическую проводимость, формируются, прежде всего, из  $2p_z$ -орбиталей.

Это обстоятельство означает, что состояния валентной зоны и зоны проводимости могут быть удовлетворительно описаны в рамках теории, использующей лишь одну атомную волновую функцию (орбital  $2p_z$ ) для одного атома углерода, что приводит к появлению матрицы  $[h(\vec{k})]$ , имеющей размер  $(2 \times 2)$ . Такая матрица находится суммированием по любой элементарной ячейке и четырем ближайшим соседям. Значение матричного элемента принимается равным  $-t$  для соседних атомов углерода и нулю для всех других атомов:

$$\begin{aligned} [h(\vec{k})] = & \begin{pmatrix} 0 & -t \\ -t & 0 \end{pmatrix} + \begin{pmatrix} 0 & -t \exp(i\vec{k} \times \vec{a}_1) \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -t \exp(i\vec{k} \times \vec{a}_2) \\ 0 & 0 \end{pmatrix} + \\ & + \begin{pmatrix} 0 & 0 \\ -t \exp(-i\vec{k} \times \vec{a}_1) & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -t \exp(i\vec{k} \times \vec{a}_2) & 0 \end{pmatrix}. \end{aligned}$$

Вводя обозначение

$$h_0 \equiv -t(1 + e^{i\vec{k} \times \vec{a}_1} + e^{i\vec{k} \times \vec{a}_2}) = -t(1 + 2e^{i\vec{k}_x a} \cos k_y b),$$

запишем данную матрицу в виде

$$[h(\vec{k})] = \begin{pmatrix} 0 & h_0 \\ h_0^* & 0 \end{pmatrix}.$$

Собственные значения этой матрицы будут определяться выражением

$$E = \pm |h_0| = \pm t \sqrt{1 + 4 \cos k_y b \cos k_x a + 4 \cos^2 k_y b}.$$

Обратите внимание на то, что каждому значению  $\vec{k}$  отвечают два собственных значения (одно положительное и одно отрицательное) и, соответственно, две ветви спектра  $E(\vec{k})$  (см. рис. 5.1.5). Этого следовало ожидать, поскольку в каждой элементарной ячейке учитывались две базисные функции. Физический смысл дисперсионного соотношения  $E(\vec{k})$  обсуждается в следующей главе в связи с расчетом энергетического спектра углеродных нанотрубок, которые можно рассматривать как слои графита, свернутые в цилиндры. В данный момент нашей задачей является обсуждение деталей расчета двумерной зонной структуры. В связи с этими расчетами необходимо обратить внимание на некоторые нетривиальные детали, не рассматривавшиеся в предыдущем параграфе.

### 5.3. Обычные полупроводники

В этом параграфе я буду близко следовать работе [50]. Все типичные полупроводники (и в том числе арсенид галлия) имеют решетку типа алмаза, элементарная ячейка которой состоит из двух атомов: катиона (галлий) и аниона (мышьяк). Для моноэлементных полупроводников типа кремния оба узла заняты атомами одного элемента. При проведении расчета необходимо учитывать, по крайней мере, четыре валентных орбитали, подобные орбиталям  $3s$ ,  $3p_x$ ,  $3p_y$  и  $3p_z$  атома кремния. Кроме этого, обычно учитывается еще одна орбиталь ( $4s$ -орбиталь в кремнии), что отражается в названии модели расчета ( $sp^3 s^*$ -модель). В ней, таким образом, имеется

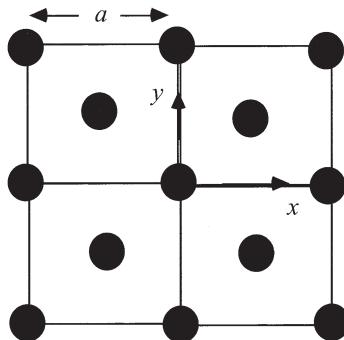


Рис. 5.3.1. ( $xy$ ) проекция гранецентрированной кубической (ГЦК) решетки, демонстрирующая расположение атомов

пять атомных орбиталей, образующих базис, состоящий из десяти орбиталей в элементарной ячейке. Следовательно, матрицы  $[h(\vec{k})]$  и  $[H_{nm}]$  в выражении (5.2.4) имеют размер  $(10 \times 10)$ . Перед выполнением суммирования в формуле (5.2.4) необходимо понять, как расположены в пространстве ближайшие соседи.

Кристаллическая решетка алмаза состоит из двух взаимопроникающих гранецентрированных кубических (ГЦК) решеток. Если в решетке GaAs, атомы галлия располагаются в узлах одной ГЦК решетки, а атомы мышьяка занимают узлы другой ГЦК решетки, сдвинутой относительно первой на четверть длины объемной диагонали, т. е. координаты этой решетки могут быть получены путем прибавления вектора  $(\vec{x} + \vec{y} + \vec{z})a/4$  к координатам в первой решетке. Если атом галлия располагается в начале координат  $(0, 0, 0)a/4$ , то в точке, отстоящей на вектор  $(\vec{x} + \vec{y} + \vec{z})a/4$ , будет располагаться атом мышьяка, который для атома галлия будет одним из ближайших соседей. Этот атом галлия будет иметь еще три ближайших соседа — атомов мышьяка. Чтобы убедиться в этом, надо рассмотреть положение в решетке ближайших атомов галлия, находящихся на грани ( $xy$ ), как показано на рис. 5.3.1. Координаты этих атомов есть  $(\vec{x} + \vec{y})a/2$ ,  $(-\vec{x} + \vec{y})a/2$ ,  $(\vec{x} - \vec{y})a/2$  и  $(-\vec{x} - \vec{y})a/2$ .

Координаты соответствующих атомов мышьяка получаются из указанных выше добавлением вектора  $(\vec{x} + \vec{y} + \vec{z})a/4$ :

$$(3\vec{x} + 3\vec{y} + \vec{z})a/4, \quad (3\vec{x} - \vec{y} + \vec{z})a/4, \quad (-\vec{x} + 3\vec{y} + \vec{z})a/4, \quad (-\vec{x} - \vec{y} + \vec{z})a/4.$$

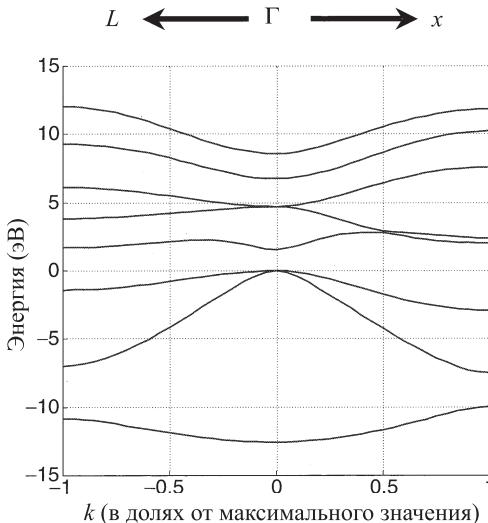


Рис. 5.3.2. Дисперсионные кривые  $E(\vec{k})$ , построенные по результатам расчета собственных значений матрицы (5.3.3) для каждого значения  $\vec{k}$  вдоль направления  $\Gamma - X$  (из точки  $\vec{k} = 0$  в точку  $\vec{k} = (\vec{x} + \vec{y} + \vec{z})\pi/a$ ) и вдоль направления  $\Gamma - L$  (из точки  $\vec{k} = 0$  в точку  $\vec{k} = (\vec{x} + 2\vec{y})\pi/a$ ). Первая зависимость построена для положительных значений на оси  $k$ , а вторая — для отрицательных значений. На графике видны только восемь, а не десять различных дисперсионных кривых, поскольку некоторые из них являются вырожденными

В решетке алмаза каждый атом имеет четыре ближайших соседа противоположного типа (катион либо анион), расположенных в вершинах тетраэдра.

Чтобы понять, как следует выполнить суммирование в выражении (5.2.5), учтем сначала только  $s$ -орбитали каждого атома. Матрицы  $[h(\vec{k})]$  и  $[H_{nm}]$  в выражении (5.2.4) в этом случае будут матрицами размера  $(2 \times 2)$ , и матрицу оператора Гамильтона  $[H_{nm}]$  можно будет записать в виде

$$\begin{array}{ccc|c} & |s_a\rangle & |s_c\rangle & \\ |s_a\rangle & E_{sa} & E_{ss} & \\ |s_c\rangle & E_{ss} & E_{sc} & \end{array} \quad (5.3.1a)$$

где  $E_{sa}$  и  $E_{sc}$  есть энергии  $s$ -орбиталей аниона и катиона соответственно,

в то время как  $E_{ss}$  обозначает интеграл перекрытия между  $s$ -орбиталью аниона и  $s$ -орбиталью катиона. Волновая функция аниона в  $n$ -й элементарной ячейке перекрывается с волновой функцией катиона в той же ячейке, а также с волновыми функциями катионов в трех других  $m$ -х ячейках, для которых выполнены соотношения

$$\vec{d}_m - \vec{d}_n =: (-\vec{x} - \vec{y})a/2, \quad (-\vec{y} - \vec{z})a/2, \quad (-\vec{z} - \vec{x})a/2.$$

Каждое такое перекрытие или связь дает вклад в матрицу  $[H_{nm}]$  в виде блока

$$\begin{pmatrix} |m, s_a\rangle & |m, s_c\rangle \\ |n, s_a\rangle & 0 & E_{ss} \\ |n, s_c\rangle & 0 & 0 \end{pmatrix}. \quad (5.3.1b)$$

Подобным образом волновая функция катиона в  $n$ -й элементарной ячейке перекрывается с волновой функцией аниона в  $n$ -й элементарной ячейке, а также с волновыми функциями анионов в трех других  $m$ -х элементарных ячейках, расположение которых удовлетворяет соотношениям

$$\vec{d}_m - \vec{d}_n =: (\vec{x} + \vec{y})a/2, \quad (\vec{y} + \vec{z})a/2, \quad (\vec{z} + \vec{x})a/2.$$

Каждое такое перекрытие дает вклад в матрицу  $[H_{nm}]$  в виде блока

$$\begin{pmatrix} |m, s_a\rangle & |m, s_c\rangle \\ |n, s_a\rangle & 0 & 0 \\ |n, s_c\rangle & E_{ss} & 0 \end{pmatrix}. \quad (5.3.1c)$$

Суммируя все эти вклады, получаем матрицу  $[h(\vec{k})]$ , имеющую вид

$$[h(\vec{k})] = \begin{pmatrix} |s_a\rangle & |s_c\rangle \\ |s_a\rangle & E_{sa} & 4E_s g_0, \\ |s_c\rangle & 4E_{ss} g_0^* & E_{sc} \end{pmatrix} \quad (5.3.2)$$

где

$$4g_0 = 1 + e^{-i\vec{k}\cdot\vec{d}_1} + e^{-i\vec{k}\cdot\vec{d}_2} + e^{-i\vec{k}\cdot\vec{d}_3}.$$

Соответственно, векторы  $\vec{d}_{1,2,3}$  определяются соотношениями

$$\vec{d}_1 \equiv (\vec{y} + \vec{z})a/2, \quad \vec{d}_2 \equiv (\vec{z} + \vec{x})a/2, \quad \vec{d}_3 \equiv (\vec{x} + \vec{y})a/2.$$

Аналогично рассчитываются и остальные блоки, образующие полную матрицу  $[h(\vec{k})]$  размера  $(10 \times 10)$ . Окончательный результат имеет следую-

щий вид:

$$\begin{array}{cccccc}
 |s_a\rangle & |s_a\rangle & |s_c\rangle & |X_a\rangle & |Y_a\rangle & |Z_a\rangle \\
 E_{sa} & 4E_{ss}g_0 & 0 & 0 & 0 & 0 \\
 |s_c\rangle & 4E_{ss}g_0^* & E_{sc} & 4E_{pasc}g_1^* & 4E_{pasc}g_2^* & 4E_{pasc}g_3^* \\
 |X_a\rangle & 0 & 4E_{pasc}g_1 & E_{pa} & 0 & 0 \\
 |Y_a\rangle & 0 & 4E_{pasc}g_2 & 0 & E_{pa} & 0 \\
 |Z_a\rangle & 0 & 4E_{pasc}g_3 & 0 & 0 & E_{pa} \\
 |X_c\rangle & 4E_{sapc}g_1^* & 0 & 4E_{xx}g_0^* & 4E_{xy}g_3^* & 4E_{xy}g_2^* \\
 |Y_c\rangle & 4E_{sapc}g_2^* & 0 & 4E_{xy}g_3^* & 4E_{xx}g_0^* & 4E_{xy}g_1^* \\
 |Z_c\rangle & 4E_{sapc}g_3^* & 0 & 4E_{xy}g_2^* & 4E_{xy}g_1^* & 4E_{xx}g_0^* \\
 |s_a^*\rangle & 0 & 0 & 0 & 0 & 0 \\
 |s_c^*\rangle & 0 & 0 & 4E_{pas^*c}g_1^* & 4E_{pas^*c}g_2^* & 4E_{pas^*c}g_3^*
 \end{array} \tag{5.3.3}$$

$$\begin{array}{ccccc}
 |X_c\rangle & |Y_c\rangle & |Z_c\rangle & |s_a^*\rangle & |s_c^*\rangle \\
 4E_{sapc}g_1 & 4E_{sapc}g_2 & 4E_{sapc}g_3 & 0 & 0 \\
 |s_c\rangle & 0 & 0 & 0 & 0 \\
 |X_a\rangle & 4E_{xx}g_0 & 4E_{xy}g_3 & 4E_{xy}g_2 & 0 & 4E_{pas^*c}g_1 \\
 |Y_a\rangle & 4E_{xy}g_3 & 4E_{xx}g_0 & 4E_{xy}g_1 & 0 & 4E_{pas^*c}g_2 \\
 |Z_a\rangle & 4E_{xy}g_2 & 4E_{xy}g_1 & 4E_{xx}g_0 & 0 & 4E_{pas^*c}g_3 \\
 |X_c\rangle & E_{pc} & 0 & 0 & 4E_{s^*apc}g_1^* & 0 \\
 |Y_c\rangle & 0 & E_{pc} & 0 & 4E_{s^*apc}g_2^* & 0 \\
 |Z_c\rangle & 0 & 0 & E_{pc} & 4E_{s^*apc}g_3^* & 0 \\
 |s_a^*\rangle & 4E_{s^*apc}g_1 & 4E_{s^*apc}g_2 & 4E_{s^*apc}g_3 & E_{s^*a} & 0 \\
 |s_c^*\rangle & 0 & 0 & 0 & 0 & E_{s^*c}
 \end{array}$$

Множители  $g_1$ ,  $g_2$  и  $g_3$  имеют вид подобный множителю  $g_0$ , который определен выше при обсуждении модели только с  $s$ -орбиталями:

$$4g_0 \equiv 1 + e^{-i\vec{k}\cdot\vec{d}_1} + e^{-i\vec{k}\cdot\vec{d}_2} + e^{-i\vec{k}\cdot\vec{d}_3}. \tag{5.3.4a}$$

Заметим, что знаки некоторых других множителей являются отрицательными:

$$4g_1 \equiv 1 + e^{-i\vec{k}\cdot\vec{d}_1} - e^{-i\vec{k}\cdot\vec{d}_2} - e^{-i\vec{k}\cdot\vec{d}_3}, \tag{5.3.4b}$$

$$4g_2 \equiv 1 - e^{-i\vec{k}\cdot\vec{d}_1} + e^{-i\vec{k}\cdot\vec{d}_2} - e^{-i\vec{k}\cdot\vec{d}_3}, \tag{5.3.4c}$$

$$4g_3 \equiv 1 - e^{-i\vec{k}\cdot\vec{d}_1} - e^{-i\vec{k}\cdot\vec{d}_2} + e^{-i\vec{k}\cdot\vec{d}_3}. \tag{5.3.4d}$$

Знак минус появляется в силу того, что волновая функция  $p$ -орбитали меняет знак вдоль одной из осей, так что интеграл перекрытия имеет различный знак для различных соседей. Это обстоятельство также влияет на

знаки интегралов перекрытия, появляющихся в выражениях для матрицы  $[h(\vec{k})]$  в (5.2.4): параметры  $E_{ss}$ ,  $E_{pasc}$  и  $E_{pas^*c}$  являются отрицательными, в то время как параметры  $E_{sapc}$ ,  $E_{xx}$ ,  $E_{xy}$  и  $E_{s^*apc}$  положительны. Обратите внимание на то, что векторы

$$\vec{d}_1 \equiv (\vec{y} + \vec{z})a/2, \quad \vec{d}_2 \equiv (\vec{z} + \vec{x})a/2, \quad \vec{d}_3 \equiv (\vec{x} + \vec{y})a/2 \quad (5.3.5)$$

соединяют катион в одной элементарной ячейке с катионом в соседней ячейке (или анион с анионом). Можно было бы определить эти векторы иначе, а именно так, чтобы они соединяли ближайших соседей в решетке. Это привело бы к умножению каждой из величин  $g_0$ ,  $g_1$ ,  $g_2$  или  $g_3$  на фазовый множитель  $\exp[i\vec{k} \cdot \vec{d}]$ , где  $\vec{d} = (\vec{x} + \vec{y} + \vec{z})a/4$ . Такая запись используется многими авторами, например, в работе [50], но указанные различия не приводят к расхождениям в ответе.

На рис. 5.3.2 представлена зонная структура  $E(\vec{k})$ , рассчитанная путем вычисления собственных значений матрицы в выражении (5.3.3) для каждого значения  $\vec{k}$ , направленного вдоль  $\Gamma - X$  и  $\Gamma - L$ . Мы использовали следующие значения параметров для GaAs, которые приведены в работе [50]:

$$\begin{aligned} E_{sa} &= -8.3431 \text{ эВ}, & E_{pa} &= 1.0414 \text{ эВ}, & E_{s^*a} &= 8.5914 \text{ эВ}, \\ E_{sc} &= -2.6569 \text{ эВ}, & E_{pc} &= 3.6686 \text{ эВ}, & E_{s^*c} &= 6.7386 \text{ эВ}, \\ 4E_{ss} &= -6.4513 \text{ эВ}, & 4E_{pasc} &= -5.7839 \text{ эВ}, & 4E_{pas^*c} &= -4.8077 \text{ эВ}, \\ 4E_{sapc} &= 4.48 \text{ эВ}, & 4E_{s^*apc} &= 4.8422 \text{ эВ}, \\ 4E_{xx} &= 1.0546 \text{ эВ}, & 4E_{xy} &= 5.0779 \text{ эВ}. \end{aligned}$$

## 5.4. Влияние спин-орбитального взаимодействия

Проведенный расчет зонной структуры является достаточно обоснованным и точным, однако он не слишком хорошо описывает структуру спектра вблизи потолка валентной зоны. Для получения более точной картины зонной структуры в этой области энергий необходимо включить в рассмотрение спин-орбитальное взаимодействие.

**Спиноры.** Прежде всего, поясним кратко, каким образом можно включить спин в уравнение Шредингера. Обычно, рассчитав уровни энергии с помощью уравнения Шредингера, заполняем их по *два* электрона на уровень. Более строго следовало бы сказать, что каждый уровень состоит из двух уровней с одинаковой энергией, но каждый уровень заполняется только *одним* электроном, как этого требует принцип запрета Паули. Каким образом следует модифицировать уравнение Шредингера, чтобы каждый

уровень превратился в два уровня с одинаковой энергией? Ответ весьма прост: заменим уравнение

$$E(\psi) = [H_{op}](\psi) \quad \text{на} \quad E \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} = \begin{pmatrix} H_{op} & 0 \\ 0 & H_{op} \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix}, \quad (5.4.1)$$

где

$$H_{op} = p^2/2m + U(\vec{r}) \quad (\vec{p} \equiv -i\hbar\vec{\nabla}).$$

Мы будем рассматривать  $\psi$  как компоненту электронной волновой функции, отвечающую проекции спина, направленной вверх, а  $\bar{\psi}$  как компоненту проекции спина, направленного вниз. Теперь изменится базис, который мы используем для получения матричного представления гамильтонiana, так, что соответствующая матрица будет иметь вдвое больший размер. К примеру, если для каждого атома будет использоваться лишь  $s$ -орбиталь, то матрица будет иметь размер  $(4 \times 4)$ , а не  $(2 \times 2)$ , как это было в уравнении (5.3.2):

$$\begin{pmatrix} |s_a\rangle & |s_a\rangle & |s_c\rangle & |\bar{s}_a\rangle & |\bar{s}_c\rangle \\ |s_a\rangle & E_{sa} & 4E_{ss}g_0 & 0 & 0 \\ |s_c\rangle & 4E_{ss}g_0^* & E_{sc} & 0 & 0 \\ |s_a\rangle & 0 & 0 & E_{sa} & 4E_{ss}g_0 \\ |s_c\rangle & 0 & 0 & 4E_{ss}g_0^* & E_{sc} \end{pmatrix}. \quad (5.4.2)$$

Соответственно, при учете всех десяти атомных орбиталей матрица размером  $(10 \times 10)$  превратится в матрицу  $(20 \times 20)$ :

$$[H_0(\vec{k})] = \begin{pmatrix} h(\vec{k}) & 0 \\ 0 & h(\vec{k}) \end{pmatrix}, \quad (5.4.3)$$

где матрица  $h(\vec{k})$  определяется выражением (5.3.3).

**Спин-орбитальное взаимодействие.** Если бы в своих расчетах зонной структуры мы использовали выражение (5.4.3) вместо (5.3.3), то получили бы тот же самый результат. Отличие было бы лишь в том, что над каждой линией в спектре располагалась бы еще одна, которую мы, вероятно, и не заметили бы при построении рисунка. Это вырождение снимется, если мы добавим в выражение (5.4.3) слагаемое, описывающее спин-орбитальное взаимодействие.

Уравнение Шредингера является нерелятивистским. Релятивистские эффекты могут стать существенными только для электронов, движущихся

с большими скоростями. Такие электроны должны описываться уравнением Дирака. Обычно в твердых телах скорости электронов не слишком высокие, но величина электрического поля вблизи атомных ядер достаточно велика. Это приводит к появлению слабых релятивистских эффектов, которые можно учесть, включив спин-орбитальные поправки  $H_{SO}$  в уравнение Шредингера:

$$E \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} = [H_0] \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} + [H_{SO}] \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix}, \quad (5.4.4)$$

где

$$H_0 = \begin{pmatrix} p^2/2m + U(\vec{r}) & 0 \\ 0 & p^2/2m + U(\vec{r}) \end{pmatrix} \quad (5.4.5)$$

и

$$H_{SO} = \frac{q\hbar}{4m^2c^2} \begin{pmatrix} E_x p_y - E_y p_x & (E_y p_z - E_z p_y) - i(E_z p_x - E_x p_z) \\ (E_y p_z - E_z p_y) + i(E_z p_x - E_x p_z) & p^2/2m + U(\vec{r}) \end{pmatrix}, \quad (5.4.6)$$

где  $c$  — скорость света в вакууме. Гамильтониан спин-орбитального взаимодействия  $H_{SO}$  часто записывают в форме

$$H_0 = \frac{q\hbar}{4m^2c^2} \vec{\sigma} \cdot [\vec{E}, \vec{p}], \quad (5.4.7)$$

где компоненты вектора  $\vec{\sigma}$  есть матрицы Паули:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.4.8)$$

Можно непосредственно показать, что два выражения гамильтониана спин-орбитального взаимодействия (5.4.7) и (5.4.6) являются идентичными. Я не буду пытаться объяснить происхождение слагаемого, отвечающего за спин-орбитальное взаимодействие, поскольку это потребовало бы обсуждения деталей, связанных с решением уравнения Дирака. Тем не менее, заинтересованный читатель может найти краткое обсуждение данной проблемы в параграфе 5.5.

**Зонная структура в присутствии спин-орбитального взаимодействия.** После того как из выражения (5.3.3) найдено матричное представление гамильтониана  $H_0$  (5.4.5), необходимо найти матричное представление  $H_{SO}$  и прибавить его к  $H_0$ . Чтобы выполнить эту процедуру, прежде всего, рассмотрим случай, когда в качестве базиса в каждом атоме используются только  $s$ -орбитали. Спин-орбитальные матричные элементы необходимо

учитывать, когда обе орбитали принадлежат одному атому. В этом случае возникает следующая структура гамильтоновской матрицы:

$$\begin{pmatrix} |s_a\rangle & |s_a\rangle & |s_c\rangle & |\bar{s}_a\rangle & |\bar{s}_c\rangle \\ |s_a\rangle & a_{11} & 0 & a_{12} & 0 \\ |s_c\rangle & 0c_{11} & 0 & c_{12} & \\ |\bar{s}_a\rangle & a_{21} & 0 & a_{22} & 0 \\ |\bar{s}_c\rangle & 0 & c_{21} & 0 & c_{22} \end{pmatrix}.$$

Нам следует записать элементы  $a_{11}$  и  $c_{11}$  этой матрицы, вычисляя матричные элементы оператора, являющегося компонентой (11) матрицы  $H_{SO}$  (см. уравнение (5.4.6)):

$$a_{11} = \langle s_a | E_x p_y - E_y p_x | s_a \rangle \quad c_{11} = \langle s_c | E_x p_y - E_y p_x | s_c \rangle.$$

Для записи элементов  $a_{11}$  и  $c_{11}$  нужно вычислить матричные элементы оператора, являющегося компонентой (12) матрицы  $H_{SO}$  (см. уравнение (5.4.6)):

$$a_{12} = \langle s_a | (E_y p_z - E_z p_y) - i(E_z p_x - E_x p_z) | \bar{s}_a \rangle,$$

$$c_{12} = \langle s_c | (E_y p_z - E_z p_y) - i(E_z p_x - E_x p_z) | \bar{s}_c \rangle.$$

Подобным образом можно поступить и с матричными элементами (21) и (22). Если считать потенциал  $U(r)$  сферически симметричным, что выглядит вполне естественно для атомных потенциалов, то все вышеупомянутые матричные элементы будут равны нулю. Этот вывод сохраняется и для атомных орбиталей  $S^*$ -типа. Однако если мы имеем дело с атомными орбиталями  $X$ -,  $Y$ - и  $Z$ -типа, то некоторые из матричных элементов будут ненулевыми. Все эти ненулевые матричные элементы могут быть выражены в единицах одного параметра  $\delta_a$  для анионных атомных орбиталей:

$$\begin{pmatrix} |X_a\rangle & |Y_a\rangle & |Z_a\rangle & |\bar{X}_a\rangle & |\bar{Y}_a\rangle & |\bar{Z}_a\rangle \\ |X_a\rangle & 0 & -i\delta_a & 0 & 0 & \delta_a \\ |Y_a\rangle & i\delta_a & 0 & 0 & 0 & -i\delta_a \\ |Z_a\rangle & 0 & 0 & 0 & -\delta_a & i\delta_a & 0 \\ |\bar{X}_a\rangle & 0 & 0 & -\delta_a & 0 & i\delta_a & 0 \\ |\bar{Y}_a\rangle & 0 & 0 & -i\delta_a & -i\delta_a & 0 & 0 \\ |\bar{Z}_a\rangle & \delta_a & i\delta_a & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (5.4.9a)$$

а также в единицах одного параметра  $\delta_c$  для катионных орбиталей:

$$\begin{pmatrix} |X_c\rangle & |Y_c\rangle & |Z_c\rangle & |\bar{X}_c\rangle & |\bar{Y}_c\rangle & |\bar{Z}_c\rangle \\ |X_c\rangle & 0 & -i\delta_c & 0 & 0 & \delta_c \\ |Y_a\rangle & i\delta_c & 0 & 0 & 0 & -i\delta_c \\ |Z_a\rangle & 0 & 0 & 0 & -\delta_c & i\delta_c & 0 \\ |\bar{X}_a\rangle & 0 & 0 & -\delta_c & 0 & i\delta_c & 0 \\ |\bar{Y}_a\rangle & 0 & 0 & -i\delta_c & -i\delta_c & 0 & 0 \\ |\bar{Z}_a\rangle & \delta_c & i\delta_c & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (5.4.9b)$$

Если мы найдем собственные значения обеих этих матриц, то получим четыре собственных значения, равных  $+\delta_c$  (или  $+\delta_a$ ), и еще два собственных значения, равных  $-2\delta_c$  (или  $-2\delta_a$ ). Разность энергий этих двух уровней, равная  $3\delta_c$  (или  $3\delta_a$ ), называется спин-орбитальным расщеплением  $\Delta_c$  (или  $\Delta_a$ ):

$$\Delta_c \text{ (или } \Delta_a) = 3\delta_c \text{ (или } 3\delta_a). \quad (5.4.10)$$

Эффекты, связанные со спин-орбитальным взаимодействием, хорошо известны. Они изучались как в работах теоретиков, так и из экспериментов с атомами различных элементов. К примеру, в галлии величина спин-орбитального расщепления равна 0.013 эВ, в то время как в мышьяке она равна 0.38 эВ. Теперь мы можем записать полное матричное представление для  $H_{SO}$ , используя выражения (5.4.8) и (5.4.9), прибавляя их к (5.3.3) и рассчитывая далее зонную структуру. Для GaAs мы приходим к результату, приведенному на рис. 5.4.1 *a*. Для сравнения, на рис. 5.4.1 *b* показаны результаты, полученные непосредственно из уравнения (5.3.3) без учета спин-орбитального взаимодействия. Данный рисунок, в основном, повторяет тот, что был приведен в предыдущем параграфе (см. рис. 5.3.2). Правда, для более четкого изображения спектра вблизи потолка валентной зоны здесь выбран более крупный масштаб по оси энергий.

**Тяжелые дырки, легкие дырки и спин-отщепленная зона.** Структура волновой функции вблизи гамма-точки ( $k_x = k_y = k_z = 0$ ) валентной зоны определяет весьма важные оптические характеристики полупроводниковыхnanoструктур. В гамма-точке матрица гамильтониана имеет относительно простой вид, поскольку лишь величина  $g_0$  отлична от нуля, в то время как параметры  $g_1$ ,  $g_2$  и  $g_3$  все равны нулю (см. выражение (5.3.3)). При учете спин-орбитального взаимодействия гамильтониан распадается в гамма-точке на четыре отдельных блока:

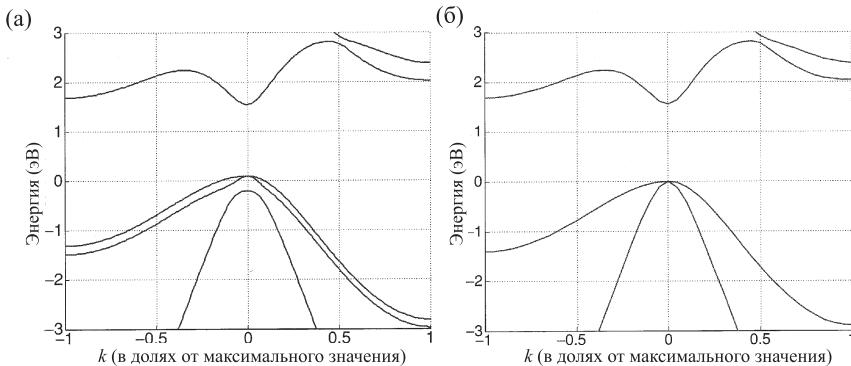


Рис. 5.4.1. Зонная структура GaAs, рассчитанная (а) с учетом спин-орбитального взаимодействия (направление  $\Gamma - X$  выбрано вдоль положительного направления оси абсцисс, в то время как направление  $\Gamma - L$  выбрано вдоль отрицательного направления) и (б) из выражения (5.3.3) без учета спин-орбитального компонента

### Блок I:

$$\begin{pmatrix} |s_a\rangle & |s_c\rangle & |\bar{s}_a\rangle & |\bar{s}_c\rangle \\ |s_a\rangle & E_{sa} & 4E_{ss} & 0 \\ |s_c\rangle & 4E_{ss} & E_{sc} & 0 \\ |\bar{s}_a\rangle & 0 & 0 & E_{sa} \\ |\bar{s}_c\rangle & 0 & 0 & 4E_{ss} \end{pmatrix}; \quad \begin{pmatrix} |s_a^*\rangle & |s_c^*\rangle & |\bar{s}_a^*\rangle & |\bar{s}_c^*\rangle \\ |s_a^*\rangle & E_{s^*a} & 0 & 0 \\ |s_c^*\rangle & 0 & E_{s^*c} & 0 \\ |\bar{s}_a^*\rangle & 0 & 0 & E_{s^*a} \\ |\bar{s}_c^*\rangle & 0 & 0 & 0 \end{pmatrix};$$

### Блок III:

$$\begin{pmatrix} |X_a\rangle & |Y_a\rangle & |\bar{Z}_a\rangle & |X_c\rangle & |Y_c\rangle & |\bar{Z}_c\rangle \\ |X_a\rangle & E_{pa} & -i\delta_a & \delta_a & 4E_{xx} & 0 \\ |Y_a\rangle & i\delta_a & E_{pa} & -i\delta_a & 0 & 4E_{xx} \\ |\bar{Z}_a\rangle & \delta_a & i\delta_a & E_{pa} & 0 & 0 \\ |X_c\rangle & 4E_{xx} & 0 & 0 & E_{pc} & -i\delta_c \\ |Y_c\rangle & 0 & 4E_{xx} & 0 & i\delta_c & E_{pc} \\ |\bar{Z}_c\rangle & 0 & 0 & 4E_{xx} & \delta_c & i\delta_c \end{pmatrix};$$

### Блок IV:

$$\begin{pmatrix} |\bar{X}_a\rangle & |\bar{Y}_a\rangle & |Z_a\rangle & |\bar{X}_c\rangle & |\bar{Y}_c\rangle & |Z_c\rangle \\ |\bar{X}_a\rangle & E_{pa} & i\delta_a & -\delta_a & 4E_{xx} & 0 \\ |\bar{Y}_a\rangle & -i\delta_a & E_{pa} & -i\delta_a & 0 & 4E_{xx} \\ |Z_a\rangle & -\delta_a & i\delta_a & E_{pa} & 0 & 0 \\ |\bar{X}_c\rangle & 4E_{xx} & 0 & 0 & E_{pc} & i\delta_c \\ |\bar{Y}_c\rangle & 0 & 4E_{xx} & 0 & -i\delta_c & E_{pc} \\ |Z_c\rangle & 0 & 0 & 4E_{xx} & -\delta_c & i\delta_c \end{pmatrix}.$$

Можно частично диагонализовать блоки III и IV, перейдя к базису функций тяжелых дырок (HH), легких дырок (LH) и спин-отщепленных состояний (SO). Соответствующая матрица перехода есть

$$[V] = \begin{pmatrix} |HH_a\rangle & |LH_a\rangle & |SO_a\rangle & |HH_c\rangle & |LH_c\rangle & |SO_c\rangle \\ |X_a\rangle & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} & 0 & 0 \\ |Y_a\rangle & i/\sqrt{2} & -i/\sqrt{6} & -i/\sqrt{3} & 0 & 0 \\ |\bar{Z}_a\rangle & 0 & \sqrt{2/3} & -1/\sqrt{3} & 0 & 0 \\ |X_c\rangle & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ |Y_a\rangle & 0 & 0 & 0 & i/\sqrt{2} & -i/\sqrt{6} & -i/\sqrt{3} \\ |\bar{Z}_c\rangle & 0 & 0 & 0 & 0 & \sqrt{2/3} & -1/\sqrt{3} \end{pmatrix},$$

и обычные правила изменения базиса, при которых  $[H_{\text{нов}}] = [V^+][H_{\text{стар}}][V]$ . Преобразованный гамильтониан блока III будет иметь вид

$$\begin{pmatrix} |HH_a\rangle E_{pa} + \delta_a & |LH_a\rangle 0 & |SO_a\rangle 0 & |HH_c\rangle 4E_{xx} & |LH_a\rangle 0 & |SO_c\rangle 0 \\ |LH_a\rangle 0 & |LH_a\rangle E_{pa} + \delta_a & |SO_a\rangle 0 & |LH_c\rangle 0 & |LH_c\rangle 4E_{xx} & |SO_c\rangle 0 \\ |SO_a\rangle 0 & |SO_a\rangle 0 & |SO_a\rangle E_{pa} - 2\delta_a & |SO_a\rangle 0 & |SO_c\rangle 0 & |SO_c\rangle 4E_{xx} \\ |HH_c\rangle 4E_{xx} & |HH_c\rangle 0 & |HH_c\rangle 0 & |HH_c\rangle E_{pc} + \delta_c & |HH_c\rangle 0 & |HH_c\rangle 0 \\ |LH_c\rangle 0 & |LH_c\rangle 4E_{xx} & |LH_c\rangle 0 & |LH_c\rangle 0 & |LH_c\rangle E_{pc} + \delta_c & |LH_c\rangle 0 \\ |SO_c\rangle 0 & |SO_c\rangle 0 & |SO_c\rangle 4E_{xx} & |SO_c\rangle 0 & |SO_c\rangle 0 & |SO_c\rangle E_{pc} - 2\delta_c \end{pmatrix}.$$

Обратите внимание на то, как четко разделены уровни энергии в самой гамма-точке. По этой причине их можно трактовать как принадлежащие тяжелым, легким и отщепленным дыркам (HH, LH и SO). По мере удаления от гамма-точки уровни уже не принадлежат в чистом виде этим типам дырок, и их волновые функции представляются суперпозициями состояний HH, LH и SO.

Подобным образом блок IV может быть преобразован с помощью матрицы перехода

$$[\bar{V}] = \begin{pmatrix} |\bar{H}H_a\rangle & |\bar{L}H_a\rangle & |\bar{S}O_a\rangle & |\bar{H}H_c\rangle & |\bar{L}H_c\rangle & |\bar{S}O_c\rangle \\ |\bar{X}_a\rangle & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} & 0 & 0 \\ |\bar{Y}_a\rangle & -i/\sqrt{2} & i/\sqrt{6} & i/\sqrt{3} & 0 & 0 \\ |\bar{Z}_a\rangle & 0 & -\sqrt{2/3} & 1/\sqrt{3} & 0 & 0 \\ |\bar{X}_c\rangle & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ |\bar{Y}_a\rangle & 0 & 0 & 0 & -i/\sqrt{2} & i/\sqrt{6} & i/\sqrt{3} \\ |\bar{Z}_c\rangle & 0 & 0 & 0 & 0 & -\sqrt{2/3} & 1/\sqrt{3} \end{pmatrix}.$$

В результате такого преобразования получается матрица

$$\begin{pmatrix} |\bar{H}H_a\rangle & |\bar{L}H_a\rangle & |\bar{SO}_a\rangle & |\bar{HH}_c\rangle & |\bar{L}H_a\rangle & |\bar{SO}_c\rangle \\ |\bar{H}H_a\rangle & E_{pa} + \delta_a & 0 & 0 & 4E_{xx} & 0 \\ |\bar{L}H_a\rangle & 0 & E_{pa} + \delta_a & 0 & 0 & 4E_{xx} \\ |\bar{SO}_a\rangle & 0 & 0 & E_{pa} - 2\delta_a & 0 & 0 \\ |\bar{HH}_c\rangle & 4E_{xx} & 0 & 0 & E_{pc} + \delta_c & 0 \\ |\bar{L}H_c\rangle & 0 & 4E_{xx} & 0 & 0 & E_{pc} + \delta_c \\ |\bar{SO}_c\rangle & 0 & 0 & 4E_{xx} & 0 & 0 \\ & & & & & E_{pc} - 2\delta_c \end{pmatrix}.$$

Необходимо отметить, что собственные векторы, которые образуют столбцы матриц  $[V]$  и  $[\bar{V}]$  соответственно, не отвечают состояниям, у которых спин в чистом виде ориентирован вверх или вниз. При этом, однако, мы можем рассматривать нижний блок  $[\bar{V}]$  как отвечающий состояниям с перевернутым спином по отношению к состояниям верхнего блока  $[V]$ . Можно непосредственно увидеть, что их собственные векторы являются ортогональными, как и должно быть для состояний со спином «вверх» и «вниз».

## 5.5. Дополнение: уравнение Дирака

Релятивистские электроны описываются уравнением Дирака:

$$E \begin{pmatrix} \psi \\ \bar{\psi} \\ \phi \\ \bar{\phi} \end{pmatrix} = \begin{pmatrix} mc^2 + U & 0 & cp_z & c(p_x - ip_y) \\ 0 & mc^2 + U & c(p_x + ip_y) & -cp_z \\ cp_z & c(p_x - ip_y) & -mc^2 + U & 0 \\ c(p_x - ip_y) & -cp_z & 0 & -mc^2 + U \end{pmatrix} \begin{pmatrix} \psi \\ \bar{\psi} \\ \phi \\ \bar{\phi} \end{pmatrix},$$

которое в более компактной форме можно записать как

$$E \begin{pmatrix} \Psi \\ \Phi \end{pmatrix} = \begin{pmatrix} (mc^2 + U)I & c\vec{\sigma} \cdot \vec{p} \\ c\vec{\sigma} \cdot \vec{p} & (mc^2 + U)I \end{pmatrix} \begin{pmatrix} \Psi \\ \Phi \end{pmatrix}, \quad (5.5.1)$$

где

$$\Psi \equiv \begin{pmatrix} \psi \\ \bar{\psi} \end{pmatrix} \quad \text{и} \quad \Phi \equiv \begin{pmatrix} \phi \\ \bar{\phi} \end{pmatrix}.$$

Предполагая, что  $U = 0$ , подставим в уравнение Дирака волновую функцию в виде плоской волны

$$\begin{pmatrix} \Psi \\ \Phi \end{pmatrix} = \begin{pmatrix} \Psi \\ \Phi \end{pmatrix} e^{i\mathbf{k} \cdot \mathbf{r}}.$$

Можно убедиться, что соответствующее дисперсионное соотношение в этом случае будет состоять из двух ветвей спектра, определяемых соотношением (см. рис. 5.5.1)

$$E(\vec{k}) = \pm \sqrt{m^2 c^4 + c^2 \hbar^2 k^2}.$$

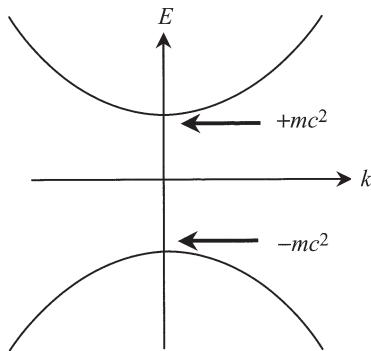


Рис. 5.5.1

Состояния на ветви с отрицательной энергией считаются полностью заполненными, чему отвечает так называемый дираковский вакуум. Расстояние между ветвями  $2mc^2$  равно приблизительно 1 МэВ, что значительно превосходит диапазон энергий, характерных для физики твердого тела. В экспериментах с высокими энергиями электроны возбуждаются при переходе с отрицательной ветви на положительную дисперсионную ветвь, т. е. рождаются электрон-позитронные пары. В физике твердого тела (в стандартных экспериментах) энергии электронов обычно ограничены величинами, не превосходящими 10 эВ; при этом ветви отрицательных энергий являются лишь инертным фоном. В области энергий, близких к  $E = mc^2$ , можно произвести биномиальное разложение выражения (5.5.1) и получить нерелятивистскую, т. е. параболическую зависимость энергии от импульса. Это разложение начинается с аддитивной постоянной, равной релятивистской энергии покоя  $mc^2$ :

$$E(\vec{k}) \approx mc^2 + (\hbar^2 k^2 / 2m).$$

В следующем разделе будут найдены релятивистские поправки, которые описывают поведение квантовых частиц при энергиях, значительно меньших  $mc^2$  (например, спин-орбитальное взаимодействие).

**Нерелятивистское приближение уравнения Дирака.** Исходя из уравнения (5.5.1), можно показать, что

$$E\{\Psi\} = (mc^2 + U)\{\Psi\} + [c\vec{\sigma} \cdot \vec{p}] \left[ \frac{1}{E + mc^2 - U} \right]^{-1} [c\vec{\sigma} \cdot \vec{p}]\{\Psi\}.$$

Заменяя энергию в правой части на  $mc^2$  и исключая компоненту  $\{\Phi\}$ , найдем в первом неисчезающем приближении

$$E\{\Psi\} = (mc^2 + U)\{\Psi\} + \frac{[\vec{\sigma} \cdot \vec{p}]^2}{2m}\{\Psi\}, \quad (5.5.2)$$

которое можно упростить и получить уравнение (5.4.1):

$$(E - mc^2)\{\Psi\} = \begin{pmatrix} U + p^2/2m & 0 \\ 0 & U + p^2/2m \end{pmatrix} \{\Psi\},$$

если учесть, что

$$[\vec{\sigma} \cdot \vec{p}]^2 = \begin{pmatrix} p^2 & 0 \\ 0 & p^2 \end{pmatrix}.$$

**Влияние магнитного поля.** В данной книге мы не собираемся подробно обсуждать влияние магнитного поля на электронные энергетические уровни. Отметим лишь, что это влияние учитывается в уравнении Дирака путем замены оператора  $\vec{p}$  на оператор  $\vec{p} + q\vec{A}$ , где  $\vec{A}$  – векторный электромагнитный потенциал:

$$E \begin{pmatrix} \Psi \\ \Phi \end{pmatrix} = \begin{pmatrix} (mc^2 + U)I & c\vec{\sigma} \cdot (\vec{p} + q\vec{A}) \\ c\vec{\sigma} \cdot (\vec{p} + q\vec{A}) & (-mc^2 + U)I \end{pmatrix} \begin{pmatrix} \Psi \\ \Phi \end{pmatrix}.$$

Как и прежде (см. уравнение (5.5.2)), можно получить нерелятивистское приближение в низшем порядке

$$E\{\Psi\} = (mc^2 + U)\{\Psi\} + \frac{[\vec{\sigma} \cdot (\vec{p} + q\vec{A})]^2}{2m}\{\Psi\},$$

которое после несложных преобразований сводится к уравнению Паули

$$(E - mc^2)\{\Psi\} = [U + (\vec{p} + q\vec{A})^2/2m][I]\{\Psi\} + \mu_B \vec{\sigma} \cdot \vec{B}\{\Psi\},$$

где  $\mu_B \equiv q\hbar/2m$  (постоянная, известная как магнетон Бора), поле  $\vec{B} = [\vec{\nabla}, \vec{A}]$ , а второе слагаемое в правой части ( $\mu_B \vec{\sigma} \cdot \vec{B}$ ) называется слагаемым Зеемана. Обратите внимание на то, что спин-орбитальное слагаемое

в уравнении (5.4.7) можно рассматривать как слагаемое Зеемана, возникающее благодаря эффективному магнитному полю, определяемому выражением

$$\vec{B}_{SO} = [\vec{E}, \vec{p}] / 2mc^2.$$

Механизм спин-орбитального взаимодействия можно пояснить следующим образом: электрон, движущийся в электрическом поле, благодаря «релятивистским эффектам» «видит» эффективное магнитное поле. Для получения спин-орбитального слагаемого непосредственно из уравнения Дирака необходимо обратиться к более высоким порядкам разложения (см., например, книгу [46]).<sup>2</sup>

## Упражнения

**У.5.1.** Запишите матрицу  $(2 \times 2)$  из уравнения (5.1.10) для одномерной модели твердого тела и постройте дисперсионную кривую  $E(k)$ , сравнив ее с рис. 5.1.5.

**У.5.2.** Запишите матрицу  $(10 \times 10)$  из уравнения (5.3.3), используя параметры GaAs, указанные в конце параграфа 5.3 и постройте дисперсионную кривую  $E(k_x, k_y, k_z)$  вдоль направлений  $\Gamma - X$  и  $\Gamma - L$ , как показано на рис. 5.3.2.

**У.5.3.** Запишите матрицу  $(20 \times 20)$ , включающую спин-орбитальное взаимодействие, как это описано в параграфе 5.4, для материала GaAs, и постройте зависимость  $E$  от  $k$  вдоль направлений  $\Gamma - X$  и  $\Gamma - L$  (для GaAs  $\Delta_c = 0.013$  эВ и  $\Delta_a = 0.38$  эВ). Сравните с результатами на рис. 5.4.1 а.

**У.5.4.** Найдите собственные значения матрицы (параметры  $a$ ,  $b$  и  $c$  являются постоянными), руководствуясь принципами расчета зонной структуры, обсуждаемыми в данной главе:

$$\begin{pmatrix} a & b & c & 0 & c & b \\ b & a & b & c & 0 & c \\ c & b & a & b & c & 0 \\ 0 & c & b & a & b & c \\ c & 0 & c & b & a & b \\ b & c & 0 & c & b & a \end{pmatrix}.$$

---

<sup>2</sup>Механизм спин-орбитального взаимодействия обсуждается также, например, в учебниках [56, 62]. — Прим. ред. перев.

# Глава 6

## Подзоны

В предыдущей главе мы пришли к выводу о том, что уровни энергии  $E_b(\vec{k})$  в твердом теле с кристаллической структурой характеризуются непрерывным квантовым числом  $\vec{k}$  и дискретным квантовым числом  $b$ , которое равно числу базисных функций и энергетических зон. Строго говоря, этот результат справедлив при наличии периодических граничных условий по всем трем пространственным направлениям. Хотя на границе реальных образцов идеальная периодичность нарушается, это обстоятельство игнорируется, поскольку ситуация на поверхности не оказывает существенного влияния на объемные свойства. Конечный размер реальных массивных твердых тел обычно не приводит к экспериментально наблюдаемым эффектам, однако по мере уменьшения их размеров дискретность спектра становится сопоставимой с тепловой энергией  $k_B T$ . В этой главе обсуждаются спектры пространственно ограниченных структур, состоящие из энергетических подзон, которые определяют физику эффектов «размерного квантования». В параграфе 6.1 на конкретных примерах рассматривается эффект размерного квантования и его связь с дисперсионной зависимостью  $E(\vec{k})$ . Далее (см. параграф 6.2) обсуждается влияние размерного квантования на экспериментально наблюдаемые величины, такие как плотность состояний  $D(E)$ , а также на число подзон, или мод  $M(E)$ . В параграфе 6.3 будет показано, что максимальная проводимость квантовой проволоки (или нити, т. е. одномерной системы) пропорциональна числу мод с энергией, близкой к уровню Ферми ( $E = \mu$ ), а максимальная проводимость одной моды равна фундаментальной постоянной величине  $G_0 = q^2/h$  (уравнение (1.1)). Наконец, в параграфе 6.4 будет рассматриваться вопрос о правильном определении скорости электрона в периодическом твердом теле. Для свободных электронов волновая функция имеет вид плоских волн  $\sim \exp(\pm ikx)$ , а соответствующая скорость равна  $\hbar k/m$ . Электроны в кристаллическом твердом теле также описываются волновыми функциями, которые задаются квантовым числом  $k$ , но эти функции не являются плоскими волнами. Естественно задаться вопросом о том, что же представляет собой та величина (если

она существует!), которая заменяет  $\hbar k/m$  в этом случае, и как можно её получить, если известна зонная структура  $E_b(\vec{k})$ ?

## 6.1. Квантовые ямы, проволоки, точки и «нанотрубки»

В главе 5 было показано, что в однородном массивном твердом теле роль квантового числа, задающего энергетические уровни, играет волновой вектор  $\mathbf{k}$ . Какие квантовые числа должны характеризовать уровни энергии в структуре нанометрового масштаба? Рассмотрим в качестве примера структуру транзистора, которая обсуждалась в главе 1, где для упрощения расчета на каждой стороне канала введены два симметрично расположенных затвора (рис. 6.1.1).

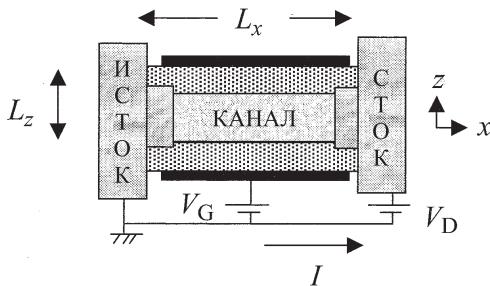


Рис. 6.1.1. Схема полевого нанотранзистора (FET) с двумя затворами. На верхний затвор подан тот же потенциал  $V_G$ , что на нижний

Размер в направлении  $x$  (равный  $L$ ) выбирается достаточно малым, но, поскольку электроны входят и выходят через контакты, необходимо ввести так называемые «открытые граничные условия», которые будут обсуждаться в главе 8. Правда, не будет большой ошибкой рассматривать такую систему и как закрытую, предполагая наличие периодических граничных условий, по крайней мере, в отсутствие напряжения стока ( $V_D = 0$ ). Движение носителей тока в направлении оси  $z$  будет финитным, что должно привести к наблюдаемым эффектам размерного квантования, которые нельзя описать, используя периодические граничные условия. В направлении  $y$ , перпендикулярном плоскости рисунка, размер системы обычно составляет несколько микрон, так, что размерное квантование можно не учитывать. Однако по мере уменьшения размеров структур такое приближение может стать неприемлемым. Каким же образом нам следует нумеровать уровни энергии в структурах, подобных рассматриваемой?

В однородном твердом теле электроны обладают свободой перемещения во всех трех направлениях, и уровни энергии могут быть описаны как  $E_b(k_x, k_y, k_z)$ , где индекс  $b$  обозначает различные энергетические зоны. Напротив, структура транзистора на рис. 6.1.1 представляет собой квантовую яму, где электроны свободно движутся лишь в плоскости  $x - y$ . Тогда уровни энергии в этой структуре можно будет найти, исходя из спектра однородного твердого тела  $E_b(k_x, k_y, k_z)$ . Потенциал, удерживающий носители и приводящий к их финитному движению вдоль оси  $z$ , можно моделировать, накладывая периодические граничные условия, ограничивать разрешенные значения величины  $k_z$  дискретным набором  $k_z = 2\pi p/L_z$ :

$$E_{b,p}(k_x, k_y) \approx E_b(k_x, k_y, k_z = 2\pi p/L_z),$$

где еще один новый индекс  $p$  может назвать индексом *подзоны*. Такое описание работает достаточно хорошо для структур, имеющих геометрию кольца, таких как углеродные нанотрубки. Но для большинства низкоразмерных структур характерна другая геометрия, задаваемая более сложными потенциалами, ограничивающими движение. Соответственно, должны использоваться более сложные граничные условия. Для модели нанотранзистора, показанной на рис. 6.1.1, слои изолятора создают потенциальную квантовую яму с бесконечно высокими стенками (см. рис. 2.1.3 б). Довольно точные оценки спектра можно получить, предполагая, что размер вдоль  $z$  ограничивает разрешенные значения  $k_z$  дискретным набором  $k_z = p\pi/L_z$ . Уровни энергии в этом случае могут классифицироваться как

$$E_{b,p}(k_x, k_y) \approx E_b(k_x, k_y, k_z = p\pi/L_z).$$

Как было установлено выше, это выражение носит лишь приближенный характер, но основная идея, которую я хочу высказать, заключается в том, что *квантовые ямы имеют одномерные подзоны*, в каждой из которых существует свой *двумерный закон дисперсии*  $E(k_x, k_y)$ .

Насколько малым должен быть размер  $L_z$ , чтобы в структуре можно было наблюдать эффекты квантования и рассматривать эту структуру как квантовую яму? Очевидно, с того момента, когда в экспериментах, выполненных на структуре, начинают наблюдаться эффекты размерного квантования. Для этого требуется, чтобы расстояние между дискретными уровнями, отвечающими квантованным значениям  $k_z = p\pi/L_z$ , было больше или порядка тепловой энергией  $k_B T$ . В противном случае все наблюдаемые эффекты размываются на этом масштабе энергии. Для элементарной оценки предположим, что дисперсионная зависимость  $E_b(\vec{k})$  в интересующей нас

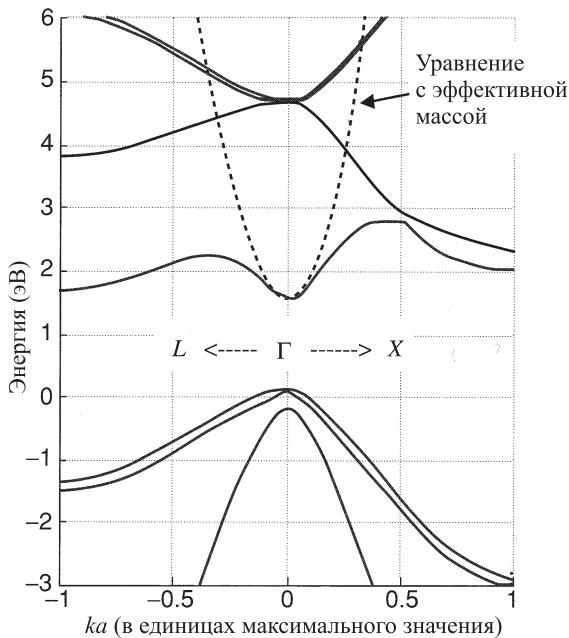


Рис. 6.1.2. Сплошными линиями показана полная зонная структура, полученная в  $sp^3s^*$ -модели, описанной в главе 5. Пунктирные кривые показывают дисперсионную зависимость, полученную из однозонной модели с эффективной массой (уравнение (6.1.1)) с параметрами, подобранными для наилучшего совпадения:  $E_c = 1.55$  эВ и  $m_c = 0.12m$  ( $m$  есть масса свободного электрона). Фактически принятые для эффективной массы GaAs значение равно  $0.07m$ , но в  $sp^3s^*$ -модели используемые нами параметры оптимизированы так, чтобы обеспечить наилучшее совпадение во всей зоне, и не обязательно очень точны вблизи края зоны

области энергий описывается параболической кривой с эффективной массой  $m_c$ :

$$E(\vec{k}) \approx E_c + \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m_c} \quad \text{Массивное твердое тело,} \quad (6.1.1)$$

где  $E_c$  и  $m_c$  являются постоянными, которые можно определять, если известна зонная структура (см. рис. 6.1.2). Эти параметры, определяющие энергию дна зоны проводимости и соответствующую эффективную массу, хорошо известны для всех широко используемых полупроводников.

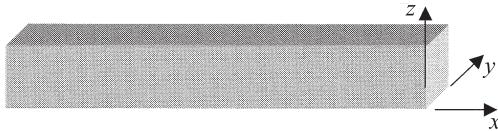
В этом приближении с дискретной  $z$ -компонентой связано появление подзон (обозначаемых индексом  $p$ ), в которых

$$E_p(k_x, k_y) \approx E_c + p^2 \varepsilon_z + \frac{\hbar^2(k_x^2 + k_y^2)}{2m_c} \quad \text{Квантовая яма,}$$

$$\varepsilon_z = \frac{\hbar^2 \pi^2}{2m_c L_z^2} = \frac{m}{m_c} \left( \frac{10 \text{ нм}}{L_Z} \right)^2 \times 3.8 \text{ мэВ.} \quad (6.1.2)$$

В слое толщиной 10 нм величина энергии подзон имеет порядок 4 мэВ, если эффективная масса  $m_c$  равна массе свободного электрона  $m$ . В материалах с меньшей электронной массой (таких как GaAs, где  $m_c = 0.07m$ ) расстояние между подзонами увеличивается, и поэтому в экспериментах эффекты размерного квантования выражены более ярко.

Можно сделать еще один шаг и рассмотреть структуры, в которых движение ограничено как вдоль  $y$ , так и вдоль  $z$ , а свободно электроны могут перемещаться лишь вдоль направления  $x$ , как показано на следующем рисунке:



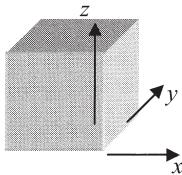
Подобная ситуация будет иметь место, например, когда ширина полевого транзистора на рис. 6.1.1 в направлении, перпендикулярном плоскости рисунка, будет иметь порядок 100 нм. Подобные *квантовые проволоки* имеют двумерные подзоны, в каждой из которых существует свое одномерное дисперсионное соотношение, которое можно приблизительно описать, предполагая, что квантуются как  $y$ -, так и  $z$ -компоненты волнового вектора  $\vec{k}$ :

$$E_{n,p}(k_x) \approx E_c + n^2 \varepsilon_y + p^2 \varepsilon_z + \frac{\hbar^2 k_x^2}{2m_c} \quad \text{Квантовая проволока,}$$

где энергия  $\varepsilon_y$  связана с размером  $L_y$  в направлении  $y$  соотношением, аналогичным (6.1.2). Наконец, можно рассматривать структуры, в которых движение электронов ограничено во всех трех направлениях (как показано на рисунке ниже), что приводит к энергетическому спектру, состоящему

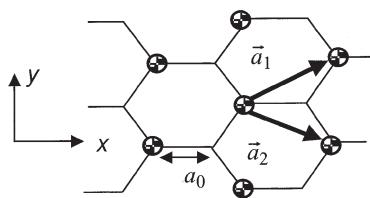
из дискретных уровней, подобному спектру атомов. Для оценки положения уровней здесь можно использовать формулы

$$E_{m,n,p} \approx E_c + \frac{m^2 \hbar^2 \pi^2}{2m_c L_x^2} + \frac{n^2 \hbar^2 \pi^2}{2m_c L_y^2} + \frac{p^2 \hbar^2 \pi^2}{2m_c L_z^2} \quad \text{Квантовая точка.}$$



Подобные структуры типа квантовых точек часто называют искусственными атомами.

(а) Исходная кристаллическая решетка  
(решетка в координатном пространстве)



(б) Обратная решетка

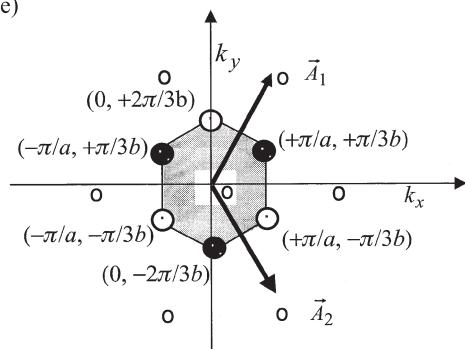


Рис. 6.1.3. (а) Расположение атомов углерода на поверхности графита. (б) Обратная решетка и зона Бриллюэна (заштрихована)

**Углеродные нанотрубки.** Углеродные нанотрубки представляют собой хороший пример, иллюстрирующий концепцию энергетических подзон (рис. 6.1.3).<sup>1</sup> В параграфе 5.2 мы видели, что уровни энергии в слое графита

<sup>1</sup>Более подробную информацию о графене и наногрубках можно найти, например, в обозрах [64, 65]. — Прим. ред. перев.

могут быть найдены путем диагонализации матрицы  $(2 \times 2)$

$$h(\vec{k}) = \begin{pmatrix} 0 & h_0 \\ h_0^* & 0 \end{pmatrix}, \quad (6.1.3)$$

где

$$h_0 \equiv -t(1 + e^{i\vec{k} \cdot \vec{a}_1} + e^{i\vec{k} \cdot \vec{a}_2}) = -t(1 + 2e^{ik_x a} \cos k_y b).$$

Собственные значения определяются выражением

$$E = \pm|h_0| = \pm t \sqrt{1 + 4 \cos k_y b \cos k_x a + 4 \cos^2 k_y b}. \quad (6.1.4)$$

Поскольку в каждой элементарной ячейке имеется две базисные функции, полное число состояний равно  $2N$ , где  $N$  есть число элементарных ячеек. Каждый атом углерода отдает по одному электрону на формирование  $\pi$ -зоны, а всего, таким образом,  $2N$  электронов заполняют ровно половину состояний. Поскольку уровни энергии расположены симметрично относительно значения  $E = 0$ , то все состояния с  $E < 0$  заняты, в то время как состояния с  $E > 0$  свободны, или, можно сказать, что уровень Ферми располагается на уровне  $E = 0$ .

Зададимся следующим вопросом: где расположены те области плоскости  $k_x - k_y$ , которые отвечают значению  $E = 0$ ? Ответ будет таким: это области, где  $h_0(\vec{k}) = 0$ . Нетрудно видеть, что это равенство выполняется в шести углах зоны Бриллюэна:

$$\begin{aligned} e^{ik_x a} \cos k_y b = -1/2 : \quad k_x a = 0, \quad k_y b = \pm 2\pi/3, \\ k_x a = \pm\pi, \quad k_y b = \pm\pi/3. \end{aligned}$$

Эти шесть точек являются выделенными, или особыми, в том смысле, что они отвечают состояниям в непосредственной близости от энергии Ферми и поэтому определяют электронные свойства материала. Эти точки можно объединить в две группы по три точки:

$$(k_x a, k_y b) = (0, -2\pi/3), (-\pi, +\pi/3), (\pi, +\pi/3); \quad (6.1.5a)$$

$$(k_x a, k_y b) = (0, +2\pi/3), (-\pi, -\pi/3), (\pi, -\pi/3). \quad (6.1.5b)$$

Все три точки в одной группе являются эквивалентными, поскольку их координаты различаются между собой на вектор обратной решетки. Каждой из шести точек соответствует *одна треть* долины вокруг нее в пределах первой зоны Бриллюэна (заштрихованная область на рис. 6.1.3б). Но

при этом можно произвести трансляцию каждой области на соответствующий вектор обратной решетки, чтобы сформировать две *полные* долины вокруг этих двух точек (по одной из каждой группы):

$$(k_x a, k_y b) = (0, \pm 2\pi/3).$$

Если лист графита свернуть в нанотрубку, то разрешенные значения  $k$  будут определяться периодическими граничными условиями (вдоль направления периметра). Обратите внимание на то, что в данном случае периодическое граничное условие *действительно* отражает физическую структуру. Вводя вектор периметра

$$\vec{c} = m\vec{a}_1 + n\vec{a}_2 = \vec{x}(m+n)a + \vec{y}(m-n)b, \quad (6.1.6)$$

который соединяет две эквивалентные точки плоскости  $x - y$ , объединяемые в одну точку при сворачивании в трубку. В этом случае требованию периодического граничного условия можно удовлетворить, полагая

$$\vec{k} \cdot \vec{c} \equiv k_c |c| = k_x a(m+n) + k_y b(m-n) = 2\pi\nu, \quad (6.1.7)$$

что определяет семейство параллельных линий, каждая из которых отвечает определенному целочисленному значению  $\nu$  (рис. 6.1.4). Одномерное дисперсионное соотношение можно записать вдоль любой из этих линий, что даст нам совокупность дисперсионных кривых  $E_\nu(k)$ , по одной для каждой подзоны  $\nu$ .

Появится ли в законе дисперсии запрещенная зона, зависит от того, будет ли какая-либо линия, определяемая уравнением (6.1.7), проходить через центр долины:

$$(k_x a, k_y b) = (0, \pm 2\pi/3),$$

где уровни энергии располагаются при  $E = 0$ . С помощью уравнения (6.1.7) нетрудно убедиться, что прямая линия пересечется с точкой  $k_x a = 0, k_y b = 2\pi/3$  при выполнении условия

$$(m - n)/3 = \nu.$$

Поскольку  $\nu$  является целым числом, этого можно добиться лишь, когда разность  $(m - n)$  кратна трем: нанотрубки, удовлетворяющие этому условию, обладают металлическими свойствами.

**Нанотрубки типа “zigzag” и “armchair”.** Рассмотрим следующий частный пример: нанотрубку, у которой вектор периметра ориентирован

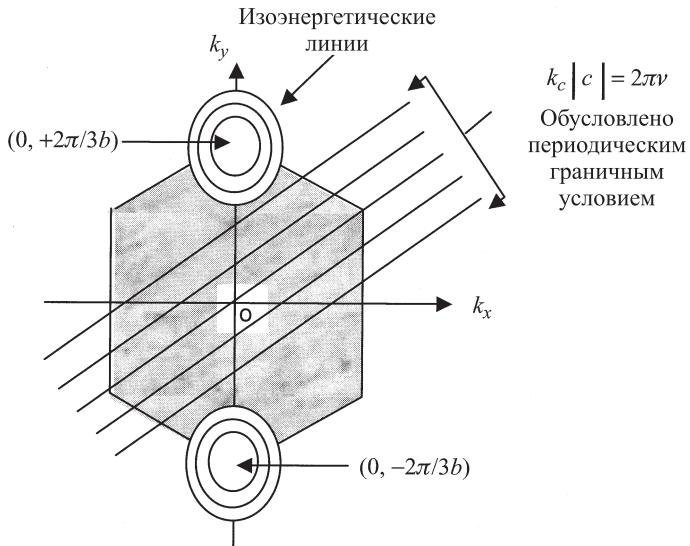


Рис. 6.1.4. Обратная решетка графита, в которой показаны прямые линии  $k_c|c| = 2\pi\nu$ , представляющие ограничения, накладываемые периодическими граничными условиями

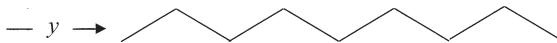


Рис. 6.1.5

вдоль направления  $y$  и имеет вид  $\vec{c} = \vec{y}2bm$ . Такая трубка называется нанотрубкой типа “zigzag”, поскольку ее край (после сворачивания в трубку) выглядит как зигзаг (рис. 6.1.5). Периодическое граничное условие в этом случае требует, чтобы разрешенные значения  $\vec{k}$  располагались параллельно оси  $k_x$  и описывались выражением

$$k_y 2bm = 2\pi\nu \rightarrow k_y = \frac{2\pi}{3b} \frac{3\nu}{2m}, \quad (6.1.8)$$

как это показано на рис. 6.1.6 а.

На рис. 6.1.7 показаны две «нижние» подзоны, отвечающие значению индекса подзоны  $\nu$  с наименьшим значением величины запрещенной зоны вблизи  $E = 0$ . Если число  $m = 66$  (т. е. кратно трем), то одна из подзон будет проходить через точку  $(k_x a, k_y b) = (0, \pm 2\pi/3)$ , и дисперсионное

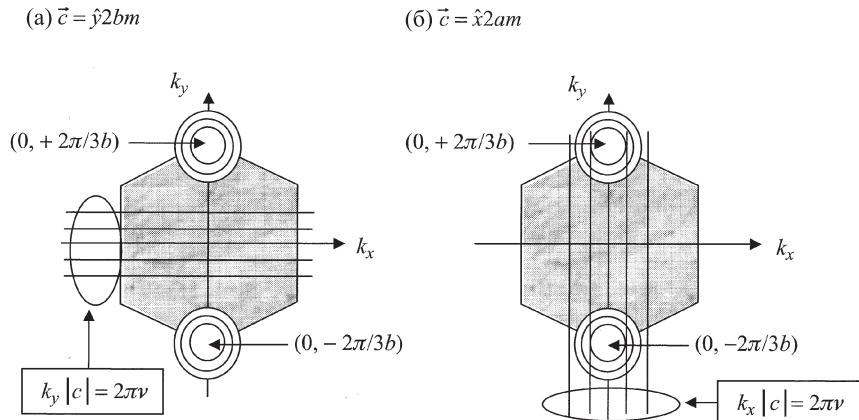


Рис. 6.1.6. (а) Зигзагообразная нанотрубка, полученная сворачиванием листа графита вдоль оси  $y$  с вектором  $\vec{c} = \hat{y}2bm$ . Она обладает разрешенными значениями  $\vec{k}$ , ограниченными направлениями линий, параллельных оси  $k_x$ , как показано на рисунке. Одна из линий будет проходить через точку  $(0, 2\pi/3b)$  лишь при условии, когда число  $t$  кратно трем. (б) Нанотрубка с краем в форме кресла, полученная сворачиванием листа графита вдоль оси  $x$  с вектором  $\vec{c} = \hat{x}2am$ . Она обладает разрешенными значениями  $\vec{k}$ , ограниченными направлениями линий, параллельных оси  $k_y$ , как показано на рисунке. Одна из линий всегда будет проходить через точку  $(0, 2\pi/3b)$  независимо от значения числа  $t$

соотношение будет иметь вид, показанный на рис. 6.1.7 а (отсутствует запрещенная зона). Но если значение  $t = 65$  (т. е. не кратно трем), то ни одна из подзон не будет проходить через точку  $(k_x a, k_y b) = (0, \pm 2\pi/3)$ , что приведет к образованию запрещенной зоны в энергетическом спектре, как это показано на рис. 6.1.7 б.

Нанотрубка, характеризуемая вектором периметра  $\vec{c} = \hat{x}2am$ , направленным, вдоль направления оси  $x$ , называется нанотрубкой типа “armchair”<sup>2</sup>. После сворачивания в трубку форма ее края напоминает кресло (“armchair”), как это показано на рис. 6.1.8 (чтобы представить это, необходимо определенное воображение!). Периодическое граничное условие в этом случае требует, чтобы разрешенные значения  $\vec{k}$  располагались

<sup>2</sup> В настоящее время терминология, обозначающая русское название данного типа нанотрубок, не сформировалась. — Прим. ред. перев.

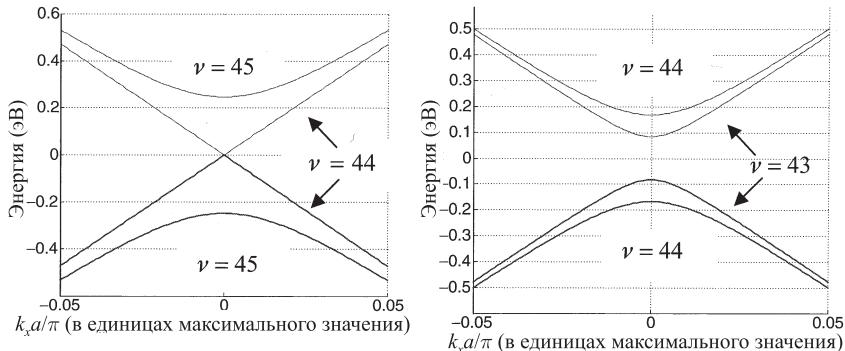


Рис. 6.1.7. Дисперсионные кривые двух «нижних» подзон в нанотрубке типа “zigzag” с параметрами (а)  $D = 5.09 \text{ нм}$ ,  $m = 66$ , где реализуется металлический спектр, т. е. отсутствует запрещенная зона и (б)  $D = 5.02 \text{ нм}$ ,  $m = 65$ , со спектром, характерным для полупроводника (присутствует запрещенная зона)

параллельно оси  $k_y$  и описывались выражением (длина периметра вновь равна  $2bm$ )

$$k_x 2am = 2\pi\nu \rightarrow k_x = \frac{2\pi\nu}{2ma}, \quad (6.1.9)$$

как это показано на рис. 6.1.6 б. В этом случае подзона с номером  $\nu = 0$  всегда будет проходить через выделенную точку  $(k_x a, k_y b) = (0, 2\pi/3b)$ , что приведет к дисперсионным соотношениям металлического типа (рис. 6.1.7 а) независимо от значения  $m$ .

**Полезное приближение.** Как известно, электрическая проводимость определяется состояниями с энергиями, близкими к энергии Ферми. Поэтому полезно вывести приближенное соотношение, описывающее спектр  $E(k)$  вблизи значения  $E = 0$ . Это можно сделать, заменив выражение для  $h_0(\vec{k}) = -t(1 + 2e^{ik_x a} \cos k_y b)$  на его разложение в виде ряда Тейлора вблизи точки  $(k_x a, k_y b) = (0, 2\pi/3b)$ , где величина запрещенной зоны равна нулю. Обратите внимание на то, что в этих точках  $h_0 = 0$ :

$$h_0 \approx k_x \left[ \frac{\partial h_0}{\partial k_x} \right]_{k_x a=0, k_y b=\pm 2\pi/3} + \left( k_y \mp \frac{2\pi}{3b} \right) \left[ \frac{\partial h_0}{\partial k_y} \right]_{k_x a=0, k_y b=\pm 2\pi/3}.$$

Частные производные вычисляются непосредственно:

$$\frac{\partial h_0}{\partial k_x} = [-2iat e^{ik_x a} \cos k_y b]_{k_x a=0, k_y b=\pm 2\pi/3} = iat = i3a_0 t/2,$$



Рис. 6.1.8

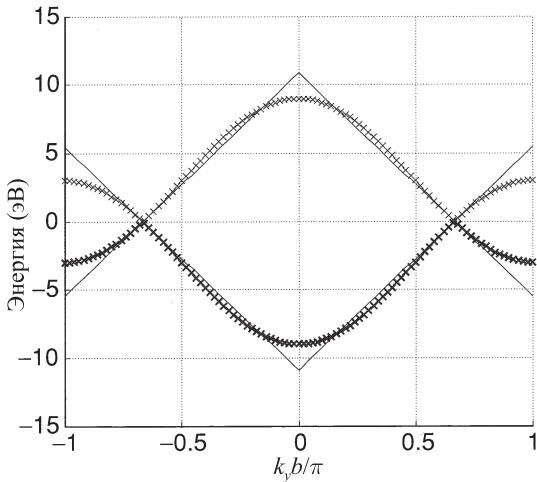


Рис. 6.1.9. Дисперсионная зависимость, построенная как функция  $k_y b$  вдоль линии  $k_x a = 0$ . Сплошная линия получена из выражения (6.1.11), а линия с поперечными штрихами — из уравнения (6.1.4)

$$\frac{\partial h_0}{\partial k_y} = [2bt e^{ik_x a} \sin k_y b]_{k_x a=0, k_y b=\pm 2\pi/3} = \pm bt\sqrt{3} = \pm 3a_0 t/2,$$

поэтому можно записать

$$h_0(\vec{k}) \approx i \frac{3a_0 t}{2} (k_x \mp i\beta_y),$$

где

$$\beta_y \equiv k_y \mp (2\pi/3b). \quad (6.1.10)$$

Соответствующая дисперсионная зависимость (сравните с выражением (6.1.4)) может быть записана в форме

$$E(\vec{k}) = \pm |h_0| = \pm \frac{3ta_0}{2} \sqrt{k_x^2 + \beta_y^2}. \quad (6.1.11)$$

Такой приближенный результат, полученный путем разложения (6.1.4) в окрестности одной из двух долин в ряд Тейлора, в достаточно широкой области энергий согласуется с точным результатом. Это явно демонстрирует рис. 6.1.9. В рамках этого приближения изоэнергетические кривые являются окружностями, равномерно распределенными вокруг центра каждой долины в точке  $(0, +2\pi/3b)$  или  $(0, -2\pi/3b)$ .

Как велика запрещенная зона в полупроводниковой нанотрубке? Ответ на этот вопрос не зависит от конкретного типа нанотрубки, коль скоро число  $(m - n)$  не кратно трем и запрещенная зона не обращается в нуль. Найдем в качестве простого примера запрещенную зону в зигзагообразной нанотрубке. Из выражений (6.1.8), (6.1.10) и (6.1.11) можно получить, что

$$E(k_x) = \pm \frac{3ta_0}{2} \sqrt{k_x^2 + \left[ \frac{2\pi}{3b} \left( \frac{3\nu}{2m} - 1 \right) \right]^2}, \quad (6.1.12)$$

откуда следует, что величина запрещенной зоны для подзоны  $\nu$  может быть записана как разность энергий между ветвями  $+$  и  $-$  при  $k_x = 0$ :

$$E_{g,\nu} = 3ta_0 \frac{2\pi}{2mb} \left( \nu - \frac{2m}{3} \right).$$

Это выражение принимает минимальное значение, равное нулю, если  $\nu = 2m/3$ . Однако если число  $m$  не кратно трем, то минимальное значение разности  $(\nu - 2m/3)$  равно  $1/3$ . Тогда это означает, что минимальное значение ширины запрещенной зоны равно

$$E_g = ta_0 \frac{2\pi}{2mb} = \frac{2ta_0}{d} \approx \frac{0.8 \text{ эВ}}{d}, \quad (6.1.13)$$

где  $d$  равно диаметру нанотрубки, выраженному в нанометрах, так что величина  $\pi d$  есть длина периметра  $2mb$ .

## 6.2. Плотность состояний

В предыдущем параграфе мы обсуждали вопрос о том, каким образом эффекты размерного квантования изменяют дисперсионные зависимости  $E(\vec{k})$  и приводят к формированию подзон. Однако необходимо отметить, что по мере уменьшения размеров твердого тела подобные эффекты не проявляются внезапно, как будто массивный образец вдруг превращается в квантовую яму. В экспериментальных измерениях эффекты понижения размерности проявляются постепенно и их можно оценить, анализируя плотности состояний  $D(E)$ .

Плотность состояний определяет число состояний, приходящихся на единичный интервал энергии. Она явно зависит от вида дисперсионного соотношения  $E(\vec{k})$ . В качестве примера найдем плотность состояний в окрестности одной из долин зоны проводимости, где энергия является параболической функцией  $E(\vec{k})$  с некоторой эффективной массой  $m_c$ :

$$E(\vec{k}) = E_c + \frac{\hbar^2 k^2}{2m_c}. \quad (6.2.1)$$

Чему равна соответствующая плотность состояний, если вектор  $\vec{k}$  может изменяться только в одном направлении (квантовая проволока), в двух направлениях (квантовая яма) или в трех направлениях (массивное твердое тело)? Рассмотрим прямоугольное твердое тело с размерами  $L_x$ ,  $L_y$ ,  $L_z$  и периодическими граничными условиями (см. рис. 2.1.4) во всех трех направлениях (сравните с выражением (2.1.17)):

$$k_x = \frac{2\pi}{L_x} \nu_x, \quad k_y = \frac{2\pi}{L_y} \nu_y, \quad k_z = \frac{2\pi}{L_z} \nu_z, \quad (6.2.2)$$

где  $\nu_x$ ,  $\nu_y$  и  $\nu_z$  — целые числа. Далее предположим, что размер потенциальной ямы настолько велик, что разрешенные значения  $k$  фактически являются непрерывными, и суммирование следует заменить интегрированием:

$$\sum_{k_x} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_x}{2\pi/L_x}, \quad \sum_{k_y} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_y}{2\pi/L_y}, \quad \sum_{k_z} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_z}{2\pi/L_z}. \quad (6.2.3)$$

Другими словами, разрешенные состояния в  $k$ -пространстве распределены с плотностью  $L/2\pi$  на единичный интервал  $k$  вдоль каждого направления в  $k$ -пространстве. Как следствие, полное число разрешенных состояний  $N(k)$  вплоть до максимального значения  $k$  определяется выражениями

$$\frac{L}{2\pi} 2k = \frac{kL}{\pi} \quad 1D \text{ c } L \equiv L_x,$$

$$\frac{L_x L_y}{4\pi^2} \pi k^2 = \frac{k^2 S}{4\pi} \quad 2D \text{ c } S \equiv L_x L_y,$$

$$\frac{L_x L_y L_z}{8\pi^3} \frac{4\pi k^3}{3} = \frac{k^3 \Omega}{6\pi^2} \quad 3D \text{ c } \Omega \equiv L_x L_y L_z.$$

Используя дисперсионное соотношение (6.2.1) и зависимость  $N(k)$ , можно найти величину  $N(E)$ , имеющую смысл числа состояний с энергией, не превосходящей  $E$ . Производная от этой величины и определит плотность состояний:

$$D(E) = \frac{d}{dE} N(E). \quad (6.2.4)$$

Таблица 6.2.1. Результат расчета плотности состояний в одномерной, двумерной и трехмерной системах с параболическим изотропным законом дисперсии (уравнение (6.2.1)). На рисунках масса  $m_c = m$  (массе свободного электрона) и приведенные плотности состояний отвечают только одной проекции спина

1D

$$N(k) = \frac{L}{2\pi} 2k = \frac{kL}{\pi}$$

$$N(E) = \frac{L[2m_c(E - E_c)]^{1/2}}{\pi\hbar}$$

$$D(E) = \frac{m_c L}{\pi\hbar} \left( \frac{1}{2m_c(E - E_c)} \right)^{1/2}$$

2D

$$\frac{S}{4\pi^2} \pi k^2 = \frac{k^2 S}{4\pi}$$

$$\frac{S 2m_c(E - E_c)}{4\pi\hbar^2}$$

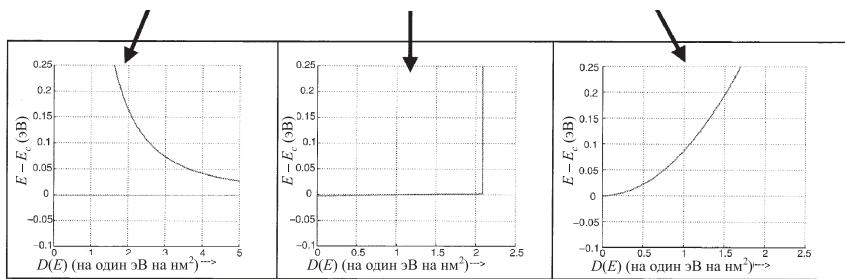
$$\frac{Sm_c}{2\pi\hbar^2}$$

3D

$$\frac{\Omega}{8\pi^3} \frac{4\pi k^3}{3} = \frac{k^3 \Omega}{6\pi^2}$$

$$\frac{\Omega [2m_c(E - E_c)]^{3/2}}{6\pi^2\hbar^3}$$

$$\frac{\Omega m_c}{2\pi^2\hbar^3} [2m_c(E - E_c)]^{1/2}$$



Сводка результатов для одного, двух и трех измерений приведена в табл. 6.2.1. Итак, параболический закон дисперсии  $E(\vec{k})$  приводит к плотности состояний, изменяющейся с энергией как  $E^{-1/2}$  в одномерной задаче, как  $E^0$  (т. е. постоянной) в двумерной задаче и как  $E^{1/2}$  в трехмерной задаче. Обратим, однако, внимание на то, что эти формулы не определяют полную плотность состояний в квантовой проволоке или квантовой яме. Они лишь описывают плотность состояний *только в одной подзоне* квантовой проволоки или квантовой ямы. Для того, чтобы получить полную плотность состояний, необходимо произвести суммирование по всем подзонам. К примеру, если в квантовой яме имеются подзоны с индексом  $p$

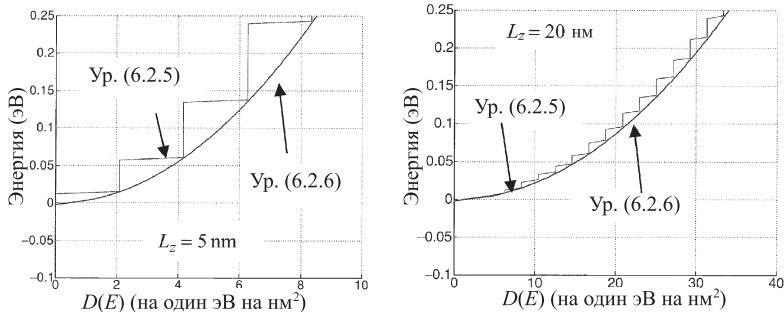


Рис. 6.2.1. Плотность состояний  $D(E)$  в двумерной потенциальной яме, найденная из выражения (6.2.5) (двумерная система с подзонами квантования), построенная для сравнения вместе с выражением для плотности состояний трехмерной системы (уравнение (6.2.6)). По мере увеличения ширины структуры  $L_z$  от 5 до 20 нм плотность состояний приближается к зависимости для трехмерной системы (эффективная масса в зоне проводимости  $m_c$  предполагается равной массе свободного электрона  $m$ )

(см. уравнение (6.1.1))

$$E_p(k_x, k_y) \approx E_c + p^2 \varepsilon_z + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m_c},$$

то плотность состояний будет выглядеть как суперпозиция нескольких двумерных плотностей состояний:

$$D(E) = \frac{m_c S}{2\pi\hbar^2} \sum_p \vartheta(E - E_c - p^2 \varepsilon_z). \quad (6.2.5)$$

На рис. 6.2.1 показана плотность состояний, найденная из выражения (6.2.5) при  $\varepsilon_z = \pi^2 \hbar^2 / 2m_c L_z^2$  в соответствии с уравнением (6.1.2). Очевидно, что по мере увеличения ширины  $L_z$  от 5 до 20 нм плотность состояний приближается к объемной, где объем  $\Omega$  равен  $S L_z$ :

$$D_{3D}(E) = \frac{S m_c}{2\pi^2 \hbar^3} [2m_c(E - E_c)]^{1/2} L_z. \quad (6.2.6)$$

**От графита к нанотрубке.** Эволюция плотности состояний по мере изменения геометрии и переходе от листа графита к нанотрубке также представляет собой поучительный пример того, как эффекты размерного квантования проявляются при понижении размерности. Для листа

Таблица 6.2.2. Результат расчета плотности состояний, приходящейся на одну проекцию спина в одной долине

Графит, ур. (6.2.7)	Зигзагообразная нанотрубка, ур. (6.2.8)
$N(k) = \frac{S}{4\pi} k^2$	$\frac{L}{\pi} k_x$
$N(E) = \frac{SE^2}{4\pi a^2 t^2}$	$\sum_{\nu} \frac{L}{\pi at} \sqrt{E^2 - E_{nu}^2}$
$D(E) = \frac{SE}{2\pi a^2 t^2}$	$\sum_{\nu} \frac{L}{\pi at} \frac{E}{\sqrt{E^2 - E_{nu}^2}}$

графита зависимость  $E(\vec{k})$  в каждой из двух долин с центрами в точках  $(k_x a, k_y b) = (0, \pm 2\pi/3)$  можно аппроксимировать выражением (см. уравнение (6.1.11) с  $a = 3a_0/2$ )

$$E(\vec{k}) = \pm \frac{3ta_0}{2} \sqrt{k_x^2 + \beta} y^2 = \pm ta |\vec{k}|. \quad (6.2.7)$$

Как мы уже видели, дисперсионные зависимости в подзонах для зигзагообразной нанотрубки определяются выражением (см. уравнение (6.1.12))

$$E(k_x) = \pm ta \sqrt{k_{nu}^2 + k_x^2} = \pm \sqrt{E_{\nu}^2 + (tak_x)^2}, \quad (6.2.8)$$

где

$$k_{\nu} \equiv \frac{2\pi}{3b} \left( \frac{3\nu}{2m} - 1 \right) \quad \text{и} \quad E_{\nu} \equiv tak_{\nu}.$$

Как ясно из табл. 6.2.2, этапы расчета плотности состояний  $D(E)$  здесь те же, что и моделях табл. 6.2.1, хотя дисперсионная зависимость иная (см. (6.2.7), (6.2.8)). Результаты, представленные на рис. 6.2.2, позволяют сравнить рассчитанные плотности состояний зигзагообразной нанотрубки, имеющей длину  $L$  и диаметр  $d$  (отметим, что периметр равен  $\pi d = 2mb$ ):

$$D_{ZNT}(E) = \sum_{\nu} \frac{2L}{\pi at} \frac{E}{\sqrt{E^2 - E_{\nu}^2}}, \quad \text{где} \quad E_{\nu} \equiv \frac{2at}{d} \left( \nu - \frac{2m}{3} \right), \quad (6.2.9)$$

с плотностью состояний в слое графита с площадью  $\pi Ld$ , которая равна

$$D_G(E) = \frac{Ld}{a^2 t^2} |E|, \quad (6.2.10)$$

для нанотрубки с  $m = 200$  (что отвечает  $d = 15.4$  нм) и для нанотрубки с  $m = 800$  ( $d = 61.7$  нм).

Очевидно, что плотность состояний в нанотрубке с меньшим радиусом сильно отличается от плотности состояний в графите. При этом в массивной нанотрубке различия заметнее, поскольку в реальных экспериментах всегда присутствует температурное уширение функции распределения. Масштаб этого уширения при комнатной температуре порядка  $k_B T \sim 0.026$  эВ (эта проблема обсуждается ниже в связи с рис. 7.3.4).

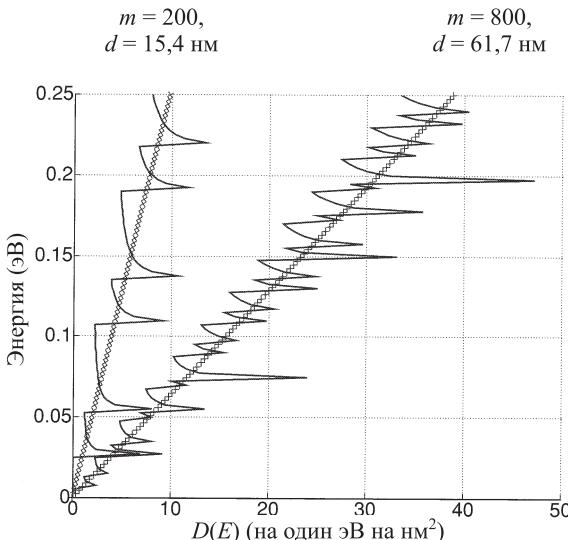


Рис. 6.2.2. Плотность состояний в зигзагообразной нанотрубке, полученная из выражения (6.2.9) (сплошные линии). Для сравнения приведена плотность состояний в графите (уравнение (6.2.10)), представленная линиями из крестиков

Аналитическое выражение для плотности состояний в зигзагообразной нанотрубке нетрудно получить, заменив суммирование по  $\nu$  интегрированием (сравните с уравнением (6.2.9)):

$$D_{ZNT}(E) \approx \int 2d\nu \frac{2L}{\pi at} \frac{|E|}{\sqrt{E^2 - E_{nu}^2}}, \quad \text{где} \quad dE_{nu} \equiv \frac{2at}{d} d\nu,$$

$$= \int_0^E dE_\nu \frac{2Ld}{\pi a^2 t^2} \frac{|E|}{\sqrt{E^2 - E_{nu}^2}} = \frac{Ld}{a^2 t^2} |E|,$$

что совпадает с плотностью состояний в графите (сравните с выражением (6.2.10)).

**Анизотропная дисперсионная зависимость.** Мы рассмотрели несколько примеров, иллюстрирующих процедуру расчета плотности состояний  $D(E)$  при известном законе дисперсии  $E(\vec{k})$ . Этой процедурой (см. табл. 6.2.1 и 6.2.2) удобно пользоваться в том случае, когда закон дисперсии  $E(\vec{k})$  является изотропным. Детали расчета, однако, усложняются, если дисперсионная зависимость является анизотропной. К примеру, у кремния имеется шесть отдельных долин в зоне проводимости, каждой из которых соответствуют эллипсоидальные изоэнергетические поверхности:

$$E = E_c + \frac{\hbar^2 k_x^2}{2m_{xx}} + \frac{\hbar^2 k_y^2}{2m_{yy}} + \frac{\hbar^2 k_z^2}{2m_{zz}}. \quad (6.2.11)$$

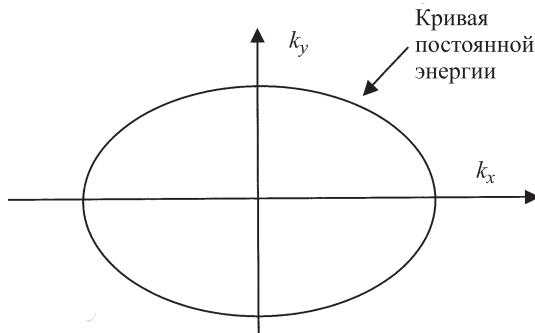


Рис. 6.2.3

При данном значении энергии  $E$  множество точек, отвечающих постоянной энергии (рис. 6.2.3), является эллипсоидом, чьи главные оси равны  $\sqrt{2m_{xx}(E - E_c)/\hbar}$ ,  $\sqrt{2m_{yy}(E - E_c)/\hbar}$  и  $\sqrt{2m_{zz}(E - E_c)/\hbar}$ . Объем этого эллипса равен

$$\frac{4\pi}{3} \frac{\sqrt{2m_{xx}(E - E_c)}}{\hbar} \frac{\sqrt{2m_{yy}(E - E_c)}}{\hbar} \frac{\sqrt{2m_{zz}(E - E_c)}}{\hbar},$$

поэтому справедлива зависимость

$$N(E) = \frac{\Omega}{8\pi^3} \frac{4\pi}{3} \frac{\sqrt{2m_{xx}(E - E_c)}}{\hbar} \frac{\sqrt{2m_{yy}(E - E_c)}}{\hbar} \frac{\sqrt{2m_{zz}(E - E_c)}}{\hbar},$$

а плотность состояний определяется выражением

$$D(E) \equiv \frac{d}{dE} N(E) = \frac{\Omega}{2\pi^2 \hbar^3} \sqrt{2m_{xx} m_{yy} m_{zz}} (E - E_c), \quad (6.2.12)$$

которое в случае изотропной массы сводится к полученному ранее результату (см. табл. 6.2.1).

**Формальное выражение для плотности состояний.** Если в общем случае мы имеем систему, состояния которой нумеруются индексом  $\alpha$ , то полное число состояний  $N_T(E)$ , энергия которых не превосходит  $E$ , записывается в виде

$$N_T(E) = \sum_{\alpha} \vartheta(E - E_{\alpha}), \quad (6.2.13)$$

где  $\vartheta(E)$  обозначает единичную ступенчатую функцию, равную нулю при  $E < 0$  и единице при  $E > 0$ . Производной от этой функции является дельта-функция, поэтому

$$D(E) \equiv \frac{d}{dE} N_T(E) = \sum_{\alpha} \delta(E - E_{\alpha}). \quad (6.2.14)$$

Можно сказать, что это выражение определяет скорее последовательность пиков, нежели непрерывную функцию. Формально мы можем получить непрерывную плотность состояний из выражения (6.2.14), представив, что размер системы становится очень большим, и заменив суммирование интегрированием. К примеру, если

$$E(\vec{k}) = E_c + \frac{\hbar^2 k^2}{2m^*},$$

то

$$\begin{aligned} D(E) &= \sum_{\vec{k}} \delta(E - \varepsilon_{\vec{k}}) = \\ &= \sum_{k_x} \sum_{k_y} \sum_{k_z} \delta \left( E - E_c - \frac{\hbar^2 (k_x^2 + k_y^2 + k_z^2)}{2m^*} \right), \end{aligned}$$

где

$$k_x = \frac{2\pi}{L_x} \nu_x, \quad k_y = \frac{2\pi}{L_y} \nu_y, \quad k_z = \frac{2\pi}{L_z} \nu_z,$$

и числа  $\nu_x$ ,  $\nu_y$  и  $\nu_z$  являются целыми. Далее, если объем системы становится очень большим, то суммирование по этим индексам заменяется

интегрированием:

$$\sum_{k_x} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_x}{2\pi/L_x}, \quad \sum_{k_y} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_y}{2\pi/L_y}, \quad \sum_{k_z} \rightarrow \int_{-\infty}^{+\infty} \frac{dk_z}{2\pi/L_z},$$

$$D(E) = \frac{\Omega}{8\pi^3} \int_0^{+\infty} dk_x dk_y dk_z \delta E - E_c - \frac{\hbar^2 k^2}{2m^*} =$$

$$= \frac{\Omega}{8\pi^3} \int_0^{+\infty} 4\pi k^2 dk \delta E - E_c - \frac{\hbar^2 k^2}{2m^*} =$$

$$= \frac{\Omega}{8\pi^3} \int_0^{+\infty} \frac{m^* dE}{\hbar^2} \frac{\sqrt{2m^*(E - E_c)}}{\hbar} \delta E - E_c - \frac{\hbar^2 k^2}{2m^*} =$$

$$= \frac{\Omega m^*}{2\pi^2 \hbar^3} [2m^*(E - E_c)]^{1/2},$$

в соответствии с полученным ранее (см. выражение для плотности состояний в табл. 6.2.1). В данной процедуре требуются несколько более глубокие математические навыки, связанные с интегрированием дельта-функцией.

Значение формального выражения (6.2.14) определяется тем, что оно верно в общем случае, независимо от вида дисперсионной кривой  $E(\vec{k})$ , или даже в отсутствии такой кривой. Разумеется, в общем случае замена суммирования интегрированием не является простой, поскольку расположение уровней энергии может не описываться простым выражением вида (6.2.2). Тем не менее, непрерывную плотность состояний можно рассчитать, исходя из выражения (6.2.14) и рассматривая уширение каждой дельта-функции в кривую Лоренца (см. уравнение (1.3.2)):

$$\delta(E - \varepsilon_\alpha) \rightarrow \frac{\gamma/2\pi}{(E - \varepsilon_\alpha)^+(\gamma/2)^2}. \quad (6.2.15)$$

Если расстояние между соседними значениями энергии  $\varepsilon_\alpha$  меньше величины уширения  $\gamma$ , что обычно справедливо для массивных систем, то плотность состояний можно считать непрерывной функцией энергии.

**Задачи с разделяющимися переменными.** Если задача на собственные значения допускает разделение переменных, то с использованием выражения (6.2.14) полная плотность состояний определяется как свертка (конволюция) плотностей состояний по каждой переменной. Пусть переменные  $x$  и  $y$  разделяются (как это демонстрирует выражение (2.3.6) в трехмерной задаче), так что полная энергия уровней является суммой энергий, отвечающих переменным  $x$  и  $y$ :

$$\varepsilon(n, m) = \varepsilon_x(n) + \varepsilon_y(m). \quad (6.2.16)$$

Плотность состояний для переменных  $x$  и  $y$  можно рассматривать отдельно:

$$D_x(E) = \sum_n \delta(E - \varepsilon_x(n)), \quad (6.2.17a)$$

$$D_y(E) = \sum_m \delta(E - \varepsilon_y(m)) \quad (6.2.17b)$$

и можно непосредственно показать, что полная плотность состояний

$$D(E) = \sum_n \sum_m \delta(E - \varepsilon_x(n) - \varepsilon_y(m)) \quad (6.2.17c)$$

может быть записана как свертка (конволюция) плотностей состояний для переменных  $x$  и  $y$ :

$$D(E) = \int_{-\infty}^{+\infty} dE' D_x(E') D_y(E - E'). \quad (6.2.18)$$

### 6.3. Минимальное сопротивление проволоки

Обсудив понятие энергетической подзоны, можно попытаться ответить на следующий чрезвычайно интересный, фундаментальный вопрос. Рассмотрим проволоку с сечением  $L_y L_z$ , вдоль которой приложено напряжение  $V$  (рис. 6.3.1). Чему равна проводимость этой проволоки, если контакты считаются идеальными и длина проволоки непрерывно уменьшается? Исходя из закона Ома, можно было бы сказать, что проводимость может возрастать неограниченно по мере уменьшения длины проволоки, поскольку сопротивление (величина, обратная проводимости) пропорциональна длине. Однако, как я отмечал в параграфе 1.3, максимальная проводимость  $G = I/V$  проводника с одним энергетическим уровнем является

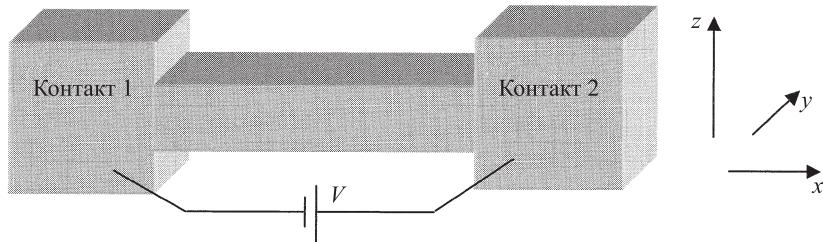


Рис. 6.3.1. Проволока с поперечным сечением  $L_y L_z$ , вдоль которой через массивные контакты приложено напряжение  $V$

фундаментальной величиной, равной

$$G_0 \equiv q^2/h = 38.7 \mu S = (25.8 k\Omega)^{-1}. \quad (6.3.1)$$

Теперь мы готовы обобщить это понятие для проволоки с конечным поперечным сечением. Экспериментально было установлено, что по мере уменьшения длины проволоки до того момента, когда электрон сможет пересечь ее без рассеяния (режим баллистического транспорта), ее проводимость стремится к постоянному значению, определяемому в предположении «идеальных» контактов величиной

$$G = [M(E)]_{E=E_F} G_0, \quad (6.3.2)$$

где  $M(E)$  равно числу «мод», или подзон с энергией  $E$ . Фактическое число мод  $M(E)$  при данной энергии зависит от деталей структуры проволоки, но максимальное значение проводимости каждой моды равно  $G_0$  независимо от этих деталей. Это можно показать следующим образом.

**Максимальный ток в проволоке с одной модой.** Рассмотрим одну моду, или подзону, с номером  $\nu$  и дисперсионной зависимостью  $E_\nu(k_x)$ . Ток, переносимый состояниями с положительной групповой скоростью, может быть записан в виде

$$I = \frac{-q}{L} \sum_{\nu_x(k_x) > 0} \nu_x(k_x).$$

Ранее мы видели, что для свободных электронов с энергией  $E = \hbar^2 k^2 / 2m$  скорость равна  $\hbar k / m$ , т. е. это импульс  $\hbar k$ , деленный на массу  $m$ . Но как определить скорости электронов в периодическом твердом

теле с некоторым дисперсионным соотношением  $E_\nu(\vec{k})$ ? Ответ на этот вопрос требует тщательного обсуждения, которое мы временно отложим (см. ниже параграф 6.4), а пока просто постулируем, что правильное значение скорости равно групповой скорости, определяемой как градиент дисперсионного соотношения:

$$\hbar \vec{v}(\vec{k}) = \vec{\nabla}_k E_\nu(\vec{k}),$$

так что в нашем одномерном примере можно записать  $\nu_x(k_x) = \frac{\partial E_\nu(k_x)}{\hbar \partial k_x}$ , и ток будет определяться выражением

$$\begin{aligned} I &= \frac{-q}{L} \sum_{\nu(k_x) > 0} \frac{1}{\hbar} \frac{\partial E_\nu(k_x)}{\partial k_x} = \\ &= -q \int \frac{dk_x}{2\pi} \frac{1}{\hbar} \frac{\partial E_\nu(k_x)}{\partial k_x} = \frac{-q}{h} \int dE_\nu, \end{aligned} \quad (6.3.3)$$

которое показывает, что каждая мода в проволоке на единицу энергии переносит ток, равный  $(q/h)$ . В равновесии полный ток равен нулю, поскольку состояния с положительной и отрицательной скоростью заполнены одинаково. Приложенное напряжение  $V$  изменяет заселенность уровней в интервале  $E_F \pm (qV/2)$ , создавая неравновесную ситуацию, детали которой зависят от природы связи с контактами. Но независимо от всех деталей очевидно, что максимальный полный ток будет течь в случае, когда состояния с положительной скоростью будут заполнены вплоть до энергии  $\nu_1 = E_F + (qV/2)$ , в то время как состояния с отрицательной скоростью будут заполнены до уровня  $\nu_2 = E_F - (qV/2)$ , так что в области энергий

$$E_F - (qV/2) < E < E_F + (qV/2)$$

остаются заполненными лишь состояния с положительной скоростью (рис. 6.3.2). Воспользовавшись рис. 6.3.2, мы можем записать ток, переносимый этими состояниями, относящимися к моде  $\nu$ , как

$$I = \frac{-q}{h} (\nu_1 - \nu_2) \rightarrow \frac{q^2}{h} V,$$

так, как было отмечено ранее, максимальная проводимость  $\nu$ -й моды равна  $q^2/h$  (см. выражение (6.3.1)). Обратите внимание на то, что этот результат является независимым от формы дисперсионного соотношения  $E_\nu(k_x)$  для данной  $\nu$ -й моды.

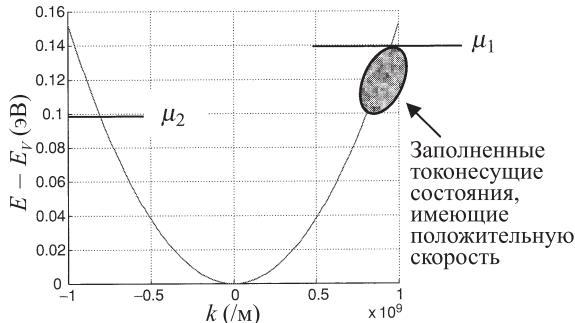


Рис. 6.3.2

**Число мод.** Количество мод  $M(E)$ , которые могут распространяться при заданной энергии, существенно зависит от условий конкретной задачи. К примеру, если соответствующий интервал энергий включает дно зоны проводимости, закон дисперсии является параболическим (см. рис. 6.1.2 и уравнение (6.1.1)). В этом случае подзоны могут быть пронумерованы двумя индексами  $(n, p)$ , как это показано на рис. 6.3.3:

$$E_{n,p}(k_x) \approx E_c + n^2 \varepsilon_y + p^2 \varepsilon_z + \frac{\hbar^2 k_x^2}{2m_c}, \quad (6.3.4)$$

где  $\varepsilon_y = \pi^2 \hbar^2 / 2m_c L_y^2$  и  $\varepsilon_z = \pi^2 \hbar^2 / 2m_c L_z^2$  в предположении, что электроны удерживаются в проволоке границами с бесконечно высокими стенками ширины  $L_y$  и  $L_z$  в направлениях  $y$  и  $z$  соответственно. В этом случае плотность мод  $M(E)$  имеет вид, изображенный на рис. 6.3.3, возрастаая с ростом энергии на одну единицу каждый раз, когда заполняется новая подзона.

Структура подзон валентной зоны более сложная, поскольку она сформирована из нескольких взаимодействующих зон, имеющих сложные дисперсионные соотношения. Для простых оценок мы можем принять простое приближение *обращенной вниз* параболы

$$h(\vec{k}) = E_\nu - \frac{\hbar^2 k_x^2}{2m_V}. \quad (6.3.5)$$

Далее мы получим дисперсионные зависимости в различных подзонах, которые имеют противоположный знак:

$$E_{n,p}(k_x) \approx E_V - n^2 \varepsilon_y - p^2 \varepsilon_z - \frac{\hbar^2 k_x^2}{2m_V}, \quad (6.3.6)$$

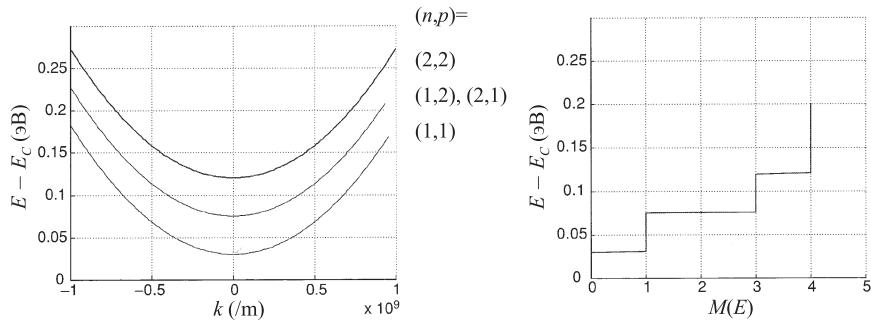


Рис. 6.3.3. Дисперсионные кривые для четырех нижних подзон зоны проводимости (см. уравнение (6.3.4)) в проволоке квадратного сечения с  $L_y = L_z = 10$  нм и  $m_c = 0.25$  м

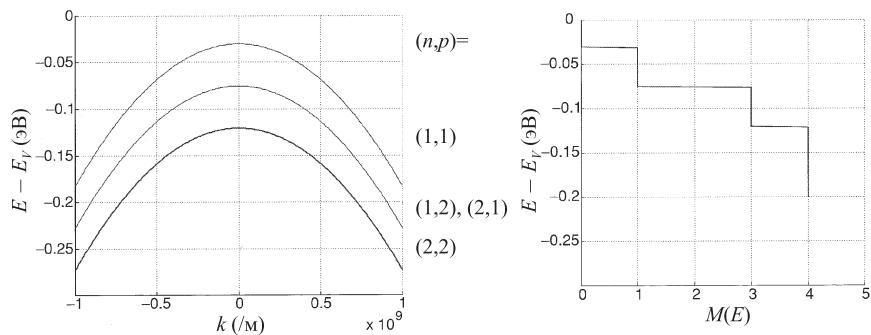


Рис. 6.3.4. Дисперсионные кривые для четырех нижних подзон валентной зоны (см. уравнения (6.3.5) и (6.3.6)) в проволоке квадратного сечения с  $L_y = L_z = 10$  нм и  $m_V = 0.25$  м

отвечающие плотности мод  $M(E)$ , возрастающей с убыванием энергии так, как это показано на рис. 6.3.4.

**Минимальное сопротивление широкого проводника.** Мы видели, что в проволоке минимальное сопротивление, приходящееся на одну моду, равно  $h/q^2$ . Отсюда следует, что существует минимальное сопротивление, которое может иметь любой проводник. К примеру, мы могли бы рассматривать полевой транзистор (FET) как двумерный проводник ширины  $W$

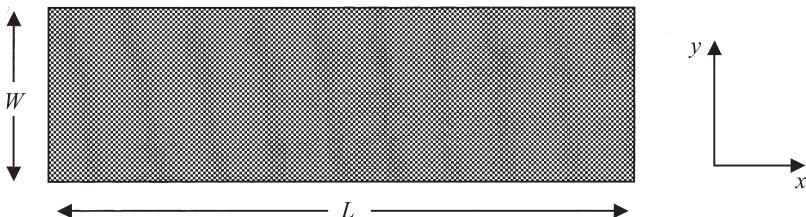


Рис. 6.3.5

и длины  $L$ , как это показано на рис. 6.3.5. Тогда, по мере того как длина проводника  $L$  уменьшается, он будет приближаться по своим свойствам к баллистическому проводнику, чье сопротивление, в основном, определяется контактами. Этот факт хорошо известен, и инженеры-приборостроители много работают над технологией контактов, минимизирующих их сопротивление. Что остается до конца не ясным, так это вопрос о существовании фундаментального минимума сопротивления контактов, достижимого для их лучших образцов.

Для оценки минимального сопротивления в контактах предположим, что имеется полупроводник  $n$ -типа, проводимость в котором осуществляют электроны зоны проводимости. Энергетический спектр считается параболическим, подобным закону дисперсии в выражении (6.3.4). Предполагая, что проводник имеет лишь одну подзону вдоль направления  $z$ , мы можем записать электронную плотность состояний на единицу площади при температуре  $T = 0$  К в виде (см. табл. 6.2.1)

$$n_s = \frac{m_c}{\pi \hbar^2} (E_F - E_c). \quad (6.3.7)$$

Максимальная проводимость в этом случае определяется выражением

$$G_{\max} = \frac{2q^2}{h} \text{Int} \sqrt{\frac{E_F - E_c}{\hbar^2 \pi^2 / 2m_c W^2}} \approx W \frac{2q^2}{h} \sqrt{\frac{2n_s}{\pi}},$$

где мы использовали уравнение (6.3.7) и функцию  $\text{Int}(x)$ , означающую ближайшее целое число, меньшее  $x$ . Минимальное сопротивление определяется величиной, обратной к  $G_{\max}$ :

$$R_{\min} W = \frac{h}{2q^2} \sqrt{\frac{\pi}{2n_s}} = \frac{16.28 k\Omega}{\sqrt{n_s}}. \quad (6.3.8)$$

При концентрации носителей  $n_s = 10^{12}/\text{см}^2$  это выражение дает величину  $R_{\min}W \approx 160 \Omega \mu\text{m}$ . Обратите внимание на то, что мы предположили наличие лишь одной зоны (подзоны), сформированной при квантовании движения вдоль направления  $z$ . В кремнии, где зона проводимости состоит из шести долин, возникает шесть наборов подзон, причем две нижние являются вырожденными. Это означает, что если размер  $L_z$  является достаточно малым, то из-за квантования движения вдоль оси  $z$  будут сформированы две вырожденных подзоны и соответствующее минимальное сопротивление контактов будет равно половине значения, полученного из нашей оценки, справедливой при наличии одной подзоны квантования вдоль  $z$ :  $R_{\min}W \approx 80 \Omega \mu\text{m}$ .

## 6.4. Скорость электрона в (под)зоне

Электроны в периодических структурах имеют энергетический спектр, состоящий из подзон, а соответствующие волновые функции имеют вид

$$\{\psi\}_n = \{\psi\}_0 \exp(\pm ikx_n), \quad (6.4.1)$$

где  $x_n = na$  обозначает узлы дискретной решетки, а  $\{\psi\}_n$  есть вектор-столбец размера  $(B \times 1)$ , где  $B$  равно числу базисных функций, описывающих элементарную ячейку (рис. 6.4.1).

Можно сказать, что такие волновые функции на масштабе элементарной ячейки ведут себя как атомные волновые функции, а на масштабах кристаллической решетки, в целом, их природа напоминает волновые функции свободного электрона, или плоские волны вида  $\sim \exp(\pm ikx)$ . Данные волновые функции часто называют блоховскими. При первом обсуждении уравнения Шредингера я упомянул, что электрону в состоянии, описываемом плоской волной  $\exp(\pm ikx)$ , отвечает скорость  $\hbar k/m$ , поскольку в этом состоянии поток вероятности равен этой скорости, умноженной на плотность вероятности (уравнение (2.1.20)). Чему же равна скорость электрона в состоянии, описываемом блоховской волной?

Вероятное предположение состоит в том, что скорость должна быть равна  $\hbar k/m^*$ , где  $m^*$  есть эффективная масса. Это и в самом деле верно в определенных условиях, но дело в том, что не все зоны или подзоны могут быть описаны в приближении эффективной массы, особенно в низкоразмерных структурах, таких как квантовые проволоки. Кроме того, не очевиден скалярный характер скорости, описываемой единственным числом. В общем случае нам придется работать с матрицей скоростей размера  $B \times B$ .

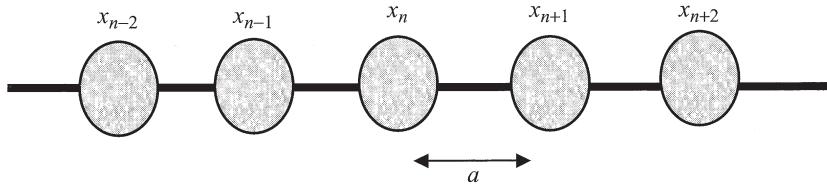


Рис. 6.4.1

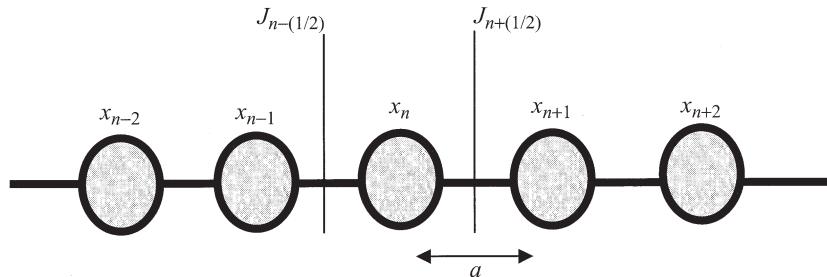


Рис. 6.4.2

**Матрица скоростей.** Чтобы установить вид матрицы скоростей, мы начнем с замечания о том, что гамильтониан квантовой проволоки может быть записан в блочном трехдиагональном виде

$$H_{\text{квантовой проволоки}} = \begin{pmatrix} \alpha & \beta & 0 & 0 & 0 & \dots \\ \beta^+ & \alpha & \beta & 0 & 0 & \dots \\ 0 & \beta^+ & \alpha & \beta & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad (6.4.2)$$

которому можно сопоставить одномерный массив единичных ячеек, в каждой из которых имеется очень большое число базисных функций  $B$ , равное числу атомов в поперечном сечении, умноженном на число базисных функций, приходящихся на один атом. Нестационарное уравнение Шредингера для квантовой проволоки может быть записано в виде

$$i\hbar \frac{d\psi_n}{dt} = \alpha\psi_n + \beta\psi_{n+1} + \beta^+\psi_{n-1}, \quad (6.4.3)$$

где  $\psi$  есть вектор-столбец размера  $(B \times 1)$ , а величины  $\alpha$  и  $\beta$  являются матрицами размера  $(B \times B)$ .

Чтобы получить правильное выражение для тока, необходимо рассмотреть скорость изменения во времени полной плотности вероятности в конкретной элементарной ячейке, к примеру, в  $n$ -й. Плотность вероятности может быть записана как  $\psi_n^+ \psi_n$ , что является матрицей  $(1 \times 1)$ , или скаляром, равным сумме квадратов модулей всех  $B$  компонент  $\psi$ -функции. Обратите внимание на то, что величина  $\psi_n \psi_n^+$  не подходит для этой цели, поскольку является матрицей порядка  $(B \times B)$ . Воспользовавшись уравнением (6.4.3), можно найти производную по времени от величины  $\psi_n^+ \psi_n$  в виде

$$i\hbar \frac{d}{dt} \psi_n^+ \psi_n = \psi_n^+ \beta \psi_{n+1} + \psi_n^+ \beta^+ \psi_{n-1} - \psi_{n-1}^+ \beta \psi_n - \psi_{n+1}^+ \beta^+ \psi_n. \quad (6.4.4)$$

Посмотрим, удастся ли нам записать правую часть в виде разности токов, текущих влево и вправо (см. рис. 6.4.2):

$$\frac{d}{dt} \psi_n^+ \psi_n = \frac{J_{n-(1/2)} - J_{n+(1/2)}}{a}, \quad (6.4.5)$$

что может быть мотивировано континуальной формой уравнения непрерывности, имеющего вид  $dn/dt = -dJ/dx$ . Непосредственно можно проверить, что если мы определим токи как

$$J_{n+(1/2)} \equiv \frac{\psi_{n+1}^+ \beta^+ \psi_n - \psi_n^+ \beta \psi_{n+1}}{i\hbar} a, \quad (6.4.6)$$

то уравнение (6.4.5) станет эквивалентно уравнению (6.4.4). Мы можем использовать уравнение (6.4.1) для электрона в (под)зоне, чтобы выразить  $\psi_{n+1}$  в уравнении (6.4.4) через  $\psi_n$ , и получить

$$J_{n+(1/2)} \equiv \psi_n^+ \frac{a}{i\hbar} [\beta^+ \exp(-ika) - \beta \exp(+ika)] \psi_n, \quad (6.4.7)$$

что наводит на мысль определить скорость электрона как

$$[v(k)] \equiv \frac{a}{\hbar} [\beta^+ \exp(-ika) - \beta \exp(+ika)], \quad (6.4.8)$$

т. е. в виде матрицы скоростей для состояния в электронной (под)зоне при данном значении  $k$ . Теперь, когда мы завершили математическую часть вывода, попробуем разобраться, в чем его физический смысл.

Отметим в первую очередь, что собственные значения энергии получаются при диагонализации этой матрицы (что следует из основного результата (5.2.4), если обратить внимание на расчет зонной структуры):

$$[h(k)] = \alpha + [\beta \exp(+ika) + \beta^+ \exp(-ika)]. \quad (6.4.9)$$

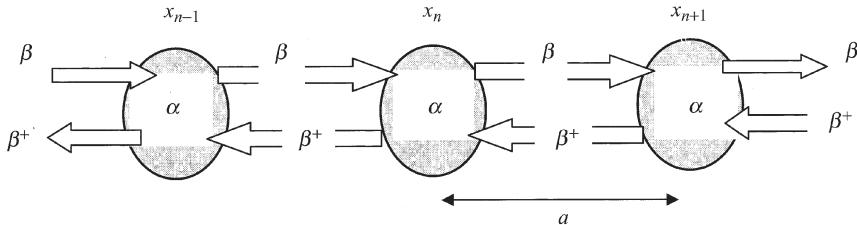


Рис. 6.4.3

Можно непосредственно проверить, что эта гамильтоновская матрица связана с найденной матрицей скоростей (уравнение (6.4.8)) соотношением

$$[\nu(k)] = \frac{1}{\hbar} \frac{d}{dk} [h(k)]. \quad (6.4.10)$$

Это означает, что если обе матрицы ( $[h]$  и  $[\nu]$ ) диагонализируются одновременно, то их собственные значения действительно будут подчиняться уравнению

$$\hbar \vec{\nu}(\vec{k}) = \vec{\nabla} E_\nu(\vec{k}),$$

которое мы использовали для вывода максимального значения тока, переносимого одной подзоной (см. уравнение (6.3.3)).

Можно ли диагонализовать матрицы  $[h]$  и  $[\nu]$  одновременно? Исходя из определений этих матриц (см. уравнения (6.4.8) и (6.4.9)), нетрудно заметить, что ответ является положительным в случае, когда матрицы  $\alpha$  и  $\beta$  могут быть диагонализованы одновременно. Можно проиллюстрировать трехдиагональную структуру гамильтониана в уравнении (6.4.2) на примере одномерного массива, в котором каждая из матриц  $\alpha$  и  $\beta$  имеет размер  $(B \times B)$ , где  $B$  есть число базисных функций, необходимых для того, чтобы описать сечение проволоки (см. рис. 6.4.3). Теперь, если мы сможем диагонализовать одновременно матрицы  $\alpha$  и  $\beta$ , то квантовую проволоку рассматривать как совокупность независимых подзон, проводящих ток в параллельном режиме, как это показано на рис. 6.4.4.

В том же представлении будет диагональной и матрица скоростей, чьи диагональные элементы окажутся равными скоростям каждой моды. Хороший пример структуры, где такая операция возможна, дает простая квадратная решетка, показанная на рис. 6.4.5 а. В этом случае мы можем рассматривать систему как одномерный массив вида, представленного на рис. 6.4.5 б.

Матрицы  $[\beta]$  и  $[\beta^+]$  описывают связь между одним из столбцов атомов с атомами следующего столбца, и могут быть записаны как единичная матрица  $[I]$ . Это означает, что, как только найдено представление, диагонализующее матрицу  $[\alpha]$ , матрица  $[\beta]$  в этом представлении также окажется диагональной, поскольку единичная матрица остается неизменной при преобразовании базиса. Следовательно, всегда можно диагонализовать матрицы  $[\alpha]$ ,  $[\beta]$ ,  $[h]$  и  $[\nu]$  одновременно и в этом представлении рассматривать квантовую проволоку как состоящую из независимых одномодовых проволок, соединенных параллельно.

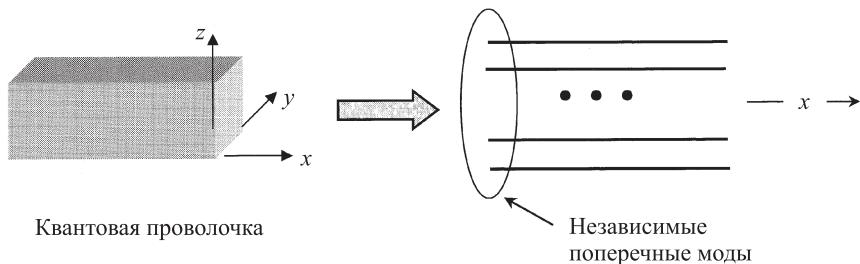


Рис. 6.4.4

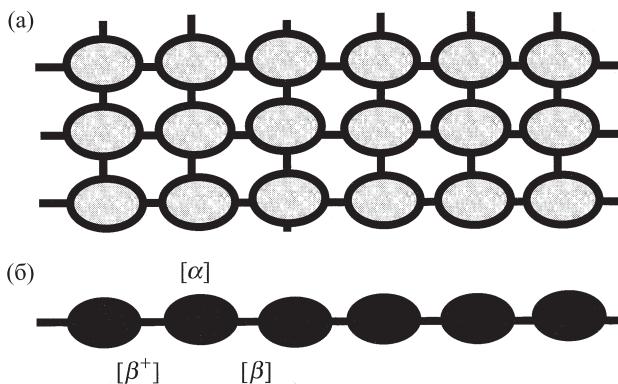


Рис. 6.4.5

**Плотность мод.** В предыдущем параграфе мы видели, что плотность мод  $M(E)$  играет в физике квантовых проволок важную роль, аналогичную

плотности состояний  $D(E)$  в объемных полупроводниках. Если проволока длины  $L$  имеет состояния, описываемые двумя индексами  $(\alpha, k)$ , то мы можем формально определить плотность мод как

$$M(E) = \sum_{\alpha} \sum_k \delta(E - \varepsilon_{\alpha,k}) \frac{\pi}{L} \frac{\partial \varepsilon_{\alpha,k}}{\partial k}. \quad (6.4.11)$$

Эту величину нетрудно вычислить, если собственные значения разделяются на сумму слагаемых с разными индексами:  $\varepsilon_{\alpha,k} = \varepsilon'_{\alpha} + \varepsilon''_k$ . Такая ситуация возможна лишь, если подзоны не взаимодействуют, как это обсуждалось выше.

## Упражнения

**У.6.1.** Используя выражение (6.2.4) и соотношение  $\hbar\nu = \partial E / \partial k$ , покажите, что для любого одномерного проводника имеет место формула  $D_{1D}(E) = L / \pi \hbar\nu$ . Этот результат имеет общий характер и не зависит от вида дисперсионного соотношения  $E(k)$ .

**У.6.2.** Постройте дисперсионную зависимость для двух нижних подзон зигзагообразной нанотрубки с параметрами  $m = 66$  и  $m = 65$  и сравните результат с рис. 6.1.7.

**У.6.3.** Сравните дисперсионные зависимости для двумерного листа графита из выражений (6.1.4) и (6.1.11), показанные на рис. 6.1.9.

**У.6.4.** Сравните трехмерную плотность состояний с плотностью состояний в двумерной потенциальной яме толщиной (а) 5 нм и (б) 20 нм, показанной на рис. 6.2.1.

**У.6.5.** Сравните плотность состояний в графите с плотностью состояний в зигзагообразной нанотрубке диаметра (а) 15.4 нм и (б) 61.7 нм, показанной на рис. 6.2.2.

**У.6.6.** Постройте дисперсионные зависимости для энергии в четырех нижних подзонах зоны проводимости (см. уравнение (6.3.4)) в проволоке прямоугольного сечения с  $L_y = L_z = 10$  нм, эффективной массой  $m_c = 0.25 m$  и сравните результат с рисунком 6.3.3.

**У.6.7.** Постройте дисперсионные зависимости для энергии в четырех нижних подзонах валентной зоны (см. уравнения (6.3.5) и (6.3.6)) в проволоке прямоугольного сечения с  $L_y = L_z = 10$  нм, эффективной массой  $m_V = 0.25m$  и сравните результат с рис. 6.3.4.

# Глава 7

## Емкость

В первой главе я упоминал, что для полного квантовомеханического описания транспорта в системе со многими уровнями необходимо заменить параметры одноуровневой модели на соответствующие матрицы. Важнейшей из физических величин является гамильтоновская матрица  $[H]$ , определяющая уровни энергии, с которой мы уже достаточно хорошо познакомились. Выводы, которые будут сделаны в этой главе, носят предварительный характер, поскольку здесь обсуждаются лишь *равновесные* задачи, используются только матрицы  $[H]$  и не учитывается уширение уровней.

В этой главе обсуждаются следующие проблемы. Нам предстоит выяснить, как электронная плотность внутри транзистора зависит от напряжения на затворе  $V_G$  в предположении, что исток и сток имеют одинаковый потенциал (потенциал стока  $V_D = 0$ , см. рис. 7.1). Строго говоря, это тоже неравновесная проблема, поскольку затвор не находится в равновесии с контактами истока и стока (которые находятся в равновесии друг с другом). Однако фактически диэлектрическая прослойка изолирует канал от затвора. Поэтому канал находится в равновесии с контактами истока и стока, имеющими одинаковый электрохимический потенциал  $\mu_1 = \mu_2 \equiv \mu$ . Матрица плотности, чьи диагональные элементы в координатном представлении определяют электронную плотность  $n(\vec{r})$ , определяется выражением

$$[\rho] = f_0([H] - \mu[I]) \quad (7.1)$$

и может быть рассчитана, если известна матрица  $[H]$ . При этом не нужно знать все детали взаимодействия источника и стока. Я предполагаю, что канал является достаточно большим, так что его уровни энергии расположены практически непрерывно и их уширение, связанное со взаимодействием с истоком и стоком, не имеет принципиального значения.

Гамильтоновская матрица  $[H]$  состоит из двух слагаемых (см. параграф 1.4):

$$H = H_0 + U([\delta \rho]), \quad (7.2)$$

где  $H_0$  отвечает изолированным друг от друга элементам устройства и определяется их зонной структурой, а слагаемое  $U$  описывает потенциал, определяемый напряжением на затворе и соответствующим *отклонением* матрицы плотности от равновесного значения. Если не принимать во внимание поправки, связанные с корреляционными эффектами (см. параграф 3.2), то величину  $U$  удается рассчитать с помощью уравнения Пуассона, описывающего кулоновские взаимодействия. Если считать, что диэлектрическая проницаемость  $\epsilon_r$  зависит от координат, то уравнение Пуассона записывается как

$$\vec{\nabla} \cdot (\epsilon_r \vec{\nabla} U) = -\frac{q^2}{\epsilon_0} [n(\vec{r}) - n_0], \quad (7.3)$$

а граничные условия примут вид:  $[U]_{\text{источника}} = [U]_{\text{стока}} = 0$ , а

$$[U]_{\text{затвора}} = -qV_G.$$

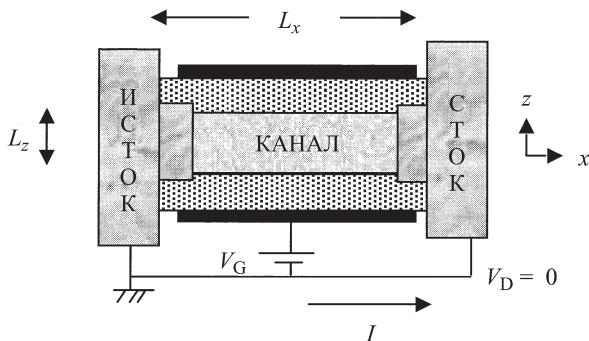


Рис. 7.1. Конденсатор в структуре металл–диэлектрик–полупроводник (МДП). Задача состоит в определении заряда, индуцированного в канале в ответ на приложенное напряжение затвора  $V_G$

В этой главе я собираюсь рассматривать данную задачу для иллюстрации того, как следует вводить гамильтониан  $[H_0]$ , описывающий неоднородную структуру, например, транзистор (параграф 7.1); каким образом вычисляется матрица плотности  $[\rho]$  (параграф 7.2). В параграфе 7.3 мы выясним, каким образом емкость  $C$ , полученная из самосогласованного решения уравнений (7.1)–(7.3), может рассматриваться как состоящая из последовательно соединенных электростатической емкости  $C_E$ , зависящей от диэлек-

трической постоянной, и квантовой емкости  $C_Q$ , которая зависит от плотности состояний в канале. Начиная с этой главы мы будем использовать матрицу гамильтониана  $[H]$ , полученную в приближении эффективной массы из однозонной модели, хотя описываемый подход является достаточно общим и может применяться вместе с другими моделями. Вспомогательный параграф 7.4 содержит начальные сведения о другом подходе, часто используемом для расчета спектра полупроводников. Речь идет о приближении эффективных масс в многозонной модели. Однако в дальнейшем в этой книге такая модель использоваться не будет.

## 7.1. Модельный гамильтониан

**Атомный гамильтониан.** Начнем с вопроса о том, как следует записать матрицы  $[H_0]$ , соответствующие различным частям рассматриваемого устройства, если известна индивидуальная зонная структура каждого материала. Можно было бы моделировать канал гамильтонионовской матрицей  $[H_0]$ , которая представлена схематично в виде сети элементарных ячеек  $[H_{nn}]$ , соединенных «связями» — матричными элементами  $[H_{nm}]$  того же размера ( $b \times b$ ). Каждая такая матрица имеет размер  $b \times b$ , где  $b$  есть число базисных функций, описывающих электроны в одной элементарной ячейке (рис. 7.1.1).

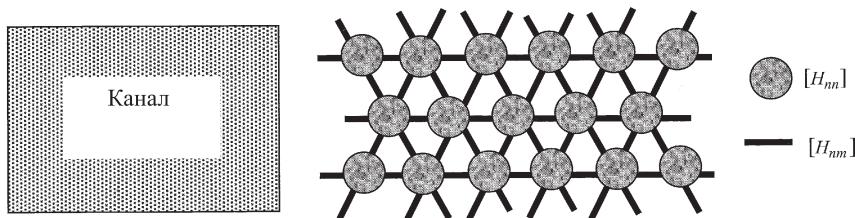


Рис. 7.1.1. Любая часть структуры (например, канал) может быть представлена схематически как трехмерная сеть элементарных ячеек, описываемых диагональными элементами матрицы  $[H_{nn}]$ , и связей, описываемых элементами  $[H_{nm}]$ ,  $n \neq m$ . Единичные ячейки находятся в узлах типичной для большинства распространенных полупроводников ГЦК решетки. Обратите внимание на то, что матрицы  $[H_{nm}]$  различных соседних узлов  $m$ , вообще говоря, различны, хотя все они обозначены одним символом

Как отмечалось в главе 5, если известны все матричные элементы  $[H_{nm}]$ , то полную зонную структуру можно рассчитать, определив при

каждом значении  $\vec{k}$  собственные значения матрицы размером  $b \times b$

$$[h(\vec{k})] = \sum_m [H_{nm}] e^{i\vec{k} \cdot (\vec{d}_m - \vec{d}_n)}, \quad (7.1.1)$$

где матричные элементы не зависят от  $n$ . И наоборот, если известна зонная структура, то можно восстановить матрицу  $[H_{nm}]$  и записать матрицу  $[H_0]$ , имеющую размер  $(Nb \times Nb)$ , описывающую периодическое твердое тело. Здесь  $N$  есть полное число элементарных ячеек.

Диэлектрическая прослойка, очевидным образом, должна описываться другой системой матриц, явный вид которых может быть найден, если известна зонная структура соответствующего материала. При решении подобных задач наиболее сложной процедурой представляется описание *границы раздела*. Это связано с тем, что нам не известна точная атомная структура реальной границы. Но даже если предположить, что микроструктура границы известна, неясно, как выглядит матрица взаимодействия  $[H_{nm}]$  между ячейками  $n$  и  $m$ , принадлежащими различным материалам А и Б. Подобная информация, разумеется, не содержится в индивидуальных зонных структурах материала А и материала Б. Если чрезмерно не углубляться в эту проблему, то наиболее простое решение можно получить, аппроксимируя величину связи А-Б в гамильтоновской матрице  $[H_{nm}]$  усредненным значением связей А-А и Б-Б.

**Гамильтониан в приближении эффективной массы.** Ранее мы видели (см. рис. 5.1.2), что уровни энергии в окрестности минимума зоны проводимости описываются простым выражением вида

$$[h(\vec{k})] = E_c + \frac{\hbar^2 k^2}{2m_c}, \quad (7.1.2)$$

где параметры  $E_c$  и  $m_c$  можно подобрать из эксперимента. Нетрудно записать дифференциальное уравнение, из которого находятся собственные значения энергии, удовлетворяющие уравнению (7.1.2). Для этого в выражении  $h(\vec{k})$  необходимо заменить вектор  $\vec{k}$  на оператор  $-i\vec{\nabla}$ :

$$\left[ E_c - \frac{\hbar^2}{2m_c} \nabla^2 \right] f(\vec{r}) = Ef(\vec{r}). \quad (7.1.3)$$

Несложно проверить, что при любом  $\vec{k}$  решения в виде плоских волн  $f(\vec{r}) = \exp(i\vec{k} \cdot \vec{r})$  являются собственными функциями этого дифференциального уравнения, отвечающими собственным значениям  $E(\vec{k}) = E_c + (\hbar^2 k^2 / 2m_c)$ . Используя метод конечных разностей (см. параграф 2.2),

можно перейти от уравнения (7.1.3) к гамильтоновской матрице, имеющей значительно более простую структуру, нежели исходный атомный гамильтониан. К примеру, в случае одного измерения мы могли бы записать трехдиагональную матрицу с элементом  $E_c + 2t_0$  на главной диагонали и элементом  $-t_0$  на нижней и верхней ближайших диагоналях (см. уравнение (2.3.1)), которая может быть представлена в виде схемы, показанной на рис. 7.1.2.

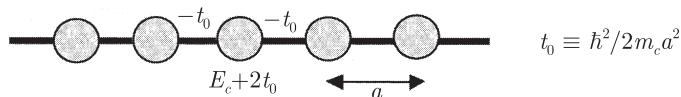


Рис. 7.1.2. Схематическое представление гамильтоновской матрицы в одномерном случае в виде одномерного массива единичных ячеек с энергией  $E_c + 2t_0$ , взаимодействующих с ближайшими соседями через матричный элемент величины  $-t_0$

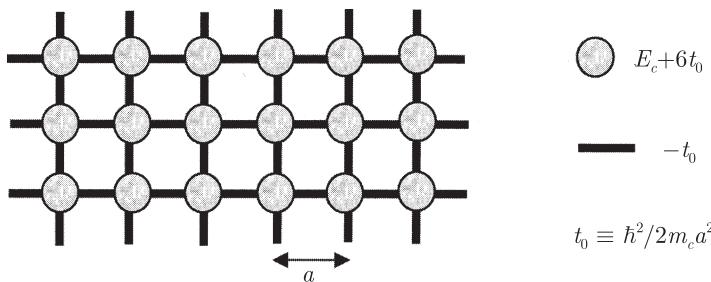


Рис. 7.1.3. Гамильтоновская матрица в приближении эффективной массы может быть изображена схематически как трехмерная сеть единичных ячеек (не имеющих отношения к реальной кристаллической структуре), каждая из которых имеет энергию  $E_c + 6t_0$  и ее взаимодействие с ближайшими соседями определяется величиной матричного элемента  $-t_0$

Чтобы записать дисперсионное соотношение, можно использовать основное уравнение исходной зонной структуры (7.1.1):

$$h(k_x) = (E_c + 2t_0) - t_0 e^{ik_x a} - t_0 e^{-ik_x a} = E_c + 2t_0(1 - \cos k_x a).$$

Для трехмерной структуры общего вида гамильтониан в приближении эффективной массы имеет вид, схематически показанный на рис. 7.1.3,

и приводит к дисперсионному соотношению

$$[h(\vec{k})] = E_c + 2t_0(1 - \cos k_x a) + 2t_0(1 - \cos k_y a) + 2t_0(1 - \cos k_z a). \quad (7.1.4a)$$

Если параметр  $k_x a$  настолько мал, что функцию  $(1 - \cos k_x a)$  можно аппроксимировать величиной  $(k_x a)^2/2$ , то мы получаем параболический закон дисперсии в уравнении (7.1.2). Аналогичное разложение и по параметрам  $k_y a$  и  $k_z a$  приводит к следующему результату:

$$[h(\vec{k})] = E_c + \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2 + k_z^2). \quad (7.1.4b)$$

Ясно, что в отличие от атомного гамильтониана, описывающего полную зонную структуру (см. уравнение (7.1.1)), этот гамильтониан определяет лишь состояния с энергией вблизи дна зоны проводимости, где уравнение (7.1.2) обеспечивает адекватную аппроксимацию. Взамен, однако, мы получаем более простое математическое описание. Действительно, получающаяся гамильтоновская матрица  $[H_0]$  по своим размерам значительно меньше атомной. Во-первых, матрицы  $[H_{nm}]$ , представляющие элементарную ячейку или связи между ними, являются скалярами, а не матрицами размера  $(b \times b)$ . Во-вторых, отдельные ячейки, показанные на рис. 7.1.3, не обязаны соответствовать атомным элементарным ячейкам (см. рис. 7.1.1). Решетка может быть не обязательно гранецентрированной, а, к примеру, простой кубической, и постоянная решетки  $a$  может быть достаточно большой в зависимости от того, в каком интервале энергий мы стремимся получить точные результаты. Здесь работает простое правило: постоянная решетки  $a$  должна быть достаточно малой для того, чтобы величина  $t_0$  была бы больше, чем интересующий нас интервал энергий над  $E_c$ . Поскольку  $t_0 \equiv \hbar^2/2m_c a^2$ , то при фиксированном интервале энергий можно выбрать большее значение параметра  $a$ , если эффективная масса  $m_c$  является малой. Но при этом необходимо помнить, что волновая функция не описывает систему на атомном масштабе, она обладает лишь информацией на усредненном пространственном масштабе. Поэтому ее зачастую называют «огибающей функцией».

**Пространственно неоднородная эффективная масса.** Уравнения в приближении эффективной массы зачастую используются для моделирования «гетероструктур», состоящих из различных материалов, так что край зоны проводимости  $E_c$  и эффективная масса  $m_c$ , появляющиеся в уравнении (7.1.3)

$$\left[ E_c - \frac{\hbar^2}{2m_c} \nabla^2 \right] f(\vec{r}) = Ef(\vec{r}),$$

изменяются в пространстве. Изменение параметра  $E_c$  не приводит к каким-либо новым трудностям, но изменение  $m_c$  нельзя учесть, просто записав

$$\left[ E_c - \frac{\hbar^2}{2m_c(\vec{r})} \nabla^2 \right] f(\vec{r}) = Ef(\vec{r}).$$

Правильная форма этого уравнения есть

$$\left[ E_c - \frac{\hbar^2}{2m_0} \vec{\nabla} \cdot \left( \frac{1}{m_c(\vec{r})} \vec{\nabla} \right) \right] f(\vec{r}) = Ef(\vec{r}). \quad (7.1.5)$$

Можно показать, что использование для решения этого уравнения метода конечных разностей на границе раздела, с эффективными массами  $m_1$  и  $m_2$ , приводит к гамильтоновской матрице, схематически представленной на рис. 7.1.4.

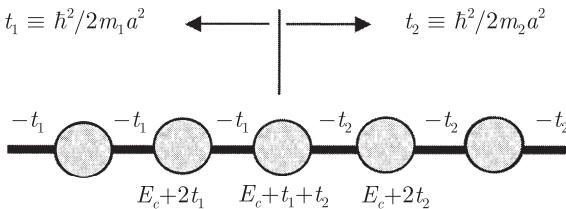


Рис. 7.1.4

Необходимо отметить, что полученная гамильтоновская матрица

$$\begin{pmatrix} E_c + 2t_1 & -t_1 & 0 \\ -t_1 & E_c + t_1 + t_2 & -t_2 \\ 0 & -t_2 & E_c + 2t_2 \end{pmatrix}$$

является эрмитовской, что обеспечивает вещественность собственных значений и сохранение тока. Напротив, использование метода конечных разностей в других моделях, например,

$$\left[ E_c - \frac{\hbar^2}{2m_c(\vec{r})} \nabla^2 \right] f(\vec{r}) = Ef(\vec{r}),$$

и привело бы к матрице вида ( $t_0 \equiv (t_1 + t_2)/2$ )

$$\begin{pmatrix} E_c + 2t_1 & -t_1 & 0 \\ -t_0 & E_c + 2t_0 & -t_0 \\ 0 & -t_2 & E_c + 2t_2 \end{pmatrix},$$

которая, очевидно, является *неэрмитовской*.

Выше отмечалось, что для правильной записи гамильтониана на границе раздела необходимо знать микроструктуру этой границы. На практике в расчетах используются наиболее простые аппроксимации. Но при этом важно отметить, что независимо от используемого приближения фундаментальным исходным требованием является *эрмитовость* матрицы гамильтониана. В противном случае мы можем столкнуться с несохранением плотности вероятности и нарушением непрерывности электронного тока.

**Пример.** Однозонные модели и приближение эффективной массы широко используются для моделирования гетероструктур, состоящих из не слишком сильно отличающихся материалов. Рассмотрим, к примеру, квантовую яму в GaAs, расположенную между барьерами из  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ . Для GaAs используем значения  $E_c = 0$  эВ,  $m_c = 0.07 m$ , а для AlAs соответствующие величины равны  $E_c = 1.25$  эВ,  $m_c = 0.15 m$ . Параметры сплава AlAs–GaAs находятся с помощью линейной интерполяции. На рис. 7.1.5 показаны энергии двух низших уровней в квантовой яме в GaAs, построенные как функции ширины ямы, а на рис. 7.1.6 приведены дисперсионные зависимости  $E(\vec{k})$  как функции модуля волнового вектора  $\vec{k} = (k_x, k_y)$  для квантовой ямы шириной  $W = 6.9$  нм.

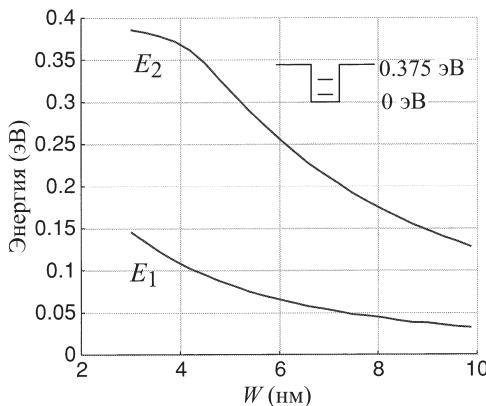


Рис. 7.1.5. Зависимость энергии двух низших уровней квантовой ямы в GaAs, расположенной в области между барьерами из  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$  как функции ширины ямы  $W$ . Расчет проведен в приближении эффективной массы

Обычный рецепт получения дифференциального уравнения в приближении эффективной массы состоит в замене  $\vec{k}$  на  $-i\vec{\nabla}$ , т. е. следует сделать

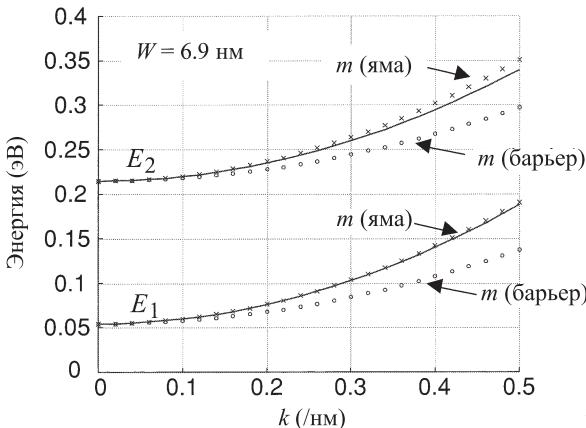


Рис. 7.1.6. Сплошные линии изображают рассчитанные в однозонной модели эффективной массы дисперсионные соотношения  $E(\vec{k})$  как функции модуля волнового вектора  $\vec{k} = (k_x, k_y)$ , лежащего в плоскости, для двух низших подзон квантовой ямы ширины  $W = 6.9$  нм. Знаки  $\times$  и  $\circ$  показывают дисперсионные зависимости, которые получаются при использовании эффективных масс, равных их значениям в яме и барьере соответственно

одновременно три замены

$$k_x \rightarrow -i\partial/\partial x, \quad k_y \rightarrow -i\partial/\partial y, \quad k_z \rightarrow -i\partial/\partial z.$$

Далее, чтобы в этом приближении получить одномерное уравнение при сохранении периодических граничных условий в плоскости  $x - y$ , следует в выражении для  $h(\vec{k})$  заменить  $k_z$  на оператор  $-i\partial/\partial z$ , не изменяя при этом  $k_x$  и  $k_y$ :

$$[h(k_x, k_y; k_z \Rightarrow -i\partial/\partial z) + U(z)]\phi(z) = E\phi(z). \quad (7.1.6)$$

Например, если рассматривать закон дисперсии (7.1.2), полученный в приближении эффективной массы для одной зоны

$$[h(\vec{k})] = E_c + \frac{\hbar^2 k^2}{2m_c},$$

то после подстановки мы придем к следующему уравнению:

$$\left[ E_c - \frac{\hbar^2}{2} \frac{\partial}{\partial z} \left( \frac{1}{m_c(z)} \frac{\partial}{\partial z} \right) + U(z) + \frac{\hbar^2}{2m_c(z)} (k_x^2 + k_y^2) \right] \phi_\alpha(z) = \varepsilon_\alpha \phi_\alpha(z), \quad (7.1.7)$$

которое является одномерным и может быть решено численно при любых значениях  $k_x$  и  $k_y$ .

Однозонная модель и приближение эффективной массы работают, когда рассматривается изотропная параболическая зона, достаточно далеко отстоящая от других зон. Это условие обычно выполняется для зоны проводимости в широкозонных полупроводниках. Напротив, валентной зоне необходимо рассматривать сразу несколько близко расположенных зон, спектр которых сильно анизотропен и которые не являются параболическими. Описание валентной зоны (структуре  $p$ -типа), а в узкозонных полупроводниках и зоны проводимости требует перехода к многозонной модели эффективной массы. Расчеты структур общего вида, основанные на многозонных моделях (см. вспомогательный материал в параграфе 7.4), а также микроскопические атомистические модели являются предметом современных исследований, и я не буду касаться их в этой книге. Для иллюстрации основных положений при описании равновесных и неравновесных задач мы будем широко использовать однозонную модель. При этом, однако, я все же попытаюсь описать и свой более общий подход, что в будущем, по мере необходимости, позволит читателям приспособить его и к более сложным гамильтонианам.

## 7.2. Электронная плотность. Матрица плотности

Как только определен вид основного гамильтониана  $H_0$ , описывающего материалы, составляющие структуру, следующим шагом будет расчет матрицы плотности и в том числе ее диагональных элементов, которые в координатном представлении определяют электронную плотность:

$$n(\vec{r}) = 2\rho(\vec{r}, \vec{r}) = 2 \sum_{\alpha} |\phi_{\alpha}(\vec{r})|^2 f_0(\varepsilon_{\alpha} - \mu), \quad (7.2.1)$$

где  $\phi_{\alpha}$  есть собственные функции матрицы  $[H]$ , отвечающие собственным значениям  $\varepsilon_{\alpha}$ , определяемым выражением

$$H = H_0 + U([\delta\rho]), \quad (7.2.2)$$

где слагаемое  $U$  есть потенциал, созданный напряжением, приложенным к затвору, а также любыми изменениями матрицы плотности относительно исходного значения, описываемого гамильтонианом  $H_0$ . В общем случае матричное представление  $[U]$  функции  $U(\vec{r})$  требует знания базисных функций, но для плавно изменяющегося потенциала можно принять простое предположение о постоянстве  $U(\vec{r} = \vec{d}_n)$  в пределах  $n$ -й ячейки и за-

писать матричные элементы как

$$[U_{nn}] = U(\vec{r} = \vec{d}_n)[I] \quad \text{и} \quad [U_{nn}] = [0] \quad \text{при} \quad m \neq n, \quad (7.2.3)$$

где  $[I]$  и  $[0]$  представляют собой единичную и нулевую матрицу соответствующего размера.

Как производится расчет потенциала  $U(\vec{r})$ ? В пренебрежении любыми поправками, связанными с корреляционными эффектами (см. параграф 3.2), для этого можно воспользоваться уравнением Пуассона, описывающим кулоновские взаимодействия (ниже  $\varepsilon_r$  обозначает относительную диэлектрическую проницаемость, которая может быть пространственно неоднородной):

$$\vec{\nabla} \cdot (\varepsilon_r \vec{\nabla} U) = -\frac{q^2}{\varepsilon_0} (n(\vec{r}) - n_0), \quad (7.2.4)$$

и удовлетворяющим граничным условиям  $[U]_{\text{источник}} = [U]_{\text{сток}} = 0$ , а для потенциала затвора выполняется

$$[U]_{\text{затвор}} = -qV_G.$$

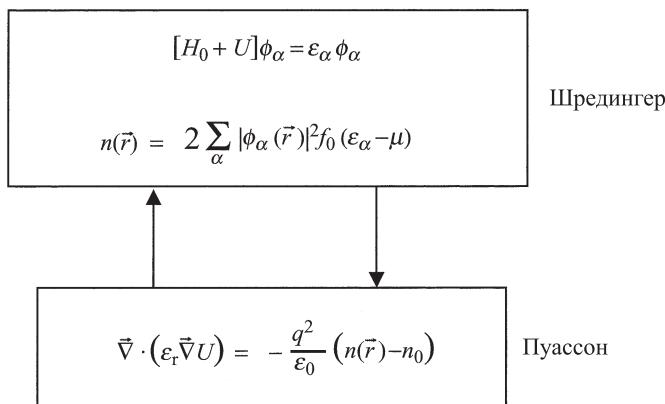


Рис. 7.2.1. Моделирование равновесного состояния структуры требует самосогласованного решения уравнения Шредингера совместно с уравнением Пуассона

Нам необходим «алгоритм решения уравнений Шредингера и Пуассона», который позволял бы решать обе части задачи самосогласованно, как это показано схематически на рис. 7.2.1. В общем случае решение трехмерных задач представляется достаточно сложным, и поэтому для иллю-

страции основных физических закономерностей мы будем рассматривать только одномерные задачи.

**Одномерный алгоритм решения уравнений Шредингера и Пуассона.** Существует большое число задач, допускающих использование одномерного алгоритма решения уравнений Шредингера и Пуассона. Так, конденсатор структуры металл-диэлектрик-полупроводник (МДП) (см. рис. 7.2.2), который упоминался в начале этой главы, представляет пример такой системы (в случае, если можно пренебречь краевыми эффектами в плоскости  $x - y$ ). Что представляет собой одномерный вариант уравнений из рис. 7.2.1? Предположим, что мы работаем с однозонным гамильтонианом, записанным в приближении эффективной массы. Можно догадаться, что прежде всего необходимо решить одномерный вариант уравнения (7.1.7)

$$\left[ E_c - \frac{\hbar^2}{2m_c} \frac{\partial^2}{\partial z^2} + U(z) \right] \phi_m(z) = \varepsilon_m \phi_m(z), \quad (7.2.5)$$

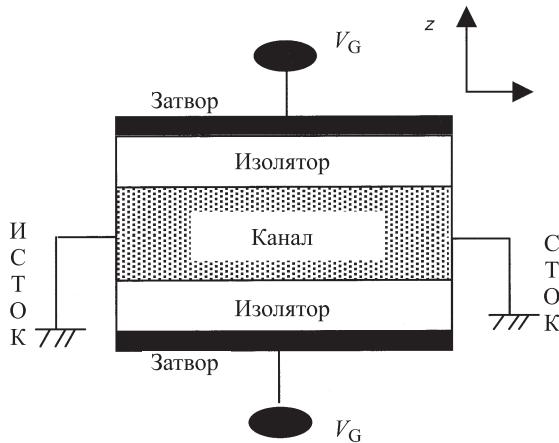


Рис. 7.2.2. Конденсатор в МДП-структуре может быть описан одномерным уравнением Шредингера при пренебрежении краевыми эффектами в плоскости  $xy$

далее вычислить электронную плотность, используя одномерный вариант уравнения (7.2.1)

$$n(z) = 2 \sum_m |\phi_m(z)|^2 f_0(\varepsilon_m - \mu) \quad (\text{НЕВЕРНО!}), \quad (7.2.6)$$

и далее проделать все это самосогласованным образом с одномерным уравнением (7.2.4), в котором начальная электронная плотность  $n_0$  принята равной нулю:

$$-\frac{d}{dz} \left( \varepsilon_r \frac{dU}{dz} \right) = \frac{q^2}{\varepsilon_0} n(z). \quad (7.2.7)$$

Все написанные выше одномерные варианты уравнений являются верными, за исключением уравнения (7.2.6). Функция распределения Ферми  $f_0$ , появляющаяся в этом уравнении, должна быть заменена новой функцией  $f_{2D}$ , определенной как

$$f_{2D}(E) = N_0 \ln[1 + \exp(-E/k_B T)] \quad \text{с} \quad N_0 = \frac{m_c k_B T}{2\pi\hbar} \quad (7.2.8)$$

(сравните с функцией  $f_0(E) = 1/[1 + \exp(E/k_B T)]$ ). Схема правильного алгоритма совместного решения одномерных уравнений Шредингера и Пуассона показана на рис. 7.2.3.

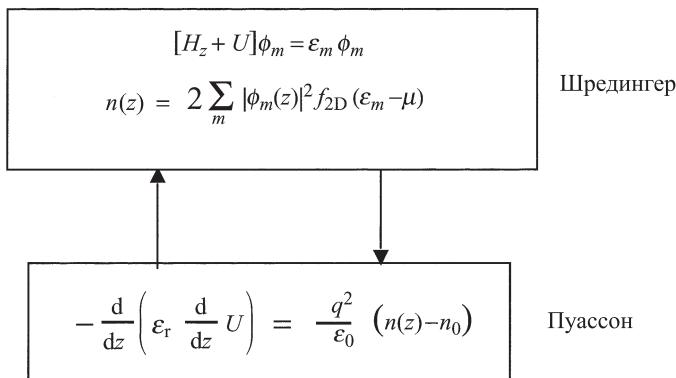


Рис. 7.2.3. Алгоритм решения одномерных уравнений Шредингера и Пуассона

**Откуда возникает новая функция  $f_{2D}$ ?** Если структура считается однородно периодической в плоскости  $xy$  и можно пренебречь краевыми эффектами, то собственные функции все еще могут быть записаны в виде плоских волн, бегущих в направлениях  $x$  и  $y$  и нормированных на масштаб  $L_x$  и  $L_y$  соответственно:

$$\phi_\alpha(\vec{r}) = \frac{\exp[ik_x x]}{\sqrt{L_x}} \frac{\exp[ik_y y]}{\sqrt{L_y}} \phi_m(z), \quad (7.2.9)$$

а электронная плотность находится из уравнения (7.2.1) путем суммирования по всем трем индексам  $\{m, k_x, k_y\} \equiv \alpha$ :

$$n(z) = 2 \sum_m \sum_{k_x, k_y} |\phi_m(z)|^2 f_0(\varepsilon_\alpha - \mu). \quad (7.2.10)$$

Выражение (7.2.6) является неверным, поскольку в нем не учтено суммирование по  $k_x$  и  $k_y$ . Правильный вариант можно получить, учитывая, что

$$\varepsilon_\alpha = \varepsilon_m + \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2), \quad (7.2.11)$$

которое следует из выражения (7.1.7) при постоянной (не зависящей от  $z$ ) эффективной массе. Обратите внимание на то, что расчет мог бы усложниться, если бы масса зависела от  $z$ , и дополнительное слагаемое

$$\frac{\hbar^2}{2m_c(z)}(k_x^2 + k_y^2)$$

не было бы аддитивной константой. При определенных условиях это приближение все еще может быть эффективно, поскольку волновая функция, в основном, сосредоточена в области, где эффективная масса постоянна (см., к примеру, упражнение 7.1(в)). Однако в общем случае такая ситуация не имеет места. Необходимо также отметить, что простой параболический закон дисперсии (7.2.11) нельзя использовать для записи уравнения многослойного приближения эффективной массы (см. рис. 7.2.4).

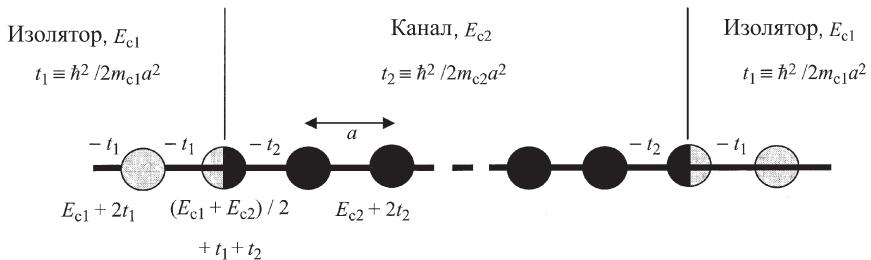


Рис. 7.2.4. Одномерный гамильтониан однозонной модели приближения эффективной массы, используемый для расчета состояний в канале, заключенном между двумя слоями изоляторов

Используя выражение (7.2.11), можно аналитически выполнить суммирование по  $k_x$  и  $k_y$  и убедиться в справедливости равенства

$$\sum_{k_x, k_y} f_0(\varepsilon_\alpha - \mu) = \sum_{k_x, k_y} f_0 \left( \varepsilon_m - \mu + \frac{\hbar^2}{2m_c} (k_x^2 + k_y^2) \right) = f_{2D}(\varepsilon_m - \mu). \quad (7.2.12)$$

Это делается следующим образом:

$$\begin{aligned} & \frac{1}{S} \sum_{k_x, k_y} f_0 \left( E + \frac{\hbar^2}{2m_c} (k_x^2 + k_y^2) \right) = \\ &= \int_0^\infty \frac{2\pi k dk}{4\pi^2} \frac{1}{1 + A \exp(\hbar^2 k^2 / 2m_c k_B T)}, \quad \text{где } A \equiv \exp(E/k_B T) = \\ &= \frac{m_c k_B T}{2\pi \hbar^2} \int_0^\infty \frac{dy}{1 + A e^y} = \frac{m_c k_B T}{2\pi \hbar^2} \{ \ln[A + e^{-y}] \}_\infty^0 = \\ &= \frac{m_c k_B T}{2\pi \hbar^2} \ln[1 + \exp(-E/k_B T)] \equiv f_{2D}(E). \end{aligned}$$

Уравнение (7.2.12) позволяет упростить выражение (7.2.10), чтобы получить выражение, показанное на рис. 7.2.3:

$$n(z) = 2 \sum_m |\phi_m(z)|^2 f_{2D}(\varepsilon_m - \mu). \quad (7.2.13)$$

**Численный пример.** Численное моделирование МДП-конденсатора, показанного на рис. 7.2.2, можно провести, используя одномерный алгоритм решения уравнений Шредингера и Пуассона. Следует также предположить, что сечение в плоскости  $xy$  является однородным и выполняются периодические граничные условия. Уравнения записываются на решетке, выбранной вдоль направления  $z$ , и одномерный гамильтониан  $H_z$  выглядит так, как показано на рис. 7.2.4. При этом допускается, что край зоны проводимости  $E_c$  и эффективная масса  $m_c$  могут иметь различные значения в диэлектрике и в канале. Тем не менее, мы будем пользоваться выражением (7.2.13) со значением  $N_0$  (см. уравнение (7.2.8)), определяемым эффективной массой в канале, поскольку волновые функции практически не проникают в область диэлектрика.

Записав гамильтониан  $H_z$ , можно непосредственно вычислить электронную плотность  $n(z)$ , которая определяется диагональными элементами матрицы плотности, определяемой следующим образом:

$$\rho(z, z') = \sum_m \phi_m(z) f_{2D}(\varepsilon_m - \mu) \phi_m^*(z). \quad (7.2.14a)$$

Как отмечалось ранее (см. обсуждение вслед за уравнением (4.3.8)), выражение (7.2.14a) можно рассматривать как координатное представление более общего определения матрицы плотности

$$\rho = f_{2D}(H_z - \mu I). \quad (7.2.14b)$$

Чтобы найти функции матрицы  $[H_z]$ , нужно, во-первых, диагонализовать ее, затем вычислить матрицу плотности в диагональном представлении и, наконец, вернуться в координатное представление на решетке. Указанные операции могут быть выполнены в среде MatLab с использованием последовательности команд

- (1)  $[V, D] = \text{eig}(Hz); D = \text{diag}(D)$
- (2)  $\rho = \log(1 + (\exp((\mu - D).kT)))$
- (3)  $\rho = V * \text{diag}(\rho) * V'; N = \text{diag}(\rho)$

Электронная плотность  $n(z)$  получается из величины  $N$  умножением на  $2N_0$  (см. уравнение (7.2.8)), где дополнительный множитель 2 учитывает спин) и делением на размер элементарной ячейки  $a$ :  $n(z) = N \times 2N_0/a$ .

Чтобы решить уравнение Пуассона, можно записать его на той же решетке,

$$-\frac{d}{dz} \left( \varepsilon_r \frac{dU}{dz} \right) = \frac{q^2}{\varepsilon_0} n(z),$$

и решить методом конечных разностей совершенно аналогичным образом:

$$[D2]\{U\} = \frac{q^2}{\varepsilon_0 a} (2N_0 a^2)\{N\} + \{U_{bdy}\}, \quad (7.2.15)$$

где  $[D2]$  есть матричный оператор второй производной. При постоянном значении диэлектрической проницаемости  $\varepsilon_r$  он имеет вид

$$[D2] \equiv \varepsilon_r \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & 0 & \dots \\ 0 & -1 & 2 & -1 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}. \quad (7.2.16)$$

Изменение в пространстве диэлектрической проницаемости  $\varepsilon_r$  можно учитывать так же, как изменение эффективной массы. Границный элемент находится из ненулевых значений  $U$  на двух границах:

$$\{U_{\text{bdy}}\}^T = \varepsilon_r \{-qV_G \quad 0 \quad \dots \quad \dots \quad 0 \quad -qV_G\}. \quad (7.2.17)$$

Зная величину  $N$ , можно с помощью уравнения (7.2.15) рассчитать потенциал  $U$ :

$$\{U\} = \frac{q^2}{\varepsilon_0 a} (2N_0 a^2) [D2]^{-1} \{N\} + [D2]^{-1} \{U_{\text{bdy}}\}. \quad (7.2.18)$$

На рис. 7.2.5 *a, б* показаны рассчитанные зависимости энергии в зонах и электронная плотность в каналах шириной 3 и 9 нм соответственно. Предполагалось, что значение края зоны проводимости  $E_c$  равно нулю в канале и равно 3 эВ в диэлектрике. Для относительной диэлектрической проницаемости и эффективной массы имеем:  $\varepsilon_r = 4$  и  $m_c = 0.25m$  во всех областях. Предполагалось также, что электрохимический потенциал  $\mu$  равен  $E_c$ . В реальных структурах значение  $\mu$  определяется работой выхода для металла. Толщина слоя оксида предполагается равной 2.1 нм, и расчеты произведены с использованием постоянной решетки  $a = 0.3$  нм. Значение напряжения на затворе полагается равным 0.25 В.

В канале ширины 3 нм распределение плотности заряда имеет пик вблизи середины канала, что характерно для волновой функции, основного состояния задачи о «частице в яме». Напротив, квазиклассическое распределение плотности зарядов сосредоточено у краев канала, что вытекает из чисто электростатических соображений. Такая ситуация является примером эффектов, связанных с *размерным квантованием*. Эти эффекты исчезают по мере того, как увеличивается ширина канала, и происходит сближение энергетических уровней потенциальной ямы, многие из которых оказываются занятами при низких температурах. Как следствие, распределение электронной плотности будет более близким к классическому в канале ширины 9 нм. На том же рисунке пунктирными линиями показана электронная плотность для случая асимметрично приложенного напряжения на затворах, т. е. когда напряжение 0 В на одном затворе и 0.25 В на другом. Обратите внимание на то, что в канале шириной 9 нм при асимметричном напряжении на затворах происходит более сильное искажение электронной плотности, чего, впрочем, можно было ожидать и интуитивно. Напротив, в канале шириной 3 нм электронное распределение лишь немногого меняет свой вид при переходе от симметричных к асимметричным потенциалам на затворах. Волновая функция мало изменяется в присутствии внешнего

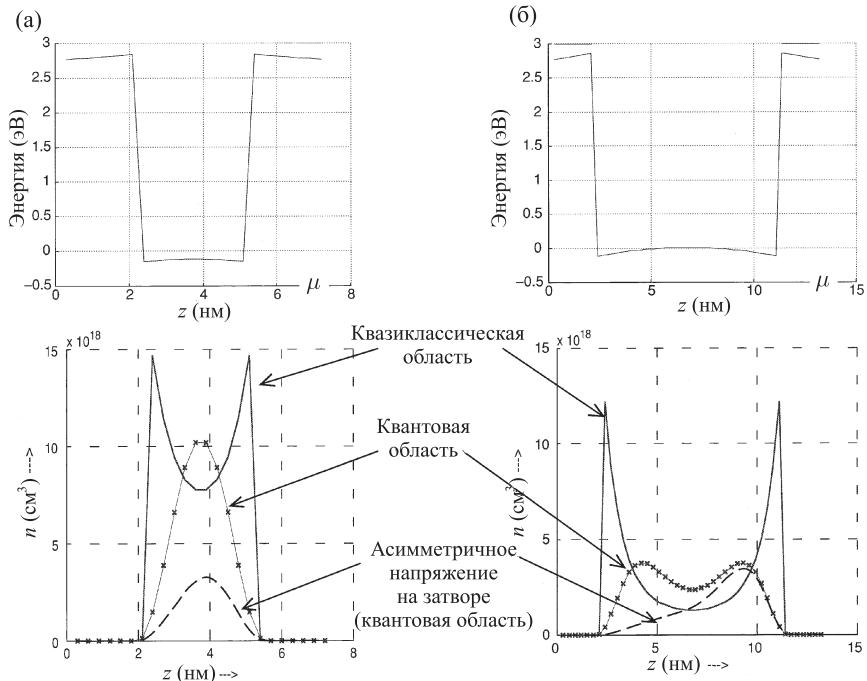


Рис. 7.2.5. МДП-конденсатор (см. рис. 7.2.2) с каналом толщины (а) 3 нм и (б) 9 нм. Предполагается, что  $\mu = 0$ ,  $E_c = 0$  в кремнии и  $E_c = 3$  эВ в оксиде. На верхних графиках представлена равновесная зонная диаграмма: сплошные линии показывают профиль зоны проводимости. Нижние графики показывают электронную плотность. Толщина оксида предполагается равной 2.1 нм, и расчеты выполнены с использованием дискретной решетки с постоянной  $a = 0.3$  нм. Напряжение на затворе  $V_G$  равно 0.25 В. Пунктирные линии показывают электронную плотность в случае, когда напряжение приложено асимметрично: 0.25 В на одном затворе и 0 В на другом затворе

напряжения, поскольку состояния значительно отделены друг от друга по шкале энергий.

**Квазиклассический метод.** На рис. 7.2.5 для сравнения приведена также электронная плотность, рассчитанная в т. н. квазиклассическом приближении. Поясним смысл такого подхода. В однородной структуре соб-

ственые функции определяются выражением

$$\phi_\alpha(\vec{r}) = \frac{\exp[ik_x x]}{\sqrt{L_x}} \frac{\exp[ik_y y]}{\sqrt{L_y}} \frac{\exp[ik_z z]}{\sqrt{L_z}}, \quad (7.2.19)$$

поэтому электронная плотность, полученная из уравнения (7.2.1), после суммирования по всем трем индексам  $\{k_x, k_y, k_z\} \equiv \alpha$  является однородной в пространстве (величина  $\Omega = L_x L_y L_z$ ):

$$n = 2(\text{для спина}) \times \frac{1}{\Omega} \sum_{k_x, k_y, k_z} f_0(\varepsilon_\alpha - \mu),$$

где

$$\varepsilon_\alpha = E_c + \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2 + k_z^2). \quad (7.2.20)$$

Суммирование может быть выполнено с помощью процедуры, использованной при обсуждении двумерной задачи (см. выражение (7.2.12)):

$$\begin{aligned} & \frac{1}{\Omega} \sum_{k_x, k_y, k_z} f_0 \left( E_c - \mu + \frac{\hbar^2}{2m_c}(k_x^2 + k_y^2 + k_z^2) \right) = \\ &= \int_0^\infty \frac{4\pi k^2 dk}{8\pi^3} \frac{1}{1 + A \exp(\hbar^2 k^2 / 2m_c k_B T)}, \quad \text{где } A \equiv \exp[(E_c - \mu)/k_B T] = \\ &= \left( \frac{m_c k_B T}{2\pi \hbar^2} \right)^{3/2} \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{dy \sqrt{y}}{1 + A e^y}, \end{aligned}$$

откуда можно получить, что

$$n = 2f_{3D}(E_c - \mu), \quad (7.2.21)$$

где

$$f_{3D}(E) \equiv \left( \frac{m_c k_B T}{2\pi \hbar^2} \right)^{3/2} \mathfrak{I}_{1/2} \left( -\frac{E}{k_B T} \right) \quad (7.2.22)$$

и

$$\mathfrak{I}_{1/2}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{d\xi \sqrt{\xi}}{1 + \exp(\xi - x)}.$$

Данное выражение для однородной электронной плотности  $n$  является верным лишь в отсутствие внешнего потенциала. При квазиклассическом подходе для расчета электронной плотности  $n(z)$ , изменяющейся в пространстве в присутствии потенциала  $U(z)$ , используется простое обобщение выражения (7.2.21):

$$n = 2f_{3D}(E_c + U(z) - \mu), \quad (7.2.23)$$

т. е. предполагается, что вблизи каждой точки  $z$  находится однородная среда с краем зоны проводимости, совпадающим с  $E_c + U(z)$ . Заменяя верхний блок рис. 7.2.3 (обозначенный как «Шредингер») данным уравнением, мы получим квазиклассический алгоритм решения уравнений Шредингера и Пуассона, широко используемый при моделировании структур.

### 7.3. Квантовая и электростатическая емкость

Электронная плотность  $n_S$  в канале (приходящаяся на единицу площади) находится интегрированием величины  $n(z)$  в выражении (7.2.13) с учетом нормировки волновых функций, удовлетворяющих условию  $\int dz |\phi_m(z)|^2 = 1$ :

$$n_S = \int dz n(z) = 2 \sum_m f_{2D}(\varepsilon_m - \mu). \quad (7.3.1)$$

На рис. 7.3.1 показана электронная плотность  $n_S$  как функция напряжения  $V_G$  на затворе, приложенного симметрично к обеим частям затвора.

Основные физические закономерности данной задачи иллюстрирует рис. 7.3.2. Положительное напряжение на затворе уменьшает полную плотность состояний и увеличивает электронную плотность  $n_S$ . Коль скоро электрохимический потенциал  $\mu$  располагается под нижним уровнем энергии, структура находится в состоянии ВЫКЛ. Если же  $\mu$  смешается в область энергий с ненулевой плотностью состояний, структура переходит в состояние ВКЛ. Из рис. 7.3.1 следует, что для включения прибора с шириной канала 3 нм требуется более высокое пороговое напряжение, нежели для структуры, имеющей канал шириной 9 нм. Это обстоятельство объясняется ростом энергии низшего уровня вследствие размерного квантования.

**Эквивалентная схема.** Интересно задаться следующим вопросом: как изменяется потенциал  $V_C$  в канале при изменении напряжения  $V_G$  на затворе? Ответ на этот вопрос легко дать в двух предельных случаях. Если канал

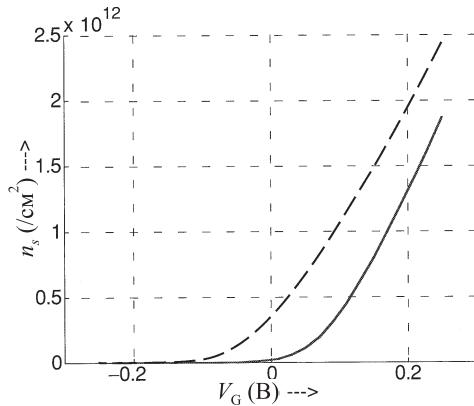


Рис. 7.3.1. Электронные плотности  $n_s$  на единицу площади в структуре с шириной канала 3 нм (сплошная линия) и 9 нм (пунктирная линия), показанные как функция напряжения  $V_G$  на затворе, приложенного симметрично к обеим частям затвора. Использован численный расчет в рамках модели, описанной в параграфе 7.2

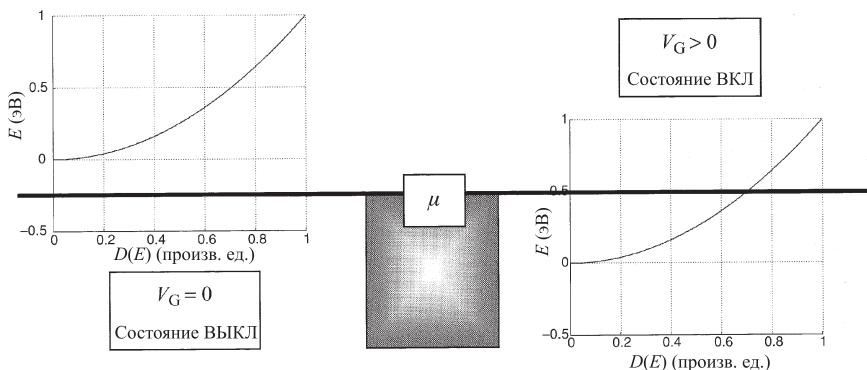


Рис. 7.3.2. Положительное напряжение  $V_G$ , приложенное к затвору, смещает плотность состояний  $D(E)$  вниз. Поскольку электрохимический потенциал  $\mu$  остается фиксированным, это увеличивает число занятых состояний и, как следствие, растет число электронов  $N$

находится в состоянии ВЫКЛ, то он ведет себя подобно изолятору, и потенциал канала  $V_C$  равен  $V_G$ . Но если канал находится в состоянии ВКЛ, то он ведет себя как отрицательно заряженная пластина плоского конденсатора,

так что потенциал канала  $V_C$  теперь равен потенциальну истока (или стока), который заземлен. Ответ на поставленный вопрос не очевиден в промежуточных ситуациях, когда канал не является ни изолятором, ни проводником. Для ответа на сформулированный вопрос полезно рассмотреть приблизительную эквивалентную схему, представленную на рис. 7.3.3. Позвольте мне объяснить, как она получена.

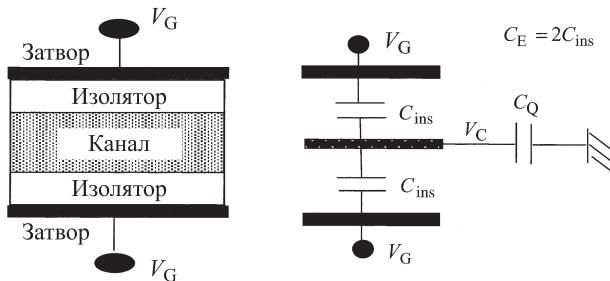


Рис. 7.3.3. Приблизительная эквивалентная схема МДП-конденсатора

Канал соединен с двумя электродами затвора через обычные плоско-параллельные конденсаторы, в которых емкость (на единицу площади) пропорциональна эффективной диэлектрической постоянной  $\epsilon$  и обратно пропорциональна расстоянию  $d$  между центром канала и электродом затвора:

$$C_{\text{изол}} \equiv \epsilon/d. \quad (7.3.2)$$

Разумеется, это лишь приближенное выражение, поскольку не все электроны в канале располагаются вблизи его центра, как мы предполагаем. Кроме того, можно задаться вопросом о том, какую именно диэлектрическую постоянную следует использовать, поскольку она должна соответствовать усредненному значению в канале и изоляторе. Учет этих «деталей» и попытка прийти к более точному выражению отвлекли бы нас от главной цели, заключающейся в выяснении основных физических закономерностей данной задачи. Чтобы получить разумные количественные результаты, достаточно воспользоваться численной схемой, описанной в конце предыдущего параграфа. Наше утверждение состоит в том, что конденсаторы  $C_{\text{изол}}$  фактически являются такими же конденсаторами, которые изучаются в начальном курсе физики.

Но откуда в таком случае берется квантовая емкость  $C_Q$ ? В предыдущем параграфе мы видели, что для определения электронной плотности и потенциала приходится одновременно решать два уравнения (см.

рис. 7.2.3): электростатическое уравнение (уравнение Пуассона) и квантовое уравнение (уравнение Шредингера). Для электростатического уравнения можно записать

$$U = U_L + [q^2(N - N_0)/C_E], \quad (7.3.3)$$

где  $U_L$  есть потенциал в канале, полученный из решения уравнения Лапласа, т. е. в отсутствие зарядов, в то время как величина  $(N - N_0)$  описывает отклонение электронной плотности зарядов от нейтрального значения  $N_0$ . В этом случае квантовое уравнение может быть записано в форме

$$N = \int_{-\infty}^{+\infty} dE D(E - U) f_0(E - \mu), \quad (7.3.4)$$

где  $D(E - U)$  есть плотность состояний (на единицу площади), аргумент которой сдвинут на величину потенциала  $U$ . Полученное уравнение является нелинейным, и его можно упростить, линеаризовав вблизи конкретной точки. К примеру, мы можем определить «нейтральный потенциал»  $U = U_N$ , для которого  $N = N_0$ , и канал находится в нейтральном состоянии:

$$N_0 = \int_{-\infty}^{+\infty} dE D(E - U_N) f_0(E - \mu).$$

Любое увеличение  $U$  поднимет энергетические уровни и уменьшит  $N$ , в то время как уменьшение  $U$  приводит к опусканию энергетических уровней и увеличению  $N$ . Таким образом, для малых отклонений от условия нейтральности можно записать

$$N - N_0 \approx C_Q[U_N - U_0]/q^2, \quad (7.3.5)$$

где величина

$$C_Q \equiv -q^2[dN/dU]_{U=U_N} \quad (7.3.6)$$

зависит от плотности состояний и называется квантовой емкостью. Это линеаризованное соотношение можно подставить в уравнение (7.3.3), откуда получим, что

$$U = U_L + \frac{C_Q}{C_E}(U_N - U) \quad \rightarrow \quad U = \frac{C_E U_L + C_Q U_N}{C_E + C_Q}. \quad (7.3.7)$$

Выражению (7.3.7) легко сопоставить схему конденсаторов, показанную на рис. 7.3.3. Реальный потенциал  $U$  в канале имеет промежуточное значение между потенциалом Лапласа  $U_L$  и нейтральным потенциалом  $U_N$ . Насколько он близок к тому или другому значению, зависит от отношения величины электростатической емкости  $C_E$  и квантовой емкости  $C_Q$ .

Пользуясь выражениями (7.3.4) и (7.3.6), можно непосредственно показать, что квантовая емкость  $C_Q$  пропорциональна плотности состояний, усредненной на масштабе порядка нескольких  $k_B T$  вблизи химического потенциала  $\mu$ :

$$C_Q \equiv q^2 D_0, \quad (7.3.8)$$

$$D_0 \equiv \int_{-\infty}^{+\infty} dE D(E + U_N) F_T(E - \mu), \quad (7.3.9)$$

где  $F_T(E)$  есть функция температурного уширения, определяемая как

$$F_T(E) \equiv -\frac{df_0}{dE} = \frac{1}{4k_B T} \sec h^2 \left( \frac{E}{2k_B T} \right). \quad (7.3.10)$$

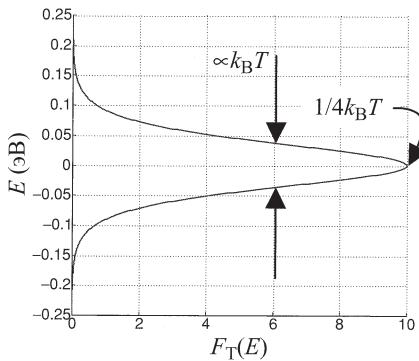


Рис. 7.3.4. График функции температурного уширения  $F_T(E)$  (см. уравнение (7.3.10)), построенный при значении  $k_B T = 0.025$  эВ

На рис. 7.3.4 показана функция температурного уширения; ее максимальное значение равно  $(1/4k_B T)$ , а ширина пропорциональна  $k_B T$ . Можно непосредственно показать, что площадь, получаемая при интегрировании этой функции, равна единице независимо от  $k_B T$ :  $\int_{-\infty}^{+\infty} dE F_T(E) = 1$ . Это означает, что при низких температурах функция  $F_T(E)$  становится

очень большой в максимуме и очень узкой, сохраняя постоянную площадь под графиком, равную единице. Такая функция является одним из представлений для дельта-функции:  $F_T(E) \rightarrow \delta(E)$ , и это позволяет упростить выражение для квантовой емкости при низких температурах:

$$C_Q \approx q^2 D(E = \mu + U_N), \quad (7.3.11)$$

которое показывает, что емкость пропорциональна плотности состояний вблизи химического потенциала  $\mu$ , сдвинутого на величину нейтрального потенциала  $U_N$ .

Пользуясь эквивалентной схемой на рис. 7.3.3, нетрудно увидеть, что

$$V_C = V_G \frac{C_E}{C_E + C_Q}, \quad \text{где} \quad C_E = 2C_{\text{изол.}} \quad (7.3.12)$$

Структуры в состоянии ВЫКЛ имеют нулевую емкость  $C_Q$ , и поэтому  $V_C = V_G$ . Напротив, в состоянии ВКЛ величина  $C_Q$  не равна нулю, и напряжение  $V_C$  при этом меньше, чем  $V_G$ . Измеряемая емкость  $C$  равна емкости последовательного соединения электростатической и квантовой емкостей,

$$C = \frac{C_E C_Q}{C_E + C_Q}, \quad (7.3.13)$$

и определяется меньшей из этих емкостей.

Мы можем приблизительно оценить величину квантовой емкости  $C_Q$  в состоянии ВКЛ, имея в виду, что двумерный проводник, описываемый параболическим законом дисперсии, имеет постоянную плотность состояний (см. табл. 6.2.1):  $D(E) = S m_c / (\pi \hbar^2)$ , так что квантовая емкость может быть аппроксимирована выражением

$$C_Q = \frac{q^2 m_c S}{\pi \hbar^2} = \frac{\varepsilon S}{a_0^*/4}, \quad (7.3.14)$$

где выражение для  $a_0^*$  подобно формуле боровского радиуса, определенного в главе 2 (см. выражение (2.1.15)). Однако параметр  $a_0^*$  по величине превосходит боровский радиус ( $= 0.052$  нм), поскольку эффективная масса в зоне проводимости ( $m_c$ ) меньше массы свободного электрона ( $m$ ), а диэлектрическая постоянная ( $\varepsilon$ ) в веществе больше, чем в вакууме ( $\varepsilon_0$ ):

$$a_0^* \equiv \frac{4\pi\varepsilon\hbar^2}{m_c q^2} = (0.053 \text{ нм}) \frac{m}{m_c} \frac{\varepsilon}{\varepsilon_0}. \quad (7.3.15)$$

Таким образом, квантовая емкость равна емкости плоскопараллельного конденсатора, расстояние между пластинами которого равно  $a_0^*/4$ . Поскольку последняя величина обычно мала, квантовая емкость в состоянии ВКЛ обычно значительно превосходит электростатическую емкость  $C_E$  и определяет измеряемую емкость (см. уравнение (7.3.13)). Однако в материалах с очень маленькой эффективной массой квантовая емкость может оказаться сравнительно малой, в особенности если изолятор очень тонкий и емкость  $C_E$  велика.

**Состояние ВЫКЛ.** В режиме ВЫКЛ плотность состояний при значениях энергии  $E$ , близких к  $\mu$  как и квантовая емкость  $C_Q$ , являются пренебрежимо малыми. Как следствие, потенциал канала  $V_C$  практически равен потенциалу затвора  $V_G$ , и можно написать, что

$$\begin{aligned} N &= \int_{-\infty}^{+\infty} dE f_0(E - \mu - qV_G) D(E) \approx \\ &\approx \int_{-\infty}^{+\infty} dE \exp\left(-\frac{E - \mu - qV_G}{k_B T}\right), \end{aligned}$$

поскольку  $E - \mu - qV_G \gg k_B T$  в интервале энергий, где  $D(E)$  отлично от нуля. В этом режиме число электронов зависит экспоненциально от приложенного напряжения:

$$\begin{aligned} N &\sim N_0 \exp\left(\frac{qV_G}{k_B T}\right), \\ \text{так что} \quad \log_{10} \left( \frac{N}{N_0} \right) &\approx \left( \frac{qV_G}{2.3k_B T} \right). \end{aligned} \quad (7.3.16)$$

Этот хорошо известный результат говорит о том, что в режиме ВЫКЛ число электронов возрастает на порядок при каждом увеличении напряжения на затворе на  $2.3k_B T$  ( $\sim 60$  мэВ при комнатной температуре). Эту закономерность можно проверить, если заново построить график, приведенный на рис. 7.3.1 в логарифмическом масштабе, и рассмотреть его наклон в режиме ВЫКЛ.

**Режим ВКЛ.** В режиме ВКЛ электрохимический потенциал  $\mu$  лежит в области, где плотность состояний  $D(E)$  (и соответствующая квантовая емкость  $C_Q$ ) имеют существенно ненулевые значения. Как отмечалось выше, фактическая емкость определяется как емкость последовательно соединенных квантовой и электростатической емкостей. Рассматривая наклон зависимости  $n_S$  от  $V_G$  в режиме ВКЛ (см. рис. 7.3.1), мы определим значение

емкости, равное приблизительно  $1.8 \times 10^{-6} \Phi/\text{см}^2$  для канала ширины 3 нм. Если приравнять это значение к величине  $2\varepsilon/d$ , найдем  $d = 3.9$  нм, что хорошо согласуется со значением, полученным суммированием половины ширины канала (1.5 нм) и толщины оксида (2.1 нм). Значит, эффективная емкость имеет, в основном, электростатическую, а не квантовую природу.

## 7.4. Дополнение: многозонный гамильтониан в приближении эффективной массы

Однозонная модель в приближении эффективной массы дает хорошие результаты в том случае, когда данная зона далеко отстоит от остальных зон и имеет изотропный параболический закон дисперсии. Такая ситуация обычно имеет место, когда мы имеем дело с зоной проводимости в широкозонных полупроводниках. При этом, однако, для описания валентной зоны требуется одновременно учитывать несколько близко расположенных зон, являющихся сильно анизотропными и непараболическими. В этом случае вблизи Г-точки закон дисперсии обычно может быть представлен в следующем виде:

$$E(\vec{k}) = E_V - Ak^2 \pm \sqrt{B^2 k^4 + C^2(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)}, \quad (7.4.1)$$

где  $A$ ,  $B$  и  $C$  являются постоянными.

Эта дисперсионная зависимость может быть найдена с помощью матрицы размера  $4 \times 4$  вида:

$$[h(\vec{k})] = -P I - T, \quad (7.4.2)$$

где  $I$  есть единичная матрица  $4 \times 4$

$$[T] \equiv \begin{pmatrix} Q & 0 - S & R & \\ 0 & Q & R^+ & S^+ \\ -S^+ & R & -Q & 0 \\ R^+ & S & 0 & -Q \end{pmatrix},$$

$$P \equiv E_V + \frac{\hbar^2 \gamma_1}{2m} (k_x^2 + k_y^2 + k_z^2),$$

$$Q \equiv \frac{\hbar^2 \gamma_2}{2m} (k_x^2 + k_y^2 - 2k_z^2),$$

$$S \equiv \frac{\hbar^2 \gamma_3}{2m} 2\sqrt{3}(k_x - ik_y)k_z.$$

Параметры Латинжера  $\gamma_1$ ,  $\gamma_2$  и  $\gamma_3$  известны для всех распространенных полупроводников (см., к примеру, книги [31, 58]). Можно предположить, что выражения (7.4.1) и (7.4.2) эквивалентны, поскольку прямой алгебраический расчет показывает, что

$$[T]^2 = (Q^2 + R^2 + S^2)I = [B^2 k^4 + C^2(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)]I.$$

Из рис. 7.4.1 можно сделать вывод, что собственные значения матрицы  $[h(\vec{k})]$  хорошо описывают две самые высокие ветви валентной зоны, отвечающие вблизи гамма-точки легким и тяжелым дыркам. Чтобы получить лучшее соответствие в более широком диапазоне значений волновых векторов, включая также спин-отщепленную зону (см. рис. 7.4.2), зачастую используют трехзонную модель для матрицы  $[h(\vec{k})]$  вида

$$[h(\vec{k})] = - \begin{pmatrix} P+Q & 0 & -S & R & -S/\sqrt{2} & \sqrt{2}R \\ 0 & P+Q & R^+ & S^+ & -\sqrt{2}R^+ & -S^+/\sqrt{2} \\ R^+ & 0 & P-Q & 0 & -\sqrt{2}Q & \sqrt{3/2}S \\ 0 & R^+ & 0 & P-Q & \sqrt{3/2}S^+ & \sqrt{2}Q^+ \\ -S^+/\sqrt{2} & -\sqrt{2}R & -\sqrt{2}Q^+ & \sqrt{3/2}S & P+\Delta & 0 \\ \sqrt{2}R^+ & -S/\sqrt{2} & \sqrt{3/2}S^+ & \sqrt{2}Q & 0 & P+\Delta \end{pmatrix}. \quad (7.4.3)$$

Чтобы написать уравнение эффективных масс в валентной зоне, можно использовать как двухзонную матрицу  $[h(\vec{k})]$  в (7.4.2), так и трехзонную матрицу  $[h(\vec{k})]$  в (7.4.3). При этом следует руководствоваться теми правилами, которые были сформулированы при рассмотрении зоны проводимости (сравните с выражениями (7.1.2), (7.1.3)). При этом мы просто заменяем вектор  $\vec{k}$  на оператор  $-i\vec{\nabla}$  в выражении для  $h(\vec{k})$  и получаем систему связанных дифференциальных уравнений вида

$$[h(\vec{k} \rightarrow -i\vec{\nabla})]\{f(\vec{r})\} = E\{f(\vec{r})\}, \quad (7.4.4)$$

где «волновая функция»  $\{f(\vec{r})\}$  теперь имеет четыре (или шесть) компонент. Нетрудно проверить, что решения в форме плоских волн вида  $\{f(\vec{r})\} = \{f_0\} \exp(i\vec{k} \cdot \vec{r})$  при любом векторе  $\vec{k}$  будут удовлетворять уравнению (7.4.4), если вектор-столбец  $\{f_0\}$  будет собственным вектором матрицы  $[h(\vec{k})]$ ,

$$[h(\vec{k})]\{f_0\} = E\{f_0\}. \quad (7.4.5)$$

Этот результат означает, что уравнение эффективной массы (7.4.4) будет генерировать ту же зонную структуру, которая следует из исходной матрицы  $[h(\vec{k})]$ .

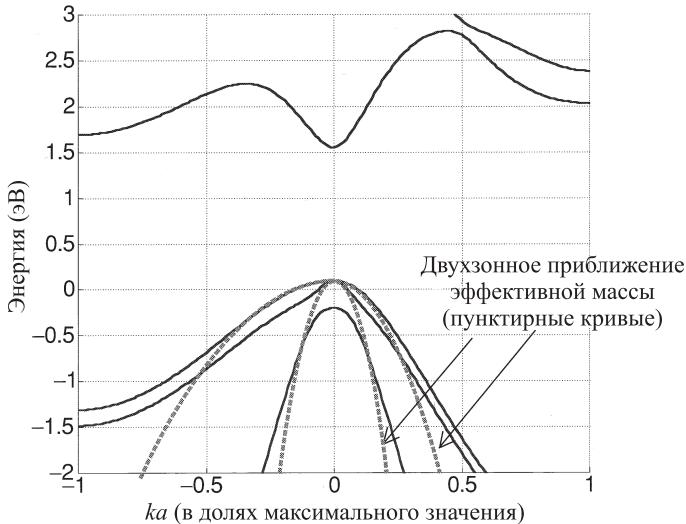


Рис. 7.4.1. Сплошными линиями показана полная зонная структура, полученная в рамках  $sp^3s^*$ -модели, описанной в главе 5. Пунктирные кривые демонстрируют дисперсионные зависимости, найденные в рамках двухзонной модели эффективных масс (выражение (7.4.2)) с параметрами, подобранными для наилучшего согласования законов дисперсии

Можно воспользоваться методом конечных разностей, для того чтобы преобразовать выражение (7.4.4) в матрицу оператора Гамильтона аналогично тому, как мы поступили с выражением (7.1.3), и получить матрицу, структура которой схематически показана на рис. 7.4.3. На этой схеме матрицы единичных ячеек  $[H_{nn}]$  и описывающие взаимодействие блоки  $[H_{nm}, n \neq m]$  так же будут иметь размер  $(4 \times 4)$  или  $(6 \times 6)$  в зависимости от того, используется ли двухзонная (выражение (7.4.2)) или трехзонная (выражение (7.4.3)) модель. К примеру, в двухзонной модели блоки  $[H_{nm}]$  будут иметь вид

$$[H_{nm}] \equiv \begin{pmatrix} -P_{nm} - Q_{nm} & 0 & S_{nm} & -R_{nm} \\ 0 & -P_{nm} - Q_{nm} & -R_{mn}^* & -S_{mn}^* \\ S_{mn}^* & -R_{nm} & -P_{nm} + Q_{nm} & 0 \\ -R_{mn}^* & -S_{nm} & -P_{nm} + Q_{nm} & -P_{nm} + Q_{nm} \end{pmatrix},$$

где отдельные элементы  $P_{nm}$ ,  $Q_{nm}$  и т. д. получаются из соответствующих функций  $P(\vec{k})$ ,  $Q(\vec{k})$  с помощью процедуры, аналогичной той, что использу-

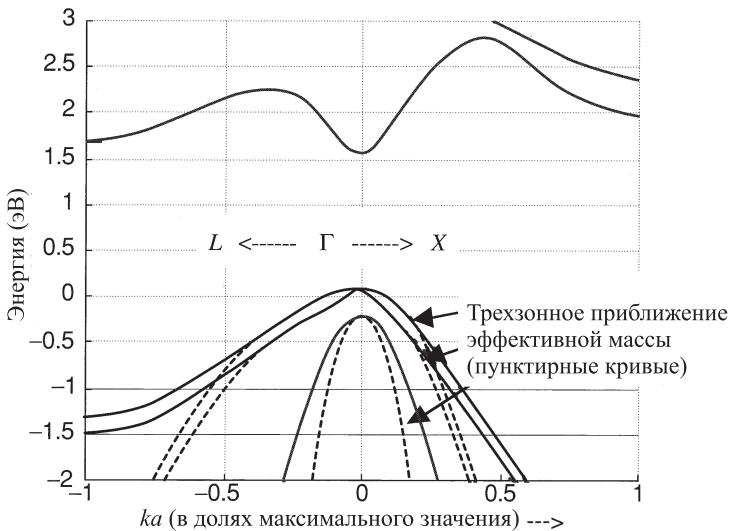


Рис. 7.4.2. Сплошные линии показывают полную зонную структуру, полученную в рамках  $sp^3s^*$ -модели, описанной в главе 5. Пунктирные кривые показывают дисперсионные зависимости, найденные из трехзонной модели эффективных масс (выражение (7.4.3)) с параметрами, подобранными для наилучшего согласования в начале дисперсии

зовалась при получении однозонного гамильтониана в приближении эффективной массы (см. рис. 7.1.3) из однозонного выражения для  $h(\vec{k})$  (см. выражение (7.1.2)).

Аналогичный подход используется для записи матричных блоков  $[H_{nm}]$  в трехзонной модели (выражение (7.4.3)). Для описания узкозонных полупроводников часто используют четырехзонную модель, в которой матрицы  $[H_{nm}]$  имеют размер  $(8 \times 8)$ . Многозонные модели эффективных масс дают не слишком большие преимущества по сравнению с микроскопическими атомными моделями, такими как  $sp^3s^*$ -модель. Но все же достигаемое при этом упрощение (вычислительное или даже методологическое) является существенным в силу двух причин:

(1) Матрицы  $[H_{nm}]$  имеют существенно меньший размер (сравните с матрицами  $(20 \times 20)$  для  $sp^3s^*$ -модели).

(2) Сетка для численного счета может быть гораздо более грубой и иметь более простую структуру по сравнению с реальной атомной ре-

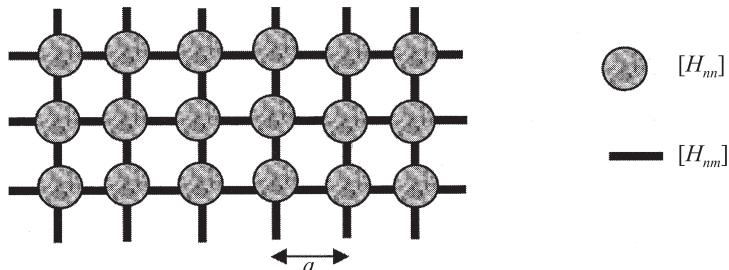


Рис. 7.4.3. Гамильтоновская матрица многозонной модели в приближении эффективной массы может быть представлена схематически как трехмерная сеть из единичных ячеек (не имеющих отношения к реальной кристаллической структуре), описываемых блоками  $[H_{nn}]$ , взаимодействующими с ближайшими соседями через блоки  $[H_{nm}]$ . Данные блоки будут иметь размер  $(4 \times 4)$  или  $(6 \times 6)$  в зависимости от того, используется ли двухзонная (выражение (7.4.2)) или трехзонная (выражение (7.4.3)) модель

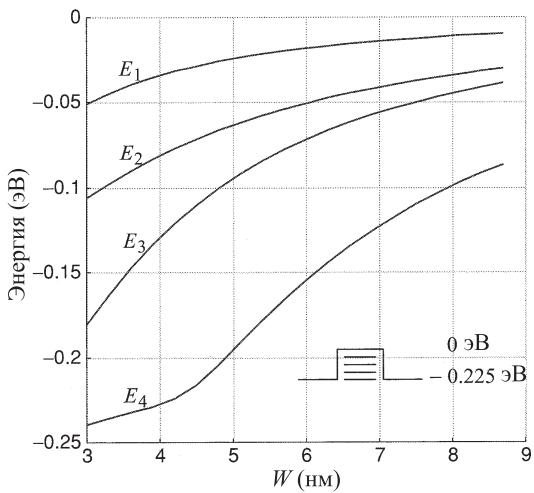


Рис. 7.4.4. Энергия четырех высших уровней в квантовой яме в GaAs, находящейся между барьерами из  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$  (структура изображена на вставке). Энергии уровней как функции ширины ямы  $W$  рассчитаны в рамках двухзонной модели эффективной массы

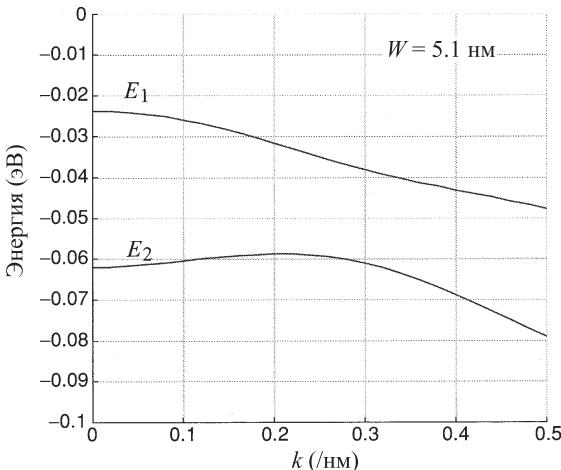


Рис. 7.4.5. Дисперсионная зависимость  $E(\vec{k})$  как функция  $k_y$  при  $k_x = 0$ . Показаны спектры двух высших подзон в квантовой яме ширины  $W = 5.1$  нм, рассчитанные в рамках двухзонной модели эффективной массы

шеткой (к примеру, простую кубическую вместо ГЦК решетки). Это приведет в итоге к гамильтониану меньшего размера, имеющему более простую структуру. Небольшой пример, иллюстрирующий основную идею, будет рассмотрен ниже, но, как я уже сказал ранее, в этой книге не будут обсуждаться многозонные модели (или любые модели, вне приближения эффективной массы в одной зоне).

**Пример.** На рис. 7.4.4 показаны энергии четырех высших уровней как функции ширины ямы  $W$  в квантовой яме в GaAs, расположенной между барьерами из  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ . Расчет проведен в рамках двухзонной модели эффективной массы при значениях параметров:

$$\text{GaAs: } E_V = 0 \text{ эВ}, \gamma_1 = 6.85, \gamma_2 = 2.1, \gamma_3 = 2.9;$$

$$\text{AlAs: } E_V = 0.75 \text{ эВ}, \gamma_1 = 3.45, \gamma_2 = 0.68, \gamma_3 = 1.29,$$

которые при помощи линейной интерполяции определяют соответствующие параметры для сплава GaAs-AlGaAs. На рис. 7.4.5 показаны дисперсионные кривые  $E(\vec{k})$  как функции  $k_y$  при  $k_x = 0$  для двух высших подзон в квантовой яме ширины  $W = 5.1$  нм.

## Упражнения

### У.7.1.

(а) Постройте зависимость  $E(k)$  вдоль направлений Г–Х и Г–Л для гамильтониана

$$h(\vec{k}) = E_c + \frac{\hbar^2 k^2}{2m_c},$$

и сравните с графиком, полученным из  $sp^3s^*$ -модели (см. упражнение У.5.2) в соответствующем интервале энергий и волновых чисел (сравните с рис. 6.1.2). При каких значениях параметров  $E_c$  и  $m_c$  получается наилучшее соответствие?

(б) Используя приближение эффективной массы в одной зоне, рассчитайте уровни энергии двух низших уровней квантовой ямы в GaAs, расположенной между барьерами из  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$  как функции ширины ямы. В расчетах используйте следующие значения параметров: для GaAs  $E_c = 0$  и  $m_c = 0.07m$ , а для AlAs  $E_c = 1.25$  эВ и  $m_c = 0.15m$ . Для сплава AlAs–GaAs используйте линейную интерполяцию и сравните результат с рис. 7.1.5.

(в) Используйте однозонную модель для расчета дисперсионной зависимости  $E(k)$  как функции модуля волнового вектора  $\vec{k} = \{k_x, k_y\}$  для двух низших подзон квантовой ямы шириной  $W = 6.9$  нм. Используйте те же параметры, как и в задаче (б), и сравните результат с рис. 7.1.6.

### У.7.2.

(а) Рассматривая модель МДП-конденсатора, показанную на рис. 7.2.2, рассчитайте самосогласованный профиль в зоне проводимости и электронную плотность, используя дискретную решетку с шагом  $a = 0.3$  нм. Расчет проводите в предположении, что (1) толщина слоя оксида равна 2.1 нм, а толщина канала равна 3 нм; (2) Значение  $\mu = 0$ , значение  $E_c = 0$  в кремнии и  $E_c = 3$  эВ в оксиде; (3) диэлектрическая постоянная  $\varepsilon = 4\varepsilon_0$  и эффективная масса  $m_c = 0.25m_0$  всюду; (4) напряжение на затворе  $V_G = 0.25$  В. Повторите расчет для канала шириной 9 нм, а также для случая асимметрично приложенного напряжения на затворе, когда к одной его половине приложено 0 В, а к другой половине приложено 0.25 В. Сравните результат с рис. 7.2.5.

(б) Рассчитайте электронную плотность на единицу площади как функцию напряжения на затворе (приложенного симметрично к обеим половинам затвора) для структуры с каналом шириной 3 нм и 9 нм. Сравните

результат с рис. 7.3.1. Рассчитайте эффективную емкость структуры, исходя из наклона кривой в состоянии ВКЛ, и определите величину эффективного расстояния  $d$  между пластинами конденсатора, приравняв найденную емкость величине  $2\varepsilon/d$ .

### У.7.3.

(а) Постройте зависимость  $E(k)$  вдоль направлений Г–Х и Г–Л в двухзонной модели (см. выражение (7.4.2)) и сравните с графиком, полученным в рамках  $sp^3s^*$ -модели (см. упражнение У.5.3), в соответствующем интервале энергий и волновых чисел (сравните с рис. 7.4.1). Используйте следующие значения параметров:  $\gamma_1 = 6.85$ ,  $\gamma_2 = 2.1$ ,  $\gamma_3 = 2.9$ .

(б) Используйте двухзонную модель для расчета энергии как функцию ширины ямы для четырех высших уровней валентной зоны в яме GaAs, расположенной между барьерами из  $Al_{0.3}Ga_{0.7}As$ . Предположите, что в GaAs  $E_V = 0$  эВ,  $\gamma_1 = 6.85$ ,  $\gamma_2 = 2.1$ ,  $\gamma_3 = 2.9$ , а в AlAs эти параметры равны  $E_V = 0.75$  эВ,  $\gamma_1 = 3.45$ ,  $\gamma_2 = 0.68$ ,  $\gamma_3 = 1.29$ . Параметры сплава GaAs-AlGaAs определите при помощи линейной интерполяции от AlAs к GaAs. Сравните полученный результат с рис. 7.4.4.

(в) Используйте двухзонную модель для расчета дисперсионной зависимости  $E(\vec{k})$  как функции  $k_y$  при  $k_x = 0$  для четырех высших подзон в квантовой яме ширины  $W = 5.1$  нм, используя те же значения параметров, что и в задаче (б). Сравните результат с рис. 7.4.5.

**У.7.4.** Небольшая структура, описываемая гамильтонианом с матрицей  $(2 \times 2)$

$$H = \begin{pmatrix} 0 & t_0 \\ t_0 & 0 \end{pmatrix}$$

и находящаяся в равновесии с резервуаром, имеет электрохимический потенциал  $\mu$ , отвечающий температуре  $T = 0$ . Запишите матрицу плотности, если (1)  $\mu = 0$ ; (2)  $\mu = 2$ , предполагая, что (а)  $t_0 = +1$  и (б)  $t_0 = -1$ .

**У.7.5.** Металлическая нанотрубка радиуса  $a$  имеет лишь одну подзону в интересующей области энергий, чья дисперсионная зависимость описывается формулой  $\varepsilon(k) = \hbar v k$ , где  $v$  есть постоянная, а  $k$  обозначает модуль волнового вектора вдоль длины нанотрубки. Предполагая, что (1) электрохимический потенциал  $\mu = 0$ , и (2) коаксиальный затвор радиуса  $b$  окружает нанотрубку, так, что электростатическая емкость  $C_E = 2\pi\varepsilon_r\varepsilon_0/\ln(b/a)$ , ответить на следующие вопросы:

(а) Как отклонение электронной плотности (на единицу длины) от равновесного значения связано с величиной напряжения  $V_G$  на затворе?

(б) Обеспечит ли это изменение хороший транзисторный эффект?

**У.7.6.** Запишите соответствующее уравнение Шредингера в приближении эффективной массы, если в одномерном случае дисперсионная зависимость задана выражением

$$E = E_c + \frac{\hbar^2 k^2}{2m_c} + \alpha k^4.$$

## Глава 8

# Уширение уровней

В первой главе уже обсуждалась задача о протекании тока через канал, соединенный с двумя контактами, характеризующимися различными значениями электрохимического потенциала. Один из контактов обеспечивает заполнение канала носителями, в то время как другой контакт обеспечивает их уход из канала. В следующей главе будет проведен квантовомеханический анализ данной транспортной задачи. В настоящей главе моя цель состоит в том, чтобы на простом примере *канала, соединенного с одним контактом*, описать некоторые ключевые понятия, которые иллюстрирует рис. 8.1.

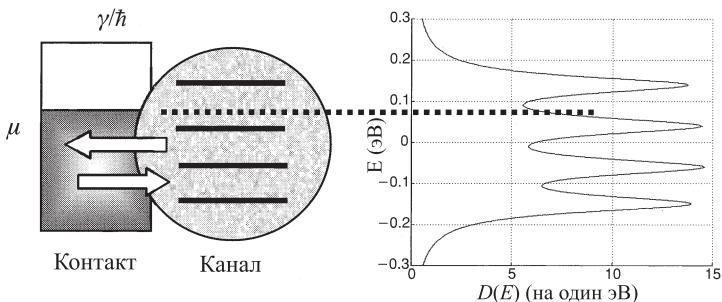


Рис. 8.1. Канал, соединенный с одним контактом. Уширение уровней формирует непрерывную плотность состояний, как это показано на графике справа

Поскольку учитывается лишь один контакт, канал просто приходит с ним в равновесие, и ток в стационарных условиях отсутствует. В подобной постановке задачи не требуется учитывать роль нескольких контактов и рассматривать неравновесные ситуации. Мы постараемся сконцентрироваться на физической стороне явлений, возникающих при соединении канала с массивным контактом, когда дискретные уровни уширяются и формируется непрерывная плотность состояний, как показано в правой части рис. 8.1.

В главе 1 я ввел это уширение без какого-либо формального обоснования, сославшись на требование правильного значения проводимости. В этой главе моей целью является квантовомеханический расчет, в котором уширение уровней возникнет как следствие соотношения «неопределеностей»  $\gamma = \hbar/\tau$ , связывающего уширение со скоростью ухода  $1/\tau$  электрона из канала в контакт. Более того, будет показано, что в общем случае уширение является не числом  $\gamma$ , как предполагалось ранее, а матрицей  $[\Gamma]$  того же размера, что и гамильтоновская матрица  $[H]$ , причем матрица уширения, в отличие от  $[H]$ , может зависеть от энергии.

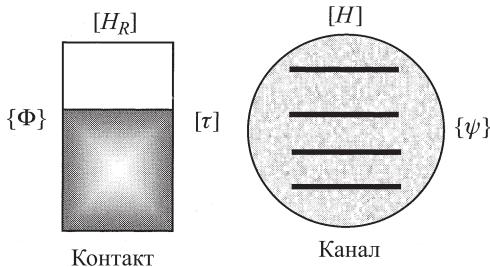


Рис. 8.2. Канал, описываемый гамильтоновской матрицей  $[H]$ , взаимодействует с контактом (взаимодействие определяется матрицей  $[\tau]$ ), которому, в свою очередь, отвечает матрица  $[H_R]$ . Можно перейти к уравнению для одного канала, придав ему форму (8.2)

Стартовав в этом параграфе с уравнения Шредингера, описывающего всю систему, а именно канал + контакт (см. рис. 8.2):

$$E \begin{pmatrix} \psi \\ \Phi \end{pmatrix} = \begin{pmatrix} H & \tau \\ \tau^+ & H_R \end{pmatrix} \begin{pmatrix} \psi \\ \Phi \end{pmatrix}, \quad (8.1)$$

я покажу, что контакт (или резервуар, имеющий размер  $R$ ) можно исключить, чтобы получить уравнение только для канала, имеющее следующий вид:

$$E\{\psi\} = [H + \Sigma]\{\psi\} + \{S\}. \quad (8.2)$$

Это основной результат, который мы будем использовать как отправную точку в параграфе 8.1 при обсуждении механизма протекания тока между двумя контактами. Здесь слагаемое  $\{S\}$  описывает возбуждение канала электронными волнами из контакта, в то время как матрица *собственно-энергетической функции*  $[\Sigma]$  определяет модификацию гамильтонiana  $[H]$ . Можно сказать, что введение матрицы  $[\Sigma]$  позволяет учесть

«границные условия» приблизительно так же, как мы делали это, добавляя пару слагаемых к  $[H]$ , чтобы учесть периодические граничные условия (см. уравнение (1.3.3)).

Существуют, впрочем, два фактора, которые не позволяют рассматривать величину  $[\Sigma]$  как простую модификацию  $[H]$ . Во-первых, матрица  $[\Sigma]$  сама зависит от энергии, и поэтому впредь мы не сможем говорить о резонансных уровнях энергии, определяемых собственными значениями  $[H]$ . Поскольку  $[\Sigma]$  зависит от энергии, каждое собственное значение  $\varepsilon_n$  матрицы  $[H + \Sigma(E = \varepsilon_n)]$  следует находить итерационным методом. При проведении расчетов удобно считать, что энергия  $E$  является *независимой переменной*, и рассматривать отклик структуры на падающие на нее электроны. Физический смысл величины  $[\Sigma]$  будет различным в зависимости от того, зависит она от энергии или нет. Ранее мы искали собственные значения энергии в изолированном канале, что напоминало поиск резонансных частот струны (см. рис. 2.1.3). Теперь мы хотим знать, как «струна» реагирует на возбуждение произвольной частоты, т. е. каким образом формируется отклик канала на возбуждение, созданное электроном с любой энергией, пришедшим из контакта.

Вторым важным отличием собственно-энергетической функции  $\Sigma$  является то, что в отличие от матрицы  $[H]$  она НЕ является эрмитовой, так что собственные значения суммы  $H + \Sigma$  комплексные. В самом деле, антиэрмитова часть оператора  $\Sigma$

$$\Gamma = i[\Sigma - \Sigma^+] \quad (8.3)$$

может рассматриваться как матричный вариант параметра уширения  $\gamma$ , введенного ранее для структуры с одним уровнем. В параграфе 8.2 будет установлена его связь с уширением плотности состояний в канале. Далее в параграфе 8.3 будет показано, что уширение связано с конечным временем жизни электронных состояний, отражающим тот факт, что электрон не остается навсегда в канале, а постепенно уходит в контакт.

Не следует удивляться тому факту, что мы получили неэрмитову матрицу  $H + \Sigma$ , исходя из эрмитовой матрицы в выражении (8.1). Фактически мы не исходили из эрмитовой матрицы, а просто прибавили бесконечно малую величину  $i0^+$  к гамильтониану резервуара  $H_R$ , сделав его «слегка» неэрмитовым. Эта бесконечно малая добавка к гамильтониану резервуара делает его «слегка» неэрмитовым и приводит к конечному, не зависящему от значения  $0^+$  Г-уширению уровней в канале. Но подчеркнем, что этот с виду простой шаг требует более аккуратного обсуждения, поскольку он фактически превращает обратимую систему в необратимую. Мы выясним в параграфе 8.4, сколь велика должна быть массивная система для того,

чтобы она могла выполнять функции резервуара, обеспечивающего необратимое поведение системы.

В этой главе будет введено понятие собственно-энергетической функции, с помощью которой учитывается связь контактов, таких как исток или сток, с каналом. Однако понятие собственно-энергетической функции является гораздо более общим и может быть использовано для описания различных взаимодействий (как обратимых, так и необратимых) не только с контактами, а с любым окружением. Это понятие является одним из краеугольных для физики систем многих частиц, где оно используется для описания сложных взаимодействий, по сравнению с которыми наша задача о контактах является относительно простой. Наша задача допускает решение и более простыми, хотя и не столь «элегантными» методами. Моя целью, однако, является не демонстрация элегантного решения простой задачи, а введение на простом примере в оборот глубокого и полезного понятия. В десятой главе мы используем это понятие для описания взаимодействия с «контактами» совершенно другой природы, а именно системами фотонов и фононов.

## 8.1. Открытые системы

В этом параграфе будет получено уравнение вида (см. уравнение (8.2))

$$E\{\psi\} = [H + \Sigma]\{\psi\} + \{S\}, \quad (8.1.1)$$

описывающее открытую систему. Такое уравнение отличается от уравнения  $E\{\psi\} = [H]\{\psi\}$ , использованного ранее для описания замкнутых систем. Основная идея может быть легко продемонстрирована на следующем простом примере.

**Простой пример.** Рассмотрим полубесконечную одномерную проволоку, описываемую однозонным гамильтонианом в приближении эффективной массы. Такая структура схематически показана на рис. 8.1.1. Будем считать, что первая точка проволоки с индексом «0» соответствует каналу, а остальные точки проволоки с индексами  $n < 0$  — контактам.

Если бы «канал» не взаимодействовал с «контактом», он описывался бы уравнением

$$E\psi = (E_c + 2t_0)\psi.$$

Как только возникает связь с «контактом», это уравнение меняется и принимает следующий вид:

$$E\psi = (E_c + 2t_0)\psi - t_0\Phi_{-1}, \quad (8.1.2)$$

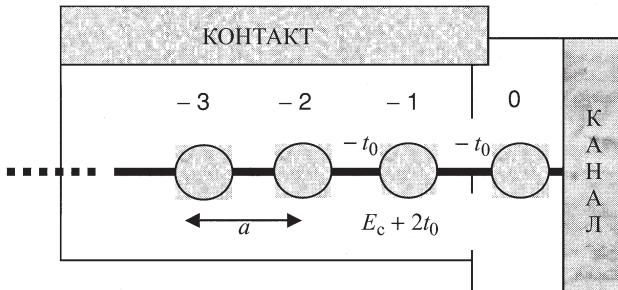


Рис. 8.1.1. Простой пример: полубесконечная проволока, описываемая однозонным гамильтонианом в приближении эффективной массы. Первая точка системы «0» рассматривается как канал, а вся оставшаяся часть как контакт

где волновые функции  $\Phi_n$  в контактах удовлетворяют бесконечной последовательности уравнений (с  $n < 0$ )

$$E\Phi_n = -t_0\Phi_{n-1} + (E_c + 2t_0)\Phi_n - t_0\Phi_{n+1}. \quad (8.1.3)$$

Поскольку все уравнения этой бесконечной последовательности имеют одинаковую структуру, можно воспользоваться приемом, часто используемым в расчетах зонной структуры (см. выражение (5.2.4)), и записать решения в форме плоских волн, задаваемых квантовым числом  $k$ . Предполагая, что решение может быть представлено волной, приходящей из контакта и волной, отраженной в канале, можно записать

$$\Phi_n = B \exp(+ikna) + C \exp(-ikna), \quad (8.1.4)$$

где

$$E = E_c + 2t_0(1 - \cos ka). \quad (8.1.5)$$

Используя выражение (8.1.4), можно записать

$$\psi \equiv \Phi_0 = B + C$$

и

$$\Phi_{-1} = B \exp(-ika) + C \exp(+ika),$$

так что

$$\Phi_{-1} = \psi \exp(+ika) + B[\exp(-ika) - \exp(+ika)].$$

Подставляя это выражение обратно в уравнение (8.1.2), получим выражение

$$E\psi = \underbrace{(E_c + 2t_0)\psi}_{H\psi} - \underbrace{t_0 \exp(+ika)\psi}_{\Sigma\psi} + \underbrace{t_0 B [\exp(+ika) - \exp(-ika)]}_S, \quad (8.1.6)$$

которое имеет именно тот вид, который нам необходим, с функциями

$$\Sigma = -t_0 \exp(+ika) \quad (8.1.7a)$$

и

$$S = it_0 2B \sin ka. \quad (8.1.7b)$$

Обратите внимание на то, что собственно-энергетическая функция  $\Sigma$  неэрмитова и не зависит от амплитуд  $B$  и  $C$  волновой функции. Она введена для того, чтобы описать уход волновой функции из канала в контакт. Слагаемое  $S$ , с другой стороны, связано с возбуждением состояний канала, волнами, идущими из контакта, и поэтому оно пропорционально амплитуде  $B$ . После изучения этого простого примера мы обратимся к обсуждению более общего случая, а именно канала, взаимодействующего с произвольным контактом.

**Общая формулировка задачи.** Рассмотрим вначале канал без электронов, не связанный с контактом, как это показано на рис. 8.1.2a. Пусть электроны в контакте описываются волновыми функциями  $\{\Phi_R\}$ , удовлетворяющими уравнению Шредингера для изолированного контакта

$$[EI_R - H_R]\{\Phi_R\} = \{0\},$$

где  $[H_R]$  есть гамильтониан контакта, а  $[I_R]$  есть единичная матрица того же размера, что и  $[H_R]$ . Изменим слегка это уравнение, записав его как

$$[EI_R - H_R + i\eta]\{\Phi_R\} = \{S_R\}, \quad (8.1.8a)$$

где  $[\eta] = 0^+[I_R]$  есть произведение положительной бесконечно малой величины, чье значение мы обсудим ниже в параграфе 8.4 на единичную матрицу. Здесь я просто отмечу, что слагаемое  $i[\eta]\{\Phi_R\}$  в левой части выражения (8.1.8a) описывает повторную инжекцию электронов из внешних источников: такое удаление и повторная инжекция существенны для поддержания контакта в равновесии при постоянном значении электрохимического потенциала.

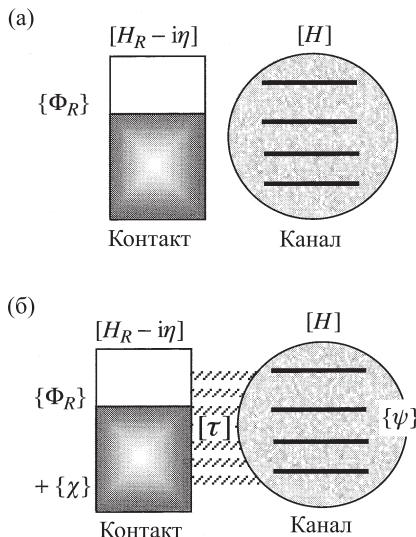


Рис. 8.1.2. (а) Канал без электронов внутри, отсоединенный от контактов, в которых электроны заполняют состояния  $\{\Phi_R\}$ . (б) При соединении с контактом волновые функции  $\{\Phi_R\}$  «проникают» вглубь структуры, формируя волновую функцию  $\{\psi\}$  в канале, которая, в свою очередь, порождает рассеянную волну  $\{\chi\}$  в контакте

Результаты, которые мы получим далее, не зависят от слагаемого  $S_R$ , которое было введено в уравнении (8.1.8a). В самом деле, может показаться, что мы могли бы приравнять его к  $i[\eta]\{\Phi_R\}$ , и тем самым свести уравнение (8.1.8a) к уравнению Шредингера  $[E_{IR} - H_R]\{\Phi_R\} = \{0\}$ . Однако, как я подчеркнул во введении к этой главе, при переходе от уравнения Шредингера к уравнению (8.1.8a) смысл величины  $E$  меняется радикально: она более не является собственным значением энергии, а становится независимой переменной, описывающей энергию возбуждения от внешних источников. В уравнении Шредингера  $\{\Phi_R\}$  есть собственная функция гамильтониана  $[H_R]$ , которая не равна тождественно нулю лишь только тогда, когда энергия  $E$  совпадает с одним из собственных матрицы  $[H_R]$ . С другой стороны, функции  $\{\Phi_R\}$  в уравнении (8.1.8a) тождественно не равны нулю при всех энергиях  $E$ , расположенных в области узкого пика вблизи собственных значений  $[H_R]$ , причем ширина пика зависит от бесконечно малой величины  $0^+$ .

Когда канал связывается с контактом, как это показано на рис. 8.1.2б, волновые функции в контакте «распиваются», и формируется волновая функция  $\{\psi\}$  в структуре, которая, в свою очередь, порождает рассеянную волну  $\{\xi\}$ . Полная волновая функция составной системы контакт-структур будет удовлетворять «блочному» уравнению Шредингера, которое может быть записано в виде двух блоков:

$$\begin{pmatrix} \text{контакт} & EI_R - H_R + i\eta & \text{структура} \\ \text{контакт} & -\tau & EI - H \\ \text{структура} & & \end{pmatrix} \begin{pmatrix} \Phi_R + \chi \\ \psi \end{pmatrix} = \begin{pmatrix} S_R \\ 0 \end{pmatrix}, \quad (8.1.8b)$$

где  $[H]$  есть гамильтониан структуры. Различные величины, фигурирующие в уравнении (8.1.8b) (за исключением энергии  $E$ ), не являются числами. Они являются матрицами различного размера, так

гамильтониан контакта  $[H_R]$ , единичная матрица  $[I_R]$ , затухание  $[\eta]$ : матрицы размера  $(R \times R)$ ;

гамильтоновская матрица в канале  $[H]$ , единичная матрица  $[I]$ : матрицы размера  $(d \times d)$ ;

гамильтониан, ответственный за связи  $[\tau]$ : матрица размера  $(d \times R)$ , а матрица  $[\tau^+]$  имеет размер  $(R \times d)$ .

Векторы-столбцы:

волновая функция в контакте  $\{\Phi_R\}$ , функция  $\{\xi\}$ , функция в источнике  $\{S_R\}$  имеют размер  $(R \times 1)$ ;

волновая функция в структуре  $\{\psi\}$  имеет размер  $(d \times 1)$ .

Обратите внимание на то, что, переходя от уравнения (8.1.8a) к уравнению (8.1.8b), мы предполагаем, что слагаемое с  $\{S_R\}$  в правой части, описывающее повторную инжекцию электронов из внешних источников, остается неизменным. Это позволяет нам использовать уравнение (8.1.8a), чтобы исключить  $\{S_R\}$  из уравнения (8.1.8b) и записать следующее:

$$[EI_R - H_R + i\eta]\{\xi\} - [\tau^+]\{\psi\} = \{0\},$$

$$[EI - H]\{\psi\} - [\tau]\{\xi\} = [\tau]\{\Phi_R\}.$$

Пользуясь правилами матричной алгебры, можно выразить  $\{\xi\}$  через  $\{\psi\}$ , что приведет нас к выражению

$$\begin{aligned} \{\xi\} &= G_R \tau^+ \{\psi\}, \\ G_R &\equiv [EI_R - H_R + i\eta]^{-1} \end{aligned} \quad (8.1.9)$$

и

$$[\eta] = 0^+[I_R],$$

и подставить это во второе уравнение, откуда получим

$$[EI - H - \Sigma]\{\psi\} = \{S\}. \quad (8.1.10)$$

Здесь введены обозначения

$$\Sigma \equiv \tau G_R \tau^+, \quad S \equiv \tau \Phi_R. \quad (8.1.11)$$

Уравнение (8.1.10) имеет тот самый вид (см. уравнение (8.1.1)), который мы желали обосновать, а уравнение (8.1.11) дает нам формальное выражение, позволяющее вычислить  $\Sigma$  и  $S$ . Из выражения (8.1.9) следует, что величина  $G_R$  описывает свойства изолированного контакта, поскольку в него входит лишь гамильтониан контакта  $H_R$ . Эта величина называется функцией Грина изолированного контакта. Физический смысл этого важного понятия обсуждается в параграфе 8.2.

**Вычисление  $\Sigma$  и  $S$ .** Из выражения (8.1.11) не вполне ясно, как им можно воспользоваться в конкретной ситуации, т. к. матрица  $G_R$  имеет очень большой размер ( $R \times R$ ), поскольку размер резервуара  $R$  зачастую является бесконечным. Однако, хотя формально матрица  $[\tau]$  имеет размер  $(d \times R)$ , в «координатном пространстве» она связывает с каналом лишь  $r$  элементов, ближайших к нему. Следовательно, требуется уменьшить ее размер до матрицы порядка  $(d \times r)$  и записать

$$\begin{matrix} \Sigma \\ (d \times d) \end{matrix} \equiv \frac{\tau}{(d \times r)(r \times r)(r \times d)} g_R \tau^+ \quad (8.1.12a)$$

и

$$\begin{matrix} S \\ (d \times 1) \end{matrix} \equiv \frac{\tau}{(d \times r)(r \times 1)} \phi_R, \quad (8.1.12b)$$

где поверхностная функция Грина  $g_R$  есть подматрица  $(r \times r)$  внутри полной функции Грина  $G_R$ , включающая лишь  $r$  точек поверхности, а вектор  $\{\phi_R\}$  представляет вектор размера  $(r \times 1)$ , принадлежащий подпространству волновой функции  $\Phi_R$  в контактах. Для простого примера, который мы обсуждали в начале параграфа, можно показать, что

$$\tau = -t_0, \quad \phi_R = -i2B \sin ka, \quad (8.1.13a)$$

а

$$g_R = -(1/t_0) \exp(ika), \quad (8.1.13b)$$

что после подстановки в выражения (8.1.12a), (8.1.12b) приводит к тем же результатам, что были получены ранее (сравните с выражениями (8.1.7a), (8.1.7b)). Выражение для функции  $\phi_R$  получено с учетом того, что она равна волновой функции  $\Phi_{-1}$  контакта (в точке, где он соединяется с каналом), при условии, что взаимодействие с каналом отсутствует. Заметим, что отсутствие взаимодействия эквивалентно граничному условию  $\Phi_0 = = 0$ , откуда следует, что  $C = -B$ , а соответствующая величина  $\Phi_{-1}$  равна  $(-i2B \sin ka)$ , как предполагалось выше. Для того чтобы найти выражение для  $g_R$ , требуется провести дополнительные чисто алгебраические расчеты, и мы оставляем их читателю (см. упражнение У.8.1, приведенное в конце главы).

Расчет величин  $\Sigma$  и  $S$  может быть выполнен и в представлении собственных состояний контакта, когда гамильтониан контакта  $H_R$  является диагональным, а функция Грина  $G_R$  выражается через собственные значения  $\varepsilon_n$  матрицы  $H_R$ :

$$[G_R(E)] = \begin{pmatrix} \frac{1}{E - \varepsilon_1 + i0^+} & 0 & \dots \\ 0 & \frac{1}{E - \varepsilon_2 + i0^+} & \dots \\ \dots & \dots & \dots \end{pmatrix}. \quad (8.1.14)$$

В этом представлении матрица взаимодействия  $[\tau]$  не сводится к матрице меньшего размера, и суммирование следует выполнять по всем состояниям резервуара, т. е. в бесконечных пределах:

$$\Sigma_{ij}(E) = \sum_n \frac{[\tau]_{in} [\tau^+]_{nj}}{E - \varepsilon_n + i0^+}, \quad (8.1.15a)$$

$$S_{ij}(E) = \sum_n [\tau]_{in} \{\Phi_R\}_{nj}. \quad (8.1.15b)$$

Однако суммирование такого вида часто может быть выполнено аналитически через переход к интегрированию. К примеру, в упражнении У.8.2 показано, как, стартуя с выражения (8.1.15a), можно для рассмотренного простого примера получить наши старые результаты (уравнение (8.1.7a)).

Прежде чем продолжить изложение, позвольте вкратце подвести итоги. Канал, описываемый гамильтоновской матрицей  $[H]$  размера  $(d \times d)$ , взаимодействует с контактом, описываемым матрицей  $[H_R]$  размера  $(R \times R)$ , причем обычно предполагается, что размер матрицы резервуара  $R$  много больше  $d$  ( $R \gg d$ ). Мы продемонстрировали, что влияние резервуара на

структуре описывается при помощи матричной собственno-энергетической функции  $\Sigma(E)$  размером  $(d \times d)$  и слагаемым, связанным с источником  $S(E)$ , описывающимся матрицей размера  $(d \times 1)$ .

## 8.2. Локальная плотность состояний

В предыдущем разделе установлено, что канал, взаимодействующий с контактом, может быть описан модифицированным уравнением Шредингера вида  $E\{\psi\} = [H + \Sigma]\{\psi\} + \{S\}$ , в котором слагаемое  $\{S\}$  описывает возбуждение, исходящее из контакта (или слагаемое источника), а собственно-энергетическая функция отвечает за изменения, связанные со взаимодействием с контактом. В отличие от гамильтониана  $[H]$ , матрица  $[H + \Sigma]$  имеет комплексные собственные значения, мнимая часть которых отвечает как за уширение плотности квантовых состояний, так и за их конечное время жизни. В этом параграфе разговор пойдет о первом из этих эффектов. Мы выясним, как следует рассчитывать плотность состояний в открытой системе, а в следующем параграфе обсудим и второй эффект.

Рассмотрим систему, состоящую из канала и контакта. Ранее при выводе выражения (6.2.14) было установлено, что в системе, имеющей собственные значения  $\varepsilon_\alpha$ , плотность состояний определена выражением

$$D(E) = \sum_{\alpha} \delta(E - \varepsilon_{\alpha}). \quad (8.2.1)$$

Возникает вопрос. Как объяснить тот факт, что различные энергетические уровни вносят не одинаковый вклад в уширение исходных дискретных уровней, как это показано на рис. 8.1? Разве выражение (8.2.1) не говорит о том, что каждый уровень отвечает одной дельта-функции, чей вес равен единице? Чтобы ответить на этот вопрос, следует принять во внимание, что плотность состояний, определенная как (8.2.1), не учитывает пространственное распределение электронной плотности. Если мы интересуемся локальной плотностью состояний в канале, то каждое состояние следует учитывать в соответствии с его весом, определяемым квадратом модуля волновой функции, в канале ширины  $d$ :

$$D(d; E) = \sum_{\alpha} |\phi_{\alpha}(d)|^2 \delta(E - \varepsilon_{\alpha}).$$

Предположим, что структура с одним энергетическим уровнем  $\varepsilon$  не взаимодействует с резервуаром, спектр которого состоит из близко распо-

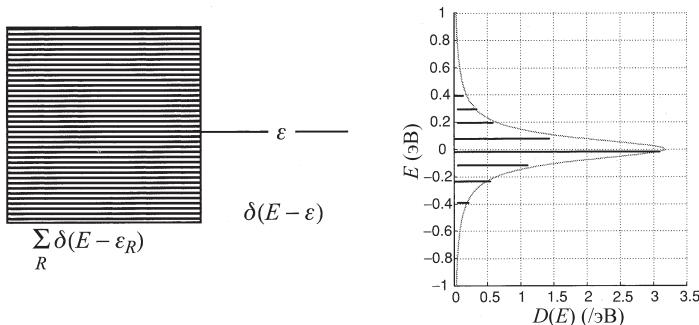


Рис. 8.2.1. Канал с одним энергетическим уровнем  $\varepsilon$ , взаимодействующий с резервуаром, с большим числом близко расположенных уровней  $\{\varepsilon_n\}$ . В отсутствие взаимодействия с резервуаром локальная плотность состояний в канале имеет одиничный резкий пик. При учете взаимодействия появляется последовательность уровней переменной высоты, отвечающих доли квадрата модуля волновой функции в области канала

ложенных уровней  $\{\varepsilon_R\}$ . В этом случае полная плотность состояний определяется выражением

$$D(E) = \sum_{\alpha} \delta(E - \varepsilon_{\alpha}) = \delta(E - \varepsilon) + \sum_n \delta(E - \varepsilon_n),$$

в то время как локальная плотность состояний в канале записывается просто так

$$D(d; E) = \sum_{\alpha} |\phi_{\alpha}(d)|^2 \delta(E - \varepsilon_{\alpha}) = \delta(E - \varepsilon),$$

поскольку волновые функции состояний резервуара имеют весьма малые амплитуды в канале. Как только будет учтено взаимодействие канала с резервуаром, описание нашей системы перестанет быть столь простым. Именно, у системы будет не один уровень с волновой функцией, сосредоточенной преимущественно в канале, но, кроме этого, появятся дополнительные состояния, волновые функции которых частично проникают в канал. Если мы обратимся к локальной плотности состояний в канале, то увидим последовательность максимумов переменной высоты, пропорциональных квадратам модуля соответствующих волновых функций канала (рис. 8.2.1).

В общем случае локальную плотность состояний (LDOS)  $D(\vec{r}; E)$  следует определить как функцию, учитывающую вес каждого уровня, пропор-

циональный квадрату модуля волновой функции в данной точке  $\vec{r}$ :

$$D(\vec{r}; E) = \sum_{\alpha} |\Phi_{\alpha}(\vec{r})|^2 \delta(E - \varepsilon_{\alpha}). \quad (8.2.2)$$

Это выражение можно рассматривать как деленный на  $2\pi$  диагональный элемент другой более общей величины, называемой спектральной функцией  $[A(E)]$ :

$$A(\vec{r}, \vec{r}'; E) = 2\pi \sum_{\alpha} \phi_{\alpha}(\vec{r}) \delta(E - \varepsilon_{\alpha}) \phi_{\alpha}^*(\vec{r}'), \quad (8.2.3)$$

аналогично тому, как электронная плотность

$$n(\vec{r}) = \sum_{\alpha} |\phi_{\alpha}(\vec{r})|^2 f_0(\varepsilon_{\alpha} - \mu) \quad (8.2.4)$$

может рассматриваться в качестве диагонального элемента матрицы плотности

$$\rho(\vec{r}, \vec{r}') = \sum_{\alpha} \phi_{\alpha}(\vec{r}) f_0(\varepsilon_{\alpha} - \mu) \phi_{\alpha}^*(\vec{r}'). \quad (8.2.5)$$

В параграфе 4.3 уже отмечалось, что выражение (8.2.5) является координатным представлением матричного соотношения

$$[\rho] = f_0([H] - \mu[I]). \quad (8.2.6)$$

Повторяя те же рассуждения, запишем спектральную функцию в виде

$$[A(E)] = 2\pi \delta(E[I] - [H]) \quad (8.2.7)$$

и будем рассматривать уравнение (8.2.3) как соответствующее координатное представление. Если использовать собственные функции оператора  $H$  в качестве базисных, то в таком представлении матрица  $[H]$  будет диагональной:

$$[H] = \begin{pmatrix} \varepsilon_1 & 0 & 0 & \dots \\ 0 & \varepsilon_2 & 0 & \dots \\ 0 & 0 & \varepsilon_3 & \dots \\ \dots & & & \dots \end{pmatrix}.$$

Диагональной будет и матрица  $[A(E)]$ :

$$[A(E)] = 2\pi \begin{pmatrix} \delta(E - \varepsilon_1) & 0 & 0 & \dots \\ 0 & \delta(E - \varepsilon_2) & 0 & \dots \\ 0 & 0 & \delta(E - \varepsilon_3) & \dots \\ \dots & & & \dots \end{pmatrix}. \quad (8.2.8)$$

Выражение (8.2.3) определяет координатное представление этой матрицы. В принципе, мы могли бы записать спектральную функцию и в любом представлении. Тогда ее диагональные элементы определят в данном представлении умноженную на  $2\pi$  локальную плотность состояний при значении энергии  $E$ , подобно тому, как диагональные элементы матрицы плотности определяют локальную электронную плотность. Полное число электронов  $N$  равно сумме всех диагональных элементов, т. е. следу этой матрицы  $[\rho]$ , значение которого не зависит от представления:

$$N = \text{Tr}[\rho] = \sum_{\alpha} f_{\alpha}(\varepsilon_{\alpha} - \mu).$$

Подобным образом полное число состояний определяется следом матричной спектральной функции  $[A]$ , деленным на  $2\pi$ , и поэтому не зависит от представления. Оно может быть получено из исходного представления собственных функций:

$$D(E) = \frac{1}{2\pi} \text{Tr}[A(E)] = \sum_{\alpha} \delta(E - \varepsilon_{\alpha}). \quad (8.2.9)$$

**Правило сумм.** Заметим, что при вычислении полного числа состояний в данной точке интегрирование по всем энергиям должно дать единицу. В исходной структуре, которая содержит один уровень и взаимодействует с резервуаром, собственный уровень расплывается в последовательность пиков (рис. 8.2.1) различной высоты. Высота каждого пика будет пропорциональной вкладу уровней в волновую функцию структуры. Но если просуммировать вклады от всех уровней, результат будет тот же, что и для одного исходного уровня. То, что теряется в структуре с одним уровнем вследствие его гибридизации с другими уровнями, возвращается вновь, так что уширенный уровень в структуре имеет точно тот же вес, что и одиничный уровень, не взаимодействующий с резервуаром. Следовательно, это правило сумм может быть сформулировано следующим образом:

$$\int_{-\infty}^{+\infty} dE D(\vec{r}; E) = 1,$$

что может быть обосновано тем, что величина  $\int_{-\infty}^{+\infty} dE D(\vec{r}; E)$  есть диагональный элемент матрицы  $\int_{-\infty}^{+\infty} dE [A(E)] / 2\pi$ , вычисленный в координат-

ном представлении. Нетрудно увидеть из выражения (8.2.8), что в представлении собственных функций

$$\int_{-\infty}^{+\infty} \frac{dE}{2\pi} [A(E)] = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \quad \text{поскольку} \quad \int_{-\infty}^{+\infty} dE \delta(E - \varepsilon) = 1.$$

Можно утверждать, что эта величина будет иметь одинаковый вид *в любом представлении*, поскольку единичная матрица не изменяется при смене базиса.

**Функция Грина.** При вычислении спектральной функции удобно использовать равенство

$$\begin{aligned} 2\pi\delta(E - \varepsilon_\alpha) &= \left[ \frac{2\eta}{(E - \varepsilon_\alpha)^2 + \eta^2} \right]_{\eta \rightarrow 0^+} = \\ &= i \left[ \frac{1}{E - \varepsilon_\alpha + i0^+} - \frac{1}{E - \varepsilon_\alpha - i0^+} \right], \end{aligned} \quad (8.2.10a)$$

при помощи которого нетрудно увидеть, что

$$2\pi\delta(EI - H) = i \{ [(E + i0^+)I - H]^{-1} - [(E - i0^+)I - H]^{-1} \}, \quad (8.2.10b)$$

где  $0^+$  обозначает положительную бесконечно малую величину, чей физический смысл будет подробно обсуждаться в параграфе 8.4. Выражение (8.2.10b) можно было бы рассматривать как простое обобщение (8.2.10a), если бы аргумент  $(EI - H)$  был простым числом. Но поскольку  $(EI - H)$  является матрицей, выражение (8.2.10b) может показаться слишком смелым обобщением формулы (8.2.10a). Однако его можно обосновать, перейдя в представление, диагонализующее матрицу  $[H]$ , в котором обе части выражения (8.2.10b) будут диагональными матрицами, а равенство соответствующих диагональных элементов обеспечивается (8.2.10a). Следовательно, справедливость матричного равенства (8.2.10b) можно установить в представлении собственных функций, что обеспечит его выполнение и в любом другом представлении.

Используя выражения (8.2.7) и (8.2.10b), запишем

$$A(E) = i[G(E) - G^+(E)], \quad (8.2.11)$$

где запаздывающая функция Грина определена как

$$G(E) = [(E + i0^+)I - H]^{-1}, \quad (8.2.12a)$$

а опережающая функция Грина есть

$$G^+(E) = [(E - i0^+)I - H]^{-1}. \quad (8.2.12b)$$

В следующем параграфе мы увидим, каким образом функции Грина (и, стало быть, спектральные функции) рассчитываются в квантовых открытых системах.

**Матрица плотности.** Воспользовавшись определением  $[\rho] = f_0([H] - \mu[I])$  (уравнение (8.2.6)), можно записать

$$\begin{aligned} [\rho] &= \int_{-\infty}^{+\infty} dE f_0(E - \mu) \delta([EI - H]) = \\ &= \int_{-\infty}^{+\infty} \frac{dE}{2\pi} f_0(E - \mu) [A(E)], \end{aligned} \quad (8.2.13)$$

которое обретет ясный смысл, если вспомнить, что  $[A(E)]/2\pi$  является матричной версией плотности состояний  $D(E)$ , аналогично тому, как матрица плотности является матричной версией полного числа электронов  $N$ . Мы можем рассматривать выражение (8.2.13) как матричную версию известного соотношения

$$N = \int_{-\infty}^{+\infty} dE f_0(E - \mu) D(E),$$

которое означает, что полное число электронов находится умножением числа состояний  $D(E)dE$  на вероятность заполнения  $f_0(E)$  их и последующего суммирования по всем энергиям.

### Почему следует использовать выражение (8.2.13), а не (8.2.6)?

В предыдущих главах матрица плотности определялась из выражения (8.2.6). Почему теперь следует пользоваться формулой (8.2.13), включающей дополнительное интегрирование по энергии? Действительно, если рассматривается вся система с матрицей  $[H]$ , то для этого нет причин. Но в том случае, когда мы имеем дело с открытой системой, описываемой матрицей вида (см. рис. 8.2)

$$\bar{H} = \begin{pmatrix} H & \tau \\ \tau^+ & H_R \end{pmatrix},$$

формула (8.2.6) потребует от нас работы со всей матрицей, размер которой всегда велик по сравнению с  $[H]$ , поскольку матрица резервуара  $[H_R]$  всегда

велика. Фигурирующие в (8.2.13) спектральная функция и функция Грина также имеют большой размер

$$\bar{A} = \begin{pmatrix} A & A_{dR} \\ A_{Rd} & A_{RR} \end{pmatrix} \quad \bar{G} = \begin{pmatrix} G & G_{dR} \\ G_{Rd} & G_{RR} \end{pmatrix},$$

но нас интересует лишь верхний блок этих матриц размером  $(d \times d)$ . Существенным преимуществом подхода, связанного с использованием функций Грина, является то, что этот блок матрицы  $[G(E)]$  и, следовательно, матрицы  $[A(E)]$  может быть вычислен без использования элементов всей матрицы. Мы продемонстрируем это ниже, при обсуждении матричной собственно-энергетической функции  $\Sigma$ , с которой мы уже имели дело в предыдущем параграфе.

**Еще раз о матричной собственно-энергетической функции.** Полная функция Грина может быть получена из выражения (8.2.12a) в виде

$$\bar{G} \equiv \begin{pmatrix} G & G_{dR} \\ G_{Rd} & G_{RR} \end{pmatrix} = \begin{pmatrix} (E + i0^+)I - H & -\tau \\ -\tau^+ & (E + i0^+)I_R - H_R \end{pmatrix}^{-1}. \quad (8.2.14)$$

Эффективность метода функций Грина обусловлена тем, что необходимый блок матрицы  $[G]$  размера  $(d \times d)$  можно вычислить точно, исходя из соотношения

$$G = [(E + i0^+)I - H - \Sigma(E)]^{-1}, \quad (8.2.15)$$

где  $\Sigma(E)$  есть матричная собственно-энергетическая функция, определяемая выражением (8.1.11).

Уравнение (8.2.15) получается из уравнения (8.2.14) после выполнения стандартных матричных преобразований. Основной результат, имеющий практическую ценность, состоит в следующем. Если ввести матрицу

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1}, \quad (8.2.16a)$$

то произведение

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix},$$

поэтому  $Aa + Bc = I$  и  $Ca + Dc = 0$ , откуда следует, что  $c = -D^{-1}Ca$ . Следовательно, справедливо равенство

$$a = (A - BD^{-1}C)^{-1}. \quad (8.2.16b)$$

Сравнивая выражение (8.2.16a) с формулой (8.2.14) и выполняя очевидные подстановки, найдем из уравнения (8.2.16b)

$$G = [(E + i0^+)I - H - \tau G_R \tau^+]^{-1}, \quad \text{где} \quad G_R = [(E + i0^+)I_R - H_R]^{-1},$$

что и приводит к результату, содержащемуся в выражении (8.2.15).

Выражение (8.2.16b), в котором исходная матрица *разбивается* на блоки меньшего размера, описывает часто используемый прием для нахождения обратных матриц большого размера. В подобных задачах необходимо находить все компоненты  $a$ ,  $b$ ,  $c$  и  $d$  обратной матрицы, которые имеют приблизительно одинаковый размер. В нашей задаче, однако, матрицы  $a$  и  $A$  гораздо меньше по размеру, чем  $d$  и  $D$ , а нам необходимо знать лишь матрицу  $a$ . Выражение (8.2.15) позволяет нам вычислить матрицу  $[G]$  путем обращения матрицы небольшого размера ( $d \times d$ ), а не полной матрицы размера  $((d + R) \times (d + R))$ , фигурирующей в (8.2.14). Это обеспечивает существенные практические преимущества, поскольку порядок матрицы  $R$  обычно намного превосходит порядок матрицы  $d$ . Идея описывать влияние внешнего окружения на структуру с помощью собственно-энергетической функции  $[\Sigma]$  не просто дает удобный вычислительный инструмент. Речь идет о важной физической величине, приложения которой обсуждаются в следующем параграфе. Теперь мы рассмотрим два примера, один из которых будет аналитическим, а другой численным.

**Аналитический расчет.** Рассмотрим однородную бесконечную одномерную проволоку, квантовые состояния в которой определяются однозонным гамильтонианом, записанным в приближении эффективной массы. Структура гамильтониана схематически показана на рис. 8.2.2. Поскольку проволока является однородной, ее состояния характеризуются квантовым числом  $k$  и соответствующим дисперсионным соотношением, что позволяет записать плотность состояний на элементарную ячейку в виде

$$D(E) = a/\pi\hbar\nu(E), \quad \text{где} \quad \nu = (1/\hbar)dE/dk.$$

Получим теперь этот же результат, используя метод функций Грина, развитый в этом параграфе. Для этого заменим бесконечную одномерную проволоку *одиночной* элементарной ячейкой и добавим слагаемые с собственной энергетической функцией. Это позволит учесть две полубесконечные проволоки на каждой стороне ячейки (рис. 8.2.3).

Функция Грина для этой одиночной ячейки является матрицей размера  $(1 \times 1)$ , т. е. числом:

$$G(E) = \frac{1}{E - (E_c + 2t_0 - t_0 \exp[ika] - t_0 \exp[-ika])}.$$

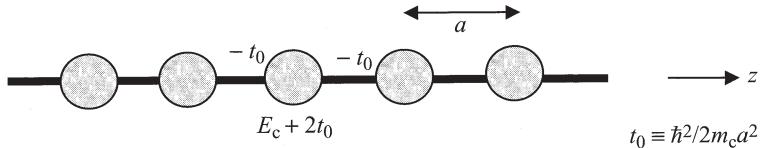


Рис. 8.2.2

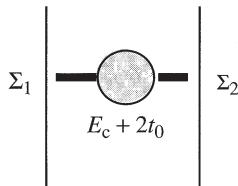


Рис. 8.2.3

Учитывая дисперсионное соотношение  $E = E_c + 2t_0(1 - \cos ka)$ , можно упростить данное выражение и получить

$$G(E) = \frac{1}{t_0 \exp[ika] - t_0 \exp[-ika]} = \frac{1}{2it_0 \sin ka},$$

откуда и находится плотность состояний:

$$D(E) = i[G - G^+]/2\pi = 1/2\pi t_0 \sin ka = a/\pi\nu.$$

Видно, что полученное выражение совпадает с найденным выше, поскольку  $\hbar\nu = dE/dk = 2at_0 \sin ka$ .

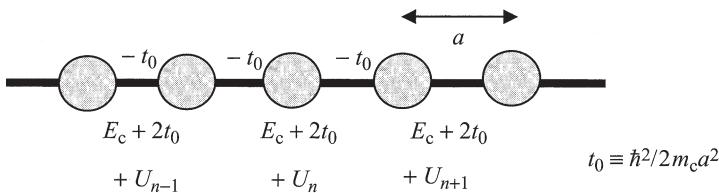


Рис. 8.2.4

**Численный пример.** Чтобы почувствовать преимущества метода собственно-энергетической функции, целесообразно еще раз, заново, решить задачу об определении равновесной электронной плотности в одномерной яме. Эта задача обсуждалась в главе 4 (см. рис. 4.3.1). Сейчас мы

рассмотрим одномерную яму с дополнительным потенциалом  $U(x)$ , который изменяется линейно от значения  $-0.5$  эВ на одном краю до  $+0.5$  эВ на другом краю. Будем использовать однозонный гамильтониан, заданный на сетке с 50 узлами, расположенными на расстоянии  $a = 2 \text{ \AA}$ , и эффективной массой  $m_c$ , равной 0.25 массы свободного электрона  $m$  (см. рис. 8.2.4).

Произведем расчет электронной плотности  $n(z)$  в яме, предполагая, что она находится в состоянии равновесия и характеризуется электрохимическим потенциалом  $\mu = E_c + 0.25$  эВ, а тепловая энергия  $k_B T = 0.025$  эВ. Электронная плотность определяется диагональными элементами матрицы плотности  $\rho$ , которые можно вычислить одним из двух способов.

(1) Можно задать периодические граничные условия:  $H(1, 100) = H(100, 1) = -t_0$ , а затем вычислить  $\rho$ , исходя из выражения (8.2.6).

(2) Можно также учесть слагаемые, связанные с собственно-энергетической функцией (см. уравнение (8.1.7a)) и имеющие ненулевое значение, равное  $-t_0 \exp[ika]$ , лишь в крайних точках  $(1, 1)$  и  $(100, 100)$  и затем, воспользовавшись выражением (8.2.15), рассчитать функцию Грина,

$$G = [(E + i0^+)I - H - \Sigma_1 - \Sigma_2]^{-1}.$$

Далее исходя из выражения (8.2.11), следует вычислить спектральную функцию, а затем, воспользовавшись уравнением (8.2.13), рассчитать равновесную матрицу плотности.

На рис. 8.2.5 показано, что оба метода дают хорошо согласующиеся между собой результаты. Некоторое расхождение, скорее всего, обусловлено ошибками дискретизации шкалы энергии, используемой при интегрировании на заключительном шаге в последнем методе. Мы использовали сетку с 250 узлами в интервале энергий  $E_c - 0.1$  эВ  $< E < E_c + 0.4$  эВ. Следует отметить, что осцилляции, полученные при использовании первого метода, связаны со стоячими волнами, образующимися в замкнутом кольце, и отсутствуют в открытой системе, моделируемой методом собственно-энергетической функции. Такие осцилляции будут менее выражены в более протяженных структурах (или при больших эффективных массах), поскольку уровни энергии в этом случае будут располагаться ближе друг к другу.

Как мы уже знаем, собственно-энергетическая функция определяет уширение энергетических уровней, но ее полный вклад в плотность состояний остается неизменным в силу правила сумм, упомянутого ранее. Как следствие, различия между двумя методами будут менее заметными, когда мы будем рассматривать электронную плотность, поскольку она включает интегрирование по энергии. Метод собственно-энергетической функции

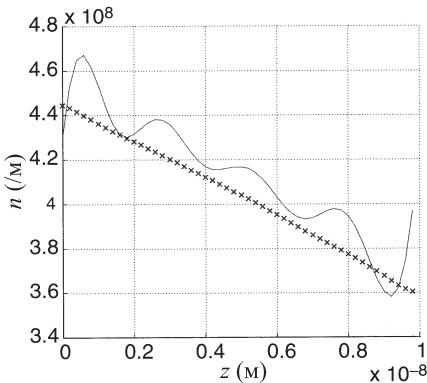


Рис. 8.2.5. График электронной плотности  $n(z)$ , рассчитанной для одномерной проволоки с линейным потенциалом  $U(z)$  при использовании периодических граничных условий (сплошная линия) и при применении метода собственной энергетической функции, отвечающего открытым граничным условиям (крестики)

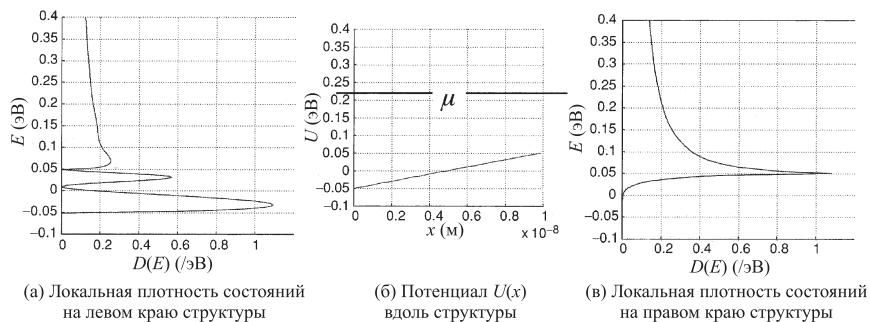


Рис. 8.2.6. Локальная плотность состояний (LDOS) на двух границах одномерной проволоки в присутствии линейного потенциала  $U(x)$ . Соответствующая электронная плотность показана на рис. 8.2.5. Значение электрохимического потенциала  $\mu = E_c + 0.25$  эВ, тепловая энергия  $k_B T = 0.0025$  эВ

позволяет подробно анализировать локальную плотность состояний на различных участках структуры (см. рис. 8.2.6).

### 8.3. Время жизни

Введенное в параграфе 8.2 понятие функции Грина  $G(E)$  открывает удобный путь расчета спектральной функции  $A(E)$ . Он основан на использовании математического тождества

$$2\pi\delta(EI - H) = i\{[(E + i0^+)I - H]^{-1} - [(E - i0^+)I - H]^{-1}\},$$

$$A(E) = i[G(E) - G^+(E)].$$

Однако функция Грина имеет и собственный физический смысл: ее можно рассматривать как решение уравнения Шредингера, описывающее отклик на *импульсное* воздействие. Такой подход поможет также понять смысл соотношения «неопределенностей», связывающего уширение уровня с конечным временем жизни. Заметим, что эти величины зависят от взаимодействия с резервуаром. Чтобы лучше понять смысл функции Грина, используем в качестве базиса собственные состояния  $H$ , так что матрица функции Грина будет диагональной:

$$[G(E)] = \begin{pmatrix} \frac{1}{E - \varepsilon_1 + i0^+} & 0 & 0 & \dots \\ 0 & \frac{1}{E - \varepsilon_2 + i0^+} & 0 & \dots \\ 0 & 0 & \frac{1}{E - \varepsilon_3 + i0^+} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (8.3.1)$$

Рассмотрим Фурье-образ функции  $G(E)$ , определяемый как

$$[\tilde{G}^R(t)] = \int_{-\infty}^{+\infty} \frac{dE}{2\pi\hbar} e^{-iEt/\hbar} [G(E)],$$

который также является диагональной матрицей вида

$$[\tilde{G}^R(t)] = \frac{-i}{\hbar} \theta(t) e^{-0^+ t} \begin{pmatrix} \exp(-i\varepsilon_1 t/\hbar) & 0 & 0 & \dots \\ 0 & \exp(-i\varepsilon_2 t/\hbar) & 0 & \dots \\ 0 & 0 & \exp(-i\varepsilon_3 t/\hbar) & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (8.3.2)$$

После выполнения определенных вычислений (включающих интегрирование по контуру на комплексной плоскости) можно перейти от матрицы (8.3.1) к ее Фурье-образу (8.3.2). Более простым является обратный переход от (8.3.2) к (8.3.1), который выполняется с использованием формулы

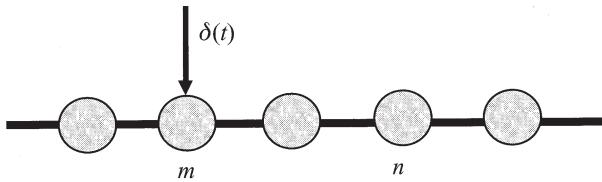


Рис. 8.3.1

обратного преобразования Фурье:

$$\begin{aligned}
 [G(E)] &= \int_{-\infty}^{+\infty} dt e^{+iEt/\hbar} [\tilde{G}^R(t)] = \\
 &= \frac{-i}{\hbar} \int_{-\infty}^{+\infty} dt e^{iEt/\hbar} \theta(t) e^{-i\varepsilon t/\hbar} e^{-0^+ t} = \\
 &= \frac{-i}{\hbar} \int_{-\infty}^{+\infty} dt e^{i(E-\varepsilon)t/\hbar} e^{-0^+ t} = \frac{1}{E - \varepsilon + i0^+}.
 \end{aligned}$$

Необходимо заметить, что введенный в этих выражениях индекс «R» не связан с резервуаром. Этим индексом обозначена «запаздывающая» (retarded) функция Грина  $\tilde{G}^R(t)$ , которая равна нулю при всех значениях времени  $t < 0$ . Нетрудно увидеть, что диагональные элементы этой функции удовлетворяют дифференциальному уравнению

$$\left( i\hbar \frac{\partial}{\partial t} - \varepsilon_\alpha \right) \tilde{G}_{\alpha\alpha}^R(t) = \delta(t).$$

Поэтому можно записать

$$\left( i\hbar \frac{\partial}{\partial t} - [H] \right) [\tilde{G}^R(t)] = [I]\delta(t), \quad (8.3.3)$$

что позволяет интерпретировать функцию  $\tilde{G}^R(t)$  как отклик уравнения Шредингера на дельта-образный импульс

$$\left( i\hbar \frac{\partial}{\partial t} - [H] \right) \{ \Psi(t) \} = 0. \quad (8.3.4)$$

Элемент  $(n, m)$  этой матрицы  $\tilde{G}_{nm}^R(t)$  описывает отклик  $n$ -й компоненты волновой функции системы, если ее  $m$ -я компонента получила импульсное возбуждение (см. рис. 8.3.1).

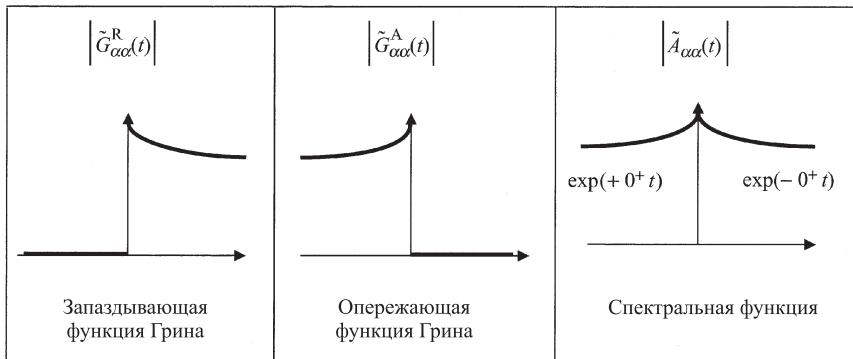


Рис. 8.3.2. Схематические изображения величины любого диагонального элемента (в представлении собственных состояний) запаздывающей и опережающей функций Грина, а также спектральной функции, рассматриваемых как функции времени

С этой точки зрения представляется естественным назвать функцию Грина «запаздывающей», поскольку отклик не может появиться раньше внешнего импульса, подействовавшего на систему в момент времени  $t = 0$ . С математической точки зрения такое решение уравнения (8.3.3), однако, не является единственным. Можно непосредственно показать, что имеется еще одно решение — «опережающая» функция Грина

$$[\tilde{G}^A(t)] = [\tilde{G}^R(-t)]^*, \quad (8.3.5)$$

удовлетворяющая тому же уравнению

$$\left( i\hbar \frac{\partial}{\partial t} - [H] \right) [\tilde{G}^A(t)] = [I]\delta(t), \quad (8.3.6)$$

но обращающаяся в нуль при всех положительных значениях времени, т. е. *после* момента  $t = 0$ . Воспользовавшись представлением собственных функций и используя выражения (8.3.2) и (8.3.5), можно записать, что

$$[\tilde{G}^A(t)] = \frac{-i}{\hbar} \theta(t) e^{+0^+ t} \begin{pmatrix} \exp(-i\varepsilon_1 t/\hbar) & 0 & 0 & \dots \\ 0 & \exp(-i\varepsilon_2 t/\hbar) & 0 & \dots \\ 0 & 0 & \exp(-i\varepsilon_3 t/\hbar) & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}. \quad (8.3.7)$$

Это выражение фактически представляет собой Фурье-преобразование функции  $G^+(E)$  (опять, обратное преобразование выполнить легче). Различие между запаздывающей и опережающей функциями Грина связано с их зависимостью от энергии

$$G(E) = [(E + i0^+)I - H]^{-1} \quad \text{Запаздывающая,}$$

$$G^+(E) = [(E - i0^+)I - H]^{-1} \quad \text{Опережающая.}$$

Две функции различаются лишь знаком бесконечно малой величины  $0^+$ . Это дает повод сделать неверный вывод о том, что эти функции мало различаются между собой. При рассмотрении их зависимости от времени, однако, разница между ними становится очевидной. Одна из функций равна нулю при  $t < 0$  (удовлетворяет принципу причинности), а другая равна нулю при  $t > 0$  (не удовлетворяет принципу причинности). Первая из этих функций интерпретируется как отклик на импульсное возбуждение при  $t = 0$ ; в то время как другая же не имеет простой физической интерпретации. Она является математически правильным решением того же уравнения, но с другим, не имеющим физического смысла, начальным условием. На рис. 8.3.2 показана величина диагональных элементов функций  $\tilde{G}_{\alpha\alpha}^R(t)$  и  $\tilde{G}_{\alpha\alpha}^A(t)$ . Обратите внимание на то, что спектральная функция пропорциональна разности между запаздывающей и опережающей функциями Грина (см. уравнение (8.2.11)):

$$\tilde{A}_{\alpha\alpha}^R(t) = i[\tilde{G}_{\alpha\alpha}^R(t) - \tilde{G}_{\alpha\alpha}^A(t)].$$

Поскольку обе функции Грина удовлетворяют одному дифференциальному уравнению, спектральная функция как разность двух решений будет удовлетворять однородному дифференциальному уравнению без источника импульсного возбуждения:

$$\left( i\hbar \frac{\partial}{\partial t} - [H] \right) [\tilde{A}(t)] = [0],$$

и поэтому, как показано на рис. 8.3.2, не имеет разрыва при  $t = 0$  в отличии от функций  $\tilde{G}_{\alpha\alpha}^R(t)$  и  $\tilde{G}_{\alpha\alpha}^A(t)$ .

**Физический смысл собственно-энергетической функции.** В параграфе 8.2 мы увидели, что можно *точно* вычислить матричный блок полной матричной функции Грина, отвечающий нашей структуре (уравнение (8.2.14)):

$$\bar{G} \equiv \begin{pmatrix} G & G_{dR} \\ G_{Rd} & G_{RR} \end{pmatrix} = \begin{pmatrix} (E + i0^+)I - H & -\tau \\ -\tau^+ & (E + i0^+)I - H_R \end{pmatrix}^{-1},$$

исходя из соотношения (8.2.15)

$$G = [(E + i0^+)I - H - \Sigma(E)]^{-1},$$

где собственно-энергетическая функция

$$\Sigma(E) = \tau g_R(E)\tau^+$$

может быть рассчитана, если известны поверхностные свойства резервуара ( $g_R$ ), а также параметры взаимодействия резервуара со структурой ( $\tau$ ).

Теперь, когда дана интерпретация зависящей от времени функции Грина как импульсному отклику уравнения Шредингера (см. уравнение (8.3.4)), можно, используя преобразование Фурье (8.2.15), записать уравнение для функции Грина — для подматрицы, отвечающей области структуры. Это преобразование легко выполняется, если собственно-энергетическая функция  $\Sigma$  не зависит от энергии  $E$ :

$$\left( i\hbar \frac{\partial}{\partial t} - [H] - [\Sigma] \right) [\tilde{G}^R(t)] = [I]\delta(t). \quad (8.3.8a)$$

Если теперь учесть зависимость от энергии, то преобразование Фурье будет более сложным: произведение функций  $\Sigma(E)$  и  $G(E)$  после Фурье-преобразования запишется как свертка этих функций:

$$\left( i\hbar \frac{\partial}{\partial t} - [H] \right) [\tilde{G}^R(t)] - \int dt' [\tilde{\Sigma}(t - t')] [\tilde{G}^R(t')] = [I]\delta(t). \quad (8.3.8b)$$

Чтобы понять физический смысл величины  $\Sigma$ , не будем пока принимать во внимание эту «деталь». Фактически можно значительно упростить задачу, рассматривая небольшую структуру, с лишь одним энергетическим уровнем  $\varepsilon$  (рис. 8.3.3). При этом величины  $[H]$  и  $[\Sigma]$  являются скалярами, а не матрицами, и уравнение имеет вид

$$\left( i\hbar \frac{\partial}{\partial t} - \varepsilon - \Sigma \right) \tilde{G}^R(t) = \delta(t).$$

Решением этого уравнения является функция

$$\tilde{G}^R(t) = -\frac{i}{\hbar} e^{-i(\varepsilon + \Sigma)t/\hbar} \theta(t),$$

которая описывает отклик волновой функции на импульсное возбуждение структуры в момент времени  $t = 0$ . Можно написать

$$\tilde{G}^R(t) = -\frac{i}{\hbar} e^{-i\varepsilon't/\hbar} e^{-\gamma t/2\hbar} \theta(t), \quad (8.3.9)$$

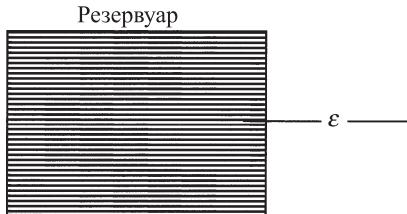


Рис. 8.3.3

где введено обозначение

$$\varepsilon' = \varepsilon + \operatorname{Re}\Sigma \quad \text{и} \quad \gamma = -2\operatorname{Im}\Sigma. \quad (8.3.10)$$

Вещественная часть собственно-энергетической функции приводит к сдвигу уровня энергии от  $\varepsilon$  к  $\varepsilon'$ , в то время как мнимая часть ответственна за конечное время жизни собственного состояния. Ясно, что зависимость квадрата модуля волновой функции от времени определяет скорость затухания после действия начального возбуждения:

$$|\tilde{G}^R(t)|^2 = \frac{1}{\hbar^2} \theta(t) \exp(-\gamma t/\hbar).$$

Очевидно, что время жизни состояния можно связать с мнимой частью собственно-энергетической функции:

$$\frac{1}{\tau} = -\frac{\gamma}{\hbar} = -\frac{2\operatorname{Im}\Sigma}{\hbar}. \quad (8.3.11)$$

Мы можем связать это равенство с соотношением «неопределенностей» между временем жизни и уширением, если вспомним, что мнимая часть собственно-энергетической функции равна уширению плотности состояний. Чтобы убедиться в этом, заметим, что Фурье-преобразование скалярной формы функции Грина в выражении (8.3.9) определяется выражением

$$G(E) = \frac{1}{E - \varepsilon' + i\gamma/2},$$

поэтому

$$\begin{aligned} \frac{A(E)}{2\pi} \equiv D(E) &= i \left( \frac{1}{E - \varepsilon' + i\gamma/2} - \frac{1}{E - \varepsilon' - i\gamma/2} \right) = \\ &= \frac{\gamma}{(E - \varepsilon')^2 + (\gamma/2)^2}, \end{aligned}$$

откуда видно, что локальная плотность состояний в структуре испытывает уширение в форме лоренцевской кривой ширины  $\gamma$ , равное удвоенной мнимой части собственно-энергетической функции. Разумеется, в общем случае форма этой линии не описывается функцией Лоренца. Мы получили результат в такой форме, поскольку пренебрегли зависимостью собственной энергетической функции от энергии и воспользовались уравнением (8.3.8a) вместо (8.3.8b). Из анализа временной зависимости следует, что этот эффект можно назвать эффектом памяти резервуара.

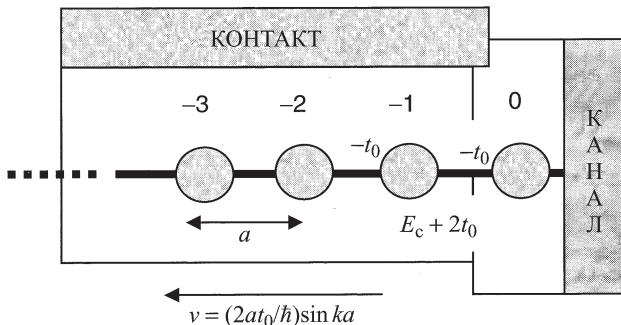


Рис. 8.3.4

В параграфе 8.2 мы видели, что собственно-энергетическая функция одномерного контакта имеет диагональный вид с двумя ненулевыми слагаемыми:

$$\Sigma(1, 1) = -t_0 \exp(ika) = -t_0 \cos(ka) - it_0 \sin(ka).$$

Используя уравнение (8.3.11), мы можем записать соответствующее время жизни для узла 1 как

$$\frac{1}{\tau} = \frac{\gamma}{\hbar} = \frac{2t_0 \sin ka}{\hbar}.$$

Интересно заметить, что связанная с конкретным волновым числом  $k$  скорость в проволоке определяется выражением

$$\nu = \frac{1}{\hbar} \frac{\partial E}{\partial k} [2t_0(1 - \cos ka)] = \frac{2at_0}{\hbar} \sin ka,$$

поэтому выражение для обратного времени жизни принимает вид

$$\frac{1}{\tau} = \frac{\gamma}{\hbar} = \frac{\nu}{a},$$

которое интуитивно выглядит удовлетворительным, поскольку скорость ухода из данной ячейки или декремент затухания должен быть равен скорости возбуждения, деленной на размер ячейки (см. рис. 8.3.4). Этим правилом можно руководствоваться и для приближенной записи мнимой части собственно-энергетической функции в структурах с более сложной геометрией, где точный расчет поверхностной функции Грина затруднен:

$$\text{Im}\Sigma(E) \approx \hbar\nu(E)/R. \quad (8.3.12)$$

В этом выражении  $R$  определяет линейный размер элементарной ячейки, а точное значение множителя перед указанным выражением зависит от геометрии конкретной структуры.

Зная мнимую часть, можно найти и вещественную составляющую, исходя из общего правила, не зависящего от конкретных деталей. Нужно воспользоваться тем, что вещественная и мнимая части связаны друг с другом преобразованием Гильберта (символ  $\otimes$  обозначает свертку):

$$\text{Re}\Sigma(E) = -\frac{1}{\pi} \int dE' \frac{\text{Im}\Sigma(E')}{E - E'} = \text{Im}\Sigma(E) \otimes \left( -\frac{1}{\pi E} \right). \quad (8.3.13)$$

Поэтому собственно-энергетическая функция может быть записана в виде

$$\Sigma(E) = [\text{Re}\Sigma(E)] + \text{Im}\Sigma(E) = i\text{Im}\Sigma(E) \otimes \left( \delta(E) + \frac{i}{\pi E} \right). \quad (8.3.14)$$

Сформулированный принцип должен соблюдаться для любой функции, чье преобразование Фурье согласуется с требованием причинности (т. е. Фурье-образ равен нулю при  $t < 0$ ). Собственно-энергетическая функция удовлетворяет требованию причинности, поскольку она пропорциональна поверхностной функции Грина резервуара (см. уравнение (8.1.11)), которая, как мы ранее отметили, обладает этим свойством. Чтобы объяснить, почему причинные функции подчиняются этому принципу, заметим, что выражение  $\delta(E) + (i/\pi E)$  является Фурье-преобразованием единичной ступенчатой функции  $\theta(t)$ . Если принять во внимание, что после преобразования Фурье произведение двух зависящих от времени функций переходит в свертку этих функций, то мы придем к заключению, что любая зависящая от времени функция вида  $\theta(t)f(t)$  обладает Фурье-образом следующего вида:

$$F(E) \otimes \left( \delta(E) + \frac{i}{\pi E} \right),$$

где  $F(E)$  обозначает Фурье-образ функции  $f(t)$ , которому соответствует функция  $i \operatorname{Im} \Sigma(E)$  в выражении (8.3.14).

**Матрица, описывающая уширение.** Рассматривая простой пример структуры с одним уровнем энергии, мы установили, что мнимая часть собственно-энергетической функции определяет величину уширения и обратное время жизни на уровне (см. уравнение (8.3.11)). В более общей ситуации собственно-энергетическая функция становится матрицей, и соответственно следует говорить о матрице уширения, равной ее антиэрмитовой части:

$$\Gamma(E) = i[\Sigma(E) - \Sigma^+(E)]. \quad (8.3.15)$$

Эта часть собственно-энергетической функции ответственна за уширение уровня, в то время как эрмитова составляющая

$$\Sigma_H(E) = \frac{1}{2}[\Sigma(E) + \Sigma^+(E)]$$

определяет поправку к гамильтониану  $[H]$ . В итоге мы можем записать

$$H + \Sigma(E) = [H + \Sigma_H(E)] - \frac{i\Gamma(E)}{2}.$$

Мы часто пользовались тем обстоятельством, что решение задачи упрощается, когда в качестве базисных функций используются собственные функции гамильтониана  $[H]$ . Для открытых систем желательно было бы использовать такое представление, в котором в интересующей нас области энергий диагонализуется и матрица  $[H + \Sigma_H]$ . Если в том же представлении диагональна и матрица  $[\Gamma]$ , то ответ формулируется просто на языке многих одноуровневых структур, соединенных параллельно. В общем случае, однако, такая ситуация не имеет места, поскольку в представлении, диагонализующем матрицу  $[H + \Sigma_H]$ , матрица  $[\Gamma]$  может быть не диагональной и наоборот. Можно, следовательно, диагонализовать одну или другую матрицу, но не обе сразу. Это и приводит к новым интересным эффектам, не описываемым в рамках одноуровневой задачи.

## 8.4. Что образует контакт (резервуар)?

Перед обсуждением новой важной концепции, я хотел бы кратко просуммировать то, что уже сделано в этой главе. В параграфе 8.1 мы исходили

из уравнения Шредингера для системы «канал + контакт»:

$$\begin{pmatrix} EI_R - H_R + i\eta & -\tau^+ \\ -\tau & EI - H \end{pmatrix} \begin{pmatrix} \Phi_R + \chi \\ \psi \end{pmatrix} = \begin{pmatrix} S_R \\ 0 \end{pmatrix} \quad (8.4.1)$$

и получили уравнение, описывающее только канал:

$$[EI - H - \Sigma]\{\psi\} = \{S\}. \quad (8.4.2)$$

Решение может быть записано в виде

$$\{\psi\} = [G]\{S\}, \quad \text{где } [G] = [EI - H - \Sigma]^{-1}. \quad (8.4.3)$$

Далее мы обсудили выражения для функции Грина  $G$  и собственно-энергетической функции  $\Sigma$ , зависящих как от энергии, так и от времени.

Если мы вернемся назад в параграф 8.1 и вспомним цепочку рассуждений, ведущую от уравнения (8.4.1) к уравнению (8.4.2), станет очевидным, что контакт также можно рассматривать как основную часть структуры, а влияние канала на него учитывать, вводя собственно-энергетическую функцию. С математической точки зрения матрицы  $[H]$  и  $[H_R]$  имеют одинаковый размер. Единственная «асимметрия» между ними связана с членами  $i\eta$  и  $S_R$ , которые мы добавили «руками», чтобы описать уход и повторную инжекцию электронов из контакта в присутствии внешнего источника. Введение величины  $i\eta$  обосновывается также требованием математической сходимости Фурье-образа. Но это бесконечно малое слагаемое могло быть добавлено из соображений «симметрии», что естественным образом поднимает вопрос о том, в чем же состоит различие между контактом и каналом. С ним связан и вопрос о том, насколько реально малой должна быть величина  $\eta$ : равен ли ее порядок, скажем, наноэлектронвольтам или милиэлектронвольтам? Имеет ли это существенное значение?

Чтобы дать ответ на эти вопросы, рассмотрим следующий простой пример. Пусть у нас есть «молекула», которая описывается гамильтоновской матрицей  $(2 \times 2)$

$$H = \begin{pmatrix} \varepsilon & \tau \\ \tau^* & \varepsilon_1 \end{pmatrix}. \quad (8.4.4)$$

Мы смогли рассчитать матрицу  $(2 \times 2)$  спектральной функции  $[A(E)]$ , исходя из нижеприведенного соотношения (где  $[\eta] = 0^+[I]$ , а  $I$  есть единичная матрица размера  $(2 \times 2)$ ):

$$[A] = i[G - G^+], \quad \text{где } [G] = [EI - H + i\eta]^{-1}, \quad (8.4.5)$$

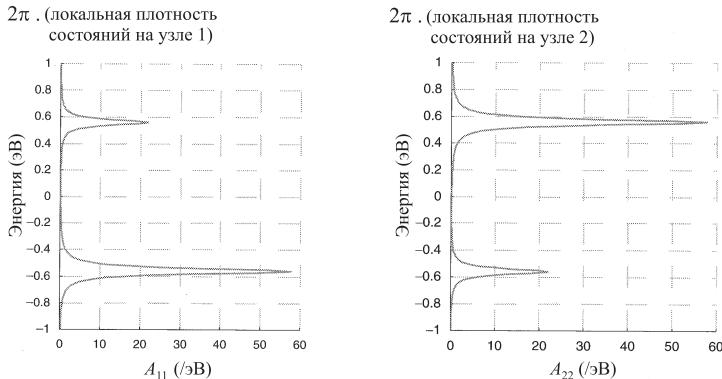


Рис. 8.4.1

и использовали ее диагональные элементы, чтобы рассчитать локальную плотность состояний (LDOS) для каждого из двух узлов. В качестве примера для системы с параметрами  $\varepsilon = -0.25$  эВ,  $\varepsilon_1 = +0.25$  эВ,  $\tau = 0.5$  эВ и  $0^+ = 0.025$  эВ мы построили графики, показанные на рис. 8.4.1. Обратите внимание на наличие двух пиков, расположенных вблизи каждого из двух собственных значений энергии, полученных из условия обращения в нуль определителя матрицы ( $EI - H$ ):

$$\det \begin{pmatrix} E - \varepsilon & \tau \\ \tau^* & E - \varepsilon_1 \end{pmatrix} = 0. \quad (8.4.6)$$

Два данных пика имеют разную высоту, и это отражает тот факт, что волновая функция, отвечающая нижнему уровню энергии, смешена к узлу «1», в то время как собственная функция верхнего уровня смешена к узлу «2».

Если нас интересует, в основном, поведение функции в области первого узла, то влияние узла 2 можно учесть с помощью собственно-энергетической функции

$$\Sigma(E) = \frac{|\tau|^2}{E - \varepsilon_1 + i0^+}, \quad (8.4.7)$$

и определить локальную плотность состояний с помощью спектральной функции с матрицей порядка  $(1 \times 1)$ :

$$[a] = i[g - g^+], \quad \text{где} \quad [g] = [E - \varepsilon - \Sigma]^{-1}. \quad (8.4.8)$$

Результат точно совпал бы с полученным ранее при использовании спектральной матричной функции размером  $(2 \times 2)$ . С другой стороны, если нас интересует поведение состояний вблизи узла 2, то влияние узла 1 можно учесть, рассматривая собственную энергетическую функцию

$$\Sigma_1(E) = \frac{|\tau|^2}{E - \varepsilon + i0^+}, \quad (8.4.9)$$

и вычислить локальную плотность состояний, вновь исходя из спектральной функции с матрицей порядка  $(1 \times 1)$ :

$$[a_1] = i[g_1 - g_1^+], \quad \text{где} \quad [g_1] = [E - \varepsilon_1 - \Sigma_1]^{-1}. \quad (8.4.10)$$

При этом результат вновь совпадет с тем, который получается при использовании спектральной функции с матрицей порядка  $(2 \times 2)$ . Говоря кратко, мы можем рассматривать точку 1 как канал, а точку 2 как контакт, либо наоборот. Любой выбор здесь является приемлемым с математической точки зрения.

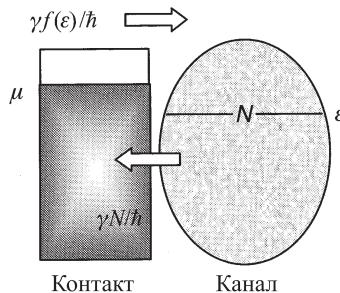


Рис. 8.4.2

Следует отметить, что ни один из этих контактов, имеющих атомный размер, не может называться «резервуаром» в широком смысле этого слова. В первой главе я привел выражение для тока, протекающего между контактом и каналом следующего вида (см. уравнение (1.2.2a) и рис. 8.4.2):

$$I = \text{Входящий поток} - \text{Выходящий поток} = (1/h)(\gamma f - \gamma N).$$

Важным свойством «резервуара» является то, что скорости  $\gamma$  и  $\gamma f$  выходящего и входящего потоков не зависят от степени заполнения состояний, а также от других динамических характеристик.

Заметим, что «контакты» атомных размеров, которые мы только что описали выше, не удовлетворяют этому критерию. К примеру, если мы рассмотрим точку 2 как контакт (уравнение (8.4.7)), то величина уширения (или частота релаксации) определится выражением

$$\gamma = i[\Sigma - \Sigma^+] = \frac{|\tau|^2 0^+}{(E - \varepsilon_1)^2 + (0^+)^2}. \quad (8.4.11)$$

Эта величина не только зависит от энергии (в форме узкого пика вблизи  $E = \varepsilon_1$ ), но также существенно зависит от конкретного значения бесконечно малой величины  $0^+$ . Это значит, что скорость релаксации ( $\gamma/h$ ) для электронов в «канале» зависит от скорости, с которой они поступают из «контакта». Таким образом, «правильно устроенный» резервуар должен обеспечивать постоянную скорость релаксации, не зависящую от величины  $0^+$ . Как мы сейчас покажем, это возможно в случае, когда резервуар содержит множество близко расположенных энергетических уровней.

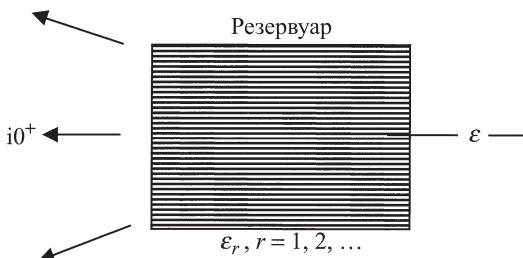


Рис. 8.4.3. Структура с единственным уровнем энергии  $\varepsilon = 0$ , взаимодействующая с резервуаром, содержащим большое число близко расположенных энергетических уровней  $\{\varepsilon_r, r = 1, 2, \dots\}$

Рассмотрим канал с единственным энергетическим уровнем  $\varepsilon = 0$  (рис. 8.4.3), взаимодействующий с контактом, спектр которого состоит из большого числа близко расположенных уровней, так что соответствующий гамильтониан задается большой матрицей вида

$$[H] = \begin{pmatrix} \varepsilon & \tau_1 & \tau_2 & \tau_3 & \dots \\ \tau_1^* & \varepsilon_1 & 0 & 0 & \dots \\ \tau_2^* & 0 & \varepsilon_2 & 0 & \dots \\ \tau_3^* & 0 & 0 & \varepsilon_3 & \dots \\ \dots & & & & \end{pmatrix}.$$

Влияние контакта можно учесть, определив собственно-энергетическую функцию, определяемую «матрицей» ( $1 \times 1$ ) (см. уравнение (8.1.15a))

$$\Sigma = \sum_r \frac{|\tau_r|^2}{(E - \varepsilon_r)^2 + (0^+)^2}, \quad (8.4.12)$$

так что уширение можно записать как

$$\gamma = i[\Sigma - \Sigma^+] = \sum_r \frac{|\tau_r|^2 0^+}{(E - \varepsilon_r)^2 + (0^+)^2}. \quad (8.4.13)$$

Если уровни расположены достаточно близко друг к другу, то суммирование заменяется интегрированием:

$$\gamma = \int d\varepsilon_r D_R(\varepsilon_r) |\tau_r|^2 \frac{0^+}{(E - \varepsilon_r)^2 + (0^+)^2}, \quad (8.4.14)$$

где  $D_R$  есть число состояний на единичный интервал энергий. Последняя функция подынтегрального выражения здесь имеет острый пик вблизи  $\varepsilon_r = E$ . Если другие функции подынтегрального выражения в этой области энергий изменяются медленно, то их можно вынести за знак интеграла, что приведет нас к выражению

$$\begin{aligned} \gamma &= D_R(E) |\tau(E)|^2 \int d\varepsilon_r \frac{0^+}{(E - \varepsilon_r)^2 + (0^+)^2} = \\ &= 2\pi D_R(E) |\tau(E)|^2. \end{aligned} \quad (8.4.15)$$

Выражение (8.4.15), часто называемое золотым правилом Ферми, широко используется при решении самых различных задач. Действительно, мы могли бы записать наш более ранний результат для контакта атомных размеров (выражение (8.4.11)) в той же форме, приняв во внимание, что плотность состояний в «резервуаре» атомных размеров есть

$$D_R(E) \rightarrow \frac{0^+/2\pi}{(E - \varepsilon_1)^2 + (0^+)^2}.$$

Ранее отмечалось, что такая модель описывает резервуар с весьма посредственной точностью, поскольку плотность состояний сильно зависит от энергии и параметра  $0^+$ . В то же время если интересующий нас уровень

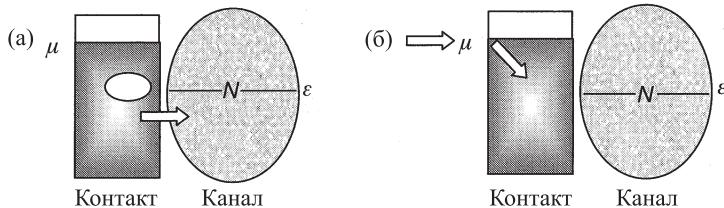


Рис. 8.4.4

взаимодействует приблизительно одинаково со всеми другими уровнями, резервуар со многими близко расположенными энергетическими уровнями должен иметь плавно изменяющуюся плотность состояний, не зависящую от  $0^+$ . При этом уровень каждого состояния уширяется на величину порядка  $0^+$ , значительно превышающую расстояние между уровнями. Обычные контакты в реальных образцах являются массивными макроскопическими объектами с расстоянием между уровнями порядка пикоэлектронвольта или менее, поэтому даже слабое внешнее возмущение может обеспечить требуемую степень уширения. При проведении аналитических расчетов собственно-энергетической функции  $\Sigma(E)$  обычно предполагают, что расстояние между уровнями и уширение уровня  $0^+$  одновременно стремятся к нулю, причем первая из этих величин ведет себя так вследствие бесконечно большого размера системы. При этом важно, чтобы уширение всегда превосходило расстояние между уровнями. В численных расчетах, чтобы сэкономить расчетное время, обычно рассматривают контакты конечных размеров. Поскольку в контактах конечных размеров расстояние между уровнями  $\Delta\varepsilon$  также конечно, важно правильно выбрать «бесконечно малую» добавку  $0^+$  так, чтобы она превосходила  $\Delta\varepsilon$ , и, соответственно, контакты вели себя как настоящие резервуары.

Хороший резервуар должен удовлетворять еще одному важному условию: не только скорость ухода из него  $\gamma$ , но скорость инжекции в нем  $\gamma f$  должны быть постоянными (здесь  $f$  есть функция Ферми на контакте). Для этого необходимо, чтобы контакт постоянно находился в состоянии статистического равновесия, что возможно, если состояния контакта взаимодействуют друг с другом. К примеру, если электрон входит в канал из контакта, то он оставляет в контакте дырку, с энергией меньше, чем электрохимическим потенциалом  $\nu$ , как это показано на рис. 8.4.4 а. Это приводит к возникновению сильно неравновесной ситуации: входящий электрон на уровне  $\nu$  быстро отдает энергию в кристаллической решетке и заполняет дырочное состояние путем экранирования или благодаря «процессам неупругого рас-

сения», описанным в главе 10. Вот почему большая часть джоулева тепла ( $I^2R$ ) выделяется в контакте, а не в канале.

Чтобы контакт вел себя как хороший резервуар, важно, чтобы равновесие восстанавливалось достаточно быстро, и можно было бы считать, что контакт все время находится в состоянии локального равновесия. То, как выполняется это требование, мы будем проверять в дальнейшем в процессе работы с «наноконтактами». Особенно в том случае, если речь идет о спиновой релаксации.

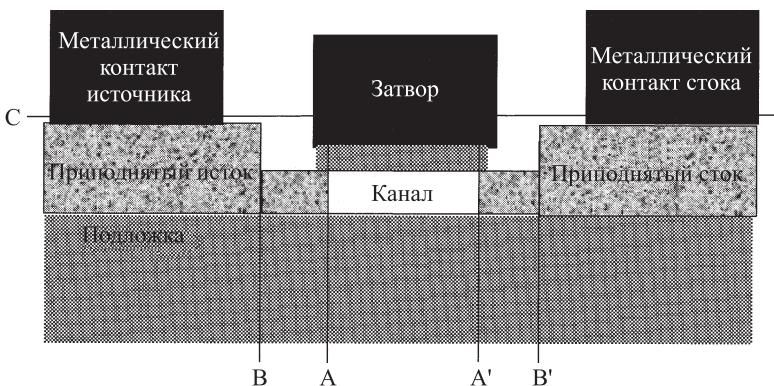


Рис. 8.4.5. Схема ультратонкого полевого транзистора типа «кремний на изоляторе». Показаны возможные положения линий между  $H$  и  $\Sigma$  (автор M. S. Lundstrom)

**Где следует проводить границу между гамильтонианом  $[H]$  и собственно-энергетической функцией  $[\Sigma]$ ?** Из практических соображений удобно, чтобы область, описываемая гамильтонианом  $[H]$ , имела наименьший возможный размер. Но при этом надо быть уверенным в том, что область, описываемая собственно-энергетической функцией  $[\Sigma]$ , будет находиться в состоянии локального равновесия. Метод собственно-энергетической функции, описанный в этой главе, оказывается точным и полезным даже тогда, когда отдельный контакт не может характеризоваться гладкой локальной плотностью состояний во всей области постоянной скорости релаксации (см., к примеру, [27]). Важнее то, чтобы процессы рассеяния были достаточно интенсивными и обеспечивали выполнение условий локального равновесия. К примеру, на рис. 8.4.5 показан ультратонкий транзистор с приподнятыми областями истока и стока. Такая конструкция разработана для того, чтобы уменьшить сопротивление. Если мы проведем линию, разделяющую  $H$  и  $\Sigma$  в сечениях А и А', то совершим

большую ошибку, поскольку область слева от сечения A (или справа от A') имеет то же сечение, что и сам канал, и нет оснований ожидать, что она будет в большей степени равновесной, чем сам канал. Поэтому необходимо сдвинуть сечения в плоскости B и B', или даже в плоскость C и за нее. Обсуждение некоторых из этих вопросов можно найти в работе [49].

Позвольте мне закончить данный параграф замечанием о том, что физическая роль контактов действительно универсальна и далеко не ограничивается приведенными здесь примерами. Когда мы вычисляем оптическое поглощение в полупроводнике, то молчаливо подразумеваем, что валентная зона и зона проводимости играют роль «истока» и «стока», находящихся в локальном равновесии (см. упражнение У.10.4). Следует заметить, что существуют примерыnanoструктур, где «контакты» находятся в сильно неравновесном состоянии. При моделировании взаимодействия электронов с колебаниями решетки, или фононами, последние обычно предполагаются находящимися в равновесии, но и этот «контакт» также может быть выведен из равновесия. Более тонким примером является эффект Оверхаузера, состоящий в том, что электронные спины выводят ядерные спины достаточно далеко из равновесия, поскольку последние, будучи изолированными от окружения, не могут быстро релаксировать (см., к примеру, работу [47]). Утверждение состоит в том, что квантовые процессы взаимодействия присутствуют во всех типах «контактов» (не обязательно металлических), по-добных контактам на рис. 8.4.5. Так что не стоит рассчитывать на то, что они будут вести себя как хорошие резервуары, не исследовав тщательно их структуру и динамику.

## Упражнения

**У.8.1.** Рассматривая одномерную модель с эффективной массой и одномерным контактом, показанным на рис. У.8.1, и учитывая выражение (8.1.9)  $G_R = [(E + i0^+)I - H_R]^{-1}$ , покажите, что  $G_R(0, 0) \equiv g_R = -\exp(ika)/t_0$ .

### У.8.2.

(а) В параграфе 8.1 мы, используя элементарную технику, получили выражение для собственно-энергетической функции одномерного контакта:

$$\Sigma(E) = -t_0 \exp(ika), \quad (8.1.7a)$$

где

$$E = E_c + 2t_0(1 - \cos ka). \quad (8.1.5)$$

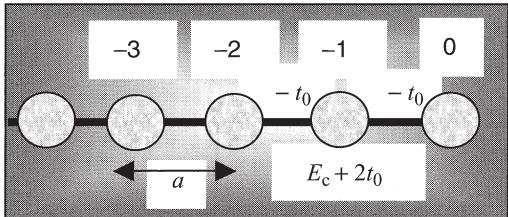


Рис. У.8.1

Исключив  $ka$  из (8.1.7a), получите выражение для собствено-энергетической функции, явно зависящее от  $E$ :

$$\Sigma(E)/t_0 = (x - 1) - i\sqrt{2x - x^2},$$

где

$$x \equiv (E - E_c)/2t_0, \quad 0 \geq x \geq 2.$$

Знак перед квадратным корнем выбирается исходя из того, чтобы мнимая часть  $\Sigma$  была отрицательной. Для значений  $x$ , при которых величина под знаком корня отрицательна (что отвечает значениям энергии вне разрешенной зоны), следует выбирать знак таким образом, чтобы величина  $\Sigma$  стремилась к нулю при больших значениях  $x$ :

$$\Sigma(E)/t_0 = (x - 1) + \sqrt{x^2 - 2x}, \quad -\infty \geq x \geq 0,$$

$$\Sigma(E)/t_0 = (x - 1) - \sqrt{x^2 - 2x}, \quad 2 \geq x \geq +\infty.$$

Постройте график собствено-энергетической функции на интервале  $-1 \geq x \geq 3$ . У вас должен получиться график, напоминающий тот, что показан на рис. У.8.2.

**(6)** Уравнение (8.1.7a) полезно вывести заново, используя более общую формулу (8.1.15a). Для полу бесконечного одномерного контакта с  $N$  узлами, отстоящими друг от друга на расстоянии  $a$ , собственные значения энергии определяются выражением

$$\varepsilon_n = 2t_0(1 - \cos k_n a).$$

Соответствующие нормированные функции записываются как

$$\psi_n(x_\alpha) = \sqrt{2/N} \sin(k_n x_\alpha), \quad \text{где} \quad k_n = n\pi/Na \quad \text{и} \quad x_\alpha = \alpha a,$$

так что  $[\tau]_{in} = [\tau]_{i\alpha}[\tau]_{\alpha n} = -t_0 \sqrt{2/N} \sin k_n a$ .

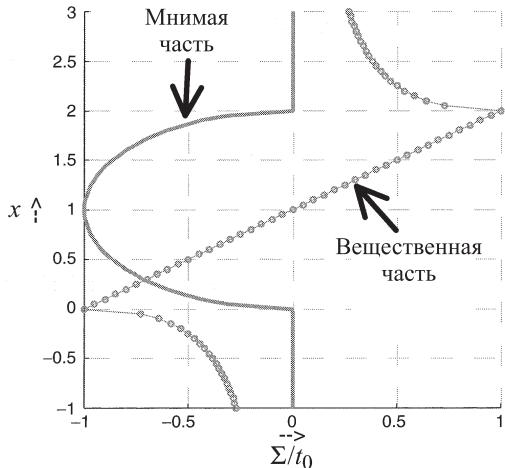


Рис. У.8.2

Из выражения (8.1.15a) следует, что

$$\Sigma(E) = \frac{2}{N} \sum_n \frac{t_0^2 \sin^2 k_n a}{E - \varepsilon_n + i0^+}.$$

Перейдите от суммирования к интегрированию и покажите, что

$$\Sigma(E) = \frac{2}{N} \int_0^{\pi/a} \frac{dk_n L}{\pi} \frac{t_0^2 \sin^2 k_n a}{E - \varepsilon_n + i0^+} = \int_0^{4t_0} \frac{d\varepsilon_n}{\pi} \frac{t_0 \sin k_n a}{E - \varepsilon_n + i0^+}.$$

**(в)** Аналитический расчет собственно-энергетической функции можно провести с использованием техники контурного интегрирования, но этот подход является достаточно сложным. Функцию уширения  $\Gamma(E)$  можно вычислить и другим способом — как удвоенную мнимую часть собственно-энергетической функции. Покажите, что функция уширения определяется выражением

$$\Gamma(E) = i(\Sigma - \Sigma^+) = 2t_0 \sin ka, \quad E_c \geq E \geq E_c + 4t_0,$$

где  $E = E_c + 2t_0(1 - \cos ka)$ , и равна нулю вне интервала  $E_c \geq E \geq E_c + 4t_0$ . Подсказка: можно воспользоваться выражением (8.2.10a).

(г) Покажите, что  $\Sigma(E) = \int_{-\infty}^{\infty} dE' \frac{\Gamma(E')/2\pi}{E - E' + i0^+}$ , и поэтому

$$\operatorname{Re}\Sigma(E) = P \left( \int_{-\infty}^{\infty} dE' \frac{\Gamma(E')/2\pi}{E - E' + i0^+} \right), \quad \operatorname{Im}\Sigma(E) = -\Gamma(E)/2,$$

где  $P$  обозначает главное значение интеграла.

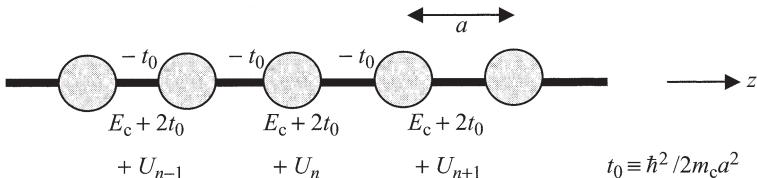


Рис. У.8.3

**У.8.3.** Рассмотрите одномерную проволоку с потенциалом  $U(x)$ , изменяющимся линейно от  $-0.1$  эВ на одном конце до  $+0.1$  эВ на другом конце (рис. У.8.3), и рассчитайте состояния в этой системе, используя однозонный гамильтониан с дискретизацией на решетке с 50 узлами, шагом  $a = 2$  Å и эффективной массой  $m_c$ , равной 0.25 массы свободного электрона  $m$ .

Рассчитайте электронную плотность  $n(z)$  в проволоке, предполагая, что она находится в состоянии равновесия с электрохимическим потенциалом  $\nu = E_c + 0.25$  эВ и  $k_B T = 0.025$  эВ, используя (а) периодические граничные условия и (б) метод собственно-энергетической функции. Сравните результат с изображенным на рис. 8.2.5. Вычислите локальную плотность состояний на двух концах проволоки, используя метод собственной энергетической функции, и сравните результат с рис. 8.2.6.

**У.8.4.** Рассмотрите задачу, обсуждаемую в начале параграфа 8.4, где речь идет о двух взаимодействующих атомах, описываемых гамильтонианом с матрицей  $(2 \times 2)$ :

$$H = \begin{pmatrix} \varepsilon & \tau \\ \tau^* & \varepsilon_1 \end{pmatrix},$$

и используйте те же параметры, что и в тексте:  $\varepsilon = -0.25$  эВ,  $\varepsilon_1 = +0.25$  эВ,  $\tau = 0.5$  эВ и  $0^+ = 0.01$  эВ.

(а) Рассчитайте матрицу  $(2 \times 2)$  спектральной функции и постройте локальную плотность состояний в точках «1» и «2» как функцию энергии.

(б) Вычислите локальную плотность состояний в точке «1», рассматривая атом «2» как собственно-энергетическую функцию, а также аналогично в точке «2», рассматривая атом «1» как собственно-энергетическую функцию, и сравните результат с ответом к пункту (а).

(в) Пусть низший (по энергии) пик на локальной плотности состояний в точке «1» имеет более высокую амплитуду, чем вышележащий пик (в точке «2» относительные величины пиков противоположны). Можете ли вы объяснить это в терминах собственных функций гамильтониана  $[H]$ ?

### У.8.5.

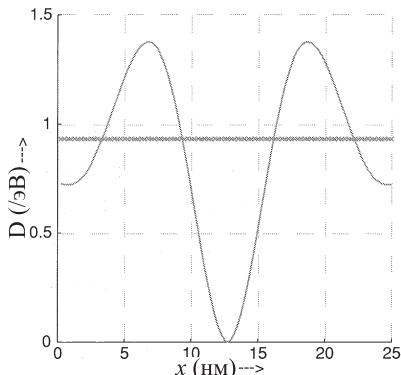
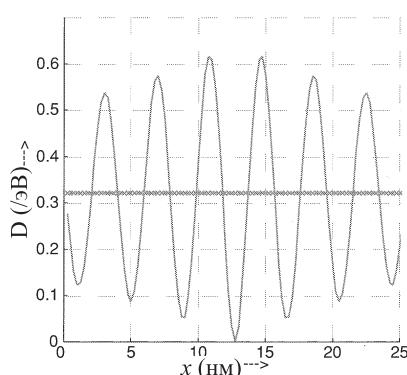
(а)  $E = 0.01$  эВ(б)  $E = 0.1$  эВ

Рис. У.8.5

Пусть имеется линейный проводник с отталкивающим потенциалом  $U(x) = U_0\delta(x)$  внутри. Рассчитайте локальную плотность состояний (LDOS) как функцию  $x$  при  $E = 0.005$  эВ и при  $E = 0.05$  эВ. Используйте однозонную модель сильной связи со 101 узлом решетки и постоянной решетки  $a = 0.25$  нм, предполагая, что  $U_0 = 5$  эВ и эффективная масса равна  $0.25m$ . У вас должны получиться графики, напоминающие показанные на рис. У.8.5.

Локальная плотность состояний обращается в нуль непосредственно в месте нахождения примеси, но при этом осциллирует возле нее с периодом, определяемым электронной длиной волны, отвечающей энергии  $E$ . Такие осцилляции LDOS наблюдались при использовании сканирующей туннельной микроскопии. Читателю было бы полезно построить контурный график зависимости  $D(x, E)$  на плоскости  $x - E$ , похожий на рис. 9 работы [29].

## Глава 9

# Когерентный транспорт

Прежде чем приступить к изучению этой главы, читателю будет полезно вспомнить содержание параграфа 1.6.

### 9.1. Предварительный обзор

Поскольку эта глава достаточно велика, я начну с подробного обзора, который одновременно можно рассматривать и как сводку основных результатов. В первой главе обсуждается очень простая модель протекания тока, а именно, модель с одним уровнем энергии  $\varepsilon$  и двумя контактами, обозначенными как исток и сток. Величина взаимодействия с истоком (или стоком) характеризуется скоростью  $\gamma_1/\hbar$  (или  $\gamma_2/\hbar$ ), с которой электрон, занимающий в начальный момент данный уровень, покидает его, уходя в контакт истока (или стока).

Я отмечал, что ток возникает лишь в том случае, если исток и сток находятся в состояниях локального равновесия, но поддерживаются при различных значениях электрохимического потенциала, и поэтому описываются двумя различными функциями Ферми:

$$f_1(E) \equiv f_0(E - \mu_1) = \frac{1}{\exp[(E - \mu_1)/k_B T] + 1}, \quad (9.1.1a)$$

$$f_2(E) \equiv f_0(E - \mu_2) = \frac{1}{\exp[(E - \mu_2)/k_B T] + 1}. \quad (9.1.1b)$$

Разность химических потенциалов пропорциональна напряжению  $V$ :  $\mu_2 - \mu_1 = -qV$ . При таких условиях число электронов в истоке определяется функцией  $f_1(E)$ , а на стоке отвечает значению функции  $f_2(E)$ . В стационарном состоянии число электронов в канале  $N$  соответствует некоторой промежуточной функции распределения. Исток «поставляет» электроны в канал, в то время как через сток они удаляются из канала. Каждый

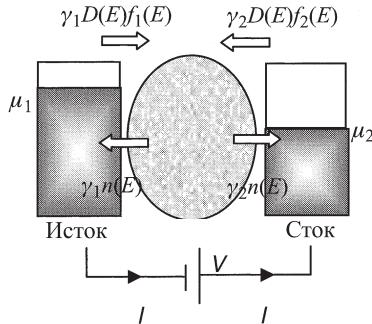


Рис. 9.1.1. Поток электронов, входящих в канал и выходящих из него: рисунок отвечает модели независимых уровней, см. уравнения (1.6.4)–(1.6.6)

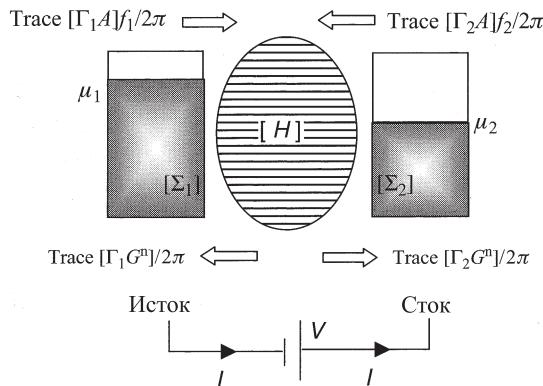


Рис. 9.1.2. Входящий и выходящий потоки в произвольной многоуровневой структуре, чьи уровни энергии описываются гамильтоновской матрицей  $[H]$ . Взаимодействие с контактами истока и стока описывается матрицами собственно-энергетических функций  $[\Sigma_1(E)]$  и  $[\Sigma_2(E)]$  соответственно

контакт стремится восстановить локальное равновесие и в результате возникает ток во внешней цепи (см. рис. 9.1.1).

В данной главе развивается обобщенный подход к описанию систем со многими уровнями энергии (см. рис. 9.1.2). Такие системы описываются матрицами, а не скалярами:

$$\varepsilon \rightarrow [H] \quad \text{Матрица гамильтониана;}$$

$$\gamma_{1,2} \rightarrow [\Gamma_{1,2}] \quad \text{Матрица уширения,} \quad \Gamma_{1,2} = i[\Sigma_{1,2} - \Sigma_{1,2}^+].$$

В главе 8 мы видели, что при соединении структуры с резервуаром ее уровни энергии уширяются (расплющиваются), так что удобно говорить о непрерывно изменяющейся энергии  $E$ , а не о дискретном наборе собственных значений. Матрицу плотности запишем в виде (см. уравнение (8.2.13))

$$[\rho] = \int_{-\infty}^{+\infty} (dE/2\pi) [G^n(E)], \quad (9.1.2a)$$

где в состоянии равновесия

$$[G^n(E)]_{\text{eq}} = [A(E)]f_0(E - \mu). \quad (9.1.2b)$$

Подобно тому, как спектральная функция  $[A]$  представляет собой матричный вариант плотности состояний на единичный интервал энергии, корреляционная функция  $[G^n]$  является матричной версией электронной плотности на единичный интервал энергии.

**Неравновесная матрица плотности.** В параграфе 9.2 мы, прежде всего, покажем, что для структуры, соединенной с двумя контактами, которым отвечают две различные функции Ферми  $f_1(E)$  и  $f_2(E)$ , матрица плотности определяется выражением (9.1.2a) (аргумент  $E$  здесь опущен)

$$[G^n] = [A_1]f_1 + [A_2]f_2, \quad (9.1.3)$$

где

$$A_1 = G\Gamma_1 G^+ \quad \text{и} \quad A_2 = G\Gamma_2 G^+, \quad (9.1.4)$$

а

$$G = [EI - H - \Sigma_1 - \Sigma_2]^{-1}. \quad (9.1.5)$$

Этот результат можно интерпретировать следующим образом: часть спектральной функции, которой отвечает матрица  $[A_1]$ , остается в равновесии с истоком, описываемым функцией Ферми  $f_1$ , в то время как другая ее часть  $[A_2]$  остается в равновесии со стоком, которому отвечает функция Ферми  $f_2$ . Будет показано, что при сложении две данные части спектральной функции действительно дают полную спектральную функцию  $[A]$ , которая была введена в восьмой главе:

$$[A] \equiv i[G - G^+] = [A_1] + [A_2]. \quad (9.1.6)$$

**Ток.** Ниже в параграфе 9.3 мы покажем, что ток  $I_i$  в  $i$ -м контакте может быть записан в виде

$$I_i = (-q/h) \int_{-\infty}^{+\infty} dE \tilde{I}_i(E), \quad (9.1.7a)$$

где

$$\tilde{I}_i(E) = \text{Tr}[\Gamma_i A]f_1 - \text{Tr}[\Gamma_i G^n] \quad (9.1.7b)$$

есть безразмерный ток, приходящийся на единичный интервал энергии. Такую ситуацию иллюстрирует рис. 9.1.2, который дает квантовую версию элементарной картины протекания тока из первой главы (см. рис. 9.1.1).

**Одноуровневая модель.** В главе 1 был рассмотрен пример системы с одним энергетическим уровнем, в которой электронная плотность и ток находились из уравнения баланса с простой моделью уширения. Далее я отмечал, что в общем случае нам потребуется матричная версия этой «скалярной модели». В последующих главах книги мы будем иметь дело именно с матричной моделью (см. рис. 1.6.5).

Полезно убедиться в том, что полная «матричная модель», о которой мы заявили выше и которую будем рассматривать в этой главе, при рассмотрении одноуровневой системы, когда все матрицы становятся обычными числами, приводит к прежним результатам (уравнения (1.6.4)–(1.6.6)):

Из выражения (9.1.5)  $G(E) = [E - \varepsilon + (i\Gamma/2)]^{-1}$ .

Из выражения (9.1.4)  $A_1(E) = \frac{\Gamma_1}{(E - \varepsilon)^2 + (\Gamma/2)^2}$ ,  $A_2(E) = \frac{\Gamma_2}{(E - \varepsilon)^2 + (\Gamma/2)^2}$ .

Из выражения (9.1.6)  $A(E) = \frac{\Gamma}{(E - \varepsilon)^2 + (\Gamma/2)^2}$ .

Из выражения (9.1.3)  $G^n(E) = A(E) \left( \frac{\Gamma_1}{\Gamma} f_1(E) + \frac{\Gamma_2}{\Gamma} f_2(E) \right)$ , что можно сравнить с формулой (1.6.4). Согласно выражению (9.1.7) ток на двух контактах определяется формулами (сравните с выражениями (1.6.5a), (1.6.5b)):

$$I_1 = \frac{q}{h} \int_{-\infty}^{+\infty} dE \Gamma_1 [A(E)f_1(E) - G^n(E)],$$

$$I_2 = \frac{q}{h} \int_{-\infty}^{+\infty} dE \Gamma_2 [A(E)f_2(E) - G^n(E)].$$

**Функция пропускания.** Выражение (9.1.7) совместно с формулами (9.1.3) и (9.1.6) приводит к результату, согласно которому

$$\bar{I}_1 = -\bar{I}_2 = \bar{T}(E)(f_1(E) - f_2(E)),$$

где коэффициент

$$\bar{T}(E) \equiv \text{Tr}[\Gamma_1 A_2] = \text{Tr}[\Gamma_2 A_1]. \quad (9.1.8)$$

Ток во внешней цепи определяется выражением

$$I = (q/h) \int_{-\infty}^{+\infty} dE \bar{T}(E)(f_1(E) - f_2(E)). \quad (9.1.9)$$

Величина  $\bar{T}(E)$ , появившаяся в выражении (9.1.9) для тока, называется функцией (или коэффициентом) *пропускания*. Она определяет скорость, с которой электроны перемещаются от истока к стоку. Зная гамильтониан структуры  $[H]$  взаимодействия с контактами, описываемого матрицами собственно-энергетических функций  $\Sigma_{1,2}$ , мы сможем вычислить ток как из выражения (9.1.7), так и из формулы (9.1.9). Этот метод может быть использован для анализа произвольной структуры, если электронный транспорт через нее является *когерентным*. Позвольте мне объяснить более подробно, что здесь имеется в виду.

Распространение электронов является когерентным, если процессы рассеяния не изменяют фазу волновой функции. Так, если электрон отражается от абсолютно твердого (т. е. неизменяемого) дефекта решетки, его распространение является когерентным<sup>1</sup>. Взаимодействие с дефектом можно учитывать введением соответствующего потенциала дефекта в гамильтониан  $[H]$ , и после этого вычислять ток, руководствуясь схемой, приведенной на рис. 9.1.2. Однако, если электрон передал часть своей энергии атомной решетке, вызвав ее колебания, произойдет сбой фазы, и такой эффект неупругого рассеяния нельзя включить в гамильтониан  $[H]$ . Вопрос о том, как этот эффект следует учитывать, обсуждается в главе 10.

Замечу, что когерентный транспорт обычно рассматривается с помощью метода коэффициентов (функций) прохождения<sup>2</sup>, в котором предполагается, что структура соединена с контактами двумя идеальными электродами или многомодовыми квантовыми проволоками. Таким образом,  $S$ -матрицу данной структуры можно рассчитать, подобно тому, как это делается в микроволновом волноводе (рис. 9.1.3). Матрица рассеяния  $s_{21}$  (или  $s_{12}$ ) имеет размер  $M \times N$  (или  $N \times M$ ), если в первом электроде присутствует  $N$  мод, а во втором  $M$  мод. Функция пропускания находится как след от матричного произведения:  $\bar{T}(E) = \text{Tr}[s_{12}s_{12}^+] = \text{Tr}[s_{21}s_{21}^+]$ . Этот подход широко распространен и особенно он удобен для тех, кто знаком с приложениями метода  $S$ -матрицы в теории микроволновых волноводов.

<sup>1</sup> В русскоязычной литературе обычно говорят об упругом рассеянии. — Прим. перев.

<sup>2</sup> Здесь в оригинале используется термин *transmission formalism*. — Прим. ред. перев.

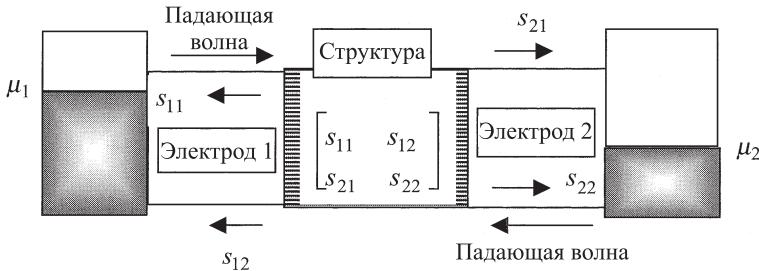


Рис. 9.1.3. Метод распространения предполагает, что структура соединяется с контактами через идеальные многомодовые квантовые проволоки (электроды), а функции (коэффициенты) пропускания связаны с  $S$ -матрицей области, лежащей между этими электродами

**Расчет функции пропускания с помощью функций Грина.** Если транспорт является когерентным, то функция пропускания может быть рассчитана с помощью метода функций Грина. Для этого нужно воспользоваться соотношением

$$\bar{T}(E) \equiv \text{Tr}[\Gamma_1 G \Gamma_2 G^+] = \text{Tr}[\Gamma_2 G \Gamma_1 G^+], \quad (9.1.10)$$

которое получается из (9.1.8) и формулы (9.1.4). В параграфах 9.2 и 9.3 будет дан вывод всех выражений, упомянутых во введении, (9.1.2a)–(9.1.7a). В данный момент, позвольте, я попытаюсь обосновать выражение (9.1.10), рассмотрев простой пример. Пусть простая одномерная проволока моделируется на дискретной решетке (рис. 9.1.4). Найдем коэффициент пропускания

$$\bar{T}(E) = (\nu_1/\nu_2)^2 |t|^2, \quad (9.1.11)$$

где отношение двух скоростей  $(\nu_1/\nu_2)$  появилось потому, что коэффициент пропускания равен отношению прошедшего тока к падающему току, а ток, в свою очередь, пропорционален скорости, умноженной на вероятность  $|\psi|^2$ .

Для расчета коэффициента пропускания по методу функций Грина обратимся к уравнению Шредингера  $[EI - H]\{\psi\} = \{0\}$ , которое описывает бесконечную систему, и, чтобы исключить из рассмотрения полубесконечные электроды, воспользуемся подходом, описанным в параграфе 8.1

$$[EI - H - \Sigma_1 - \Sigma_2]\{\psi\} = \{S\} \rightarrow \{\psi\} = [G]\{S\}, \quad (9.1.12)$$

где матрица  $[G]$  определяется выражением (9.1.5). Величины  $\Sigma_1$  и  $\Sigma_2$  являются матрицами, описывающими влияние двух электродов и имеющими по

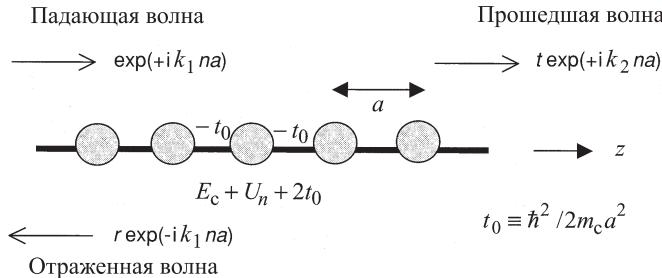


Рис. 9.1.4

одному ненулевому элементу (см. выражение (8.1.7a)):

$$\Sigma_1(1,1) = -t_0 \exp(ik_1 a), \quad \Sigma_2(N,N) = -t_0 \exp(ik_2 a),$$

что отвечает крайней точке канала с индексом 1 или  $N$ , к которой подключен электрод. Слагаемое, описывающее источника  $\{S\}$ , — это вектор-столбец с одной ненулевой компонентой, отвечающей крайней точке (1), на которую падает электронная волна (см. выражение (8.1.7b)):

$$S(1) = i2t_0 \sin k_1 a = i(\hbar\nu_1/a).$$

Заметим, что в общем случае при одной и той же энергии  $E$  волновые числа  $k$  (а стало быть, и скорости) могут быть различными на краях цепочки, поскольку потенциальная энергия  $U$  в этих точках может иметь различные значения:

$$E = E_c + U_1 + 2t_0 \cos k_1 a = E_c + U_N + 2t_0 \cos k_2 a.$$

Из выражения (9.1.12) имеем  $t = \psi(N) = G(N, 1)S(1)$ , откуда с помощью (9.1.11) найдем

$$\bar{T}(E) = (\hbar\nu_1/a)(\hbar\nu_2/a)|G(1, N)|^2,$$

что полностью соответствует результату, полученному из общего выражения (9.1.10).

Этот простой пример был предназначен для иллюстрации связи между подходом, основанным на использовании функции Грина, и методом коэффициентов пропускания. На мой взгляд, метод функций Грина имеет следующие преимущества.

(1) Общий, универсальный характер данного подхода говорит о том, что основные результаты могут быть применимы к каналам произвольной формы, описываемым матрицей  $[H]$  и имеющим произвольные контакты, описываемые матрицами  $[\Sigma_1]$  и  $[\Sigma_2]$ . Подобное разделение каналов и контактов очень полезно при работе с более сложными структурами.

(2) Метод функций Грина позволяет произвести расчет матрицы плотности (а значит, и электронной плотности). Этот расчет можно провести и в рамках формализма функции пропускания (см. работу [7]).

(3) Как будет показано в главе 10, подход, основанный на использовании функций Грина, позволяет описывать и некогерентный транспорт с расщеплением, нарушающим фазу. Напротив, в методе коэффициентов пропускания процессы со сбоем фазы описываются лишь феноменологически (см. работу [5]).

В параграфе 9.2 будет получено выражение для матрицы плотности (9.1.3)–(9.1.6), а в параграфе 9.3 будут найдены выражения, описывающие ток, (9.1.7a), (9.1.7b). Далее в параграфе 9.4 мы обсудим связь с методом коэффициентов пропускания, и в заключение в параграфе 9.5 приведем несколько иллюстрирующих примеров. Подводя итог, можно сказать, что параграфы 9.2–9.4 посвящены выводу результатов, анонсированных в данном введении, а в параграфе 9.5 обсуждаются некоторые приложения. Вначале читателю будет полезно заглянуть в параграф 9.5, чтобы познакомиться с примерами использования развивающегося аппарата, и увидеть, ради чего ему стоит затрачивать усилия на изучение предшествующих параграфов.

## 9.2. Матрица плотности

В этом параграфе будет приведен вывод ряда результатов, анонсированных в параграфе 9.1 и относящихся к неравновесной матрице плотности в канале, соединенном с двумя контактами (уравнения (9.1.3)–(9.1.6)). В следующем параграфе будет дан вывод уравнений, определяющих ток (уравнения (9.1.7)–(9.1.9)).

**Канал с одним контактом.** Прежде чем приступить к решению реальной задачи о канале, соединенном с двумя контактами, вернемся к задаче о канале с одним контактом и напомним некоторые принципиальные положения. В параграфе 8.1 мы с помощью уравнения Шредингера проанализировали состояния структуры, состоящей из одного контакта и канала:

$$\begin{pmatrix} EI_R - H_R + i\eta & -\tau^+ \\ -\tau & EI - H \end{pmatrix} \begin{pmatrix} \Phi_R + \chi \\ \psi \end{pmatrix} = \begin{pmatrix} S_R \\ 0 \end{pmatrix}, \quad (9.2.1)$$

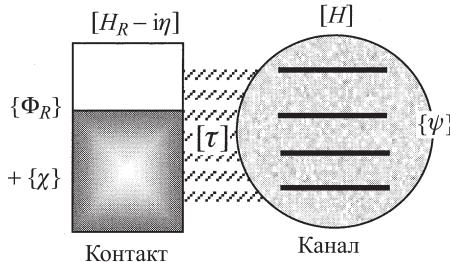


Рис. 9.2.1

и показали, что рассеянные волны  $\{\psi\}$  и  $\{\chi\}$  связаны с «проникновением» волновой функции  $\{\Phi_R\}$  в область изолированного контакта (рис. 9.2.1). Несложные матричные вычисления привели нас к следующему результату:

$$\{\chi\} = G_R \tau^+ \{\psi\}, \quad (9.2.2)$$

где

$$G_R \equiv [EI_R - H_R + i\eta]^{-1}, \quad (9.2.3)$$

$$\{\psi\} = [G]\{S\}, \quad (9.2.4)$$

$$G \equiv [EI - H - \Sigma]^{-1}, \quad (9.2.5)$$

$$\Sigma \equiv \tau G_R \tau^+, \quad (9.2.6)$$

$$\{S\} = \tau\{\Phi\}. \quad (9.2.7)$$

Поскольку имеется лишь один контакт, задача является равновесной, и матрица плотности выражается через спектральную функцию

$$A(E) = i[G - G^+] \quad (9.2.8)$$

и функцию Ферми (см. формулы (9.1.2a), (9.1.2b)). Этот результат можно получить и совершенно иным путем. Именно, будем предполагать, что состояния источника  $\{\Phi_R\}$  заполняются в соответствии с функцией Ферми, а состояния в канале заполняются вследствие просачивания из контакта. Я покажу, что получающаяся матрица плотности в канале совпадает с полученной ранее. Как только станет ясен метод решения этой задачи, мы применим его к реальной задаче с двумя kontaktами.

До того как канал будет соединен с контактом, электроны займут состояния контакта  $\alpha$  в соответствие с распределением Ферми, так что матрица плотности контакта примет вид

$$\rho_R(\vec{r}, \vec{r}') = \sum_{\alpha} \phi_{\alpha}(\vec{r}) f_0(\varepsilon_{\alpha} - \mu) \phi_{\alpha}^*(\vec{r}'),$$

или, в наших матричных обозначениях,

$$[\rho_R] = \sum_{\alpha} f_0(\varepsilon_{\alpha} - \mu) \{\phi_{\alpha}\} \{\phi_{\alpha}\}^+. \quad (9.2.9)$$

Теперь рассчитаем матрицу плотности канала, вычисляя отклик на возбуждение  $\tau\{\Phi\}$ , исходящее от контакта. Мы можем записать вклад в источник от каждого состояния  $\alpha$  в виде  $\{S_{\alpha}\} = \tau\{\phi_{\alpha}\}$ , затем найти волновую функцию в канале, используя (9.2.4) и (9.2.7),  $\{\psi_{\alpha}\} = G\tau\{\psi_{\alpha}\}$ . Далее матрица плотности в канале может быть найдена суммированием отдельных компонентов с весом, равным фактору Ферми для исходного состояния  $\alpha$  в контакте:

$$\begin{aligned} [\rho] &= \sum_{\alpha} f_0(\varepsilon_{\alpha} - \mu) \{\psi_{\alpha}\} \{\psi_{\alpha}\}^+ = \\ &= \int dE f_0(E - \mu) \sum_{\alpha} \delta(E - \varepsilon_{\alpha}) \{\phi_{\alpha}\} \{\phi_{\alpha}\}^+ = \\ &= \int dE f_0(E - \mu) G\tau \left[ \sum_{\alpha} \delta(E - \varepsilon_{\alpha}) \{\phi_{\alpha}\} \{\phi_{\alpha}\}^+ \right] \tau^+ G^+ = \\ &= \int \frac{dE}{2\pi} f_0(E - \mu) G\tau A_R \tau^+ G^+. \end{aligned} \quad (9.2.10)$$

Здесь используется выражение (8.2.3) для спектральной функции в контакте:

$$A_R(E) = 2\pi \sum_{\alpha} \delta(E - \varepsilon_{\alpha}) \{\phi_{\alpha}\} \{\phi_{\alpha}\}^+. \quad (9.2.11)$$

Из выражения (9.2.6) получим

$$\Gamma = i[\Sigma - \Sigma^+] = \tau A_R \tau^+, \quad (9.2.12)$$

и при помощи выражения (9.2.10) найдем, что

$$[G]^n = [G\Gamma G^+] f_0(E - \mu), \quad (9.2.13)$$

где мы учли (9.1.2a). Чтобы убедиться в том, что этот результат совпадает с полученным ранее (формулы (9.1.2a), (9.1.2b)), нам потребуется следующее *важное тождество*. Если

$$[G] = [EI - H - \Sigma]^{-1} \quad \text{и} \quad \Gamma = i[\Sigma - \Sigma^+],$$

то

$$A \equiv i[G - G^+] = G\Gamma G^+ = G^+\Gamma G. \quad (9.2.14)$$

Это можно показать, записывая  $(G^+)^{-1} - G^{-1} = \Sigma - \Sigma^+ = -i\Gamma$  и далее умножая слева на  $G$  и справа на  $G^+$ , что приведет к выражению

$$G - G^+ = -iG\Gamma G^+ \rightarrow A = G\Gamma G^+.$$

С другой стороны, если умножить на  $G^+$  слева и на  $G$  справа, то получим

$$G - G^+ = -iG^+\Gamma G \rightarrow A = G^+\Gamma G.$$

Я воспользовался этой простой задачей с одним контактом для того, чтобы проиллюстрировать важный физический принцип, согласно которому различные собственные состояния являются *некоррелированными*. Поэтому их вклады в матрицу плотности следует вычислять независимо, а затем суммировать.

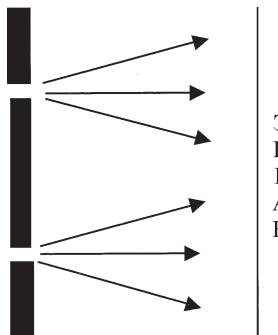


Рис. 9.2.2

Обнаруженное нами свойство в известном смысле напоминает интерференционный эксперимент Янга с двумя щелями (рис. 9.2.2). Если обе щели освещаются когерентным источником, то распределение интенсивности на экране имеет интерференционный характер. Но в том случае, когда щели освещены источником некогерентного излучения, интенсивность на экране будет просто суммой интенсивностей от каждой щели. В нашем случае можно сказать, что каждое состояние  $\alpha$  играет роль «щели», каждая из которых «освещает» структуру, причем эти «щели» не обладают фазовой когерентностью. Вот почему мы вычисляем матрицу плотности для каждой «щели»  $\alpha$  независимо и далее суммируем результаты.

Теперь, когда мы разобрали этот пример, можно сформулировать следующее правило работы с волновыми функциями в контакте и канале:

$$\{\Phi_R\}\{\Phi_R^+\} \Rightarrow \int (dE/2\pi) f_0(E - \mu)[A_R(E)], \quad (9.2.15a)$$

$$\{\psi\}\{\psi^+\} \Rightarrow \int (dE/2\pi)[G^n(E)], \quad (9.2.15b)$$

отражающие тот факт, что электроны в контакте распределены в соответствии с функцией Ферми  $f_0(E - \mu)$ , а непрерывное распределение собственных состояний описывается спектральной функцией  $A_R(E)$ . Это правило дает возможность существенно сократить алгебраические выкладки. К примеру, чтобы рассчитать матрицу плотности, следует, прежде всего, записать результат для одного собственного значения:

$$\{\psi\} = G\tau\{\Phi_R\} \rightarrow \{\psi\}\{\psi\}^+ = G\tau\{\Phi_R\}\{\Phi_R\}^+\tau^+G^+,$$

и далее используя выражение (9.2.11), получить  $[G^n] = [G\tau A_R \tau^+ G^+] \times \times f_0(E - \mu)$ , что с учетом выражения (9.2.12) сводится к уравнению (9.2.11).

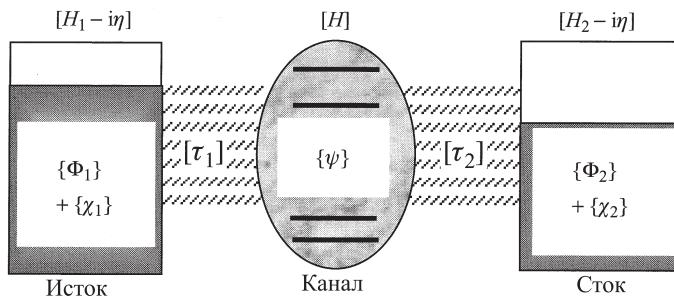


Рис. 9.2.3. Канал, соединенный с двумя контактами

**Канал с двумя контактами.** Теперь мы готовы взяться за реальную задачу с двумя контактами. Предположим, что до соединения с каналом электроны в контактах истока и стока имели волновые функции  $\{\Phi_1\}$ ,  $\{\Phi_2\}$ , подчиняющиеся «уравнениям Шредингера» для изолированных контактов:

$$[EI - H_1 + i\eta]\{\Phi_1\} = \{S_1\} \quad \text{и} \quad [EI - H_2 + i\eta]\{\Phi_2\} = \{S_2\}, \quad (9.2.16)$$

где  $[H_1]$  и  $[H_2]$  — это гамильтонианы первого и второго контактов соответственно. Чтобы учесть диссипацию, как это делалось ранее, в уравнения

введена бесконечно малая положительная величина, умноженная на единичную матрицу:  $[\eta] = 0^+[I]$ . Когда канал соединяется с контактами, как это показано на рис. 9.2.3, электронные состояния «просачиваются» из контактов, увеличивая амплитуду волновой функции  $\{\psi\}$ , а та, в свою очередь, порождает рассеиваемые волны  $\{\chi_1\}$  и  $\{\chi_2\}$  в контактах истока и стока соответственно.

Полная волновая функция будет удовлетворять комбинированному уравнению Шредингера для составной системы, состоящей из первого контакта, канала (или структуры) и второго контакта. Соответствующее уравнение может быть записано в виде трехблочной матрицы:

$$\begin{pmatrix} EI - H_1 + i\eta & -\tau_1^+ & 0 \\ -\tau_1 & EI - H & -\tau_2 \\ 0 & -\tau_2^+ & EI - H_2 + i\eta \end{pmatrix} \begin{pmatrix} \Phi_1 + \chi_1 \\ \psi \\ \Phi_2 + \chi_2 \end{pmatrix} = \begin{pmatrix} S_1 \\ 0 \\ S_2 \end{pmatrix}, \quad (9.2.17)$$

где матрица  $[H]$  описывает гамильтониан канала. Выполняя стандартные матричные преобразования, мы получаем из первого и последнего уравнений этой системы

$$\{\chi_1\} = G_1 \tau_1^+ \{\psi\} \quad \text{и} \quad \{\chi_2\} = G_2 \tau_2^+ \{\psi\}, \quad (9.2.18)$$

где

$$G_1 = [EI - H_1 + i\eta]^{-1} \quad \text{и} \quad G_2 = [EI - H_2 + i\eta]^{-1} \quad (9.2.19)$$

есть функции Грина изолированных резервуаров. Используя уравнение (9.2.18) для того, чтобы исключить  $\{\chi_1\}$  и  $\{\chi_2\}$  из второго уравнения системы (9.2.17), придем к уравнению

$$[EI - H - \Sigma_1 - \Sigma_2]\{\psi\} = \{S\}, \quad (9.2.20)$$

где

$$\Sigma_1 = \tau_1 G_1 \tau_1^+ \quad \text{и} \quad \Sigma_2 = \tau_2 G_2 \tau_2^+ \quad (9.2.21)$$

есть матрицы собственно-энергетических функций, которые были введены в главе 8. Соответствующие матрицы уширения (см. (9.2.12)) даются выражениями

$$\Gamma_1 = \tau_1 A_1 \tau_1^+ \quad \text{и} \quad \Gamma_2 = \tau_2 A_2 \tau_2^+, \quad (9.2.22)$$

где  $A_1 \equiv i[G_1 - G_1^+]$  и  $A_2 \equiv i[G_2 - G_2^+]$  — это спектральные функции первого и второго изолированных контактов соответственно. Наконец, вектор-столбец

$$\{S\} \equiv \tau_1 \{\Phi_1\} + \tau_2 \{\Phi_2\} \quad (9.2.23)$$

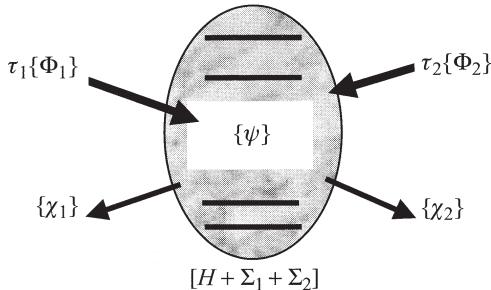


Рис. 9.2.4. Канал, в котором состояния возбуждаются слагаемыми  $\tau_1\{\Phi_1\}$  (от контакта истока) и  $\tau_2\{\Phi_2\}$  (от контакта стока). Отклик канала, определяемый уравнением (9.2.20), в свою очередь, порождает волновые функции  $\{\chi_1\}$  и  $\{\chi_2\}$  в контактах (см. уравнение (9.2.18))

представляет собой сумму слагаемых-источников  $\tau_1\{\Phi_1\}$  (от контакта истока) и  $\tau_2\{\Phi_2\}$  (от контакта стока), как показано на рис. 9.2.3.

Чтобы рассчитать матрицу плотности, определим функцию Грина в канале

$$G \equiv [EI - H - \Sigma_1 - \Sigma_2]^{-1}, \quad (9.2.24)$$

и используем ее для того, чтобы выразить волновую функцию в канале в терминах слагаемых источников  $\{S\}$  из уравнения (9.2.20):

$$\{\psi\} = G\{S\} \rightarrow \{\psi\}\{\psi\}^+ = G\{S\}\{S\}^+G^+. \quad (9.2.25)$$

Обратите внимание на то, что выделенные перекрестные слагаемые в выражении

$$SS^+ = \tau_1\Phi_1\Phi_1^+\tau_1^+ + \tau_2\Phi_2\Phi_2^+\tau_2^+ + \underbrace{\tau_1\Phi_1\Phi_2^+\tau_2^+ + \tau_2\Phi_2\Phi_1^+\tau_1^-}_{\text{перекрестные слагаемые}}$$

равны нулю, поскольку  $\Phi_1$  и  $\Phi_2$  — это волновые функции в контактах истока и стока, не связанных с каналом. Они физически удалены друг от друга и не взаимодействуют. Диагональные слагаемые вычисляются согласно ранее данному определению (см. уравнения (9.2.15a), (9.2.15b)):

$$\{\Phi_1\}\{\Phi_1^+\} \Rightarrow \int (dE/2\pi)f_1(E)[A_1(E)], \quad (9.2.26a)$$

$$\{\Phi_2\}\{\Phi_2^+\} \Rightarrow \int (dE/2\pi)f_2(E)[A_2(E)], \quad (9.2.26b)$$

откуда при помощи выражения  $\{\psi\}\{\psi\}^+ = G\{S\}\{S\}^+G^+$  матрица плотности записывается как

$$\rho = \int (dE/2\pi) \{ [G\tau_1 A_1 \tau_1^+ G^+] f_1 + [G\tau_2 A_2 \tau_2^+ G^+] f_2 \}.$$

Используя уравнение (9.2.22), это выражение можно упростить и записать функции

$$G^n = G\Sigma^{\text{in}}G^+ \quad (9.2.27)$$

и

$$\Sigma^{\text{in}} = [\Gamma_1]f_1 + [\Gamma_2]f_2, \quad (9.2.28)$$

имея в виду, что согласно определению (9.1.2a) имеем  $[\rho] = \int (dE/2\pi)[G^n]$ . Подобно тому как функция  $G^n$  определяется произведением  $\{\psi\}\{\psi\}^+$ , функция  $\Sigma^{\text{in}}$  связана с произведением  $\{S\}\{S\}^+$ . Исходя из этого, можно утверждать, что выражение (9.2.27) связывает «электронную плотность» в канале со слагаемым  $\{S\}$ , описывающим «просачивание» электронов из контактов.

**Частная спектральная функция (на левом и правом контактах).**  
Подставляя (9.2.28) в (9.2.27), можно получить, что

$$[G^n] = [A_1]f_1 + [A_2]f_2, \quad (9.2.29)$$

где

$$A_1 = G\Gamma_1 G^+ \quad \text{и} \quad A_2 = G\Gamma_2 G^+.$$

Сравнивая эти выражения с результатом, справедливым в состоянии равновесия (см. уравнение (9.1.2b))  $[G^n] = [A]f_0$ , можно заметить, что полную спектральную функцию  $[A(E)]$  можно рассматривать как состоящую из двух частей. Слагаемое  $[A_1(E)]$  появляется вследствие просачивания (или распространения) состояний из левого контакта, в то время как  $[A_2(E)]$  появляется от просачивания состояний из правого контакта. Заполнение первых описывается функцией Ферми  $f_1(E)$ , а вторых состояний — функцией Ферми  $f_2(E)$ . Чтобы показать, что две группы состояний суммируются и дают правильную полную спектральную функцию  $A = A_1 + A_2$ , мы заметим в соответствии с (9.2.14), что собственно-энергетическая функция  $\Sigma$  состоит из двух составляющих  $\Sigma_1$  и  $\Sigma_2$ , отвечающих двум контактам. Поэтому  $A = G[\Gamma_1 + \Gamma_2]G^+ = A_1 + A_2$  в соответствии с уравнением (9.1.6).

**Принцип запрета?** Прежде чем мы двинемся дальше, необходимо обсудить одно важное обстоятельство. В нашем подходе для расчета эволюции квантового состояния  $\Phi_\alpha$ , связанного с каким-либо контактом, используется уравнение Шредингера и далее, чтобы установить правила, сформулированные в (9.2.15a), (9.2.15b), (9.2.26a), (9.2.26b), рассматривается

суперпозиция таких состояний. Может показаться, что при суммировании всех индивидуальных вкладов мы забываем о принципе запрета Паули. Не выглядит ли это так, как будто присутствие электронов, эволюционирующих из одного состояния, не блокирует эволюцию других состояний? Ответ на этот вопрос будет отрицательным, коль скоро эволюция электронов является когерентной. Ответ легче всего обосновать, рассматривая временную эволюцию двух электронов, находящихся в различных состояниях:  $\{\Phi_1\}$  и  $\{\Phi_2\}$ . В начальный момент времени вопрос о блокировании одного состояния другим не возникает, поскольку эти состояния являются ортогональными:  $\{\Phi_1\}^+ \{\Phi_2\} = 0$ . В последующие моменты времени их волновые функции могут быть записаны как

$$\{\psi_1(t)\} = \exp[-iHt/\hbar]\{\Phi_1\},$$

$$\{\psi_2(t)\} = \exp[-iHt/\hbar]\{\Phi_2\},$$

если оба состояния эволюционируют когерентно в соответствии с уравнением Шредингера

$$i\hbar d\{\psi\}/dt = [H]\{\psi\}.$$

Тогда можно непосредственно показать, что перекрытие между любыми двумя состояниями не меняется в результате такой эволюции, т. е.  $\{\psi_1(t)\}^+ \{\psi_2(t)\} = \{\Phi_1\}^+ \{\Phi_2\}$ . Отсюда следует, что начальные ортогональные состояния эволюционируют так, что остаются ортогональными во все последующие моменты времени и никогда не блокируют друг друга в смысле принципа запрета Паули. Отметим, однако, что подобные аргументы не могут быть использованы, если в процессе эволюции происходит сбой фазы, поскольку такая эволюция не описывается одночастичным уравнением Шредингера.

### 9.3. Входящий и исходящий токи

Теперь, когда получены основные выражения для неравновесной матрицы плотности (см. (9.1.3)–(9.1.6)), мы можем обсудить расчеты тока на контактах (выражение (9.1.7)–(9.1.9)). Как и прежде, начнем с анализа системы «с одним контактом», схема которой приведена на рис. 9.2.1.

**Канал с одним контактом.** Рассмотрим вновь канал, соединенный с одним контактом и описывающийся уравнением

$$E \begin{pmatrix} \psi \\ \Phi \end{pmatrix} = \begin{pmatrix} H & \tau \\ \tau^+ & H_R - i\eta \end{pmatrix} \begin{pmatrix} \psi \\ \Phi \end{pmatrix},$$

совпадающим по форме с уравнением (9.2.1) с  $\{\Phi\} \equiv \{\Phi_R + \chi\}$ . Каким образом рассчитывается ток, протекающий между каналом и контактом? Как и в параграфе 6.4, где обсуждалось понятие скорости электрона в энергетической зоне, обратимся к зависящей от времени версии этого уравнения:

$$i\hbar \frac{d}{dt} \begin{pmatrix} \psi \\ \Phi \end{pmatrix} = \begin{pmatrix} H & \tau \\ \tau^+ & H_R - i\eta \end{pmatrix} \begin{pmatrix} \psi \\ \Phi \end{pmatrix}$$

и с его помощью найдем скорость изменения во времени электронной плотности вероятности внутри канала, определяемой как

$$\text{Tr}[\psi\psi^+] = \text{Tr}[\psi^+\psi] = \psi^+\psi.$$

Обратите внимание на то, что произведение  $\psi^+\psi$  является скаляром, и поэтому не важно, берется ли от него след или нет. Далее можно записать выражение для потока:

$$I \equiv \frac{d}{dt} \psi^+ \psi = \frac{\text{tr}[\psi^+\tau\Phi - \Phi^+\tau^+\psi]}{i\hbar}. \quad (9.3.1)$$

Имея в виду, что  $\{\Phi\} \equiv \{\Phi_R + \chi\}$ , мы можем мысленно разделить полный поток на входящий, пропорциональный «падающей» волне  $\{\Phi_R\}$ , и исходящий, пропорциональный «рассеянной» волне  $\{\chi\}$ :

$$I = \underbrace{\left( \frac{\text{tr}[\psi^+\tau\Phi_R - \Phi_R^+\tau^+\psi]}{i\hbar} \right)}_{\text{Входящий поток}} - \underbrace{\left( \frac{\text{tr}[\chi^+\tau^+\psi - \psi^+\tau\chi]}{i\hbar} \right)}_{\text{Исходящий поток}}. \quad (9.3.2)$$

Используя уравнения (9.2.4) и (9.2.7), а также соотношение  $i[G - G^+] = [A]$ , запишем исходящий поток следующим образом:

$$\text{Входящий поток} = \text{Tr}[S^+G^+S - S^+GS]/i\hbar = \text{Tr}[SS^+A]/\hbar,$$

Чтобы получить полный входящий поток, следует просуммировать входящие потоки от различных состояний  $\alpha$  контакта. Как мы убедились, все эти состояния будут учтены в силу соотношения (9.2.15a)

$$\{\Phi_R\}\{\Phi_R^+\} \Rightarrow \int \frac{dE}{2\pi} f_0(E - \mu)[A_R(E)].$$

Поскольку  $S = \tau\Phi_R$ , будем иметь

$$SS^+ \Rightarrow \int \frac{dE}{2\pi} f_0(E - \mu)\tau A_R \tau^+ = \int \frac{dE}{2\pi} f_0(E - \mu)[\Gamma],$$

и, следовательно, слагаемое с входящим потоком записывается как

$$\text{Входящий ток} = \frac{1}{\hbar} \int \frac{dE}{2\pi} f_0(E - \mu) \text{Tr}[\Gamma A]. \quad (9.3.3a)$$

Аналогично, с учетом выражений (9.2.2) и (9.2.12), можно записать слагаемое, определяющее исходящий поток:

$$\text{Исходящий ток} = \text{Tr}[\psi^+ \tau G_R^+ \tau^+ \psi - \psi^+ \tau G_R \tau^+ \psi] / i\hbar = \text{Tr}[\psi \psi^+ \Gamma] / \hbar.$$

Суммирование по всем состояниям эквивалентно интегрированию по энергии,  $\psi \psi^+ \Rightarrow \int dE G^n / 2\pi$ , поэтому

$$\text{Исходящий ток} = \frac{1}{\hbar} \int \frac{dE}{2\pi} \text{Tr}[\Gamma G^n]. \quad (9.3.3b)$$

Нетрудно увидеть, что в равновесии входящий и исходящий потоки равны, поскольку  $G^n = A f_0$  (см. выражение (9.1.2b)).

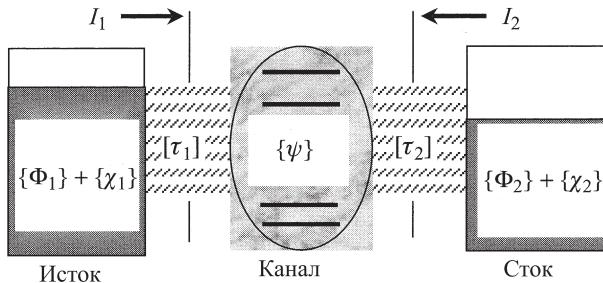


Рис. 9.3.1

**Канал с двумя контактами.** Теперь мы готовы к расчету входящего и исходящего токов в канале, имеющем два контакта (см. рис. 9.3.1). Будем рассматривать одну из границ, к примеру, относящуюся к контакту истока, и запишем входящий поток (сравните с уравнением (9.3.2)) в виде

$$I_1 = \underbrace{\left( \frac{\text{tr}[\psi^+ \tau_1 \Phi_1 - \Phi_1^+ \tau_1^+ \psi]}{i\hbar} \right)}_{\text{Входящий поток}} - \underbrace{\left( \frac{\text{tr}[\chi_1^+ \tau_1^+ \psi - \psi^+ \tau_1 \chi_1]}{i\hbar} \right)}_{\text{Исходящий поток}}.$$

Используя соотношения  $\psi = GS$  (где  $S$  и  $G$  определены в уравнении (9.2.23) и (9.2.24)) и  $\{S\} \equiv \tau_1 \{\Phi_1\}$ , сможем записать

$$\text{Входящий ток} = \text{Tr}[S^+ G^+ S_1 - S_1^+ GS] / i\hbar = \text{Tr}[S_1 S_1^+ A] / \hbar,$$

поскольку  $S = S_1 + S_2$  и  $S_1^+ S_2 = S_2^+ S_1 = 0$ .

Далее просуммируем входящие потоки от всех состояний  $\alpha$ , используя подстановку (см. выражение (9.2.26a))

$$\{\Phi_1\}\{\Phi_1^+\} \Rightarrow \int \frac{dE}{2\pi} f_1(E) [A_1(E)],$$

откуда  $\{S_1\}\{S_1^+\} = [\tau_1 \Phi_1 \Phi_1^+ \tau_1^+]$  и

$$\Rightarrow \int \frac{dE}{2\pi} [\tau_1 A_1 \tau_1^+] f_1(E) = \int \frac{dE}{2\pi} [\Gamma_1] f_1(E),$$

так что

$$\text{Входящий ток} = \frac{1}{\hbar} \int \frac{dE}{2\pi} f_1(E) \text{Tr}[\Gamma_1 A]. \quad (9.3.4a)$$

Аналогично выражения (9.2.18) и (9.2.22) можно применить для записи исходящего тока

$$\text{Исходящий ток} = \text{Tr}[\psi^+ \tau_1 G_1^+ \tau_1^+ \psi - \psi^+ \tau_1 G_1 \tau_1^+ \psi] / i\hbar = \text{Tr}[\psi \psi^+ \Gamma_1] / \hbar.$$

Суммирование по всем состояниям с учетом  $\{\psi\}\{\psi^+\} \Rightarrow \int (dE/2\pi)[G^n]$  дает

$$\text{Исходящий ток} = \frac{1}{\hbar} \int \frac{dE}{2\pi} \text{Tr}[\Gamma_1 G^n]. \quad (9.3.4b)$$

Полный электрический ток  $I_i$  в  $i$ -м контакте равен разности между входящим и исходящим потоками, умноженной на заряд электрона  $-q$  в соответствии с уравнением (9.1.7a):

$$I_i = (-q/\hbar) \int_{-\infty}^{+\infty} \frac{dE}{2\pi} \tilde{I}_i(E),$$

где

$$\tilde{I}_i(E) = \text{Tr}[\Gamma_i A] f_i - \text{Tr}[\Gamma_i G^n], \quad (9.3.5)$$

что иллюстрирует рис. 9.1.2.

## 9.4. Функция пропускания

В предыдущем параграфе мы определили ток в контакте как разность между входящим и исходящим токами. Сейчас будет получено другое выражение, которое позволяет рассмотреть задачу о протекании тока с другой

точки зрения и установить связь с широко используемым в литературе методом функции пропускания. Исходя из выражений (9.1.7b), (9.1.5) (9.1.6), можно получить

$$\bar{I}_1 = \bar{T}_{12}(E)[f_1(E) - f_2(E)], \quad \text{где} \quad \bar{T}_{12}(E) \equiv \text{Tr}[\Gamma_1 A_2];$$

$$\bar{I}_2 = \bar{T}_{21}(E)[f_2(E) - f_1(E)], \quad \text{где} \quad \bar{T}_{21}(E) \equiv \text{Tr}[\Gamma_2 A_1].$$

Если выполняется равенство  $\text{Tr}[\Gamma_1 A_2] = \text{Tr}[\Gamma_2 A_1]$ , то можно ожидать, что токи в двух контактах будут равны по величине и противоположны по направлению. Чтобы убедиться в этом, воспользуемся уравнением (9.2.14), откуда следует, что

$$\text{Tr}[\Gamma_1 A] = \text{Tr}[\Gamma_1 G \Gamma G^+] = \text{Tr}[\Gamma G^+ \Gamma_1 G] = \text{Tr}[\Gamma A_1].$$

Вычитая  $\text{Tr}[\Gamma_1 A_1]$  из обеих частей уравнения, мы получим желаемый результат  $\text{Tr}[\Gamma_1 A_2] = \text{Tr}[\Gamma_2 A_1]$  (заметим, что  $\Gamma = \Gamma_1 + \Gamma_2$  и  $A = A_1 + A_2$ ). Это позволяет записать следующее выражение для тока (с постоянной Планка  $h = 2\pi\hbar$ )

$$I = \frac{q}{h} \int_{-\infty}^{+\infty} dE \bar{T}(E)[f_1(E) - f_2(E)], \quad (9.4.1)$$

где величина

$$\begin{aligned} \bar{T}(E) &\equiv \text{Tr}[\Gamma_1 A_2] = \text{Tr}[\Gamma_2 A_1] = \\ &= \text{Tr}[\Gamma_1 G \Gamma_2 G^+] = \text{Tr}[\Gamma_2 G \Gamma_1 G^+] \end{aligned} \quad (9.4.2)$$

называется функцией или коэффициентом пропускания (прозрачности). С физической точки зрения мы можем рассматривать ток в выражении (9.4.1) как разность двух встречных потоков, идущих из истока и стока, как это изображено на рис. 9.4.1. Нашу структуру можно рассматривать как «полупроницаемую» мембрану, разделяющую два электронных резервуара (исток и сток), а функция  $\bar{T}(E)$  при таком рассмотрении будет определять проницаемость этой мембраны для электронов с энергией  $E$ . Мы увидим, что та же функция  $\bar{T}(E)$  управляет двумя встречными потоками, если транспорт является когерентным.

**Метод функции пропускания.** В методе функции пропускания (иногда называемом методом Ландауэра) предполагается, что канал соединяется с контактами двумя однородными электродами, которые можно рассматривать как квантовые проволоки с большим числом мод, или энергетических подзон (см. главу 6), с хорошо определенными дисперсионными зависимостями  $E(k)$  (см. рис. 9.4.2). Данное обстоятельство позволяет определить

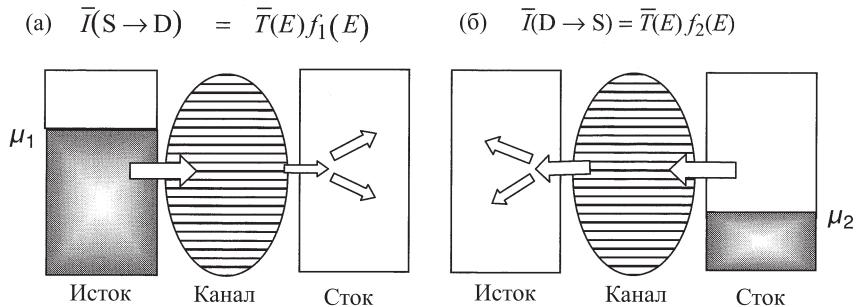


Рис. 9.4.1. Полный ток, протекающий через структуру (уравнение (9.4.1)), может рассматриваться как разность между двумя противоположно направленными потоками электронов: первый поток на рис. (а), для которого  $f_2 = 0$ , направлен от истока к стоку, а второй поток на рис. (б), для которого  $f_1 = 0$ , направлен от стока к истоку

$S$ -матрицу структуры аналогично тому, как это делается в задачах о микроволновых волноводах, где элемент  $t_{nm}$  матрицы  $T$  описывает амплитуду преобразования  $m$ -й моды, падающей на электрод 1 и проходящей через структуру в  $n$ -ю моду на электроде 2. Можно показать, что при таком подходе ток будет определяться выражением (9.4.1) с функцией пропускания, определяемой формулой

$$\bar{T}(E) = \sum_m \sum_n |t_{nm}|^2 = \text{Tr}[tt^+]. \quad (9.4.3)$$

Такой подход, основанный на довольно простых предположениях, весьма популярен и позволяет, изучая рассеяние, рассчитать функцию пропускания. В следующем параграфе, рассматривая простые примеры, мы покажем, что этот подход приводит к тем же результатам, что были получены из выражения  $\bar{T} = \text{Tr}[\Gamma_2 G \Gamma_1 G^+]$ , применяемого к структурам с однородными электродами.

**Формула Ландауэра.** Ландауэр впервые использовал теорию рассеяния в качестве основополагающего подхода к проблеме транспорта и указал на фундаментальную связь электрической проводимости с функцией пропускания: «проводимость есть пропускание». Основное соотношение можно получить, исходя из выражения (9.4.1) и используя при этом (9.1.1):

$$I = (q/h) \int_{-\infty}^{+\infty} dE \bar{T}(E) [f_0(E - \mu_1) - f_0(E - \mu_2)],$$

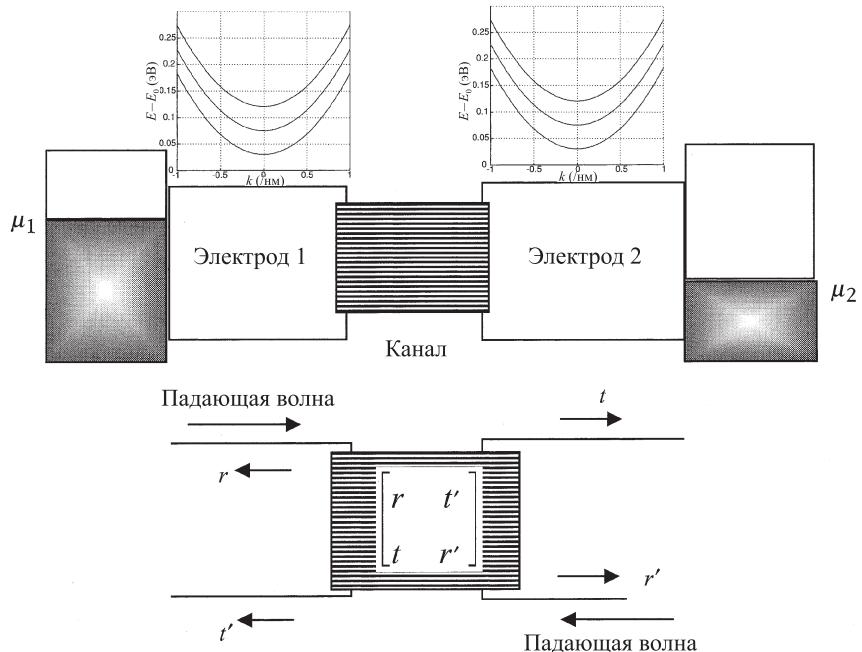


Рис. 9.4.2. В методе коэффициентов пропускания предполагается, что канал соединяется с контактами двумя однородными электродами, которые можно рассматривать как квантовые проволоки с большим числом энергетических подзон (см. главу 6), характеризующихся законами дисперсии  $E(k)$ , как это показано на рисунке. Данное обстоятельство позволяет определить  $S$ -матрицу структуры аналогично тому, как это делается при рассмотрении микроволновых волноводов

(обратите внимание на то, что в равновесном состоянии ток равен нулю, поскольку  $\mu_1 = \mu_2$ ). Небольшое приложенное напряжение  $V$  меняет каждую из функций  $\bar{T}$ ,  $\mu_1$  и  $\mu_2$ , так, что в результате в первом приближении ток записывается в виде

$$I \approx (q/h) \int_{-\infty}^{+\infty} dE \delta \bar{T}(E) [f_0(E - \mu_1) - f_0(E - \mu_2)] + \\ + (q/h) \int_{-\infty}^{+\infty} dE \bar{T}(E) \delta [f_0(E - \mu_1) - f_0(E - \mu_2)],$$

где символ  $\delta$  означает малое изменение. Первое слагаемое равно нулю, а второе может быть записано как

$$I \approx (q^2 V/h) \int_{-\infty}^{+\infty} dE (-\partial f_0(E)/\partial E)_{E=\mu}.$$

Поэтому проводимость дается формулой

$$G = (q^2/h)T_0, \quad \text{где } T_0 \equiv \int_{-\infty}^{+\infty} dE \bar{T}(E) F_T(E - \mu), \quad (9.4.4)$$

где функция  $F_T$  определяет температурное уширение, о котором речь шла в главе 7. Она имеет острый пик шириной  $k_B T$  вблизи значения  $E = \mu$  (см. рис. 7.3.4). Следовательно, проводимость пропорциональна функции пропускания, усредненной по интервалу энергий в несколько  $k_B T$  вблизи равновесного электрохимического потенциала  $\mu$ , подобно тому, как квантовая емкость пропорциональна усредненной плотности состояний (см. уравнение (7.3.8)).

Максимальное значение функции пропускания (а стало быть, и проводимости) достигается тогда, когда каждая из  $M$  подзон или мод одного электрода осуществляет идеальный перенос зарядов в другой электрод (см. рис. 9.4.2). Матрица  $[tt^+]$  в этом случае является диагональной матрицей размера  $(M \times M)$  с единицами на главной диагонали, и поэтому коэффициент пропускания равен  $M$ . Это рассуждение наводит на мысль о том, что максимальный коэффициент пропускания равен числу мод в электроде. Но что в таком случае произойдет, если структура окажется более узкой, чем электрод, в ней могут распространяться лишь  $N$  мод, где  $N < M$  (рис. 9.4.3 *a*)?

Можно утверждать, что подобная структура не может иметь коэффициент пропускания больший чем структура, подводящие электроды которой имеют тот же размер, что и сам канал (рис. 9.4.3 *b*); в любом случае электроны должны пройти через узкую область самой структуры (считается, что длина структуры не настолько мала, чтобы происходило прямое туннелирование). Поскольку вторая структура на рис. 9.4.3 *b* имеет максимальный коэффициент пропускания, равный  $N$ , это утверждение должно быть справедливым и для первой структуры на рис. 9.4.3 *a*, что подтверждается детальными расчетами. В общем случае следует ожидать, что максимальный коэффициент пропускания равен числу мод в наиболее узкой

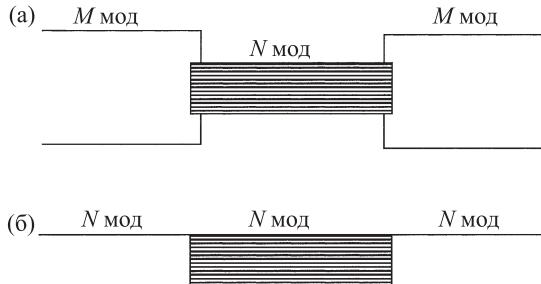


Рис. 9.4.3

части системы. Ранее в главе 6, исходя из проделанных расчетов максимального тока, который теоретически способна пропускать проволока, мы высказали мысль о том, что максимальная проводимость проволоки с  $N$  модами равна  $(q^2/h)N$ .

**Уравнения Бюттикера.** Измерения проводимости часто проводятся в структуре с четырьмя контактами (рис. 9.4.4), и их интерпретация для небольших структур была неясной до тех пор, пока Бюттикер не высказал свою элегантную идею (см. работу [5]). Он предположил, что формула Ландауэра

$$G = (q^2/h)\tilde{T} \rightarrow I = (q/h)\tilde{T}[\mu_1 - \mu_2]$$

может быть обобщена на случай структур со многими контактами. Для этого следует записать выражения для тока  $I_i$  в  $i$ -м контакте в виде

$$I_i = (q/h) \sum_j \tilde{T}_{ij} [\mu_i - \mu_j], \quad (9.4.5)$$

где величина  $\tilde{T}_{ij}$  есть усредненный коэффициент пропускания между  $j$ -м и  $i$ -м контактами. Нам известны значения электрохимических потенциалов  $\nu$  в контактах, через которые протекает ток (контакты 1 и 2), но мы не знаем их значений на контактах, к которым приложено напряжение и где потенциал формируется так, чтобы ток стал равным нулю. Каким образом с помощью уравнения (9.4.5) можно рассчитывать токи, если не известны все потенциалы? Идея расчета состоит в том, что при заданных любых четырех неизвестных величинах из восьми (четыре потенциала и четыре тока) найти оставшиеся четыре можно путем простых алгебраических вычислений. Фактически в этой задаче только шесть независимых переменных. Поскольку существенна лишь разность потенциалов, приводящая к появлению тока, одно из значений потенциала всегда можно принять за нуль.

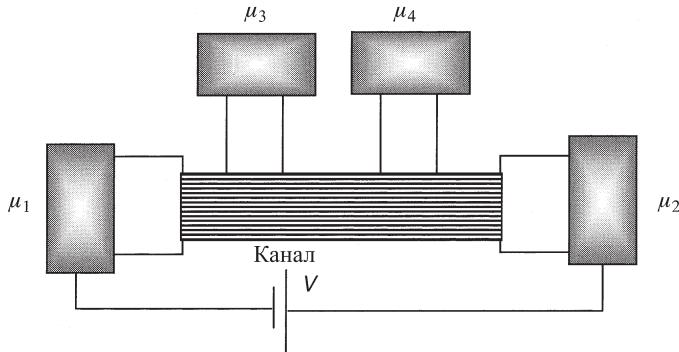


Рис. 9.4.4. Для измерения проводимости обычно используется четырехзондовая схема, описываемая уравнениями Бюттикера

Кроме того, поскольку законы Кирхгофа требуют равенства нулю полной алгебраической суммы всех токов, по известной величине трех токов можно найти четвертый ток. Таким образом, удобно задавать значение потенциала на одном контакте (к примеру, на контакте № 2), равное нулю, и далее записывать уравнение (9.4.5) в виде матричного уравнения размера  $(3 \times 3)$ :

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \frac{q}{h} \begin{pmatrix} \tilde{T}_{12} + \tilde{T}_{13} + \tilde{T}_{14} & -\tilde{T}_{13} & -\tilde{T}_{14} \\ -\tilde{T}_{31} & \tilde{T}_{31} + \tilde{T}_{32} + \tilde{T}_{34} & -\tilde{T}_{34} \\ -\tilde{T}_{41} & -\tilde{T}_{43} & \tilde{T}_{41} + \tilde{T}_{42} + \tilde{T}_{43} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}.$$

Зная величину  $\mu_1$ , а также то, что  $I_3 = 0$  и  $I_4 = 0$ , мы можем найти  $I_1$ ,  $\nu_3$  и  $\nu_4$ , а следовательно, и проводимость в геометрии четырех контактов  $G_4$ :

$$G_4 = (\nu_3 - \nu_4)/qI_1.$$

Если матричные элементы функции пропускания обладают свойством взаимности, т. е.  $\tilde{T}_{ij} = \tilde{T}_{ji}$ , то можно построить несложную эквивалентную схему, наглядно иллюстрирующую смысл уравнений Бюттикера. В этом случае уравнения Бюттикера будут эквивалентны законам Кирхгофа, применяемым к системе проводников с проводимостями  $G_{ij} \propto \tilde{T}_{ij} = \tilde{T}_{ji}$ , между парами контактов с индексами  $i$  и  $j$  (см. рис. 9.4.5). Такое описание, однако, невозможно в случае, когда матричные элементы функции пропускания не обладают свойством взаимности, т. е. когда  $\tilde{T}_{ij} \neq \tilde{T}_{ji}$ . Эти условия нарушаются в присутствии магнитного поля и, в частности, при холловских измерениях (заметим, что интерпретация результатов этих измерений

с помощью уравнений Бюттикера (9.4.5) является одним из наиболее известных примеров их применения).

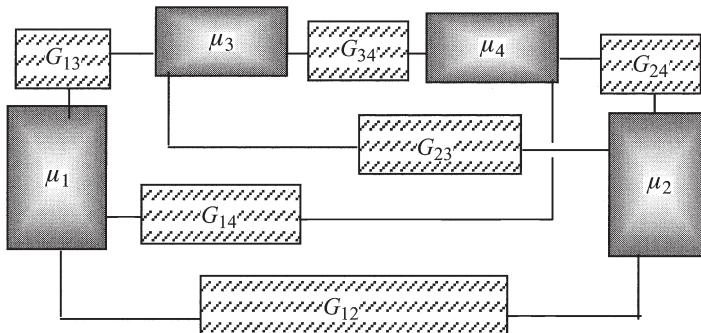


Рис. 9.4.5. Если матричные элементы функции пропускания между контактами обладают свойством взаимности, то системе уравнений Бюттикера может быть сопоставлена схема эквивалентного электрического контура

**Зонды Бюттикера.** Выше отмечалось, что описание транспорта на языке теории рассеяния возможно лишь тогда, когда он имеет когерентный характер и можно определить  $S$ -матрицу рассеяния. Применяемые в процессе измерений «плавающие» электроды (зонды) с нефиксированным потенциалом фактически извлекают электроны из структуры и затем возвращают их обратно со случайно измененной фазой. Они действуют подобно рассеивателям, нарушающим фазу. Это важнейшее наблюдение, сделанное Бюттикером, дает простой феноменологический способ учета процессов сбоя фазы при расчете тока. Именно, можно просто присоединять один или более воображаемых «плавающих» (зондов) электродов к нашей структуре и далее, используя уравнения Бюттикера рассчитать полный ток.

Более того, включив множественные плавающие контакты-рассеиватели, для расчета тока можно использовать и более общее выражение (см. уравнение (9.4.1)) вместо формулы, работающей только при малых напряжениях,

$$I_i = (q/h) \int_{-\infty}^{+\infty} dE \bar{I}_i(E), \quad (9.4.6)$$

где

$$\bar{I}_i(E) = \sum_j \bar{T}_{ij}(E)[f_i(E) - f_j(E)]. \quad (9.4.7)$$

Далее можно подобрать химический потенциал  $\nu_j$  так, чтобы значение  $j$ -го тока при любой энергии было равно нулю:  $\bar{I}_j(E) = 0$ . При этом, в принципе, химический потенциал  $\mu_j$  может иметь различные значения при различных энергиях. С другой стороны, можно потребовать, чтобы потенциал  $\mu_j$  имел фиксированное значение при всех энергиях, подобранное так, чтобы был равен нулю полный ток при всех энергиях:  $\int dE \bar{I}_j(E) = 0$ . В этом случае положительные значения  $\bar{I}_j(E)$  при одних энергиях уравновешиваются отрицательными значениями при других энергиях, обеспечивая равенство полного тока нулю. В такой ситуации можно сказать, что различные энергетические каналы оказываются связанными друг с другом, так как процессы рассеяния в «зонде» изменяют величину тока при различных энергиях.

Коэффициенты пропускания при данной энергии обычно вычисляются, если известна  $S$ -матрица составной структуры, включающей воображаемые зонды:

$$\bar{T}_{ij} = \text{Tr}[s_{ij}(E)s_{ij}^+(E)]. \quad (9.4.8)$$

Но с тем же успехом этот феноменологический подход можно сочетать с методом функций Грина. При этом плавающие электроды можно описать, используя собственно-энергетические функции  $[\Sigma_i]$  в выражении для коэффициента пропускания

$$\bar{T}_{ij}(E) = \text{Tr}[\Gamma_i G \Gamma_j G^+]. \quad (9.4.9)$$

Подобное выражение можно получить с помощью той же процедуры, что была описана ранее при изучении структур с двумя контактами: ток в  $i$ -м контакте равен разности между входящим и исходящим потоками:

$$I_i(E) = (1/\hbar) \text{Tr}\{[\Gamma_i(E)][[A(E)]f_i - [G^n(E)]]\}.$$

Используя соотношения  $A = \sum_j G \Gamma_j G^+$  и  $G^n = \sum_j G \Gamma_j G^+ f_j$  (см. уравнения (9.2.14), (9.2.27) и (9.2.28)), можем написать

$$I_i(E) = (1/\hbar) \sum_j \text{Tr}[\Gamma_i G \Gamma_j G^+](f_i - f_j),$$

так, что ток определится выражением

$$I_i(E) = (1/\hbar) \sum_j \tilde{T}_{ij}(f_i - f_j), \quad (9.4.10)$$

т. е. будет выражаться только через коэффициенты пропускания, определенные выражением (9.4.9).

**Правило сумм.** Согласно весьма полезному правилу, вытекающему из теории рассеяния, сумма строк или столбцов матрицы прохождения должна равняться числу мод:

$$\sum_j \bar{T}_{ij} = \sum_j \bar{T}_{ji} = M_i, \quad (9.4.11)$$

где  $M_i$  есть число мод в  $i$ -м контакте. Одним из важных следствий этого правила сумм является то, что в структуре с двумя контактами даже в присутствие магнитного поля выполняется соотношение  $\bar{T}_{12} = \bar{T}_{21}$ , поскольку для  $\bar{T}$ -матрицы размера  $(2 \times 2)$

$$\begin{pmatrix} \bar{T}_{11} & \bar{T}_{12} \\ \bar{T}_{21} & \bar{T}_{22} \end{pmatrix}$$

имеет место равенство

$$\bar{T}_{11} + \bar{T}_{12} = M_1 = \bar{T}_{11} + \bar{T}_{21},$$

откуда

$$\bar{T}_{12} = \bar{T}_{21}.$$

Обратите внимание на то, что подобные утверждения не будут иметь места для структуры с более чем двумя контактами. К примеру, для структуры с тремя контактами можно показать, что  $\bar{T}_{12} + \bar{T}_{13} = \bar{T}_{21} + \bar{T}_{31}$ , но условия  $\bar{T}_{12} = \bar{T}_{21}$  или  $\bar{T}_{13} = \bar{T}_{31}$  не будут выполнены.

Выражение для коэффициентов пропускания, основанное на использовании функции Грина (см. уравнение (9.4.9)), приводит к аналогичному правилу сумм:

$$\sum_j \bar{T}_{ij} = \sum_j \bar{T}_{ji} = \text{Tr}[\Gamma_i A]. \quad (9.4.12)$$

Справедливость последнего равенства можно установить, заметив, что

$$\sum_j \bar{T}_{ij} = \sum_j \text{Tr}[\Gamma_i G \Gamma_j G^+] = \text{Tr}[\Gamma_i G \Gamma G^+] = \text{Tr}[\Gamma_i A],$$

где на последнем шаге мы воспользовались уравнением (9.2.14). Подобным образом преобразуется и вторая сумма:

$$\sum_j \bar{T}_{ji} = \sum_j \text{Tr}[\Gamma_j G \Gamma_i G^+] = \text{Tr}[\Gamma_i G^+ \Gamma G] = \text{Tr}[\Gamma_i A].$$

Следует заметить, что здесь величина  $\text{Tr}[\Gamma_i A]$  играет ту же роль, что и число мод  $M_i$  в теории транспорта, основанной на теории рассеяния.

Интересно, что если  $M_i$  по определению является целым числом, то след матрицы  $[\Gamma_i A]$  может быть равен любому, не обязательно целому числу. К примеру, если структура очень мала и имеет только один энергетический уровень  $E = \varepsilon$ , но при этом взаимодействует с многочисленными резервуарами, то

$$\Gamma_i A = \frac{\Gamma_i \Gamma}{(E - \varepsilon)^2 + (\Gamma/2)^2}, \quad \text{где} \quad \Gamma = \sum_i \Gamma_i.$$

Заметим еще раз, что в том случае, когда рассеяние не зависит от энергии, уширение имеет форму кривой Лоренца и очевидно, что эта величина может иметь любое дробное значение.

## 9.5. Примеры

### 9.5.1. Аналитический пример

**Теория рассеяния.** Чтобы убедиться в том, что метод функций Грина приводит к тем же результатам, что и метод теории рассеяния, полезно рассмотреть простой пример, допускающий аналитическое решение, так что соответствующие результаты легко получаются на бумаге. Пусть мы имеем одномерный проводник с отталкивающим потенциалом  $U(z) = U_0\delta(z)$ , размещенным в точке  $z = 0$  (рис. 9.5.1). Коэффициенты  $r$  и  $t$  можно определить, исходя из требования непрерывности волновой функции при  $z = 0$  (см. рис. 9.5.1):

$$[\psi]_{z=0+} - [\psi]_{z=0-} = 0 \rightarrow t - (1 + r) = 0, \quad (9.5.1a)$$

а скачок производной удовлетворял бы следующему уравнению:

$$\left[ \frac{d\psi}{dz} \right]_{z=0+} - \left[ \frac{d\psi}{dz} \right]_{z=0-} = \frac{2mU_0}{\hbar^2} \rightarrow ik[t - (1 + r)] = \frac{2mU_0t}{\hbar^2}. \quad (9.5.1b)$$

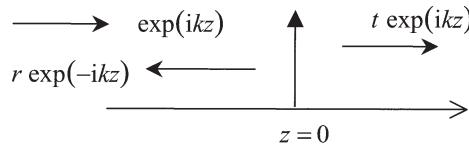


Рис. 9.5.1

Решение уравнения (9.5.1a) и (9.5.1b) приводит к следующему результату:

$$t = \frac{i\hbar\nu}{i\hbar\nu - U_0} \rightarrow \bar{T} = |t|^2 = \frac{\hbar^2\nu^2}{\hbar^2\nu^2 + U_0^2}.$$

**Теория рассеяния на дискретной решетке.** Решим эту задачу еще раз, используя дискретную решетку с узлами, отстоящими на расстоянии  $a$ . Пусть в центральной ячейке размещен дополнительный дельта-образный потенциал с амплитудой  $(U_0/a)$  (рис. 9.5.2). Можно проделать описанные выше расчеты на дискретной решетке, стартуя с записи выражения

$$E\psi_0 = [E_c + 2t_0 + (U_0/a)]\psi_0 - t_0\psi_{-1} - t_0\psi_{+1}, \quad (9.5.2)$$

и далее записать

$$\psi_0 = 1 + r = t,$$

$$\psi_{+1} = t \exp(+ika),$$

$$\psi_{-1} = \exp(-ika) + r \exp(+ika).$$

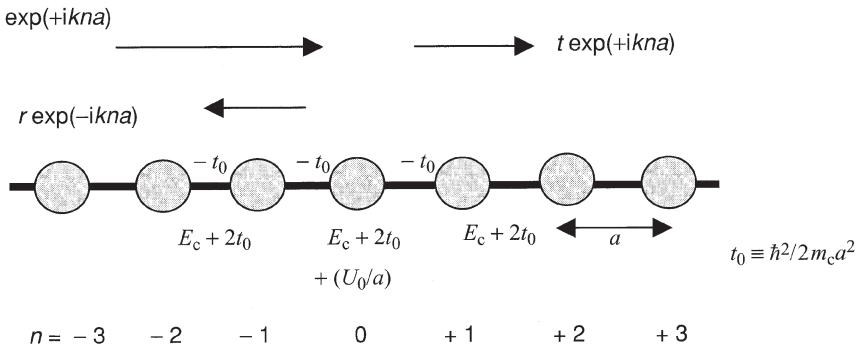


Рис. 9.5.2

Отсюда

$$\begin{aligned} \psi_{+1} &= \psi_0 \exp(+ika), \\ \psi_{-1} &= -2i \sin ka + \psi_0 \exp(+ika). \end{aligned} \quad (9.5.3)$$

Подставляя эти выражения в уравнение (9.5.2), найдем, что

$$[E - E_c - 2t_0 - (U_0/a) + 2t_0 \exp(+ika)]\psi_0 = 2it_0 \sin ka.$$

Используя далее дисперсионное соотношение

$$E = E_c + 2t_0(1 - \cos ka) \rightarrow \hbar\nu(E) = 2at_0 \sin ka, \quad (9.5.4)$$

придем к уравнению  $[-(U_0/a) + 2it_0 \sin ka]\psi_0 = 2it_0 \sin ka$ , откуда

$$\psi_0 = \frac{i\hbar\nu}{i\hbar\nu - U_0}. \quad (9.5.5)$$

Следовательно, в данной задаче коэффициент пропускания будет определяться выражением

$$\bar{T}(E) = |t|^2 = |\psi_0|^2 = \frac{\hbar^2 \nu^2(E)}{\hbar^2 \nu^2(E) + U_0^2}. \quad (9.5.6)$$

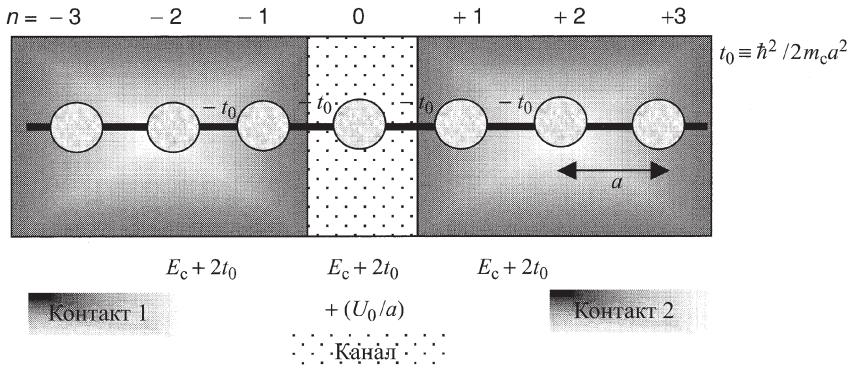


Рис. 9.5.3

**Метод функций Грина.** В заключение решим ту же задачу, применяя метод функций Грина, описанный в этой главе. Пусть роль «структурой» играет единственная точка, описываемая гамильтонианом, имеющим размер  $(1 \times 1)$  и определяемым выражением (см. рис. 9.5.3)

$$[H] = E_c + 2t_0 + (U_0/a),$$

в то время как влияние двух полубесконечных электродов (по одному с каждой стороны) описывается матрицами собственно-энергетических функций:

$$[\Sigma_1(E)] = -t_0 \exp(ika) \quad \text{и} \quad [\Sigma_2(E)] = -t_0,$$

где  $ka$  связан с энергией  $E$  дисперсионным соотношением (см. уравнение (9.5.4)), так что

$$[\Gamma_{1,2}(E)] = i[\Sigma_{1,2} - \Sigma_{1,2}^+] = 2t_0 \sin ka = \hbar\nu/a.$$

Поскольку все матрицы здесь имеют размер  $(1 \times 1)$ , нетрудно записать и функцию Грина:

$$G = [EI - H - \Sigma_1 - \Sigma_2]^{-1} = [E - E_c - 2t_0 + 2t_0 \exp(ika) - (U_0/a)]^{-1}.$$

Как и прежде, используя дисперсионное соотношение, упростим это выражение и получим функцию Грина

$$G = [i2t_0 \sin ka - (U_0/a)]^{-1} = a/(i\hbar\nu - U_0),$$

откуда следует, что коэффициент пропускания определяется формулой

$$\bar{T}(E) = \text{Tr}[\Gamma_1 G \Gamma_2 G^+] = \frac{\hbar^2 \nu^2(E)}{\hbar^2 \nu^2(E) + U_0^2},$$

что согласуется с более ранним результатом (см. выражение (9.5.6)).

### 9.5.2. Численный пример

Полные возможности метода функций Грина проявляются, разумеется, при решении более сложных задач. Коль скоро известна матрица гамильтониана  $[H]$  и матрицы собственno-энергетических функций  $\Sigma_1(E)$  и  $\Sigma_2(E)$ , процедура решения по сути становится стандартной. С использованием выражений (9.1.5) и (9.1.10) расчет коэффициента пропускания проводится стандартным образом. Разумеется, если контакты устроены сложным образом, то при расчете собственno-энергетических функций требуется некоторая дополнительная работа, но ее необходимо проделать лишь один раз. Кроме того, как мы отмечали ранее, без проведения подробных вычислений можно принять разумное приближение, основанное на уравнениях (8.3.12) и (8.3.14). Эта процедура может быть оправдана и с общефизической точки зрения, поскольку точная структура контактов все равно не известна в деталях. В обсуждаемых ниже примерах рассматриваются только одномерные контакты, для которых собственно-энергетическая функция записывается без труда.

Чтобы моделировать различные типы структур, мы используем одномерную дискретную решетку с шагом  $a = 0.3$  нм. Предполагается, что

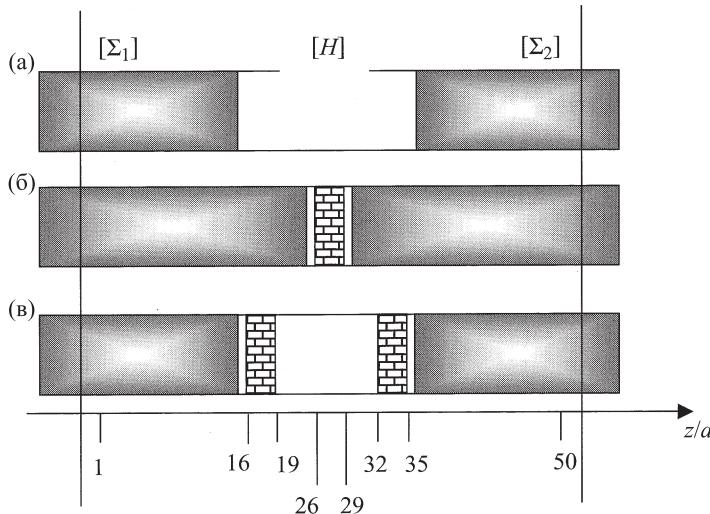
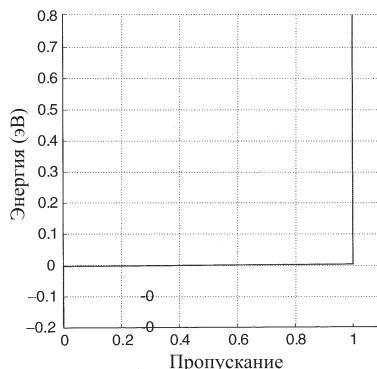
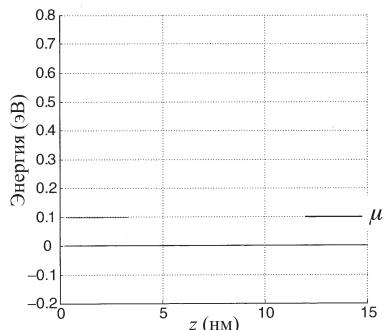


Рис. 9.5.4. Три примера структур: (а) баллистическая структура; (б) туннельная структура; (в) резонансная структура

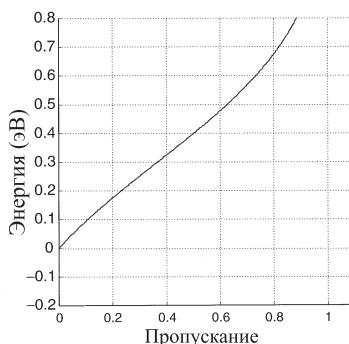
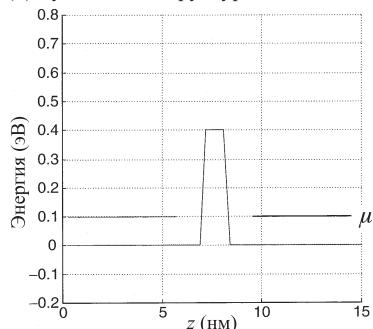
во всех структурах в поперечных направлениях  $x$  и  $y$  может распространяться только одна мода (см. рис. 9.5.4). Барьерные области обозначаются «стеной из кирпичиков». Зона проводимости в них лежит на 0.4 эВ выше, чем в остальных областях. Мы предполагаем, что эффективная масса везде имеет одно и то же значение  $m_c = 0.25m$ . На рис. 9.5.5 показаны несамосогласованные равновесные зонные диаграммы и функции пропускания  $\bar{T}(E)$ , рассчитанные численно с использованием гамильтоновской матрицы  $[H]$  и матриц собственно-энергетических функций  $\Sigma_{1,2}(E)$  для каждой из этих структур.

В баллистической структуре при энергиях, меньших края зоны  $E_c$ , пропускание отсутствует, а при энергии много большей, чем этот порог, пропускание стремится к единице. В туннельной структуре пропускание также растет от нуля до единицы, хотя и более медленно. С другой стороны, коэффициент пропускания резонансной туннельной структуры демонстрирует существенно иную зависимость от энергии. Здесь присутствуют два узких резонанса, появляющиеся вследствие того, что два барьера образуют прямоугольную потенциальную яму с (почти) дискретными уровнями энергии (рис. 9.5.6; см. также параграф 2.1). Коэффициент пропускания при прохождении частицы слева направо имеет пик при энергии частицы, сов-

(а) Баллистическая структура



(б) Туннельная структура



(в) Резонансная туннельная структура

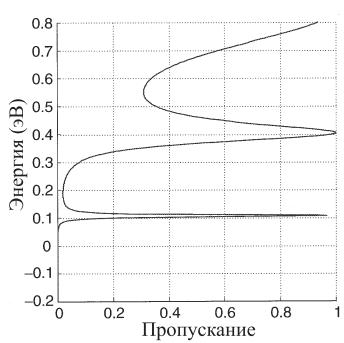
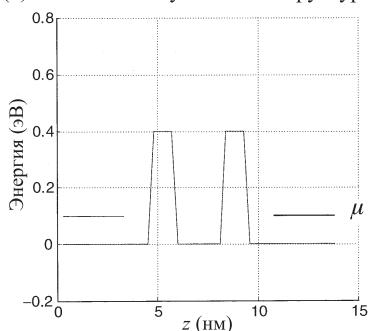


Рис. 9.5.5. Равновесные зонные диаграммы и коэффициенты пропускания для трех структур, показанных на рис. 9.5.4

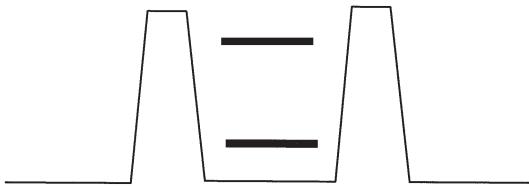


Рис. 9.5.6

падающей с одним из этих уровней потенциальной ямы. Этот результат можно получить, сшивая волновые функции и их производные на различных сечениях структуры; однако соответствующий расчет будет достаточно громоздким. С другой стороны, спектры потенциалов произвольного профиля легко включаются в матрицу гамильтониана  $[H]$ , после чего коэффициент пропускания рассчитывается непосредственно в рамках формализма функции Грина:  $\bar{T}(E) = \text{Tr}[\Gamma_1 G \Gamma_2 G^+]$ .

При расчете функции пропускания через структуру, имеющую острые резонансные пики (например, резонансную туннельную структуру), зачастую удобно включить в модель зонд Бюттикера (см. параграф 9.4). Причина состоит в том, что в численных расчетах можно пропустить острые резонансы, если дискретизация шкалы энергий недостаточно точная. Зонд Бюттикера играет ту же роль, что и процессы рассеяния, разрушающие fazу и приводящие к уширению резонанса. В таком подходе эффективный коэффициент пропускания определяется из решения уравнений Бюттикера (см. уравнение (9.4.5)), как это было проделано в предыдущем параграфе. В этом случае коэффициенты пропускания между различными контактами обладают свойством взаимности по отношению друг к другу, и поэтому эффективную проводимость нужно рассчитывать, используя простую модель сети сопротивлений (см. рис. 9.5.7), примененную к схеме с тремя контактами.

Учитывая, что проводимость пропорциональна пропусканию, можно записать формулу эффективной проводимости, используя элементарное правило расчета сопротивления последовательно и параллельно соединенных проводников:

$$\bar{T}(E)_{\text{eff}} = \bar{T}(E)_{12} + \frac{\bar{T}(E)_{13}\bar{T}(E)_{23}}{\bar{T}(E)_{13} + \bar{T}(E)_{23}}. \quad (9.5.7)$$

На рис. 9.5.8 приведен коэффициент эффективного пропускания резонансной туннельной структуры с одним зондом Бюттикера, находящемся

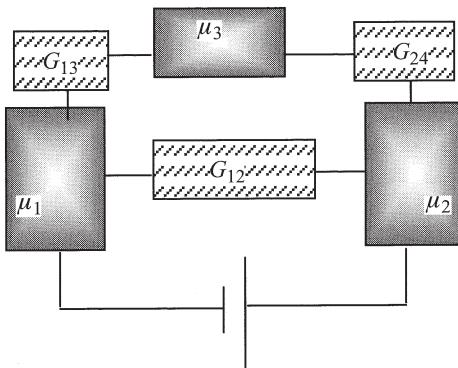


Рис. 9.5.7

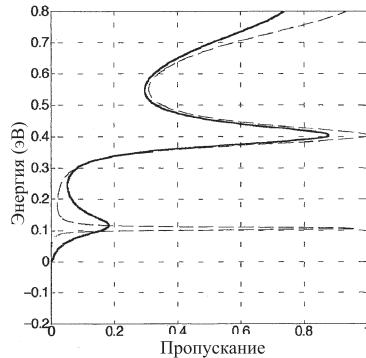
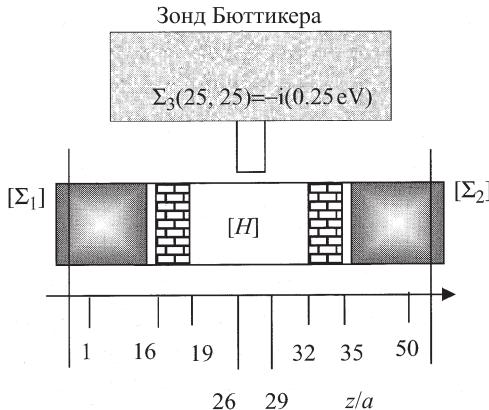


Рис. 9.5.8. Эффективная функция пропускания для резонансной туннельной структуры, в которой на узле решетки под номером 25 в центре структуры расположен зонд Бюттикера, предназначенный для феноменологического моделирования процессов нарушения фазы. Штриховой линией показан результат из рис. 9.5.5 в, когда зонд Бюттикера отсутствует

в центре структуры. Можно заметить, что по сравнению с предыдущим результатом (без зонда) резонансы стали более широкими, особенно самый узкий из них.

**Вольт-амперные характеристики  $I(V)$ .** Уравнение (9.1.9) может быть использовано для расчета вольта-амперных характеристик любой

структурой с когерентным транспортом, коль скоро известно падение напряжения на структуре. Это не столь важно, если нас интересует проводимость только при малых напряжениях (или «линейный отклик»), но является принципиально важным при определении полного профиля вольт-амперных характеристик, как это обсуждалось в параграфе 1.4.

В общем случае, для того чтобы получить правильные и точные результаты, необходимо сначала с помощью *самосогласованной* процедуры рассчитать профиль потенциала. Как и при рассмотрении равновесной задачи (см. рис. 7.2.1), самосогласованно рассчитанный потенциал  $U$  следует включить в полный гамильтониан  $H = H_0 + U([\delta\rho])$ .

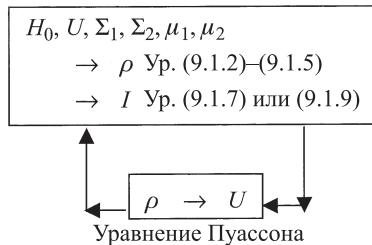


Рис. 9.5.9

Здесь  $U$  есть усредненный потенциал, связанный с изменением электронной плотности  $\delta\rho$ , или, точнее, матрицы плотности. На первом шаге рассчитывается электронная плотность, для чего используются диагональные элементы матрицы плотности. Далее электронная плотность включается в уравнение Пуассона, из которого находится потенциал, который вновь включается в гамильтониан, и электронная плотность рассчитывается заново. Этот процесс повторяется до тех пор, пока результат не сойдется, как это схематически показано на рис. 9.5.9. Полный самосогласованный расчет, подобный вышеописанному, может занять много времени (простой пример будет дан в параграфе 11.4). Поэтому зачастую предполагают наличие некоего «правдоподобного» потенциала. Возникает вопрос, как следует выбирать профиль такого потенциала?

Основной принцип выбора состоит в следующем. Если канал изолирован, и, следовательно, имеет низкую квантовую емкость, (см. выражение (7.3.8)), то профиль потенциала  $U_L(\vec{r})$  можно найти, решая уравнение Лапласа. В другом предельном случае, когда канал обладает металлическими свойствами и большой квантовой емкостью, профиль задается «нейтральным потенциалом»  $U_N(\vec{r})$ , который можно найти из транспорт-

ного уравнения в предположении о полной зарядовой нейтральности. Реальный потенциал будет иметь некоторое промежуточное значение. В областях с низкой плотностью состояний и малой квантовой емкостью профиль будет вести себя как функция  $U_L(\vec{r})$ , в то время как в областях с большой плотностью состояний и большой квантовой емкостью потенциал будет скорее напоминать функцию  $U_N(\vec{r})$ . Распространенная практика выбора «правдоподобного профиля» заключается в предположении, что потенциал имеет вид функции  $U_N(\vec{r})$  (необходимой для выполнения условия нейтральности) на краях структуры, которые должны быть областями с большой плотностью состояний, в то время как в центральной области канала потенциальный профиль должен по предположению иметь вид потенциала Лапласа  $U_L(\vec{r})$ .

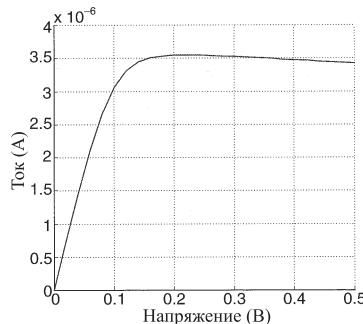
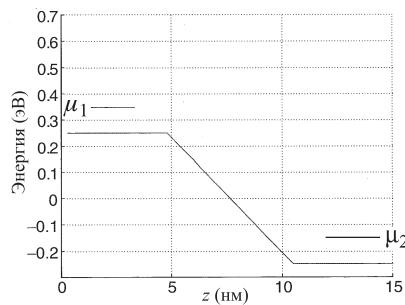
На рис. 9.5.10 показаны вольт-амперные характеристики (а) баллистической структуры, (б) туннельной структуры и (в) резонансной туннельной структуры, рассчитанные в предположении, что потенциал спадает линейно вдоль центральной незаштрихованной области из рис. 9.5.4. Сделанные предположения о форме потенциала приводят к качественно правильным выводам, но при этом нетрудно проверить, что результаты могут количественно измениться, если выбрать другие профили потенциала. К обсуждению этого вопроса мы еще вернемся в параграфе 11.4 при изучении факторов, влияющих на ток нанотранзистора.

## Упражнения

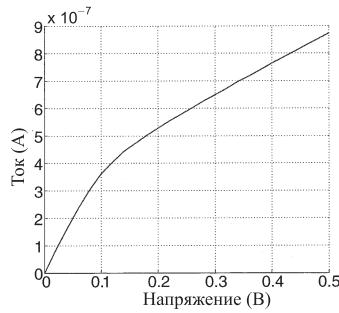
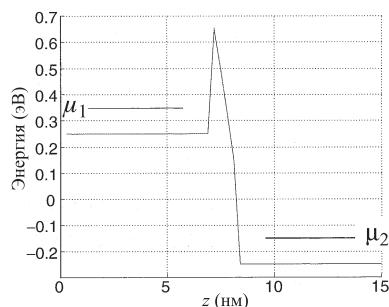
**Y.9.1.** Используя одномерную дискретную решетку с периодом  $a = 0.3$  нм, проведите моделирование каждой из показанных на рис. 9.5.4 структур, которые по предположению имеют единственную моду в попечевых направлениях  $x$  и  $y$ . Предположите, что эффективная масса везде имеет одно и то же значение  $m_c = 0.25m$  и области барьера, показанные в виде «стены из кирпичиков», имеют край зоны проводимости, располагающийся на 0.4 эВ выше, чем в остальных областях. (а) Выберите шаг по энергии на интервале  $-0.2$  эВ  $< E < 0.8$  эВ и постройте вероятность пропускания как функцию энергии. Сравните результат с приведенным на рис. 9.5.5. (б) Постройте, используя зонд Бюттика, зависимость вероятности пропускания от энергии для резонансной туннельной структуры, как это показано на рис. 9.5.8.

**Y.9.2.** (а) Рассчитайте вольт-амперные характеристики  $I = I(V)$  в интервале напряжений  $0 < V < 1$  В, предполагая, что приложенное

(а) Баллистическая структура



(б) Туннельная структура



(в) Резонансная туннельная структура

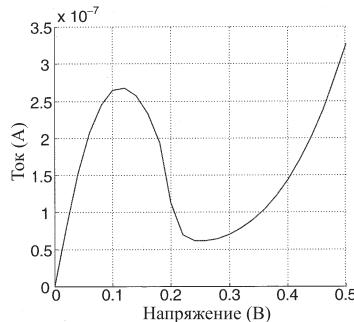
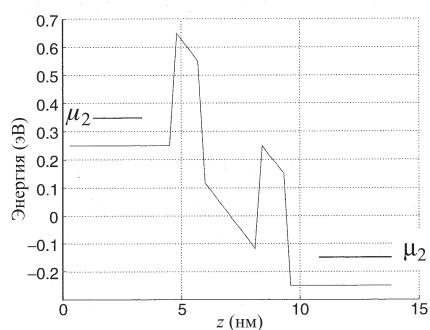


Рис. 9.5.10. Зависимости тока ( $I$ ) от напряжения ( $V$ ) для трех типов структур, показанных на рис. 9.5.4, рассчитанные в предположении о линейном профиле потенциала. На рисунках слева показаны соответствующие зонные диаграммы при напряжении, равном 0,5 В

к структуре напряжение падает в соответствии с профилем, показанным на рис. 9.5.10. Считайте, что равновесное значение энергии Ферми  $E_F = 0.1$  эВ, а при наличии внешнего напряжения химические потенциалы контактов  $\mu_1 = E_F + qV/2$  и  $\mu_2 = E_F - qV/2$ . Интегрирование по энергии должно проводиться лишь в интервале  $\mu_1 + 4k_B T < E < \mu_2 - 4k_B T$ . Используйте шаг по шкале энергии, равный  $\Delta E \approx 0.2k_B T$ . (б) Рассчитайте электронную плотность  $n(x)$  на единицу длины, предполагая, что приложенное напряжение 0.5 В спадает вдоль *туннельной структуры* в соответствие с профилем, показанным на рис. 9.5.10.

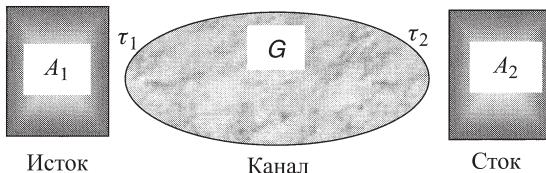


Рис. У.9.3

**У.9.3. Гамильтониан переноса.** Обратитесь к рис. У.9.3. Исходя из выражения (9.4.2) для коэффициентов пропускания  $\bar{T}(E) = \text{Tr}[\Gamma_1 G \Gamma_2 G^+]$  и от выражения (9.2.22) для матриц уширения, покажите, что

$$\bar{T}(E) = \text{Tr}[A_1 M A_2 M^+],$$

где  $A_1$  и  $A_2$  есть спектральные функции двух контактов, а матричный элемент  $M$  определяется выражением

$$M = \tau_1^\dagger G \tau_2.$$

Эта форма записи напоминает выражение, встретившееся в методе гамильтониана переноса (см., например, выражение (2.3.5) на с. 69 книги [9]). При использовании этого метода предполагается, что матричный элемент  $M$  не подвержен малому влиянию взаимодействия с контактами. Однако в нашей формулировке функция  $G$ , а значит, и  $M$  испытывают влияние контактов через их собственно-энергетическую функцию.

**У.9.4. Двумерное сечение.** В примерах, рассмотренных в параграфе 9.5, мы предполагали, что структура является одномерной. Нетрудно обобщить метод решения и на двумерные наноструктуры. При этом следу-

ет потребовать выполнения периодических граничных условий и предположить, что все поперечные моды не взаимодействуют друг с другом, как в главе 7 при расчете емкости. Далее можно попросту просуммировать наш одномерный результат по всем поперечным модам, задаваемым двумерным волновым вектором  $\vec{k}$ , и получить (с учетом  $\varepsilon_{\vec{k}} = \hbar^2 k^2 / 2m_c$ ):

$$\frac{I}{S} = \frac{q}{\pi \hbar} \int_{-\infty}^{+\infty} dE \bar{T}(E) [f_{2D}(E - \mu_1) - f_{2D}(E - \mu_2)].$$

Это означает, что ток в структуре с двумерным сечением вычисляется в рамках той же процедуры, которую мы применяли в одномерном случае, и далее проводится суммирование двумерной функции Ферми  $f_{2D}$  по  $k$  (см. выражение (7.2.12)). Проделайте вновь упражнение У.9.2, используя для расчета структуры с двумерным сечением — функцию  $f_{2D}$  (см. выражение (7.2.12)) вместо функции Ферми  $f_0$ . В этом случае ток должен иметь размерность  $A/m^2$ , а электронная плотность —  $m^{-3}$ .

**У.9.5. Одномерное сечение.** При анализе полевых транзисторов нам зачастую приходится иметь дело с суммированием по одномерному сечению (в направлении оси  $y$ ), в то время как функция пропускания должна рассчитываться в геометрии двумерной задачи (в плоскости  $z - x$ ) (см. рис. У.9.5). Предполагая выполнение периодических граничных условий в направлении  $y$ , покажите, что одномерное суммирование по  $k$  может быть выполнено аналитически, что даст

$$\frac{I}{W} = \frac{q}{\pi \hbar} \int_{-\infty}^{+\infty} dE \bar{T}(E) [f_{1D}(E - \mu_1) - f_{1D}(E - \mu_2)],$$

где одномерная функция Ферми под знаком интегрирования по  $k$  определяется выражением

$$f_{1D} \equiv \left( \frac{m_c k_B T}{2\pi \hbar^2} \right) I_{-1/2} \left( -\frac{E}{k_B T} \right),$$

а

$$I_{-1/2}(x) \equiv \frac{1}{\sqrt{\pi}} \int_0^{+\infty} \frac{dy}{1 + \exp(y - x)} \frac{1}{\sqrt{x}} = \frac{d}{dx} I_{1/2}(x),$$

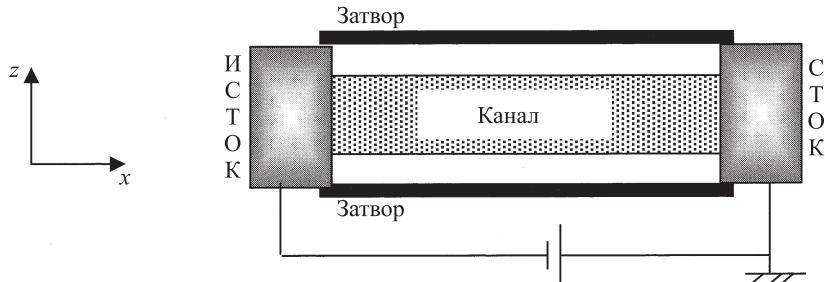


Рис. У.9.5

где функция  $I_{1/2}(x)$  определена в выражении (7.2.22).

**У.9.6.** В параграфе 9.3 мы получили выражения для токов в контактах, рассматривая полный входящий и исходящий из канала токи (уравнения (9.1.7)–(9.1.9)). Такой подход, однако, не дает информацию о детальном распределении тока внутри канала. Коль скоро корреляционная функция  $G^n$  получена из уравнения (9.1.3), ток, протекающий из точки «*а*» в точку «*б*» внутри канала, может быть найден из следующего выражения:

$$I_{a \rightarrow b} = (-q/h) \int_{-\infty}^{+\infty} dE \tilde{I}_{a \rightarrow b}(E),$$

где  $\tilde{I}_{a \rightarrow b}(E) = i[H_{ab}G_{ab}^n - G_{ab}^n H_{ba}]$ .

Получите это выражение, обобщив формулу для одномерного оператора тока из формулы (6.4.6):

$$J_{a \rightarrow b} = [\Psi_b^+ H_{ba} \Psi_a - \Psi_a^+ H_{ab} \Psi_b]/i\hbar$$

и используя далее формулу (9.2.15b). Кстати, это выражение дает альтернативный способ расчета тока в данном контакте: постройте плоскость, отделяющую этот контакт от остальной части структуры, и просуммируйте  $I_{a \rightarrow b}$  по всем точкам «*а*» и «*б*», лежащим на противоположных сторонах этой плоскости.

**У.9.7.** Для знакомства со многими другими расчетами когерентного транспорта читатель отсыпается к работе [14], где приведены программы, работающие в среде MATLAB, с помощью которых можно строить соответствующие графики.

## Глава 10

# Некогерентный транспорт

В главе 9 обсуждалась квантовомеханическая модель, описывающая *когерентное* протекание электронов через канал. Роль диссипативных процессов, приводящих к сбоям фаз, была ограничена лишь областью контактов, где они поддерживали электроны в состоянии локального равновесия. На практике такие процессы идут и в самом канале, причем их роль возрастает с увеличением длины канала. До возникновения мезоскопической физики роль контактов считалась минимальной, и теория транспорта рассматривала диссипативные процессы только в самом канале. В рамках принятого подхода «от малого к большому» мы выяснили, как следует описывать когерентный транспорт в небольшой структуре, и теперь подошли к обсуждению диссипативных процессов, идущих с нарушением фазы во всей системе.

Подобные процессы возникают в результате взаимодействия одного электрона с резервуаром, включающим фотоны и фононы, а также другие электроны. Существенное различие между когерентным и некогерентным транспортом состоит в том, что последний сопровождается переходами с нарушением фазы волновой функции. При когерентных взаимодействиях среда остается неизменной, а электрон взаимодействует с ней подобно тому, как мячик для настольного тенниса взаимодействует с грузовиком. Он сталкивается и отлетает от него, а движение грузовика остается неизменным. Реально окружающая среда не столь жесткая и неизменная, «как грузовик». При перемещении электрона она приводится в «движение»: в ней рождаются фононы, фотоны и т. д. В общем случае точное описание таких процессов невозможно и поэтому мы будем обсуждать только приближение низшего порядка, иногда называемое самосогласованным борновским приближением. В рамках этого приближения «стандартное» состояние канала с  $\{N_\omega\}$  фотонами или фононами, принадлежащими модам с различными частотами  $\{\omega\}$ , взаимодействует с соседними состояниями, в которых имеется на один фотон или фонон меньше (процесс поглощения), либо на один фотон или фонон больше (испускание), как это схематически показано на рис. 10.1.

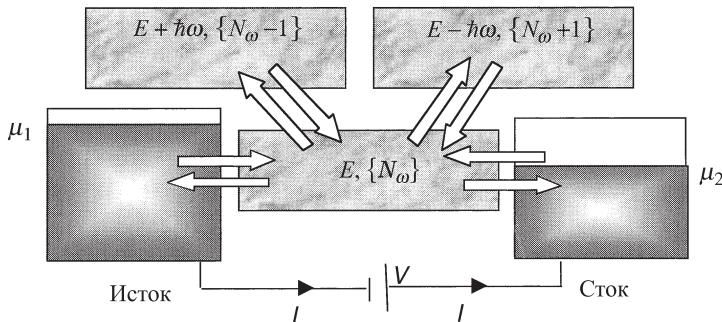


Рис. 10.1. Нарушение фазы происходит при взаимодействии с соседними состояниями, изменяющим число возбуждений  $\{N_\omega\}$  на единицу

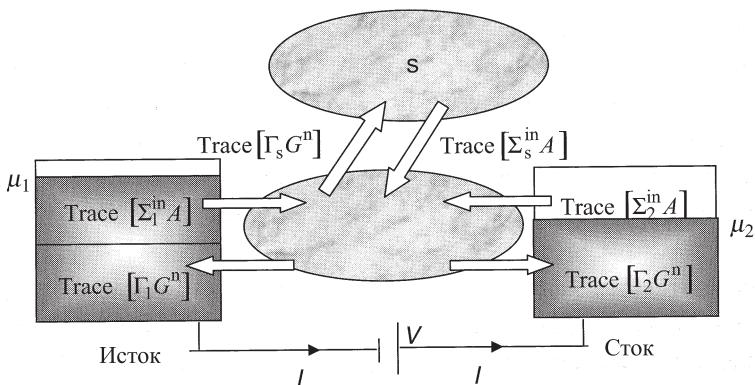


Рис. 10.2. Некогерентный квантовый транспорт: приход и уход

Данное взаимодействие с соседними состояниями вызывает отток электронов из рассматриваемой области и последующий их возврат, или входящий поток. В общей модели квантового когерентного транспорта, развитой в главе 9, учитывались как входящие, так и исходящие потоки. Их учет производится с помощью дополнительного контакта (терминала) «*s*», описываемого слагаемыми  $\Sigma_s^{in}$  и  $\Sigma_s$  (см. рис. 10.2). Моя главная цель в этой главе состоит в том, чтобы объяснить, каким образом можно рассчитать эти дополнительные слагаемые. Нам уже известно, что в обычных контактах рассеяние падающих электронов приводит к уширению:

$$\Sigma_1^{in} = \Gamma_1 f_1 \quad \text{и} \quad \Sigma_2^{in} = \Gamma_2 f_2.$$

Для контакта, моделирующего рассеяние, величины  $\Sigma_s^{in}$  и  $\Sigma_s$  должны быть определены независимым образом, поскольку не существует функции Ферми  $f_s$ , описывающей рассеивающий «электрод». Поэтому нет и простой связи между  $\Sigma_s^{in}$  и  $\Sigma_s$  (или  $\Gamma_s$ ). Можно, конечно, использовать феноменологический подход и рассматривать третий электрод как дополнительный контакт с химическим потенциалом, определяемым из условия обращения тока на этом контакте в нуль. Такой подход по духу соответствовал бы подходу с использованием «зонда Бюттикера», который, как показано в главе 9, оказывается адекватным во многих приложениях. Я, однако, собираюсь обсуждать микроскопические расчеты  $\Sigma_s^{in}$  и  $\Sigma_s$  (или  $\Gamma_s$ ), а не феноменологические. Впоследствии читатель сможет использовать эти микроскопические модели для проверки различных феноменологических моделей, а также в качестве отправной точки при рассмотрении различных механизмов рассеяния.

Понятия входящего и исходящего потоков позволяют описывать более тонкие диссипативные процессы, нежели те, с которыми мы встречались при изучении когерентного транспорта. Полное описание требует применения достаточно сложных методов, описанных в Приложении, но в этой главе я приведу основные результаты, не используя эти методы. В следующей главе, где будут подводиться итоги, мы запишем полную систему уравнений квантового диссипативного транспорта и проиллюстрируем их на нескольких интересных примерах.

Изложение материала в параграфе 10.1 начнем с объяснения двух подходов, которые используются для описания взаимодействия электрона с его окружением, к примеру, с электромагнитными колебаниями, или фотонами. Один из этих подходов основан на одночастичной картине, в которой рассматривается электрон, находящийся в поле потенциала рассеяния  $U_s$ . Однако мы убедимся в том, что для объяснения известных равновесных свойств этот потенциал  $U_s$  приходится наделять весьма необычными свойствами, исключающими его использование в уравнении Шредингера. Вместо этого мы можем принять другую точку зрения и считать, что электрон и фотоны рассматриваются как одна большая система, описываемая одним «очень длинным» уравнением Шредингера. К сожалению, такое описание оказывается чрезвычайно сложным. В общем случае важно соблюдать баланс между этими подходами, сочетая простоту одночастичной картины со строгостью многочастичного описания. В параграфе 10.2 нам предстоит рассмотреть примеры, иллюстрирующие основные идеи и понятия, используемые в данной главе. Затем в параграфе 10.3 будут получены общие выражения для входящего и исходящего потоков. Следующий далее параграф 10.4 обращен к читателям, которых интересуют расчеты колебаний

решетки, или фононы, распространяющихся в некоторых широко используемых полупроводниках.

## 10.1. Почему атом излучает свет?

В начале книги было сделано замечание о том, что объяснение наблюдавшихся оптических спектров атома водорода стало первым большим успехом квантовой теории, основанной на решении уравнения Шредингера. Было обнаружено, что спектр, излучаемый нагретыми парами атомов водорода, состоит из дискретных линий с частотами  $\omega = 2\pi\nu$ , которые оказались связанными с собственными значениями уравнения Шредингера:  $\hbar\omega = \varepsilon_n - \varepsilon_m$ . Подобное совпадение связано с тем, что электрон, находящийся в возбужденном состоянии  $|2\rangle$ , может релаксировать в основное состояние  $|1\rangle$ , так, что разность энергий излучается в виде порций (квантов) света, или фотонов (рис. 10.1.1). Интересно, что описанное поведение системы на самом деле не следует из уравнения Шредингера, до тех пор пока мы не добавим к нему кое-что еще.

Чтобы продемонстрировать это, запишем нестационарное уравнение Шредингера (уравнение (2.1.8)) как матричное уравнение

$$i\hbar \frac{d}{dt} \{\psi\} = [H] \{\psi\}. \quad (10.1.1)$$

Если в качестве базисных выбраны базисные функции гамильтониана  $[H]$ , то это уравнение примет следующий вид:

$$i\hbar \frac{d}{dt} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \dots \end{pmatrix} = \begin{pmatrix} \varepsilon_1 & 0 & 0 & \dots \\ 0 & \varepsilon_2 & 0 & \dots \\ 0 & 0 & \varepsilon_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \dots \end{pmatrix}.$$

Данное матричное уравнение распадается на совокупность независимых уравнений:

$$i\hbar \frac{d}{dt} \{\psi_n\} = [\varepsilon_n] \{\psi_n\}, \quad (10.1.2)$$

по одному на каждое собственное значение энергии  $\varepsilon_n$ . Решение каждого уравнения (10.1.2) легко находится, если известен набор начальных (при  $t = 0$ ) условий:

$$\psi_n(t) = \psi_n(0) \exp(-i\varepsilon_n t/\hbar). \quad (10.1.3a)$$

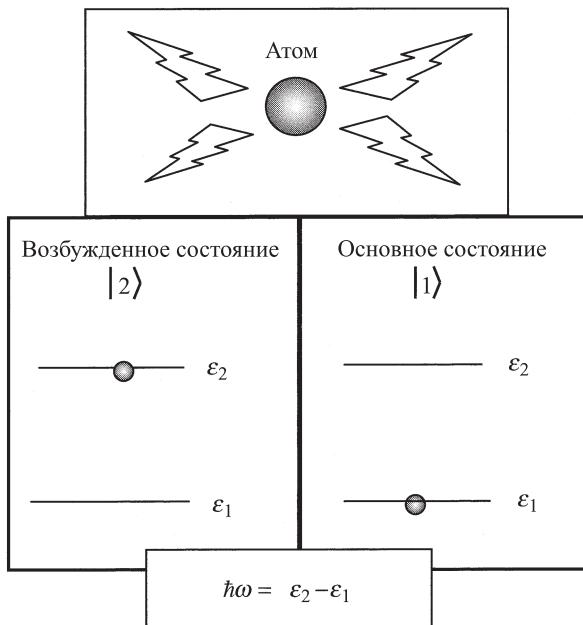


Рис. 10.1.1. Если электрон находится в возбужденном состоянии  $|2\rangle$ , он будет терять энергию, излучая свет и релаксируя в основное состояние  $|1\rangle$ . Это поведение, однако, не следует из уравнения Шредингера до тех пор, пока мы не модифицируем его соответствующим образом

Последнее означает, что вероятность  $P_n$  обнаружить электрон в состоянии  $n$  не изменяется со временем:

$$P_n(t) = |\psi_n(t)|^2 = |\psi_n(0)|^2 = P_n(0). \quad (10.1.3b)$$

Получается, что в соответствии с уравнением Шредингера электрон, помещенный в возбужденное состояние, будет находиться там вечно! Каков бы ни был механизм релаксации из возбужденного состояния в основное, он определенно не описывается уравнением (10.1.1) или (10.1.2).

Как же решается эта проблема? На этот вопрос можно дать два варианта ответа, и ниже мы рассмотрим их последовательно.

**Одиночастичный подход.** В рамках этого подхода рассматривают модель, в которой электрон ощущает воздействие случайного внешнего потен-

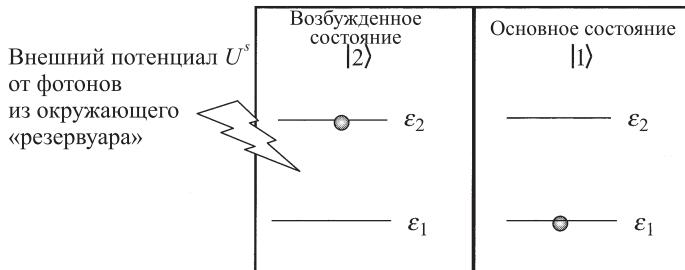


Рис. 10.1.2. При одночастичном подходе говорят, что электрон чувствует внешний потенциал  $U^s$  фотонов в окружавшем систему «резервуаре», который и вызывает релаксацию из состояния  $|2\rangle$  в состояние  $|1\rangle$

циала  $U^s$ , обусловленного фотонами из окружающего систему «резервуара». Этот случайный потенциал и вызывает релаксацию в основное состояние (рис. 10.1.2). Этот потенциал приводит к появлению недиагональных элементов в гамильтониане, связывающих между собой различные состояния. Если рассматриваются лишь два состояния, то следует исходить из уравнения

$$i\hbar \frac{d}{dt} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 & U_{12}^S \\ U_{21}^S & \varepsilon_2 \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \quad (10.1.4)$$

Не вдаваясь в детали, можно утверждать, что если электрон в начальный момент времени находился в состоянии  $|2\rangle$ , то случайный потенциал, которому отвечает матричный элемент  $U_{12}^S$ , будет связывать различные состояния и вызывать переходы в состояние  $|1\rangle$ . Но здесь возникает ряд вопросов. Во-первых, почему при нулевой температуре могут появиться какие-либо внешние потенциалы, когда все тепловые возбуждения выморожены. Ответ обычно заключается в том, что даже при нулевой температуре в окружающей среде присутствует некоторый шум, связанный с т. н. нулевыми флуктуациями электромагнитного поля, которые стимулируют электронную релаксацию из состояния  $|2\rangle$  в состояние  $|1\rangle$ . Но этот ответ порождает второй вопрос: почему эти нулевые флуктуации не приводят также к переходам из состояния  $|1\rangle$  обратно в состояние  $|2\rangle$ ? Каким же образом следует постулировать присутствие рассеивающего потенциала, для которого (учитывая, что  $\varepsilon_2 > \varepsilon_1$ )

$$U_{21}^S = 0, \quad \text{но} \quad U_{1s}^S \neq 0$$

при нулевой температуре.

При конечных температурах матричный элемент  $U_{21}^S$  не равен нулю, но он должен быть все еще значительно меньше, чем  $U_{12}^S$ , для того чтобы интенсивность переходов из состояния 2 в состояние 1 была велика по сравнению с интенсивностью обратных переходов. К примеру, мы можем написать, что

$$S(2 \rightarrow 1) = K_{2 \rightarrow 1} f_2 (1 - f_1) \quad \text{и} \quad S(1 \rightarrow 2) = K_{1 \rightarrow 2} f_1 (1 - f_2),$$

где  $f_1 (1 - f_2)$  есть вероятность того, что система находится в состоянии  $|1\rangle$  (уровень 1 занят, а уровень 2 свободен), и аналогично  $f_2 (1 - f_1)$  есть вероятность того, что система находится в состоянии  $|2\rangle$  (уровень 2 занят, а уровень 1 свободен). В состоянии равновесия обе интенсивности переходов должны быть равны, откуда вытекает требование

$$\frac{K_{1 \rightarrow 2}}{K_{2 \rightarrow 1}} = \frac{f_2 (1 - f_1)}{f_1 (1 - f_2)} = \frac{(1 - f_1)/f_1}{(1 - f_2)/f_2}. \quad (10.1.5)$$

В состоянии равновесия, однако, числа заполнения  $f_1$  и  $f_2$  определяются функцией Ферми:

$$f_n = \frac{1}{1 + \exp[(\varepsilon_n - \mu)/k_B T]} \rightarrow \frac{1 - f_n}{f_n} = \exp\left(\frac{\varepsilon_n - \mu}{k_B T}\right).$$

Как следствие из уравнения (10.1.5) получим

$$\left(\frac{K_{1 \rightarrow 2}}{K_{2 \rightarrow 1}}\right)_{\text{в равновесии}} = \exp\left(-\frac{\varepsilon_2 - \varepsilon_1}{k_B T}\right). \quad (10.1.6)$$

Очевидно, что в равновесии  $K_{2 \rightarrow 1} \gg K_{1 \rightarrow 2}$ , коль скоро разность энергий  $(\varepsilon_2 - \varepsilon_1) \gg k_B T$ .

В начале двадцатого века Эйнштейн высказал предположение о том, что если число фотонов с энергией  $\hbar\omega$  в резервуаре равно  $N$ , то интенсивность переходов с уменьшением энергии пропорциональна числу  $(N + 1)$ , в то время как интенсивность переходов с увеличением энергии пропорциональна  $N$ :

$$\begin{aligned} K(1 \rightarrow 2) &= \alpha N \quad \text{поглощение фотона} \\ K(2 \rightarrow 1) &= \alpha(N + 1) \quad \text{излучение фотона.} \end{aligned} \quad (10.1.7)$$

Это условие обеспечивает выполнение равенства (10.1.6) в состоянии равновесия, поскольку число фотонов определяется распределением Бозе–Эйнштейна

$$[N]_{\text{в равновесии}} = \frac{1}{\exp(\hbar\omega/k_B T) - 1}, \quad (10.1.8)$$

и нетрудно проверить, что

$$\left( \frac{K_{1 \rightarrow 2}}{K_{2 \rightarrow 1}} \right)_{\text{в равновесии}} = \left( \frac{N}{N+1} \right)_{\text{в равновесии}} = \exp \left( -\frac{\hbar\omega}{k_B T} \right). \quad (10.1.9)$$

Поскольку  $\hbar\omega = \varepsilon_2 - \varepsilon_1$ , выражения (10.1.9) и (10.1.6) согласуются между собой.

Остается неясным, почему внешний потенциал должен стимулировать нисходящие ( $2 \rightarrow 1$ ), а не восходящие ( $1 \rightarrow 2$ ) по энергии переходы. Но так должно быть, поскольку в соответствии с функцией Ферми нижележащие состояния всегда характеризуются более высокой вероятностью заполнения, чем вышележащие состояния. Реально не существует однозначной процедуры, позволяющей с использованием внешнего потенциала  $U^S$  дать описание такого эффекта с помощью уравнения Шредингера. Для любого эрмитового оператора  $U^S$  должно выполняться условие  $U_{12}^S = \bar{U}_{21}^S$ , и, следовательно, интенсивности восходящих и нисходящих переходов должны быть равны.

**Многочастичный подход.** Здесь мы переходим к описанию другого подхода, который дает естественное объяснение разницы между интенсивностями переходов, идущих с уменьшением энергии, и переходов, при которых энергия возрастает. Такой подход является принципиально более сложным. Теперь будем рассматривать электрон и фотоны как одну многочастичную систему, чья динамика описывается уравнением, с формальной точки зрения напоминающим уравнение Шредингера (10.1.1):

$$i\hbar \frac{d}{dt} \{\Psi\} = [H] \{\Psi\}. \quad (10.1.10)$$

Теперь, однако,  $\{\Psi\}$  — это вектор состояния в многочастичном гильбертовом пространстве, включающем как электронную, так и фотонную подсистемы. Базисные состояния в этом многочастичном пространстве могут быть записаны как прямые произведения состояний в электронных и фотонных подпространствах (см. рис. 10.1.3):

$$|n, N\rangle = |n\rangle_{\text{электрон}} \otimes |N\rangle_{\text{фотон}},$$

подобно тому, как базисные функции двумерной задачи записываются в виде произведения базисных состояний одномерной задачи:

$$|k_x, k_y\rangle = |k_x\rangle \otimes |k_y\rangle \sim \exp(ik_x x) \exp(ik_y y).$$

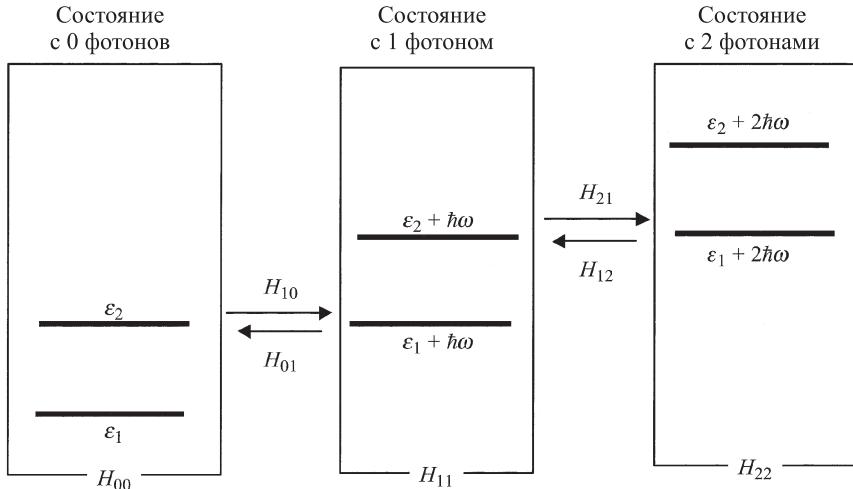


Рис. 10.1.3. В многочастичной картине электрон-фотонное взаимодействие вызывает переходы между состояниями  $|2, N\rangle$  и  $|1, N-1\rangle$ , являющимися вырожденными состояниями составной системы

Запишем (10.1.10) как матричное уравнение:

$$i\hbar \frac{d}{dt} \begin{pmatrix} \Psi_0 \\ \Psi_1 \\ \Psi_2 \end{pmatrix} = \begin{pmatrix} H_{00} & H_{01} & 0 & \dots \\ H_{10} & H_{11} & H_{12} & \dots \\ 0 & H_{21} & H_{22} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \Psi_0 \\ \Psi_1 \\ \Psi_2 \end{pmatrix},$$

где  $\{\Psi_N\}$  есть  $N$ -фотонная компонента волновой функции. Если электронное подпространство включает два состояния  $|1\rangle$  и  $|2\rangle$ , как показано на рис. 10.1.3, то  $\{\Psi_N\}$  есть вектор-столбец с двумя компонентами

$$\{\Psi_N\} = \begin{pmatrix} \psi_{1,N} \\ \psi_{2,N} \end{pmatrix},$$

а каждая из матриц  $H_{NM}$  имеет размер  $(2 \times 2)$ :

$$\begin{aligned} H_{NN} &= \begin{pmatrix} \varepsilon_1 + N\hbar\omega & 0 \\ 0 & \varepsilon_2 + N\hbar\omega \end{pmatrix}, \\ H_{N,N+1} &= \begin{pmatrix} 0 & 0 \\ K^* \sqrt{N+1} & 0 \end{pmatrix}, \end{aligned} \tag{10.1.11}$$

где выполнено условие  $H_{N+1,N} = H_{N,N+1}^+$ .

**Уширение.** Сделаем следующее утверждение:  $N$ -фотонное подпространство можно рассматривать как открытую систему, взаимодействующую с  $(N + 1)$ - и  $(N - 1)$ -фотонными подпространствами, подобно тому как проводящая структура взаимодействует с контактами истока и стока. В главе 8 мы видели, что эффекты взаимодействия с контактами истока и стока могут быть описаны матрицей собственно-энергетической функции (см. уравнение (8.1.11)):

$$\Sigma = \tau g \tau^+,$$

чья мнимая (точнее, антиэрмитова часть) описывает уширение

$$\Gamma \equiv i[\Sigma - \Sigma^+] = \tau a \tau^+,$$

где  $a \equiv i[g - g^+]$  — это спектральная функция изолированного резервуара. Аналогичное соотношение можно было бы использовать и для расчета собственно-энергетической функции, описывающей влияние остальной части фотонного резервуара на наше  $N$ -фотонное подпространство. Фактические детали этой модели являются более сложными, поскольку, в отличие от случая когерентных взаимодействий, нам приходится учитывать принцип запрета. В данный момент, однако, мы вычислим уширение (или исходящий поток), предполагая, что все прочие состояния являются «пустыми», так что не нужно беспокоиться о выполнении принципа запрета. Кроме того, для простоты сосредоточимся лишь на диагональном элементе матрицы уширения:

$$\begin{aligned} [\Gamma_{nn}(E)]_{N,N} &= [H_{nm}]_{N,N+1}[a_{mm}]_{N+1,N+1}[H_{mn}]_{N+1,N} + \\ &+ [H_{nm}]_{N,N-1}[a_{mm}]_{N-1,N-1}[H_{mn}]_{N-1,N}. \end{aligned}$$

Предполагая, что взаимодействие одного фотонного подпространства с другим является слабым, будем аппроксимировать спектральные функции а их невозмущенными значениями:

$$\begin{aligned} [\Gamma_{nn}(E)]_{N,N} &= |K_{mn}^{em}|^2(N+1)2\pi\delta[E - \varepsilon_m - (N+1)\hbar\omega] + \\ &+ |K_{mn}^{ab}|^2N2\pi\delta[E - \varepsilon_m - (N-1)\hbar\omega], \end{aligned}$$

где

$$K_{mn}^{em} \equiv [H_{mn}]_{N+1,N}/\sqrt{N+1} \quad (10.1.12a)$$

и

$$K_{mn}^{ab} \equiv [H_{mn}]_{N-1,N}/\sqrt{N}. \quad (10.1.12b)$$

Снова предположим, что при слабом взаимодействии между фотонными подпространствами состояние  $|n, N\rangle$  с некоторой точностью остается собственным состоянием с энергией  $\varepsilon_n + N\hbar\omega$ , и поэтому уширение можно рассчитать при энергии  $E = \varepsilon_n + N\hbar\omega$ :

$$\begin{aligned}\Gamma_{nn} = & 2\pi|K_{mn}^{em}|^2(N+1)\delta[\varepsilon_n - \varepsilon_m - \hbar\omega] \quad \text{испускание} \\ & + 2\pi|K_{mn}^{ab}|^2N\delta[\varepsilon_n - \varepsilon_m + \hbar\omega] \quad \text{поглощение.}\end{aligned}\quad (10.1.13)$$

Первое слагаемого здесь связано со взаимодействием  $N$ -фотонного подпространства с  $(N+1)$ -фотонным подпространством, т. е. оно ответственно за процесс *испускания фотона*. В самом деле, оно пропорционально дельта-функции, которая обращается в бесконечность при энергии фотона  $\hbar\omega$ , удовлетворяющей уравнению  $\varepsilon_n - \hbar\omega = \varepsilon_m$ . Таким образом, первое слагаемое описывает процесс, в котором электрон переходит из состояния  $n$  в состояние  $m$ , и излучает избыток энергии в виде фотона. Второе слагаемое в выражении (10.1.13) связано с взаимодействием  $N$ -фотонного подпространства с  $(N-1)$ -подпространством, т. е. описывает процесс *поглощения фотона*. Действительно, соответствующий дельта-образный максимум находится при энергии фотона  $\hbar\omega$ , определяемой условием  $\varepsilon_n + \hbar\omega = \varepsilon_m$ . Ясно, что это переход электрона из состояния  $n$  в состояние  $m$ , сопровождающийся поглощением фотона.

**Константы связи.** Как определить константу связи  $K$  в уравнении (10.1.13)? Вернемся для этого вновь к одноэлектронной точке зрения (см. рис. 10.1.2). Нам необходимо записать «потенциал»  $U_S$ , действующий на электрон и обусловленный фотоном или фононами, связанными с конкретными модами с частотой  $\omega$ :

$$U_S(\vec{r}, t) = U^{ab}(\vec{r}) \exp(-i\omega t) + U^{em}(\vec{r}) \exp(+i\omega t), \quad (10.1.14)$$

где  $U^{ab}(\vec{r}) = U^{em}(\vec{r})^*$ .

После того как мы определили этот «потенциал взаимодействия», константы связи для процессов испускания и поглощения, связанные с матричными элементами  $U^{em}$  и  $U^{ab}$ , будут найдены без труда:

$$K_{mn}^{em} = \int dr \phi_m^*(r) U^{em} \phi_n(r) \equiv \langle |U^{em}| \rangle$$

и

$$K_{mn}^{ab} = \int dr \phi_m^*(r) U^{ab} \phi_n(r) \equiv \langle |U^{ab}| \rangle, \quad (10.1.15)$$

где  $\phi_m$  и  $\phi_n$  есть волновые функции состояний  $m$  и  $n$  соответственно.

**Электрон-фононное взаимодействие.** В параграфе 10.4 будет подробно рассматриваться физический смысл понятия квантов упругих колебаний или фононов. Сейчас же достаточно просто связать их с колебаниями атомной решетки, подобно тому, как фотоны сопоставляются электромагнитным колебаниям. Чтобы записать потенциал взаимодействия фононов с электронами, необходимо смещение атомов (или напряжение), связанное с одним фононом в mode с частотой  $\omega$ , умножить на величину изменения электронной энергии  $D$ , приходящуюся на единичное смещение (или напряжение). Величина  $D$ , называемая деформационным потенциалом, известна из экспериментов для большинства объемных материалов. Следует отметить, что известно сравнительно небольшое число работ, посвященных изучению фононных мод вnanoструктурах. Поэтому при описании упругих колебаний в образцах очень малых размеров используют модели плоских волн с волновым вектором  $\vec{\beta}$  и берут значения деформационного потенциала в объемных материалах. Присутствие продольно-поляризованного фона в плоской волне приводит к появлению напряжений (здесь  $\rho$  обозначает массовую плотность, а  $\Omega$  есть нормировочный объем)

$$S = \beta \sqrt{2\hbar/\rho\omega\Omega} \cos(\vec{\beta} \cdot \vec{r} - \omega(\beta)t). \quad (10.1.16)$$

Отсюда ясно, что потенциалы взаимодействия в выражении (10.1.14) равны

$$U_{\vec{\beta}}^{ab}(\vec{r}) \equiv (U_{\vec{\beta}}/2) \exp(i\vec{\beta} \cdot \vec{r}) \quad \text{и} \quad U_{\vec{\beta}}^{em}(\vec{r}) = U_{\vec{\beta}}^{ab}(\vec{r})^*, \quad (10.1.17)$$

где

$$U_{\vec{\beta}} = D\beta \sqrt{2\hbar/\rho\omega\Omega}.$$

**Взаимодействие электронов с фотонами.** Чтобы найти константу электрон-фотонного взаимодействия, необходимо, прежде всего, записать потенциал, который действует на электрон в присутствии одного фотона, принадлежащего конкретной mode. Детали расчета в этом случае усложняются, поскольку электромагнитное поле представлено в уравнении Шредингера векторным, а не скалярным потенциалом.

Прежде всего, запишем выражение для электрического поля одиночного фотона, принадлежащего mode  $(\vec{\beta}, \vec{\nu})$ :

$$\vec{E} = \vec{\nu} E_0 \sin(\vec{\beta} \cdot \vec{r} - \omega(\beta)t),$$

чья амплитуда  $E_0$  определяется из условия равенства энергии фотона величине  $\hbar\omega$ . Здесь  $\Omega$  обозначает нормировочный объем «резервуара»:

$$\varepsilon E_0^2 \Omega / 2 = \hbar\omega \rightarrow |E_0| = \sqrt{2\hbar\omega/\varepsilon\Omega}.$$

Соответствующий векторный электромагнитный потенциал  $\vec{A}$ , связанный с вектором напряженности электрического поля соотношением  $\vec{E} = -\partial \vec{A}/\partial t$ , записывается в виде

$$\vec{A} = \vec{\nu} A_0 \cos(\vec{\beta} \cdot \vec{r} - \omega(\beta)t), \quad \text{где } |A_0| = \sqrt{2\hbar/\omega\varepsilon\Omega}. \quad (10.1.18)$$

Далее представим векторный потенциал одного фотона в виде двух слагаемых (сравните с выражением (10.1.14)):

$$\vec{A}(\vec{r}, t) = \vec{A}^{ab}(\vec{r}) \exp(-i\omega(\beta)t) + \vec{A}^{em}(\vec{r}) \exp(+i\omega(\beta)t), \quad (10.1.19)$$

где  $\vec{A}^{ab}(\vec{r}) \equiv \vec{\nu}(A_0/2) \exp(i\vec{\beta} \cdot \vec{r})$  и  $\vec{A}^{em}(\vec{r}) = \vec{A}^{ab}(\vec{r})^*$ .

Константа взаимодействия для процесса поглощения определяется матричным элементом оператора  $(q/m)\vec{A}^{ab}(\vec{r}) \cdot \vec{p}$ , в то время как для процесса испускания эта константа равна матричному элементу оператора  $(q/m)\vec{A}^{em}(\vec{r}) \cdot \vec{p}$ . Отсюда, учитывая, что  $\vec{p} \equiv -i\hbar\vec{\nabla}$ , получим

$$K_{mn}(\vec{\beta}, \vec{\nu}) = (qA_0/2m) \langle m | \exp(i\vec{\beta} \cdot \vec{r}) \vec{p} \cdot \vec{\nu} | \rangle \quad \text{Поглощение}, \quad (10.1.20a)$$

$$K_{mn}(\vec{\beta}, \vec{\nu}) = (qA_0/2m) \langle m | \exp(-i\vec{\beta} \cdot \vec{r}) \vec{p} \cdot \vec{\nu} | \rangle \quad \text{Испускание}. \quad (10.1.20b)$$

Обратите внимание на то, что символ  $m$  обозначает массу, и его не следует путать с индексом  $m$ , нумерующим базисные функции.

Выражения (10.1.20a), (10.1.20b) требуют более подробного обоснования, поскольку до сих пор мы не имели дела с векторным потенциалом. Мы знаем, что скалярный потенциал  $\phi(\vec{r})$  входит в уравнение Шредингера аддитивным образом:

$$p^2/2m \rightarrow (p^2/2m) - q\phi(\vec{r}).$$

Если бы фотон описывался скалярным потенциалом, то константы взаимодействия легко определялись через матричные элементы операторов  $(-q)\phi^{ab}(\vec{r})$  и  $(-q)\phi^{em}(\vec{r})$  процессов поглощения и испускания. Именно так описывалось электрон-фононное взаимодействие. Фотоны, однако, описываются векторным потенциалом, который входит в уравнение Шредингера через подстановку

$$p^2/2m \rightarrow (\vec{p} + q\vec{A}) \cdot (\vec{p} + q\vec{A})/2m,$$

поэтому изменение гамильтониана в присутствии фотона равно

$$(q/2m)(\vec{A} \cdot \vec{p} + \vec{p} \cdot \vec{A}) + (q^2/2m)\vec{A} \cdot \vec{A} \approx (q/2m)(\vec{A} \cdot \vec{p} + \vec{p} \cdot \vec{A}),$$

где предполагается, что векторный потенциал достаточно мал и можно пренебречь квадратичным слагаемым. Наконец, заметим, что для любой скалярной функции  $\phi(\vec{r})$

$$\vec{p} \cdot (\vec{A}\phi) = \vec{A} \cdot (\vec{p}\phi) + \phi(\vec{p} \cdot \vec{A}),$$

поэтому мы можем записать

$$(q/2m)(\vec{A} \cdot \vec{p} + \vec{p} \cdot \vec{A}) = (q/m)\vec{A} \cdot \vec{p},$$

которое скоро выполнено условие  $\vec{p} \cdot \vec{A} = 0$ . Можно убедиться в том, что для фотонного векторного потенциала в уравнении (10.1.19) это действительно имеет место вследствие поперечного характера электромагнитных волн, у которых волновой вектор  $\vec{\beta}$  и вектор поляризации  $\vec{\nu}$  ортогональны:  $\vec{\beta} \cdot \vec{\nu} = 0$ . Это обстоятельство позволяет выражать константу связи через матричный элемент оператора  $(q/m)\vec{A}(\vec{r}) \cdot \vec{p}$ , используя подстановку  $\vec{A} \rightarrow \vec{A}^{ab}(\vec{r})$  для процесса поглощения и подстановку  $\vec{A} \rightarrow \vec{A}^{em}(\vec{r})$  для процесса испускания.

## 10.2. Примеры

В этом параграфе мы рассмотрим несколько примеров, иллюстрирующих основные идеи, используемые при описании некогерентных взаимодействий. Я рассмотрю общий случай входящих и исходящих потоков в параграфе 10.3. В этом параграфе я буду предполагать, что все прочие состояния являются «свободными» и, следовательно, не нужно беспокоиться о выполнении принципа запрета. Уширение (или исходящий поток) будет рассчитываться по формуле  $\hbar/\tau$ , где  $\tau$  есть время жизни данного состояния. В общем случае под этой величиной можно понимать: (1) время жизни относительно атомных переходов с излучением; (2) время жизни для межзонных излучательных переходов в полупроводниках; (3) фононное время жизни для безызлучательных внутризонных переходов в полупроводниках (рис. 10.2.1). Чтобы рассчитать все эти времена, необходимо записать потенциал взаимодействия (см. выражение (10.1.14)), вычислить константы связи (см. выражение (10.1.15)) и далее с помощью уравнения (10.1.13) найти уширение, а значит, и время жизни.

### 10.2.1. Атомные переходы

Из выражения (10.1.13) очевидно, что уширение будет большим, если аргумент дельта-функции стремится к нулю. Насколько оно будет велико,

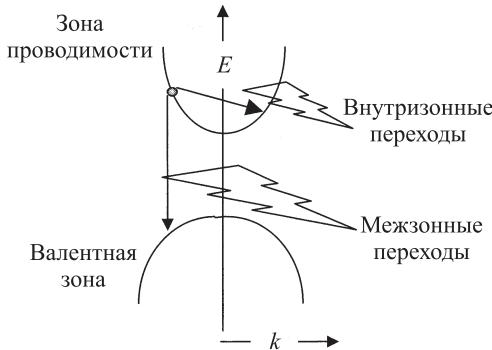


Рис. 10.2.1. Электронные переходы в полупроводниках могут быть как межзонными, так и внутризонными. Первые связаны, в основном, с электрон-фотонным, а вторые — с электрон-фононным взаимодействием

зависит от малого параметра  $0^+$  (см. параграф 8.4), который вводится для описания уширения каждого состояния резервуара. Как было установлено в главе 8, точное значение параметра  $0^+$  обычно не имеет значения, коль скоро система взаимодействует с непрерывно распределенными состояниями резервуара. Подобная ситуация имеет место и в данном случае, поскольку существует много фотонов с различными волновыми векторами  $\vec{\beta}$  и энергиями, определяемыми равенством

$$\hbar\omega(\vec{\beta}) = \hbar\bar{c}\beta, \quad (10.2.1)$$

где  $\bar{c}$  есть скорость света в твердом теле. Как следствие, состояния в конкретном подпространстве не являются дискретными, как показано на рис. 10.1.3, но скорее напоминают состояния, приведенные на рис. 10.2.2.



Рис. 10.2.2

Величина уширения находится из выражения (10.1.13) после суммирования по всем модам  $\vec{\beta}$  и по двум поляризациям  $\vec{\nu}$ , отвечающим каждой

моде  $\vec{\beta}$ :

$$\begin{aligned}\Gamma_{nn} = & \sum_{\vec{\beta}, \vec{\nu}} 2\pi |K_{mn}(\vec{\beta}, \vec{\nu})|^2 (N_{\vec{\beta}, \vec{\nu}} + 1) \delta(\varepsilon_n - \varepsilon_m - \hbar\omega(\beta)) + \\ & + \sum_{\vec{\beta}, \vec{\nu}} 2\pi |K_{mn}(\vec{\beta}, \vec{\nu})|^2 N_{\vec{\beta}, \vec{\nu}} \delta(\varepsilon_n - \varepsilon_m + \hbar\omega(\beta)).\end{aligned}\quad (10.2.2)$$

Если фотонный резервуар находится в состоянии равновесия, то число фотонов в моде  $\vec{\beta}$  и с поляризацией  $\vec{\nu}$  определяется функцией распределения Бозе-Эйнштейна (выражение (10.1.8)):

$$N_{\vec{\beta}, \vec{\nu}} = \frac{1}{\exp(\hbar\omega(\beta)/k_B T) - 1}. \quad (10.2.3)$$

Если рассматриваются переходы с энергиями, значительно превосходящими тепловую энергию  $k_B T$ , то число  $N_{\vec{\beta}, \vec{\nu}}$  можно приравнять к нулю, так что уширение, пропорциональное обратному излучательному времени жизни  $\tau_\Gamma$ , определится выражением

$$\Gamma_{nn} = \left( \frac{\hbar}{\tau_\Gamma} \right) = \sum_{\vec{\beta}, \vec{\nu}} |K_{mn}(\vec{\beta}, \vec{\nu})|^2 \delta(\varepsilon_n - \varepsilon_m - \hbar\omega(\beta)). \quad (10.2.4a)$$

Далее предполагая выполнение периодических граничных условий для фотонных мод, необходимо перейти от суммирования к интегрированию. Мы следуем здесь тому же рецепту, который использовали ранее в расчетах электронных состояний:

$$\sum_{\vec{\beta}} \rightarrow \frac{\Omega}{8\pi^3} \int_0^\infty d\beta \beta^2 \int_0^{+\pi} d\theta \sin \theta \int_0^{2\pi} d\phi,$$

где опять  $\Omega$  обозначает нормировочный объем «резервуара». Выполняя интегрирование, находим

$$\Gamma_{nn} = \frac{\Omega}{8\pi^3} \sum_{\vec{\nu}} \int_0^\infty d\beta \beta^2 \int_0^{+\pi} d\theta \sin \theta \int_0^{2\pi} d\phi 2\pi |K_{mn}(\vec{\beta}, \vec{\nu})|^2 \delta(\varepsilon_n - \varepsilon_m - \hbar\omega(\beta)). \quad (10.2.4b)$$

Чтобы продвинуться далее, необходимо подставить сюда константы  $K$  электрон-фотонного взаимодействия из выражений (10.1.20a), (10.1.20b).

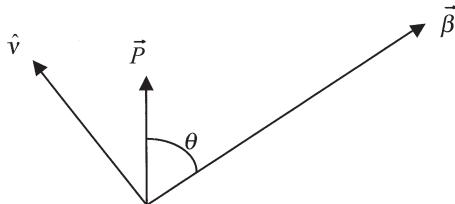


Рис. 10.2.3. Испускание фотона с волновым вектором  $\vec{\beta}$  и поляризацией  $\vec{v}$  в процессе внутриатомного перехода, характеризующегося эквивалентным дипольным моментом  $\vec{P}$

При работе с атомными волновыми функциями, которые локализованы на малых масштабах (гораздо меньших, чем оптическая длина волны), можно пренебречь зависимостью от координаты в множителе  $\exp(i\vec{\beta} \cdot \vec{r})$ , и с помощью выражений (10.1.20a), (10.1.20b) записать

$$K_{mn}(\vec{\beta}, \vec{v}) = \frac{q}{m} \sqrt{\frac{\hbar}{2\varepsilon\omega\Omega}} P \sin \theta, \quad (10.2.5)$$

где  $\vec{P} \equiv \langle m | \vec{p} | n \rangle$  и угол  $\theta$  равен дополнительному углу к углу между вектором дипольного момента и вектором поляризации фотона (см. рис. 10.2.3).

Используя выражения (10.2.5) и (10.2.1), можно найти время жизни до момента испускания фотона из уравнения (10.2.4b):

$$\Gamma = \frac{\Omega}{8\pi^3} \int_0^\infty \frac{d\omega\omega^2}{\bar{c}^3} \int_0^{+\pi} d\theta \sin^3 \theta \int_0^{2\pi} d\phi \frac{q^2\hbar}{2m^2\varepsilon\omega\Omega} P^2 2\pi\delta(\varepsilon_n - \varepsilon_m - \hbar\omega),$$

откуда

$$\frac{1}{\tau_r} = \frac{\Gamma}{\tau} = \frac{q^2}{4\pi\varepsilon\hbar\bar{c}} \frac{2(\varepsilon_n - \varepsilon_m)}{3\hbar m\bar{c}^2} \frac{2P^2}{m}. \quad (10.2.6)$$

Обратите внимание на то, что ответ был получен без какого-либо учета точного значения высоты пика дельта-функции (который, как обсуждалось в параграфе 8.4, определяется параметром  $0^+$ ). Однако, если спектр фотонных мод не является квазинепрерывным, что может иметь место, например, в небольшихnanoструктурах, то вполне возможно, что ответ будет зависеть от величины параметра  $0^+$ .

**Аналогия с классической дипольной антенной.** Мощность, излучаемая одним электроном, определяется из следующего соотношения (имея

в виду, что  $\hbar\omega = \varepsilon_n - \varepsilon_m$ ):

$$W = \frac{\hbar\omega}{\tau} = \frac{q^2}{4\pi\varepsilon\hbar c} \frac{2(\varepsilon_n - \varepsilon_m)}{3\hbar m c^2} \frac{2P^2}{m} = \frac{\omega^2}{12\pi\varepsilon\bar{c}^3} \left( \frac{2qP}{m} \right)^3. \quad (10.2.7)$$

Отметим, что мощность, излучаемая классической дипольной антенной длины  $d$ , по которой протекает ток  $I \cos \omega t$ , определяется выражением

$$W = \frac{\omega^2}{12\pi\varepsilon\bar{c}^3} (Id)^2,$$

что говорит об аналогии между атомной излучающей системой и классическим диполем

$$\vec{Id}_{\text{(антенна)}} = 2q\vec{P}/m_{\text{(атомный излучатель)}} = (2q/m)\langle m|\vec{p}|n\rangle. \quad (10.2.8)$$

В самом деле, не только полная мощность излучения, но его поляризация и угловое распределение излучения выглядят одинаково для классической дипольной антенны и атомной излучающей системы. Излучаемый свет поляризован в той плоскости, где лежит вектор, указывающий направление на точку наблюдения и вектор  $\vec{P}$ , а его интенсивность пропорциональна  $\sin^2 \theta$ , где  $\theta$  есть угол между направлением распространения и дипольным моментом, как это показано на рис. 10.2.3.

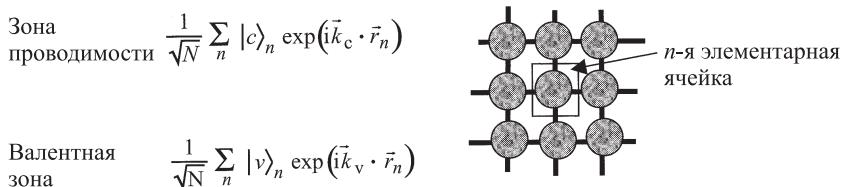


Рис. 10.2.4

### 10.2.2. Межзонные переходы в полупроводниках

Для определения константы связи с делокализованными электронными состояниями можно использовать выражения (10.1.20a), (10.1.20b), но теперь нельзя, как при выводе уравнения (10.2.5), пренебречь пространственной зависимостью множителя  $\exp(i\vec{\beta} \cdot \vec{r})$ . К примеру, в полупроводниках (см. рис. 10.2.4) состояния зоны проводимости (с) и валентной зоны (v)

обычно распределены по всему объему образца, состоящему из большого числа элементарных ячеек, как это изображено на рис. 10.2.4. Здесь  $|c\rangle_n$  и  $|v\rangle_n$  это атомные части волновой функции в  $n$ -й ячейке, относящиеся к зоне проводимости или валентной зоне. Эти функции зависят от волнового вектора  $\vec{k}_c$  или  $\vec{k}_v$ , но остаются неизменными в любой элементарной ячейке, испытывая лишь пространственное смещение. Это позволяет, исходя из (10.1.20a), (10.1.20b), записать матричные элементы, описывающие поглощение и испускание, в виде

$$\langle \nu | \vec{p} \cdot \vec{\nu} | c \rangle \sum_n \frac{1}{N} \exp[i(\vec{k}_c \pm \vec{\beta} - \vec{k}_v) \cdot \vec{r}_n],$$

где верхний знак (+) отвечает поглощению, нижний знак (-) — испусканию. Угловые скобки  $\langle \dots \rangle$  обозначают интегрирование по элементарной ячейке, причем на ее масштабе можно пренебречь пространственным изменением множителя  $\exp(i\vec{\beta} \cdot \vec{r})$ . В результате приходим к выводу, что данное выражение не равно нулю лишь при выполнении условия

$$\vec{k}_v = \vec{k}_c \pm \vec{\beta}. \quad (10.2.9)$$

Условие (10.2.9) может рассматриваться как закон сохранения импульса, если помнить о том, что  $\hbar\vec{k}$  есть импульс электрона, а  $\hbar\vec{\beta}$  есть импульс фотона. Конечный импульс электрона  $\hbar\vec{k}_v$  равен начальному импульсу  $\hbar\vec{k}_c$  плюс или минус импульс фотона  $\hbar\vec{\beta}$ , в зависимости от того, поглощается фотон или излучается. Волновой вектор фотона обычно очень мал по сравнению с электронным волновым вектором, и поэтому излучательные переходы являются почти «вертикальными» с конечным волновым вектором  $\vec{k}_v = \vec{k}_c \pm \vec{\beta} \approx \vec{k}_c$ . Это нетрудно увидеть, если вспомнить, что область значений  $k$  включает в себя зону Бриллюэна, размер которой имеет характерный масштаб  $\sim 2\pi$ , деленный на постоянную атомной решетки, в то время как волновой вектор фотона равен величине  $2\pi$ , деленной на оптическую длину волны, которая в тысячи раз превышает атомный размер.

Предполагая, что выполнен закон сохранения импульса, можно, воспользовавшись (10.1.20a), (10.1.20b), записать константы связи как

$$K_{mn}(\vec{\beta}, \vec{\nu}) = \frac{q}{m} \sqrt{\frac{\hbar}{2\varepsilon\omega\Omega}} P \sin \theta, \quad \text{где } \vec{P} \equiv \langle \nu | \vec{p} | c \rangle, \quad (10.2.10)$$

откуда следует, что «вертикальные» излучательные переходы в полупроводниках должны описываться по аналогии с атомными переходами (см. выражение (10.2.5)), но атомные волновые функции следует заменить атомными

функциями зоны проводимости и валентной зоны. К примеру, если подставить численные параметры, характеризующие переход между валентной зоной и зоной проводимости в типичном полупроводнике в выражение (10.2.6), то окажется, что в GaAs, где  $\varepsilon_n - \varepsilon_m = 1.5$  эВ,  $2P^2/m = 20$  эВ,  $\varepsilon = 10\varepsilon_0$ , излучательное время жизни равно  $\tau_r = 0.7$  нс.

**Поляризация и угловое распределение.** Известно, что электронные состояния зоны проводимости, расположенные вблизи  $\Gamma$ -точки, являются изотропными состояниями s-типа и обозначаются как  $|S\rangle$ . Состояния, расположенные вблизи потолка валентной зоны, — это состояния  $p_x$ -,  $p_y$ - или  $p_z$ -типа, обозначаемые как  $|X\rangle$ ,  $|Y\rangle$  или  $|Z\rangle$ . Можно сказать, что такая система электронов подобна системе трех независимых антенн с дипольными моментами, ориентированными вдоль  $x$ ,  $y$  или  $z$ ,

$$\langle s|\vec{p}|X\rangle = \vec{x}P, \quad \langle s|\vec{p}|Y\rangle = \vec{y}P, \quad \langle s|\vec{p}|Z\rangle = \vec{z}P.$$

Можно показать, что полное излучение такой системы будет неполяризованным и изотропным. Реально, однако, состояния потолка валентной зоны имеют более сложную структуру: они включают в себя функции легких и тяжелых дырок, каждая из которых является линейной комбинацией состояний со спином, ориентированным вверх и вниз. В этом случае эквивалентные дипольные моменты каждой пары состояний из зоны проводимости и валентной зоны могут быть записаны в виде, представленном в табл. 10.2.1.

Таблица 10.2.1

$HH$	$\overline{HH}$	$LH$	$\overline{LH}$
$C \begin{Bmatrix}  S\rangle \\ 0 \end{Bmatrix}$	$\frac{\hat{x} + i\hat{y}}{\sqrt{2}}P$	$0$	$-\sqrt{\frac{2}{3}}\hat{z}P$
$\overline{C} \begin{Bmatrix} 0 \\  S\rangle \end{Bmatrix}$	$0$	$\frac{\hat{x} - i\hat{y}}{\sqrt{2}}P$	$\frac{\hat{x} + i\hat{y}}{\sqrt{6}}P$
			$\sqrt{\frac{2}{3}}\hat{z}P$

Таким образом, имеется восемь независимых излучающих «антенн», по одной на соответствующую пару состояний зоны проводимости и валентной зоны. Если состояние  $C$  занято, то первый ряд из четырех типов излучателей (антенн) является активным. В направлении оси  $z$  распространя-

няется излучение, связанное с переходами  $C - HH$  и  $C - \bar{L}H$ . При переходах  $C - HH$  будет излучаться свет с правой круговой поляризацией (ПКП), интенсивность которого в три раза будет превышать интенсивность излучения с левой круговой поляризацией (ЛКП) при переходах  $C - \bar{L}H$ . Если состояние  $\bar{C}$  также занято, то переход  $\bar{C} - \bar{H}H$  с ЛКП будет в три раза более интенсивным, чем переход  $\bar{C} - LH$  с ПКП. В итоге интенсивности излучения с ЛКП и ПКП будут равны. Но если занято только состояние  $C$ , то интенсивность излучения с ПКП будет в три раза выше. На практике степень циркулярной поляризации излучения в условиях данного эксперимента часто используется в качестве меры, определяющей степень спиновой поляризации.

### 10.2.3. Внутризонные переходы в полупроводниках

Мы обсудили поляризацию излучения, связанного с межзонными переходами в полупроводниках (см. рис. 10.2.1). Но как обстоит дело с внутризонными переходами? Могут ли они приводить к излучению света? Мы покажем, что излучение света свободным электроном запрещено законами сохранения импульса и энергии, если скорость электрона не превосходит скорость света:  $\hbar k/m > c$ . Такая ситуация невозможна в вакууме, но может иметь место в твердом теле. Излучение, производимое электронами, движущимися со сверхсветовой скоростью, действительно наблюдалось и получило название излучение Черенкова. Данный эффект, однако, обычно не имеет отношения к работе твердотельных структур, поскольку типичные скорости электронов там обычно имеют порядок одной тысячной от скорости света. Более существенным является черенковское излучение акустических волн, или фононов, скорость которых на пять порядков меньше скорости света. Электронные скорости обычно значительно превосходят скорости фононов, и это приводит к появлению значительного черенковского излучения (и поглощения) фононов, подобно тому как сверхзвуковые самолеты генерируют ударные волны в воздухе.

**Критерии появления излучения Черенкова.** При внутризонных переходах начальное и конечное состояния описываются одной и той же атомной волновой функцией. Для простоты я не буду выписывать ее явно. Вместо этого запишем функции начального и конечного состояний в виде плоских волн, как будто мы имеем дело с электронами в вакууме:

$$|\vec{k}\rangle \equiv (1/\sqrt{\Omega}) \exp(i\vec{k} \cdot \vec{r}) \quad \text{и} \quad |\vec{k}_f\rangle \equiv (1/\sqrt{\Omega}) \exp(i\vec{k}_f \cdot \vec{r}). \quad (10.2.11)$$

Из выражений (10.1.20a), (10.1.20b) найдем, что излучательная кон-

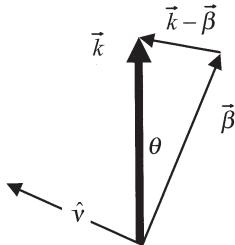


Рис. 10.2.5. Излучение Черенкова: показано начальное состояние  $\vec{k}$  с волновым вектором фотона  $\vec{\beta}$  и конечное состояние  $\vec{k} - \vec{\beta}$ . Направление поляризации фотона обозначено вектором  $\hat{\nu}$

станта связи равна

$$K(\vec{k}_f, \vec{k}, \vec{\beta}, \vec{\nu}) = (qA_0/2m)\hbar\vec{k} \cdot \vec{\nu}, \quad (10.2.12)$$

если выполнены условия

$$\begin{aligned} \vec{k}_f &= \vec{k} - \vec{\beta} && \text{испускание,} \\ \vec{k}_f &= \vec{k} + \vec{\beta} && \text{поглощение,} \end{aligned} \quad (10.2.13)$$

и равна нулю в противном случае. Как и (10.2.9), условия (10.2.13) можно интерпретировать как условия сохранения импульса. Выполнение закона сохранения энергии, свою очередь, обеспечивается наличием дельта-функции в выражении (10.1.13):

$$\begin{aligned} \varepsilon(\vec{k}_f) &= \varepsilon(\vec{k}) - \hbar\omega(\vec{\beta}) && \text{испускание,} \\ \varepsilon(\vec{k}_f) &= \varepsilon(\vec{k}) + \hbar\omega(\vec{\beta}) && \text{поглощение.} \end{aligned} \quad (10.2.14)$$

Из уравнений (10.2.13) и (10.2.14) мы имеем, что для процесса *испускания*

$$\varepsilon(\vec{k}_f) - \varepsilon(\vec{k}) + \hbar\omega(\vec{\beta}) = 0 = \frac{\hbar^2}{2m}([\vec{k} - \vec{\beta}] \cdot [\vec{k} - \vec{\beta}] - k^2) + \hbar c \beta,$$

поэтому

$$\frac{\hbar^2}{2m}(-2k\beta \cos \theta + \beta^2) + \hbar c \beta = 0,$$

откуда следует, что

$$\cos \theta = \frac{\bar{c}}{\hbar k/m} + \frac{\beta}{2k}. \quad (10.2.15)$$

Утверждение состоит в том, что, поскольку  $\cos \theta$  по модулю всегда меньше единицы, излучение Черенкова не может иметь места до тех пор, пока электронная скорость  $\hbar k/m$  не превысит скорость света  $c$ . Как отмечалось ранее, это условие нельзя выполнить в вакууме, но можно в твердом теле, или в жидкости, где черенковское излучение действительно наблюдалось экспериментально. Направление излучаемого света формирует конус вокруг электронного волнового вектора  $\vec{k}$  с максимальным углом  $\theta_{max} = \cos^{-1}(m\bar{c}/\hbar k)$  (см. рис. 10.2.5).

**Черенковское излучение акустических фононов.** Ранее отмечалось, что черенковское излучение света не слишком актуально для физики твердотельных структур. Напротив, испускание и поглощение звуковых волн, т. е. акустических фононов, в ряде задач является определяющим процессом. Скорость акустических волн на пять порядков меньше, чем скорость света, и скорость электронов обычно превышает скорость звука. Поскольку акустические фононы обычно обладают энергией, меньше чем  $k_B T$ , их число в состоянии равновесия обычно значительно:

$$N_{\vec{\beta}} + 1 - N_{\vec{\beta}} = [\exp(\hbar\omega(\beta)/k_B T) - 1]^{-1} \cong k_B T / \hbar\omega(\beta) \quad (\text{акустические фононы}).$$

Теперь оба слагаемых в выражении (10.1.13) дают вклад в уширение или обратное время жизни: черенковское поглощение («*ab*») столь же важно, как и черенковское испускание («*em*»):

$$(\Gamma)_{\vec{k}, em} = \sum_{\vec{\beta}} 2\pi \frac{k_B T}{\hbar\omega(\beta)} |K(\vec{\beta})|^2 \delta[\varepsilon(\vec{k}) - \varepsilon(\vec{k} - \vec{\beta}) - \hbar\omega(\vec{\beta})], \quad (10.2.16a)$$

$$(\Gamma)_{\vec{k}, ab} = \sum_{\vec{\beta}} 2\pi \frac{k_B T}{\hbar\omega(\beta)} |K(\vec{\beta})|^2 \delta[\varepsilon(\vec{k}) - \varepsilon(\vec{k} + \vec{\beta}) + \hbar\omega(\vec{\beta})]. \quad (10.2.16b)$$

Константа связи  $K(\vec{\beta})$  пропорциональна значению потенциала, который чувствует электрон в присутствии одного фонона (см. уравнение (10.1.18)). Не вдаваясь в детали расчета правых частей выражений (10.2.16a), (10.2.16b), угол излучения  $\theta$  нетрудно связать с модулем волнового вектора фонона  $\beta$ , для чего следует приравнять аргумент дельта-функции к нулю и действовать так, как при выводе уравнения (10.2.15):

$$\begin{aligned} \cos \theta &= \frac{c_s}{\hbar k/m} + \frac{\beta}{2k} \quad \text{испускание,} \\ \cos \theta &= \frac{c_s}{\hbar k/m} - \frac{\beta}{2k} \quad \text{поглощение,} \end{aligned} \quad (10.2.17)$$

где  $c_s$  есть скорость звуковых волн и  $\omega = c_s\beta$ . Детальный расчет электронного времени жизни по отношению к процессам испускания и поглощения акустического фонона описан с использованием выражений (10.2.16a), (10.2.16b) в упражнении У.10.1.

**Черенковское излучение оптических фононов.** Кроме акустических фононов в твердых телах присутствуют оптические фононы (см. вспомогательные примечания к параграфу 10.4), чья частота почти постоянна, т. е.  $\omega = \omega_0$ , а энергия  $\hbar\omega_0$  обычно имеет порядок в несколько десятых мэВ, так что число таких фононов при комнатной температуре в равновесном состоянии будет порядка единицы:

$$N_{\vec{\beta}} = [\exp(\hbar\omega_0/k_B T) - 1]^{-1} \equiv N_0 \quad (\text{оптические фононы}).$$

Теперь с помощью выражения (10.1.13) можно получить интенсивности испускания и поглощения:

$$(\Gamma)_{\vec{k},em} = (N_0 + 1) \sum_{\vec{\beta}} 2\pi |K(\vec{\beta})|^2 \delta[\varepsilon(\vec{k}) - \varepsilon(\vec{k} - \vec{\beta}) - \hbar\omega_0], \quad (10.2.18a)$$

$$(\Gamma)_{\vec{k},em} = N_0 \sum_{\vec{\beta}} 2\pi |K(\vec{\beta})|^2 \delta[\varepsilon(\vec{k}) - \varepsilon(\vec{k} + \vec{\beta}) + \hbar\omega_0], \quad (10.2.18b)$$

которые можно рассчитать как описано в упражнении У.10.2.

### 10.3. Входящий и выходящий токи

В этом параграфе будет показано, как развитые в данной главе методы позволяют найти дополнительные слагаемые  $\Sigma_s^{in}$  и  $\Gamma_s$ , фигурирующие на схеме, приведенной на рис. 10.2.

**Дискретные уровни.** Рассмотрим сначала, как формируются входящий и исходящий потоки для электронов, совершающих переходы на данный уровень. Пусть переходы происходят с уровней  $a$  и  $b$ , один из которых расположен выше исходного, а второй — ниже на величину  $\hbar\omega$ :  $\varepsilon_b - \varepsilon = \varepsilon - \varepsilon_a = \hbar\omega$  (см. рис. 10.3.1). В параграфе 10.1 было показано, что интенсивность процессов поглощения ( $K^{ab}$ ) пропорциональна числу фононов ( $N_\omega$ ), в то время как параметр ( $K^{em}$ ), характеризующий процессы испускания, пропорционален ( $N_\omega + 1$ ). Предположим, что температура настолько мала, что можно пренебречь энергией теплового движения по сравнению с энергией  $\hbar\omega$ . Тогда  $N_\omega \ll 1$  и нужно учитывать только процессы испускания.

Используя обозначения  $N$ ,  $N_a$  и  $N_b$  для записи числа электронов на каждом из этих уровней, мы сможем записать для уровня  $\varepsilon$

$$\text{Входящий поток} = K^{em}(1-N)N_b \quad \text{и} \quad \text{Исходящий поток} = K^{em}N(1-N_a).$$

Теперь мы можем записать входящий ток как разность двух членов:

$$\text{Входящий поток} = K^{em}N_b - K^{em}NN_b,$$

где второе слагаемое представляет собой связанное с принципом запрета уменьшение входящего потока. Из более детального рассмотрения следует (см. параграф П.4 Приложения), что эта часть входящего тока на самом деле не блокируется. Скорее следует говорить о том, что на эту величину ( $K^{em}NN_b$ ) возрастает исходящий поток:

$$\text{Входящий ток} = K^{em}N_b \quad \text{и} \quad \text{Выходящий ток} = K^{em}N(1 - N_a + N_b). \quad (10.3.1a)$$

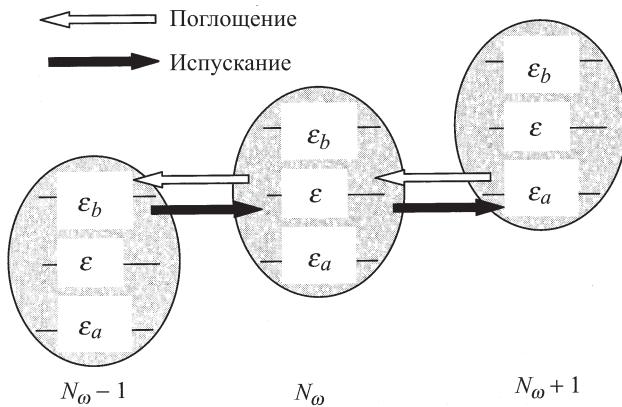


Рис. 10.3.1. Идеализированная структура с тремя уровнями энергии, используемая для описания входящего и исходящего потоков в процессах поглощения и испускания фотонов и фононов

Выражение для разности между входящим и исходящим потоками не изменилось по сравнению с тем, которое мы интуитивно предполагали ранее. Но это различие не является чисто академическим, поскольку именно исходящий ток (а не разность токов) определяет уширение уровня. Мы можем сравнить эту ситуацию с той, что обсуждалась ранее в задаче о входящем и исходящем потоках между контактом и дискретным уровнем (см.

рис. 8.1):

$$\text{Входящий ток} = \gamma f \quad \text{и} \quad \text{Выходящий ток} = \gamma N. \quad (10.3.1b)$$

Здесь выражение для входящего потока напоминает полученное выше, где роль  $K^{em}$  играет  $\gamma$ , но выражение для исходящего потока включает в себя дополнительный множитель  $(1 - N_a + N_b)$ , который сводится к единице лишь при  $N_a = N_b$ , но не в общем случае. Выражению (10.3.1b) ранее был придан следующий вид (см. рис. 9.1.2):

$$\text{Входящий ток} = \text{Tr}[\Gamma A] = \text{Tr}[\Sigma^{in} A] \quad \text{и} \quad \text{Исходящий ток} = \text{Tr}[\Gamma G^n].$$

Аналогично мы можем в данном случае обобщить и выражение (10.3.1a):

$$\text{Входящий ток} = \text{Tr}[\Sigma_s^{in} A] = \text{Tr}[\Sigma^{in} A] \quad \text{и} \quad \text{Выходящий ток} = \text{Tr}[\Gamma_s G^n], \quad (10.3.2)$$

где выражения для  $\Sigma_s^{in}$  и  $\Gamma_s$  могут быть найдены эвристическим путем по аналогии с написанным выше. Ниже мы проделаем этот вывод шаг за шагом.

**Входящий поток для состояний непрерывного спектра.** Влияние рассматриваемых процессов рассеяния подобно действию обычных контактов. Основное различие состоит в том, что обычные контакты поддерживаются в состоянии равновесия внешними источниками, в то время как в процессах рассеяния сама структура играет роль «контакта» (см. рис. 10.3.1), причем ее состояние может быть весьма далеким от состояния равновесия. В главе 8 мы видели, что для обычных контактов функция рассеяния для входящего потока определяется выражением

$$\Sigma^{in}(E) = [\tau A \tau^+] f = [\tau G^n \tau^+],$$

где  $[G^n] = [A]f$  — корреляционная функция контакта, которая описывает рассеяние при данной энергии  $E$  внутрь структуры. В этой ситуации при рассмотрении процесса испускания корреляционная функция структуры при энергии  $E + \hbar\omega$  отвечает рассеянию внутрь структуры при энергии  $E$ , что позволяет записать нам выражение

$$\Sigma_s^{in}(E) = \sum_{\vec{\beta}} (N_{\beta} + 1) \left[ U_{\vec{\beta}}^{em} G^n(E + \hbar\omega(\vec{\beta})) (U_{\vec{\beta}}^{em})^+ \right],$$

в котором  $U^{em}$  есть часть потенциала взаимодействия, отвечающая за испускание (см. выражение (10.1.17)). Записывая умножение матриц более

подробно, находим, что

$$\begin{aligned}\Sigma_s^{in}(i, j; E) &= \sum_{p', q', \vec{\beta}} (N_{\vec{\beta}} + 1) U_{\vec{\beta}}^{em}(i, r) G^n(r, s; E + \hbar\omega(\vec{\beta})) U_{\vec{\beta}}^{em}(j, s) = \\ &= \int_0^\infty \frac{d(\hbar\omega)}{2\pi} \sum_{r, s} D^{em}(i, r; j, s; \hbar\omega) G^n(r, s; E + \hbar\omega),\end{aligned}\quad (10.3.3)$$

где

$$D^{em}(i, r; j, s; \hbar\omega) \equiv \sum_{\vec{\beta}} (N_{\vec{\beta}} + 1) \delta(\hbar\omega - \hbar\omega(\vec{\beta})) U_{\vec{\beta}}^{em}(i, r) U_{\vec{\beta}}^{em}(j, s)^*. \quad (10.3.4)$$

Если учтем и процесс поглощения, будем иметь

$$\Sigma_s^{in}(i, j; E) = \int_0^\infty \frac{d(\hbar\omega)}{2\pi} \left( \begin{array}{l} D^{em}(i, r; j, s; \hbar\omega) G^n(r, s; E + \hbar\omega) \\ + D^{ab}(i, r; j, s; \hbar\omega) G^n(r, s; E - \hbar\omega) \end{array} \right), \quad (10.3.5)$$

где слагаемое, ответственное за поглощение, отличается от эмиссионного слагаемого только заменой  $(N + 1)$  на  $N$ :

$$D^{ab}(i, r; j, s; \hbar\omega) = 2\pi \sum_{\vec{\beta}} N_{\vec{\beta}} \delta(\hbar\omega - \hbar\omega(\vec{\beta})) U_{\vec{\beta}}^{em}(i, r) U_{\vec{\beta}}^{em}(j, s)^*. \quad (10.3.6)$$

В общем случае функции  $D^{em}$  и  $D^{ab}$ , определяющие процессы испускания и поглощения, являются тензорами четвертого ранга, но в данном параграфе мы ограничимся упрощенным изложением, и будем считать, что они являются скалярами. Полное тензорное описание представлено в параграфе П.4 Приложения. Сейчас же мы дадим упрощенные выражения для функций, определяющих испускание и поглощение:

$$D^{em}(\hbar\omega) \equiv (N_{\omega} + 1) D_0(\hbar\omega), \quad (10.3.7a)$$

$$D^{ab}(\hbar\omega) \equiv N_{\omega} D_0(\hbar\omega). \quad (10.3.7b)$$

В этих обозначениях функцию рассеяния можно записать в следующем виде:

$$\Sigma_s^{in}(E) = \int_0^\infty \frac{d(\hbar\omega)}{2\pi} D_0(\hbar\omega) ((N_{\omega} + 1) G^n(E + \hbar\omega) + N_{\omega} G^n(E - \hbar\omega)). \quad (10.3.8)$$

**Исходящий поток для случая непрерывного спектра состояний.** Для того чтобы записать исходящий поток для структуры с непрерывным спектром, нужно в выражении (10.3.1a) заменить  $N_b$  на  $G^n(E + \hbar\omega)$ , а  $1 - N_a$  на  $G^p(E - \hbar\omega)$  и получить

$$\Gamma_s(E) = \int \frac{d(\hbar\omega)}{2\pi} D^{em}(\hbar\omega)[G^p(E - \hbar\omega) + G^n(E + \hbar\omega)],$$

где я определил новую величину  $G^p \equiv A - G^n$ , которая несет информацию о плотности (на один эВ) свободных состояний, или дырок, также как величина  $G^n$  несет информацию о заполненных состояниях, или электронах. Сумма этих величин определяет спектральную функцию, описывающую плотность состояний. Как и ранее (см. выражения (10.3.5)–(10.3.10)), этот результат распространяется на случай, когда присутствуют слагаемые, отвечающие за поглощение. В результате получим

$$\Gamma_s(E) = \int \frac{d(\hbar\omega)}{2\pi} D_0(\hbar\omega) \left( \begin{array}{l} (N_\omega + 1)[G^p(E - \hbar\omega) + G^n(E + \hbar\omega)] \\ + N_\omega [G^n(E - \hbar\omega) + G^p(E + \hbar\omega)] \end{array} \right). \quad (10.3.9)$$

Выражения (10.3.8)–(10.3.9) определяют интересующие нас функции рассеяния. Собственно-энергетическую функцию  $\Sigma_s$  можно записать как  $\text{Re}(\Sigma_s) + i\text{Im}(\Sigma_s)/2$ , где, как отмечалось в параграфе 8.3, вещественная часть может быть получена путем гильбертова преобразования мнимой части, представленной в уравнении (10.3.9).

**«Теорема» Мигдала.** Если распределение электронов описывается равновесной функцией Ферми, то имеем

$$G^n(E') \equiv f(E')A(E') \quad \text{и} \quad G^p(E') \equiv (1 - f(E'))A(E').$$

Отсюда находим

$$\begin{aligned} \Gamma_s(E) &= \int d(\hbar\omega) \int dE' D_0(\hbar\omega) \cdot A(E') \times \\ &\times (N_\omega + 1 - f(E'))\delta(E - E' - \hbar\omega) + (N_\omega + f(E'))\delta(E - E' + \hbar\omega). \end{aligned}$$

Это выражение часто встречается в литературе, посвященной электронно-фононному взаимодействию в металлах, и носит название «теоремы» Мигдала (см., к примеру, [1]<sup>1</sup>). Вблизи состояния равновесия не требуется записывать отдельное выражение для функции  $\Sigma_s^{in}(E)$ , так как подобно выражению для обычного контакта она просто равна  $f(E)\Gamma_s(E)$ .

<sup>1</sup> В российской литературе под теоремой Мигдала понимается несколько иное по форме утверждение: см., например, [54]. — Прим. ред. перев.

**Почти упругие процессы.** Заметьте, что приведенные выражения существенно упрощаются в случае низкоэнергетических процессов рассеяния ( $\hbar\omega \rightarrow 0$ ), когда можно положить  $E + \hbar\omega \approx E \approx E - \hbar\omega$ :

$$\Sigma_s^{in}(E) = \int_0^{\infty} d(\hbar\omega) (D^{em}(\hbar\omega) + D^{ab}(\hbar\omega)) G^n(E), \quad (10.3.10)$$

$$\Sigma_s(E) = \int_0^{\infty} d(\hbar\omega) (D^{em}(\hbar\omega) + D^{ab}(\hbar\omega)) A(E). \quad (10.3.11)$$

В самом деле, в этом случае для вывода вещественной части собственно-энергетической функции  $\Sigma_s$  не требуется проводить преобразование Гильберта

$$\Sigma_s(E) = \int_0^{\infty} d(\hbar\omega) (D^{em}(\hbar\omega) + D^{ab}(\hbar\omega)) G(E), \quad (10.3.12)$$

поскольку  $\text{Re}(\Sigma_s)$  связана с  $\Gamma_s$  также, как  $\text{Re}(G)$  с величиной  $A$ , а именно, через гильбертово преобразование.

Обсуждавшиеся в этом параграфе полные уравнения квантового диссипативного транспорта, включающие слагаемые с входящим или исходящим токами, приведены в главе 11 вместе с иллюстрирующими их примерами.

## 10.4. Дополнение: фононы

Как было упомянуто ранее, фононы — это кванты (порции) колебаний кристаллической решетки, подобно тому, как фотоны есть кванты электромагнитных колебаний. В этом параграфе я попытаюсь обосновать это утверждение. При низких температурах атомы, составляющие молекулу или твердое тело, «заморожены» вблизи узлов кристаллической решетки. С ростом температуры амплитуда колебаний атомов увеличивается, и благодаря «электрон-фононному взаимодействию» ускоряется обмен энергией с электронами. Чтобы понять, как описываются колебания решетки, начнем рассмотрение с простого примера, а именно, с молекулы водорода.

Как отмечалось в главе 3, колебания молекулы можно рассматривать как систему масс и соединенных пружиной (рис. 10.4.1). Масса  $M$  представляет собой массу одного из атомов водорода, а жесткость пружины  $K$



Рис. 10.4.1. Молекула водорода может рассматриваться как состоящая из двух масс, соединенных пружиной

равна второй производной потенциальной энергии как функции расстояния. Из начального курса физики известно, что подобная система ведет себя как осциллятор с частотой  $\omega = \sqrt{K/M}$ . Экспериментаторы измерили эту частоту и установили, что она имеет порядок  $\omega = 2\pi(10^{14}/c)$ . Кстати, зная чему равна масса  $M$ , мы можем отсюда вычислить константу  $K$  и сравнить ее с теоретическим значением, но это уже другая задача. Вопрос, который мы хотим сейчас сформулировать, звучит следующим образом. По мере того, как растет температура, молекула начинает колебаться с возрастающей амплитудой и при очень высоких температурах колебания становятся настолько сильными, что молекула распадается. Будем предполагать, что температура гораздо ниже этого порога, и амплитуда колебаний атомов много меньше, чем длина химической связи  $a$ . Какова зависимость этой амплитуды колебаний от температуры?

Чтобы дать количественное описание, введем переменную  $u(t)$ , определяющую отклонение координаты атома от равновесного значения  $u_0$ . Предположим, что атомы осциллируют с частотой  $\omega$ , так что

$$u(t) = A \cos(\omega t + \phi).$$

Удобно ввести также комплексную амплитуду  $\tilde{u} = (A/2) \exp[i\phi]$  и записать

$$u(t) = \tilde{u} \exp[-i\omega t] + \tilde{u}^* \exp[+i\omega t]. \quad (10.4.1)$$

Кинетическая энергия, отвечающая этому виду колебаний, записывается как

$$KE = \frac{M}{2} \left( \frac{du}{dt} \right)^2 = \frac{M\omega^2 A^2}{2} \sin^2(\omega t + \phi),$$

в то время как потенциальная энергия равна

$$PE = \frac{K}{2} u^2 = \frac{M\omega^2 A^2}{2} \cos^2(\omega t + \phi).$$

Обратите внимание на то, что  $K = M\omega^2$  и полная энергия, как и следовало ожидать, не зависит от времени:

$$E = KE + PE = M\omega^2 A^2 / 2 = 2M\omega^2 |\tilde{u}|^2. \quad (10.4.2)$$

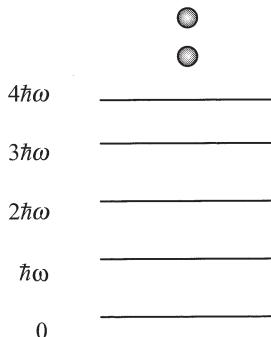


Рис. 10.4.2. Разрешенные уровни энергии осциллятора с резонансной частотой  $\omega$

В эпоху классической физики считалось, что энергия  $E$  может принимать любые значения. На заре двадцатого века Планк показал, что экспериментально наблюдаемые спектры излучения можно объяснить, если постулировать, что энергия электромагнитных осцилляторов принимает квантованные в единицах  $\hbar\omega$  значения. Сейчас известно, что энергии любых гармонических осцилляторов всегда кратны величине  $\hbar\omega$ , как это поясняет рис. 10.4.2 ( $n = 0, 1, 2, \dots$ ):

$$E = 2M\omega^2|\tilde{u}|^2 = n\hbar\omega \rightarrow |\tilde{u}| = \sqrt{n\hbar/2M\omega}. \quad (10.4.3)$$

Используя уравнение (10.4.3), можно переписать выражение (10.4.1) в виде

$$u(t) = \sqrt{\frac{\hbar}{2M\omega}} (a \exp[-i\omega t] + a^* \exp[+i\omega t]), \quad (10.4.4)$$

где  $|a|^2 = n = 0, 1, 2, \dots$

Поскольку энергия осциллятора равна целому кратному  $\hbar\omega$ , можно интерпретировать ее как одиночный уровень энергии  $\hbar\omega$ , на котором располагается целое число  $\nu = 0, 1, 2, \dots$  частиц, называемых фононами. Если осциллятор находится в состоянии с энергией  $4\hbar\omega$ , то мы говорим, что имеется четыре фонона. Таким образом, колебания решетки, как и электроны, наряду со свойствами волн приобретают свойства частицы. Различие между ними состоит в том, что электроны являются фермионами и подчиняются принципу запрета. Как следствие, число частиц  $n_k$ , которое можно разместить в данном состоянии  $k$ , равно нулю или единице. Фононы являются бозонами, и на них не распространяется принцип запрета Паули, так что в данном состоянии может присутствовать любое число фононов.

Чем больше это число заполнения  $\nu$ , тем большее значение имеет энергия колебаний и, стало быть, их амплитуда  $a$  в выражении (10.4.4). Для когерентных колебаний амплитуды  $a$  являются комплексными числами с хорошо определенными фазами, но для тепловых колебаний фазы нельзя ввести фазы с определенным значением.

Чему равно среднее число фононов при данной температуре? Для электронов, как известно, число частиц  $n_k$  в состоянии  $k$  может быть равно 0 или 1, а среднее их число в состоянии равновесия определяется функцией Ферми (формула (1.1.1)). Число фононов в конкретной моде колебаний может быть равно любому положительному числу, а среднее число  $\langle n \rangle$  в равновесии определяется функцией Бозе:

$$\langle n \rangle = \frac{1}{\exp(\hbar\omega/k_B T) - 1}. \quad (10.4.5)$$

Я не буду обсуждать физические принципы, исходя из которых получена эта функция (мы не обсуждали их и для функции Ферми), а замечу лишь, что функции Бозе и Ферми следуют из общего принципа равновесной статистической механики, содержащегося в выражении (3.4.5).

Определим, чему равен средний квадрат амплитуды колебаний при данной температуре  $T$ ? Из выражения (10.4.3) следует, что

$$\langle |\tilde{u}|^2 \rangle = \langle n \rangle \hbar / 2M\omega,$$

т. е. амплитуда колебаний растет вместе с числом фононов и поэтому можно ожидать, что должен наблюдаться ее рост с увеличением температуры. При температурах, намного меньших температуры плавления, эта амплитуда должна составлять лишь малую часть равновесной длины химической связи.

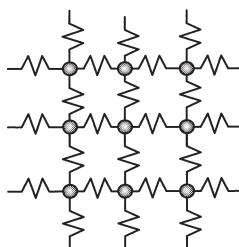
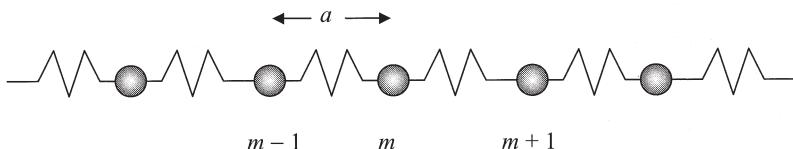


Рис. 10.4.3. Кристаллическая решетка как система масс и пружин

**Фононы.** В главе 3 мы сделали утверждение, согласно которому молекулу водорода можно представлять себе как систему двух масс, соединенных пружиной, чья длина в состоянии равновесия равна длине межатомной связи. Развивая эту точку зрения, будем рассматривать кристаллическую решетку твердого тела как периодический массив атомных масс, соединенных пружинами (рис. 10.4.3).

Каким образом следует описывать колебания этой системы, состоящей из масс и пружин? Можно приравнять полную силу, действующую на данную массу, произведению величины этой массы на ускорение и получить бесконечную систему уравнений, по одному на одну массу. Такая задача выглядит неразрешимой и не только в квантовой механике. Однако для решения можно использовать тот же метод, что применялся в главе 5 для расчета зонной структуры.

(a)



(б)

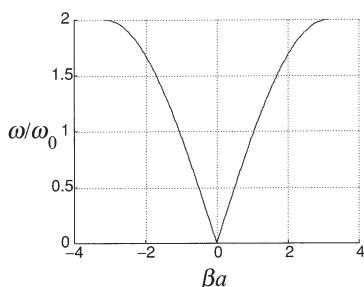


Рис. 10.4.4. (а) Бесконечная одномерная цепочка атомов. (б) Дисперсионная зависимость  $\omega/\omega_0$  от  $\beta a$

**Одномерное твердое тело.** Рассмотрим для простоты одномерный массив атомов, составленный из масс и пружин (рис. 10.4.4 a). Предполагая, что  $u_m$  есть смещение  $m$ -го атома из его положения равновесия в решетке, определим действующую на него силу слева  $K(u_m - u_{m-1})$  и справа  $-K(u_m - u_{m+1})$ . Записывая второй закон Ньютона, получаем, что

$$M \frac{d^2 u_m}{dt^2} = K[u_{m+1} - 2u_m + u_{m-1}], \quad (10.4.6)$$

где  $M$  — масса атома. Предполагая, что колебания имеют синусоидальный характер с частотой  $\omega$  и комплексной амплитудой  $\tilde{u}_m$ , так что  $u_m = \tilde{u}_m \exp(-i\omega t)$ , мы можем записать это уравнение в алгебраической форме:

$$M\omega^2 \tilde{u}_m = K[2\tilde{u}_m - \tilde{u}_{m-1} - \tilde{u}_{m+1}],$$

которое может быть представлено в форме матричного уравнения

$$\omega^2 \{\tilde{u}\} = [\Omega] \{\tilde{u}\}, \quad (10.4.7)$$

где  $\omega_0 \equiv \sqrt{K/M}$  и

$$\Omega = \begin{pmatrix} & 1 & 2 & \dots & \dots & N-1 & N \\ 1 & 2\omega_0^2 & -\omega_0^2 & & & 0 & -\omega_0^2 \\ 2 & -\omega_0^2 & 2\omega_0^2 & & & 0 & 0 \\ & \dots & \dots & & & & \\ N-1 & 0 & 0 & & & 2\omega_0^2 & -\omega_0^2 \\ N & -\omega_0^2 & 0 & & & -\omega_0^2 & 2\omega_0^2 \end{pmatrix}.$$

При записи этой матрицы были использованы периодические граничные условия, т. е. предполагалось, что твердое тело имеет форму кольца, в котором  $N$ -й атом соединяется с первым атомом. Обращаем внимание на то, что эта матрица имеет точно такой же вид, что и гамильтоновская матрица одномерной модели твердого тела (выражение (5.1.1)).

Данная аналогия, разумеется, является чисто математической, поскольку уравнения (5.1.1) и (10.4.7) описывают различные физические явления: первое определяет квантовую механику электронов в периодической решетке, а второе описывает классическую динамику периодической системы масс и пружин. Но, поскольку матрица  $[H]$  в выражении (5.1.1) имеет точно такой вид, как и матрица  $[\Omega]$  в уравнении (10.4.7), для их диагонализации можно использовать единый подход, который приводит к следующему дисперсионному соотношению:

$$\omega^2 = 2\omega_0^2(1 - \cos \beta a), \quad \text{где } \beta a = n2\pi/N. \quad (10.4.8)$$

Значения  $\beta a$  пробегают интервал от  $-\pi$  до  $+\pi$  с шагом  $2\pi/N$ , аналогично значениям  $ka$  на рис. 5.1.2. Смещения масс, которым отвечают компоненты собственного вектора  $\tilde{u}_m$  в данной моде  $\beta$ , определяются выражением

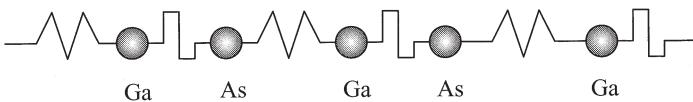
$$\sum_{\beta} \tilde{u}_{\beta} \exp[i\beta ma] = \sum_{\beta} \tilde{u}_{\beta} \exp[i(\beta ma - \omega(\beta))t].$$

Фактические мгновенные значения смещений равны вещественной части этих комплексных функций. Здесь та же ситуация, что и при расчете напряжений и токов в электрическом контуре

$$u_m(t) = 2\operatorname{Re} \left\{ \sum_{\beta} \tilde{u}_{\beta} \exp[i(\beta m a - \omega(\beta))t] \right\} = \\ = \sum_{\beta} \tilde{u}_{\beta} \exp[i(\beta m a - \omega(\beta))t] + \sum_{\beta} \tilde{u}_{\beta}^* \exp[-i(\beta m a - \omega(\beta))t]. \quad (10.4.9)$$

Точное значение амплитуд  $u_{\beta}$  зависит от конкретного вида воздействия, возбуждающего колебания в решетке или от начальных условий. Однако, если известен соответствующий набор амплитуд  $u_{\beta}$ , то любая функция  $u_m(t)$  на решетке может в соответствии с (10.4.9) быть представлена как суперпозиция нормальных колебаний.

(a)



(б)

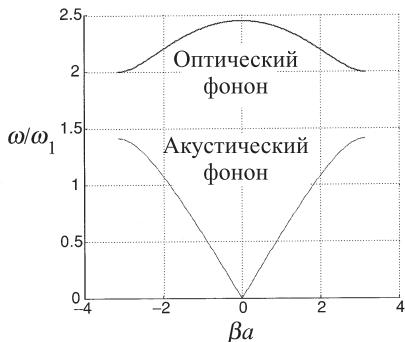


Рис. 10.4.5. (а) Мгновенный снимок движения атомов в решетке GaAs вдоль направления (111). (б) Дисперсионная зависимость  $\omega/\omega_1$  как функция  $\beta a$  для акустических и оптических фононов, построенная при  $\omega_2 = \sqrt{2}\omega_1$

**Оптические фононы.** Мы уже знаем, что в кристаллах с элементарной ячейкой, состоящей из двух атомов (см. рис. 5.1.4), график зависимости

$E(k)$  состоит из двух ветвей (см. рис. 5.1.5). Рассмотрим решетку с двумя различными массами  $M_{1,2}$  и/или константами жесткости  $K_{1,2}$  в одной элементарной ячейке, как это показано на рис. 10.4.5 *a*. Записав законы Ньютона, придем к матричному уравнению

$$\omega^2 \{\tilde{u}\} = [\Omega] \{\tilde{u}\}, \quad (10.4.10)$$

где  $\omega_1 \equiv \sqrt{K_1/M_1}$  и  $\omega_2 \equiv \sqrt{K_2/M_2}$ , а матрица  $\Omega$  устроена следующим образом:

$$\Omega = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots \\ 2 & \omega_1^2 + \omega_2^2 & -\omega_1^2 & 0 & 0 & \dots \\ 3 & -\omega_1^2 & \omega_1^2 + \omega_2^2 & -\omega_2^2 & 0 & \dots \\ 4 & 0 & -\omega_1^2 & \omega_1^2 + \omega_2^2 & -\omega_1^2 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Обратите внимание на то, что эта матрица напоминает матрицу из формулы (5.1.6), и ее собственные значения могут быть найдены с помощью аналогичного метода, который приводит к следующим уравнениям (сравните с уравнением (5.1.10)):

$$\begin{pmatrix} \omega_1^2 + \omega_2^2 & \omega_1^2 + \omega_2^2 e^{-i\beta a} \\ \omega_1^2 + \omega_2^2 e^{+i\beta a} & \omega_1^2 + \omega_2^2 \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix} = \omega^2(\beta) \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}. \quad (10.4.11)$$

Приравнивая определитель этой системы к нулю:

$$\det \begin{pmatrix} \omega_1^2 + \omega_2^2 & \omega_1^2 + \omega_2^2 e^{-i\beta a} \\ \omega_1^2 + \omega_2^2 e^{+i\beta a} & \omega_1^2 + \omega_2^2 \end{pmatrix} = 0,$$

мы получаем дисперсионное соотношение:

$$\omega^2(\beta) = \omega_1^2 + \omega_2^2 \pm (\omega_1^4 + \omega_2^4 + 2\omega_1^2\omega_2^2 \cos \beta a)^{1/2}. \quad (10.4.12)$$

Аналогично уравнению (5.1.10), уравнение (10.4.12) приводит к появлению двух ветвей фононного спектра, показанного на рис. 10.4.5 *b* (сравните его с рис. 5.1.5). Ветвь с более низкой частотой называется ветвью акустических фононов, поскольку в области низких частот она отвечает обычным акустическим волнам. Для этих волн смещения обоих атомов в элементарной ячейке (Ga и As) имеют один и тот же знак. Напротив, в оптической ветви эти смещения имеют противоположные знаки. Название «оптические» происходит от того, что этот вид колебаний сравнительно

легко возбуждается падающим электромагнитным излучением, чье электрическое поле заставляет атомы Ga и As двигаться в противоположных направлениях, поскольку один из них заряжен отрицательно, а второй положительно.

Нетрудно увидеть причину, по которой частота  $\omega$  растет с ростом  $\beta$  в акустической ветви, но остается почти постоянной в оптической ветви. Мы знаем, что резонансная частота системы из масс и пружины возрастает по мере роста жесткости. Возбуждение колебаний в акустической моде с малым значением  $\beta$  приводит к очень небольшому растяжению соответствующей воображаемой пружины, поскольку атомы в целом движутся в одном направлении. Таким образом, эффективная жесткость уменьшается при малых смещениях  $\beta$ , что и приводит к уменьшению частот колебаний. Напротив, для оптических мод степень растяжения пружины почти не зависит от  $\beta$ . Даже если  $\beta$  равно нулю, атомы Ga и As будут двигаться в противоположных направлениях, приводя к значительному растяжению воображаемой пружины. В этой ситуации частота  $\omega$  как функция волнового вектора  $\beta$  меняется очень слабо.

**Трехмерные твердые тела.** В реальных трехмерных твердых телах атомное смещение каждого из двух атомов в единичной ячейке задается вектором с тремя компонентами  $u_x, u_y, u_z$ . Соответственно появляются шесть ветвей спектра, три из которых акустические, а три оптические. Одна из трех ветвей соответствует продольной моде, в которой смещение параллельно вектору  $\vec{\beta}$ , а две другие — поперечным модам, в которых атомные смещения перпендикулярны  $\vec{\beta}$ . Таким образом, в трехмерном кристалле распространяется продольная акустическая мода (LA), две поперечные акустические моды (TA), а также одна продольная оптическая (LO) и две поперечные оптические моды (TO). Полное векторное смещение на данном узле решетки и в данный момент времени  $(\vec{r}, t)$  может быть записано в виде

$$\begin{aligned} \{\vec{u}(\vec{r}, t)\} = \sum_{\nu, \vec{\beta}} \{u_{\nu, \vec{\beta}}\} \exp[i(\beta \cdot \vec{r} - \omega_{\nu}(\beta)t)] + \\ + \{u_{\nu, \vec{\beta}}\}^* \exp[-i(\beta \cdot \vec{r} - \omega_{\nu}(\beta)t)]. \quad (10.4.13) \end{aligned}$$

Здесь индекс  $\nu$  нумерует все шесть ветвей дисперсионного спектра (LA, TA и т. д.),  $\{u_{\nu, \vec{\beta}}\}$  представляют собой векторы-столбцы размера  $(6 \times 1)$ , удовлетворяющие системе уравнений на собственные значения с матрицей  $(6 \times 6)$ . Такая система получается как обобщение (10.4.9) с включением сюда трех компонент двух векторов, описывающих смещения

атомов разного сорта. Собственные векторы размером  $(2 \times 1)$  для каждой из трех поляризаций  $\vec{\nu}$  записываются следующим образом:

$$\{\vec{u}(\vec{r}, t)\}_\nu = \sum_{\vec{\beta}} \{u_{\vec{\beta}}\}_\nu \exp[i(\vec{\beta} \cdot \vec{r} - \omega_\nu(\beta)t)] + \{u_{\vec{\beta}}\}_\nu^* \exp[-i(\vec{\beta} \cdot \vec{r} - \omega_\nu(\beta)t)]. \quad (10.4.14)$$

Здесь две компоненты описывают смещение атомов разного сорта в одной элементарной ячейке. Повторяя те же аргументы, что были использованы при выводе уравнения (10.4.4), можно записать вектор смещения в виде

$$\{\vec{u}_{\vec{\beta}}\}_\nu = \vec{\nu} \sqrt{\frac{\hbar}{2\rho\Omega\omega}} a_{\nu,\vec{\beta}} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad (10.4.15)$$

где компоненты  $(u_1, u_2)$  равны  $(1, 1)/\sqrt{2}$  для акустических фононов и  $(1, -1)/\sqrt{2}$  для оптических фононов. Квадрат амплитуды определяет число фононов в данной моде, а их среднее значение задается функцией Бозе–Эйнштейна:

$$|a_{\nu,\vec{\beta}}|^2 = n_{\nu,\vec{\beta}}, \quad \langle n_{\nu,\vec{\beta}} \rangle = \frac{1}{\exp(\hbar\omega_\nu(\beta)/k_B T) - 1}. \quad (10.4.16)$$

Используя выражения (10.4.14)–(10.4.16), можно записать средний квадрат смещения атома как

$$\langle u^2 \rangle = \sum_{\vec{\beta}} \frac{\hbar}{2\rho\omega\Omega} \frac{1}{\exp(\hbar\omega/k_B T) - 1}. \quad (10.4.17)$$

**Деформации в решетке, обусловленные одним фононом.** Сейчас будет получен результат, сформулированный ранее как выражение (10.1.16). Продольная деформация в атомной решетке определяется как дивергенция вектора смещения:

$$\begin{aligned} S = \vec{\nabla} \cdot \vec{u} &= \sum_{\vec{\beta}} \vec{\beta} \cdot \vec{\nu} \frac{\hbar}{2\rho\omega\Omega} \left( a_{\vec{\beta},\vec{\nu}} \exp[i(\vec{\beta} \cdot \vec{r} - \omega_\nu(\beta)t)] + \right. \\ &\quad \left. + a_{\vec{\beta},\vec{\nu}}^* \exp[-i(\vec{\beta} \cdot \vec{r} - \omega_\nu(\beta)t)] \right). \end{aligned}$$

Для одного фона на с волновым вектором  $\vec{\beta}$  и поляризацией  $\vec{\nu}$  в соответствии с выражением (10.1.16) можем положить  $a_{\vec{\beta},\vec{\nu}} = 1$

$$S = (\vec{\beta} \cdot \vec{\nu}) \sqrt{2\hbar/\rho\omega\Omega} \cos(\vec{\beta} \cdot \vec{r} - \omega(\beta)t).$$

## Упражнения

**У.10.1.** Предполагая, что электронная скорость равна  $\hbar k/m$ , а скорость звука равна  $c_s$ , вычислите правые части выражений (10.2.16a), (10.2.16b). Действуйте в следующей последовательности: (а) выполните переход от суммирования к интегрированию; и (б) запишите аргумент дельта-функции в виде

$$\delta\left(\frac{\hbar^2 k \beta}{m} \left[ \cos \theta - \frac{\bar{c}}{\hbar k/m} \pm \frac{\beta}{2k} \right] \right);$$

(в) выполните интегрирование по  $\cos \theta$  для того, чтобы избавиться от дельта-функции и задать конечные пределы интегрирования по  $\beta$ :  $\beta_{\min} < \beta < \beta_{\max}$ ; (г) выполните интегрирование по  $\beta$ . Покажите, что время жизни, обусловленное поглощением и излучением акустического фонона, определяется выражением

$$\frac{1}{\tau(k)} = \frac{m D^2 k_B T}{\pi \hbar^3 \rho c_s^2} k.$$

(д) Какой вид имеет угловое распределение испускаемых фононов?

**У.10.2.** Известно, что электрон находится в зоне с параболическим законом дисперсии, массой  $m$  и энергией  $E$ , превышающей энергию оптического фона  $\hbar\omega_0$ . Приравняв аргумент дельта-функции в формуле (10.1.13) к нулю, получите выражение, связывающее амплитуду  $\beta$  волнового вектора *испускаемого* оптического фона с углом  $\theta$ , под которым он излучается (отсчитываемым от вектора  $\vec{k}$ ) так, как это показано на рис. У.10.2). Чему равен интервал углов  $\theta$ , вне которого оптические фононы не излучаются?

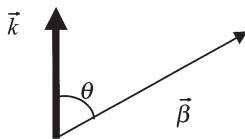


Рис. У.10.2

**У.10.3.** Среднее значение квадрата смещения атома, обусловленное фононами, в соответствии с формулой (10.4.17) равно

$$\langle u^2 \rangle = \sum_{\beta} \frac{\hbar}{2\rho\omega\Omega} \frac{1}{\exp(\hbar\omega/k_B T) - 1}.$$

Перейдите от суммирования к интегрированию, предполагая наличие периодических граничных условий, и вычислите значение интеграла численно, построив зависимость  $\sqrt{\langle u^2 \rangle}$  от температуры  $T$  в интервале  $0 \text{ K} < T < 1000 \text{ K}$ . Расчет произведите в предположении о наличии акустических фононов с законом дисперсии  $\omega_\beta = c_s\beta$  ( $c_s = 5 \times 10^3 \text{ м/с}$  и  $\rho_s = 5 \times 10^3 \text{ кг/м}^3$ ).

**У.10.4.** В этой книге рассматриваются, в основном, задачи квантового транспорта, а оптические процессы обсуждались только в связи с их влиянием на электронный транспорт. Интересно, однако, продемонстрировать, что развитые в этой книге понятия теории транспорта могут быть использованы также и для описания оптических процессов. Такой подход, насколько мне известно, не освещается в литературе достаточно подробно. Рассмотрим, к примеру, процесс оптического поглощения, при котором электрон переходит из валентной зоны в зону проводимости. В принципе этот процесс допустимо рассматривать и как упругий транспорт из состояний валентной зоны с  $N$  фотонами в состояние зоны проводимости с  $(N - 1)$  фотонами, как схематически показано ниже (параметр  $K$  есть матричный элемент процесса поглощения, определенный уравнением (10.1.20a)).

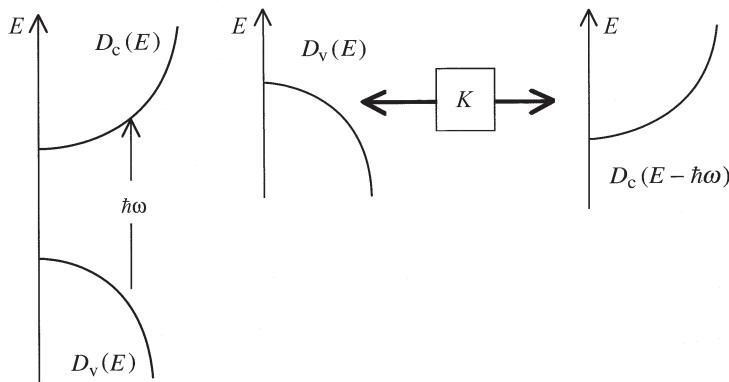


Рис. У.10.4

Объясните, почему интенсивность, с которой фотоны поглощаются за одну секунду, может быть записана как

$$R_{\text{abs}} = \frac{1}{2\pi\hbar} \int dE \bar{T}(E) [f_V(E) - f_C(E)],$$

где коэффициент «пропускания» легко находится из альтернативного выра-

жения, приведенного в упражнении У.9.3:

$$\bar{T}(E) = \text{Tr}[A_1 M A_2 M^+].$$

При решении следует связать спектральные функции  $A_1$  и  $A_2$  с плотностями состояний  $2\pi D_V(E)$  и  $2\pi D_C(E - \hbar\omega)$  в двух «контактах», а также учесть, что матричный элемент  $M$  в данном случае равен  $K$ , где  $K$  определено выражением (10.1.20а). В итоге должно получиться

$$R_{\text{abs}} = \frac{2\pi}{\hbar} \int dE |K|^2 D_V(E) D_C(E - \hbar\omega) [f_V(E) - f_C(E)].$$

**У.10.5.** Другие упражнения, иллюстрирующие теорию диссипативного квантового транспорта, читатель может найти в главе 11.

# Глава 11

## От атома к транзистору

В главе 1, обсуждая схему транзистора, показанную на рис. 11.1.1 *a*, я постарался пояснить, какие проблемы мотивировали написание данной книги. В последующих главах (со второй по седьмую) мы рассматривали структуру гамильтоновской матрицы  $[H_0]$ , относящейся к активной области транзистора и описывающей соответствующие разрешенные уровни энергии (см. рис. 11.1.1 *b*). В главе 8 были введены величины  $[\Gamma_1]$  и  $[\Gamma_2]$ , которые характеризуют взаимодействие с контактами истока и стока и определяют уширение. Далее в главе 9 я ввел новые понятия, используемые в расчетах электронного тока, *пренебрегая* при этом процессами сбоя фазы. В главе 10 обсуждалась природа процессов, идущих с нарушением фазы (некогерентных процессов), а также то, какую роль играют входящий и исходящий потоки в теории электронного транспорта. Теперь у нас есть полный «инструментарий», необходимый для описания квантового диссипативного транспорта в модели самосогласованного поля, которая обсуждалась в главе 3. Напомним, что при этом каждый электрон рассматривается как независимая частица, движущаяся в усредненном потенциале, созданном другими электронами. Впрочем, заметим, что модель независимых электронов не учитывает «сильные корреляционные эффекты» (см. параграф 1.5), которые до сих пор еще не слишком хорошо изучены. Вопрос о том, в какой степени подобные эффекты могут быть учтены, остается открытым (см. Приложение, параграф П.5).

В этой главе мы дадим сводку всех основных положений развитой теории (параграф 11.1), а также проиллюстрируем ее применение к конкретным задачам. Я полагаю, что эти примеры могут стать отправной точкой для тех читателей, которые хотели бы применить развитые подходы для решения новых задач в их профессиональной области. В то же время я выбрал эти примеры для иллюстрации фундаментальных положений, принципиально важных для понимания природы электрического сопротивления проводников, имеющих атомные масштабы. Мы выясним, как возникает закон Ома  $V = IR$ , согласно которому сопротивление  $R$  пропорционально

длине проводника (параграф 11.2), обсудим пространственное распределение тепла, выделяющегося обратимым и необратимым образом при протекании тока (параграф 11.3) и, наконец, выясним характер пространственного распределение напряжения в структуре и на контактах (параграф 11.4).

## 11.1. Уравнения квантового транспорта

Позвольте мне привести основные формулы теории квантового транспорта, обсуждавшейся ранее:

$$G^n = G\Sigma^{\text{in}}G^+; \quad (11.1.1)$$

$$G = [EI - H_0 - U - \Sigma]^{-1}; \quad (11.1.2)$$

$$A = i[G - G^+], \quad \Gamma = i[\Sigma - \Sigma^+], \quad (11.1.3)$$

где

$$\begin{aligned} \Sigma^{\text{in}} &= \Sigma_1^{\text{in}} + \Sigma_2^{\text{in}} + \Sigma_3^{\text{in}}, \\ \Sigma &= \Sigma_1 + \Sigma_2 + \Sigma_3. \end{aligned} \quad (11.1.4)$$

Приведенные уравнения могут быть использованы для расчета корреляционной функции  $G^n$ , а стало быть, и матрицы плотности  $\rho$ , чьи диагональные элементы определяют электронную плотность:

$$\rho = \int dE G^n(E)/2\pi. \quad (11.1.5)$$

Ток для одного состояния с определенным значением спина в  $i$ -м контакте находится с помощью выражения

$$I_i = (q/\hbar) \int_{-\infty}^{+\infty} dE \tilde{I}_i(E)/2\pi, \quad (11.1.6)$$

где

$$\tilde{I}_i = \text{Tr}[\Sigma^{\text{in}} A] - \text{Tr}[\Gamma_i G^n], \quad (11.1.7)$$

где смысл входящего ( $\Sigma^{\text{in}} A$ ) и исходящего ( $\Gamma_i G^n$ ) потоков в типичной транзисторной структуре ясен из рис. 11.1.1 $a$  и 11.1.1 $b$ .

**Исходные параметры.** Для того чтобы воспользоваться уравнениями, записанными выше, необходимо знать гамильтониан  $[H_0]$ , описывающий

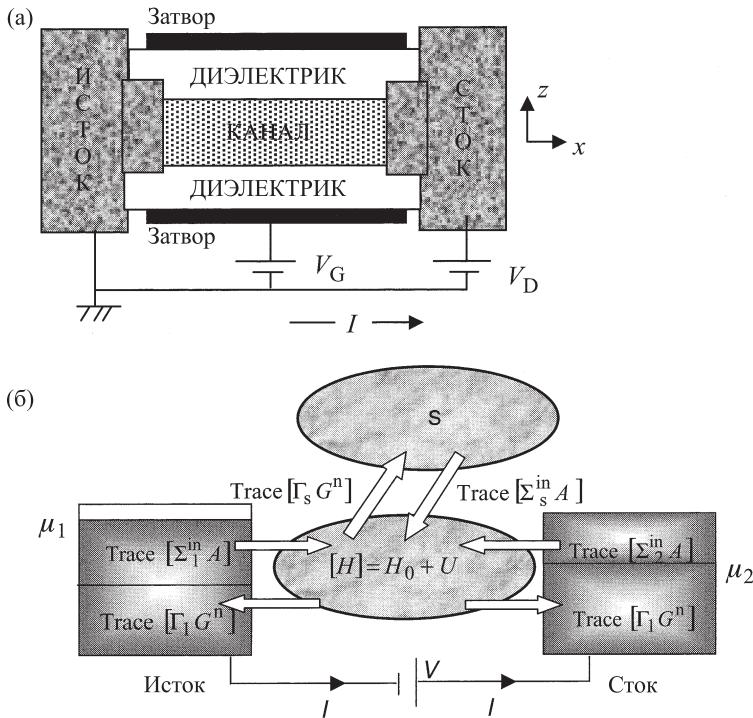


Рис. 11.1.1. (а) Типичная структура транзистора, обсуждавшаяся во введении. (б) Входящий и исходящий потоки электронов в транзисторной структуре, представленной на рис. (а)

канал, а также функцию  $[\Sigma^{\text{in}}]$ , описывающую рассеяние, и  $[\Gamma]$ , определяющую уширение. Для системы с двумя контактами между этими функциями существует следующая связь:

$$\Sigma_1^{\text{in}} = \Gamma_1 f_1 \quad \text{и} \quad \Sigma_2^{\text{in}} = \Gamma_2 f_2, \quad (11.1.8)$$

а значение уширения или собственно-энергетической функции для каждого контакта может быть определено через поверхностную спектральную функцию  $a$ , поверхностную функцию Грина  $g$  для контакта, а также матриц  $[\tau]$ , описывающих взаимодействие канала и контакта:

$$\Gamma = \tau a \tau^+ \quad \text{и} \quad \Sigma = \tau g \tau^+. \quad (11.1.9)$$

Все численные результаты, представленные в этой главе, будут получены в рамках простой однозонной модели и в приближении эффективной массы (равной массе свободного электрона). В этой модели будет записан гамильтониан канала  $[H_0]$  и собственно-энергетические функции контактов  $\Sigma_1, \Sigma_2$ .

Для рассеивающей «границной области», в отличие от контакта, не существует простой связи между  $\Sigma_s^{\text{in}}$  и  $\Sigma_s$  (или  $\Gamma_s$ ). Если рассеивание практически упругое ( $E \approx E + \hbar\omega$ ), то, исходя из уравнений (10.3.10)–(10.3.12), можно записать следующее:

$$\begin{aligned}\Sigma_s^{\text{in}}(E) &= D_0 G^n(E); \\ \Gamma_s(E) &= D_0 A(E); \\ \Sigma_s(E) &= D_0 G(E).\end{aligned}\tag{11.1.10}$$

Для процессов же неупругого рассеяния общего вида имеют место следующие выражения (см. уравнения (10.3.8), (10.3.9)):

$$\begin{aligned}\Sigma_s^{\text{in}}(E) &= \int_0^\infty \frac{d(\hbar\omega)}{2\pi} \left( D^{em}(\hbar\omega) \cdot G^n(E + \hbar\omega) \right. \\ &\quad \left. + D^{ab}(\hbar\omega) \cdot G^n(E - \hbar\omega) \right); \\ \Gamma_s(E) &= \int_0^\infty \frac{d(\hbar\omega)}{2\pi} \left( D^{em}(\hbar\omega) \cdot [G^p(E - \hbar\omega) + G^n(E + \hbar\omega)] \right. \\ &\quad \left. + D^{ab}(\hbar\omega) \cdot [G^n(E - \hbar\omega) + G^p(E - \hbar\omega)] \right).\end{aligned}$$

В наших примерах мы будем учитывать только фононы с фиксированной частотой  $\omega_0$ , и поэтому воспользуемся более простыми выражениями

$$\begin{aligned}\Sigma_s^{\text{in}}(E) &= D_0^{em} G^n(E + \hbar\omega_0) + D_0^{ab} G^n(E - \hbar\omega_0); \\ \Gamma_s(E) &= D_0^{em} [G^p(E - \hbar\omega_0) + G^n(E + \hbar\omega_0)] \\ &\quad + D_0^{ab} [G^n(E - \hbar\omega_0) + G^p(E + \hbar\omega_0)].\end{aligned}\tag{11.1.11}$$

Кроме того, мы будем пренебрегать эрмитовой частью функции  $\Sigma_s(E)$ , которая находится путем преобразования Гильберта функции  $\Gamma_s(E)$ . Обратите внимание на то, что функция рассеяния для входящего потока и функция уширения в уравнениях (11.1.10) и (11.1.11) зависят от корреляционных функций, в отличие от случая когерентного транспорта. Это обстоятельство усложняет решение уравнений транспорта (11.1.1)–(11.1.7), для самосогласованного решения которых необходимо использовать итерационную процедуру.

Обратите внимание на то, что в уравнениях (11.1.10) и (11.1.11) мы рассматриваем тензоры  $D^{em}$  и  $D^{ab}$ , описывающие испускание и поглощение как скалярные величины. Полную тензорную запись можно найти в Приложении (см. уравнения (П.4.9) и (П.4.10)).

**Уравнение диффузии.** Если «фононная спектральная функция»  $D_0$  – это константа, умноженная на единичную матрицу, то из уравнения (11.1.10) следует, что  $\Sigma_s^{\text{in}}$  есть диагональная матрица:

$$\Sigma_s^{\text{in}}(r, r; E) = D_0 G^n(r, r; E).$$

Для проводников большой длины можно пренебречь влиянием контактов и записать уравнение (11.1.10) в форме  $G^n = G \Sigma_s^{\text{in}} G^+$ , откуда следует, что диагональные элементы корреляционной матрицы, которые можно отождествить с электронной плотностью  $n(r; E) = G^n(r, r; E)/2\pi$ , удовлетворяют уравнению

$$n(r; E) = \sum_{r'} D_0 |G(r, r'; E)|^2 n(r'; E).$$

Можно показать, что если функция Грина медленно меняется в пространстве, то это уравнение сводится к уравнению диффузии  $\nabla^2 n(r; E) = 0$  [12].

**Функция пропускания.** В главе 9 мы убедились в том, что метод функции пропускания является весьма продуктивным, и с помощью этого метода получили выражение для тока (см. уравнение (9.1.9)). Упомянутый метод остается полезным и для рассмотрения диссипативного квантового транспорта на качественном уровне, однако рассчитать ток через функцию пропускания не представляется возможным. Дело в том, что в отличие от области контактов, где  $\Sigma_i^{\text{in}} = \Gamma_i f_i$ , не существует простой связи между  $\Sigma_s^{\text{in}}$  и  $\Gamma_s$ . В этой ситуации более удобным является расчет тока непосредственно из уравнения (11.1.6). Мы можем определить величину эффективного пропускания, сравнивая выражение (11.1.6) с величиной (9.1.9):

$$\bar{T}_{\text{eff}}(E) = \frac{\tilde{I}_i(E)}{f_1(E) - f_2(E)} = \frac{\text{Tr}[\Sigma_i^{\text{in}} A] - \text{Tr}[\Gamma_i G^n]}{f_1 - f_2}, \quad (11.1.12)$$

которую можно сравнить с соответствующим выражением для когерентного транспорта.

**Самосогласованный расчет.** В общем случае необходимо находить самосогласованное решение уравнений транспорта и уравнения «Пуассона», которое учитывает электрон-электронные взаимодействия через самосогласованный потенциал  $U(\vec{r})$  (см. рис. 11.1.2). Здесь название уравнения

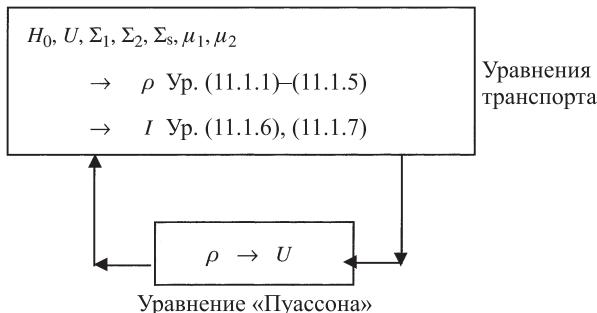


Рис. 11.1.2. В общем случае задача об электронном транспорте должна решаться самосогласованно с учетом уравнения «Пуассона», описывающего электрон-электронные взаимодействия через потенциал  $U$

заключено в кавычки в связи с тем, что в этой части задачи может оказаться необходимым учет поправок, связанных с корреляционными эффектами (см. главу 3). Это обстоятельство зачастую упускается из виду (как постутили и мы в параграфах 11.2 и 11.3) при расчетах «линейного отклика» на внешнее напряжение, малое по сравнению с тепловой энергией  $k_B T$  и/или по сравнению с масштабом энергии, на котором существенно меняется плотность состояний. Линейный отклик не зависит от конкретного профиля потенциала  $U(\vec{r})$ , сформированного напряжением, приложенным к стоку. Но если нас интересует вольт-амперная характеристика в широком интервале напряжений, например, при расчете тока в открытом состоянии транзистора (параграф 1.4), то конкретный профиль потенциала чрезвычайно важен. Этот обстоятельство уже было отмечено в параграфе 1.4.

**Простой пример системы с одним уровнем.** В главе 1 мы рассматривали систему с одним энергетическим уровнем, в которой электронная плотность и ток могли быть найдены из обычного уравнения баланса с простой функцией уширения уровня. Далее я отмечал, что в общем случае потребуется матричная версия этой «скалярной модели». Этой проблеме и посвящена оставшаяся часть книги (см. рис. 1.6.5).

Теперь, когда в нашем распоряжении имеется полная «матричная модель», мы сможем выполнить все необходимые расчеты. Они будут описаны в оставшейся части этой главы. Но, перед тем как мы займемся вычислениями, полезно остановиться на обсуждении одноуровневой системы с упругим процессом нарушения фазы (рис. 11.1.3), где все матрицы сводятся к обычным числам, а результаты легко получаются аналитически. Из

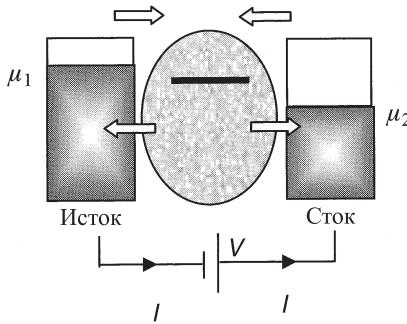


Рис. 11.1.3

уравнений (11.1.1)–(11.1.4) находим, что

$$G(E) = (E - \varepsilon + (i\Gamma/2))^{-1};$$

$$A(E) = \frac{\Gamma}{(E - \varepsilon)^2 + (\Gamma/2)^2};$$

$$\Gamma = \Gamma_1 + \Gamma_2 + D_0 A;$$

$$G^n(E) = \Sigma^{\text{in}}(E) A(E) / \Gamma;$$

$$\Sigma^{\text{in}} = \Gamma_1 f_1 + \Gamma_2 f_2 + D_0 G^n.$$

Отсюда получаем, что

$$\frac{G^n}{A} = \frac{\Gamma_1 f_1 + \Gamma_2 f_2 + D_0 G^n}{\Gamma_1 + \Gamma_2 + D_0 A} = \frac{\Gamma_1 f_1 + \Gamma_2 f_2}{\Gamma_1 + \Gamma_2},$$

независимо от амплитуды рассеяния  $D_0$ . Из уравнений (11.1.6) и (11.1.7) находим

$$\tilde{I}_s = D_0 G^n a - D_0 A G^n = 0$$

и

$$\tilde{I}_1 = \Gamma_1 A (f_1 - (G^n/A)) = \frac{\Gamma_1 \Gamma_2 A}{\Gamma_1 + \Gamma_2} (f_1 - f_2) = -\tilde{I}_2,$$

откуда

$$I = \frac{q}{h} \int dE \frac{\Gamma_1 \Gamma_2 A}{\Gamma_1 + \Gamma_2} (f_1 - f_2),$$

что говорит о том, что упругий процесс нарушения фазы такого вида, присутствующий в одноуровневой системе, не сказывается на величине тока, которая не зависит от  $D_0$ .

Даже для одноуровневой системы процесс неупругого изменения фазы является более сложным, поскольку функция рассеяния для входящего тока  $\Sigma_s^{\text{in}}$  при энергии  $E$  зависит от корреляционной функции  $G^n$  при энергии  $E + \hbar\omega$  (см. упражнение У.11.6).

## 11.2. Закон Ома

В этом параграфе я покажу, как общие уравнения квантового транспорта могут быть использованы для описания проводников, в которых прохождение тока сопровождается нарушением фазы. Мы выясним как очень короткий баллистический проводник превращается в макроскопический проводник, подчиняющийся закону Ома, и проводимость оказывается прямо пропорциональной площади поперечного сечения  $S$  и обратно пропорциональна длине проводника  $L$ . В главе 1, используя элементарные аргументы, я объяснял, почему в баллистическом проводнике проводимость должна возрастать с ростом  $S$ . Теперь, когда мы познакомились в главе 6 с понятием энергетической подзоны, можно сформулировать нашу точку зрения более точно. А именно, проводимость проводника возрастает с ростом поперечного сечения, поскольку растет число подзон, которые могут переносить ток, причем для массивных проводников это число прямо пропорционально величине  $S$ .

Но почему проводимость должна уменьшаться с ростом длины образца  $L$ ? В баллистическом проводнике без рассеяния проводимость, как известно, не зависит от его длины. Но в длинных проводниках проводимость уменьшается с длиной, также как средняя вероятность прохождения пути от истока до стока. В главе 9 мы видели, что проводимость вблизи энергии Ферми (см. уравнение (9.4.4)) пропорциональна полному коэффициенту пропускания  $\bar{T}$ , который может быть записан как произведение числа мод  $M$  на усредненную вероятность прохождения (на одну моду  $T$ ):

$$G = (2q^2/h)\bar{T} = (2q^2/h)MT \quad (\text{формула Ландауэра}). \quad (11.2.1)$$

Для массивных проводников  $M \sim S$  и  $T \sim 1/L$ , что и приводит к закону Ома  $G \sim S/L$ . Возникает и следующий вопрос: почему вероятность прохождения уменьшается с ростом длины образца?

### 11.2.1. Классический транспорт

Если электрон рассматривается как классическая частица, то нетрудно понять, почему вероятность прохождения должна иметь вид  $T \sim 1/L$ .

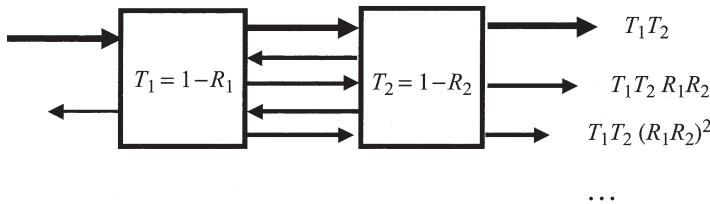


Рис. 11.2.1. Классическое «сложение» вероятностей прохождения для двух последовательно соединенных секций позволяет определить полную вероятность прохождения

Чтобы получить этот закон, рассмотрим проводник, состоящий из двух последовательно соединенных секций, как это показано на рис. 11.2.1. Первая секция характеризуется вероятностью прохождения  $T_1$ , а вторая, соответственно,  $T_2$ . Чему равна вероятность  $T$  прохождения электрона через обе секции? Очевидный ответ:  $T = T_1 T_2$  был бы неверен. Это произведение описывает только вероятность прохождения электроном двух секций с первой попытки. Но после отражения электрона от секции 2 при первой попытке он сможет пройти через эту секцию и во второй раз после отражения от первой секции, что будет характеризоваться вероятностью  $T_1 T_2 R_1 R_2$ , как это показано на рис. 11.2.1 (коэффициенты отражения  $R_1 = 1 - T_1$  и  $R_2 = 1 - T_2$ ).

Аналитическое суммирование вероятностей всех возможных последовательных попыток прохождения приведет к следующему результату:

$$T = T_1 T_2 ((R_1 R_2) + (R_1 R_2)^2 + (R_1 R_2)^3 + \dots),$$

$$\frac{T_1 T_2}{1 - R_1 R_2} = \frac{T_1 T_2}{T_1 + T_2 - T_1 T_2},$$

откуда следует

$$\frac{1}{T} = \frac{1}{T_1} + \frac{1}{T_2} - 1. \quad (11.2.2)$$

Это соотношение определяет итоговое значение коэффициента пропускания  $T$  двух секций с коэффициентами прохождения  $T_1$  и  $T_2$ . С его помощью нетрудно получить выражение для зависимости проводимости образца от его длины  $T(L)$ , определив функцию, удовлетворяющую уравнению

$$\frac{1}{T(L_1 + L_2)} = \frac{1}{T(L_1)} + \frac{1}{T(L_2)} - 1.$$

Нетрудно проверить, что такой функцией является

$$T = \frac{\Lambda}{L + \Lambda}, \quad (11.2.3)$$

где постоянная  $\Lambda$  имеет порядок длины свободного пробега, которую можно трактовать как масштаб длины, на котором вероятность пропускания уменьшается приблизительно вдвое.

Уравнение (11.2.3) описывает вероятность пропускания для классических частиц как функцию длины проводника. Объединяя его с уравнением (11.2.1), получим

$$G = (2q^2 M/h) \frac{\Lambda}{L + \Lambda}.$$

Значит, в соответствии с законом Ома проводимость может быть записана как сумма двух слагаемых, или сумма двух последовательно соединенных сопротивлений: постоянного сопротивления границы раздела и сопротивления «структуры», которое линейно растет с ее длиной:

$$\frac{1}{G} = \frac{h}{2q^2 M} \underset{\text{сопротивление границы}}{} + \frac{h}{2q^2 M} \frac{L}{\Lambda} \underset{\text{сопротивление «структурой»}}{} . \quad (11.2.4)$$

Выражение (11.2.4) наводит на мысль о том, что проводимость структуры должна записываться в виде

$$G = \frac{2q^2 M}{h} \frac{T}{1 - T}, \quad (11.2.5)$$

который был впервые получен Ландауэром. Как и следовало ожидать, выражение (11.2.5) приводит к нулевому сопротивлению для баллистического проводника с  $T = 1$ , в то время как уравнение (11.2.1) приводит к ненулевому значению сопротивления. Физический смысл этого выражения бурно обсуждался в 1980-х годах. Какое выражение является правильным:

$$1/G = (h/2q^2 M) \quad \text{уравнение (11.2.1)}$$

или

$$1/G = 0 \quad \text{уравнение (11.2.5)?}$$

В настоящее время общепринятой считается точка зрения, согласно которой ненулевое сопротивление отвечает сопротивлению границы раздела между низкоразмерным проводником с  $M$  подзонами и двумя массивными контактами с очень большим числом подзон. В свете того, что измерения

в схеме с двумя контактами дают значение полного сопротивления, а не сопротивления одной лишь «структурой», наблюдается тенденция использования уравнения (11.2.1), понимая при этом, что оно включает в себя сопротивление границы раздела и сопротивление структуры. С другой стороны, при измерениях в схеме с четырьмя контактами результаты обычно интерпретируются на основе обсуждавшейся в главе 9 формулы Бюттикера для схемы измерений со многими контактами.

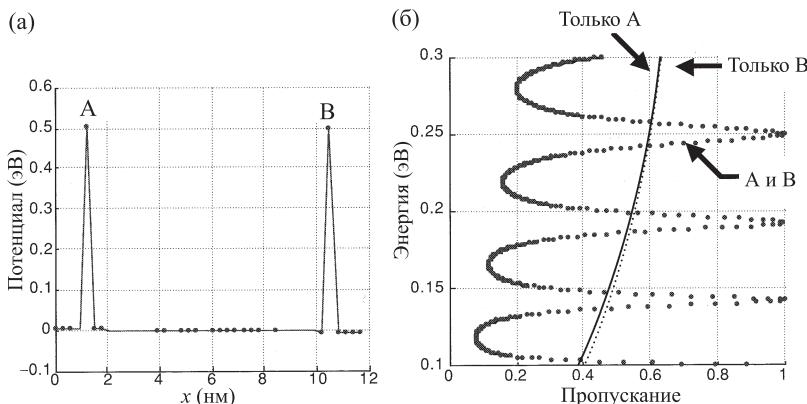


Рис. 11.2.2. (а) Короткая структура с двумя рассеивателями А и В. В расчетах была использована модель дискретной решетки с 40 узлами и шагом  $a = 3 \text{ \AA}$ , и приближение эффективной массы (с массой, равной массе свободного электрона  $m$ ). Предполагалось наличие слабого потенциала, линейно спадающего вдоль структуры от +5 до −5 мэВ. Каждый рассеиватель описывался как пик потенциала со значением 0.5 эВ на одном узле решетки. (б) Зависимость коэффициента пропускания от энергии, рассчитанная при наличии только одного рассеивателя (А или В), а также двух рассеивателей (А и В)

### 11.2.2. Когерентный транспорт (одна подзона)

Как было установлено, транспорт классических электронов описывается законом Ома вместе с последовательно включенным сопротивлением от границы раздела. Но будут ли эти выводы справедливы, если мы станем рассматривать электроны как квантовые частицы, подчиняющиеся волновому уравнению Шредингера? Ответ на этот вопрос оказывается отрицательным, в чем нетрудно убедиться, выполнив численный расчет сопротивления структуры, содержащей один рассеиватель (А или В), а затем два рассеивателя (А и В), как это показано на рис. 11.2.2.

В подобной задаче пропускание рассчитывается на базе уравнений, приведенных в начале параграфа 11.2 и с равными нулю слагаемыми, отвечающими за сбой фазы ( $\Sigma_s$  и  $\Sigma_s^{in}$ ). Поскольку рассматривается когерентный транспорт, коэффициент пропускания можно было бы рассчитать непосредственно из уравнения (9.1.10). Однако мы воспользуемся уравнением (11.1.12), поскольку оно применимо и для описания некогерентного транспорта, что позволяет использовать его в последующих примерах.

Важный вывод, который можно сделать на основании расчетов, представленных на рис. 11.2.2, состоит в том, что из квантовой задачи о прохождении через два рассеивателя не следуют простые правила, подобные тем, что получались для классических частиц из выражения (11.2.2). Основной причиной этого расхождения является интерференция, вызванная двумя рассеивателями. Если расстояние между ними равно половине длины волны, то интерференция приводит к резкому провалу на зависимости коэффициента пропускания от энергии. Если же расстояние между рассеивателями равно четверти длины волны, то отраженные волны интерферируют так, что гасят друг друга, что отражается в появлении пика на зависимости пропускания. Эти эффекты проявляются в сильных осцилляциях пропускания, как функции энергии, от которой зависит длина волны Де Броиля электронов. Очевидно, что при последовательном соединении двух секций нет оснований ожидать, что проводимость будет вести себя аддитивным образом, как этого требует закон Ома. В самом деле, в зависимости от положения уровня Ферми коэффициент прохождения через систему двух рассеивателей может даже *превзойти* свое значение для одиночного рассеивателя. Это никогда бы не произошло с классическими частицами, которые не могут иметь более высокую вероятность прохождения через два препятствия, чем через одно. Но этот эффект, обусловленный интерференцией, хорошо известен в физике волн: так, свет более эффективно проходит через линзу, если на ее поверхность нанесено антиотражающее покрытие. Нечто похожее наблюдается и для электронных волн: одна из последовательных секций сопротивления может рассматриваться как антиотражающее покрытие для другой секции, обеспечивая более высокое пропускание и, как следствие, более низкое сопротивление для двух секций по сравнению с одной секцией.

### 11.2.3. Когерентный транспорт (много подзон)

Электроны с заданной энергией имеют фиксированную длину волны только в чисто одномерной задаче. Напротив, в проводнике с большим числом мод электроны при данной энергии имеют различные значения  $k$  и,

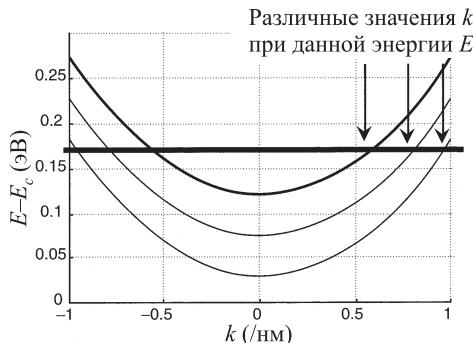


Рис. 11.2.3

стало быть, различные длины волн в продольном направлении, как это изображено на рис. 11.2.3.

В результате можно ожидать, что эффекты интерференции будут сглаживаться при наложении большого числа осцилляций, отвечающих различным длинам волн. Такую ситуацию иллюстрирует рис. 11.2.4, на котором представлена зависимость коэффициента пропускания через двумерную проволоку шириной 75 Å и длиной 200 Å, которая моделируется дискретной решеткой с расстоянием между узлами, равным 5 Å. В отсутствие рассеивателя коэффициент пропускания при данной энергии  $E$  равен числу распространяющихся мод  $M(E)$ , которое скачкообразно изменяется в указанном интервале энергий от четырех до шести. Коэффициент пропускания при наличии одного рассеивателя с поперечным профилем, показанным на рис. 11.2.5, возрастает монотонно с энергией. Наконец, мы можем найти коэффициент пропускания в квантизированной модели двух рассеивателей, используя уравнение (11.2.2) с  $T = \bar{T}/M$  и учитывая, что  $T_2 \approx T_1$ :

$$\frac{M}{\bar{T}} = \frac{2M}{\bar{T}_1} - 1 \rightarrow \bar{T} = \frac{\bar{T}_1}{2 - (\bar{T}_1/M)}.$$

Очевидно, что коэффициент пропускания в квантовой задаче с двумя рассеивателями испытывает флуктуации с амплитудой порядка единицы в окрестности квантизированного значения. Подобные флуктуации кондактанса узких проволок как функции напряжения на затворе (сдвигающим энергию Ферми относительно уровней структуры) наблюдались экспериментально и часто называются *универсальными флуктуациями кондактанса*. Флуктуации наблюдались также при изменении магнитного поля, которое меняет эффективную длину волны при данной энергии.

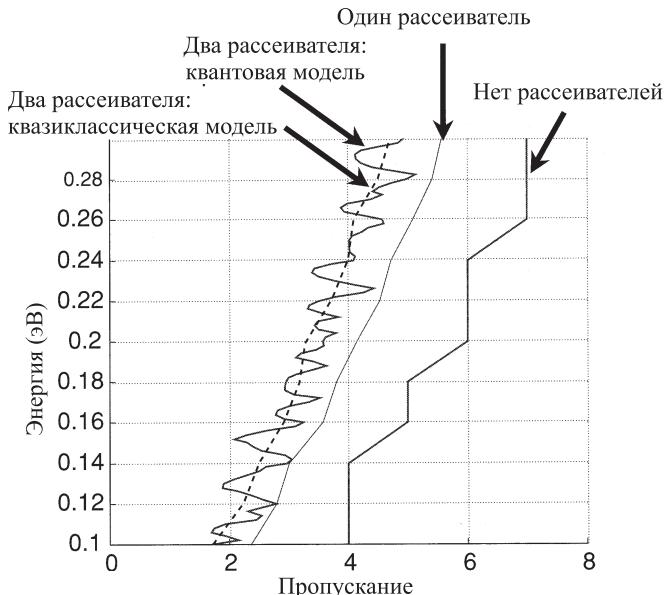


Рис. 11.2.4. Зависимость пропускания от энергии, рассчитанная для проволоки со многими подзонами в случаях, когда рассеиватели отсутствуют, а также когда имеется один или два рассеивателя. Каждый рассеиватель характеризуется максимумом потенциала в центре проволоки, равным 0.25 эВ

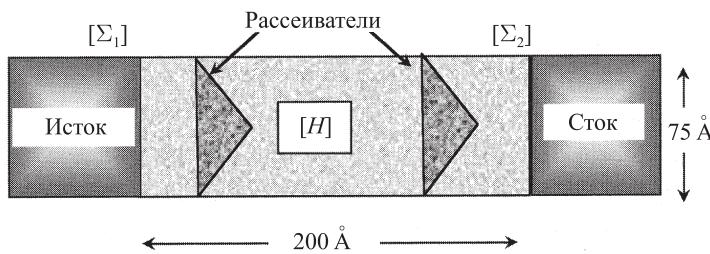


Рис. 11.2.5

Поскольку величина флуктуаций порядка единицы, читатель может задаться вопросом о том, что будет, если коэффициент пропускания упадет ниже единицы. Такая ситуация будет иметь место, если вероятность прохождения  $T$  на одну моду  $M$  меньше, чем  $1/M$ , так что полный коэффи-

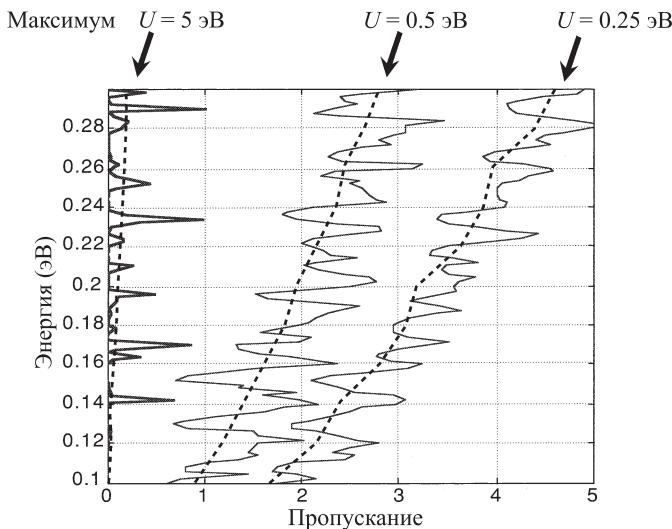


Рис. 11.2.6. Зависимость пропускания от энергии, рассчитанная для проволоки со многими энергетическими подзонами и двумя рассеивателями с максимальной амплитудой потенциала 5, 0.5 и 0.25 эВ. Пунктирные линии показывают квазиклассический результат для задачи с двумя рассеивателями, найденный из зависимости пропускания в задаче с одним рассеивателем

коэффициент пропускания  $\bar{T} = MT$  меньше единицы. На рис. 11.2.6 показаны результаты расчета коэффициентов пропускания для нескольких амплитуд потенциала рассеивателей, возрастающего от 0.25 до 5 эВ. Очевидно, что в последнем случае на графике проводимости появляются пики с большой амплитудой, разделенные интервалами энергии, где коэффициент пропускания становится пренебрежимо малым. Такое поведение говорит о *сильной локализации* электронных волновых функций.

По-иному выглядит коэффициент пропускания в присутствии слабых рассеивающих потенциалов, когда его квазиклассическое значение больше единицы. В этом случае квантовый коэффициент пропускания демонстрирует флуктуации с амплитудой порядка единицы вблизи квазиклассического значения. При наличии многих рассеивателей усредненное значение квантового коэффициента пропускания оказывается приблизительно на единицу *меньше*, чем квазиклассическое значение. Этот результат совсем не следует из рис. 11.2.6, где рассматривается случай лишь двух рассеивателей. В присутствии большого числа независимых рассеивателей рассеяние

назад возрастает для каждой моды на величину  $(1/M)$ . Это есть эффект положительной интерференции, приводящий к одинаковому уменьшению пропускания каждой моды и, как следствие, к общему уменьшению коэффициента пропускания на величину порядка единицы. Присутствие магнитного поля разрушает интерференцию и увеличивает пропускание, что наблюдалось экспериментально как отрицательное магнетосопротивление (т. е. уменьшение сопротивления в магнитном поле). В связи с этим говорят о подавлении в магнитном поле эффекта *слабой локализации*. Отметим, что фундаментальный эффект, связанный с когерентным возрастанием рассения назад, наблюдался и при изучении электромагнитных волн в ряде задач, не имеющих отношения к электронному транспорту.

#### 11.2.4. Квантовый транспорт с дефазировкой

Можно подвести некоторые итоги проведенного анализа: по мере увеличения длины проводника значение квазиклассического коэффициента пропускания будет уменьшаться в соответствии с законом Ома, а квантовый коэффициент пропускания будет испытывать флуктуации вблизи квазиклассического значения с амплитудой порядка единицы. С увеличением длины проводника мы постепенно придем к тому, что квазиклассический коэффициент пропускания станет меньше единицы (см. уравнение (11.2.3)):

$$\bar{T} \approx \frac{M\Lambda}{L + \Lambda} < 1, \quad \text{если } L > M\Lambda.$$

Квантовый коэффициент пропускания в проводнике с длиной, превосходящей длину локализации ( $\sim M\Lambda$ ), будет испытывать большие флуктуации, характерные для режима сильной локализации. Можно ожидать, что даже медная проволока достаточно большой длины рано или поздно войдет в этот режим и перестанет подчиняться правилу, напоминающему закон Ома! Реально, однако, этого не происходит, и было бы интересно понять, почему.

Причина состоит в том, что наши наблюдения справедливы для проводников, где транспорт имеет когерентный характер, и не происходит нарушения фазы, размыивающего квантовые интерференционные эффекты. В проволоке сильная локализация будет наблюдаться лишь в случае, если длина локализации  $M\Lambda$  станет короче, чем расстояние, на котором происходит нарушение фазы. Поскольку последняя обычно весьма мала, особенно при комнатной температуре, то маловероятно, что в медной проволоке с очень большим числом мод  $M$  такой режим будет когда-либо возможен.

На рис. 11.2.7 показана зависимость от энергии эффективного коэффициента пропускания для одномерной проволоки с двумя когерентными рассеивателями как при отсутствии дефазировки, так и при ее наличии. На рис. 11.2.8 показано, как в одномерной проволоке с рассеивателями, разрушающими фазу, зависимость сопротивления от длины образца напоминает закон Ома. Разумеется, в этом предельном случае не требуется обращаться к полной квантовой модели транспорта и можно использовать квазиклассическую модель, вовсе не учитывающую эффекты интерференции и рассматривающую электроны как частицы. В общем случае, однако, в системе будут представлены как когерентные, так и нарушающие фазу процессы рассеяния, и тогда развитая в параграфе 11.1 модель позволит включить оба этих механизма.

При выполнении расчетов предполагалось, что рассеиватели, вносящие дефазировку, уносят пренебрежимо малую долю энергии ( $\hbar\omega \rightarrow 0$ ), так что допустимо пользоваться упрощенным вариантом уравнений (сравните с уравнением (11.1.10))

$$\Sigma_s^{\text{in}}(E) = D_0[G^n(E)] \quad \text{и} \quad \Sigma_s(E) = D_0[G(E)], \quad (11.2.6)$$

чтобы рассчитать собственно-энергетическую функцию и функцию рассеяния для входящего тока. Предполагается, что рассеяние описывается диагональной матрицей в реальном пространстве и не зависит от координат, т. е.  $D_0$  является постоянной величиной, для которой принято значение  $0.01 \text{ эВ}^2$  на рис. 11.2.7 и  $0.05 \text{ эВ}^2$  на рис. 11.2.8.

### 11.3. Где происходит выделение тепла?

Мы видели, что в присутствии процессов дефазировки сопротивление образца в соответствии с законом Ома растет линейно с его длиной (рис. 11.2.8). Но по мере того, как длина проводника стремится к нулю, т. е. при переходе к баллистическому режиму, сопротивление стремится к постоянному значению, равному сопротивлению границы раздела между низкоразмерным проводником и трехмерным контактом. Возникает вопрос, где происходит выделение джоулева тепла  $I^2 R$ ?

Ответ на этот вопрос зависит от природы процесса рассеяния. Если сопротивление связано с потерей электронами импульса в результате рассеивания, то выделение джоулева тепла вызвано потерей энергии. К примеру, в расчетах, представленных на рис. 11.2.7 и 11.2.8, учитывались упругие процессы дефазировки и не принимались во внимание потери энергии.

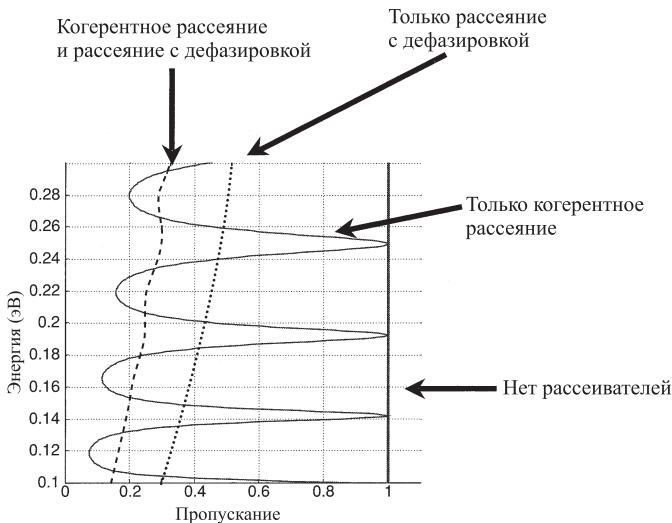


Рис. 11.2.7. Зависимость коэффициента пропускания от энергии для одномерной проволоки с различными типами рассеяния, указанными на рисунке

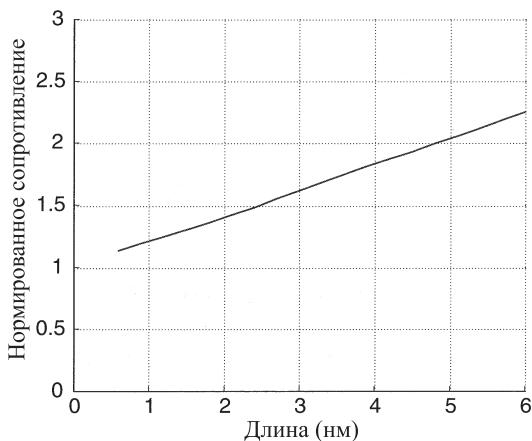


Рис. 11.2.8. Нормированное сопротивление (обратное коэффициенту пропускания) как функция длины образца для одномерной проволоки с учетом только процессов рассеяния, идущих с дефазировкой

Последнее означает, что в рамках данной модели внутри структуры энергия вообще не диссирирует. Тем не менее, упругие процессы вносят свой вклад в омическое сопротивление, поскольку с ним связана потеря импульса электронов, а не потеря фазы. В самом деле, в небольших проводниках существенная доля джоулева тепла  $I^2R$ , связываемого с сопротивлением  $R$ , может диссирировать в контактах, а не в самом проводнике. В пользу этого утверждения говорит и тот факт, что в реальных экспериментах через небольшие проводники удается «пропускать» токи, превышающие те, что достаточны для джоулева нагрева и разрушения в случае, когда все тепло выделяется внутри проводника.

Тот факт, что упругое рассеяние не связано с какой-либо потерей энергии, находит свое подтверждение в том, что нормированный на энергию ток  $\tilde{I}_i(E)$  (см. уравнение (11.1.6)) имеет одинаковое значение в контактах истока и стока, как это показано на рис. 11.3.1. Можно утверждать, что поток энергии в каждом из контактов определяется выражением (сравните с выражением (11.1.6))

$$I_{E,i} = \int_{-\infty}^{+\infty} dE E \tilde{I}_i(E) / 2\pi\hbar, \quad (11.3.1)$$

поэтому если мощность ( $P_d$ ) диссириует внутри структуры, то эта диссириация должна отражаться в разнице потоков энергии в контактах истока и стока:

$$P_d = I_{E,\text{сток}} - I_{E,\text{исток}}. \quad (11.3.2)$$

Поскольку закон сохранения требует равенства токов в контактах истока и стока, потоки энергии могут различаться лишь, когда их распределение по энергии различно. Очевидно, что в структуре, представленной на рис. 11.3.1, нет диссириации, поскольку распределение токов по энергии в контактах истока и стока одинаково. Но если предположить, что рассеяние происходит на фонах с энергией  $\hbar\omega = 20$  мэВ, то из-за диссириации энергии в виде джоулева тепла  $I^2R$  распределение тока по энергии в истоке и стоке будет различным. Электроны в среднем входят в контакт истока при более высокой энергии, чем та, с которой они покидают структуру через сток, как это показано на рис. 11.3.2.

### 11.3.1. Эффект «Пельтье»

Интересно отметить, что мощность диссириует не обязательно во всех точках структуры. В неоднородной структуре могут существовать локальные области, где энергия поглощается электронами и образец локально

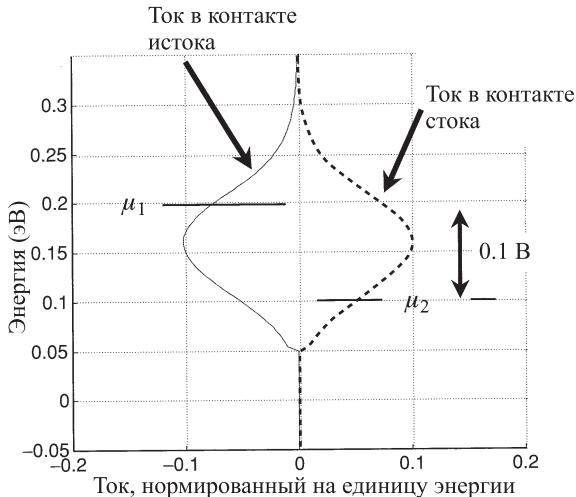


Рис. 11.3.1. Нормированный на единицу энергии ток  $\tilde{I}_i(E)$  в одномерной проволоке с упругим разрушающим фазу рассеянием

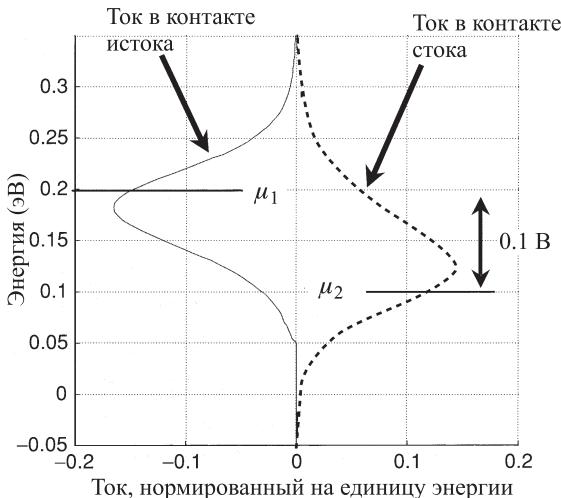


Рис. 11.3.2. Нормированный на единицу энергии ток  $\tilde{I}_i(E)$  в одномерной проволоке с неупругим рассеянием при наличии фононов с энергией  $\hbar\omega = 20 \text{ мэВ}$  (значение  $D_0 = 0.1 \text{ эВ}^2$ ). Из-за релаксации энергии внутри структуры ток в контакте стока отвечает более низкой энергии, чем ток истока

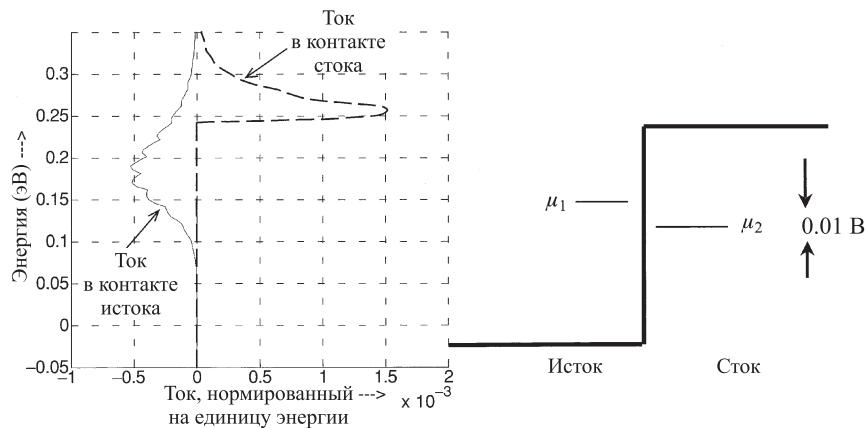


Рис. 11.3.3. Нормированный на единицу энергии  $E$  ток в одномерной проволоке с неупругим рассеянием на фононах с энергией  $\hbar\omega = 20$  мэВ (значение  $D_0 = 0.1$  эВ<sup>2</sup>) и скачком потенциала в середине структуры. Ток стока отвечает более высокой энергии по сравнению с током истока, что приводит к охлаждению структуры протекающим током

**охлаждается.** Рассмотрим, к примеру, задачу о протекании тока в одномерной структуре со скачком потенциала в центре (рис. 11.3.3). Тогда электроны в области стока будут иметь более высокую энергию, чем вблизи истока. В результате, как это следует из уравнения (11.3.2), данная структура будет охлаждаться при протекании тока. Этот эффект можно назвать микроскопическим аналогом хорошо известного эффекта Пельтье, где при протекании электрического тока охлаждается один контакт за счет нагревания другого. Так, если ток протекает через потенциальный барьер, где скачок потенциала вверх сменяется скачком вниз, то область, где происходит скачок вверх, будет охлаждаться, а область со скачком вниз, соответственно, нагреваться. Этот вид энергетического обмена является обратимым, поскольку он линеен по току, в отличие от необратимого джоулева нагрева, который пропорционален квадрату тока [30].

## 11.4. Где происходит падение напряжения?

В параграфе 1.1 я подчеркнул необходимость самосогласованного подхода к расчетам квантового транспорта с использованием уравнения Пуас-

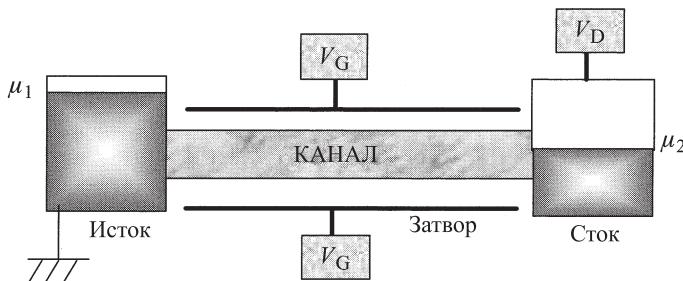


Рис. 11.4.1. Нанотранзистор, состоящий из квантовой проволоки, формирующей канал, и окружающего канал коаксиального затвора, используемого для управления электронами, входящими в канал

сона (см. рис. 11.1.2) и определения самосогласованного потенциала  $U$ , учитывавшего влияние других электронов. Этот вопрос приобретает особую важность при вычислении тока при «большом» приложенном напряжении, поскольку форма вольт-амперной характеристики сильно зависит от профиля потенциала, а также от распределения напряжения внутри канала. К примеру, при определении максимального тока транзистора (т. е. тока в состоянии ВКЛ) важно знать, где расположены области наиболее сильного падения напряжения.

Рассмотрим «нанотранзистор», состоящий из узкой квантовой проволоки, обозначенной на рис. 11.4.1 как «канал» радиуса  $a$ , который окружен коаксиальным затвором радиуса  $b$ . Ясно, что затвор, как это обсуждалось в главе 7, используется для изменения концентрации электронов в канале. Пусть электроны в канале принадлежат одной энергетической подзоне с параболическим законом дисперсии

$$E = E_c + (\hbar^2 k^2 / 2m). \quad (11.4.1)$$

В состоянии равновесия, когда  $\mu_2 = \mu_1$ , а температура низка ( $T \rightarrow 0$  K), электронная плотность на единицу длины в канале может быть записана следующим образом с учетом дополнительного множителя, учитывающего двукратное вырождение по спину (см. таблицу 6.2.1):

$$n_L = 2\sqrt{2m(\mu_1 - E_c)/\pi\hbar}. \quad (11.4.2)$$

Если увеличить положительное напряжение на затворе  $V_G$ , то это приведет к увеличению числа электронов в канале, в то время как отрицательное напряжение  $V_G$  вызовет обеднение канала. Этот вопрос обсуждался

нами в параграфе 7.3, речь шла о плоском двумерном проводнике, а теперь мы рассматриваем цилиндрический одномерный проводник. В главе 7 рассматривался только равновесный случай с  $\mu_2 = \mu_1$ . Задача, которую мы будем обсуждать сейчас, является неравновесной. Пусть к контакту стока приложено напряжение  $V_D$ , и поэтому  $\mu_2 = \mu_1 - qV_D$ . Чему в этой ситуации будет равен ток  $I$ ? С формальной точки зрения мы можем найти его, следя самосогласованной процедуре, изображенной на рис. 11.1.2. Результаты численного расчета будут представлены позднее на рис. 11.4.4, а сейчас попробуем понять физический смысл описываемых процессов.

**Баллистический нанотранзистор.** Рассмотрим, прежде всего, баллистический транзистор (без рассеяния), имеющий идеальные контакты. Если контакты хорошо проводящие и в них нет рассеяния, то можно ожидать, что проводимость в области малых напряжений будет равна удвоенному (из-за наличия спина) значению кванта проводимости:

$$I = (2q^2/h)V_D. \quad (11.4.3a)$$

Состояния с положительной проекцией  $+k$  заполняются электронами из левого контакта с электрохимическим потенциалом  $\mu_1$ , а состояния с отрицательной проекцией  $-k$  заполняются электронами из правого контакта с электрохимическим потенциалом  $\mu_2$ . В интервале энергий между  $\mu_1$  и  $\mu_2$  (плюс добавка в несколько  $k_B T$  на каждой границе) состояния с  $+k$  почти заполнены и переносят ток, который не компенсируется током состояний с  $-k$ , поскольку они почти пустые. В силу того, что в одномерной проволоке ток на единицу энергии равен  $(2q/h)$ , полный ток, протекающий через структуру, будет равен величине  $(2q/h)$ , умноженной на величину интервала энергии  $(\mu_1 - \mu_2)$ . В соответствии с упомянутым выше будем иметь  $(2q^2/h)V_D$ .

Как только химический потенциал  $\mu_2$  опускается ниже дна зоны, с энергией  $E_c$ , ток перестает возрастать. Поэтому можно ожидать, что в состоянии ВКЛ максимальный ток транзистора равен

$$I_{\text{ВКЛ}}^L = (2q^2/h)(\mu_1 - E_c). \quad (11.4.3b)$$

Уравнения (11.4.3a), (11.4.3b) наводят на мысль о том, что ток должен сначала возрастать линейно с напряжением, а затем, как только напряжение достигнет величины  $(\mu_1 - E_c)$ , выходить на насыщение. Действительно, ток растет линейно с напряжением, а затем выходит на насыщение. Но при определенных электростатических параметрах системы этот ток в состоянии ВКЛ может быть в *четыре* раза больше максимального значения, определяемого из уравнения (11.4.3b), которому приписан индекс «L»,

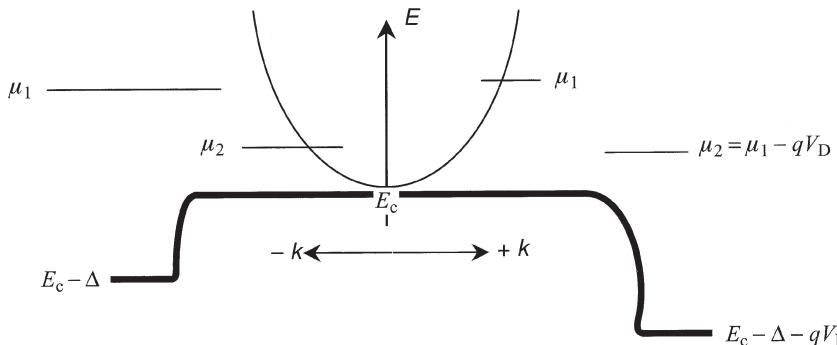


Рис. 11.4.2. Приложенное напряжение понижает энергию уровней и электрохимический потенциал в области стока. Состояния  $+k$  заполняются из контакта истока вплоть до энергии, равной  $\mu_1$ , а состояния  $-k$  заполняются со стороны стока до энергии  $\mu_2$ , что создает ненулевой полный ток, как это обсуждалось в главе 6 (см. параграф 6.3)

означающий приближение Лапласа. Сейчас я объясню, что означает этот символ и чему соответствует другой предел.

Результат, который следует из уравнения (11.4.3b), основан на схеме, представленной на рис. 11.4.2. Здесь предполагается, что единственный эффект, от возрастающего напряжения стока, состоит в понижении величины  $\mu_2$ , в то время как положения уровней энергии в канале остаются неизменными относительно электрохимического потенциала истока. Однако при определенных электростатических параметрах возможна ситуация, при которой потенциальная энергия  $U$  канала понизится на величину  $-qV_D$ , составляющую некоторую часть потенциала стока. Это приведет к понижению дна зоны и увеличению тока транзистора в состоянии ВКЛ на величину

$$I_{\text{ВКЛ}} = (2q^2/h)(\mu_1 - E_c - U). \quad (11.4.4)$$

Обратите внимание на то, что  $U$  является отрицательной величиной при положительных напряжениях на стоке.

Здесь предполагается, что  $U$  остается меньше, чем  $\Delta$  (см. рис. 11.4.2). Таким образом, при оценке тока в состоянии ВКЛ наиболее важным является следующий вопрос: как происходит падение напряжения? Если исток находится при нулевом потенциале, а сток — при потенциале  $-qV_D$ , то как ведет себя потенциальная энергия внутри канала?

В общем случае определяем потенциал канала путем самосогласованного решения задачи электростатики и транспорта. Для задачи, которую мы

рассматриваем, электростатический потенциал записывается в виде

$$U = U_L + (q^2/C_E)\delta n_L, \quad (11.4.5)$$

где  $C_E$  представляет собой емкость на единицу длины коаксиального конденсатора, внутренний и внешний радиусы которого равны  $a$  и  $b$  соответственно:

$$C_E = 2\pi\epsilon_r\epsilon_0 / \ln(b/a). \quad (11.4.6)$$

Здесь  $\delta n_L$  есть изменение электронной плотности на единицу длины, а  $U_L$  — решение, в котором  $\delta n_L = 0$ , т. е. решение уравнения Лапласа, показанное на рис. 11.4.3. Поскольку разность потенциалов прикладывается только к стоку, а не к затвору, потенциала, отвечающий решению Лапласа, выглядит так, как показано на рис. 11.4.3. Фактически он равен потенциалу истока на всем протяжении канала, а затем быстро спадает до значения потенциала стока на другом конце канала. Насколько быстрым будет этот спад, зависит от того, как близок затвор к каналу.

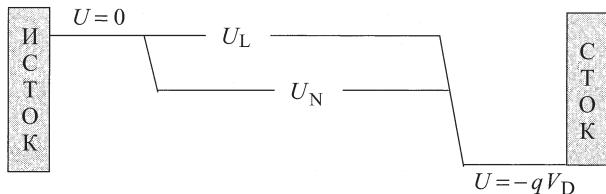


Рис. 11.4.3. То же, что и на рис. 11.4.2, но за вычетом профиля зоны проводимости. Здесь показана только зависимость потенциала  $U$  от приложенного напряжения. Смысл лапласовского потенциала  $U_L$  и нейтрального потенциал  $U_N$  разъясняется в тексте

Если емкость  $C_E$  велика, и вторым слагаемым в уравнении (11.4.5) можно пренебречь, то истинное распределение потенциала внутри канала ближе к электростатическому пределу Лапласа. В этом случае можно считать, что диэлектрическая постоянная очень большая:  $\epsilon_r = 100$  (см. уравнение (11.4.6), в котором принято  $b = 2a$ ). Но если  $\epsilon_r = 2$ , то ток увеличивается почти в два раза (см. рис. 11.4.4).

Если емкость еще меньше, то, в принципе, можно прийти к другому предельному случаю, когда  $\delta n_L \rightarrow 0$ , а второе слагаемое является конечным. Это т. н. нейтральный предельный случай, которому отвечает нейтральный потенциал  $U_N$  (см. параграф 7.3). Как определить потенциал  $U_N$ ? Для этого нужно записать электронную плотность в состоянии транзистора

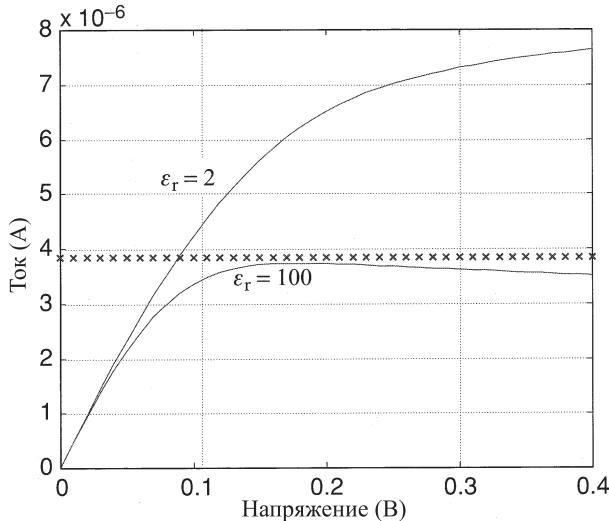


Рис. 11.4.4. Зависимость тока  $I$  от напряжения стока  $V_D$  в транзисторе с баллистической квантовой проволокой в качестве канала, построенная для различных значений емкости  $C_E$ . В лапласовском пределе  $U = U_L$  зависимость тока от напряжения показана крестиками при  $U = U_L$  (рисунок С. Салахуддина)

ВКЛ в виде

$$[n_L]_{\text{ВКЛ}} = \sqrt{2m(\mu_1 - E_c - U)} / \pi \hbar, \quad (11.4.7)$$

где учтено, что отрицательная потенциальная энергия  $U$  смешает зону вниз на величину  $E_c + U$ . При этом теряется множитель, равный двум, поскольку лишь половина состояний (с проекцией импульса  $+k$ ) являются занятыми. Вычитая (11.4.7) из выражения (11.4.2), запишем изменение электронной плотности в виде

$$\delta n_L = (\sqrt{2m}/\pi \hbar) \left( 2\sqrt{\mu_1 - E_c} - \sqrt{\mu_1 - E_c - U} \right). \quad (11.4.8)$$

Полагая  $\delta n_L = 0$ , находим нейтральный потенциал:

$$2\sqrt{\mu_1 - E_c} = \sqrt{\mu_1 - E_c - U} \rightarrow \mu_1 - E_c - U_N = 4(\mu_1 - E_c), \quad (11.4.9)$$

откуда следует, что ток транзистора в состоянии ВКЛ, найденный в нейтральном пределе из уравнения (11.4.4), в четыре раза превышает значение,

ожидаемое в лапласовском пределе (сравните с выражением (11.4.3b)):

$$I_{\text{ВКЛ}}^{(N)} = (2q^2/h)(\mu_1 - E_c - U_N) = 4I_{\text{ВКЛ}}^{(L)}. \quad (11.4.10)$$

Должен отметить, что нейтральный предел для значения тока в состоянии ВКЛ не всегда превышает предел Лапласа в четыре раза. Множитель, равный четырем, появляется только в рассматриваемой нами одномерной задаче и связан с тем, что электронная плотность пропорциональна квадратному корню из энергии (см. уравнение (11.4.2)). Для двумерного проводника электронная плотность растет линейно с энергией, и здесь можно показать, что ток в нейтральном пределе превысит ток в пределе Лапласа в два раза. Важно то, что существуют два этих предельных значения — предел Лапласа и нейтральный предел, а фактическое значение тока может располагаться в любой точке между ними в зависимости от величины емкости  $C_E$ .

**Электростатические граничные условия.** На рис. 11.4.3 показан сформированный внешним источником напряжения профиль потенциала, стремящийся к асимптотическим значениям 0 и  $-qV_D$  на контактах истока и стока. Такое поведение действительно имеет место, когда эти области контактов достаточно длинные. Однако, если транспорт в граничных областях является баллистическим, то потенциал в них может и не достигать асимптотических значений вне зависимости от их длины.

Причина такого поведения в том, что в этих проводящих граничных областях потенциал  $U$  приближается к нейтральному значению  $U_N$ , которое обеспечивает выполнение условия  $\delta n_L = 0$ . Рассмотрим, к примеру, область вблизи истока. Если бы потенциал  $U$  в ней был равен нулю, то величина  $\delta n_L$  была бы отрицательной, поскольку часть плотности состояний в этой области, определяемая электрохимическим потенциалом  $\mu_2$  стока, теперь оказывается занятой. Эта часть описывается частной спектральной функцией  $[A_2]$ , которая была определена в главе 9. Чтобы сохранить значение  $\delta n_L = 0$ , нейтральный потенциал в этой области принимает отрицательное значение (см. рис. 11.4.5). Ситуация не изменится, если мы просто удлиним краевые области контактов. Как следствие, нейтральный потенциал в области истока будет меньше, чем его асимптотическое значение, равное нулю. Применяя аналогичные аргументы, можно показать, что в области стока потенциал будет иметь значение большее, чем асимптотическое  $-qV_D$ .

Потенциал  $U$  будет принимать правильные асимптотические значения  $\pm qV/2$ , когда лишь малая доля плотности состояний (или спектральная функция) на одном конце канала будет подвержена влиянию контакта на



Рис. 11.4.5. Профиль нейтрального потенциала: в областях контактов потенциал не стремится к правильному асимптотическому значению, равному нулю слева и  $-qV_D$  справа, поскольку часть плотности состояний в этих областях контролируется контактом на противоположном конце канала

другом конце. Такая ситуация должна иметь место при достаточно сильном рассеянии внутри структуры, либо когда число мод в контактах значительно превышает число мод в канале, т. е.  $M \gg N$  (см. рис. 11.4.6).

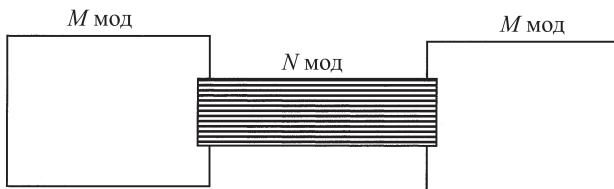


Рис. 11.4.6

Из приведенного обсуждения следует, что при моделировании квази-баллистических структур в отсутствие существенного эффекта геометрического уменьшения числа мод на контактах не следует фиксировать значение потенциала на двух концах, приравнивая его к асимптотическим значениям, как это делалось при расчете емкости (см. уравнение (7.2.17)). Одна из возможностей состоит в использовании граничных условий для уравнения Пуассона с нулевым полем и последующим самосогласованным расчетом потенциал  $U$ .

Пространственный профиль нейтрального потенциала  $U_N$ , который может отличаться от профиля истинного потенциала  $U$ , можно рассматривать как индикатор пространственного распределения сопротивления. Он (нейтральный потенциал), естественно, остается постоянным в любой баллистической области, что говорит о ее нулевом сопротивлении. Из

рис. 11.4.5 следует, что только часть приложенного напряжения  $V_D$  падает в области между двумя концами одномерной структуры; и это говорит о том, что сопротивление лишь частично формируется в канале, а оставшаяся его часть должна быть связана влиянием границы раздела между узкими областями канала и широкими областями контактов.

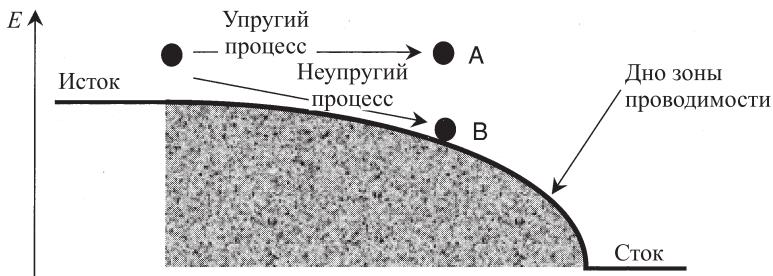


Рис. 11.4.7. Как только электрон потеряет в неупругих процессах значительную часть энергии (точка В), он не сможет повернуть назад к истоку. Вследствие этого ток не ослабляется при дальнейшем рассеянии. Но если электрон еще не потерял достаточноную энергию (точка А), остается существенным рассеяние назад в область контакта источника

**Нанотранзистор с рассеянием.** Мы предполагаем, что ток в состоянии ВКЛ, приобретая множитель  $T = \Lambda / (L + \Lambda)$ , уменьшается в присутствии рассеяния. Дело в том, что коэффициент пропускания, зависящий от соотношения между длиной канала и длиной свободного пробега, становится меньше единицы (см. уравнение (11.2.3)). На практике, однако, если процессы рассеяния имеют неупругий характер, ток в состоянии ВКЛ больше значения, следующего из (11.4.10).

Чтобы понять, почему так происходит, заметим, что неупругие процессы приводят к уменьшению энергии электронов, приходящих от истока к стоку (см. рис. 11.4.7). Как только электроны потеряют достаточно большую часть энергии (чему соответствует точка В на рисунке), их движения уже не смогут переориентироваться в сторону истока, поскольку в истоке более нет свободных состояний со столь низкой энергией. Как следствие, электроны продолжают движение к контакту стока и ток, вследствие такого рассеяния, не уменьшается. Но если рассеяние имеет упругий характер, и нет релаксации энергии (это соответствует точке А на рисунке), то электроны могут и повернуть назад в исток, уменьшая тем самым величину тока. Эти процессы могут быть приближенно описаны путем замены

длины структуры  $L$  в уравнении (11.2.3) на расстояние, которое электрон проходит до момента неупругой релаксации энергии  $L_i$ , т. е. до уменьшения энергии на величину порядка нескольких  $k_B T$ . Эта величина должна быть достаточно велика для того, чтобы исключить возможность движения назад. Следовательно,

$$T = \Lambda / (L_i + \Lambda). \quad (11.4.11)$$

В этих условиях фактическое значение тока в состоянии ВКЛ оказывается намного больше, чем можно было бы ожидать. Отметим, что в двухмерных структурах даже чисто упругое рассеяние приводит к аналогичному возрастанию тока в состоянии ВКЛ, поскольку оно приводит к релаксации *продольной* части кинетической энергии (отвечающей движению от истока к стоку), оставляя неизменной сумму энергий продольного и поперечного движений. В самом деле, в серийно выпускаемых транзисторах в состоянии ВКЛ на протяжении многих лет наблюдаются токи, величина которых составляет приблизительно 50 процентов от соответствующих баллистических значений, даже если длина структуры превышает десять длин свободного пробега. Причина состоит в том, что длина релаксации энергии  $L_i$  стремится к длине свободного пробега  $\Lambda$ , и в этих условиях вероятность прохождения становится приблизительно равной 0.5 независимо от фактической длины канала.

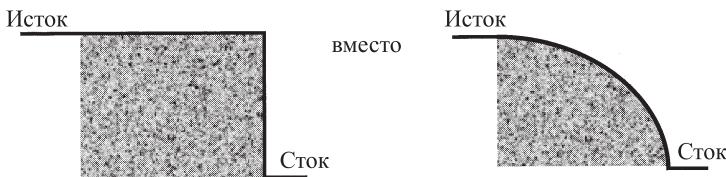


Рис. 11.4.8

Причина, по которой обсуждается этот вопрос, связана с тем, что он демонстрирует, насколько важен самосогласованный расчет профиля потенциала при больших напряжениях. Действительно, описанный эффект не может наблюдаться, если потенциал имеет профиль, показанный на рис. 11.4.8, и в этом случае область релаксации энергии была бы слишком малой. Следует ожидать, что фактическая величина обсуждаемого эффекта в реальной структуре будет зависеть от профиля потенциала, рассчитанного самосогласованно, как показано на рис. 11.1.2. Это обстоятельство часто игнорируют, поскольку отмеченная на схеме рис. 11.1.2 часть расчетов, от-

носящихся к транспорту, требует больших усилий по сравнению с блоком решения уравнения Пуассона. Подводя итог, хотелось бы еще раз напомнить, что в задачах, подобных расчету тока в состоянии транзистора ВКЛ, именно блок, связанный с решением уравнения Пуассона, определяет ключевые моменты физической картины, а расчет транспорта — это всего лишь техническая процедура!

## Упражнения

**У.11.1** (а) Пусть имеется короткий одномерный канал с двумя рассеивателями А и В, моделируемый на дискретной решетке с 40 узлами и постоянной решетки  $a = 3 \text{ \AA}$ . Предполагается, что к структуре приложена малая разность потенциалов, которая понижает потенциал от +5 до -5 мэВ вдоль структуры. Каждый рассеиватель описывается точечным скачком потенциала на +0.5 эВ на соответствующем узле решетки. Рассчитайте зависимость коэффициента пропускания от энергии в случае, когда присутствует только один рассеиватель (А или В), а также оба рассеивателя (А и В), и сравните результаты с рис. 11.2.2. (б) Повторите проделанный расчет с учетом процессов упругого рассеяния и с дефазировкой в соответствии с уравнением (11.2.6). Сравните результат с рис. 11.2.7. (в) Постройте зависимость обратного коэффициента пропускания (при фиксированном значении энергии 0.1 эВ) от длины одномерной проволоки при учете только упругих процессов, идущих с нарушением фазы, и сравните результат с рис. 11.2.8.

**У.11.2.** (а) Рассматривая канал ширины 75  $\text{\AA}$ , содержащий большое число мод, вычислите коэффициент прохождения через один рассеиватель и через два рассеивателя, предполагая, что каждый рассеиватель в поперечном направлении описывается треугольным потенциалом с максимумом 0.25 эВ в центре, а в продольном направлении характеризуется точечным скачком потенциала на одном узле. Получите также квазиклассический результат для случая двух рассеивателей, считая известным результат для одного рассеивателя, и сравните найденную зависимость с рис. 11.2.4. (б) Повторите расчеты для различных значений рассеивающего потенциала и сравните результат с рис. 11.2.6.

**У.11.3.** (а) Рассматривая одномерный канал с учетом упругих процессов, идущих с нарушением фазы, постройте зависимость тока на единицу энергии, как показано на рис. 11.3.1. (б) Повторите расчет для канала, в котором имеется неупругое рассеяние, и сравните результат

с рис. 11.3.2. (в) Проделайте тот же расчет для канала с неупругим рассеянием, в середине которого имеется скачок потенциала, и сравните результат с рис. 11.3.3.

**У.11.4.** Рассчитайте самосогласованно вольт-амперную характеристику зависимости тока  $I$  от напряжения стока  $V_D$  для транзистора с каналом из баллистической квантовой проволоки и сравните результат с рис. 11.4.4.

**У.11.5. Туннелирование в присутствии дефазировки.** Рассчитайте зависимость величины обратного коэффициента пропускания при низких значениях напряжения стока при  $E = \mu$  для проводника, характеризующегося равновесным значением электрохимического потенциала  $\mu$ , расположенным при 0.1 эВ. Считайте, что дно зоны проводимости в контакте находится при 0 эВ, а в канале при 0.5 эВ, как это показано на рис. У.11.5а.

Проводимость канала определяется туннельными процессами и поэтому, если нет дефазировки, обратный коэффициент пропускания зависит от длины барьера  $L$  экспоненциально. Постройте график логарифма обратного коэффициента пропускания (от нормированного сопротивления) как функцию длины барьера в интервале 1 нм  $< L <$  7 нм при двух различных амплитудах рассеивающего потенциала  $D_0 = 2$  эВ<sup>2</sup> и  $D_0 = 3$  эВ<sup>2</sup>. У вас должен получиться график, подобный приведенному на рис. У.11.5б. Обратите внимание на то, что при большей амплитуде рассеивающего потенциала ожидаемая экспоненциальная зависимость сопротивления (т. е. линейная зависимость логарифма сопротивления) не имеет места (см., к примеру, [41]).

**У.11.6. Туннельная спектроскопия с неупругим рассеянием электронов (НЕТС).** В конце параграфа 11.1 мы рассмотрели простой пример системы с одним уровнем и упругими процессами дефазировки. Полезно рассмотреть вновь случай одноуровневой системы, но уже при наличии неупругого рассеяния. Это более сложная задача, поскольку здесь взаимодействуют состояния с различной энергией. Основные уравнения по-прежнему имеют тот же вид, что и в параграфе 11.1, где матрицы становятся скалярами (матрицами размером  $(1 \times 1)$ ). Матрицы, описывающие рассеяние входящих электронов, и матрицы уширения определяются следующими выражениями:

$$\gamma_s^{\text{in}}(E) = \int d(\hbar\omega) D^{\text{ph}}(\hbar\omega) G^n(E + \hbar\omega)$$

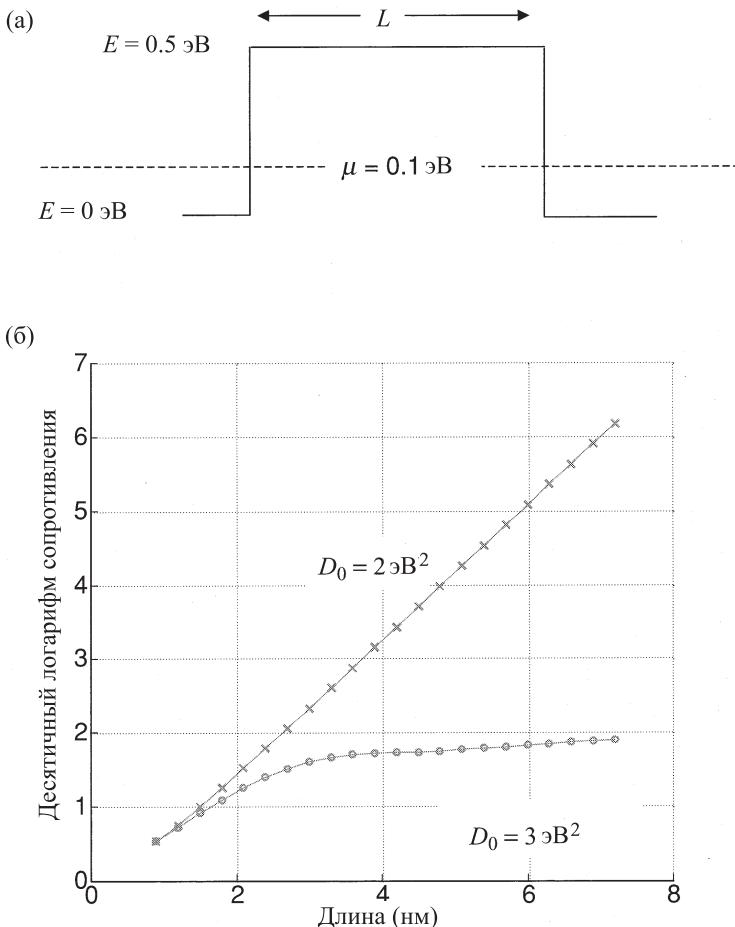


Рис. У.11.5

и

$$\Gamma_s(E) = \int d(\hbar\omega) D^{\text{ph}}(\hbar\omega) (G^n(E + \hbar\omega) + G^p(E - \hbar\omega)),$$

где «фононная» спектральная функция может быть записана в виде суммы эмиссионного слагаемого (с положительными частотами) и слагаемого,

отвечающего поглощению (отрицательные частоты):

$$D^{\text{ph}}(\hbar\omega) = \sum_i D_i [(N_i + 1)\delta(\hbar\omega - \hbar\omega_i) + N_i\delta(\hbar\omega + \hbar\omega_i)],$$

где  $N_i$  представляет собой число фононов с частотой  $\omega_i$  и амплитудой взаимодействия  $D_i$ .

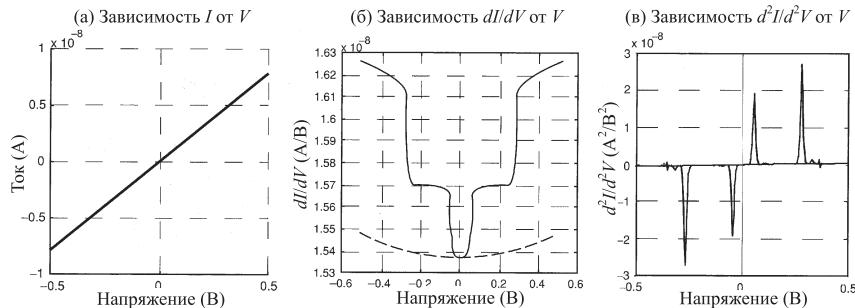


Рис. У.11.6. (а) Ток ( $I$ ), (б) проводимость ( $dI/dV$ ) и (в) вторая производная  $d^2I/dV^2$  как функции напряжения, рассчитанные без учета рассеяния на фонах (пунктирные линии) и с учетом этого рассеяния (сплошные линии) при двух различных частотах, которым отвечают слегка различные константы взаимодействия ( $D_1 = 0.5$ ,  $\hbar\omega_1 = 0.075$  эВ и  $D_2 = 0.7$ ,  $\hbar\omega_2 = 0.275$  эВ). Рисунок воспроизведен с разрешения С. Датта (2004) [15]

Рассчитайте зависимость тока от напряжения без учета самосогласованного потенциала, предполагая, что значение энергии уровня  $\varepsilon = 5$  эВ располагается гораздо выше равновесного электрохимического потенциала  $\mu = 0$ , так что протекание тока носит туннельный характер. У вас должен получиться график, показанный на рис. У.11.6.

Токи, рассчитанные в отсутствии рассеяния на фонах (все  $D_i = 0$ ) и при учете этого рассеяния ( $D_1 = 0.5$ ,  $\hbar\omega_1 = 0.075$  эВ и  $D_2 = 0.7$ ,  $\hbar\omega_2 = 0.275$  эВ), различаются не сильно. Отличие, однако, проявляется в поведении дифференциальной проводимости  $dI/dV$ . В точке, где потенциал равен энергии фона  $\hbar\omega_1$ , наблюдается разрыв, величина которого пропорциональна  $D_i$ . В точке разрыва появляются пики на зависимости  $d^2I/dV^2$ , расположенные на оси напряжений, в точках, отвечающих квантам молекулярных колебаний. Изучение этих особенностей, связанных с неупругим рассеянием электронов, лежит в основе метода туннельной спектроскопии (IETS) (см., к примеру, [52]).

**У.11.7. Рассеяние с переворотом спина.** В конце параграфа 11.1 мы видели, что упругие процессы дефазировки не оказывают заметного влияния на ток в одноуровневой системе. Напротив, в случае двухуровневой системы эти процессы могут приводить к существенным эффектам, если два уровня взаимодействуют с контактами по-разному. Рассмотрим уровень со спином, ориентированным вверх и взаимодействующим с левым контактом, и пусть уровень со спином, ориентированным вниз, взаимодействует с правым контактом, так что

$$[H] = \begin{pmatrix} \varepsilon & \Delta \\ \Delta^* & \varepsilon \end{pmatrix}, \quad [\Gamma_1] = \begin{pmatrix} \gamma_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad [\Gamma_2] = \begin{pmatrix} 0 & 0 \\ 0 & \gamma_2 \end{pmatrix}.$$

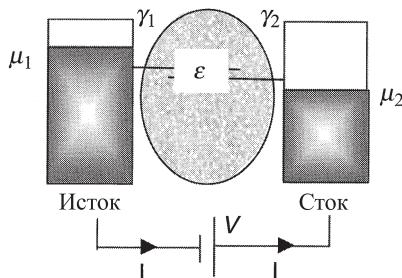


Рис. У.11.7

(а) Получите выражение для коэффициента пропускания  $\bar{T}(E)$ . Объясните физический смысл того, что он равен нулю при  $\Delta = 0$ .

(б) Рассмотрите случай, когда  $\Delta = 0$  и ток в отсутствие рассеяния равен нулю. Предположите далее, что имеются упругие процессы с дефазировкой, описываемые величинами  $\Sigma_{ij} = \sum_{k,l} D_{ijkl} G_{kl}$  и  $\Sigma_{ij}^{\text{in}} = \sum_{k,l} D_{ijkl} G_{kl}^n$ . Рассчитайте зависимость тока от  $D_0$  при малом фиксированном напряжении стока, предполагая, что единственными ненулевыми элементами матрицы  $D$  являются (а)  $D_{1122} = D_{2211} = D_0$  и (б)  $D_{1111} = D_{2222} = D_0$  (дальнейшее обсуждение см. в [16]).

## Глава 12

### Эпилог

В начале этой книги рассматривалась следующая «простая» задача: пусть к структуре очень малых размеров присоединены два контакта, причем в интересующей нас области энергий в спектре структуры есть всего один энергетический уровень. К контактам приложено внешнее напряжение  $V$ , так что их электрохимические потенциалы имеют различные значения:  $\mu_1 - \mu_2 = qV$  (см. рис. 1.2.2). Чему будет равен ток, идущий через структуру? Мы видели, что ответ на этот вопрос можно дать, рассматривая два уравнения, описывающие протекание тока через границы структуры с контактами (см. уравнения (1.2.3a), (1.2.3b)), при условии, что учтено уширение энергетических уровней.

Необходимо также учесть входящие и исходящие из контактов токи (рис. 12.1), разность между которыми равна скорости изменения числа электронов:

$$\hbar \frac{dN}{dt} + (\gamma N)_{\text{Входящий ток}} = (\gamma f)_{\text{Выходящий ток}}. \quad (12.1)$$

В главах 8 и 9 мы обсуждали квантовомеханический подход к этой задаче, базирующийся на одноэлектронном уравнении Шредингера

$$i\hbar \frac{d\psi}{dt} - H\psi - (\Sigma\psi)_{\text{Выходящий ток}} = (S)_{\text{Входящий ток}}, \quad (12.2)$$

в которое входит дополнительное слагаемое с собственно-энергетической функцией  $\Sigma\psi$  и слагаемое  $S$ , отвечающие исходящему и входящему потокам. Заметим, что уравнение (8.2) в главе 8 может рассматриваться как преобразование Фурье этого уравнения. Как и в традиционных курсах квантовой механики, в главах 2–7 был сделан акцент на описание внутренней динамики электронов в замкнутой системе с гамильтонианом  $[H]$ . Начиная с главы 8, однако, наше обсуждение базировалось на понятиях входящего и исходящего потоков, которые играют определяющую роль в задаче об

электронном транспорте. Далее, были введены уравнения квантового транспорта, весь набор которых, представленный в параграфе 11.1, определяет входящий и исходящий потоки через корреляционные функции величин, фигурирующих в матричных уравнениях (12.2):  $G^n \sim \psi\psi^+$ ,  $\sigma^{\text{in}} \sim SS^+$  и т. д.

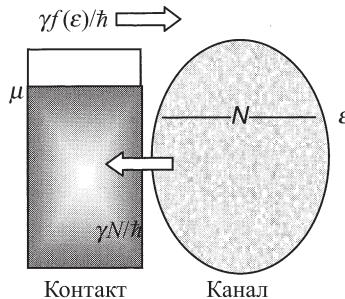


Рис. 12.1

Эти уравнения описывают идентичные физические явления, протекающие как в отдельном атоме, так и в нанотранзисторе. Причем здесь можно говорить как о единых моделях, так и методах расчета, объединяющих обратимую физику микромира, описываемую гамильтонианом  $[H]$ , с необратимой физикой макромира с ее токами и диссипативными потоками. Анализ необратимых процессов, опирающийся на обратимые законы физики, является одной из наиболее острых и глубоких проблем, постоянно обсуждаемой в литературе. Ее детальное обсуждение увело бы нас в сторону от основной темы этой книги, однако я позволю себе все же сказать несколько слов о задачах классического транспорта, которые могут оказаться полезными для изучения задач квантового транспорта. Чтобы упростить задачу, запишем соответствующие уравнения для одномерного случая, имея в виду, что их обобщение на трехмерный случай не представляет труда.

**Уравнение Ланжевена.** Классическая механика, основанная на законе Ньютона  $m(dv/dt) = F(t)$ , как известно, прекрасно описывает движение в системах без трения (к примеру, движение планет вокруг Солнца). Для описания движения тел, окружающих нас в повседневной жизни (например, автомобилей), это уравнение необходимо дополнить, включив в него силу трения  $\gamma v$ . Тогда оно примет вид

$$m \frac{dv}{dt} + \gamma v = F(t).$$

Из приведенного уравнения следует, что скорость тела будет стремиться к нулю в отсутствие каких-либо внешних сил, что не совсем верно, поскольку микроскопические объекты в состоянии равновесия находятся в состоянии случайного движения со средней кинетической энергией  $k_B T/2$  на каждую степень свободы (т. е. на каждый квадратичный по импульсу член гамильтониана). Это случайное движение не заметно на макроскопических масштабах, но весьма существенно для электронов и даже для больших молекул (примером является т. н. броуновское движение). Чтобы описать такую динамику с помощью уравнений движения, следует добавить в их правую часть случайную силу  $N(t)$ . Тогда динамика будет описываться уравнением Ньютона со стохастической силой

$$m \frac{dv}{dt} + \gamma v = F(t) + N(t). \quad (12.3)$$

Иногда это уравнение называют уравнением Ланжевена. Выбрав подходящие статистические параметры силы  $N(t)$  и коэффициента трения  $\gamma$ , с его помощью можно адекватно описывать необратимые явления и в том числе транспорт электронов в твердом теле. Так, уравнения дрейфа и диффузии, которые широко используются при моделировании полупроводниковых структур, могут рассматриваться как следствия одночастичного уравнения (12.3).

**Уравнение Больцмана.** Рассматривая динамику отдельной частицы, мы следим за ее координатой  $x(t)$  и скоростью  $v(t)$ . При описании коллектива частиц величины  $x$  и  $v$  становятся независимыми переменными, от которых зависят функции распределения  $f(x, v, t)$ , описывающие среднее число частиц, движущихся со скоростью  $v$  в точке  $x$ . Можно показать, что одночастичные уравнения

$$dx/dt = v(t) \quad \text{и} \quad m(dv/dt) = F(t)$$

могут быть записаны в эквивалентной форме на языке коллективной функции распределения (см., например, [34]):

$$\frac{\partial f}{\partial t} + \frac{F}{m} \frac{\partial f}{\partial v} + v \frac{\partial f}{\partial x} = 0.$$

Стохастическую силу  $N(t)$  и силу трения  $\gamma v$  можно эффективно учесть, рассматривая с помощью этого уравнения процессы, в которых электрон изменяет значение  $v$ , что описывается слагаемым  $S^{\text{in}}$ , определяющим приход, и слагаемым  $S^{\text{out}}$ , описывающим уход, для каждой точки

фазового пространства  $(x, v)$ :

$$\frac{\partial f}{\partial t} + \frac{F}{m} \frac{\partial f}{\partial v} + v \frac{\partial f}{\partial x} = S^{\text{in}} - S^{\text{out}}. \quad (12.4)$$

Перед нами знаменитое уравнение Больцмана, которое широко используется для описания явлений транспорта в различных системах, включая нейтроны и электроны. Оно записано для одномерного случая, но без труда обобщается и на трехмерный. Уравнения диффузии и дрейфа можно получить из уравнения (12.4) рассматривая т. н. «первые моменты», которые находятся умножением всего уравнения на  $v$  и последующим интегрированием по всем возможным значениям  $v$ .

Современные варианты описанного подхода, получившего название полуклассической теории транспорта, учитывают структуру энергетических зон, модифицируя при этом законы Ньютона следующим образом:

$$dx/dt = v \rightarrow dx/dt = (1/\hbar) \partial E / \partial k \equiv v,$$

$$m(dv/dt) = F \rightarrow \hbar dk/dt = F.$$

Для параболических зон с законом дисперсии  $E = \hbar^2 k^2 / 2m^*$  обе формы являются эквивалентными, нужно только заменить массу свободного электрона  $m$  на эффективную массу  $m^*$ . Если зоны не параболические, то две записи закона Ньютона становятся не эквивалентными. Можно показать, что правильной в этом случае является модифицированная версия, в которой фигурирует величина  $k$ . Соответствующим образом модифицированное уравнение Больцмана для функции распределения  $f(x, k, t)$  (а не для  $f(x, v, t)$ ) имеет вид

$$\frac{\partial f}{\partial t} + \frac{F}{\hbar} \frac{\partial f}{\partial k} + v \frac{\partial f}{\partial x} = S^{\text{in}} - S^{\text{out}}. \quad (12.5)$$

**«Стрела» времени.** Функциональная зависимость слагаемых  $S^{\text{in}}$  и  $S^{\text{out}}$  от функции распределения  $f$  должна отражать присутствие различных видов рассеяния, таких как рассеяние на неоднородностях, на фонах или рассеяние электронов на электронах. В общем случае, однако, эти слагаемые должны удовлетворять следующему условию: равновесная функция распределения  $f_0(E)$ , для которой  $E = U(x) + \varepsilon(k)$ , должна обращать разность  $S^{\text{in}} - S^{\text{out}} = 0$  в нуль. В этом случае функция  $f_{\text{eq}} = f_0(E)$  с  $E = U(x) + \varepsilon(k)$  является решением уравнения Больцмана, поскольку имеют место следующие равенства (обратите внимание на то, что  $F = -dU/dx$ ):

$$\partial f_{\text{eq}} / \partial t = 0,$$

$$\partial f_{\text{eq}} / \partial k = (df_0 / dE)(\partial E / \partial k) = (df_0 / dE)hv,$$

$$\partial f_{\text{eq}} / \partial x = (df_0 / dE)(\partial E / \partial x) = -(df_0 / dE)F,$$

откуда следует уравнение  $\frac{\partial f_{\text{eq}}}{\partial t} + \frac{F}{h} \frac{\partial f_{\text{eq}}}{\partial k} + v \frac{\partial f_{\text{eq}}}{\partial x} = 0$ .

Предполагается, что произвольное начальное распределение будет релаксировать, переходя со временем в равновесное распределение  $f_{\text{eq}}(x, k)$ , что находится в согласии с обычно наблюдаемыми процессами. Это свойство определяет предпочтительное направление (или «стрелу») во времени, т. е. принцип, согласно которому неравновесные распределения всегда эволюционируют в равновесные и никогда — в обратную сторону.

С другой стороны, если записать систему  $N$  связанных уравнений для совокупности  $N$  взаимодействующих частиц

$$m \frac{d^2 x_n}{dt^2} = m \frac{dv_n}{dt} = - \frac{\partial}{\partial x_n} U_{\text{int}}(x_1, \dots, x_N), \quad (12.6)$$

где  $U_{\text{int}}$  — это энергия взаимодействия, а градиенты по переменной  $x_n$ , взятые со знаком минус, — силы, действующие на  $n$ -ю частицу, то эти уравнения будут обратимыми во времени. Другими словами, если снять на видео эволюцию во времени любого решения системы (12.6) и воспроизвести его в обратном направлении, она будет выглядеть вполне естественно, поскольку замена  $t$  на  $-t$  не меняет само уравнение. Но если заснять и затем показать в обратном направлении эволюцию, описываемую уравнением Больцмана, то картина будет выглядеть абсурдно. Вместо того, чтобы увидеть, как неравновесное распределение эволюционирует в равновесное, мы будем наблюдать обратную картину, которая противоречит нашему повседневному опыту.

Современники критиковали Больцмана за то, что уравнения вида (12.5) с предпочтительным направлением времени не могут следовать из обратимых во времени уравнений типа (12.6). Теперь мы понимаем, что при наличии достаточно большого числа частиц система уравнений (12.6) будет также демонстрировать поведение, аналогичное тому, что предсказывает уравнение Больцмана. Такая динамика должна иметь место на практически значимом интервале времени, меньшем т. н. времени возврата (см., например, книгу [38] и имеющиеся там ссылки). Но это утверждение не является общепризнанным и можно встретить немало работ, где оно подвергается сомнению. Описать переход к равновесному состоянию, исходя из законов Ньютона (12.6), не просто. Больцман обошел эти трудности, просто выбрав подходящий вид рассеяния, часто называемый «гипотезой

о молекулярном хаосе» (Stosszahlansatz), в котором игнорируются тонкие корреляционные поправки к скоростям различных частиц. В задачах квантового транспорта при попытке описать равновесные состояния с помощью уравнения Шредингера (и даже его многочастичной версии вида (3.2.1)) также могут возникнуть подобные проблемы. Поэтому естественно выглядит попытка проделать с уравнением Шредингера то, что Больцман сделал с законом Ньютона.

**Квантовое уравнение Больцмана.** Поставленная выше задача решается с помощью метода неравновесных функций Грина (NEGF), развитого в работах Швингера, Бейма, Каданоффа и Келдыша (см. [24, 25, 37]). Основной величиной в этом методе является корреляционная функция (обычно записываемая в литературе как  $-iG^<(x, x'; t, t')$ )

$$G^n(x, x'; t, t') \equiv \langle \psi^+(x', t') \psi(x, t) \rangle,$$

которая играет ту же роль, что функция распределения  $f(x, k, t)$  в теории Больцмана. Для транспорта в стационарном состоянии корреляционная функция зависит только от разности  $(t - t')$ , а Фурье-компоненты от энергии  $E$ . Обратите внимание на то, что в квантовом формализме появляется одна дополнительная переменная; и в дополнение к временной координате рассматривается набор переменных  $(x, x', E)$ , а не  $(x, k)$ , как в классическом подходе Больцмана. Такая ситуация возникает потому, что в подходе Больцмана состояния с индексом « $k$ » считаются собственными, и поэтому индексы « $k$ » и « $E$ » однозначно связаны между собой. При более общем квантовом подходе такая связь может и не иметь места. Для более глубокого знакомства с данной точкой зрения и смежными вопросами, такими как преобразование Вигнера, читателю можно рекомендовать, например, параграф 8.8 книги [13].

**Модель Ландауэра.** До начала 80-х годов прошлого века в работах, посвященных изучению электронного транспорта (квантового или классического), в основном изучалось электрон-электронное, электрон-фононное или электрон-примесное взаимодействия, а также роль этих взаимодействий в формировании необратимости, диссиpации и в конечном счете электрического сопротивления. В экспериментах с реальными образцами влияние контактов обычно рассматривалось как небольшое возмущение. Эта точка зрения, однако, изменилась в 1980-х годах, когда в экспериментах с мезоскопическими структурами была установлена важная роль контактов. Это стимулировало появление работ Ландауэра, которые приобрели в настоящее время большую известность. В модели Ландауэра считается,

что в проводнике — проволоке или канале — нет каких-либо взаимодействий, а необратимость и диссипация появляются при его соединении с контактами. Эта модель позволяет адекватно описать электронные структуры малых размеров и объяснить микроскопическую природу сопротивления. Транспорт в такой системе мы называли в данной книге когерентным (см. главу 9).

В этой книге в качестве отправной точки я использовал еще более простую версию этой модели: рассматривалась система малых размеров, соединенная с двумя массивными контактами. Приведенное в главе 1 уравнение (1.2.1) можно рассматривать как некий вариант уравнения Больцмана, записанного для системы малых размеров, в которой роль функции распределения играет число  $N$ . Контакты при этом за счет необратимых взаимодействий поддерживаются в состоянии локального равновесия. Фактически рассмотрение сложной динамики здесь заменяется предположением о том, что приход и уход описываются простыми выражениями  $S^{\text{in}} = \gamma f_1$  и  $S^{\text{out}} = \gamma N$ , которые приводят к правильному равновесному решению  $N = f_1$ . Такое рассмотрение по духу близко к подходу Больцмана, заменившего рассмотрение сложной ньютоновской динамики многих тел удивительной подстановкой слагаемых  $S^{\text{in}}$  и  $S^{\text{out}}$ . В главе 9 мы развили эти представления и построили полную модель когерентного транспорта, которая эквивалентна модели Ландауэра. Далее в главе 10 была введена диссипация и получен полный набор уравнений диссипативного квантового транспорта (при слабом взаимодействии). Более подробно эти уравнения обсуждались и были проиллюстрированы примерами в главе 11.

**Теория за рамками когерентного транспорта.** На протяжении всей книги я пытался сочетать физически ясный подход Ландауэра с мощным теоретическим методом неравновесных функций Грина, который, как я убежден, необходим для описания некогерентного квантового транспорта. Так, было сделано утверждение о том, что слагаемое  $\Sigma\psi$  в выражении (12.2) отвечает уходу электронов, поэтому естественно потребовать, чтобы собственно-энергетическая функция  $\Sigma$  (чья мнимая часть определяет уширение уровней, или обратное время жизни) зависела от того, является ли конечное состояние, в которое направляется исходящий поток, пустым или заполненным. В главе 1 и далее в главе 9 я утверждал, что такой принцип запрета не действует до тех пор, пока рассматриваются лишь строго когерентные процессы. В присутствии некогерентных взаимодействий, как мы видели в главе 10, принципы запрета проявляются неявно, что трудно обосновать в рамках используемой нами одноэлектронной картины. Уже

отсюда ясно, что правильное описание возможно вне рамок одноэлектронного приближения.

При одноэлектронном подходе отдельные электроны описываются волновой функцией  $\psi$ , а электронная плотность получается при суммировании величины  $\psi^*\psi$  по всем отдельным электронам. В многоэлектронных задачах электроны описываются на языке полевых операторов  $c$ , так что оператор  $c^\dagger c$  определяет число частиц в данном состоянии, которое может быть равно 0 или 1 в соответствии с тем, свободно состояние или заполнено. Полевые операторы называют операторами «вторичного квантования». Это делается для того, чтобы подчеркнуть, что переход от ньютоновского описания к волновому уравнению Шредингера — это «первичное квантование». Метод вторичного квантования завершает построение квантового формализма, объединяя корпускулярную и волновую картину.

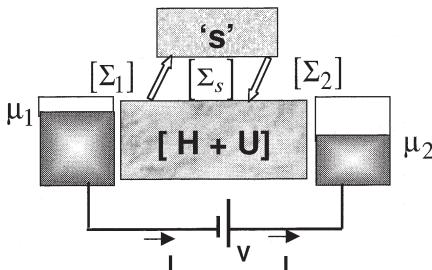
Можно показать, что эти операторы удовлетворяют дифференциальному уравнению

$$i\hbar \frac{d}{dt} - Hc - \Sigma c = S, \quad (12.7)$$

которое напоминает уравнение, использованное нами в одноэлектронных задачах (см. уравнение (12.2)). При этом, в отличие от величины  $\psi^*\psi$ , которая может принимать любое значение, операторы вида  $c^\dagger c$  могут принимать лишь значение 0 или 1 в соответствии с принципом запрета, который не содержится в уравнении Шредингера. Этот более совершенный подход необходим для выхода за рамки теории когерентного транспорта в область, где появляются неупругие взаимодействия, а также для описания таких тонких эффектов, как эффект Кондо, лежащий, впрочем, за рамками этой книги. При описании транспорта в системах сильно взаимодействующих частиц операторы  $c$  соответствуют сложным, или «одетым» частицам, таким как поляроны. В Приложении я попытался ввести читателя в круг более общих методов вторичного квантования, повторно рассматривая результаты глав 8–10. Здесь приведена полная, не обсуждавшаяся ранее система уравнений, зависящих от времени, и продемонстрировано, как эти уравнения применяются в системах с сильным электрон–электронным взаимодействием для описания кулоновской блокады и резонанса Кондо. Последний экспериментально наблюдаемый эффект невозможно объяснить с помощью простой одноэлектронной картины.

По мере развитияnanoэлектроники можно ожидать открытия новых эффектов, в которых определяющую роль будет играть взаимодействие контактов с каналом. При этом, как я упоминал в параграфе 8.4, понятие «контакт» включает в себя не только металлические электроды, но и фононы, описываемые собственно-энергетическими функциями  $[\Sigma]$ .

различного вида. При таком подходе, возможно, потребуется вводить дополнительные уравнения, определяющие состояния каждого «контакта». С другой стороны, более полезным может оказаться вариант теории, в которой транспортные уравнения описывают не отдельные электроны, а более сложные объекты, перенормированные взаимодействием с «контактами». В данный момент можно лишь предполагать тот или иной вариант развития.



Подводя итог, я еще раз высажу уверенность в том, что обсуждавшаяся здесь область физики конденсированного состояния находится на подъеме. Здесь можно ожидать существенного прогресса как в плане прикладных исследований, так и в развитии фундаментальных представлений. С одной стороны, будет совершенствоваться количественное описание физических явлений вnanoструктурах различной геометрии, изготовленных из различных материалов. Хотя сейчас многие экспериментальные результаты удовлетворительно описываются в рамках обсуждавшейся модели самосогласованного поля, предстоит еще многое сделать в плане выбора более удобных наборов базисных функций, представляющих гамильтониан  $[H]$  и собственно-энергетическую функцию  $[\Sigma]$  в присутствии неупругих процессов рассеяния, а также разработать более эффективные алгоритмы решения уравнений квантового транспорта. В то же время следует ожидать, что в области квантового транспорта будут открыты новые эффекты (как стационарные, так и зависящие от времени) в условиях сильного электрон-фононного и электрон-электронного взаимодействий. Обнаруженные до сей поры явления, вероятно, представляют собой лишь верхушку айсберга. По мере того как экспериментаторы будут совершенствовать методы контроля за взаимодействием между локализованными состояниями «канала» и делокализованными состояниями «контактов», можно ожидать открытия принципиально новых квантовых эффектов, некоторые из которых могут даже оказаться полезными на практике! Я надеюсь, что та общая

идея, которую я пытался развить здесь, поможет читателю присоединиться к небольшой, но активной части общества, которая, по словам Фейнмана, совершает «величайшее путешествие, которое когда-либо предпринимал человеческий разум» [17].

## Приложение: формализм вторичного квантования

При развитии общей «матричной» модели, схема которой показана на рис. 1.6.5 б, мы исходили из одноэлектронного уравнения Шредингера

$$i\hbar \frac{d\psi}{dt} - H\psi = 0, \quad (\text{П.1})$$

описывающего эволюцию во времени одноэлектронной волновой функции  $\psi$ . Электронная плотность находится как сумма  $\psi^*\psi$  по всем электронам. Хотя и непросто объяснить, что следует понимать под «различными электронами», эта процедура является адекватной при решении простых задач с когерентными взаимодействиями, где структура системы не подвержена влиянию тока электронов. Но если требуется продвинуться за рамки этого приближения в область более сложных процессов, включающих взаимодействия, нарушающие фазу, или сильные электрон-электронные взаимодействия, то в этой ситуации желательно встать на более общую точку зрения, рассматривающую электроны на языке полевых операторов  $c$ . Для системы невзаимодействующих электронов эти операторы вторичного квантования подчиняются системе дифференциальных уравнений

$$i\hbar \frac{d}{dt} c - Hc = 0, \quad (\text{П.2})$$

которая напоминает уравнение (П.1), описывающее эволюцию одноэлектронной волновой функции. При этом, в отличие от величины  $\psi^*\psi$ , которая может принимать любое значение, оператору числа частиц  $c^+c$  может отвечать только число 0 либо число 1, что отражает важный факт, никак не описываемый уравнением Шредингера. Эти два значения для оператора  $c^+c$  говорят нам, свободно ли данное состояние или занято. В состоянии равновесия среднее значение оператора  $c^+c$  в одноэлектронном состоянии с энергией  $\varepsilon$  определяется соответствующей функцией Ферми  $\langle c^+c \rangle = f_0(\varepsilon - \mu)$ . Это справедливо, однако, лишь для случая, когда наш канал состоит из

невзаимодействующих электронов (возможно, со взаимодействием в рамках самосогласованного поля), находящихся в состоянии равновесия. Для неравновесных задач необходима запись уравнений транспорта, которые позволяют найти значение  $c^{+}c$ , основываясь на нашем знании слагаемых типа источника, относящихся к контактам, которые предполагаются находящимися в состоянии локального равновесия и поэтому описываются равновесными соотношениями.

Я построю данное Приложение следующим образом: в параграфе П.1 будут введены полевые операторы и определены корреляционные функции; далее в параграфе П.2 будет получено выражение (11.1.1) для неравновесной матрицы плотности; в параграфе П.3 будет заново получено выражение (11.1.6) для тока; в параграфе П.4 будет приведен вывод выражений для функции  $\Sigma_S^{in}(E)$  и уширения уровней  $\Gamma_S(e)$  для некогерентных процессов (11.1.11); наконец, в параграфе П.5 мы покажем, каким образом данный подход может быть обобщен для задач, включающих сильные электрон-электронные взаимодействия, приводящие к эффектам кулоновской блокады и резонансу Кондо.

## П.1. Корреляционные функции

**Операторы рождения и уничтожения.** Рассмотрим систему одноэлектронных состояний, определяемых индексом  $j$ . В многоэлектронной картине каждому одноэлектронному состоянию  $j$  соответствует два состояния  $0_j$  и  $1_j$ . Оператор рождения  $c_j^{+}$ , который переводит  $j$ -й электрон из состояния  $0_j$  в состояние  $1_j$ , может быть представлен в виде матрицы

$$c_j^{+}(t) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \exp(+i\varepsilon_j t/\hbar), \quad (\text{П.1.1a})$$

матричные элементы которой записаны в базисе  $(1, 0)$ ,  $(0, 1)$ , где первый вектор отвечает состоянию  $0_j$ , а второй — состоянию  $1_j$ .

Оператор уничтожения  $c_j$ , который можно в матричной форме записать как

$$c_j(t) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \exp(-i\varepsilon_j t/\hbar), \quad (\text{П.1.1b})$$

переводит  $j$ -й электрон из состояния  $1_j$  в состояние  $0_j$ . Можно непосредственно убедиться в том, что в любой момент времени  $t$

$$c_j^{+}(t)c_j(t) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{и} \quad c_j(t)c_j^{+}(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

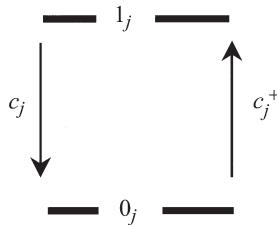


Рис. П.1.1

Первый из этих операторов называется оператором числа частиц, поскольку его собственные значения 0 и 1 равны числу электронов в соответствующем состоянии, а второй оператор определяет число незанятых состояний. Средние значения этих операторов интерпретируются как число электронов и число дырок в  $j$ -м состоянии:

$$\langle c_j^+(t)c_j(t) \rangle = f_j \quad \text{и} \quad \langle c_j(t)c_j^+(t) \rangle = 1 - f_j. \quad (\text{П.1.2})$$

Можно убедиться в справедливости следующих равенств:

$$c_j(t)c_j^+(t) + c_j^+(t)c_j(t) = 1;$$

$$c_j(t)c_j(t) = 0;$$

$$c_j^+(t)c_j^+(t) = 0.$$

Менее интуитивно понятными являются соотношения между полевыми операторами двух различных состояний  $i$  и  $j$ :

$$c_i(t)c_j^+(t) + c_j^+(t)c_i(t) = \delta_{ij}; \quad (\text{П.1.3a})$$

$$c_i(t)c_j(t) + c_j(t)c_i(t) = 0; \quad (\text{П.1.3b})$$

$$c_i^+(t)c_j^+(t) + c_j^+(t)c_i^+(t) = 0. \quad (\text{П.1.3c})$$

Обычно считается, что операторы двух различных состояний являются независимыми друг от друга, так что при  $i \neq j$  должно выполняться соотношение  $c_i(t)c_j^+(t) = c_j^+(t)c_i(t)$ . Исходя из уравнения (П.1.3a) можно было бы предположить, что каждое произведение операторов тождественно равно нулю. Однако в силу принципа запрета два различных состояния не являются по-настоящему независимыми и имеет место равенство  $c_i(t)c_j^+(t) = -c_j^+(t)c_i(t)$ . Можно показать, что оно будет выполнено, если мы слегка изменим определение (П.1.1), записав его как

$$c_j(t) = \begin{pmatrix} 0 & (-1)^\nu \\ 0 & 0 \end{pmatrix} \exp(-i\varepsilon_j t/\hbar)$$

и

$$c_j^+(t) = \begin{pmatrix} 0 & 0 \\ (-1)^\nu & 0 \end{pmatrix} \exp(-i\varepsilon_j t/\hbar),$$

где  $\nu$  есть число занятых состояний «слева» от  $j$ -го состояния. Последнее означает, что при работе с многими одноэлектронными состояниями необходимо установить определенный порядок в их расположении (при этом сам порядок нумерации значения не имеет), и далее придерживаться его. В этих условиях при выполнении расчетов выражение «слева» от  $j$ -го состояния имеет четко определенный смысл. На практике, однако, не нужно беспокоиться об этом, поскольку мы работаем с операторами, удовлетворяющими алгебраическим соотношениям (П.1.3a)–(П.1.3c). Необходимо лишь отметить, что из операторной алгебры, математически отражающей принцип запрета, следует, что на электрон в  $j$ -м состоянии влияет присутствие или отсутствие другого электрона в  $i$ -м состоянии, даже если не учитывается взаимодействие между электронами.

**Корреляционная и спектральная функции.** В общем случае можно определить двухвременную электронную корреляционную функцию как

$$G_{ij}^n(t, t') \equiv \langle c_j^+(t') c_i(t) \rangle \equiv -i G_{ij}^<(t, t'), \quad (\text{П.1.4a})$$

а также двухвременную дырочную корреляционную функцию

$$G_{ij}^p(t, t') \equiv \langle c_i(t) c_j^+(t') \rangle \equiv +i G_{ij}^>(t, t'). \quad (\text{П.1.4b})$$

При  $t = t'$  эти функции сводятся к операторам числа частиц (П.1.2). Сумма этих операторов определяет спектральную функцию:

$$A_{ij}(t, t') \equiv G_{ij}^p(t, t') + G_{ij}^n(t, t') = i[G_{ij}^>(t, t') - G_{ij}^<(t, t')]. \quad (\text{П.1.4c})$$

**Преобразования Фурье.** В стационарных состояниях, когда корреляционные функции зависят только от разности временных координат, удобно работать с их Фурье-образами, по переменной  $(t - t')$ , зависящими от энергии  $E$ :

$$G_{ij}^n \equiv \int_{-\infty}^{+\infty} (d\tau/\hbar) \exp(iE\tau/\hbar) G_{ij}^n(t, t - \tau). \quad (\text{П.1.5a})$$

Обратное преобразование дается формулой

$$G_{ij}^n(t, t - \tau) \equiv \langle c_j^+(t - \tau) c_i(t) \rangle = \quad (\text{П.1.5b})$$

$$= \int_{-\infty}^{+\infty} (dE/2\pi) G_{ij}^n(E) \exp(-iE\tau/\hbar), \quad (\text{П.1.5c})$$

и при совпадающих временных аргументах корреляционная функция записывается как

$$G_{ij}^n(t, t) \equiv \langle c_j^+(t) c_i(t) \rangle = \int_{-\infty}^{+\infty} (dE/2\pi) G_{ij}^n(E). \quad (\text{П.1.6})$$

Аналогичные соотношения можно ввести и для дырочной корреляционной функции  $G^p$ , а также для спектральной функции  $A$ .

**Состояние равновесия.** Электронная и дырочная корреляционные функции могут принимать любые значения, если их сумма, в соответствии с уравнением (П.1.4c), будет равна спектральной функции. В состоянии равновесия, однако, электронная корреляционная функция пропорциональна функции Ферми (я использую строчные буквы для функций в состоянии равновесия)

$$g_{ij}^n(E) = a_{ij}(E) f_0(E - \mu), \quad (\text{П.1.7a})$$

$$g_{ij}^p(E) = a_{ij}(E)(1 - f_0(E - \mu)). \quad (\text{П.1.7b})$$

В случае изолированной системы, описываемой гамильтонианом  $[h]$ , спектральная функция имеет простой вид в представлении собственных функций гамильтониана:

$$a_{rs}(E) = \delta_{rs} \delta(E - \varepsilon_r), \quad (\text{П.1.7c})$$

где индексы  $r$  и  $s$  нумеруют собственные состояния  $[h]$ , так что

$$g_{rs}^n(E) = \delta_{rs} \delta(E - \varepsilon_r) f_r, \quad \text{где } f_r \equiv f_0(\varepsilon_r - \nu); \quad (\text{П.1.8a})$$

$$g_{rs}^n(t, t') = \delta_{rs} f_r \exp[-i\varepsilon(t - t')/\hbar] \exp(-\eta|t - t'|/\hbar), \quad (\text{П.1.8b})$$

а  $\eta$  есть бесконечно малая положительная величина. Наша цель состоит в том, чтобы записать эти соотношения с соответствующими функциями Ферми в контактах и затем рассчитать корреляционные функции в интересующей нас области, а именно, в канале.

**Бозонные операторы.** При обсуждении задач, в которых фигурирует излучение фононов или фотонов, необходимо рассматривать операторы  $b_\alpha^+$  и  $b_\alpha$ , описывающие фононныe или фотонные поля. В отличие от электронных эти операторы удовлетворяют другим алгебраическим соотношениям:

$$b_\beta(t) b_\alpha^+(t) - b_\alpha^+(t) b_\beta(t) = \delta_{\alpha\beta}, \quad (\text{П.1.9a})$$

$$b_\alpha(t) b_\beta(t) - b_\beta(t) b_\alpha(t) = 0, \quad (\text{П.1.9b})$$

$$b_\alpha^+(t) b_\beta^+(t) - b_\beta^+(t) b_\alpha^+(t) = 0, \quad (\text{П.1.9c})$$

где индексы  $\alpha$  и  $\beta$  нумеруют различные фононные или фотонные моды. Сравнивая эти соотношения с уравнениями (П.1.3a)–(П.1.3c) для электронных операторов, нетрудно видеть, что различие состоит в замене положительного знака в сумме на отрицательный. Это «незначительное» изменение, однако, делает свойства этих операторов гораздо более понятными на интуитивном уровне, поскольку теперь различные моды с различными  $\alpha \neq \beta$  в соответствии со «здравым смыслом» становятся независимыми

$$b_\beta(t)b_\alpha^+(t) = b_\alpha^+(t)b_\beta(t) \quad \text{при } \alpha \neq \beta.$$

Действительно, читатель, уже прослушавший курс квантовой механики, сразу узнает здесь операторный подход, используемый для описания гармонического осциллятора, где также вводятся операторы рождения и уничтожения, являющиеся линейными комбинациями операторов координаты ( $x$ ) и импульса ( $p$ ) и подчиняющиеся тем же коммутационным соотношениям, что и операторы в выражениях (П.1.9a)–(П.1.9c). Это означает, что в отличие от электронных операторов в этом случае мы можем выразить фононные или фотонные операторы через обычные дифференциальные операторы. В настоящем Приложении мы, тем не менее, не будем работать в каком-либо конкретном представлении. Достаточно просто использовать алгебраические соотношения (П.1.9a)–(П.1.9c). Одним из следствий этих изменений в операторной алгебре будет то, что вместо выражения (П.1.2) мы получим

$$\langle b_\alpha^+(t)b_\alpha(t) \rangle = N_\alpha \quad \text{и} \quad \langle b_\alpha(t)b_\alpha^+(t) \rangle = 1 + N_\alpha, \quad (\text{П.1.10})$$

где  $N_\alpha$  есть число фононов в соответствующей моде. Как и для электронных операторов (см. уравнение (П.1.4a)–(П.1.4c)), определим двухвременные корреляционные функции

$$G_{\alpha\beta}^{ab}(t, t') \equiv \langle b_\beta^+(t')b_\alpha(t) \rangle \quad (\text{П.1.11a})$$

и

$$G_{\alpha\beta}^{em}(t, t') \equiv \langle b_\beta(t)b_\alpha^+(t') \rangle, \quad (\text{П.1.11b})$$

которым в отличие от электронных и дырочных индексов «п» и «р» приписаны индексы «аб» и «ем», означающие поглощение и испускание. В стационарных условиях, когда эти функции зависят только от разности времен ( $t - t'$ ), после применения к ним преобразования Фурье мы придем к Фурье-образам, зависящим от частоты. В состоянии равновесия будем иметь (сравните с уравнениями (П.1.7a)–(П.1.7c))

$$\begin{aligned} g_{\alpha\beta}^{ab}(\hbar\omega) &= a_{\alpha\beta}^{ph}(\hbar\omega)N(\hbar\omega), \\ g_{\alpha\beta}^{em}(\hbar\omega) &= a_{\alpha\beta}^{ph}(\hbar\omega)(1 + N(\hbar\omega)), \end{aligned} \quad (\text{П.1.12})$$

где распределение Ферми заменяется распределением Бозе:

$$N(\hbar\omega) = [\exp(\hbar\omega/k_B T) - 1]^{-1}, \quad (\text{П.1.13})$$

а фононная спектральная функция в представлении собственных состояний определяется выражением

$$a_{\alpha\beta}^{\text{ph}}(\hbar\omega) = \delta_{\alpha\beta}\delta(\hbar\omega - \hbar\omega_\alpha), \quad (\text{П.1.14})$$

откуда следует

$$\begin{aligned} g_{\alpha\beta}^{\text{ab}}(E) &= \delta_{\alpha\beta}\delta(\hbar\omega - \hbar\omega_\alpha)N_\alpha, \quad \text{где } N_\alpha \equiv N(\hbar\omega_\alpha), \\ g_{\alpha\beta}^{\text{ab}}(t, t') &= \delta_{\alpha\beta}N_\alpha \exp[-i\omega(t - t')] \exp(-\eta|t - t'|/\hbar) \end{aligned} \quad (\text{П.1.15})$$

и

$$\begin{aligned} g_{\alpha\beta}^{\text{em}}(E) &= \delta_{\alpha\beta}\delta(\hbar\omega - \hbar\omega_\alpha)(N_\alpha + 1), \\ g_{\alpha\beta}^{\text{em}}(t, t') &= \delta_{\alpha\beta}(N_\alpha + 1) \exp[-i\omega(t - t')] \exp(-\eta|t - t'|/\hbar). \end{aligned}$$

**Соотношение для энергии и времени.** Необходимо упомянуть о том, что волновая функция  $\psi(t)$  и полевой оператор  $c(t)$  не являются наблюдаемыми величинами. В качестве примера наблюдаемых величин можно привести корреляционные функции вида  $\psi^*(t')\psi(t)$  или  $c^+(t')c(t)$ . Эти величины подобны шумовым флуктуациям напряжения  $V(t)$  на контактах со-противления, которые описываются корреляционной функцией  $V(t')V(t)$ . В стационарных условиях подобные двухвременные корреляционные функции зависят только от разности аргументов  $(t - t')$ , а соответствующие образы Фурье зависят от переменной энергии  $E$ . Так, электрон с волновой функцией  $\psi(t) = \psi_0 \exp(-i\varepsilon t/\hbar)$  описывается корреляционной функцией вида

$$\psi^*(t')\psi(t) = \psi^*\psi \exp[-i\varepsilon(t - t')/\hbar],$$

а соответствующий Фурье-образ по переменной  $(t - t')$  пропорционален дельта-функции  $\delta(E - \varepsilon)$  при  $E = \varepsilon$ . Эффекты, происходящие в стационарном состоянии, могут быть описаны на языке подобных Фурье-образов с энергией  $E$  в качестве независимой переменной, как, например, рассмотренный простейший механизм уширения уровней в параграфе 1.3. В последующих параграфах будут получены выражения, зависящие от времени, а затем после преобразования Фурье по переменной  $(t - t')$  найдены зависящие от энергии выражения для анализа стационарных состояний.

## П.2. Неравновесная матрица плотности

**Разбиение.** Операторы, удовлетворяющие соотношениям (П.1.9а)–(П.1.9с), обычно называют бозонными операторами, а соотношения (П.1.3а)–(П.1.3с) — фермиевские операторы. В состоянии *равновесия* соответствующие корреляционные функции определяются выражениями (П.1.8а)–(П.1.8б) и (П.1.15) соответственно. Задача состоит в том, чтобы вычислить их в области, где равновесие отсутствует. Поэтому мы примем предположение о том, что внешние воздействия на канал исходят от контактов, а также от фононных или фотонных резервуаров, поддерживаемых в своих состояниях равновесия внешней средой.

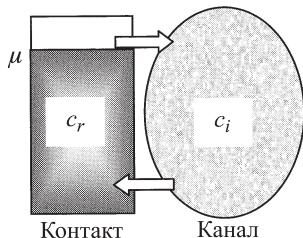


Рис. П.2.1

Чтобы реализовать этот план, разобьем всю структуру на интересующую нас часть, обозначенную как канал, и на резервуар, обозначенный как контакт, имеющий большую и непрерывно распределенную плотность состояний (рис. П.2.1). Начнем с уравнений, описывающих эту составную систему:

$$i\hbar \frac{d}{dt} \begin{pmatrix} c_i \\ c_r \end{pmatrix} = \begin{pmatrix} \varepsilon_i & \tau_{ir} \\ [\tau_{ir}^+]_{ri} & \varepsilon_r \end{pmatrix} \begin{pmatrix} c_i \\ c_r \end{pmatrix}, \quad (\text{П.2.1})$$

и получим уравнение вида (12.7), описывающее канал, избавившись при этом от переменных с индексами  $r$ , отвечающими резервуару (предполагается, что  $r$  и  $i$  описывают состояния изолированного канала и резервуара соответственно). Мы сделаем это, предполагая, что резервуар поддерживается внешней средой в состоянии равновесия, поэтому соответствующие поля описываются выражением (П.1.7). В этом параграфе я хотел бы проиллюстрировать обсуждаемый метод, предполагая, что взаимодействие с контактом включает только упругие столкновения, как мы уже предполагали в наших обсуждениях когерентного транспорта в главах 8 и 9. Мы увидим, что математические выкладки в этом случае выглядят так, как ес-

ли бы величины  $c$  были обычными комплексными числами, такими как используемые нами волновые функции. Нам по-настоящему не потребуется использовать тот факт, что в действительности они являются фермионными операторами. В последующих параграфах, однако, мы будем обсуждать более общие задачи, включающие электрон-фононное и электрон-электронное взаимодействия, где будут использованы свойства фермионных и бозонных операторов, чтобы получить результаты, которые было бы сложно обосновать, оставаясь в рамках одноэлектронной картины.

**Устранение переменных, описывающих состояния резервуара.** Подмножество состояний в выражении (П.2.1), описывающих резервуар, отвечает системе уравнений (причем  $[\tau^+]_{rj} = \tau_{jr}^*$ )

$$i\hbar \frac{d}{dt} c_r = (\varepsilon_r - i\eta) c_r + \sum_j \tau_{jr}^* c_j + S_r, \quad (\text{П.2.2})$$

где, как и при выводе уравнения (8.1.8a), мы ввели бесконечно малую минимую добавку  $\eta$  к энергии, а в правую часть добавили слагаемое типа источника  $S_r$ . Решение для полевых операторов резервуара можно записать в виде

$$c_r(t) = C_r(t) + \sum_j \int_{-\infty}^{+\infty} dt_1 g_{rr}(t, t') \tau_{jr}^* c_j(t_1), \quad (\text{П.2.3})$$

где первое слагаемое отвечает изолированному контакту, а второе обусловлено влиянием канала. Функция  $g_{rr}$ , являющаяся решением уравнения для полевых операторов резервуара, определенная как

$$g_{rr}(t, t') = (1/i\hbar)\vartheta(t - t') \exp(-i\varepsilon_r - \eta)(t - t'), \quad (\text{П.2.4})$$

является функцией Грина (или «импульсным откликом») уравнения (П.2.2):

$$L_r g_{rr}(t, t') = \delta(t - t'),$$

где

$$L_r \equiv i\hbar \frac{d}{dt} - (\varepsilon_r - i\eta), \quad (\text{П.2.5})$$

а  $C_r(t)$  является решением однородного уравнения  $L_r C_r(t) = 0$ .

Функция  $C_r(t)$  определяет эволюцию полевого оператора *изолированного* контакта (см. рис. П.2.2) до момента его соединения с каналом, которое происходит в момент времени  $t = 0$ . Это обстоятельство позволяет

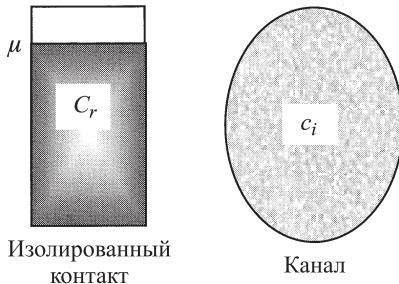


Рис. П.2.2

записать корреляционные функции ( $f_r \equiv f_0(\varepsilon_r - \mu)$ ) для состояния равновесия (см. выражения (П.1.8а), (П.1.8б)):

$$g_{rs}^n(E) = \delta_{rs}\delta(E - \varepsilon_r)f_r, \quad (\text{П.2.6а})$$

$$\begin{aligned} g_{rs}^n(t, t') &\equiv \langle C_s^+(t')C_r(t) \rangle = \\ &= \delta_{rs}f_r \exp[-i\varepsilon_r(t - t')] \exp(-\eta|t - t'|). \end{aligned} \quad (\text{П.2.6б})$$

При этом предполагается, что в канале до момента времени  $t = 0$  электронов нет вообще, и далее они проникают в него, «просачиваясь» из контакта после того, как между ним и каналом при  $t = 0$  установлена связь. Этот процесс можно описать, отталкиваясь от уравнения для подмножества состояний канала в уравнении (П.2.1):

$$i\hbar \frac{d}{dt}c_i = \varepsilon_i c_i + \sum_j \tau_{ir} c_r,$$

и подставив далее выражение для  $c_r(t)$  из (П.2.3):

$$i\hbar \frac{d}{dt}c_i - \varepsilon_i c_i - \sum_j \int_{-\infty}^{+\infty} dt_1 \Sigma_{ij}(t, t_1) c_j(t_1) = S_i(t), \quad (\text{П.2.7})$$

где

$$\Sigma_{ij}(t, t_1) \equiv \sum_r \tau_{ir} g_{rr}(t, t') \tau_{jr}^* \quad (\text{П.2.8})$$

и

$$S_i(t) \equiv \sum_r \tau_{ir} C_r(t). \quad (\text{П.2.9})$$

**Уравнения для корреляционных функций.** Определив функцию Грина уравнения (П.2.7) с интегродифференциальным оператором как

$$i\hbar \frac{d}{dt} G_{ik}(t, t') - \varepsilon_i G_{ik}(t, t') - \sum_j \int_{-\infty}^{+\infty} dt_1 \Sigma_{ij}(t, t') G_{jk}(t_1, t') = \delta_{ik} \delta(t - t'),$$

запишем решение уравнения (П.2.7) в виде

$$c_i(t) = \sum_k \int_{-\infty}^{+\infty} dt_1 G_{ik}(t, t_1) S_k(t_1). \quad (\text{П.2.10})$$

Тогда корреляционная функция будет определяться выражением

$$\begin{aligned} G_{ij}^n(t, t') &\equiv \langle c_j^+(t') c_i(t) \rangle = \\ &= \sum_{k,l} \int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 G_{ik}(t, t_1) G_{jl}^*(t', t_2) \langle S_l^+(t_2) S_k(t_1) \rangle. \end{aligned}$$

Определив далее функции

$$\Sigma_{kl}^{\text{in}}(t_1, t_2) \equiv \langle S_l^+(t_2) S_k(t_1) \rangle \quad (\text{П.2.11})$$

и

$$[G^+]_{ij}(t_2, t') \equiv [G]_{jl}^*(t', t_2) \quad (\text{П.2.12})$$

и используя матричные обозначения, мы можем записать следующее уравнение:

$$[G^n(t, t')] = \int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 [G(t, t_1)] [\Sigma^{\text{in}}(t_1, t_2)] [G^+(t_2, t')], \quad (\text{П.2.13})$$

где

$$L[G(t, t')] = [I] \delta(t - t'),$$

$$L \equiv i\hbar \frac{d}{dt} [I] - [h] - \int_{-\infty}^{+\infty} dt_1 [\Sigma(t, t_1)] \quad (\text{П.2.14})$$

и

$$[\Sigma(t, t_1)] \equiv [\tau][g(t, t')][\tau^+], \quad (\text{П.2.15})$$

а  $[h]$  — это гамильтониан, описывающий изолированный канал, чьи состояния нумеруются индексом  $i$ . Выражение (П.2.13) связывает корреляционную функцию канала с собственно-энергетической функцией  $\Sigma^{\text{in}}$ , описывающей рассеяние входящих в канал электронов и определяющей степень корреляции с источником.

В стационарных условиях все функции зависят только от разности двух временных координат, поэтому интегралы по времени в (П.2.13) являются интегралами типа свертки, Фурье-образы которых превращаются в простые произведения:

$$[G^n(E)] \equiv [G(E)][\Sigma^{\text{in}}(E)][G^+(E)], \quad (\text{П.2.16})$$

где

$$L[G(E)] = [I],$$

$$L \equiv E[I] - [h] - [\Sigma(E)], \quad (\text{П.2.17})$$

$$[\Sigma(E)] \equiv [\tau][G(E)][\tau^+]. \quad (\text{П.2.18})$$

**Собственно-энергетическая функция, описывающая рассеяние в канал.** Чтобы вычислить функцию рассеяния для входящих электронов

$$\Sigma_{ij}^{\text{in}}(t, t') \equiv \langle S_j^+(t') S_i(t) \rangle,$$

мы подставим для слагаемого типа источника выражение (П.2.9), откуда

$$\begin{aligned} \Sigma_{ij}^{\text{in}}(t, t') &= \sum_s \tau_{js}^* \sum_r \tau_{ir} \langle C_s^+(t') C_r(t) \rangle = \\ &= \sum_{r,s} \tau_{ir} g_{rs}^n(t, t') \tau_{js}^*, \end{aligned}$$

где  $g_{rs}^n(t, t')$  определяется выражением (П.2.6), поскольку она отвечает корреляционной функции изолированного контакта в состоянии равновесия, когда он не взаимодействует с каналом. В матричных обозначениях этому соответствует уравнение

$$[\Sigma^{\text{in}}(t, t')] = [\tau][g^n(t, t')][\tau^+]. \quad (\text{П.2.19})$$

Необходимо еще раз подчеркнуть, что поскольку эта корреляционная функция зависит лишь от разности двух временных аргументов, преобразование Фурье по этой переменной приводит к уравнению

$$[\Sigma^{\text{in}}(E)] = [\tau][g^n(E)][\tau^+]. \quad (\text{П.2.20})$$

**Рассеяние в канал и уширение уровней.** Интересно отметить, что антиэрмитова часть собственно-энергетической функции (см. выражение (П.2.18)), называемая также функцией уширения, определяется выражением

$$[\Gamma(E)] = i[\Sigma(E) - \Sigma^+(E)] = [\tau][a(E)][\tau^+],$$

где  $[a(E)]$  есть спектральная функция изолированного контакта:  $[a] = i[g - g^+]$ . Если предположить, что контакт находится в состоянии равновесия, описываемом функцией Ферми  $f(E)$ , то функция рассеяния внутри канала из уравнения (П.2.20) может быть записана как

$$[\Sigma^{\text{in}}] = [\tau][a(E)][\tau^+]f(E),$$

что устанавливает следующую связь между функцией рассеяния внутри канала с функцией уширения:

$$[\Sigma^{\text{in}}(E)] = [\Gamma(E)]f(E). \quad (\text{П.2.21})$$

### П.3. Входящий и исходящий потоки

Сейчас мы получим выражение для входящего и выходящего электронных потоков, выбрав в качестве отправной точки выражение для тока (заметив, что  $\langle c_i^+ c_i \rangle$  определяет число электронов в  $i$ -м состоянии):

$$\begin{aligned} I(t) &\equiv \sum_i \frac{d}{dt} \langle c_i^+(t) c_i(t) \rangle = \\ &= \frac{1}{i\hbar} \sum_i \langle c_i^+(t) (i\hbar \frac{d}{dt} c_i(t)) \rangle - \langle (-i\hbar \frac{d}{dt} c_i^+(t)) c_i(t) \rangle, \end{aligned}$$

и подставим в него  $c_i(t)$  из уравнения (П.2.7), чтобы получить явные выражения для входящего и выходящего потоков. Говоря более общими словами, мы можем определить вариант функции  $I(t, t')$ , зависящий от двух временных координат, как

$$I(t, t') \equiv \sum_i \left( \frac{d}{dt} + \frac{d}{dt'} \right) \langle c_i^+(t') c_i(t) \rangle, \quad (\text{П.3.1})$$

чые «диагональные» элементы ( $t = t'$ ) дают значения полного тока. Преимущество такого варианта записи с двумя переменными состоит в том, что для описания транспорта в стационарном состоянии мы можем выполнить преобразование Фурье по переменной ( $t - t'$ ), чтобы получить энергетический спектр тока. Подставляя выражение для  $c_i(t)$  из (П.2.7) в (П.3.1), получим

$$I(t, t') = I^{\text{in}}(t, t') - I^{\text{out}}(t, t'),$$

где

$$\begin{aligned} & I^{\text{out}}(t, t') = \\ & = (-1/i\hbar) \sum_{ij} \int_{-\infty}^{+\infty} dt_1 (\sigma_{ij}(t, t_1) \langle c_i^+(t') c_j(t_1) \rangle - \sigma_{ij}^*(t', t_1) \langle c_j^+(t_1) c_i(t) \rangle) = \\ & = (-1/i\hbar) \sum_{ij} \int_{-\infty}^{+\infty} dt_1 (\sigma_{ij}(t, t_1) G_{ji}^n(t_1, t') - \sigma_{ij}^*(t', t_1) G_{ij}^n(t, t_1)) \\ \text{и} \quad & I^{\text{in}}(t, t') = (1/i\hbar) \sum_i (c_i^+(t') S_i(t) - S_i^+(t') c_i(t)) = \\ & = (1/i\hbar) \sum_{ij} \int_{-\infty}^{+\infty} dt_1 G_{ij}^*(t', t_1) \langle S_j^+(t_1) S_i(t) \rangle - G_{ij}(t, t_1) \langle S_i^+(t') S_j(t_1) \rangle. \end{aligned}$$

Здесь при выводе было использовано уравнение (П.2.10), или

$$I^{\text{in}}(t, t') = (1/i\hbar) \sum_{i,j} \int_{-\infty}^{+\infty} dt_1 G_{ij}^*(t', t_1) \sigma_{ij}^{\text{in}}(t, t_1) - G_{ij}(t, t_1) \sigma_{ji}^{\text{in}}(t_1, t'),$$

где использовалось уравнение (П.2.11).

В матричных обозначениях эти результаты записываются следующим образом:

$$I^{\text{in}}(t, t') = (1/i\hbar) \text{Tr} \int_{-\infty}^{+\infty} dt_1 [\Sigma^{\text{in}}(t, t_1)] [G^+(t_1, t')] - [G(t, t_1)] [\Sigma^{\text{in}}(t_1, t')], \quad (\text{П.3.2})$$

$$I^{\text{out}}(t, t') = (-1/i\hbar) \text{Tr} \int_{-\infty}^{+\infty} dt_1 [\Sigma(t, t_1)] [G^n(t_1, t')] - [G^n(t, t_1)] [\Sigma^+(t_1, t')]. \quad (\text{П.3.3})$$

Обратите внимание на то, что здесь мы определили обычные токи проводимости, к которым для того, чтобы получить полный ток в контакте, следует добавить еще токи смещения.

Хотелось бы подчеркнуть еще раз, что в стационарных условиях каждая из величин зависит только от разности двух временных аргументов, и при выполнении преобразования Фурье по этой переменной интеграл типа свертки дает Фурье-образ, равный простому произведению:

$$\begin{aligned} I^{\text{in}}(E) &= \text{Tr}[\Sigma^{\text{in}}(E)G^+(E) - G(E)\Sigma^{\text{in}}(E)]/i\hbar = \\ &= \text{Tr}[\Sigma^{\text{in}}(E)A(E)]/\hbar, \end{aligned} \quad (\text{П.3.4})$$

$$\begin{aligned} I^{\text{out}}(E) &= \text{Tr}[G^n(E)\Sigma^+(E) - \Sigma(E)G^n(E)]/i\hbar = \\ &= \text{Tr}[\Gamma(E)G^n(E)]/\hbar. \end{aligned} \quad (\text{П.3.5})$$

**Структуры с несколькими контактами.** Нами получены основные выражения для входящего (уравнение (П.3.4)) и исходящего (уравнение (П.3.5)) электронных потоков, о которых речь шла еще в связи с обсуждением рис. 1.6.5 и уравнения для функции  $G^n$  (уравнение (П.2.16)). Для простоты мы рассматривали случай, когда канал соединен только с одним контактом, но результаты могут быть распространены и на реальные структуры с двумя или большим числом контактов, обозначаемых индексами  $p, q$ . К примеру, уравнения (П.2.13)–(П.2.15) могут быть модифицированы следующим образом:

$$[G^n(t, t')] = \sum_p \int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{+\infty} dt_2 [G(t, t_1)][\Sigma^{\text{in}}(t_1, t_2)]^{(p)} [G^+(t_2, t')], \quad (\text{П.3.6})$$

где

$$L[G(t, t')] = [I]\delta(t - t')$$

и

$$L \equiv i\hbar \frac{d}{dt}[I] - [h] - \sum_p \int_{-\infty}^{+\infty} dt_1 [\Sigma(t, t_1)]^{(p)}, \quad (\text{П.3.7})$$

$$[\Sigma(t, t')]^{(p)} \equiv [\tau]^{(p)} [g(t, t')]^{(p)} [\tau^+]^{(p)}, \quad (\text{П.3.8})$$

в то время как уравнения (П.2.19), (П.3.2) и (П.3.3) принимают вид

$$[\Sigma^{\text{in}}(t, t')]^{(p)} = [\tau]^{(p)} [g^n(t, t')]^{(p)} [\tau^+]^{(p)}, \quad (\text{П.3.9})$$

$$I^{\text{in}}(t, t')^{(p)} = (1/i\hbar) \text{Tr} \int_{-\infty}^{+\infty} dt_1 [\Sigma^{\text{in}}(t, t_1)]^{(p)} [G^+(t_1, t')] - [G(t, t_1)] [\Sigma^{\text{in}}(t_1, t')]^{(p)}, \quad (\text{П.3.10})$$

$$I^{\text{out}}(t, t')^{(p)} = (1/i\hbar) \text{Tr} \int_{-\infty}^{+\infty} dt_1 [\Sigma(t, t_1)]^{(p)} [G^n(t_1, t')] - [G^n(t, t_1)] [\Sigma^+(t_1, t')]^{(p)}.$$

В стационарных условиях зависящие от энергии величины в структурах со многими контактами принимают следующий вид:

$$[G^n(E)] = \sum_p [G(E)] [\Sigma^{\text{in}}(E)]^{(p)} [G^+(E)], \quad (\text{П.3.11})$$

$$L[G] = [I], \quad L \equiv E[I] - [h] - \sum_p [\Sigma(E)]^{(p)}, \quad (\text{П.3.12})$$

$$[\Sigma(E)]^{(p)} = [\tau]^{(p)} [g(E)]^{(p)} [\tau^+]^{(p)}, \quad (\text{П.3.13})$$

$$[\Sigma^{\text{in}}(E)]^{(p)} = [\tau]^{(p)} [g^n(E)]^{(p)} [\tau^+]^{(p)}, \quad (\text{П.3.14})$$

$$I^{\text{in}}(E)^{(p)} = \text{Tr} [\Sigma^{\text{in}}(E)]^{(p)} [A(E)] / \hbar, \quad (\text{П.3.15})$$

$$I^{\text{out}}(E)^{(p)} = \text{Tr} [\Gamma(E)]^{(p)} [G^n(E)]. \quad (\text{П.3.16})$$

Полный ток в контакте равен разности между входящим и исходящим токами:

$$I(E)^{(p)} = (1/\hbar) \text{Tr} \left( [\Sigma^{\text{in}}(E)]^{(p)} [A(E)] - [\Gamma(E)]^{(p)} [G^n(E)] \right).$$

Используя выражение (П.3.10) и соотношение

$$A = \sum_q G \Gamma^q G^+ = \sum_q G^+ \gamma^{(q)} G,$$

следующее из уравнений (П.3.11) и (П.3.12), при условии  $\Sigma^{\text{in}(p)} = \Gamma^{(p)} f_p$  будем иметь

$$I(E)^{(p)} = (1/\hbar) \sum_q \text{Tr} \left( \Sigma^{\text{in}(p)} G \Gamma^{(q)} G^+ - \Gamma^p G \Sigma^{\text{in}(q)} G^+ \right) =$$

$$= (1/\hbar) \sum_q \text{Tr} [\Gamma^{(p)} G \Gamma^{(q)} G^+] (f_p - f_q).$$

Далее выражение для тока запишем в виде

$$I(E)^{(p)} = (1/\hbar) \sum_q T^{(pq)} (f_p - f_q) \quad (\text{П.3.17})$$

с использованием функции пропускания

$$T^{(pq)} \equiv \text{Tr}[\Gamma^{(p)} G \Gamma^{(q)} G^+]. \quad (\text{П.3.18})$$

Линеаризуя далее уравнение (П.3.17) вблизи равновесного значения электрохимического потенциала, мы получим стандартные уравнения Бютткера, в которых для отклонений от равновесных значений использовано обозначение  $\delta$

$$\delta I(E)^{(p)} = (1/\hbar) \sum_q T^{(pq)} (\delta \mu_p - \delta \mu_q). \quad (\text{П.3.19})$$

Заметим, что эти обозначения широко используются в физике мезоскопических систем (см., например, [5] и главу 2 книги [13]).

## П.4. Электронный ток и неупругие процессы

Ранее мы видели (см. рис. 10.3.1), что неупругие процессы внутри канала могут быть проанализированы в рамках модели, согласно которой *канал сам по себе обладает свойствами контакта*. Имея это в виду, рассмотрим задачу, в которой электрон может войти из контакта в канал с испусканием (рождением) или поглощением (уничтожением) фонара (рис. П.4.1).

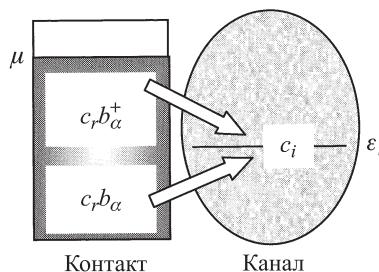


Рис. П.4.1

Отметим, что в общем случае мы не можем предполагать, что «контакт» находится в состоянии равновесия, поскольку он в данном случае

представляет также и область канала, другими словами, нам не следует пользоваться уравнениями (П.2.6a), (П.2.6b). Мы начнем с записи уравнений движения для полевых операторов в области канала:

$$i\hbar \frac{d}{dt} c_i - \varepsilon_i c_i = s_i,$$

где

$$s_i = \sum_{r,\alpha} \tau_{ir\alpha} c_r b_\alpha + \tau_{ri\alpha}^* c_r b_\alpha^+. \quad (\text{П.4.1})$$

Подобные уравнения движения, описывающие эволюцию во времени полевых операторов, находятся по общим правилам записи гайзенберговских уравнений движения произвольного оператора  $P$ :

$$i\hbar dP/dt = PH - HP,$$

где  $H$  — гамильтониан в представлении вторичного квантования:

$$\begin{aligned} H = & \sum_i \varepsilon_i c_i^+ c_i + \sum_i \varepsilon_i c_i^+ c_i + \sum_\alpha \hbar \omega_\alpha (b_\alpha^+ b_\alpha + 1/2) + \\ & + \sum_{i,r,\alpha} (\tau_{ir\alpha} c_i^+ c_r b_\alpha + \tau_{ir\alpha}^* c_r^+ c_i b_\alpha^+ + \tau_{ri\alpha} c_r^+ c_i b_\alpha + \tau_{ri\alpha}^* c_i^+ c_r b_\alpha^+). \end{aligned} \quad (\text{П.4.2})$$

Последнее слагаемое в уравнении (П.4.2), отвечающее электрон-фононному взаимодействию, определяется потенциалом  $U(x)$ , который описывает действие на электрон со стороны фононной моды  $\alpha$  (величина  $\tau_\alpha(\vec{x})$  подобна потенциалу  $U_\beta^{ab}(\vec{r})$  в уравнении (10.1.17)):

$$U(\vec{x}) = \sum_\alpha b_\alpha \tau_\alpha(\vec{x}) + b_\alpha^+ \tau_\alpha^+(\vec{x})$$

и последующей записи его в представлении вторичного квантования, как это делается для любого потенциала  $U(\vec{x})$ :

$$H_{\text{el-ph}} = \sum_{i,r} U_{ir} c_i^+ c_r + U_{ri} c_r^+ c_i.$$

Наша ближайшая цель состоит в последовательном вычислении этих двух слагаемых, описывающих источник в уравнении (П.4.1).

**Вычисление  $c_r b_\alpha$ .** Стартуем, записав уравнения движения для операторов одного электрона и одного фонона:

$$i\hbar \frac{d}{dt} c_r = \varepsilon_r c_r + \sum_{i,\alpha} \tau_{ri\alpha} c_i b_\alpha + \tau_{ir\alpha}^* c_i b_\alpha^+,$$

$$i\hbar \frac{d}{dt} b_\alpha = \hbar \omega_\alpha b_\alpha + \sum_{i,r} \tau_{ri\alpha}^* c_i^+ c_r + \tau_{ir\alpha}^* c_r^+ c_i,$$

и запишем уравнение для эволюции произведения операторов:

$$\begin{aligned} i\hbar \frac{d}{dt} c_r b_\alpha &= \left( i\hbar \frac{d}{dt} c_r \right) b_\alpha + c_r \left( i\hbar \frac{d}{dt} b_\alpha \right) = \\ &= \left( \varepsilon_r c_r + \sum_{j,\beta} \tau_{rj\beta} c_j b_\beta + \tau_{jr\beta}^* c_j b_\beta^+ \right) b_\alpha = \\ &\quad + c_r \left( \hbar \omega_\alpha b_\alpha + \sum_{j,s} (\tau_{sj\alpha}^* c_j^+ c_s + \tau_{js\alpha}^* c_s^+ c_j) \right), \end{aligned}$$

которое можно упростить, заменив произведения с переменными, отвечающими резервуару, на их средние значения (которые не зависят от времени). Это с учетом того, что  $\langle b_\beta b_\alpha \rangle = 0$  и  $\langle c - r c_s \rangle = 0$  приведет нас к уравнениям

$$\begin{aligned} i\hbar \frac{d}{dt} c_r b_\alpha - (\varepsilon_r + \hbar \omega_\alpha) c_r b_\alpha &= \\ = \sum_j \sum_{s,\beta} \tau_{js\beta}^* (\langle b_\beta^+ b_\alpha \rangle \delta_{rs} + \langle c_r c_s^+ \rangle \delta_{\alpha\beta}) c_j &= \\ = \sum_j \sum_{s,\beta} \tau_{js\beta}^* (\langle b_\beta^+ b_\alpha \rangle \langle c_s c_r^+ \rangle + \langle c_r c_s^+ \rangle \langle b_\alpha b_\beta^+ \rangle) c_j. & \end{aligned}$$

При выполнении последнего преобразования были использованы соотношения  $\delta_{rs} = c_r^+ c_s + c_s c_r^+$  (см. уравнение (П.1.3а)) и  $\delta_{\alpha\beta} = b_\alpha b_\beta^+ - b_\beta^+ b_\alpha$  (см. уравнение (П.1.9а)).

Аналогично проделанному ранее с уравнением (П.2.3), мы можем записать полное решение в виде суммы решения для изолированной системы и отклика, обусловленного слагаемым типа источника:

$$c_r(t) b_\alpha(t) = C_r(t) B_\alpha(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{js\beta}^* g_{r\alpha}(t, t_1) \times$$

$$\times \left( \langle b_\beta^+(t_1) b_\alpha(t_1) \rangle \langle c_s^+(t_1) c_r(t_1) \rangle + \langle c_r(t_1) c_s^+(t_1) \rangle \langle b_\alpha(t_1) b_\beta^+(t_1) \rangle \right),$$

где

$$i\hbar \frac{d}{dt} g_{r\alpha}(t, t_1) - (\varepsilon_r - i\eta + \hbar\omega_\alpha) g_{r\alpha}(t, t_1) = \delta(t - t_1).$$

В низшем порядке имеем  $c_r(t)b_{\alpha(t)} \approx g_{r\alpha}(t, t_1)c_r(t_1)b_\alpha(t_1)$ , откуда

$$\begin{aligned} c_r(t)b_\alpha(t) &= C_r(t)B_\alpha(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{js\beta}^* \times \\ &\times \left( \langle b_\beta^+(t_1) b_\alpha(t_1) \rangle \langle c_s^+(t_1) c_r(t_1) \rangle + \langle c_r(t_1) c_s^+(t_1) \rangle \langle b_\alpha(t_1) b_\beta^+(t_1) \rangle \right) = \\ &= C_r(t)B_\alpha(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{js\beta}^* \times \\ &\times \left( G_{\alpha\beta}^{ab}(t, t_1) G_{r,s}^n(t, t_1) + G_{r,s}^p(t, t_1) G_{\alpha\beta}^{em}(t, t_1) \right). \quad (\Pi.4.3a) \end{aligned}$$

**Расчет  $c_r b_\alpha^+$ .** После того, как рассчитано произведение  $c_r b_\alpha$ , следующим шагом будет расчет  $c_r b_\alpha^+$ . Имея в виду схему расчетов, проделанных выше, начнем с уравнений

$$i\hbar \frac{d}{dt} c_r = \varepsilon_r c_r + \sum_{i,\alpha} \tau_{ri\alpha} c_i b_\alpha + \tau_{ir\alpha}^* c_i b_\alpha^+$$

и

$$-i\hbar \frac{d}{dt} b_\alpha^+ = \hbar\omega_\alpha b_\alpha^+ + \sum_{i,r} \tau_{ir\alpha} c_i^+ c_r + \tau_{ri\alpha} c_r^+ c_i$$

и запишем уравнения эволюции их произведения:

$$\begin{aligned} i\hbar \frac{d}{dt} c_r b_\alpha^+ &= \left( i\hbar \frac{d}{dt} c_r \right) b_\alpha^+ - c_r \left( -i\hbar \frac{d}{dt} b_\alpha^+ \right) = \\ &= \left( (\varepsilon_r - i\eta) c_r + \sum_{j,\beta} \tau_{rj\beta} c_j b_\beta + \tau_{jr\beta}^* c_j b_\beta^+ \right) b_\alpha^+ - \\ &- c_r \left( \hbar\omega_\alpha b_\alpha^+ + \sum_{j,s} (\tau_{js\alpha} c_j^+ c_s + \tau_{sj\alpha} c_s^+ c_j) \right), \end{aligned}$$

которое опять можно упростить, заменив произведения переменных резервуара их средними значениями:

$$\begin{aligned} i\hbar \frac{d}{dt} c_r b_\alpha^+ - (\varepsilon_r - \hbar\omega_\alpha) c_r b_\alpha^+ &= \sum_j \sum_{s,\beta} \tau_{sj\beta} (\langle b_\beta b_\alpha^+ \rangle \delta_{rs} - \langle c_r c_s^+ \rangle \delta_{\alpha\beta}) c_j = \\ &= \sum_j \sum_{s,\beta} \tau_{sj\beta} (\langle b_\beta b_\alpha^+ \rangle \langle c_s^+ c_r \rangle + \langle c_r c_s^+ \rangle \langle b_\alpha^+ b_\beta \rangle) c_j. \end{aligned}$$

Далее по аналогии с расчетами, проведенными в предыдущем пункте, запишем решение в виде суммы решения для изолированной системы и отклика, индуцированного слагаемым типа источника:

$$\begin{aligned} c_r(t) b_\alpha^+(t) &= C_r(t) B_\alpha^+(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{sj\beta} \bar{g}_{r\alpha}(t, t_1) \times \\ &\quad \times (\langle b_\beta(t_1) b_\alpha^+(t_1) \rangle \langle c_s^+(t_1) c_r(t_1) \rangle + \langle c_r(t_1) c_s^+(t_1) \rangle \langle b_\alpha^+(t_1) b_\beta(t_1) \rangle), \end{aligned}$$

где

$$i\hbar \frac{d}{dt} \bar{g}_{r\alpha}(t, t_1) - (\varepsilon_r - i\eta - \hbar\omega_\alpha) \bar{g}_{r\alpha}(t, t_1) = \delta(t - t_1).$$

Выполняя преобразования по той же схеме, что и в предыдущем пункте, найдем, что в низшем порядке теории возмущений

$$\begin{aligned} c_r(t) b_\alpha^+(t) &\approx C_r(t) B_\alpha^+(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{sj\beta} \times \\ &\quad \times (\langle b_\beta(t_1) b_\alpha^+(t_1) \rangle \langle c_s^+(t_1) c_r(t_1) \rangle + \langle c_r(t_1) c_s^+(t_1) \rangle \langle b_\alpha^+(t_1) b_\beta(t_1) \rangle) = \\ &= C_r(t) B_\alpha^+(t) + \int_{-\infty}^{+\infty} dt_1 \sum_j c_j(t_1) \sum_{s,\beta} \tau_{sj\beta} \times \\ &\quad \times (G_{\beta\alpha}^{\text{em}}(t_1, t) G_{r,s}^n(t, t_1) + G_{r,s}^p(t, t_1) G_{\beta\alpha}^{\text{ab}}(t_1, t)). \quad (\text{П.4.3b}) \end{aligned}$$

**Исключение переменных, описывающих резервуар.** Подставляя выражения (П.4.3a), (П.4.3b) в уравнение (П.4.1), получим уравнение для по-левых операторов в канале (сравните с уравнениями (П.2.7)–(П.2.9)):

$$i\hbar \frac{d}{dt} c_i - \varepsilon_i c_i - \sum_j \int_{-\infty}^{+\infty} dt_1 \Sigma_{ij}(t, t_1) c_j(t) = S_i(t), \quad (\text{П.4.4})$$

где

$$S_i(t) \equiv \sum_{r,\alpha} \tau_{ir\alpha} C_r(t) B_\alpha(t) + \tau_{ri\alpha}^* C_r(t) B_\alpha^+(t), \quad (\text{П.4.5})$$

$$\Sigma_{ij}(t, t_1) \equiv \vartheta(t - t_1) \Gamma_{ij}(t, t_1) \quad (\text{П.4.6})$$

и

$$\begin{aligned} \Gamma_{ij}(t, t_1) = & \sum_{r,s,\alpha,\beta} \tau_{ir\alpha} \tau_{js\beta}^* (G_{\alpha\beta}^{\text{ab}}(t, t_1) G_{rs}^{\text{n}}(t, t_1) + G_{rs}^{\text{p}}(t, t_1) G_{\alpha\beta}^{\text{em}}(t, t_1)) \\ & + \tau_{sj\beta} \tau_{ri\alpha}^* (G_{\beta\alpha}^{\text{em}}(t_1, t) G_{rs}^{\text{n}}(t, t_1) + G_{rs}^{\text{p}}(t, t_1) G_{\beta\alpha}^{\text{ab}}(t_1, t)). \end{aligned} \quad (\text{П.4.7})$$

**Собственно-энергетическая функция, описывающая рассеяние в канал.** Сейчас мы вычислим функцию, описывающую рассеивание электронов, входящих в канал:

$$\Sigma_{ij}^{\text{in}}(t, t') \equiv \langle S_j^+(t') S_i(t) \rangle,$$

следуя тому же плану, что и ранее, т. е. подставив для слагаемого типа источника выражение (П.4.5):

$$\begin{aligned} \Sigma_{ij}^{\text{in}}(t, t') = & \sum_{r,s,\alpha,\beta} \tau_{ir\alpha} \tau_{js\beta}^* \langle C_s^+(t') C_r(t) \rangle \langle B_\beta^+(t') B_\alpha(t) \rangle + \\ & + \tau_{sj\beta} \tau_{ri\alpha}^* \langle C_s^+(t') C_r(t) \rangle \langle B_\beta(t') B_\alpha^+(t) \rangle, \end{aligned}$$

откуда найдем

$$\Sigma_{ij}^{\text{in}}(t, t') = \sum_{r,s,\alpha,\beta} \tau_{ir\alpha} \tau_{js\beta}^* G_{rs}^{\text{n}}(t, t') G_{\alpha\beta}^{\text{ab}}(t, t') + \tau_{sj\beta} \tau_{ri\alpha}^* G_{rs}^{\text{n}}(t, t') G_{\beta\alpha}^{\text{em}}(t', t). \quad (\text{П.4.8})$$

**Стационарное состояние.** В стационарных условиях можно выполнить преобразование Фурье по переменной, равной разности временных координат. В результате из уравнения (П.4.7) найдем

$$\begin{aligned} \Gamma_{ij}(E) = & \int_0^\infty d(\hbar\omega)/2\pi \times \\ & \times \sum_{r,s,\alpha,\beta} \tau_{ir\alpha} \tau_{js\beta}^* (G_{rs}^{\text{n}}(E - \hbar\omega) G_{\alpha\beta}^{\text{ab}}(\hbar\omega) + G_{rs}^{\text{p}}(E - \hbar\omega) G_{\alpha\beta}^{\text{em}}(\hbar\omega)) + \\ & + \tau_{sj\beta} \tau_{ri\alpha}^* (G_{rs}^{\text{n}}(E + \hbar\omega) G_{\beta\alpha}^{\text{em}}(\hbar\omega) + G_{rs}^{\text{p}}(E + \hbar\omega) G_{\beta\alpha}^{\text{ab}}(\hbar\omega)), \end{aligned} \quad (\text{П.4.9})$$

а из уравнения (П.4.8) получим Фурье-образ  $\Sigma_{ij}^{\text{in}}(t, t')$ :

$$\begin{aligned}\Sigma_{ij}^{\text{in}}(E) = \int_0^\infty d(\hbar\omega)/2\pi \sum_{r,s,\alpha,\beta} \tau_{ir\alpha} \tau_{js\beta}^* G_{rs}^{\text{n}}(E - \hbar\omega) G_{\alpha\beta}^{\text{ab}}(\hbar\omega) + \\ + \tau_{sj\beta} \tau_{ri\alpha}^* G_{rs}^{\text{n}}(E + \hbar\omega) G_{\beta\alpha}^{\text{em}}(\hbar\omega).\end{aligned}\quad (\text{П.4.10})$$

Если мы примем предположение о том, что фононы остаются в состоянии равновесия, то мы можем использовать уравнения (П.1.15), чтобы получить уравнение (10.3.5) из уравнения (П.4.10), заметив, что роль величины  $\tau_\alpha(\vec{x})$  аналогична потенциалу  $U_\beta^{\text{ab}}(\vec{r})$ , встречавшемуся в параграфе 10.3. Выражение (П.4.9) дает нам, таким образом, полный тензорный вариант формулы (10.3.9).

**Ток.** Заметим, что нам нет необходимости заново получать выражения для входящего и выходящего токов (см. уравнения (П.3.3)–(П.3.5)), поскольку они были получены с использованием уравнения (П.2.7), которое все еще является справедливым (см. уравнение (П.4.4)). Изменения затрагивают выражение для собственно-энергетической функции  $[\Sigma]$  (сравните выражения (П.2.8) и (П.4.7)), а также собственно-энергетическую функцию  $[\Sigma^{\text{in}}]$  (сравните выражения (П.2.19) и (П.4.8)). При этом, однако, основные формулы для тока, выражаемого через эти величины, остаются теми же самыми.

## П.5. Кулоновская блокада и резонанс Кондо

Далее мы рассмотрим задачу, где фигурирует строго когерентное взаимодействие с контактом, но при этом учитывается кулоновское взаимодействие между двумя спиновыми уровнями внутри канала, обозначаемыми как  $c$  и  $d$  (рис. П.5.1). Для простоты мы будем рассматривать лишь равновесную задачу с одним контактом и учтем в наших расчетах лишь один уровень каждого типа:

$$i\hbar \frac{d}{dt} c = \varepsilon c + \sum_r \tau_r c_r + U d^+ d c; \quad (\text{П.5.1a})$$

$$i\hbar \frac{d}{dt} c_r = \varepsilon_r c_r + \tau_r^* c. \quad (\text{П.5.1b})$$

Далее расчет продолжается по той же схеме, что и при выводе уравнения (П.2.7) из уравнения (П.2.1). Результат будет аналогичным, за исключением

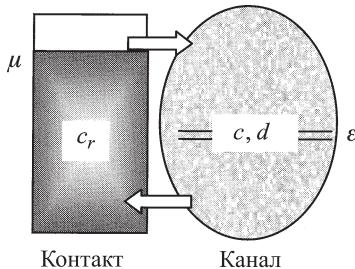


Рис. П.5.1

чением дополнительного слагаемого  $Ud^+dc$ :

$$i\hbar \frac{d}{dt}c - \varepsilon c - \int_{-\infty}^{+\infty} dt_1 \Sigma_0(t, t_1)c(t_1) = S^c(t) + U d^+(t)d(t)c(t). \quad (\text{П.5.2})$$

Выполняя преобразование Фурье, приходим к уравнению

$$(E - \varepsilon - \Sigma_0)c = S^c + U\{d^+dc\}(E), \quad (\text{П.5.3})$$

$$\Sigma_0 \equiv \sum_r \frac{|\tau_r|^2}{E - \varepsilon_r + i\eta}, \quad S^c \equiv \sum_r \tau_r C_r,$$

где фигурные скобки, в которые заключено слагаемое  $d^+dc$ , означают, что от данного произведения берется Фурье-образ, не сводящийся к произведению Фурье-образов сомножителей.

Аналогично можно преобразовать и уравнения для состояния с отрицательной проекцией спина

$$i\hbar \frac{d}{dt}d = \varepsilon d + \sum_r \tau_r d_r + U c^+ cd \quad (\text{П.5.4a})$$

и

$$i\hbar \frac{d}{dt}d_r = \varepsilon_r d_r + \tau_r^* d, \quad (\text{П.5.4b})$$

откуда для  $d$  следует уравнение

$$(E - \varepsilon - \Sigma_0)d = S^d + U \{c^+ cd\}(E), \quad (\text{П.5.5})$$

в котором  $S^d \equiv \sum_r \tau_r D_r$ .

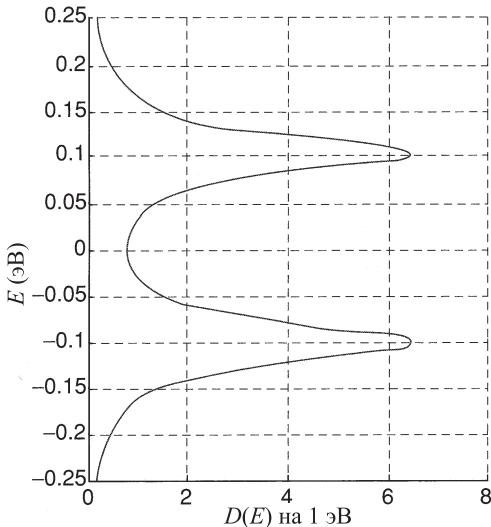


Рис. П.5.2. Плотность состояний  $D(E) = i[G - G^+]$ , рассчитанная с помощью функции Грина  $G(E)$ , удовлетворяющей уравнению (П.5.20) с параметрами  $\epsilon = -0.1$  эВ,  $U = +0.2$  эВ,  $\langle d^+d \rangle = 0.5$  и величиной  $\Sigma$ , найденной из уравнения (П.5.4). Предполагалось, что резервуар состоит из большого числа эквидистантно расположенных уровней, расстояние между которыми 0.4 мэВ на интервале энергии от  $-1$  эВ до  $+1$  эВ,  $\eta = 1$  мэВ и  $\tau_r = 1.8$  мэВ

В этом параграфе мы покажем, что при учете в рамках метода самосогласованного поля новых слагаемых  $Ud^+dc$  и  $Uc^+cd$  на графике спектральной функции появляется один максимум, как это было показано на рис. 1.5.1, а при использовании более точной аппроксимации для этих слагаемых возникают два пика, как это показано на рис. П.5.2. Эти пики следует рассматривать как комбинацию двух возможных вариантов, представленных на рис. 1.5.2. Если еще улучшить наше приближение, будет обнаружен локальный максимум и в центре при  $E = \mu$ , как это видно на рис. П.5.3. Последний пик является ответственным за возрастание при понижении температуры сопротивления объемных металлических образцов, содержащих магнитные примеси. Объяснение этого явления было дано Кондо в 1960-х годах (см., например, [55, 28]). В более поздних экспериментах, выполненных в 1990-х годах, было также обнаружено возрастание проводимости одиночных квантовых точек и молекул, что связано с тем же максимумом Кондо в зависимости спектральной функции от энергии ([40]).

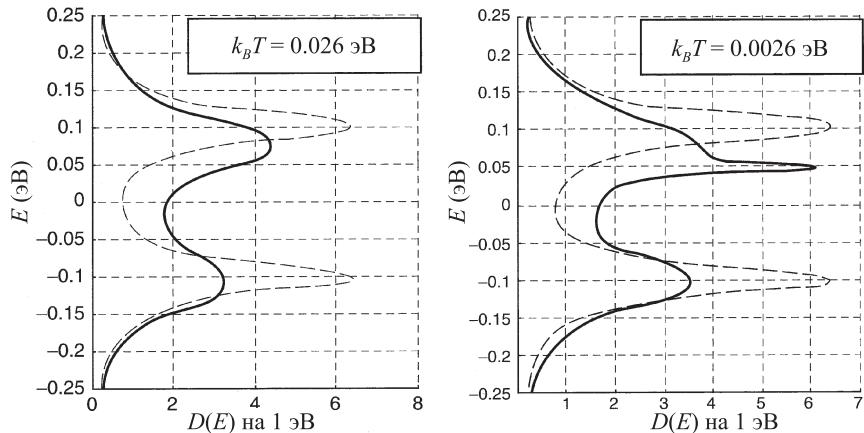


Рис. П.5.3. Плотность состояний  $D(E) = i[G - G^+]$ , рассчитанная с помощью функции Грина, удовлетворяющей уравнению  $G(E)$  (П.5.20) при  $\mu = 0.05$  эВ и теми же значениями остальных параметров, что на рис. П.5.2 (для сравнения на рисунке показан результат расчета с функцией  $G(E)$  из уравнения (П.5.12)). Обратите внимание на то, что при более низкой температуре «пик Кондо» лежит при  $\mu = 0.05$  эВ (автор признателен А. В. Гошу и Д. Сену за их содействие при подготовке данного примера)

**Приближение самосогласованного поля (CCP).** В наиболее простом приближении запишем

$$\{c^+cd\}(E) \approx \langle c^+c \rangle d(E), \quad (\text{П.5.6}a)$$

$$\{d^+dc\}(E) \approx \langle d^+d \rangle c(E), \quad (\text{П.5.6}b)$$

откуда с помощью (П.5.4) и (П.5.5) находим, что

$$c = \frac{1}{E - \varepsilon - \Sigma_0 - U \langle d^+d \rangle} \quad \begin{matrix} S^c \\ \text{отвечает } G_c(E) \end{matrix} \quad (\text{П.5.7}a)$$

и

$$d = \frac{1}{E - \varepsilon - \Sigma_0 - U \langle c^+c \rangle} \quad \begin{matrix} S^d \\ \text{отвечает } G_d(E) \end{matrix} \quad (\text{П.5.7}b)$$

Данные выражения фактически представляют собой полное приближение CCP, хотя для полноты описания нам следовало бы обсудить то, каким

образом мы будем вычислять средние значения  $\langle c^+ c \rangle$  и  $\langle d^+ d \rangle$ . Мы не будем обсуждать здесь этот вопрос, а вместо этого примем предположение о том, что величины  $\langle c^+ c \rangle$  и  $\langle d^+ d \rangle$  являются заданными параметрами, и обсудим то, как можно улучшить наши приближения для величин  $G_c$  и  $G_d$ , определяемых соотношениями  $c = G_c S^c$  и  $d = G_d S^d$ .

**Кулоновская блокада.** Используя уравнение эволюции для  $d^+ dc$ , заменим уравнение (П.5.6б) более точным выражением:

$$i\hbar \frac{d}{dt} d^+ dc = (\varepsilon + U) d^+ dc + \sum_r (\tau_r^* d_r^+ cd + \tau_r d^+ d_r c + \tau_r d^+ dc_r) \approx \quad (\text{П.5.8})$$

$$\begin{aligned} & \approx (\varepsilon + U) d^+ dc + \sum_r \tau_r d^+ dc_r \rightarrow \\ & \rightarrow (E - \varepsilon - U) \{d^+ dc\} = \sum_r \tau_r d^+ dc_r. \end{aligned} \quad (\text{П.5.9})$$

Используя (П.5.2), можно записать

$$\{d^+ dc\}(E) = \sum_r \frac{\tau_r \langle d^+ d \rangle C_r}{E - \varepsilon - U - \Sigma_0} = \frac{\langle d^+ d \rangle}{E - \varepsilon - U - \Sigma_0} S^c. \quad (\text{П.5.10})$$

Подставляя это в уравнение (П.5.4), придем к уравнению

$$(E - \varepsilon - U - \Sigma_0) c = \left( 1 + \frac{U \langle d^+ d \rangle}{E - \varepsilon - U - \Sigma_0} \right) S^c, \quad (\text{П.5.11})$$

с помощью которого функцию Грина  $G(E)$  можно записать в виде

$$\begin{aligned} c &= \left( \frac{1}{E - \varepsilon - \Sigma_0} + \frac{U \langle d^+ d \rangle}{(E - \varepsilon - \Sigma_0)(E - \varepsilon - U - \Sigma_0)} \right) S^c = \\ &= \left( \frac{1 - \langle d^+ d \rangle}{E - \varepsilon - \Sigma_0} + \frac{\langle d^+ d \rangle}{E - \varepsilon - U - \Sigma_0} \right) S^c. \quad (\text{П.5.12}) \end{aligned}$$

Выражение в круглых скобках, стоящее в нижней строке перед  $S^c$ , и есть функция Грина  $G(E)$ . Пример плотности состояний  $D(E) = i[G - G^+]$ , рассчитанный с помощью найденной выше функции  $G(E)$ , показан на рис. П.5.2 для системы с типичными значениями параметров. Можно

сделать вывод, что найденная зависимость ведет себя аналогично тому, что ожидалось нами в условиях кулоновской блокады (см. рис. 1.5.2).

**Резонанс Кондо.** Чтобы двинуться дальше, нам необходимо вернуться на шаг назад к уравнению (П.5.8):

$$\begin{aligned} i\hbar \frac{d}{dt} d^+ dc &= (\varepsilon + U) d^+ dc + \sum_r (\tau_r^* d_r^+ cd + \tau_r d^+ d_r c + \tau_r d^+ d_c r) \rightarrow \\ \rightarrow (E - \varepsilon - U) \{d^+ dc\} &= \sum_r (\tau_r^* \{d_r^+ cd\} + \tau_r \{d^+ d_r c\} + \tau_r \{d^+ d_c r\}), \quad (\text{П.5.13}) \end{aligned}$$

и проделать более точный расчет по сравнению с тем, который привел нас к выражению (П.5.9). Начнем с рассмотрения зависящих от времени уравнений для каждой из трех величин, представленных в (П.5.13):

$$\begin{aligned} i\hbar \frac{d}{dt} d^+ d_c r &= (\varepsilon_r - i\eta) d^+ d_c r + \sum_r (\tau_r^* d^+ dc + \tau_r d^+ d_r c + \tau_r^* d_r^+ c_r d) \approx \\ &\approx (\varepsilon_r - i\eta) d^+ d_c r + \sum_r \tau_r^* d^+ dc; \\ \{d^+ d_c r\} &= \langle d^+ d \rangle C_r + \sum_r \frac{\tau_r^*}{E - \varepsilon_r + i\eta} \{d^+ dc\}; \quad (\text{П.5.14}) \end{aligned}$$

$$\begin{aligned} i\hbar \frac{d}{dt} d^+ d_r c &= \varepsilon d^+ d_r c + \sum_r (\tau_r^* d^+ dc + \tau_r d^+ d_r c - \tau_r^* d_r^+ d_r c) \approx \\ &\approx \varepsilon_r d^+ d_r c + \sum_r \tau_r^* d^+ dc - \tau_r^* f_r c; \\ \{d^+ d_r c\} &= \sum_r \frac{\tau_r^*}{E - \varepsilon_r + i\eta} \{d^+ dc\} - \frac{\tau_r^* f_r}{E - \varepsilon_r + i\eta} c; \quad (\text{П.5.15}) \end{aligned}$$

$$\begin{aligned} i\hbar \frac{d}{dt} d_r^+ cd &= (2\varepsilon + U - \varepsilon_r - i\eta) d_r^+ cd + \sum_r (\tau_r d^+ dc - \tau_r d_r^+ d_r c - \tau_r d_r^+ c_r c) \approx \\ &\approx (2\varepsilon_r - \varepsilon_r - i\eta) d_r^+ cd + \sum_r (\tau_r d^+ dc - \tau_r f_r c); \\ \{d^+ cd\} &\approx \sum_r \frac{\tau_r}{E - 2\varepsilon - U + \varepsilon_r + i\eta} \{d^+ dc\} - \frac{\tau_r f_r}{E - 2\varepsilon - U + \varepsilon_r + i\eta} c. \quad (\text{П.5.16}) \end{aligned}$$

Подставляя выражения (П.5.14), (П.5.15) и (П.5.16) в (П.5.13), мы найдем более точное по сравнению с (П.5.10) приближение для величины  $\{d^+dc\}$ :

$$(E - \varepsilon - U - 2\Sigma_0 - \Sigma_1)\{d^+dc\} = \langle d^+d \rangle S^c - (\Sigma_2 + \Sigma_3)c$$

$$\rightarrow \{d^+dc\} = \frac{\langle d^+d \rangle}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} S^c - \frac{\Sigma_2 + \Sigma_3}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} c, \quad (\text{П.5.17})$$

где

$$\Sigma_1 \equiv \sum_r \frac{|\tau_r|^2}{E - 2\varepsilon - U + \varepsilon_r + i\eta},$$

$$\Sigma_2 \equiv \sum_r \frac{|\tau_r|^2 f_r}{E - \varepsilon_r + i\eta}, \quad (\text{П.5.18})$$

$$\Sigma_3 \equiv \sum_r \frac{|\tau_r|^2 f_r}{E - 2\varepsilon - U + \varepsilon_r + i\eta}.$$

Подставляя (П.5.17) в (П.5.13), приходим к уравнению

$$\left( E - \varepsilon - E_0 + \frac{U(\Sigma_2 + \Sigma_3)}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} \right) c =$$

$$= \left( 1 + \frac{U \langle d^+d \rangle}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} \right) S^c, \quad (\text{П.5.19})$$

откуда следует, что функция Грина  $G(E)$ , удовлетворяющая условию  $c = GS^c$ , определяется выражением

$$G = \left[ 1 - \langle d^+d \rangle / \left( E - \varepsilon - E_0 + \frac{U(\Sigma_2 + \Sigma_3)}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} \right) \right] +$$

$$+ \left[ \langle d^+d \rangle / \left( E - \varepsilon - \Sigma_0 - U - \frac{U(\Sigma_0 + \Sigma_1 - \Sigma_2 - \Sigma_3)}{E - \varepsilon - U - 2\Sigma_0 - \Sigma_1} \right) \right]. \quad (\text{П.5.20})$$

Плотность состояний  $D(E) = i[G - G^+]$ , рассчитанная с помощью найденной функции  $G(E)$  из выражения (П.5.20), представлена на рис. П.5.3 при тех же значениях параметров, что и на рис. П.5.2 (также для сравнения

показан результат с функцией  $G(E)$ , рассчитанной по формуле (П.5.12)). Обратите внимание на то, как возрастают амплитуда центрального «пика Кондо» при понижении температуры. В этой связи следует отметить, что при низких температурах и при больших амплитудах взаимодействия с резервуаром применимость использованного нами метода может оказаться под вопросом. В подобных условиях может оказаться необходимым использование более высоких порядков теории возмущений (вплоть до бесконечности!), поэтому данная задача оставалась одной из наиболее сложных в многоэлектронной теории на протяжении долгих лет.

**Выводы.** Мы получили выражения для тока, отправляясь от описания (на языке вторичного квантования) составной системы, включающей канал и контакт. В ходе расчетов были исключены из рассмотрения переменные, отвечающие контакту, и получено уравнение только для канала:

$$i\hbar \frac{d}{dt}c - Hc - (\Sigma c)_{\text{входящий ток}} = S_{\text{выходящий ток}}.$$

Мы обнаружили, что квантовомеханический расчет можно производить и с одноэлектронной функцией, т. е. вне рамок представления вторичного квантования, но такой подход не позволяет описать эффекты электронной блокады, связанной с принципом Паули. Например, возникает вопрос о том, зависит ли величина  $\Sigma$  в слагаемом, отвечающем исходящему току, от того, занято конечное состояние или свободно. Для когерентных процессов ответ обычно отрицательный, но это не всегда так, как, например, для эффекта Кондо. На подобные вопросы трудно дать ответ, оставаясь в рамках одноэлектронной картины, однако требуемые ответы появляются естественным путем при использовании метода вторичного квантования. В сильно взаимодействующих системах подобные подходы могут потребоваться для записи уравнений, описывающих транспорт не исходной «голой» частицы — электрона, которому отвечает оператор  $c$ , а «одетой», или композитной, квазичастицы, получаемой с помощью подходящих унитарных преобразований, включающих также часть состояний из «контакта».

# Программы на языке MATLAB для построения иллюстраций

Все права на представленные здесь программы принадлежат Суприйо Датта. Планируется разместить электронные версии этих программ на моем Интернет-сайте и на файловом сервере, где можно запускать их работу удаленно без установки на свой компьютер.

## Глава 1

### %Рис.1.1.1

```
clear all
```

```
%Параметры (Все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; eps0=8.854E-12; epsr=4; m=0.25*9.11e-31; %Эффективная масса I0=q*q/hbar;
```

```
%Параметры
```

```
W=1e-6; L=10e-9; %W=Ширина, L=Длина активной области, t=толщина слоя оксида
```

```
Cg=epsr*eps0*W*L/t; Cs=0.05*Cg; Cd=0.05*Cg; CE=Cg+Cs+Cd; U0=q/CE;
alphag=Cg/CE, alphad=Cd/CE
```

```
%alphag=1; alphad=0.5; U0=0.25;
```

```
kT=0.025; mu=0; ep=0.2;
```

```
v=1e5; %Скорость покидания
```

```
g1=hbar*v/(q*L); g2=g1; g=g1+g2;
```

```
%g1=0.005; g2=0.005; g=g1+g2;
```

```
%Energy grid
```

```
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
```

```
D0=m*q*W*L/(pi*hbar*hbar); % Шаговая плотность состояний на 1 эВ
```

```
D=D0*[zeros(1,251) ones(1,250)];
```

```
%D=(2*g/(2*pi))./((E.^2)+((g/2).^2)); %Лоренцевская плотность состояний
```

```
1 эВ %D=D./(dE*sum(D)); %Нормировка на единицу
```

```

%Число электронов
f0=1./(1+exp((E+ep-mu)./kT)); N0=2*dE*sum(D.*f0); ns=N0/(L*W*1e4),
%/cm^2

%Напряжение
IV=6; VV=linspace(0,0.6,IV);
for iV=1:IV
    Vg=0.5; Vd=VV(iV);
    %Vd=0.5; Vg=VV(iV);
    mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

    U=0; %Самосогласованное поле
    dU=1;
    while dU>1e-6
        f1=1./(1+exp((E+UL+U+ep-mu1)./kT));
        f2=1./(1+exp((E+UL+U+ep-mu2)./kT));
        N(iV)=dE*sum(D.*((f1.*g1/g)+(f2.*g2/g)));
        Unew=U0*(N(iV)-N0); dU=abs(U-Unew);
        U=U+0.1*(Unew-U);
    end
    I(iV)=dE*I0*(sum(D.*(f1-f2)))*g1*g2/g;
end

hold on
h=plot(VV,I,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('Voltage (V) ->')
ylabel('Current (A) ->')
grid on

```

**%Рис.1.1.3**

```

clear all

E=linspace(-.25,.25,501); dE=E(2)-E(1); kT=0.025; Ef=0;
V=0; mu1=Ef+(V/2); mu2=Ef-(V/2);
f1=1./(1+exp((E-mu1)./kT)); f2=1./(1+exp((E-mu2)./kT));
%dE*(sum(f1-f2))/V

hold on
h=plot(f1,E,'g');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])

```

```
xlabel('Fermi function ->')
ylabel('E - mu (eV) ->')
grid on
```

**%Рис.1.3.3, 1.5.1**

```
clear all

E=linspace(-.5,.5,50001); dE=E(2)-E(1); gam=0.05;
D=(gam/(2*pi))./((E.^2)+((gam/2)^2));
%D=(gam/(2*pi))./(((E-0.25).^2)+((gam/2)^2)); %Для Рис.1.5.2
%D=D+((gam/(2*pi))./(((E+0.25).^2)+((gam/2)^2))); %Для Рис.1.5.2
%dE*sum(D)

hold on
h=plot(D,E,'g');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('D(E) ->')
ylabel('E(eV) ->')
grid on
```

**%Рис.1.4.6**

```
clear all

%Константы (все в МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; l0=q*q/hbar;

%Константы
U0=0.025; kT=0.025; mu=0; ep=0.2;
g1=0.005; g2=0.005; g=g1+g2;
alphag=1; alphad=0.5;

%Cетка по энергии
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
D=(g/(2*pi))./((E.^2)+((g/2)^2)); %Лоренцевская плотность состояний на 1 эВ
D=D./(dE*sum(D)); %Нормировка на единицу

%Напряжение
IV=101; VV=linspace(0,1,IV);
for iV=1:IV
    Vg=0; Vd=VV(iV);
    %Vd=0; Vg=VV(iV);
```

```

mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

U=0; %Самосогласованное поле
dU=1;
while dU>1e-6
    f1=1./(1+exp((E+ep+UL+U-mu1)./kT));
    f2=1./(1+exp((E+ep+UL+U-mu2)./kT));
    N(iV)=dE*sum(D.*((f1.*g1/g)+(f2.*g2/g)));
    Unew=U0*N(iV); dU=abs(U-Unew);
    U=U+0.1*(Unew-U);
end
I(iV)=dE*I0*(sum(D.*(f1-f2)))*(g1*g2/g);
end

hold on
h=plot(VV,N,'b');
%h=plot(VV,I,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('Voltage (V) ->')
%ylabel('Current (A) ->')
ylabel('Number of electrons ->')
grid on

```

**%Рис.Е.1.3**

```

clear all

%Константы (все в системе MKC, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; I0=q*q/hbar;

%Константы
U0=0.025; kT1=0.026; kT2=0.025; ep=0.2;
g1=0.005; g2=0.005; g=g1+g2;
alphag=1; alphad=0.5;

%Cетка по энергии
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
g1=0.005*(E+abs(E))./(E+E+1e-6); % нуль для отрицательных E
g2=0.005*ones(1,NE); g1=g2;
g=g1+g2;

%Напряжение
IV=101; VV=linspace(-0.25,0.25,IV);

```

```

for iV=1:IV
    mu1=ep+VV(iV); mu2=mu1;
    f1=1./(1+exp((E-mu1)./kT1));
    f2=1./(1+exp((E-mu2)./kT2));
    D=(g./((2*pi))./(((E-ep).^2)+(g./2).^2));
    D=D./((dE*sum(D)));
    I(iV)=dE*2*I0*(sum(D.*(f1-f2).*g1.*g2./g));
end

hold on
%h=plot(VV,N/2,'b'); % Part (a)
h=plot(VV,I,'b');
set(h,'linewidth',[2.0])
set(gca,'FontSize',[25])
xlabel('Voltage (V) ->')
ylabel('Current (A) ->')
%ylabel('Number of electrons ->')
grid on

```

#### %Рис.Е.1.4

```

clear all

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; I0=q*hbar;

%Константы
U0=0.025; kT=0.025; mu=0; ep=0.2; N0=0;
g1=0.005; g2=0.005; g=g1+g2;
alphag=1; alphad=0.5;

%Cетка по энергии
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
g1=0.005*(E+abs(E))./(E+E+1e-6); % zero for negative E
g2=0.005*ones(1,NE);
g=g1+g2;

%Напряжение
IV=101; VV=linspace(-.6,.6,IV);
for iV=1:IV
    Vg=0; Vd=VV(iV);
    %Vd=0; Vg=VV(iV);
    mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

```

```

U=0; %Самосогласованное поле dU=1;
while dU>1e-6
    f1=1./(1+exp((E-mu1)./kT));
    f2=1./(1+exp((E-mu2)./kT));
    D=(g./(2*pi))./(((E-ep-UL-U).^2)+((g./2).^2));
    D=D./((dE*sum(D)));
    N(iV)=dE*2*sum(D.*((f1.*g1./g)+(f2.*g2./g)));
    Unew=U0*(N(iV)-N0); dU=abs(U-Unew);
    U=U+0.1*(Unew-U);
end
I(iV)=dE*2*I0*(sum(D.*(f1-f2).*g1.*g2./g));
end

hold on
%h=plot(VV,N/2,'b'); % Часть (а)
h=plot(VV,I,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('Voltage (V) ->')
ylabel('Current (A) ->')
%ylabel('Number of electrons ->')
grid on

```

## Глава 2

### %Рис.2.3.2а,б

clear all

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)  
 $\hbar=1.055e-34$ ;  $m=9.110e-31$ ;  $\epsilon_{\text{р}}=8.854e-12$ ;  $q=1.602e-19$ ;

%Решетка

$N_p=100$ ;  $a=1e-10$ ;  $X=a*[1:N_p]$ ;  $t_0=(\hbar^2)/(2*m*(a^2))/q$ ;  $L=a*(N_p+1)$ ;  
 $T=(2*t_0*diag(ones(1,N_p)))-(t_0*diag(ones(1,N_p-1),1))-(t_0*diag(ones(1,N_p-1),-1))$ ;

[V,D]=eig(T); D=diag(D); [Enum,ind]=sort(D);

E1=D(ind(1)); psi1=abs(V(:,ind(1))); P1=psi1.\*conj(psi1);
E2=D(ind(25)); psi2=abs(V(:,ind(25))); P2=psi2.\*conj(psi2);

%Аналитический расчет собственных значений

```
Ean=(((hbar*pi)^2)/(2*m*(L^2))/q)*[1:Np].*[1:Np];

hold on
%h=plot(Enum,'bx'); % Part (a)
%h=plot(Ean,'b'); % Part (a)
h=plot(P1,'b'); % Part (b)
h1=plot(P2,'b'); % Part (b)
set(h,'linewidth',[3.0])
set(h1,'linewidth',[1.0])
set(gca,'Fontsize',[25])

%xlabel('Eigenvalue Number, alpha ->'); % Part (a)
%ylabel('E (eV) ->'); % Part (a)
xlabel('Lattice site # ->'); % Part (b)
ylabel('Probability ->'); % Part (b)
grid on
```

### %Рис.2.3.5

```
clear all
```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)  
 $hbar=1.055e-34$ ;  $m=9.110e-31$ ;  $epsil=8.854e-12$ ;  $q=1.602e-19$ ;

%Решетка

```
Np=100; a=1e-10; X=a*[1:1:Np]; t0=(hbar^2)/(2*m*(a^2))/q; L=a*(Np+1);
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
T(1,Np)=-t0; T(Np,1)=-t0;
```

```
[V,D]=eig(T); D=diag(D); [Enum,ind]=sort(D);
```

```
E1=D(ind(1)); psi1=abs(V(:,ind(1))); P1=psi1.*conj(psi1);
E2=D(ind(50)); psi2=abs(V(:,ind(50))); P2=psi2.*conj(psi2);
```

%Аналитические выражения для собственных значений  
 $Ean=(((hbar*pi)^2)/(2*m*(L^2))/q)*[1:Np].*[1:Np];$

```
hold on
h=plot(Enum,'bx');
set(h,'linewidth',[3.0])
set(gca,'Fontsize',[25])
xlabel('Eigenvalue Number, alpha ->');
ylabel('E (eV) ->');
grid on
```

**%Рис.2.3.6, 2.3.7**

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; epsil=8.854e-12; q=1.602e-19;
a0=4*pi*epsil*hbar*hbar/(m*q*q), E0=q/(8*pi*epsil*a0)
```

```
%Решетка
```

```
Np=100; a=(5e-10^2/Np); % *1 for Fig.1.3.6 and *2 for Fig.1.3.7
R=a*[1:1:Np]; t0=(hbar^2)/(2*m*(a^2))/q;
```

```
%Квантовые числа
```

```
n=1; l=0; % для 1s, n=1 и для 2s, n=2
```

```
%Гамильтониан,H=кинетическая энергия,K+потенциальная,U
K=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),
-1));
U=(-q/(4*pi*epsil)./R)+(l*(l+1)*hbar*hbar/(2*m*q))./(R.*R)); U=diag(U);
[V,D]=eig(K+U); D=diag(D); [DD,ind]=sort(D);
E=D(ind(n-1)); psi=V(:,ind(n-1));
P=psi.*conj(psi); [-E0/(n^2)E]
```

```
%аналитические решения
```

```
P1s=(4*a/(a0^3))*R.*R.*exp(-2*R./a0);
P2s=(4*a/(2*4*4*(a0^3)))*R.*R.*((2-(R./a0)).^2).*exp(-2*R./(2*a0));
P3s=(4*a/(3*81*81*(a0^3)))*R.*R.*((27-(18*R./a0)+(2*(R./a0).^2)).^2).*exp(-2*R./(3*a0));
P2p=(4*a/(3*32*(a0^3)))*R.*R.*((R./a0).^2).*exp(-2*R./(2*a0));
P3p=(8*a/(3*81*81*(a0^3)))*R.*R.*((6-(R./a0)).^2).*((R./a0).^2).*exp(-2*R./(3*a0));
```

```
hold on
```

```
h=plot(R,P,'b');
```

```
h=plot(R,P1s,'bx'); % use P1s for '1s' and P2s for '2s'
```

```
set(h,'linewidth',[2.0])
```

```
set(gca,'Fontsize',[25])
```

```
xlabel('x (м) —>');
```

```
ylabel('Probability —>');
```

```
grid on
```

## Глава 3

### %Рис.3.1.4

```

clear all

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; epsilon=8.854e-12; q=1.602e-19;

%Решетка
Np=200; a=(10e-10/Np); R=a*[1:Np]; t0=(hbar^2)/(2*m*(a^2))/q;

%Гамильтониан,H=кинетическая энергия,T+потенциальная,U+Uscf
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
UN=(-q^2/(4*pi*epsilon))./R; % Z=2 for Helium

Uscf=zeros(1,Np); change=1;
while change>0.01
    [V,D]=eig(T+diag(UN+Uscf)); D=diag(D); [DD,ind]=sort(D);
    E=D(ind(1)); psi=V(:,ind(1)); P=psi.*conj(psi); P=P';
    Unew=(q/(4*pi*epsilon))*((sum(P./R)-cumsum(P./R))+(cumsum(P)./R));
    change=sum(abs(Unew-Uscf))/Np, Uscf=Unew;
end

%аналитические решения для 1s состояний атома водорода
a0=4*pi*epsilon*hbar*hbar/(m*q*q);
P0=(4*a/(a0^3))*R.*R.*exp(-2*R./a0);

hold on
%h=plot(R,UN,'b'); % Part (a)
%h=plot(R,Uscf,'b'); % Part (a)
h=plot(R,P,'b'); % Part (b)
h=plot(R,P0,'bx'); % Part (b)
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('R (m) —>');
%ylabel('U (eV) —>'); % Part (a)
%axis([0 1e-9-100 20]); % Part (a)
ylabel('Probability —>'); % Part (b)
axis([0 1e-9 0 0.1]); % Part (b)
grid on

```

**%Рис.3.1.5**

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; epsil=8.854e-12; q=1.602e-19;
```

```
%Решетка
```

```
Np=200; a=(10e-10/Np); R=a*[1:1:Np]; t0=(hbar^2)/(2*m*(a^2))/q;
```

```
%Гамильтониан,Н=кинетическая энергия,T+потенциальная,U+Ul+Uscf
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
UN=(-q.^14/(4*pi*epsil))./R; % Z=14 для кремния
l=1; Ul=(l*(l+1)*hbar*hbar/(2*m*q))./(R.*R);
```

```
Uscf=zeros(1,Np); change=1;
```

```
while change>0.1
```

```
[V,D]=eig(T+diag(UN+Uscf)); D=diag(D); [DD,ind]=sort(D);
```

```
E1s=D(ind(1)); psi=V(:,ind(1)); P1s=psi.*conj(psi); P1s=P1s';
```

```
E2s=D(ind(2)); psi=V(:,ind(2)); P2s=psi.*conj(psi); P2s=P2s';
```

```
E3s=D(ind(3)); psi=V(:,ind(3)); P3s=psi.*conj(psi); P3s=P3s';
```

```
[V,D]=eig(T+diag(UN+Ul+Uscf)); D=diag(D); [DD,ind]=sort(D);
```

```
E2p=D(ind(1)); psi=V(:,ind(1)); P2p=psi.*conj(psi); P2p=P2p';
```

```
E3p=D(ind(2)); psi=V(:,ind(2)); P3p=psi.*conj(psi); P3p=P3p';
```

```
n0=(2*(P1s+P2s+P3s))+(6*P2p)+(2*P3p);
```

```
n=n0*(13/14);
```

```
Unew=(q/(4*pi*epsil))*((sum(n./R)-cumsum(n./R))+(cumsum(n)./R));
```

```
%Uex=(-q/(4*pi*epsil))*((n./((4*pi*a*R.*R)).^(1/3))); %Unew=Unew+Uex;
```

```
change=sum(abs(Unew-Uscf))/Np, Uscf=Unew;
```

```
end
```

```
[E1s E2s E2p E3s E3p]
```

```
%аналитические решения для состояний 1s водорода
```

```
a0=4*pi*epsil*hbar*hbar/(m*q*q);
```

```
P0=(4*a/(a0^3))*R.*R.*exp(-2*R./a0);
```

```
hold on
```

```
h=plot(R,P1s,'b');
```

```
h=plot(R,P0,'bx');
```

```
h=plot(R,P3p,'bo');
```

```
set(h,'linewidth',[2.0])
```

```
set(gca,'FontSize',[25])
xlabel('R (m) →');
ylabel('Probability →');
axis([0 5e-10 0 0.08]);
grid on
```

**%Рис.3.3.4**

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; epsil=8.854e-12; q=1.602e-19;
a0=4*pi*epsil*hbar*hbar/(m*q*q); E0=q/(8*pi*epsil*a0);
```

```
R0=.05*[1:200];
a=(-2*E0)*(1-(exp(-2*R0).*(1+R0)))./R0;
b=(-2*E0)*exp(-R0).*(1+R0);
s=(1+R0+((R0.^2)/3)).*exp(-R0);
Uee=(2*E0)./sqrt(1+(R0.^2)); UNN=(2*E0)./R0;
```

```
EB0=(a+b)./(1+s); R=a0*R0;
```

```
hold on
h=plot(R,EB0,'b-');
h=plot(R,Uee,'bx');
h=plot(R,UNN,'b');
h=plot(R,(2*EB0)+UNN+Uee,'b+');
set(h,'LineWidth',[2.0])
set(gca,'FontSize',[25])
grid on
xlabel('R (m) →');
ylabel('Energy (eV) →');
axis([0 4e-10 -25 25])
```

**%Рис.3.4.2**

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; I0=q*q/hbar;
```

```
%Константы
U0=0.5; % U0 is 0.25 for part(a), 0.1 for part(b)
kT=0.025; mu=0; ep=0.2;
```

```

g1=0.005; g2=0.005; g=g1+g2;
alphag=1; alphad=0.5;

%Напряжение
IV=101; VV=linspace(0,1,IV);
for iV=1:IV
    Vd=0; Vg=VV(iV);
    mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);
    f1=1/(1+exp((ep+UL-mu1)/kT)); f2=1/(1+exp((ep+UL-mu2)/kT));
    f1U=1/(1+exp((ep+UL+U0-mu1)/kT)); f2U=1/(1+exp((ep+UL+U0-mu2)/kT));

    P1=((g1*f1)+(g2*f2))/(1e-6+(g1*(1-f1))+(g2*(1-f2)));
    P2=P1*((g1*f1U)+(g2*f2U))/(1e-6+(g1*(1-f1U))+(g2*(1-f2U)));
    P0=1/(1+P1+P1+P2); P1=P1*P0; P2=P2*P0;
    p0(iV)=P0; p1(iV)=P1; p2(iV)=P2;
end

hold on
h=plot(VV,p0,'bo');
h=plot(VV,p1,'b');
h=plot(VV,p2,'bx');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
grid on
xlabel('Gate voltage, VG (volts) ->')
ylabel('Current (Amperes) ->')
axis([0 1 0 1])

```

**%Рис.3.4.3**

clear all

% Константы (все в системе MKC, кроме энергии, которая дана в эВ)  
 $\hbar=1.055\text{e-}34$ ;  $q=1.602\text{e-}19$ ;  $I_0=q/\hbar$ ;

%Константы  
 $U_0=0.1$ ; %  $U_0$  is 0.25 for part(a), 0.025 for part(b)  
 $kT=0.025$ ;  $\mu_0=0$ ;  $ep=0.2$ ;  
 $g1=0.005$ ;  $g2=0.005$ ;  $g=g1+g2$ ;  
 $alphag=1$ ;  $alphad=0.5$ ;

%Напряжение  
IV=101; VV=linspace(0,1.5,IV);  
for iV=1:IV

```

Vg=0; Vd=VV(iV);
%Vd=0; Vg=VV(iV);
    mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

%Многоэлектронный метод
f1=1/(1+exp((ep+UL(-U0/2)-mu1)/kT));
f2=1/(1+exp((ep+UL(-U0/2)-mu2)/kT));
f1U=1/(1+exp((ep+UL+(U0/2)-mu1)/kT));
f2U=1/(1+exp((ep+UL+(U0/2)-mu2)/kT));

P1=((g1*f1)+(g2*f2))/(1e-6+(g1*(1-f1))+(g2*(1-f2)));
P2=P1*((g1*f1U)+(g2*f2U))/(1e-6+(g1*(1-f1U))+(g2*(1-f2U)));
P0=1/(1+P1+P1+P2); P1=P1*P0; P2=P2*P0;

I1(iV)=2*I0*((P0*g1*f1)-(P1*g1*(1-f1))+(P1*g1*f1U)-(P2*g1*(1-f1U)));
I2(iV)=2*I0*((P0*g2*f2)-(P1*g2*(1-f2))+(P1*g2*f2U)-(P2*g2*(1-f2U)));
end

%RSCF метод (то же, что и на Рис.1.4.6 с дополнительным множителем, равным двум)
%Сетка по энергии
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
D=(g/(2*pi))./((E.^2)+((g/2).^2)); % Лоренцевская плотность состояний на 1 эВ
D=D./ (dE*sum(D)); % Нормировка на единицу

%Напряжение
for iV=1:IV
    Vg=0; Vd=VV(iV);
    %Vd=0; Vg=VV(iV);
        mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

U=0; %Самосогласованное поле
dU=1;
while dU>1e-6
    F1=1./(1+exp((E+ep+UL+U-mu1)./kT));
    F2=1./(1+exp((E+ep+UL+U-mu2)./kT));
    N(iV)=dE*2*sum(D.*((F1.*g1/g)+(F2.*g2/g)));
    Unew=U0*N(iV);
    dU=abs(U-Unew); U=U+0.1*(Unew-U);
end
I(iV)=dE*2*I0*(sum(D.*(F1-F2)))*(g1*g2/g);
end

```

```

hold on
h=plot(VV,I1,'b');
h=plot(VV,I,'b-');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
grid on
xlabel('Drain Voltage, VD (volts) ->')
ylabel('Current (Amperes) ->')
axis([0 1.5 0 1.4e-6])

%E.3.5c: неограниченное приближение scf
clear all

% Константы (все в системе MKS, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; I0=q*q/hbar;

%Константы
U0=0.25; kT=0.025; mu=0; ep=0.2;
g1=0.005; g2=0.005; g=g1+g2;
alphag=1; alphad=0.5;

%Cетка по энергии
NE=501; E=linspace(-1, 1, NE); dE=E(2)-E(1);
D=(g/(2*pi))./((E.^2)+((g/2)^2)); % Lorentzian Density of states per eV
D=D./ (dE*sum(D)); % Normalizing to one

%Напряжение
IV=101; VV=linspace(0,1,IV);
for iV=1:IV
    Vg=0; Vd=VV(iV);
    %Vd=0; Vg=VV(iV);
    mu1=mu; mu2=mu1-Vd; UL=-(alphag*Vg)-(alphad*Vd);

    Uup=0; Udn=0.1; %Неограниченный метод самосогласованного поля
    dU=1; while dU>.001
        f1up=1./(1+exp((E+ep+UL+Uup-mu1)./kT));
        f2up=1./(1+exp((E+ep+UL+Uup-mu2)./kT));
        Nup(iV)=dE*sum(D.*((f1up.*g1)+(f2up.*g2))./(g1+g2));
        f1dn=1./(1+exp((E+ep+UL+Udn-mu1)./kT));
        f2dn=1./(1+exp((E+ep+UL+Udn-mu2)./kT));
        Ndn(iV)=dE*sum(D.*((f1dn.*g1)+(f2dn.*g2))./(g1+g2));
        Udnnew=2*U0*(Nup(iV)-0.5); Udn=Udn+0.1*(Udnnew-Udn);
        Uupnew=2*U0*(Ndn(iV)-0.5); Uup=Uup+0.1*(Uupnew-Uup);
        dU=abs(Uup-Uupnew);
    end
end

```

```

dU=abs(Uup-Uupnew)+abs(Udn-Udnnew);
end
Iup(iV)=dE*I0*sum(D.*(f1up-f2up))*(g1*g2/g);
Idn(iV)=dE*I0*sum(D.*(f1dn-f2dn))*(g1*g2/g);
end

hold on
%h=plot(VV,Nup,'bo'); %Part (b)
%h=plot(VV,Nup,'bx'); %Part (b)
h=plot(VV,Iup+Idn,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('Voltage (V) ->')
ylabel('Current (A) ->')
%ylabel('Number of electrons ->'); %Part (b)
grid on

```

## Глава 4

### %Рис.4.1.4

```
clear all
```

```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; epsil=8.854e-12; q=1.602e-19;
a0=4*pi*epsil*hbar/(m*q*q); E0=q/(8*pi*epsil*a0);

```

```

%Базис
L=.074e-9/a0; s=exp(-L)*(1+L+((L^2)/3));
r=linspace(-2e-10,+2e-10,101); r0=r/a0;
psi=sqrt(1/(pi*(a0^3)))*(exp(-abs(r0-(L/2)))+exp(-abs(r0+(L/2))));
n=2*psi.*conj(psi)./(2*(1+s));

```

```

a=-2*E0*(1-((1+L)*exp(-2*L))/L;
b=-2*E0*(1+L)*exp(-L);
EB0=-E0+((a+b)/(1+s));
[a b s EB0]

```

```

hold on
h=plot(r,n,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
grid on

```

```

xlabel('R (m) ->')
ylabel('Electron density (/m^3) ->')
axis([-2e-10 2e-10 0 2e30])

```

### %Рис.4.3.1

```
clear all
```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)

```
hbar=1.055e-34; m=9.110e-31; q=1.602e-19; mu=0.25;
```

```
kT=0.025; % 0.025 для части (с), (е) и 0.0025 для частей (д), (ф)
```

```
%Решетка
```

```
Np=100; a=2e-10; X=a*[1:1:Np]; t0=(hbar^2)/(2*m*(a^2))/q; U=linspace(0,0,Np);
```

```
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
```

```
T(1,Np)=-t0; T(Np,1)=-t0; % Периодические граничные условия для частей (д),
```

```
(f)
```

```
U(Np/2)=U(Np/2)+10; % Потенциал примеси для частей (д), (ф)
```

```
[V,D]=eig(T+diag(U)); E=sort(diag(D)');
```

```
D=diag(D)-mu;
```

```
rho=1./(1+exp(D./kT)); rho=V*diag(rho)*V'; rho=diag(rho)/a;
```

```
hold on
```

```
grid on
```

```
%h=plot(E,'b'); h=plot(mu*ones(Np/2,1),'b'); %Часть (б)
```

```
h=plot(X,pho,'b'); %Части (с)-(ф)
```

```
set(h,'linewidth',[2.0])
```

```
set(gca,'Fontsize',[25])
```

```
grid on
```

```
% xlabel('Eigenvalues number ->'); %Часть (б)
```

```
% ylabel('Energy (eV) ->'); %Часть (б)
```

```
xlabel('x (m) ->'); %Части (с)-(ф)
```

```
ylabel('Electron density (/m^3) ->'); %Части (с)-(ф)
```

```
%axis([0 100 0 4]); %Часть (б)
```

```
axis([0 2e-8 0 1e9]); %Части (с)-(ф)
```

### %Рис.4.4.1, 4.4.2

```
clear all
```

% Параметры указаны в системе МКС (кроме энергии, которая дана в эВ)

```
hbar=1.055e-34; m=9.110e-31; epsil=8.854e-12; q=1.602e-19;
```

```

a0=4*pi*epsil*hbar*hbar/(m*q*q); E0=q/(8*pi*epsil*a0);

F=linspace(0,1e9,11); A=(a0*128*sqrt(2)/243)*F; B=(-3*a0)*F;
for kF=1:11
M=[-E0 0 A(kF);0 -E0/4 B(kF);A(kF) B(kF) -E0/4];
[V,D]=eig(M); D=diag(D); [DD,ind]=sort(D);
E1(kF)=D(ind(1)); E2(kF)=D(ind(2)); E3(kF)=D(ind(3));
end

%Результаты применения теории возмущений
E1s=-E0-((A.^2)/(3*E0/4));
E2s=-(E0/4)+B;
E2p=-(E0/4)-B;

hold on
%h=plot(F,E1,'b'); %Fig.3.4.1
%h=plot(F,E1s,'bx'); %Fig.3.4.1
h=plot(F,E2,'b'); %Fig.3.4.2
h=plot(F,E3,'b'); %Fig.3.4.2
h=plot(F,E2s,'bx'); %Fig.3.4.2
h=plot(F,E2p,'bo'); %Fig.3.4.2
set(h,'linewidth',[2.0])
set(gca,'FontSize',[25])
grid on
xlabel('Field (V/m) ->');
ylabel('Energy (eV) ->');
%axis([0 2e-8 0 1e9]);

```

## Глава 5

### %Рис.5.1.5

```

clear all

k=linspace(-1,1,21); a=2; b=1;
E1=sqrt((a^2)+(b^2)+(2*a*b*cos(pi*k)));

hold on
h=plot(k,E1,'b');
h=plot(k,-E1,'b');
set(h,'linewidth',[2.0])
set(gca,'FontSize',[25])
xlabel('k (in units of pi/a) ->')

```

```
ylabel('Energy (eV) ->')
grid on
```

### %Рис.5.3.2

```
clear all
```

```
Esa=-8.3431; Epa=1.0414; Esc=-2.6569; Epc=3.6686; Esea=8.5914;
Esec=6.7386; Vss=-6.4513; Vxx=1.9546; Vxy=5.0779; Vsapc=4.4800;
Vpasc=5.7839; Vseapc=4.8422; Vpasec=4.8077;
```

%Любой из следующих вариантов для d1,d2,d3 и d4 должен приводить к одному и тому же результату.

```
d1=[1 1 1]/4; d2=[1 -1 -1]/4; d3=[-1 1 -1]/4; d4=[-1 -1 1]/4;
d1=[0 0 0]/2; d2=[0 -1 -1]/2; d3=[-1 0 -1]/2; d4=[-1 -1 0]/2;
```

```
l=1; m=1; n=1; kmax=pi; Nt=21; %L-direction
%l=1; m=0; n=0; kmax=2*pi; Nt=21; %X-direction
```

```
for Nk=1:Nt
```

```
k=[l m n]*kmax*(Nk-1)/(Nt-1);
p1=exp(i*sum(k.*d1)); p2=exp(i*sum(k.*d2));
p3=exp(i*sum(k.*d3)); p4=exp(i*sum(k.*d4));
g0=(p1+p2+p3+p4)/4; g1=(p1+p2-p3-p4)/4;
g2=(p1-p2+p3-p4)/4; g3=(p1-p2-p3+p4)/4;
```

```
h=[Esa/2 Vss*g0 0 0 0 Vsapc*g1 Vsapc*g2 Vsapc*g3 0 0;
    Esc/2 -Vpasc*conj(g1) -Vpasc*conj(g2) -Vpasc*conj(g3) 0 0 0 0 0;
    0 0 Epa/2 0 0 Vxx*g0 Vxy*g3 Vxy*g2 0 -Vpasec*g1;
    0 0 0 Epa/2 0 Vxy*g3 Vxx*g0 Vxy*g1 0 -Vpasec*g2;
    0 0 0 0 Epa/2 Vxy*g2 Vxy*g1 Vxx*g0 0 -Vpasec*g3;
    0 0 0 0 Epc/2 0 0 Vseapc*(g1) 0;
    0 0 0 0 0 Epc/2 0 Vseapc*(g2) 0;
    0 0 0 0 0 0 Epc/2 Vseapc*(g3) 0;
    0 0 0 0 0 0 Esea/2 0 0;
    0 0 0 0 0 0 0 0 Esec/2];
```

```
H=h+h';
```

```
[V,D]=eig(H);
```

```
eigst=sum(D);
```

```
E(Nk,:)=sort(real(eigst));
```

```
X(Nk)=-(Nk-1)/(Nt-1); %L-direction
```

```
X1(Nk)=(Nk-1)/(Nt-1); %X-direction
```

```
end
```

```

hold on
h=plot(X,E,'b');
%h=plot(X1,E,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
 xlabel('k (as fraction of maximum value) ->')
 ylabel('Energy (eV) ->')
grid on

```

**%Примечание: Часть оси X от 0 до +1 отвечает направлению -X,  
%в то время как часть оси X от 0 до -1 отвечает направлению -L**

### Рис.5.4.1а

```

clear all

soa=.3787/3; soc=.0129/3; Esa=-8.3431; Epa=1.0414; Esc=-2.6569; Epc=3.6686;
Esea=8.5914; Esec=6.7386; Vss=-6.4513; Vxx=1.9546; Vxy=5.0779;
Vsapc=4.4800; Vpasc=5.7839; Vseapc=4.8422; Vpasec=4.8077;
d1=[1 1 1]/4; d2=[1 -1 -1]/4; d3=[-1 1 -1]/4; d4=[-1 -1 1]/4;
d1=[0 0 0]/2; d2=[0 -1 -1]/2; d3=[-1 0 -1]/2; d4=[-1 -1 0]/2;

l=1; m=1; n=1; kmax=pi; Nt=101; %L-direction
%l=1; m=0; n=0; kmax=2*pi; Nt=101; %X-direction

for Nk=1:Nt
k=[1 m n]*kmax*(Nk-1)/(Nt-1);
p1=exp(i*sum(k.*d1)); p2=exp(i*sum(k.*d2));
p3=exp(i*sum(k.*d3)); p4=exp(i*sum(k.*d4));
g0=(p1+p2+p3+p4)/4; g1=(p1+p2-p3-p4)/4;
g2=(p1-p2+p3-p4)/4; g3=(p1-p2-p3+p4)/4;

h=[Esa/2 Vss*g0 0 0 0 Vsapc*g1 Vsapc*g2 Vsapc*g3 0 0;
    0 Esc/2 -Vpasc*conj(g1) -Vpasc*conj(g2) -Vpasc*conj(g3) 0 0 0 0 0;
    0 0 Epa/2 0 0 Vxx*g0 Vxy*g3 Vxy*g2 0 -Vpasec*g1;
    0 0 0 Epa/2 0 Vxy*g3 Vxx*g0 Vxy*g1 0 -Vpasec*g2;
    0 0 0 0 Epa/2 Vxy*g2 Vxy*g1 Vxx*g0 0 -Vpasec*g3;
    0 0 0 0 Epc/2 0 0 Vseapc*(g1) 0;
    0 0 0 0 0 Epc/2 0 Vseapc*(g2) 0;
    0 0 0 0 0 0 Epc/2 Vseapc*(g3) 0;
    0 0 0 0 0 0 0 Esea/2 0;
    0 0 0 0 0 0 0 0 Esec/2];
H=[h+h'*zeros(10)];

```

```

zeros(10) h+h'];
hs0=zeros(20);
hs0(3,4)=-i*soa; hs0(3,15)=soa;
hs0(4,15)=-i*soa;
hs0(5,13)=-soa; hs0(5,14)=i*soa;
hs0(6,7)=-i*soc; hs0(6,18)=soc;
hs0(7,18)=-i*soc;
hs0(8,16)=-soc; hs0(8,17)=i*soc;
hs0(13,14)=i*soa;
hs0(16,17)=i*soc;
Hso=hs0+hs0';
[V,D]=eig(H+Hso);
eigst=sum(D);
E(Nk,:)=sort(real(eigst));
X(Nk)=-(Nk-1)/(Nt-1); %Направление L
X1(Nk)=(Nk-1)/(Nt-1); %Направление X
end

hold on
h=plot(X,E,'b');
%h=plot(X1,E,'b');
axis([-1 1 -3 3])
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[25])
xlabel('k (as fraction of maximum value) ->')
ylabel('Energy (eV) ->')
grid on

```

## Глава 6

### %Рис.6.1.2

```
clear all
```

```
z=zeros(5); Z=zeros(10);
```

```
% Константы (все в системе MKC, кроме энергии, которая дана в эВ)
```

```
hbar=1.055e-34; q=1.602e-19; a=2.45e-10*4/sqrt(3); m=9.110e-31;
```

```
d1=[1 1 1]/4; d2=[1 -1 -1]/4; d3=[-1 1 -1]/4; d4=[-1 -1 1]/4;
```

```
%Параметры для sp3s*-модели
```

```
soa=.3787/3; soc=.0129/3;
```

```
Esa=-8.3431; Epa=1.0414; Esc=-2.6569; Epc=3.6686; Esea=8.5914; Esec=6.7386;
Vss=-6.4513; Vpasc=-5.7839; Vpasec=-4.8077;
Vsapc=4.4800; Vseapc=4.8422; Vxx=1.9546; Vxy=5.0779;
```

%Параметры для зоны проводимости в приближении эффективной массы  
 $E_c = 1.55$ ;  $m_{eff} = .12 * m$ ;

```
Nt=101; kk=l*linspace(0,1,Nt);
l=0.5; m=0.5; n=0.5; %Направление L
%l=1; m=0; n=0; %Направление X

for Nk=1:Nt
    k=2*pi*kk(Nk)*[1 m n];

    %sp3s*-модель
    p1=exp(i*sum(k.*d1)); p2=exp(i*sum(k.*d2));
    p3=exp(i*sum(k.*d3)); p4=exp(i*sum(k.*d4));
    g0=(p1+p2+p3+p4)/4; g1=(p1+p2-p3-p4)/4;
    g2=(p1-p2+p3-p4)/4; g3=(p1-p2-p3+p4)/4;

    a1=diag([Esa Epa Epa Epa Esea]); A1=[a1 z;z a1];
    a2=diag([Esc Epc Epc Epc Esec]); A2=[a2 z;z a2];
    b=[Vss*g0 Vsapc*g1 Vsapc*g2 Vsapc*g3 0;
        Vpasc*g1 Vxx*g0 Vxy*g3 Vxy*g2 Vpasec*g1;
        Vpasc*g2 Vxy*g3 Vxx*g0 Vxy*g1 Vpasec*g2;
        Vpasc*g3 Vxy*g2 Vxy*g1 Vxx*g0 Vpasec*g3;
        0 Vseapc*conj(g1) Vseapc*conj(g2) Vseapc*conj(g3) 0];
    B=[b z;z b]; h=[a1 b;b' a2]; H=[A1 B;B' A2];

    aso=soa*[0 0 0 0 0 0 0 0 0 0;
               0 0 -i 0 0 0 0 0 1 0;
               0 i 0 0 0 0 0 0 -i 0;
               0 0 0 0 0 0 -1 i 0 0;
               0 0 0 0 0 0 0 0 0 0;
               0 0 0 0 0 0 0 0 0 0;
               0 0 0 -1 0 0 0 i 0 0;
               0 0 0 -i 0 0 -i 0 0 0;
               0 1 i 0 0 0 0 0 0 0;
               0 0 0 0 0 0 0 0 0 0];
    cso=soc*[0 0 0 0 0 0 0 0 0 0;
               0 0 -i 0 0 0 0 0 1 0;
               0 i 0 0 0 0 0 0 -i 0;
```

```

0 0 0 0 0 0 -1 i 0 0;
0 0 0 0 0 0 0 0 0 0;
0 0 0 0 0 0 0 0 0 0;
0 0 0 -1 0 0 0 i 0 0;
0 0 0 -i 0 0 -i 0 0 0;
0 1 i 0 0 0 0 0 0 0;
0 0 0 0 0 0 0 0 0 0]; H=H+[aso Z;Z cso];

[V,D]=eig(H);
eiglst=sum(D);
E(Nk,:)=sort(real(eiglst));

%Модель зоны проводимости в приближении эффективной массы
Em(Nk)=Ec+((hbar^2)*sum(k.*k)/(2*meff*q*(a^2)));
end

kk=-kk; %Направление L

hold on
h1=plot(kk,E,'b');
h2=plot(kk,Em,'b-');
axis([-1 1 -3 3])
set(h1,'linewidth',[1.0])
set(h2,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('ka (fraction of maximum value) ->')
ylabel('Energy (eV) ->')
grid on

%Рис.6.1.7

clear all

t=3; m=65; %66 for (a), 65 for (b)
D=2*m*0.14*sqrt(3)/(2*pi);
Eg=2*t*0.14/D; nu=round(2*m/3)+0; % +1 is used for higher mode
kyb=2*pi*nu/(2*m);
kxa=0.05*linspace(-pi,pi,101);

E1=(3*t/2)*sqrt(((kxa*2/3).^2)+(((abs(kyb)-(2*pi/3))*2/sqrt(3)).^2));
%a0=b*2/sqrt(3)=a*2/3;
E2=t*sqrt(1+(4*cos(kyb).*cos(kxa))+(4*cos(kyb).^2));

```

```
k=kxa./pi; [D Eg nu min(E1)]  
  
hold on  
h=plot(k,E1,'b');  
h=plot(k,-E1,'b');  
axis([-0.05 0.05 -0.6 0.6])  
set(h,'linewidth',[1.0])  
set(h,'linewidth',[2.0])  
set(gca,'Fontsize',[24])  
 xlabel('kxa/pi (fraction of maximum value) →')  
 ylabel('Energy (eV) →')  
 grid on
```

**%Рис.6.1.9**

```
clear all  
  
t=3; kxa=0;  
kyb=linspace(-pi,pi,101);  
E1=(3*t/2)*sqrt(((kxa*2/3).^2)+((abs(kyb)-(2*pi/3))*2/sqrt(3)).^2));  
%a0=b*2/sqrt(3)=a*2/3;  
E2=t*sqrt(1+(4*cos(kyb).*cos(kxa))+(4*cos(kyb).^2));  
  
k=kyb./pi;  
  
hold on  
h=plot(k,E1,'b');  
h=plot(k,-E1,'b');  
h=plot(k,E2,'bx');  
h=plot(k,-E2,'bx');  
axis([-1 1 -15 15])  
set(h,'linewidth',[1.0])  
set(h,'linewidth',[2.0])  
set(gca,'Fontsize',[24])  
 xlabel('kyb/pi →')  
 ylabel('Energy (eV) →')  
 grid on
```

**%Рис.6.2.1**

```
clear all
```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)  
 $\hbar=1.055\text{e-}34$ ;  $m=9.110\text{e-}31$ ;  $q=1.602\text{e-}19$ ;  $L=1\text{e-}9$ ;

```

D2=zeros(1,101);
Lz=20e-9; %5e-9 for (a), 20e-9 for (b)
E0=(hbar^2)*(pi^2)/(2*q*m*Lz^2);
for p=1:25
E=linspace(-0.1,0.25,101); theta=(E+abs(E))./(2*E);
EE=E-(p*p*E0); theta=(EE+abs(EE))./(2*EE);
D1=(L)^q*m*theta.*real((2*m*E*q).^( -0.5))./(pi*hbar);
D2=D2+((L^2)*q*m*theta./(2*pi*hbar*hbar));
D3=(L^3)*q*m*theta.*real((2*m*E*q).^0.5)./(2*pi*pi*hbar*hbar*hbar);
end

hold on
h=plot(D2,E,'b');
h=plot(D3.*Lz/L,E,'b');
%axis([0 10 -0.1 0.25]); %Часть (а)
axis([0 40 -0.1 0.25]); %Часть (б)
set(h,'linewidth',[1.0])
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('D(E) (per eV per nm^2) ->')
ylabel('Energy (eV) ->')
grid on

```

## %Рис.6.2.2

```
clear all
```

```
t=3; m=800; % Используйте значения 200 и 800 для двух вариантов графиков
a0=0.14; D=2*m*a0*sqrt(3)/(2*pi); Eg=2*t*0.14/D; c=pi*D; L=1;D
nu0=round(2*m/3); a=3*a0/2;
```

```
E=linspace(0,0.25,101);
DG=(2*c*L/(2*pi*a*a*t*t))*E;

DN=zeros(1,101);
for nu=nu0-100:nu0+100
Ek=((t*2*pi/sqrt(3))*(3*nu/(2*m))-1)+(i*1e-12);
    DN=DN+((2*L/(pi*a*t))*abs(real(E./sqrt((E.^2)-(Ek.^2))))));
end
```

```
hold on
h1=plot(DG,E,'bx');
h2=plot(DN,E,'b');
hold on
```

```
axis([0 50 0 0.25]);
set(h1,'linewidth',[1.0])
set(h2,'linewidth',[2.0])
set(gca,'Fontsize',[24])
 xlabel('D(E) (per eV per nm) ->')
 ylabel('Energy (eV) ->')
grid on
```

**%Рис.6.3.3**

```
clear all
```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)  
 $\hbar=1.055\text{e-}34$ ;  $m=9.110\text{e-}31$ ;  $q=1.602\text{e-}19$ ;  $a=5\text{e-}10$ ;  $L=10\text{e-}9$ ;

```
k=0.5*linspace(-1,1,201)/a;
Ek=(hbar^2)*(k.^2)/(2*0.25*m*q);
EE=linspace(0,0.2,201);
```

```
%Подзона (1,1)
E1=2*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=((EE-E1)+abs(EE-E1))./(2*abs(EE-E1));
```

```
%Подзоны (1,2) и (2,1)
E2=5*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+(((EE-E2)+abs(EE-E2))./(abs(EE-E2)));
```

```
%Подзона (2,2)
E3=8*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+(((EE-E3)+abs(EE-E3))./(2*abs(EE-E3)));
```

```
hold on
h=plot(k,E1+Ek,'b'); %Часть (а)
h=plot(k,E2+Ek,'b'); %Часть (а)
h=plot(k,E3+Ek,'b'); %Часть (а)
%h=plot(M,EE,'b'); %Часть (б)
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
 xlabel('k(/m)'); %Часть (а)
% xlabel('M(E)'); %Часть (б)
 ylabel('E-Ec (eV) ->')
axis([-1e9 1e9 0 0.3]); %Часть (а)
%axis([0 5 0 0.3]); %Часть (б)
grid on
```

**%Рис.6.3.4**

```

clear all

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; q=1.602e-19; a=5e-10; L=10e-9;

k=0.5*linspace(-1,1,201)/a;
Ek=-(hbar^2)*(k.^2)/(2*0.25*m*q);
EE=linspace(0,-0.2,201);

%Подзона (1,1)
E1=-2*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=((E1-EE)+abs(E1-EE))./(2*abs(E1-EE));

%Подзоны (1,2) и (2,1)
E2=-5*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+(((E2-EE)+abs(E2-EE))./(abs(E2-EE)));

%Подзона (2,2)
E3=-8*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+(((E3-EE)+abs(E3-EE))./(2*abs(E3-EE)));

hold on
%h=plot(k,E1+Ek,'b'); %Часть (а)
%h=plot(k,E2+Ek,'b'); %Часть (а)
%h=plot(k,E3+Ek,'b'); %Часть (а)
h=plot(M,EE,'b'); %Часть (б)
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
% xlabel('k/(m)');
xlabel('M(E)');
% ylabel('E-Ev (eV) ->');
% axis([-1e9 1e9 -0.3 0]);
axis([0 5 -0.3 0]);
grid on

```

**Глава 7****%Рис.7.1.5**

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; a=3e-10; m=9.110e-31;
```

```
%Параметры зоны проводимости
```

```
mw=.07*m; ma=.22*m; mb=(.7*mw)+(.3*ma); kk=0*.l*pi;
Ec=0; Eb=(.7*0)+(.3*1.25);
```

```
for nk=1:24
```

```
Nw=nk+10; Nb=2*Nw; Np=Nb+Nw+Nb; W(nk)=(Nw-1)*a*1e9;
tb=(hbar^2)/(2*mb*(a^2)*q); tw=(hbar^2)/(2*mw*(a^2)*q);
t=[tb*ones(1,Nb) tw*ones(1,Nw-1) tb*ones(1,Nb)];
tt=[0 t]+[t 0];
Ebk=Eb+(tb*(kk^2)); Ewk=tw*(kk^2); Ebwk=(Eb/2)+((tb+tw)*(kk^2)/2);
U=Ec+[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
H=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);
[V,D]=eig(H); D=diag(D); D=(sort(real(D)));
E1(nk)=D(1); E2(nk)=D(2);
end
```

```
hold on
```

```
h1=plot(W,E1,'b');
h2=plot(W,E2,'b-');
set(h1,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('W (nm) -> ')
ylabel('Energy (eV) -> ')
axis([2 10 0 .4])
grid on
```

## %Рис.7.1.6

```
clear all
```

```
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; a=3e-10; m=9.110e-31;
```

```
%Параметры зоны проводимости
```

```
mw=.07*m; ma=.22*m; mb=(.7*mw)+(.3*ma);
Nw=24; Nb=2*Nw; Np=Nb+Nw+Nb; W=(Nw-1)*a*1e9;
Ec=0; Eb=(.7*0)+(.3*1.25);
```

```
for nk=1:26
```

```
kk=(nk-1)*a*1e10/500; k(nk)=kk/(a*1e9);
tb=(hbar^2)/(2*mb*(a^2)*q); tw=(hbar^2)/(2*mw*(a^2)*q);
```

```
t=[tb*ones(1,Nb) tw*ones(1,Nw-1) tb*ones(1,Nb)];
tt=[0 t]+[t 0];
Ebk=Eb+(tb*(kk.^2)); Ewk=tw*(kk.^2); Ebwk=(Eb/2)+((tb+tw)*(kk.^2)/2);
U=Ec+[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
H=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);
[V,D]=eig(H); D=diag(D); D=(sort(real(D)))';
E1(nk)=D(1); E2(nk)=D(2);
end

E1w=E1(1)+(hbar.^2)*(k.^2)./(2*mw*1e-18*q);
E2w=E2(1)+(hbar.^2)*(k.^2)./(2*mw*1e-18*q);
E1b=E1(1)+(hbar.^2)*(k.^2)./(2*mb*1e-18*q);
E2b=E2(1)+(hbar.^2)*(k.^2)./(2*mb*1e-18*q);

hold on
h=plot(k,E1,'b');
h=plot(k,E2,'b');
h=plot(k,E1w,'b:');
h=plot(k,E2w,'b:');
h=plot(k,E1b,'b-');
h=plot(k,E2b,'b-');
set(h,'linewidth',[1.0])
set(gca,'FontSize',[24])
xlabel('k (/nm) → ')
ylabel('Energy (eV) → ')
axis([0 .5 0 0.4])
grid on
```

### %Рис.7.2.5

```
clear all
```

```
% Константы (все в системе MKC, кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; eps0=8.85E-12; epsr=4; m=.25*9.1e-31;
mu=0; kT=.025; n0=m*kT*q/(2*pi*(hbar.^2));n0
```

%Входные значения параметров

```
a=3e-10; t0=(hbar.^2)/(2*m*(a.^2)*q); e0=q*a/eps0;
```

Nox=7; Nc=10; %используйте Nc=10,30 для параметра канала 3,9 нм соответственно

```
Np=Nox+Nc+Nox; XX=a.*1e9*[1:1:Np];
```

```
Ec=[3*ones(Nox,1); 0*ones(Nc,1); 3*ones(Nox,1)];
```

%Матрица гамильтониана

```

T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));

%Матрица диэлектрической проницаемости
D2=epsr*((2*diag(ones(1,Np)))-(diag(ones(1,Np-1),1))-(diag(ones(1,Np-1),-1)));
iD2=inv(D2);

Vg=-.25; Ubdy=-epsr*[Vg; zeros(Np-2,1); Vg];
%Ubdy=[0; zeros(Np-2,1); Vg]; %for asymmetric bias
U0=iD2*Ubdy;

%Самосогласованный расчет
U1=1e-6*ones(Np,1); UU=U1; change=1;
while change>1e-3
    U1=U1+(0.1*(UU-U1));
    [P,D]=eig(T+diag(Ec)+diag(U1)); D=diag(D);
    rho=ilog(1+exp((mu-D)./kT)); rho=P*diag(rho)*P';
    n=2*n0*diag(rho);
    for kp=1:Np
        ncl(kp)=a*2*(n0^1.5)*Fhalf((mu-Ec(kp)-U1(kp))/kT);
    end
    %n=ncl'; %use for semiclassical calculation
    UU=U0+(iD2*e0*n);
    change=max(max(abs(UU-U1)));
    U=Ec+U1; %self-consistent band profile
end

%Электронная плотность в канале на квадратный сантиметр
ns=1e-4*sum(sum(n.*[zeros(Nox,1);ones(Nc,1);zeros(Nox,1)]));Vg,ns
nn=1e-6*n./a; %electron density per cm3
Fn=mu*ones(Nc+Nox+Nox,1);

hold on
h=plot(XX,nn,'g');
%h=plot(XX,Ec,'g-');
%h=plot(XX,Ec+U1,'g');
%h=plot(XX,En,'g:');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('z (nm) → ')
%ylabel('E (eV) → ')
ylabel('n (/cm3) → ')
%axis([0 8 -.5 3])
axis([0 8 0 15e18])

```

```
grid on
```

### %Рис.7.3.1

```
clear all
```

% Константы (все в системе МКС, кроме энергии, которая дана в эВ)  
 $\hbar=1.06e-34$ ;  $q=1.6e-19$ ;  $\epsilon_0=8.85E-12$ ;  $\epsilon_{sr}=4$ ;  $m=.25*9.1e-31$ ;  
 $\mu_0=0$ ;  $kT=.025$ ;  $n_0=m*kT*q/(2*pi^*(\hbar^2))$ ;

%Входящие значения параметров

```
a=3e-10; t0=(\hbar^2)/(2*m*(a^2)*q); e0=q*a/\epsilon_0;
```

Nox=7; Nc=10; %используйте Nc=10,30 для канала с параметром 3,9 нм соответственно

```
Np=Nox+Nc+Nox; XX=a*1e9*[1:1:Np];
```

```
Ec=[3*ones(Nox,1); 0*ones(Nc,1); 3*ones(Nox,1)];
```

%Матрица гамильтониана

```
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
```

%Матрица диэлектрической проницаемости

```
D2=epsr*((2*diag(ones(1,Np)))-(diag(ones(1,Np-1),1))-(diag(ones(1,Np-1),-1)));  
iD2=inv(D2);
```

```
Vg=linspace(-.25,.25,26);
```

```
for kg=1:26
```

```
Ubdy=-epsr*[Vg(kg); zeros(Np-2,1); Vg(kg)];kg;
```

```
%Ubdy=-epsr*[0; zeros(Np-2,1); Vg(kg)]; %Для асимметричного напряжения
```

```
U0=iD2*Ubdy;
```

%Самосогласованный расчет

```
U1=1e-6*ones(Np,1); UU=U1; change=1;
```

```
while change>1e-3
```

```
    U1=U1+(0.1*(UU-U1));
```

```
    [P,D]=eig(T+diag(Ec)+diag(U1)); D=diag(D);
```

```
    rho=log(1+exp((mu-D)./kT)); rho=P*diag(rho)*P';
```

```
    n=2*n0*diag(rho);
```

```
    for kp=1:Np
```

```
        ncl(kp)=a*2*(n0^1.5)*Fhalf((mu-Ec(kp)-U1(kp))/kT);
```

```
    end
```

```
    %n=ncl'; %use for semiclassical calculation
```

```
    UU=UU+(iD2*e0*n);
```

```
    change=max(max((abs(UU-U1))));
```

```
    U=Ec+U1; %self-consistent band profile
```

```

end

%Электронная плотность в канале на квадратный сантиметр
ns(kg)=1e-4*sum(sum(n.*[zeros(Nox,1); ones(Nc,1); zeros(Nox,1))]);
nn(:,kg)=1e-6*n./a; %электронная плотность на кубический сантиметр
Fn(:,kg)=mu*ones(Nc+Nox+Nox,1);
end
C=q*(ns(26)-ns(25))/(Vg(26)-Vg(25));
d=1e-4*epsr*eps0*2/C;d,C
%ns=log10(ns)

hold on
h=plot(Vg,ns,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('Vg (V) → ')
ylabel(' ns (/cm2) → ')
%axis([0 .3 0 3.5e12])
grid on

```

### %Рис.7.3.2

```

clear all

E=linspace(-.5,1,1001);
D=sqrt(E);

hold on
h=plot(D,E,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('D (E) (arb. units) ')
ylabel('E (eV)')
grid on

```

### %Рис.7.3.4

```

clear all

E=linspace(-.25,.25,501); dE=E(2)-E(1); kT=0.025; Ef=0;
V=0; mu1=Ef+(V/2); mu2=Ef-(V/2);
f1=1./(1+exp((E-mu1)./kT)); f2=1./(1+exp((E-mu2)./kT));
FT=(0 diff(f1)); FT=FT.*(-1/dE);

```

```
%dE*(sum(f1-f2))/V
```

```
hold on
h=plot(FT,E,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
grid on
```

### **%Рис.7.4.1, Рис.7.4.2**

```
clear all
```

```
z=zeros(5); Z=zeros(10);
% Константы (все в системе МКС, кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; a=2.45e-10*4/sqrt(3); m=9.110e-31;
d1=[1 1 1]/4; d2=[1 -1 -1]/4; d3=[-1 1 -1]/4; d4=[-1 -1 1]/4;
```

```
%параметры для sp3s*-модели
soa=.3787/3; soc=.0129/3;
Esa=-8.3431; Epa=1.0414; Esc=-2.6569;
Epc=3.6686; Esea=8.5914; Esec=6.7386;
Vss=-6.4513; Vpasc=-5.7839; Vpasec=-4.8077;
Vsapc=4.4800; Vseapc=4.8422; Vxx=1.9546; Vxy=5.0779;
```

```
% Параметры Латинжера-Кона для валентной зоны
Ev=-.1; del=.3; g1=6.85; g2=2.1; g3=2.9;
t1=(hbar^2)*g1/(2*m*q*(a^2));
t2=(hbar^2)*g2/(2*m*q*(a^2));
t3=(hbar^2)*g3/(2*m*q*(a^2));
```

```
Nt=101; kk=linspace(0,1,Nt);
l=1; m=0; n=0; %направление X
l=0.5; m=0.5; n=0.5; %направление L
```

```
for Nk=1:Nt
k=2*pi*kk(Nk)*[l m n];
```

```
%sp3s*-модель
p1=exp(i*sum(k.*d1)); p2=exp(i*sum(k.*d2));
p3=exp(i*sum(k.*d3)); p4=exp(i*sum(k.*d4));
g0=(p1+p2+p3+p4)/4; g1=(p1+p2-p3-p4)/4;
g2=(p1-p2+p3-p4)/4; g3=(p1-p2-p3+p4)/4;
```

```
a1=diag([Esa Epa Epa Epa Esea]); A1=[a1 z;z a1];
```

```

a2=diag([Esc Epc Epc Epc Esec]); A2=[a2 z;z a2];
b=[Vss*g0 Vsapc*g1 Vsapc*g2 Vsapc*g3 0;
    Vpasc*g1 Vxx*g0 Vxy*g3 Vxy*g2 Vpasec*g1;
    Vpasc*g2 Vxy*g3 Vxx*g0 Vxy*g1 Vpasec*g2;
    Vpasc*g3 Vxy*g2 Vxy*g1 Vxx*g0 Vpasec*g3;
    0 Vseapc*conj(g1) Vseapc*conj(g2) Vseapc*conj(g3) 0];
B=[b z;z b]; h=[a1 b;b' a2]; H=[A1 B;B' A2];

aso=soa*[0 0 0 0 0 0 0 0 0 0 0 0;
            0 0 -i 0 0 0 0 0 1 0;
            0 i 0 0 0 0 0 -i 0;
            0 0 0 0 0 -1 i 0 0;
            0 0 0 0 0 0 0 0 0 0;
            0 0 0 0 0 0 0 0 0 0;
            0 0 0 -1 0 0 0 i 0 0;
            0 0 0 -i 0 0 -i 0 0 0;
            0 1 i 0 0 0 0 0 0 0;
            0 0 0 0 0 0 0 0 0 0];

cso=soc*[0 0 0 0 0 0 0 0 0 0 0 0;
            0 0 -i 0 0 0 0 0 1 0;
            0 i 0 0 0 0 0 -i 0;
            0 0 0 0 0 0 -1 i 0 0;
            0 0 0 0 0 0 0 0 0 0;
            0 0 0 0 0 0 0 0 0 0;
            0 0 0 -1 0 0 0 i 0 0;
            0 0 0 -i 0 0 -i 0 0 0;
            0 1 i 0 0 0 0 0 0 0;
            0 0 0 0 0 0 0 0 0 0]; H=H+[aso Z;Z cso];

[V,D]=eig(H);
eiglst=sum(D);
E(Nk,:)=sort(real(eiglst));

%Модель Латинжера-Кона для валентной зоны
P=Ev+(t1*sum(k.*k)); Q=t2*((k(1)^2)+(k(2)^2)-(2*(k(3)^2)));
R=-(sqrt(3)*t2*((k(1)^2)-(k(2)^2)))+(i*2*t3*sqrt(3)*k(1)*k(2));
S=2*t3*sqrt(3)*((k(1)-(i*k(2)))*k(3));

H4=-[P+Q -S R 0;
       -S' P-Q 0 R;
       R' 0 P-Q S;
       0 R' S' P+Q]; [V,D]=eig(H4);

```

```

eiglst=sum(D);
ELK4(Nk,:)=sort(real(eiglst));

H6=[P+Q -S R 0 -S/sqrt(2) sqrt(2)*R;
     -S' P-Q 0 R -sqrt(2)*Q sqrt(1.5)*S';
     R' 0 P-Q S sqrt(1.5)*S' sqrt(2)*Q';
     0 R' S' P+Q -sqrt(2)*R' -S'/sqrt(2);
     -S'/sqrt(2) -sqrt(2)*Q' sqrt(1.5)*S -sqrt(2)*R P+del 0;
     sqrt(2)*R' sqrt(1.5)*S' sqrt(2)*Q' -S/sqrt(2) 0 P+del];
[V,D]=eig(H6);
eiglst=sum(D);
ELK6(Nk,:)=sort(real(eiglst));
end

kk=-kk;%Направление L

hold on
h1=plot(kk,E,'b');
%h2=plot(kk,ELK4,'b-');%Рис.6.4.1
h2=plot(kk,ELK6,'b-');%Рис.6.4.2
set(h1,'linewidth',[2.0])
set(h2,'linewidth',[3.0])
set(gca,'FontSize',[24])
xlabel('ka (fraction of maximum value) → ')
ylabel('Energy (eV) → ')
axis([-1 1 -2 3])
grid on

```

**%Рис.7.4.4**

clear all

%Все параметры в системе MKC (кроме энергии, которая дана в эВ)  
 $\hbar = 1.055 \times 10^{-34}$ ;  $q = 1.602 \times 10^{-19}$ ;  $a = 3 \times 10^{-10}$ ;  $m = 9.110 \times 10^{-31}$ ;

```

Eb=.15;
%Параметры Латинжера-Кона
g1=6.85; g2=2.1; g3=2.9;%Для GaAs
w1=(hbar^2)*g1/(2*m*q*(a^2));
w2=(hbar^2)*g2/(2*m*q*(a^2));
w3=(hbar^2)*g3/(2*m*q*(a^2));
g1=3.45; g2=0.68; g3=1.29;%Для AlAs
a1=(hbar^2)*g1/(2*m*q*(a^2)); b1=(.7*w1)+(3*a1);
a2=(hbar^2)*g2/(2*m*q*(a^2)); b2=(.7*w2)+(3*a2);

```

```

a3=(hbar^2)*g3/(2*m*q*(a^2)); b3=(.7*w3)+(3*a3);
Ev=0; Evb=(0.7*0)+(0.3*0.75); kx=0*pi; ky=0*pi; k2=(kx^2)+(ky^2);

for nk=1:20
Nw=nk+10; Nb=Nw; Np=Nb+Nw+Nb; W(nk)=(Nw-1)*a*1e9; Z=zeros(Np);nk
X(nk)=Nw-1;
t=[b1*ones(1,Nb) w1*ones(1,Nw-1) b1*ones(1,Nb)]; tt=[0 t]+[t 0];
Ebk=Evb+(b1*k2); Ewk=(w1*k2); Ebwk=(Ebk+Ewk)/2;
U=Ev+[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
P=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);

t=-2*[b2*ones(1,Nb) w2*ones(1,Nw-1) b2*ones(1,Nb)]; tt=[0 t]+[t 0];
Ebk=b2*k2; Ewk=w2*k2; Ebwk=(Ebk+Ewk)/2;
U=[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
Q=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);

Ebk=-(sqrt(3)*b2*((kx^2)-(ky^2)))+(i*2*b3*sqrt(3)*kx*ky);
Ewk=-(sqrt(3)*w2*((kx^2)-(ky^2)))+(i*2*w3*sqrt(3)*kx*ky);
Ebwk=(Ebk+Ewk)/2;
U=[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
R=diag(U);

t=2*i*sqrt(3)*(kx-(i*ky))*[b3*ones(1,Nb) w3*ones(1,Nw-1) b3*ones(1,Nb)];
S=diag(t,1)-diag(t,-1);

H=[P+Q Z;Z P+Q]; HL=[P-Q Z;Z P-Q];
HC=[-S R;R' S'];
H=-[H HC;HC' HL];

[V,D]=eig(H); D=diag(D); D=-(sort(real(-D)))';
E1(nk)=D(1); E2(nk)=D(2); E3(nk)=D(3); E4(nk)=D(4);
E5(nk)=D(5); E6(nk)=D(6); E7(nk)=D(7); E8(nk)=D(8);
end

%Аналитические результаты для бесконечно глубокой ямы
Ean1=-(w1-(2*w2))*(pi^2)./(X.^2);
Ean2=-(w1+(2*w2))*(pi^2)./(X.^2);

hold on
%h=plot(W,Ean1,'b');
%h=plot(W,Ean2,'b');
h=plot(W,E1,'b');
%h=plot(W,E2,'bx');

```

```

h=plot(W,E3,'b');
%h=plot(W,E4,'b+');
h=plot(W,E5,'b');
%h=plot(W,E6,'bx');
h=plot(W,E7,'b');
%h=plot(W,E8,'b+');
set(h,'linewidth',[2.0])
set(gca,'FontSize',[24])
 xlabel('W (nm) → ')
 ylabel('Energy (eV) → ')
axis([3 9 -.25 0])
grid on

```

**%Рис.7.4.5**

```
clear all
```

```
% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.055e-34; q=1.602e-19; a=3e-10; m=9.110e-31;
```

```
%Параметры Латинжера-Кона
```

```

g1=6.85; g2=2.1; g3=2.9;%Для GaAs
w1=(hbar^2)*g1/(2*m*q*(a^2));
w2=(hbar^2)*g2/(2*m*q*(a^2));
w3=(hbar^2)*g3/(2*m*q*(a^2));
g1=3.45; g2=0.68; g3=1.29;%Для AlAs
a1=(hbar^2)*g1/(2*m*q*(a^2)); b1=(.7*w1)+(3*a1);
a2=(hbar^2)*g2/(2*m*q*(a^2)); b2=(.7*w2)+(3*a2);
a3=(hbar^2)*g3/(2*m*q*(a^2)); b3=(.7*w3)+(3*a3);
Ev=0; Evb=(0.7*0)+(0.3*0.75);
```

```
Nw=18; Nb=Nw; Np=Nb+Nw+Nb; W=(Nw-1)*a*1e9; Z=zeros(Np);
```

```
for nk=1:26
```

```
k(nk)=(nk-1)/500;% in A^-1
```

```
l=0; m=1; lm=sqrt((l^2)+(m^2));
```

```
kx=(1/lm)*k(nk)*a*1e10; ky=(m/lm)*k(nk)*a*1e10;
```

```
k2=(kx^2)+(ky^2);
```

```
t=[b1*ones(1,Nb) w1*ones(1,Nw-1) b1*ones(1,Nb)]; tt=[0 t]+[t 0];
```

```
Ebk=Evb+(b1*k2); Ewk=(w1*k2); Ebwk=(Ebk+Ewk)/2;
```

```
U=Ev+[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
```

```
P=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);
```

```
t=-2*[b2*ones(1,Nb) w2*ones(1,Nw-1) b2*ones(1,Nb)]; tt=[0 t]+[t 0];
```

```

Ebk=b2*k2; Ewk=w2*k2; Ebwk=(Ebk+Ewk)/2;
U=[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
Q=-diag(t,1)-diag(t,-1)+diag(tt)+diag(U);

Ebk=-(sqrt(3)*b2*((kx^2)-(ky^2)))+(i*2*b3*sqrt(3)*kx*ky);
Ewk=-(sqrt(3)*w2*((kx^2)-(ky^2)))+(i*2*w3*sqrt(3)*kx*ky);
Ebwk=(Ebk+Ewk)/2;
U=[Ebk*ones(1,Nb) Ebwk Ewk*ones(1,Nw-2) Ebwk Ebk*ones(1,Nb)];
R=diag(U);

t=-2*i*sqrt(3)*(kx-(i*ky))*[b3*ones(1,Nb) w3*ones(1,Nw-1) b3*ones(1,Nb)]/2;
S=diag(t,1)-diag(t,-1);

H=[P+Q Z;Z P+Q]; HL=[P-Q Z;Z P-Q];
HC=[-S R;R' S'];
H=-[H HC;HC' HL];
[nk sum(sum(abs(H-H')))]

[V,D]=eig(H); D=diag(D); D=-(sort(real(-D)))';
E1(nk)=D(1); E2(nk)=D(2); E3(nk)=D(3); E4(nk)=D(4);
end

k=k*10;% перевод из масштаба на 1 ангстрем в масштаб на 1 нанометр
hold on
%h=plot(W,Ean1,'b');
%h=plot(W,Ean2,'b');
h=plot(k,E1,'b');
%h=plot(k,E2,'bx');
h=plot(k,E3,'b');
%h=plot(k,E4,'b+');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('k (/nm) -> ')
ylabel('Energy (eV) -> ')
axis([0 .5 -.1 0])
grid on

```

## Глава 8

### %Рис.8.1

```
clear all
```

```
E=linspace(-.3,.3,50001); dE=E(2)-E(1); gam=0.05;
```

```

D=(gam/(2*pi))./(((E-.14).^2)+((gam/2)^2));
D=D+(gam/(2*pi))./(((E-.04).^2)+((gam/2)^2));%Используйте для рис. 1.5.2
D=D+((gam/(2*pi))./(((E+.06).^2)+((gam/2)^2)));// Используйте для рис. 1.5.2
D=D+(gam/(2*pi))./(((E+.15).^2)+((gam/2)^2));
dE*sum(D)

hold on
h=plot(D,E,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('D(E) (per eV) ->')
ylabel('E (eV) -> ')
grid on

```

### %Рис.8.2.5

```

clear all

% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=0.25*9.1e-31; mu=0.25; kT=0.025;

%Входящие значения параметров
a=2e-10; t0=(hbar^2)/(2*m*(a^2)*q); Np=50; t0
X=a*linspace(0,Np-1,Np); U=linspace(-0.05,0.05,Np);
H=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
-(t0*diag(ones(1,Np-1),-1));
H=H+diag(U); HP=H;
HP(1,Np)=-t0; HP(Np,1)=-t0;

[V,D]=eig(HP); D=diag(D);
rho=1./(1+exp((D-mu)./kT));
rho=V*diag(rho)*V'; rho=diag(rho)/a;

%Cетка по энергии для метода функций Грина
Emin=-0.1; Emax=0.4; NE=250; E=linspace(Emin,Emax,NE);
dE=E(2)-E(1); zplus=i*1e-12;
f=1./(1+exp((E-mu)./kT));

%Метод функций Грина
sig1=zeros(Np,Np); sig2=zeros(Np,Np); n=zeros(Np,1);
for k=l:NE
    ck=(1-((E(k)+zplus-U(1))/(2*t0))); ka=acos(ck);
    sigma=-t0*exp(i*ka); sig1(1,1)=sigma;
    ck=(1-((E(k)+zplus-U(Np))/(2*t0))); ka=acos(ck);

```

```

sigma=-t0*exp(i*ka); sig2(Np,Np)=sigma;
G=inv(((E(k)+zplus)*eye(Np))-H-sig1-sig2);
n=n+(f(k)*(dE*diag(i*(G-G'))/(2*pi*a)));
end

hold on
h=plot(X,rho,'b');
h=plot(X,n,'bx');
grid on
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('X (m) → ')
ylabel('n (/m) → ')

```

### %Рис.8.2.6

```
clear all
```

```
% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=0.25*9.1e-31; mu=0.25; kT=0.025;
```

```
%Входящие значения параметров
```

```
a=2e-10; t0=(hbar^2)/(2*m*(a^2)*q); Np=50; t0
X=a*linspace(0,Np-1,Np); U=linspace(-0.05,0.05,Np);
H=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
-(t0*diag(ones(1,Np-1),-1));
H=H+diag(U);
```

```
%Сетка по энергии для метода функций Грина
```

```
Emin=-0.1; Emax=0.4;NE=250; E=linspace(Emin,Emax,NE);
dE=E(2)-E(1); zplus=i*1e-12;
f=1./(1+exp((E-mu)./kT));
```

```
%Метод функций Грина
```

```
sig1=zeros(Np,Np); sig2=zeros(Np,Np);
for k=1:NE
    ck=(1-((E(k)+zplus-U(1))/(2*t0))); ka=acos(ck);
    sigma=-t0*exp(i*ka); sig1(1,1)=sigma;
    ck=(1-((E(k)+zplus-U(Np))/(2*t0))); ka=acos(ck);
    sigma=-t0*exp(i*ka); sig2(Np,Np)=sigma;
    G=inv(((E(k)+zplus)*eye(Np))-H-sig1-sig2);
    D0=diag(i*(G-G'))/(2*pi); D1(k)=D0(1); D2(k)=D0(Np);
end
```

```

hold on
%h=plot(X,U,'b');
h=plot(D1,E,'b');
%h=plot(D2,E,'b');
grid on
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('X (m) → ')
ylabel('U (eV) → ')
%axis([0 1e-8 -.1 .4])
axis([0 1.2 -.1 .4])

```

**%Рис.8.4.1**

```

ep=-0.25; ep1=0.25; t=0.5; eta=0.025;
H=[ep t;t ep1];

```

```

E=linspace(-1,1,201);
for kE=1:201
G=inv(((E(kE)+(i*eta))*eye(2,2))-H);
A=diag(i*(G-G')); D(kE)=A(1)/2*pi;
end

```

```

hold on
h=plot(D,E,'gx');
set(h,'linewidth',[3.0])
set(gca,'Fontsize',[24])
grid on
xlabcl('LDOS (/eV ) → ')
ylabel('Energy (eV) → ')

```

**%Рис.Е.8.2**

```

clear all
t0=1; zplus=1e-10;

NE=81; X=linspace(-1,3,NE);
for kE=1:NE
    E=2*X(kE);
    ck=1-((E+zplus)/(2*t0)); ka=acos(ck);
    if imag(ka) < 0
        ka=ka'; end
    sig(kE)=-i0*exp(i*ka);

```

```

end

hold on
h1=plot(real(sig),X,'g');
h2=plot(imag(sig),X,'g');
h1=plot(real(sig),X,'go');
set(h1,'linewidth',[2.0])
set(h2,'linewidth',[4.0])
set(gca,'Fontsize',[24])
grid on
xlabel(' → ')
ylabel(' → ')

```

### %Рис.Е.8.5

```
clear all
```

```
% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=0.25*9.1e-31; zplus=i*5e-3;
```

```
%Входящие значения параметров
```

```
a=2.5e-10; t0=(hbar^2)/(2*m*(a^2)*q); Np=100; t0
X=a*linspace(0,Np-1,Np); U=zeros(1,Np);U(Np/2)=5/(a*1e10);
H=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
-(t0*diag(ones(1,Np-1),-1));
H=H+diag(U);
E=0.1;
```

```
%Метод функций Грина
```

```
sig1=zeros(Np,Np); sig2=zeros(Np,N);
ck=(1-((E+zplus-U(1))/(2*t0))); ka=acos(ck);
sigma=-t0*exp(i*ka); sig1(1,1)=sigma;
ck=(l-((E+zplus-U(Np))/(2*t0))); ka=acos(ck);
sigma=-t0*exp(i*ka); sig2(Np,Np)=sigma;
G=inv(((E+zplus)*eye(Np))-H-sig1-sig2);
D0=diag(i*(G-G'))/(2*pi);
```

```
hold on
h=plot(X,D0,'b');
grid on
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('X (m) → ')
ylabel('DOS (/eV) → ')
```

## Глава 9

### %Рис.9.4.2

```

clear all

% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.055e-34; m=9.110e-31; q=1.602e-19; a=5e-10; L=10e-9;

k=0.5*linspace(-l,l,201)/a;
Ek=(hbar^2)*(k.^2)/(2*0.25*m*q);
EE=linspace(0,0.2,201);

%Подзона (1,1)
E1=2*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=((EE-E1)+abs(EE-E1))./(2*abs(EE-E1));

%Подзона (1,2) и (2,1)
E2=5*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+((EE-E2)+abs(EE-E2))./(abs(EE-E2)));

%Подзона (2,2)
E3=8*(hbar^2)*(pi^2)/(2*0.25*m*q*L^2);
M=M+((EE-E3)+abs(EE-E3))./(2*abs(EE-E3)));
k=k*1e-9;

hold on
h=plot(k,E1+Ek,'b');
h=plot(k,E2+Ek,'b');
h=plot(k,E3+Ek,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('k (/nm)');
ylabel('E-Ec (eV) ->');
axis([-1 1 0 0.3]);
grid on

```

### %Рис.9.5.5

```
clear all
```

```
% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31; IE=(q*q)/(2*pi*hbar);
```

```

Ef=0.1; kT=.025;

%Входящие значения параметров
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q);
NS=15; NC=30; ND=15; Np=NS+NC+ND;

%Матрица гамильтониана
%NS=15; NC=20; ND=15; Np=NS+NC+ND; UB=0*ones(Np,1); %no barrier
%NS=23; NC=4; ND=23; Np=NS+NC+ND;
    %UB=[zeros(NS,1); 0.4*ones(NC,1); zeros(ND,1)]; %tunneling barrier
NS=15; NC=16; ND=15; Np=NS+NC+ND;
    UB=[zeros(NS,1); .4*ones(4,1); zeros(NC-8,1); .4*ones(4,1); zeros(ND,1)];
    %RT-барьер
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
T=T+diag(UB);

%Напряжение
V=0; mu1=Ef+(V/2); mu2=Ef-(V/2);
U1=V*[.5*ones(1,NS) linspace(0.5,-0.5,NC) -.5*ones(1,ND)];
U1=U1'; %Профиль приложенного потенциала

%Сетка по энергии для метода функций Грина
NE=501; E=linspace(-.2,.8,NE); zplus=i*1e-12; dE=E(2)-E(1);
f1=1./(1+exp((E-mu1)./kT));
f2=1./(1+exp((E-mu2)./kT));

%Коэффициент пропускания
I=0; %Ток
for k=1:NE
    sig1=zeros(Np); sig2=zeros(Np); sig3=zeros(Np);
    ck=1-((E(k)+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
    sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
    ck=1-((E(k)+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
    sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');
    G=inv(((E(k)+zplus)*eye(Np))-T-diag(U1)-sig1-sig2-sig3);
    TM(k)=real(trace(gam1*G*gam2*G'));
    I=I+(dE*IE*TM(k)*(f1(k)-f2(k)));
end
V,I

XX=a*1e9*[1:1:Np];
XS=XX([1:NS-4]); XD=XX([NS+NC+5:Np]);

```

```

hold on
%h=plot(TM,E,'b');
h=plot(XX,U1+UB,'b');
h=plot(XS,mu1*ones(1,NS-4),'b-');
h=plot(XD,mu2*ones(1,ND-4),'b-');
%axis([0 1.1 -.2 .8])
axis([0 15 -.2 .8])
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
%xlabel('Transmission -> ')
 xlabel('z (nm) ->')
 ylabel('Energy (eV) -> ')
grid on

```

**%Рис.9.5.8**

```

clear all

% Все параметры в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31; IE=(q*q)/(2*pi*hbar);
Ef=0.1; kT=.025;

%Входящие значения параметров
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q);
NS=15; NC=30; ND=15; Np=NS+NC+ND;

%Матрица гамильтониана
NS=15; NC= 16; ND=15; Np=NS+NC+ND;
UB=[zeros(NS,1); .4*ones(4,1); zeros(NC-8,1); .4*ones(4,1); zeros(ND,1)];
%RT barrier
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
T=T+diag(UB);

%Напряжение
V=0; mu1=Ef+(V/2); mu2=Ef-(V/2);
U1=V*[.5*ones(1,NS) linspace(0.5,-0.5,NC) -.5*ones(1,ND)];
U1=U1'; %Профиль приложенного потенциала

%Cетка по энергии для метода функций Грина
NE=501; E=linspace(-.2,.8,NE); zplus=i*le-12; dE=E(2)-E(1);
f1=1./((1+exp((E-mu1)./kT)));
f2=1./((1+exp((E-mu2)./kT)));

%Коэффициент пропускания

```

```

I=0; %Ток
for k=1:NE
    sig1=zeros(Np); sig2=zeros(Np); sig3=zeros(Np);
    ck=l-((E(k)+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
    sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
    ck=1-((E(k)+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
    sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');
    sig3(Np/2,Np/2)=-i*0.25; gam3=i*(sig3-sig3'); %Büttiker probe
    G=inv(((E(k)+zplus)*eye(Np))-T-diag(U1)-sig1-sig2-sig3);
    T12=real(trace(gam1*G*gam2*G'));
    T13=real(trace(gam1*G*gam3*G'));
    T23=real(trace(gam2*G*gam3*G'));
    TM(k)=T12+(T13*T23/(T13+T23));
    I=I+(dE*IE*TM(k)*(f1(k)-f2(k)));
end
V,I

XX=a*1e9*[1:1:Np];
XS=XX([1:NS-4]); XD=XX([NS+NC+5:Np]);

hold on
h=plot(TM,E,'b');
axis([0 1.1 -.2 .8])
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('Transmission -> ')
ylabel('Energy (eV) -> ')
grid on

```

### %Рис.9.5.10

```
clear all
```

```
% Все параметры в системе MKC (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31; IE=(q*q)/(2*pi*hbar);
Ef=0.1; kT=.025;
```

```
%Входящие значения параметров
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q);
NS=15; NC=30; ND=15; Np=NS+NC+ND;
```

%Матрица гамильтониана

```
%NS=15; NC=20; ND=15; Np=NS+NC+ND; UB=0*ones(Np,1); %Случай, когда
нет барьера
```

```

%NS=23; NC=4; ND=23; Np=NS+NC+ND;
%UB=[zeros(NS,1); 0.4*ones(NC,1); zeros(ND,1)]; %tunneling barrier
NS=15; NC=16; ND=15; Np=NS+NC+ND;
UB=[zeros(NS,1); 0.4*ones(4,1); zeros(NC-8,1); 0.4*ones(4,1); zeros(ND,1)];
%RT-барьер
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-(t0*diag(ones(1,Np-1),-1));
T=T+diag(UB);

%Напряжение
NV=26; VV=linspace(0,.5,NV);
for iV=1:NV
    V=VV(iV); mu1=Ef+(V/2); mu2=Ef-(V/2);
    U1=V*[.5*ones(1,NS) linspace(0.5,-0.5,NC) -.5*ones(1,ND)];
    U1=U1'; %Профиль приложенного потенциала

    %Сетка по энергии для метода функций Грина
    NE=101; E=linspace(-.2,.8,NE); zplus=i*1e-12; dE=E(2)-E(1);
    f1=1./((1+exp((E-mu1)./kT)));
    f2=1./((1+exp((E-mu2)./kT));
    %For infinite 2-D cross-section
    %f1=(2*m*kT*q/(2*pi*hbar^2)).*log(1+exp((mu1-E)./kT));
    %f2=(2*m*kT*q/(2*pi*hbar^2)).*log(1+exp((mu2-E)./kT));

    %Коэффициент пропускания
    I=0; %Ток
    for k=1:NE
        sig1=zeros(Np); sig2=zeros(Np); sig3=zeros(Np);
        ck=1-((E(k)+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
        sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
        ck=1-((E(k)+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
        sig2(Np,Np)=-t0*exp(j*ka); gam2=i*(sig2-sig2');
        sig3(Np/2,Np/2)=-i*0.00025; gam3=j*(sig3-sig3'); %Büttiker probe
        G=inv(((E(k)+zplus)*eye(Np))-T-diag(U1)-sig1-sig2-sig3);
        T12=real(trace(gam1*G*gam2*G'));
        T13=real(trace(gam1*G*gam3*G'));
        T23=real(trace(gam2*G*gam3*G'));
        TM(k)=T12+(T13*T23/(T13+T23));
        I=I+(dE*IE*TM(k)*(f1(k)-f2(k)));
    end
    II(iV)=I;V,I
end

XX=a*1e9*[1:1:Np];

```

```
XS=XX([1:NS-4]); XD=XX([NS+NC+5:Np]);
```

```
hold on
h=plot(VV,II,'b');
%h=plot(XX,U1+UB,'b');
%h=plot(XS,mu1*ones(1,NS-4),'b-');
%h=plot(XD,mu2*ones(1,ND-4),'b-');
axis([0 .5 0 3.5e-7])
%axis([0 15 -.3 .7])
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel(' Voltage ( V ) → ')
%xlabel('z (nm) →')
%ylabel('Energy (eV) → ')
ylabel('Current (A) → ')
grid on
```

## Глава 10

### %Рис. 10.4.4

```
clear all
```

```
beta=pi*linspace(-1,1,201); w0=1;
y=sqrt(2*w0*(1-cos(beta)));
```

```
hold on
h=plot(beta,y,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
grid on
```

### %Рис. 10.4.5

```
clear all
```

```
beta=pi*linspace(-1,1,201); w1=1; w2=2;
for n=1:201
    A=[w1+w2 w1+(w2*exp(-i*beta(n))); w1+(w2*exp(i*beta(n))) w1+w2];
    [V,D]=eig(A); D=sort(real(diag(D)));
    D1(n)=real(sqrt(D(1))); D2(n)=real(sqrt(D(2)));
end
```

```

hold on
h=plot(beta,D1,'b');
h=plot(beta,D2,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
grid on

```

## Глава 11

### %Рис.11.2.2, 11.2.7

```
clear all
```

%1-D случай с упругим механизмом нарушения фазы и/или когерентным, зависимость T от E, фиксированная длина

```
% Все параметры в системе MKC (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=1*9.1e-31; IE=(q*q)/(2*pi*hbar); kT=.025;
```

```
%Входящие значения параметров
```

```
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q); D=0.01*0;
```

```
%Сетка по энергии
```

```
NE=401; E=linspace(.1,.3,NE); zplus=i*1e-12; dE=E(2)-E(1);
```

```
%Напряжение
```

```
V=0.01; f1=1/(1+exp((-V/2)/kT));
f2=l/(l+exp((V/2)/kT));
```

```
%Гамильтониан
```

```
Np=40; UB=zeros(Np,1); UB(5)=0.5*1; UB(36)=0.5*1;
U1=V*linspace(0.5,-0.5,Np)'; XX=a*linspace(0,Np-1,Np);
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
-(t0*diag(ones(1,Np-1),-1));
```

```
%Итерационное решение
```

```
for k=1:NE
```

```
    sig1=zeros(Np); sig2=zeros(Np);
    ck=1-((E(k)+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
    sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
    ck=1-((E(k)+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
    sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');
```

```

%calculating the Green function, G sell-consistently
G=inv(((E(k)+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2); change=1;
    while(change>1e-4)
        sigp=diag(D.*diag(G));
S=inv(((E(k)+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2-sigp);
change=sum(sum(abs(G-S)))/(sum(sum(abs(G)+abs(S))));
        G=(0.5*G)+(0.5*S);
    end
    G=S; A=i*(G-G');
    M=D*(G.*conj(G));

%calculating the electron density,n(r;E)
gamp=i*(sigp-sigp'); gamma=gam1+gam2+gamp;
    sigin1=f1*gam1; sigin2=f2*gam2;
    n=(inv(eye(Np)-M))*diag(G*(sigin1+sigin2)*G');
    siginp=D*diag(n);

%calculating the correlation function Gn
    Gn=G*(sigin1+sigin2+siginp)*G';
%calculating the effective transmission
    I1(k)=(1/(f2-f1))*real(trace(gam1*Gn)-trace(sigin1*A));
    I2(k)=(1/(f1-f2))*real(trace(gam2*Gn)-trace(sigin2*A));
end

hold on
h=plot(I1,E,'b');
%h=plot(I2,E,'bx');
%h=plot(1e9*XX,U1+UB,'b');
%h=plot(1e9*XX,U1+UB,'bo');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
%xlabel('x (nm) → ')
%ylabel('Potential (eV) → ')
xlabel('Transmission → ')
ylabel('Energy (eV) → ')
grid on
axis([0 1.1 .1 .3])

```

### %Рис. 11.2.4, 11.2.6

clear all

%многомодовый когерентный транспорт, зависимость T от E

```
% Параметры (все даны в системе МКС, кроме энергии, которая указана в эВ)
hbar=1.06e-34; q=1.6e-19; m=9.1e-31; IE=(q*q)/(2*pi*hbar); kT=.025;

%Входящие значения параметров
a=5e-10; t0=(hbar^2)/(2*m*(a^2)*q);
%Cетка по энергии
NE=11; % значение 11 для одного рассеивателя, значение 101 для двух рассеивателей
E=linspace(0.1,0.3,NE); zplus=i*1e-12; dE=E(2)-E(1);

%Напряжение
V=0.01; f1=1/(1+exp((-V/2)/kT));
f2=1/(1+exp((V/2)/kT));

%Поперечные моды
NW=15; NT=7;
alpha=(4*t0*diag(ones(1,NW)))-(t0*diag(ones(1,NW-1),1))
-(t0*diag(ones(1,NW-1),-1));
[VT,D]=eig(alpha); [D ind]=sort(diag(D));
in=[ ]; for k=1:NT
    in=[in ind(k)]; end
VT=VT(:,in); D=diag(VT'*alpha*VT);

%Гамильтониан
Np=40; UB=zeros(Np,1); UB(5)=0.25*1; UB(36)=0.25*0;
impshape=[linspace(0,1,7) linspace(1,0,NW-7)];
U1=V*linspace(0.5,-0.5,Np)';
al=alpha+(U1(1)*eye(NW,NW));
H=VT'*al*VT; H1=H;
Z=zeros(NT,NT); bet=-t0*eye(NT,NT);
for N=2:Np
    al=alpha+(U1(N)*eye(NW,NW)); al1=al;
    al=al+(diag(UB(N))*impshape);
    al=VT'*al*VT; H=[H bet;bet' al];
    al1=VT'*al1*VT; H1=[H1 bet;bet' al1]; %Используйте для случая одного рассеивателя
    bet=[Z;bet];
end

%Вычисление коэффициента пропускания
for k=1:NE
    ck=(D-E(k)-zplus+U1(1))/(2*t0); ka=acos(ck);
    s1=-t0*exp(i.*ka);
```

```

sig1=[diag(s1)zeros(NT,NT*(Np-1)); zeros(NT*(Np-1),NT*Np)];
ck=(D-E(k)-zplus+U1(Np))/(2*t0); ka=acos(ck);
s2=-t0*exp(i.*ka);
sig2=[zeros(NT*(Np-1),NT*Np); zeros(NT,NT*(Np-1))diag(s2)];
gam1=i*(sig1-sig1'); gam2=i*(sig2-sig2');

G=inv(((E(k)+zplus)*eye(NT*Np))-H-sig1-sig2);
T(k)=real(trace(gam1*G*gam2*G'));
G1=inv(((E(k)+zplus)*eye(NT*Np))-H1-sig1-sig2); %Используйте для случая одного рассеивателя
M(k)=real(trace(gam1*G1*gam2*G1'));[k T(k) M(k)];
%Используйте для случая одного рассеивателя
end

Tsc=T./(2-(T./M)); %Квазиклассическая добавка, использовать для случая одного рассеивателя

%save condfluct2
XX=a*linspace(0,Np-1,Np);

hold on
%h=plot(T,E,'b');
h=plot(Tsc,E,'b-');
%h=plot(M,E,'b');
set(h,'linewidth',[2,0]);
set(gca,'Fontsize',[24]);
xlabel('Transmission -> ')
ylabel('Energy (eV) -> ')
axis([0 5 .1 .3])
grid on

```

### %Рис.11.2.8

clear all

%1-D -случай упругого когерентного процесса и/или процесса с разрушением фазы, зависимость R от L, фиксированное значение E

% Все значения параметров в системе МКС (кроме энергии, которая дана в эВ)  
 $\text{hbar}=1.06e-34$ ;  $q=1.6e-19$ ;  $m=.25*9.1e-31$ ;  $IE=(q*q)/(2*pi*hbar)$ ;  $kT=.025$ ;

%Входящие значения параметров  
 $a=3e-10$ ;  $t0=(hbar^2)/(2*m*(a^2)*q)$ ;  $D=0.05$ ;  $V=0.01$ ;

```

%Напряжение

%Сетка для энергии
E=0.1; zplus=i*1e-12;
f1=1/(1+exp((-V/2)/kT));
f2=1/(1+exp((V/2)/kT));

%Ток
for k=2:21
    Np=k; UB=zeros(Np,1); U1=V*linspace(0.5,-0.5,Np)';k
    T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
    -(t0*diag(ones(1,Np-1),-1));

    sig1=zeros(Np); sig2=zeros(Np); sig3=zeros(Np);
    ck=1-((E+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
    sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
    ck=1-((A+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
    sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');

    %Расчет функции Грина, функция G вычисляется самосогласованно
    G=inv(((E+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2); change=1;
    while(change>1e-4)
        sigp=diag(D*diag(G));
        S=inv(((E+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2-sigp);
        change=sum(sum(abs(G-S)))/(sum(sum(abs(G)+abs(S))));
        G=(0.5*G)+(0.5*S);
    end
    G=S; A=i*(G-G');
    M=D*(G.*conj(G));

    %расчет функций, описывающих рассеяние внутрь структуры из контактов F1,F2
    gam1=i*(sig1-sig1'); gam2=i*(sig2-sig2');
    gamp=i*(sigp-sigp'); gamma=gam1+gam2+gamp
    sign1=f1*gam1; sign2=f2*gam2;
    n=(inv(eye(Np)-M))*diag(G*(sign1+sign2)*G');
    signp=D*diag(n);

    %расчет корреляционной функции Gn
    Gn=G*(sign1 +sign2+signp)*G';

    %расчет тока
    I1(k-1)=(1/(f1-f2))*real(trace(gam1*Gn)-trace(sign1*A));

```

```

I2(k-1)=(1/(f1-f2))*real(trace(gam2*Gn)-trace(sigin2*A));
L(k-1)=k*a*1e10;
end

L=L./10; %Переход от ангстремов к нанометрам
hold on
h=plot(L,I2,'b');
%h=plot(I1+I2,'g-');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('Length (nm) → ')
ylabel('Normalized resistance → ')
axis([0 6 0 3])
grid on

```

### %Рис.11.3.1, 11.3.2, 11.3.3

```
clear all
```

```
%1-D -случай с неупругим рассеянием
```

```
%Значения параметров в системе МКС (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31; Ef=0.15; kT=0.025;
```

```
%Входящие значения параметров
```

```
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q);
```

```
%Матрица гамильтониана
```

```
Np=40; UB=0*[zeros(10,1); 0.25*ones(Np-10,1)];
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
-(t0*diag(ones(1,Np-1),-1));
```

```
%Напряжение
```

```
V=0.1; mu1=Ef+(V/2); mu2=Ef-(V/2);
```

```
U1=V*[.5*ones(1,1) linspace(0.5,-0.5,Np-2) -.5*ones(1,1)];
```

```
%Профиль приложенного потенциала
```

```
D=1e-1; %Амплитуда рассеивателя
```

```
%Сетка по энергии
```

```
NE=101; E=linspace(-.05,.35,NE); zplus=i*1e-12; dE=E(2)-E(1);
```

```
f1=1./(1+exp((E-mu1)./kT));
```

```
f2=1./(1+exp((E-mu2)./kT));
```

```
%Исходное предположение
```

```

sigin=0*ones(Np,NE); sigout=0*ones(Np,NE);

%Итерационное решение уравнений транспорта
change=1; it=1; n=zeros(Np,NE); p=zeros(Np,NE);
while change>1e-3
for k=1:NE
    sig1=zeros(Np); sig2=zeros(Np);
    ck=1-((E(k)+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
    sig1(1,1)=-t0*exp(i*ka); gam1=i*(sig1-sig1');
    ck=1-((E(k)+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
    sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');
    sigin1(:,k)=f1(k)*diag(gam1); sigin2(:,k)=f2(k)*diag(gam2);
    sigout1(:,k)=(1-f1(k))*diag(gam1); sigout2(:,k)=(1-f2(k))*diag(gam2);
    gamp=sigin(:,k)+sigout(:,k);

    G=inv(((E(k)+zpius)*eye(Np))-T-diag(U1+UB)-sig1-sig2+(i*0.5*diag(gamp)));
    A=diag(i*(G-G'));
    n(:,k)=real(diag(G*((f1(k)*gam1)+(f2(k)*gam2)+diag(sigin(:,k)))*G'));
    p(:,k)=A-n(:,k);
end

off=0; %Меньше, чем NE-1, и равно нулю в случае упругих взаимодействий
C=exp(-dE*off/kT);
ne=n(:,[off+1:NE]); ne=[ne zeros(Np,off)];
na=n(:,[1:NE-off]); na=[zeros(Np,off) na];
pa=p(:,[off+1:NE]); pa=[pa zeros(Np,off)];
pe=p(:,[1:NE-off]); pe=[zeros(Np,off) pe];

siginnew=(D*ne)+(C*D*na);
sigoutnew=(D*pe)+(C*D*pa);
change=sum(sum(abs(siginnew-sigin)));
change=change+sum(sum(abs(sigoutnew-sigout)));
sigin=((1-it)*sigin)+(it*siginnew);
sigout=((1-it)*sigout)+(it*sigoutnew);
end

I1=real((sigout1.*n)-(sigin1.*p)); I1=sum(I1);
I2=real((sigout2.*n)-(sigin2.*p)); I2=sum(I2);
I3=real((sigout.*n)-(sigin.*p)); I3=sum(I3);

I123=(dE/V)*[sum(I1) sum(I2) sum(I3)], %Nonnalized Conductance
IE=(dE/(V*V))*[sum(E.*I1) sum(E.*I2) sum(E.*I3)], %Norrnalized Power
kirchoff=[sum(I123) sum(IE)],

```

```
%проверка закона сохранения тока и потока энергии
```

```
save inel0
hold on
h=plot(I1,E,'b');
h=plot(I2,E,'b-');
%h=plot(I3,E,'c');
set(h,'linewidth',[2,0])
set(gca,'Fontsize',[24])
xlabel('Normalized current / energy → ')
ylabel('Energy (eV) → ')
axis([-2 .2 -.05 .35])
```

### %Рис. 11.4.4

```
clear all
```

```
%Баллистическое самосогласованное решение
```

```
%Значения параметров в системе МКС (кроме энергии, которая дана в эВ)
```

```
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31;
```

```
kT=0.0259; zplus=i*1e-12; eps0=8.854e-12;
```

```
%Входящие значения параметров
```

```
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q); N=40;
```

```
Ef=0.1; Ec=-0.5; Vg=0;
```

```
r=5e-9; tox=5e-9; K=2; %Используйте большое значение проницаемости для
предела Лапласа
```

```
U0=q/2/pi/a/K/eps0.*log((r+tox)/r)
```

```
%Матрица гамильтониана
```

```
Np=40;
```

```
H0=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))
```

```
-(t0*diag(ones(1,Np-1),-1));
```

```
%Сетка по энергии
```

```
NE=401; E=linspace(-0.5,0.3,NE); dE=E(2)-E(1); I0=(q^2)/hbar/2/pi
```

```
%Напряжение
```

```
iV=41; V=linspace(0,0.4,iV); n0=0; UL=-Vg*ones(Np,1); U=UL;
```

```
for kk=1:iV
```

```
    Vd=V(kk); mu1=Ef; mu2=Ef-Vd;
```

```
    sig1=zeros(Np); sig2=zeros(Np);
```

```
    epsilon=1;
```

```

while (epsilon>0.001)
    rho=0;
    for k=1:NE
        f1=1/(1+exp((E(k)-mu1)/kT)); f2=1/(1+exp((E(k)-mu2)/kT));

        cka1=1-(E(k)+zplus-Ec)/2/t0; ka1=acos(cka1);
        sig1(1,1)=-t0*exp(i*ka1); gam1=i*(sig1-sig1');
        cka2=1-(E(k)+zplus-Ec+Vd)/2/t0; ka2=acos(cka2);
        sig2(N,N)=-t0*exp(i*ka2); gam2=i*(sig2-sig2');
        G=inv((E(k)+zplus)*eye(N)-H0-diag(U)-sig1-sig2);
        A=i*(G-G');
        sigin1=f1*gam1; sigin2=f2*gam2;
        Gn=G*(sigin1+sigin2)*G'; rho=rho+dE/2/pi*Gn;
        T(k)=trace(gam1*G*gam2*G');
        I1(k)=real(trace(sigin1*A)-trace(gam1*Gn));
        I2(k)=-real(trace(sigin2*A)-trace(gam2*Gn));
        end
        n=real(diag(rho)); Unew=UL+(U0*(n-n0));
        dU=Unew-U; epsilon=max(abs(dU))
        U=U+0.25*dU;
        if Vd==0
            n0=n; epsilon=0; end
        end
        ID1=2*I0*dE*sum(I1); ID2=2*I0*dE*sum(I2); %множитель 2 отвечает
спину
        I(kk)=ID1;
    end

    save IV2
    IonL=I0*Ef
    hold on
    h=plot(V,I,'b');
    h=plot(V,IonL*ones(iV,1),'bx');
    set(h,'linewidth',[2.0])
    set(gca,'Fontsize',[24])
    grid on
    xlabel('Voltage (V) → ')
    ylabel('Current (A) → ')

```

**%Рис.Е.11.5**

```

    clear all
    %1-D -случай туннелирования и/или упругий механизм нарушения фазы, за-
    висимость R от L, фиксированное значение E

```

```
%Параметры указаны в системе MKC (кроме энергии, которая дана в эВ)
hbar=1.06e-34; q=1.6e-19; m=.25*9.1e-31; IE=(q*q)/(2*pi*hbar); kT=.025;
zplus=i*1e-51;
```

%Входящие значения параметров

```
a=3e-10; t0=(hbar^2)/(2*m*(a^2)*q);
```

```
D=3e-1; %Амплитуда рассеивателя: 2e-1 (x's) and 3e-1 (o's) eV^2
```

```
V=0.001; %Приложенное напряжение
```

```
mu=0.1; %Энергия Ферми
```

```
f1=1/(1+exp((-V/2)/kT)); %Распределение Ферми в контакте 1 при E=mu
```

```
f2=1/(1+exp((V/2)/kT)); %Распределение Ферми в контакте 2 при
```

```
E=mu
```

%Фактический расчет

```
E=mu;
```

```
for k=5:26
```

```
Np=k; %Length of barrier = (Np-2)*a
```

```
UB=[0; 0.5*ones(Np-2,1); 0]; %Высота барьера
```

```
U1=V*linspace(0.5,-0.5,Np)'; %Профиль приложенного потенциала
```

```
T=(2*t0*diag(ones(1,Np)))-(t0*diag(ones(1,Np-1),1))-...
```

```
(t0*diag(ones(1,Np-1),-1)); %Гамильтониан в приближении сильной
```

связи

```
sig1=zeros(Np); sig2=zeros(Np); sig3=zeros(Np);
```

```
ck=l-((E+zplus-U1(1)-UB(1))/(2*t0)); ka=acos(ck);
```

sig1(1,1)=-t0\*exp(i\*ka); gam1=i\*(sig1-sig1'); %Собственная энергетическая функция для контакта 1

```
ck=l-(E+zplus-U1(Np)-UB(Np))/(2*t0)); ka=acos(ck);
```

```
sig2(Np,Np)=-t0*exp(i*ka); gam2=i*(sig2-sig2');
```

%Self-energy for contact 2

```
%calculating the Green function, G self-consistently
```

```
G=inv(((E+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2); change=1;
```

```
while(change>1e-15)
```

```
sigp=diag(D*diag(G)); %Self-energy due to scattering
```

```
S=inv(((R+zplus)*eye(Np))-T-diag(U1+UB)-sig1-sig2-sigp);
```

```
change=sum(sum(abs(G-S)))/(sum(sum(abs(G)+abs(S))));
```

```
G=(0.5*G)+(0.5*S);
```

```
end
```

```
G=S; A=i*(G-G'); M=D*(G.*conj(G));
```

%Расчет функций, описывающий рассеяние внутри структуры из контактов F1,F2

```

gam1=i*(sig1-sig1'); gam2=i*(sig2-sig2');
gamp=i*(sigp-sigp'); gamma=gam1+gam2+gamp;
    sigin1=f1*gam1; sigin2=f2*gam2;
n=(inv(eye(Np)-M))*diag(G*(sigin1+sigin2)*G');
siginp=D*diag(n); %Входящий ток, обусловленный рассеиванием

%Вычисление корреляционной функции Gn
Gn=G*(sigin1+sigin2+siginp)*G';

%Вычисление тока
I1(k-4)=(1/(f1-f2))*real(trace(gam1*Gn)-trace(sigin1*A));
I2(k-4)=(1/(f1-f2))*real(trace(gam2*Gn)-trace(sigin2*A));
L(k-4)=(k-2)*a* 1e9; %в нанометрах
end

hold on
h=plot(L,log10(-1./I1),'g'); %Ток на левом конце
h=plot(L,log10(1./I2),'go'); %Ток на правом конце
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
xlabel('Length (nm) → ')
ylabel('log10 (resistance) → ')
grid on

```

### %Рис.Е.11.6

```

clear all
%0-D -случай с неупругим рассеянием

```

% Параметры указаны в системе MKC (кроме энергии, которая дана в эВ)  
 $\hbar = 1.06e-34$ ;  $q = 1.6e-19$ ;  $10 = q^*q/(2*pi^*\hbar)$ ;

%Параметры

```

H0=5; Ef=0; kT=0.0025; dE=0.0005; zplus=i*1e-12; gamma=0.1;
D0=0; Dnu=0*[0.5 0.7]; Nph=size(Dnu,2);
hnu=[100 550]; %Умножаем на dE для получения фактического
значения hnu

```

$$Nhnu=1./((exp(dE*hnu./kT))-1);$$

%Напряжение

```

NV=203; VV=linspace(-0.51,0.5,NV); dV=VV(2)-VV(1);
for iV=1:NV
    V=VV(iV); mu1=Ef; mu2=Ef-V; U1=(-0.5)*V;

```

```

%Сетка по энергии
E=[mu2-(10*kT)-(10*dE):dE:mu1+(10*kT)+(10*dE)];
if V<0
    E=[mu1-(10*kT)-(10*dE):dE:mu2+(10*kT)+(10*dE)];
end
NE=size(E,2);[iV NE]
f1=1./(1+exp((E-mu1)./kT));
f2=1./(1+exp((E-mu2)./kT));

%Исходное предположение
n=zeros(1,NE); p=zeros(1,NE);
sigin1=zeros(1,NE); sigout1=zeros(1,NE);
sigin2=zeros(1,NE); sigout2=zeros(1,NE);
sigin=0*ones(1,NE); sigout=0*ones(1,NE);

%Итерационное решение уравнения транспорта
change=1; it=1;
while change>1e-3
    for k=1:NE
        sig1=-i*gamma/2; gam1=i*(sig1-sig1');
        sig2=-i*gamma/2; gam2=i*(sig2-sig2');
        sigin1(k)=f1(k)*gam1; sigin2(k)=f2(k)*gam2;
        sigout1(k)=(1-f1(k))*gam1; sigout2(k)=(1-f2(k))*gam2;
        gamp=sigin(k)+sigout(k);
        G=inv((E(k)+zplus)-H0-U1-sig1-sig2+(i*0.5*gamp));
        A=i*(G-G');
        n(k)=real(G*((f1(k)*gam1)+(f2(k)*gam2)+sigin(k)*G'));
        p(k)=A-n(k);
    end

    siginnew=D0*n; sigoutnew=D0*p;
    for iph=1:Nph
        inu=hnu(iph);
        if inu<NE
            ne=n([inu+1:NE]); ne=[ne zeros(1,inu)];
            na=n([1:NE-inu]); na=[zeros(1,inu) na];
            pe=p([inu+1:NE]); pe=[pe zeros(1,inu)];
            pa=p([1:NE-inu]); pa=[zeros(1,inu) pa];
            siginnew=siginnew+((Nhnu(iph)+1)*Dnu(iph)*ne)+...
                (Nhnu(iph)*Dnu(iph)*na);
            sigoutnew=sigoutnew+(Nhnu(iph)*Dnu(iph)*pe)+...
                ((Nhnu(iph)+1)*Dnu(iph)*pa);
        end
    end

```

```

    end

    change=sum(sum(abs(siginnew-sigin)));
    change=change+sum(sum(abs(sigoutnew-sigout)));
        sigin=((1-it)*sigin)+(it*siginnew);
        sigout=((1-it)*sigout)+(it*sigoutnew);
end

I1=real((sigout1.*n)-(sigin1.*p)); I1=sum(I1);
I2=real((sigout2.*n)-(sigin2.*p)); I2=sum(I2);
I3=real((sigout.*n)-(sigin.*p)); I3=sum(I3);

I123=dE*[sum(I1) sum(I2) sum(I3)], %Нормированная проводимость
%IE=(dE/(V*V))*[sum(E.*I1) sum(E.*I2) sum(E.*I3)], %Нормированная мощн-
ность
%kirchoff=[sum(I123) sum(IE)], %Проверка сохранения полного тока
and energy current

II(iV)=sum(I2)*dE*I0;
end

G1=diff(II)./dV; VG=VV([2:NV]);
IETS=diff(G1)./dV; VETS=VV([3:NV]);

hold on
%h=plot(VV,I1,'rx');
h5=plot(VG,G1,'b-');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
% xlabel('Voltage (V) -> ')
% ylabel('d2I/dV2 -> ')

```

## Приложения

### %Рис.П.5.2, П.5.3

```
clear all
```

```

NE=1001; E=linspace(-.25,.25,NE); zplus=i*1e-3;dE=E(2)-E(1); kT=.00026;
Nep=5001; ep=linspace(-1,1,Nep); tau=0.05; dep=ep(2)-ep(1); delta=3.117*tau*tau/2
ep0=-25*delta; U=50*delta;[U/pi abs(ep0)]/delta
D=ones(1,Nep); f=1./(1+exp(ep./kT)); fK=1./(1+exp(E./kT)); tau=0.06;

```

```

for kE=1:NE
s0(kE)=dep*tau*tau*sum(D./(E(kE)-ep+zplus));
s1(kE)=dep*tau*tau*sum(D./(E(kE)-ep0-ep0-U+ep+zplus));
s2(kE)=dep*tau*tau*sum(D.*f./(E(kE)-ep+zplus));
s3(kE)=dep*tau*tau*sum(D.*f./(E(kE)-ep0-ep0-U+ep+zplus));
end

g=U./(E-ep0-U-s0-s1);
GK=(1+(0.5*g))./(E-ep0-s0+(g.*(s2+s3)));
G=(1+(0.5*U./(E-ep0-U-s0)))./(E-ep0-s0);
A=i*(G-conj(G))/(2*pi); dE*sum(A)
AK=i*(GK-conj(GK))/(2*pi); dE*sum(AK)
dE*sum(AK.*fX)
del=-dE*sum(imag(s0))

hold on
%h=plot(E,-imag(s0));
%h=plot(E,imag(s1));
%h=plot(E,imag(s2),'mx');
%h=plot(E,imag(s3),'m');
h=plot(A,E,'b-');
h=plot(AK,E,'b');
set(h,'linewidth',[2.0])
set(gca,'Fontsize',[24])
grid on
xlabel(' D(E) per eV -> ')
ylabel(' E(eV) -> ')

```

# **Дополнительная литература**

Приведенный ниже список дополнительной литературы, безусловно, не является исчерпывающим. В него попала лишь малая часть статей и книг, призванных помочь читателю в изучении материала данной книги.

## **Общие руководства:**

В настоящей книге использованы лекции, читавшиеся мной на протяжении ряда лет для студентов старших, а с недавнего времени младших и средних курсов. Читателю будет полезно ознакомиться и с видеолекциями (презентациями), посвященными главным разделам данной книги. Соответствующая информация доступна на моем сайте

<http://dynamo.ecn.purdue.edu/~datta>

и через сервер “nanohub” на сайте [www.nanohub.org](http://www.nanohub.org).

## **1. Пролог**

Содержание этой главы, а также Приложения, переработаны с изменениями на основе книги [15].

Пример, использованный в параграфе 1.4 для иллюстрации того, как важно учитывать конкретный профиль потенциала при определении вольт-амперных характеристик (см. рис. 1.4.2, 1.4.3), появился благодаря экспериментам, обсуждаемым в работе

Datta S., Tian W., Hong S., Reifenberger R., Henderson J., and Kubiak C.P. (1997). STM current-voltage characteristics of self-assembled monolayers (SAMs). *Physical Review Letters*, **79**, 2530.

В этой книге не обсуждается подробно режим кулоновской блокады. Исключение составляют параграфы 1.5 и 3.4. Для дальнейшего чтения можно рекомендовать, например, книгу [4], а также следующие работы:

- Kastner M. (1993). Artificial atoms. *Physics Today*, **46**, 24.
- Likharev K. (1999). Single-electron devices and their applications. *Proceedings of IEEE*, **87**, 606.
- Kouwenhoven L.P. and McEuen P.L. (1997). Single electron transport through a quantum dot. In *Nano-Science and Technology*, ed. G. Timp, AIP Press.
- Ferry D.K. and Goodnick S.M. (1997). *Transport in Nanostructures*. Cambridge University Press. Chapter 4.

Простая модель нанотранзистора, кратко описанная в параграфе 1.6, фактически совпадает с моделью, обсуждаемой в работе

Rahman A., Guo J., Datta S., and Lundstrom M. (2003). Theory of ballistic transistors. *IEEE Transistors and Electronic Devices*, **50**, 1583.

В параграфе 1.6 я упомянул о том, что анализ транспорта, в режиме, при котором  $t \leq U_0$ , сталкивается с существенными трудностями теоретического характера (см., например, [20]). В связи с этим в литературе широко обсуждалась концепция т. н. «латинжеровской жидкости», о чём можно прочитать, например, в [21].

Для изучения вопросов, связанных с эффектом Кондо, читателю можно рекомендовать книгу [28].

## 2. Уравнение Шредингера

Описание основных принципов и методов квантовой механики, включая проблему атома водорода, читатель может найти в любом вводном курсе квантовой механики и в том числе в таких книгах, как:

- Eisberg R. and Resnick R. (1974, 1985). *Quantum Physics*. Wiley.
- Cox P.A. (1996). *Introduction to Quantum Theory and Atomic Structure*. Oxford Chemistry Primers.

Примечание редактора перевода. С основами квантовой механики можно также познакомиться по следующим книгам, изданным на русском языке: [63, 62, 57].

В этой главе обсуждался метод конечных разностей. Читателю, интересующемуся методом конечных элементов, можно обратиться к руководствам [44] и [51].

### 3. Самосогласованное поле

Для подробного изучения метода самосогласованного поля можно рекомендовать книгу [23], а также следующие руководства:

Slater J.C. (1963-1974). *Quantum Theory of Molecules and Solids*, Vols. I-IV. McGraw-Hill.

Grant G.H. and Richards W.G. (1995). *Computational Chemistry*. Oxford Chemistry Primers.

Подробное описание структур, молекул и твердых тел дано в книге:

Pettifor D. (1995). *Bonding and Structure of Molecules and Solids*. Oxford University Press.

Примечание редактора перевода. Читателю можно рекомендовать также следующие книги по физикеnanoструктур на русском языке: [59, 60, 61].

Обсуждение, представленное в параграфе 3.4, основано на методах, описанных в работе:

Beenakker C.W.J. (1991). Theory of Coulomb blockade oscillations in the conductance of a quantum dot. *Physical Review B*, **44**, 1646.

Одночастичные модели энергетических уровней обсуждаются, например, в работе:

Brus L.E. (1983). A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *Journal of Chemical Physics*, **79**, 5566.

Читателю можно рекомендовать и более поздние источники:

Bakkers E.P.A.M., Hens Z., Zunger A., Franceschetti A., Kouwenhoven L.P., Gurevich L., and Vanmaekelberg D. (2001). Shell-tunneling spectroscopy of the single-particle energy levels of insulating quantum dots. *Nano Letters*, **1**, 551.

Niquet Y.M., Delerue C., Allan G., and Lannoo M. (2002). Interpretation and theory of tunneling experiments on single nanostructures. *Physical Review B*, **65**, 165334.

## 4. Базисные функции

Читатель может насладиться блестящим обсуждением природы «базисных состояний», приведенным в книге Фейнмана [17], и особенно в ее восьмой главе.

Системы ортогональных базисных функций подробно обсуждаются во всех учебниках квантовой механики, в том числе в следующих книгах:

Cohen-Tannoudji C., Diu B., and Laoloe F. (1977). *Quantum Mechanics*. Wiley.

Schiff L.I. (1955). *Quantum Mechanics*. McGraw-Hill. [Имеется русский перевод: Л. Шифф, *Квантовая механика*, М.: Мир, 1959.]

Неортогональные системы базисных функций более подробно обсуждаются, например, в следующей книге по квантовой химии:

Szabo A. and Ostlund N.S. (1996). *Modern Quantum Chemistry*. Dover.

## 5. Зонная структура

Вопросы, связанные с зонной структурой, рассматриваются в книгах по физике твердого тела, в частности, в следующих руководствах:

Ziman J.M. (1972). *Principles of the Theory of Solids*. Cambridge University Press. [Имеется русский перевод: Д. Займан, *Принципы теории твердого тела*, М.: Мир, 1974.]

Kittel C. (1976). *Introduction to Solid State Physics*. Wiley. [Имеется русский перевод: Ч. Киттель, *Введение в физику твердого тела*, М.: Мир, 1978.]

Параграф 5.3 близко воспроизводит содержание фундаментальной работы [50]. Для ознакомления с более новыми результатами и дальнейшими ссылками читатель может обратиться к следующим работам:

Boykin T.B., Klimeck G., Chris Bowen R., and Oyafuso F. (2002). Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory. *Physical Review B*, **66**, 125207, см. также ссылки в данной работе.

В нашей книге мы описали только эмпирический метод сильной связи. В зонной теории используется большое число других методов, наиболее известными из которых являются метод псевдопотенциала и  $k \cdot p$ -метод. Для ознакомления с ними читатель может обратиться к руководствам [19] и [48].

## 6. Подзоны

Подробное обсуждение природы углеродных нанотрубок можно найти в книге:

Dresselhaus M.S., Dresselhaus G., and Eklund P.C. (1996). *Science of Fullerenes and Carbon Nanotubes*. Academic.

## 7. Емкость

Впервые термин «квантовая емкость», вероятно, был введен в работе:

Luryi S. (1988). Quantum capacitance devices. *Applied Physics Letters*, **52**, 101.

Он использовался и другими авторами, например, в работе:

Katayama Y. and Tsui D.C. (1993). Lumped circuit model of two-dimensional tunneling transistors. *Applied Physics Letters*, **62**, 2563.

Для более подробного ознакомления с моделью эффективной массы для большого числа зон читатель отсыпается к книге [48].

## 8. Уширение уровней

С интересной дискуссией, посвященной описанию необратимых явлений, читатель может познакомиться, обратившись к книге:

Prigogine I. (1980). *From Being to Becoming*. Freeman.

Feynman R.P. (1965). The distinction of past and future. *Character of Physical Law*. MIT Press. Глава 5.

## 9. Когерентный транспорт

Для детального изучения метода функции пропускания читатель может обратиться к книгам [13] и [5], а также к следующим руководствам:

Imry Y. (1997). *Introduction to Mesoscopic Physics*. Oxford University Press.

Ferry D.K. and Goodnick S.M. (1997). *Transport in Nanostructures*. Cambridge University Press.

Для знакомства с более поздними работами можно рекомендовать работы [3] и [6].

Расчеты когерентного транспорта часто выполняются несамосогласованно. Так, например, был получен результат, представленный на рис. 9.5.10. Однако это обстоятельство может существенно сказаться на форме вольт-амперных характеристик. Читателю можно рекомендовать работу [7], где для расчета электронной плотности использован метод коэффициентов пропускания. Впрочем, в большей мере мне представляется оправданным применение здесь уравнений для неравновесных функций Грина, описанных в этой главе.

## 10. Некогерентный транспорт

Читателю, желающему получить более подробную информацию об электрон-фотонных взаимодействиях, рекомендуются следующие книги:

Feynman R.P. (1985). *QED: The Strange Theory of Light and Matter*. Princeton University Press. [Имеется русский перевод: Р. Фейнман, *КЭД: странная теория света и вещества*, М.: Мир, 1988.]

Marcuse D. (1980). *Principles of Quantum Electronics*. Academic.

Квазиклассические методы, в которых не учитывается квантовая интерференция, но подробно рассматриваются некогерентные процессы рассеяния, широко используются для моделирования электронных приборов. В качестве введения в данный круг вопросов читателю можно рекомендовать книгу:

Lundstrom M.S. (2000). *Fundamentals of Carrier Transport*. Cambridge University Press.

Описание электрон-фононных взаимодействий в низшем порядке, обсуждавшееся в параграфе 10.3, аналогично представленному в работе [8].

В параграфе 10.3 я упоминал о том, что принцип запрета нельзя интерпретировать на чисто интуитивном уровне. Данный круг вопросов обсуждается в параграфе 4.4 книги [24] и параграфе 8.2 руководства [13].

## 11. От атома к транзистору

Читатель, интересующийся проблемами проводимости, может обратиться к разделу «Когерентный транспорт» в седьмой главе руководства [13], где обсуждаются аналогии между электронным транспортом и распространением света.

Пример эффекта Пельтье, приведенный в параграфе 11.3.1, был предложен Ландауэром и основан на работе [30]. Обсуждение в параграфе 11.4 основано на методе, описанном в [33].

Квантовые модели, основанные на использовании неравновесных функций Грина, описывающие одномерные структуры, интенсивно развиваются исследовательской группой «NEMO». Дополнительные сведения по этому вопросу читатель может найти, например, в [27].

Данные материалы открыты для свободного доступа на сайте этого коллектива: <http://hpc.jpl.nasa.gov/PEP/gekco/nemo>.

Необходимо отметить работы, в которых метод неравновесных функций Грина использовался для описания молекулярных электронных приборов: см., например, [53]. В Приложении приведено краткое описание системы уравнений для неравновесных функций Грина с использованием *неортогональных базисов*.

## 12. Приложение: формализм вторичного квантования

Я полагаю, что в ряде задач методы, обсуждаемые в Приложении, могут оказаться недостаточными для систематического изучения высших поправок. Несколько обязательных ссылок на оригинальные работы и обзорные статьи, в которых метод неравновесных функций Грина использовался для описания бесконечных однородных сред, читатель сможет отыскать в библиографическом списке к настоящей книге под номерами [37, 24, 25, 10, 45, 35, 26].

Многие авторы используют метод неравновесных функций Грина в задачах, связанных со структурами конечных размеров, следуя работе [8]. Как уже отмечалось в параграфе 1.6, основные результаты, представленные в настоящей книге, основаны на работах [11] и [39]. Восьмая глава

работы [13] содержит обсуждение метода неравновесных функций Грина и его связи с методом Ландауэра.

Временные уравнения, представленные в данной книге, аналогичны уравнениям, рассматривавшимся в следующих работах:

Jauho A.P., Wingreen N.S., and Meir Y. (1994). Time-dependent transport in interacting and non-interacting resonant tunneling systems. *Physical Review B*, **50**, 5528.

Haug H. and Jauho A.P. (1996). *Quantum Kinetics in Transport and Optics of Semiconductors*. Springer-Verlag.

Приведенные в данной книге уравнения для функций Грина, описывающие резонанс Кондо, аналогичны использованным в работе [40].

Электронный транспорт в сильно коррелированных системах рассматривался в ряде недавно опубликованных работ. Полезными для читателя могут оказаться работы [20, 21], а также следующие книги:

Fulde P. (1991). *Electron Correlations in Molecules and Solids*. Springer-Verlag.

Mahan G.D. (1991). *Many-Particle Physics*. Plenum.

Hewson A.C. (1993). *The Kondo Problem to Heavy Fermions*. Cambridge University Press.

# Литература

- [1] Allen P. B. and Mitrovic B. (1982). Theory of superconducting  $T_c$ . In *Solid State Physics*, ed. by H. Ehrenreich *et al.*, vol. 37. Academic Press.
- [2] Ashcroft N. W. and Mermin D. (1976). *Solid State Physics*. Holt, Rinehart and Winston. [Имеется русский перевод: Н. Ашкрофт, Н. Мермин, *Физика твердого тела*, т. 1–2, М.: Мир, 1978.]
- [3] Beenakker C.W.J. (1997). Random matrix theory of quantum transport. *Review of Modern Physics*, **69**, 731.
- [4] Bonet E., Deshmukh M.M. and Ralph D.C. (2002). Solving rate equations for electron tunneling via discrete quantum states. *Physical Review B*, **65**, 045317.
- [5] Buttiker M. (1988). Symmetry of electrical conduction. *IBM Journal of Research Devices*, **32**, 317.
- [6] Buttiker M. (2001). Capacitance, charge fluctuations and dephasing in Coulomb coupled conductors. In *Interacting Electrons in Nanostructures*, ed. H. Schoeller and R. Haug. *Lecture Notes in Physics*, vol. 579. Springer-Verlag. P. 149.
- [7] Cahay M., McLennan M., Datta S., and Lundstrom M.S. (1987). Importance of space-charge effects in resonant tunneling devices. *Applied Physics Letters*, **50**, 612.
- [8] Caroli C., Combescot R., Nozieres P., and Saint-James D. (1972). A direct calculation of the tunneling current: IV. Electron-phonon interaction effects. *Journal of Physics C: Solid State Physics*, **5**, 21.
- [9] Chen C.J. (1993). *Introduction to Scanning Tunneling Microscopy*. Oxford University Press.
- [10] Danielewicz P. (1984). Quantum theory of non-equilibrium processes. *Annals of Physics, New York*, **152**, 239.
- [11] Datta S. (1989). Steady-state quantum kinetic equation. *Physical Review B* **40**, 5830.
- [12] Datta S. (1990). A simple kinetic equation for steady-state quantum transport. *Journal of Physics: Condensed Matter*, **2**, 8023.
- [13] Datta S. (1995). *Electronic transport in Mesoscopic Systems*. Cambridge University Press.
- [14] Datta S. (2000). Nanoscale Device Simulation: The Green's Function Formalism. *Superlattices and Microstructures*, **28**, 253.

- [15] Datta S. (2004). Electrical resistance: an atomistic view. *Nanotechnology*, **15**, S433.
- [16] Datta S. (2005). Spin Dephasing and «Hot Spins». *Proceedings of the International School of Physics «Enrico Fermi»*. Course CLX. Societa, Italiana di Fisica.
- [17] Feynman R.P. (1965). *Lectures in Physics*, vol. III. Addison-Wesley.
- [18] Feynman R.P. (1972). Statistical Mechanics. *Frontiers in Physics*. Addison-Wesley. [Имеется русский перевод: Р. Фейнман, *Статистическая механика*, М.: Мир, 1975.]
- [19] Fischetti M.V. (1991). Monte Carlo simulation of transport in significant semiconductors of the diamond and zinc-blend structures - Part I: homogeneous transport. *IEEE Trans. Electron Devices*, **38**, 634.
- [20] Georges A. (2004). Strongly correlated electron materials: dynamical mean-field theory and electronic structure. *AIP Conference Proceedings*, 715; arXiv.org/cond-mat.0403123
- [21] Giamarchi T. (2004). *Quantum Physics in One Dimension*. Oxford University Press.
- [22] Harrison W.A. (1999). *Elementary Electronic Structure*. World Scientific.
- [23] Herman F. and Skillman S. (1963). *Atomic Structure Calculations*. Prentice-Hall. pp. 3–9.
- [24] Kadanoff L.P. and Baym G. (1962). Quantum Statistical Mechanics. *Frontiers in Physical Lecture Notes*. Benjamin/Cummings.
- [25] Келдыш Л. В., *Диаграммная техника для неравновесных процессов*, ЖЭТФ, т. 47, с. 1515 (1964).
- [26] Khan F.S., Davies J.H., and Wilkins J.W. (1987). Quantum transport equations for high electric fields. *Physical Review B*, **36**, 2578.
- [27] Klimeck G., Lake R., Bowen R.C., Frenksley W.R., and Moise T.S. (1995). Quantum device simulation with a generalized tunneling formula. *Applied Physics Letters*, **67**, 2539.
- [28] Kouwenhoven L. and Glazman L. (2001). Revival of the Kondo effect. *Physics World*, January, p.33.
- [29] Lake R. and Datta S. (1992). Non-equilibrium Green's function method applied to double-barrier resonant-tunneling diodes. *Physical Review B*, **45**, 6670.
- [30] Lake R. and Datta S. (1992). Energy balance and heat exchange in mesoscopic systems. *Physical Review B*, **46**, 4757.
- [31] Lawaetz P. (1971). Valence band parameters in cubic semiconductors. *Physical Review B*, **4**, 3460.
- [32] Lindley D. (2001). *Boltzmann's Atom: The Great Debate that Launched a Revolution in Physics*. Free Press.
- [33] Lundstrom M.S. (1997). Elementary scattering theory of the MOSFET. *IEEE Electron Device Letters*, **18**, 361.

- [34] Lundstrom M.S. (2000). *Fundamentals of Carrier Transport*. Cambridge University Press.
- [35] Mahan G.D. (1987). Quantum transport equations for electric and magnetic fields. *Physics Reports*, **145**, 251.
- [36] Mann J.B. (1967). *Atomic Structure Calculations, 1: Hartree-Fock Energy Results for Elements Hydrogen to Lawrencium*. Distributed by Clearinghouse for Technical Information, Springfield, VA.
- [37] Martin P.C. and Schwinger J. (1959). Theory of many-particle systems. *Physical Review*, **115**, 1342.
- [38] McQuarrie D.A. (1976). *Statistical Mechanics*. Harper and Row.
- [39] Meir Y. and Wingreen S. (1992). Landauer formula for the current through an interacting electron region. *Physical Review Letters*, **68**, 2512.
- [40] Meir Y., Wingreen N.S., and Lee P.A. (1991). Transport through a strongly interacting electron system: theory of periodic conductance oscillations. *Physical Review Letters*, **66**, 3048.
- [41] Neofotistos G., Lake R., and Datta S. (1991). Inelastic scattering effects on single barrier tunneling. *Physical Review B*, **43**, 2242.
- [42] Paulsson M. and Datta S. (2003). Thermoelectric effects in molecular electronics. *Physical Review B*, **67**, 241403.
- [43] Rakshit T., Liang G.-C., Ghosh A., and Datta S. (2004). Silicon-based molecular electronics. *Nano Letters*, **4**, 1803.
- [44] Ramdas Ram-Mohan L. (2002). *Finite Element and Boundary Element Applications in Quantum Mechanics*. Oxford University Press.
- [45] Rammer J. and Smith H. (1986). Quantum field-theoretical methods in transport theory of metals. *Review of Modern Physics*, **58**, 323.
- [46] Sakurai J.J. (1967). *Advanced Quantum Mechanics*. Addison-Wesley.
- [47] Salis G., Fuchs D.T., Kikkawa J.M., Awschalom D.D., Ohno Y., and Ohno H. (2001). Optical manipulation of nuclear spin by a two-dimensional electron gas. *Physical Review Letters*, **86**, 2677.
- [48] Singh J. (1993). *Physics of Semiconductors and Their Heterostructures*. McGraw-Hill.
- [49] Venugopal R., Goasguen S., Datta S., and Lundstrom M.S. (2004). A quantum mechanical analysis of channel access, geometry and series resistance in nanoscale transistors. *Journal of Applied Physics*, **95**, 292.
- [50] Vogl P., Hjalmarson H.P., and Dow J. (1983). A semi-empirical tight-binding theory of the electronic structure of semiconductors. *Journal of Physics and Chemistry of Solids*, **44**, 365.
- [51] White S.R., Wilkins J.W., and Teter M.P. (1989). Finite-element method for electronic structure. *Physical Review B*, **39**, 5819.

- [52] Wolf E.I. (1989). *Principles of Electron Tunneling Spectroscopy*. Oxford Science Publications.
- [53] Zahid F., Paulsson M., and Datta S. (2003). Electrical conduction in molecules. In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Morkoc. Academic Press.

#### Литература, добавленная редактором перевода

- [54] Абрикосов А. А., Горьков Л. П., Дзялошинский И. Е. *Методы квантовой теории поля в статистической физике*, М.: Физматгиз, 1962.
- [55] Абрикосов А. А. *Основы теории металлов*, М.: Наука, 1987.
- [56] Берестецкий В. Б., Лифшиц Е. М., Питаевский Л. П. *Квантовая электродинамика (Теоретическая физика, том IV)*, М.: Наука, 1989.
- [57] Галицкий В. М., Карнаков Б. М., Коган В. И. *Сборник задач по квантовой механике*, М.: Наука, 2001.
- [58] Гантмахер В. Ф., Левинсон И. Б. *Рассеяние носителей тока в металлах и полупроводниках*, М.: Наука, 1984.
- [59] Гантмахер В. Ф. *Электроны в неупорядоченных средах*, М.: Физматлит, 2005.
- [60] Демиховский В. Я., Вугальтер Г. А. *Физика низкоразмерных квантовых структур*, М.: Логос, 2000.
- [61] Драгунов В. П., Неизвестный И. Г., Гридчин В. А. *Основы наноэлектроники*, Новосибирск: Издательство НГТУ, 2000.
- [62] Елютин П. В., Кривченков В. Д. *Квантовая механика с задачами*, М.: Физматлит, УНЦ довузовского образования МГУ, 2001.
- [63] Ландау Л. Д., Лифшиц Е. М. *Квантовая механика. Нерелятивистская теория (Теоретическая физика, том III)*, М.: Наука, 1989.
- [64] Castro Neto A. H., Guinea F., Peres N. M. R., Novoselov K. S., Geim A. K. *The electronic properties of graphene*, arXiv:0709.1163 (2007)
- [65] Geim A. K., Novoselov K. S. *The rise of graphene*, Nature Materials **6**, 183 (2007).

# Предметный указатель

- $sp^3s^*$ -модель, 170–175
- атомные
- переходы и дипольные антенны, 356
  - уровни энергии, 103, 106
- базисные функции, 127–154
- баллистика, 56, 57, 59, 63, 402, 411
- блокада Паули, 36, 311
- боровский радиус, 22, 68, 243
- валентные электроны, 103, 106
- внутризонные переходы, 358–361
- волновая функция, 72, 73, 236
- волновые функции атома водорода, 84–86
- вольтамперная характеристика, 27–42
- время
- возврата, 419
  - жизни, 276–284
- входящий ток, 36, 60, 63, 255–292, 312–315, 347–366, 381, 437–441, 453
- высокотемпературные сверхпроводники, 52
- выходящий ток, 36, 60, 63, 255–292, 312–315, 347–366, 381, 437–441, 453
- гамильтониан, 24, 74–75, 220–228, 297
- переноса, 335
- гелий, 91–92, 96–97
- граничные условия, 79
- графитовый лист, графен, 167–170
- двумерный проводник, 55
- дельта-символ Кронекера, 24, 73, 74, 143
- дельта-функция Дирака, 24
- деформационный потенциал, 343, 350
- джоулев нагрев, 57, 63, 396–398
- диаграмма энергетических уровней, 27, 97
- дисперсионное соотношение, 70
- диссипация, 57, 63, 396–398
- длина волны Де Бройля, 68
- емкость, 220–246
- закон
- Ома, 25, 27, 51, 56, 57, 63, 386–395
  - равновесия, 119
- зарядка, 27
- затвор, 26–42, 44
- золотое правило Ферми, 288
- зона Бриллюэна, 159, 164–167, 192
- зонд Бюттикера, 59, 321–322, 335, 340
- zonная
- структура, 156–182
  - структура, основной результат, 163

- излучение Черенкова, 360–361
- интенсивность (скорость) покидания, 34, 55, 255
- ионизация, 91–92, 98–101, 113
- источник, 26–42, 44, 53
- канал, 26, 27
- квазиклассический метод, 236
- квантовая
  - емкость, 221, 238–246
  - проводимость, 22, 27–48, 207
  - квантовые ямы, проволоки и точки, 187–190, 411
  - когерентное рассеяние, 57
  - когерентный транспорт, 390–394
  - колебания, 69–70, 113, 350, 367–376
  - контакт, 27, 28, 59, 282–289
  - корреляционная функция, 24, 59, 299, 381, 427–428
  - коэффициенты Эйнштейна, 345
  - кремний, 97
  - кулоновская блокада, 47–49, 51, 52, 99–101, 426, 451
  - линейный отклик, 42, 56
  - локализация, 56, 394
  - локальная плотность состояний, 264–274
  - лоренцева кривая, 39, 52
  - матрица плотности, 24, 59, 142–149, 220, 228–238, 266, 269, 298, 299, 304–311, 381, 425, 431–436
  - межзонные переходы, 356–357
  - метод
    - «от больших масштабов к малым», 25, 26
    - «от малых масштабов к большим», 25, 339
  - Каданоффа–Байма, см. неравновесная функция Грина
  - Ландауэра, основанная формула, 387
  - конечных разностей, 73–77, 325
  - многоэлектронная (многочастичная) картина, 48, 97–103, 114–123, 341, 346–347
  - многоэлектронное уравнение эволюции, 51, 114–123
  - модель Бора, 68
  - модель частицы в ящике, 69–70, 73, 77–86, 236
  - модели, 209–212, 217
  - нанотранзистор, 26, 27, 55, 402, 408, 411
  - нанотрубки, 190–198, 202–203
  - напряжение на затворе, 26, 27
  - напряжение стока, 26–27
  - насыщение тока, 56
  - нейтральный потенциал, 242
  - некогерентное рассеяние, 51–57, 60, 300, 381–383, 437–441, 453
  - некогерентный транспорт, 339–366
  - неравновесная функция Грина, 61, 419
  - обмен, 48
  - обратная решетка, 167
  - одноуровневая модель, 52–55, 299–300, 380–381
  - одноэлектронная (одночастичная) картина, 343, 350
  - одноэлектронная зарядка, 47–48, 51, 52, 99–101, 426, 451
  - оператор
    - рождения, 425–427
    - уничтожения, 425–427
    - числа частиц, 425–427

- оптические матричные элементы, 357–358
- оптическое испускание, 341
- ортогонализация, 137–138
- отрицательное дифференциальное сопротивление, 30, 64
- параметр Латинжера, 22, 246, 251
- периодическая таблица, 90, 106
- периодические граничные условия, 70–72, 79–82, 157–159, 188
- плотность состояний, 24, 39–49, 52, 53, 56, 59, 61, 198–207, 264
- подзоны, 187–217
- полевой
  - оператор, 425–427
  - транзистор, 26
- полупроводник
  - n-типа, 31
  - p-типа, 31
- поляризация оптического излучения, 357–358
- пороговое напряжение, 27
- постоянная Ридберга xiii, 68
- поток энергии, 398
- правило сумм, 39, 267–268, 311, 323–324
- преобразование Гильberta, 283
- приближение Хартри, 47, 101–102
- принцип
  - запрета, 31, 36, 311
  - неопределенности, 40, 255
- проводимость, 22, 25, 56
- пропускание, 24, 53, 300–304, 327, 383, 387–389, 408
- пространственно неоднородная эффективная масса, 225–226
- профиль потенциала, 400–409
- работа выхода, 30
- разбиение, 272
- размерное квантование, 236
- рассеяние, 56, 59, 408
  - внутрь структуры, 24, 381, 426, 436–437, 446–447
  - с нарушением фазы, 51–57, 60, 300, 381–383, 395–398, 411, 426
- расщепление электрохимического потенциала, 32–49
- резервуар, 27, 28, 59
- резонансное туннелирование, 335
- самовоздействие, 49
- самосогласованное поле, 24, 46–49, 51, 52, 59, 90–124, 383–385, 403–405, 450
- самосогласованный потенциал, 24, 46–49, 51, 52, 59, 90–124, 383–385, 403–405, 450
- связь (химическая), 104–114
- скорость электрона, 213–217
- собственная энергетическая функция, 24, 59, 256, 257, 259, 262–264, 271
- соотношение для энергии и времени, 431
- сопротивление, 22
  - контакта, минимальное, 207–212, 389
- сопротивление xiii, 25–63
- состояния симметричной и антисимметричной связи, 129–133
- спектральная функция, 24, 59, 266, 267, 299, 381, 427–428
- спин, 27, 48, 52
- спин-орбитальное взаимодействие, 175–182
- спиновые матрицы Паули, 177
- сток, 26–42, 44, 53

- схема решения уравнения Шредингера–Пуассона, 230
- теорема Блоха, 162
- теория
- возмущений, 149–154
  - функционала плотности, 102
- теормоэлектрический эффект, 63, 398–400
- ток
- в состоянии ВКЛ, 28, 402, 403, 405, 408–409
  - вероятности, плотность, 72–73
- трансформация базиса, 139–140
- туннелирование, 335, 411
- тяжелые и легкие дырки, 180–182
- уменьшение масштаба, 26
- унитарное преобразование, 140
- уравнение
- Больцмана, 416–419
  - Бюттикера, 319–322, 335, 340
  - Дирака, 183–185
  - Лапласа, 44
  - Пуассона, 44–47, 102, 221, 230, 235
  - Шредингера, 67–87, 139, 341
  - диффузии, 383
  - непрерывности, 73
  - эволюции, 51, 114–123
- уровень Ферми (энергия Ферми), 27
- уширение, 22, 24, 38–40, 48, 49, 59, 283, 297, 348, 381, 426, 437
- фононы, 69–70, 113, 350, 367–376
- формула для тока, 299, 315, 321, 336, 440
- фотоиспускание, 28
- функция
- Бозе–Эйнштейна, 345
  - Грина, 24, 262, 268–269, 299, 327, 381
  - Ферми, 24, 30, 53, 120, 297
  - — в двумерном случае, 231–233, 336
  - — в одномерном случае, 337
  - — в трехмерном случае, 238
  - теплового уширения, 242
- электрон-фононное взаимодействие, 69–70, 113, 350, 367–376
- электрон-фотонное взаимодействие, 350
- электрон-электронное взаимодействие, 47
- электронная плотность, 24, 61, 72, 92–93, 228–238, 299
- электронное сродство, 98–101
- электростатика, 27
- электрохимический потенциал, 24, 27–30, 42, 49, 53, 59
- энергия
- корреляции, 48, 102–103
  - связи, 113–114
- эрмитовский оператор, 140–142, 226, 257, 346
- эффект
- Кондо, 25, 51, 426, 450–551
  - Оверхаузера, 289
- эффективная масса, 22, 25, 222, 226, 246–251

*Суприйо Датта*

## КВАНТОВЫЙ ТРАНСПОРТ: ОТ АТОМА К ТРАНЗИСТОРУ

*Дизайнер Л. Н. Загуменова  
Технический редактор А. В. Широбоков  
Корректор Г. Г. Тетерина*

---

Подписано в печать 28.07.2009. Формат 60 × 84<sup>1</sup>/<sub>16</sub>.  
Печать офсетная. Усл. печ. л. 30,69. Уч. изд. л. 30,12.  
Гарнитура Таймс. Бумага офсетная №1. Заказ №38.

АНО «Ижевский институт компьютерных исследований»  
426034, г. Ижевск, ул. Университетская, 1.

<http://shop.rcd.ru> E-mail: [mail@rcd.ru](mailto:mail@rcd.ru) Тел./факс: (+73412) 500–295

---