



МИР материалов и технологий

ДЖ.М. МАРТИНЕС-ДУАРТ,
Р.ДЖ. МАРТИН-ПАЛМА,
Ф. АГУЛЛО-РУЕДА

Нанотехнологии
для микро- и
оптоэлектроники

Издание 2-е, дополненное

Перевод с английского
А.В. Хачояна
под ред. д.ф.-м.н.
Е.Б. Якимова

ТЕХНОСФЕРА

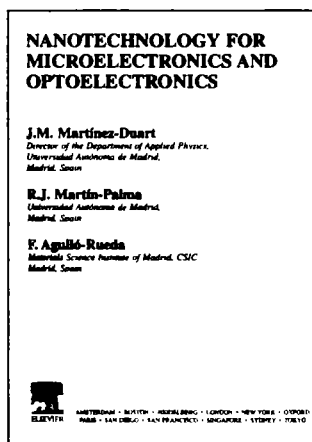
Москва

2009

Мартинес-Дуарт Дж.М., Мартин-Палма Р.Дж., Агулло-Руеда Ф.
Нанотехнологии для микро- и оптоэлектроники
Издание 2-е, дополненное
Москва:
Техносфера, 2009. –368с. ISBN 978-5-94836-209-0

В данной книге подробно описаны основные физические концепции, связанные с нанонаукой и нанотехнологиями, и возможности создания на их основе микроэлектронных и оптоэлектронных приборов нового поколения.

В настоящее время издается много книг по новейшим разделам нанонауки, но почти отсутствуют учебники и пособия для студентов-старшекурсников и аспирантов, связанных с нанонаукой. Предлагаемая книга восполняет этот пробел, так как представляет собой ценное учебное и справочное пособие для студентов, специализирующихся в физике, материаловедении и некоторых других технических дисциплинах. Кроме того, книга может представить интерес для ученых и инженеров-практиков, желающих глубже понять принципы нанонауки и нанотехнологии.



© Elsevier Ltd, The Boulevard, Lanford Lane,
Kidlington, OX5 1GB, England

© 2007, ЗАО «РИЦ «Техносфера», перевод на русский язык

© 2009, ЗАО «РИЦ «Техносфера», оригинал-макет, оформление

© 2009, ОАО «НИЦПВ», дополнение

ISBN 978-5-94836-209-0

ISBN 0-080-44553-5 (англ.)

СОДЕРЖАНИЕ

Предисловие	11
Структура книги	15
Глава 1. Мезоскопическая физика и нанотехнологии	17
1.1. Содержание книги.....	17
1.2. Основные тенденции развития нано- и оптоэлектроники.....	18
1.3. Характеристические длины в мезоскопических системах.....	26
1.4. Квантово-механическая когерентность.....	31
1.5. Квантовые ямы, проволоки и точки.....	32
1.6. Плотность состояний и размерность системы.....	33
1.7. Полупроводниковые гетероструктуры.....	36
1.8. Квантовые процессы переноса.....	38
Литература.....	39
Дополнительная литература.....	39
Задачи и упражнения.....	40
Глава 2. Введение в физику твердого тела	43
2.1. Введение.....	43
2.2. Краткие сведения из квантовой механики.....	45
2.2.1. Корпускулярно-волновой дуализм и принцип Гейзенберга.....	45
2.2.2. Уравнение Шрёдингера.....	47
2.2.3. Распределения Ферми — Дирака и Бозе — Эйнштейна.....	50
2.2.4. Методы теории возмущений.....	52

2.3. Модель свободных электронов в твердых телах. Функция плотности состояний	54
2.4. Теорема Блоха	58
2.5. Электроны в кристаллических твердых телах	60
2.5.1. Модель почти свободных электронов	60
2.5.2. Приближение сильной связи	62
2.6. Динамика электронов в энергетических зонах	65
2.6.1. Уравнение движения	65
2.6.2. Эффективная масса	67
2.6.3. Дырки	70
2.7. Колебания решетки	71
2.7.1. Одномерная решетка	72
2.7.2. Трехмерная решетка	77
2.8. Фононы	78
Литература	79
Дополнительная литература	80
Задачи и упражнения	80
Глава 3. Общие сведения из физики полупроводников	83
3.1. Введение	83
3.2. Энергетические зоны электронов в типичных полупроводниках	84
3.3. Собственные и примесные полупроводники	87
3.4. Концентрации электронов и дырок в полупроводниках	92
3.5. Элементарные процессы переноса в полупроводниках	98
3.5.1. Движение носителей заряда в электрическом поле. Подвижность зарядов	99
3.5.2. Диффузионная проводимость	101
3.5.3. Уравнения непрерывности. Время жизни носителей и длина диффузии	102
3.6. Вырожденные полупроводники	107
3.7. Оптические свойства полупроводников	108
3.7.1. Оптические процессы в полупроводниках	108

3.7.2. Межзонное поглощение	110
3.7.3. Экситонные эффекты.....	113
3.7.4. Спектр излучения.....	117
3.7.5. Стимулированное излучение	119
Литература	122
Дополнительная литература.....	122
Задачи и упражнения.....	123
Глава 4. Физика полупроводников с пониженной размерностью.....	126
4.1. Введение	126
4.2. Основные характеристики двумерных полупроводниковых наноструктур	127
4.3. Прямоугольная потенциальная яма конечной глубины.....	134
4.4. Параболическая и треугольная квантовые ямы.....	136
4.4.1. Параболическая потенциальная яма.....	136
4.4.2. Треугольная потенциальная яма	137
4.5. Квантовые проволоки.....	139
4.6. Квантовые точки.....	142
4.7. Напряженные слои.....	144
4.8. Влияние напряжений на валентную зону	146
4.9. Зонная структура в квантовых ямах	150
4.10. Экситонные эффекты в квантовых ямах	152
Литература	156
Дополнительная литература.....	156
Задачи и упражнения.....	157
Глава 5. Полупроводниковые квантовые наноструктуры и сверхрешетки.....	160
5.1. Введение	160
5.2. Структуры полевых МОП-транзисторов (MOSFET)	161
5.3. Гетеропереходы	166
5.3.1. Гетеропереходы с модулированным легированием	166

5.3.2. Напряженные гетероструктуры на основе SiGe.....	170
5.4. Квантовые ямы	172
5.4.1. Модулированно-легированные квантовые ямы.....	172
5.4.2. Множественные квантовые ямы (MQW).....	174
5.5. Сверхрешетки.....	177
5.5.1. Концепция сверхрешеток	177
5.5.2. Модель сверхрешетки Кронига — Пенни. Расщепление зон.....	178
5.5.3. Приближение сильной связи в теории сверхрешеток.....	184
5.5.4. Сверхрешетки типа <i>nipi</i>	187
Литература.....	189
Дополнительная литература	190
Задачи и упражнения.....	190

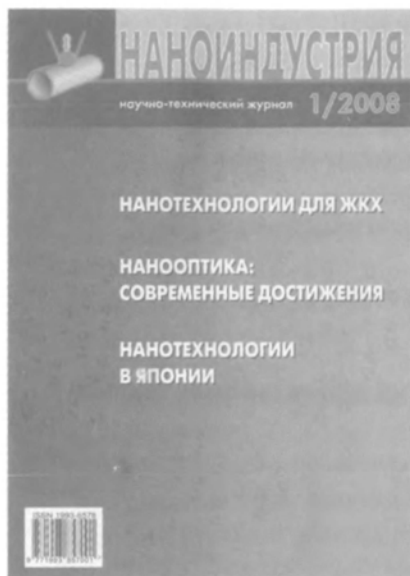
Глава 6. Процессы переноса в наноструктурах

в электрических полях.....	192
6.1. Введение	192
6.2. Продольный перенос	193
6.2.1. Механизмы рассеяния электронов	193
6.2.2. Экспериментальные данные по продольному переносу	197
6.2.3. Продольный перенос горячих электронов.....	200
6.3. Поперечный перенос.....	203
6.3.1. Резонансное туннелирование	204
6.3.2. Влияние поперечных электрических полей на свойства сверхрешеток	206
6.4. Квантовый перенос в наноструктурах	212
6.4.1. Квантовая проводимость. Формула Ландауэра	213
6.4.2. Формула Ландауэра — Бюттикера для квантового переноса в многозондовых структурах	218
6.4.3. Кулоновская блокада	220

Литература.....	224
Дополнительная литература.....	225
Задачи и упражнения.....	225
Глава 7. Перенос в магнитных полях и квантовый эффект Холла.....	229
7.1. Введение.....	229
7.2. Воздействие магнитного поля на кристаллы.....	231
7.3. Поведение систем пониженной размерности в магнитных полях.....	233
7.4. Плотность состояний двумерных систем в магнитных полях.....	234
7.5. Эффект Аронова — Боме.....	236
7.6. Эффект Шубникова — де Гааза.....	239
7.7. Квантовый эффект Холла.....	242
7.7.1. Экспериментальные данные и элементарная теория целочисленного квантового эффекта Холла (IQHE).....	242
7.7.2. Краевые состояния и IQHE.....	244
7.7.3. Протяженные и локализованные состояния.....	247
7.7.4. Использование квантового эффекта Холла (IQHE) в метрологии.....	249
7.7.5. Дробный квантовый эффект Холла (FQHE).....	251
Литература.....	254
Дополнительная литература.....	255
Задачи и упражнения.....	255
Глава 8. Оптические и электрооптические процессы в квантовых гетероструктурах.....	258
8.1. Введение.....	258
8.2. Оптические свойства квантовых ям и сверхрешеток.....	259
8.3. Оптические характеристики квантовых точек и нанокристаллов.....	265

8.3.1. Методы выращивания кристаллов. Самоорганизация квантовых точек	265
8.3.2. Оптические свойства	268
8.4. Электрооптические эффекты в квантовых точках. Эффект квантово-размерный Штарка	277
8.5. Электрооптические эффекты в сверхрешетках. Лестницы Штарка и осцилляции Блоха	282
Литература	288
Дополнительная литература	289
Задачи и упражнения	289
Глава 9. Электронные приборы на наноструктурах	292
9.1. Введение	292
9.2. Модуляционно-легированные полевые транзисторы (MODFET)	295
9.3. Биполярные транзисторы на гетеропереходах	298
9.4. Резонансный туннельный эффект	302
9.5. Транзисторы на горячих электронах	307
9.6. Транзисторы с резонансным туннелированием	311
9.7. Одноэлектронные транзисторы	314
Литература	319
Дополнительная литература	319
Задачи и упражнения	319
Глава 10. Оптоэлектронные устройства на основе наноструктур	322
10.1. Введение	322
10.2. Лазеры на полупроводниковых гетероструктурах	323
10.3. Лазеры на полупроводниковых квантовых ямах	327
10.4. Поверхностные лазеры с вертикальным резонатором (VCSEL)	332
10.5. Лазеры на напряженных структурах с квантовыми ямами	335
10.6. Лазеры на квантовых точках	338

10.7. Фотодетекторы на квантовых ямах и сверхрешетках	343
10.7.1. Фотодетекторы на подзонах квантовых ям	343
10.7.2. Лавинные фотодетекторы на сверхрешетках	345
10.8. Модуляторы на квантовых ямах	346
Литература	349
Дополнительная литература	350
Задачи и упражнения	350
Дополнение. Метрологический и стандартизационный базис нанотехнологий	353



Научно-технический журнал
«Наноиндустрия»
издается с февраля 2007 года

Выходит
1 раз в 2 месяца

Выпускается при содействии
Министерства промышленности
и торговли
Российской Федерации

Издатель журнала:
ЗАО РИЦ «Техносфера»
<http://www.technosphera.ru>

Главный редактор:
П.П. Мальцев

<http://www.nanoindustry.su>

Журнал «Наноиндустрия» поддерживает и популяризует оригинальные работы отечественных и русскоязычных зарубежных специалистов, знакомит читателей с перспективами развития новых направлений нанотехнологий и наноматериалов, освещает вопросы производства, экономики и бизнеса в области наноиндустрии.

Подписка

- по каталогу «Газеты и журналы» агентства «Роспечать» – 80939 (полугодовой индекс)
- в ООО «Вся пресса» по тел.: (495) 787-34-49
- в редакции журнала по тел.: (495) 234-01-10

Предисловие

Авторы ставят своей целью описание основных физических концепций, связанных с нанонаукой и нанотехнологией, и возможностей их применения для создания полупроводниковых материалов и приборов. Нанотехнология основана на том, что в очень небольших по размеру твердых телах и структурах (размеры которых сопоставимы с характеристическими для данного материала длинами, например, с длиной волны де Бройля, длиной когерентности, длиной локализации и т. д.) начинают проявляться новые физические свойства, обусловленные квантовыми эффектами. К таким эффектам можно отнести квантовые осцилляции проводимости, квантовый эффект Холла, резонансное туннелирование, одноэлектронный перенос и другие явления, которые могут проявляться и использоваться в специально создаваемых объемных наносистемах (например, в полупроводниковых гетеропереходах, квантовых потенциальных ямах, сверхрешетках и т. п.). Многие из этих эффектов подробно описаны в книге. Явления в твердотельных квантовых структурах очень интересны не только с точки зрения фундаментальной науки (свидетельством чего служат Нобелевские премии, регулярно присуждаемые за такие исследования в последние десятилетия), но они имеют также и огромное практическое значение, поскольку на их основе создаются новые поколения микроэлектронных и оптоэлектронных приборов.

Прошло лишь около тридцати лет с тех пор, как в начале 70-х годов группа исследователей из фирмы ИБМ Эсаки, Тцу и Чанг (Esaki, Tsu, Chang) получила Нобелевскую премию за открытие новых эффектов в квантовых ямах и сверхрешетках. Экспериментальное наблюдение таких эффектов оказалось возможным только благодаря тому, что были развиты достаточно сложные технологические методы, такие, как молекулярно-лучевая эпитаксия, позволяющая наращивать кристаллы «послойно», способы легирования полупроводниковых наноструктур и др. При этом разработка новой аппаратуры и новых методик зачастую происходила настолько быстро, что их описание не удавалось даже ввести соответствующим

образом в лекционные курсы по экспериментальной физике. Однако в самые последние годы ряд университетов приступил к значительному расширению своих учебных планов и программ для студентов всех уровней, разрабатывая специальные курсы по нанотехнологиям, приборам с размерами в нанометровом диапазоне, квантовым приборам и наноструктурам. В настоящее время вводится ученая степень магистра для специальностей, связанных с нанонаукой. Новые учебные курсы, относящиеся к нанотехнологиям, читаются во многих учебных заведениях, готовящих физиков, материаловедов и инженеров.

В настоящее время ощущается явная нехватка учебников по нанонауке и нанотехнологии общего плана для студентов старших курсов, желающих получить представление о современном состоянии исследований в данной области. Возникла практика включения сведений о нанотехнологиях в виде дополнений к уже существующим курсам или учебникам физики твердого тела, при этом новые разделы (по нанотехнологии) просто добавляются к известным учебникам при переиздании, практически не связывая нанонауку с остальным содержанием книги. Несколько лучше обстоит дело с книгами узкой специализации, так как за последние пятнадцать лет было издано много прекрасных монографий, посвященных отдельным разделам нанотехнологии и физики полупроводниковых объектов с пониженной размерностью. Перечни таких книг читатель найдет в списках «Дополнительной литературы» в конце соответствующих глав.

Предлагаемая книга представляет собой учебный курс (рассчитанный на один семестр), предназначенный в первую очередь для студентов-старшекурсников и аспирантов, специализирующихся в области физики, материаловедения и некоторых технических дисциплин (электроника и др.). Кроме того, книга может представить интерес для ученых и инженеров-практиков, желающих глубже понять принципы нанонауки и нанотехнологии. Авторы ставили своей целью создание вводного курса, начинающегося с общих сведений о полупроводниках с пониженной размерностью и квантовых гетеропереходах. Ясное представление о физической сущности происходящих в таких структурах процессов должно позволить студентам понять принципы квантования энергии

электронов в потенциальных ямах и наноструктурах с пониженной размерностью (двумерных, одномерных и нульмерных), а затем и причины возникновения новых электронных, транспортных и оптических свойств в твердых телах. Для понимания физических основ описываемых явлений студенты должны знать квантовую механику и статистическую физику лишь в объеме общего курса физики для университетов. Авторы надеются, что такой подход позволит читателям легче воспринять общие идеи и закономерности нанонауки (некоторые сложности, возможно, будут возникать лишь при рассмотрении квантового эффекта Холла), а также связать их со свойствами объемных твердых тел и полупроводников. Авторы считают, что введение некой общей схемы строения наноструктур (главы 1—3) позволит читателю оценить возможности их технологического использования в микроэлектронике и оптоэлектронике.

Помимо сказанного авторы в процессе написания книги старались выполнить следующие требования и условия:

1. Объем книги ограничен упомянутым условием изучения учебного курса в течение одного семестра (следует учесть, что студенты, уже имеющие достаточную подготовку в физике твердого тела и теории полупроводников, могут пропустить материал глав 2 и 3). Следствием такого ограничения стало исключение из курса данных по некоторым разделам нанотехнологии (например, по углеродным нанотрубкам и бимолекулярным структурам).

2. Авторы старались вводить новые физические представления с достаточной научной строгостью, но использовать лишь наиболее простые математические описания. При сложных расчетах мы иногда просто приводим конечный результат вычислений (давая, естественно, соответствующие ссылки) или ссылаемся на известные аналогичные формулы из курса физики твердого тела (например, при использовании так называемого «золотого правила» Ферми).

3. Книга представляет собой вводный учебный курс, вследствие чего литература в конце глав включает лишь необходимый минимум ссылок. Во всех случаях мы старались включить в этот список труды тех авторов, которые были первооткрывателями новых эффектов или первыми сформулировали описывающие эти эффекты теории.

4. Для лучшего понимания новых идей и представлений, в конце каждой главы прилагается небольшой набор задач и упражнений. Они не являются слишком сложными, и их решение не требует сложных теоретических построений (во многих случаях текст содержит «намёки» и подсказки, помогающие решению). Мы считаем важным, чтобы студенты научились получать точные численные результаты, что позволит им лучше представить приблизительные значения физических величин и понять значимость приближенных вычислений в нанотехнологических исследованиях.

Структура книги

Книга состоит из десяти глав, которые концептуально разделены на четыре части. Первая часть (главы 1—3) начинается с описания современного состояния исследований в области микроэлектроники и оптоэлектроники (глава 1). Читателю предлагаются общие сведения о физической картине явлений в полупроводниках с низкой размерностью и различные определения, относящиеся к наноструктурам. Главы 2 и 3 содержат основы физики твердого тела и физики полупроводников. Эти две главы могут быть полезны студентам нефизических специальностей, которые не изучали указанные предметы достаточно глубоко и серьезно. Кроме того, материал этих глав позволяет легче понять общие основы нанотехнологических процессов.

Вторая часть книги (главы 4 и 5) посвящена физическим механизмам, определяющим поведение электронов в наноструктурах и их связи с функцией плотности состояний и квантованием энергии электронов в различных потенциальных ямах. Описываются самые распространенные квантовые наноструктуры, имеющие наиболее важное значение в исследованиях и практических приложениях (в частности, полупроводниковые гетеропереходы, квантовые ямы, сверхрешетки и т. п.).

В третьей части книги (главы 6—8) рассмотрены оптические свойства наноструктур явления переноса в таких структурах. В главе 6 описаны процессы переноса в электрическом поле, когда интересующие нас материалы демонстрируют некоторые необычные и очень важные особенности (например, их проводимость приобретает квантовый характер, наблюдается так называемая кулоновская блокада, резонансное туннелирование и т. д.). При описанном в главе 7 одновременном воздействии электрических и магнитных полей наблюдается целочисленный и дробный квантовый эффект Холла. Пока мы не имеем полного теоретического понимания физической сущности этих эффектов, и они продолжают оставаться объектом самого пристального изучения, так как открытие двух этих явлений можно считать наиболее важным событием в физике твердого тела за последние десятилетия. Описываемые в главе 8 оптические свойства наноструктур (подобно упомянутым ранее процессам

переноса) также являются весьма необычными. Они поразительно отличаются от привычных свойств объемных полупроводниковых материалов (в качестве наглядных примеров можно упомянуть возможность «настройки» запрещенной зоны, микроволновое излучение Блоха (блоховская эмиссия), зависимость оптических характеристик от размеров и т. п.).

Материал этих глав подготавливает студентов к последнему разделу (главы 9 и 10), посвященному новейшим разработкам в области полупроводниковых приборов на наноструктурах. В главе 9 описаны одноэлектронные и высокочастотные приборы (транзисторы с высокой подвижностью электронов, приборы с резонансным туннелированием и т. д.). В главе 10 описываются оптоэлектронные и фотонные приборы с квантовыми гетероструктурами (лазеры на квантовых ямах и квантовых точках, фотодетекторы с использованием кристаллических сверхрешеток, высокоскоростные оптические модуляторы и т. д.). Изучение этих глав позволит студентам понять, каким образом можно создавать устройства с заранее запланированными электронными или оптическими характеристиками (такая деятельность уже получила название «конструирование (инженерия) зонных структур». Кроме того, авторы старались в двух последних главах внушить студентам представление о нанотехнологии как об уже существующей и активно развивающейся области науки и техники. Действительно, очень многие из описываемых приборов (например, транзисторы MODFET, биполярные транзисторы на гетеропереходах, лазеры на квантовых ямах, фотодетекторы на сверхрешетках, высокоскоростные оптические модуляторы и т. п.) уже перестали быть только объектом научного исследования. Эти физические структуры уже стали коммерческими продуктами, находящими все более широкое практическое применение в самых разнообразных электронных и оптоэлектронных приборах.

ГЛАВА I

МЕЗОСКОПИЧЕСКАЯ ФИЗИКА И НАНОТЕХНОЛОГИИ

1.1. Содержание книги

Постоянно возрастающий интерес к исследованию физических свойств полупроводниковых объектов очень небольшого размера (прежде всего, в нанометровом диапазоне) обусловлен несколькими факторами, главным из которых является общая тенденция к использованию микроэлектронных интегральных схем, в которых малые размеры сочетаются с повышенными рабочими частотами, улучшенными характеристиками и низкой стоимостью производства. Дополнительным фактором (возможно, более значимым с научной точки зрения) выступает то, что работа таких приборов основана на фундаментально новых квантовых физических эффектах (таких, как резонансное туннелирование, квантование проводимости, кулоновская блокада, квантовый эффект Холла и др.). Кроме того, очень часто новые вещества и характеристики оказываются пригодными к непосредственному использованию в новых приборах, имеющих практическое и коммерческое применение (лазеры на квантовых ямах, одноэлектронные транзисторы, оптические модуляторы на квантовом эффекте Штарка в ограниченных по размерам (confined quantum Stark effect) системах и др.

Первая глава предлагаемой книги начинается с описания современных микроэлектронных и оптоэлектронных структур, ставших основой новейших направлений нанотехнологии. После этого вводится некоторый набор параметров и концепций так называемой мезоскопической физики (прежде всего, речь идет о характеристических длинах), необходимых для описа-

ния наноструктур. Затем рассматриваются физические основы существования и создания полупроводниковых объектов с «пониженной размерностью»: двумерных (2D) квантовых ям, одномерных (1D) квантовых проволок и нульмерных квантовых точек (0D). Конечно, некоторые из вводимых представлений могут показаться читателю непонятными или сложными, особенно потому, что их сложно представить наглядно. Однако авторы считают, что именно введение этих представлений с самого начала позволит читателю легче понять их роль в описываемых позднее процессах.

Главы 2 и 3 содержат обзор наиболее общих положений и самые необходимые сведения из квантовой механики, физики твердого тела и физики полупроводников. В главе 4 рассмотрены вопросы физики полупроводников с низкой размерностью (т. е. упомянутых выше квантовых ям, проволок и точек. В главе 5 обсуждаются некоторые наиболее часто используемые гетероструктуры и сверхрешетки, а в главах 6 и 7 рассмотрено воздействие электрических и магнитных полей на такие наноструктуры, особенно в связи с эффектами квантования проводимости материалов и эффекта Холла. Многообразие оптических явлений в полупроводниковых наноструктурах описано в главе 8, которая завершает описание основных физических свойств мезоскопических систем. Последние две главы книги непосредственно относятся к новейшим электронным и оптоэлектронным полупроводниковым приборам. В главе 9 описаны некоторые типы высокочастотных диодов и транзисторов, работа которых основана на использовании одноэлектронных переходов и резонансного туннелирования. В 10-й главе читатель найдет описание различных оптоэлектронных и фотонных приборов на квантовых гетероструктурах (лазеры на квантовых ямах, фотодетекторы, оптические модуляторы).

1.2. Основные тенденции развития нано- и оптоэлектроники

Эволюция микроэлектронных приборов определяется многими факторами, но важнейшими из них выступают требования к постоянному росту объема памяти интегральных схем и скорости передачи информации, повышение эффективности оптической связи и др. Этим требованиям удовлетворяют электрон-

ные приборы с повышенным быстродействием и уменьшенными размерами, как в случае кремниевых интегральных схем, в частности, динамических ОЗУ (оперативное запоминающее устройство, DRAM). На рис. 1.1 показано уменьшение критических размеров распространенных МОП-транзисторов (и соответственно, плотности записи на них, измеряемой в числе битов на чип) за период с 1970 до 2000 годов, а также предполагаемая экспертами эволюция таких схем на ближайшие двадцать лет. Например, в настоящее время характерные размеры элементов L стандартных DRAM-чипов на 256 Мбит (содержащих в себе около 10^9 транзисторов) составляют примерно 100 нм. В структурах с такими размерами описание процессов переноса носителей заряда ещё может рассматриваться в рамках классической теории, однако это рассмотрение уже явно находится на самой «границе» проявления квантовых эффектов. Поэтому исследователи должны вводить какие-то новые представления о квантовых процессах переноса (quantum transport), что и обсуждается в разделе 1.8. В настоящее время предполагается, что современная кремниевая технология будет использоваться до размеров элементов $L \sim 10$ нм, но ниже этого предела должны быть созданы транзисторы, основанные на новых принципах (одноэлектронные

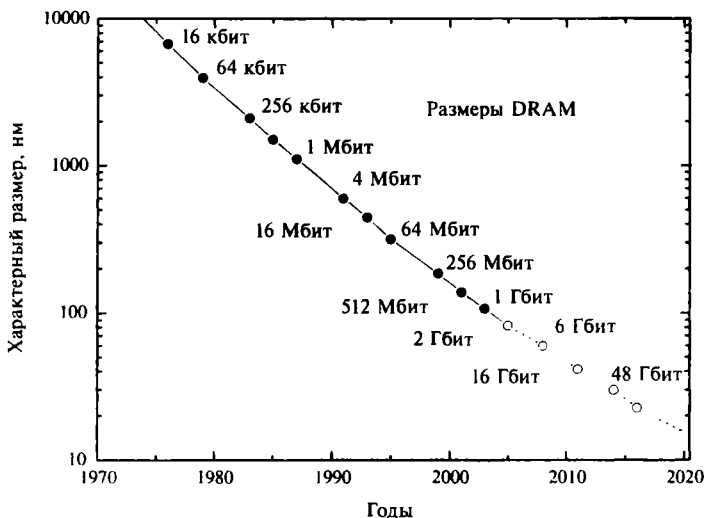


Рис. 1.1. Изменение минимальных размеров элементов кремниевых чипов типа DRAM (по данным работы [1])

транзисторы, приборы с резонансным туннелированием и т. п.). Новейшие приборы такого типа (и их теория, основанная на принципах мезоскопической и квантовой физики) рассматриваются в главе 9. Интересно отметить, что квантовые эффекты в полупроводниках $A_{III}B_V$, т. е. полупроводниковых соединениях, состоящих из элементов III и V групп, иногда проявляются и при больших размерах структур вследствие уменьшения эффективной массы электронов и соответствующего увеличения длины волны де Бройля (раздел 1.3).

Современная промышленность, средства связи, коммуникационные информации, военная и космическая техника постоянно совершенствуются, поэтому можно с уверенностью предсказать, что в близком будущем микроэлектронику в науке и промышленности сменит нанoeлектроника, с характерными размерами элементов электронных приборов около 10 нм. Несмотря на поразительные достижения в области изготовления полупроводниковых приборов и развития технологии в течение последнего десятилетия, прогресс в развитии новых систем архитектуры, позволяющих объединять работу миллиардов транзисторов, пока не соответствует развитию технологии. В этой связи в настоящее время разрабатываются новые архитектурные решения для параллельной обработки данных и т. п.

Значительный прогресс в развитии нанoeлектроники оказался возможным, прежде всего, из-за развития техники осаждения очень тонких пленок, образующих гетероструктуры, в которых электроны как бы оказываются «заключенными» (локализованными) внутри двумерных (2D) мезоскопических систем. Существовавшие ранее стандартные методы получения тонких пленок (осаждение, напыление) не позволяли формировать гетероструктуры требуемого качества. Однако уже в 80-е годы (см. раздел 1.7) были разработаны новые технологии, такие, как молекулярно-пучковая эпитаксия МПЭ (molecular beam epitaxy, MBE) и газофазная эпитаксия на основе металлоорганических соединений МОС ГФЭ (metal organic chemical vapour deposition, MOCVD). В методе МПЭ осуществляется эпитаксиальное выращивание пленок требуемого соединения на подложках в условиях сверхвысокого вакуума (при давлении менее 10^{-10} мбар). Требуемое вещество испаряется в цилиндрической эффузионной ячейке Кнудсена при достаточно высоких температурах. Вылетающие из ячейки атомы или молеку-

лы формируют молекулярный пучок,двигающийся по прямым линиям к подложке, на которой они конденсируются.

Метод МПЭ позволяет не только получать сверхтонкие пленки одноэлементных полупроводников (Si, Ge) и соединений $A_{III}B_V$ (Al—Ga, AlGaAs, InP) или $A_{II}B_{VI}$ (CdTe, PbS), но и послойно выращивать пленки и сверхрешетки, а также легировать их атомами заданного типа (B, Al, As и др.). Этот метод позволяет получать высококачественные пленки, он является дорогим и малопроизводительным, и с промышленной точки зрения для получения полупроводниковых гетероструктур более привлекательным является метод MOCVD. Последний метод позволяет выращивать слой арсенида GaAs на подложке при температуре около 500°C за счет химической реакции газообразного арсина AsH_3 с триметил-галлием $(\text{CH}_3)_3\text{Ga}$ при пониженных давлениях. В некоторых случаях осаждением из газовой фазы удается вводить в полупроводники легирующие добавки, например используя SiH_4 (растворенный в газообразном водороде, выступающем в качестве носителя), в GaAs можно ввести донорную примесь Si. Описываемая техника позволяет осуществлять осаждение на несколько пластин одновременно, достаточно легко наносить покрытие на большое количество пластин, что уже используется, например, в коммерческом производстве полупроводниковых лазеров.

Вообще говоря, мезоскопические системы требуют формирования наноструктур размером около 100 нм, что требует снижения существующих промышленных стандартов примерно на порядок (исходя из состояния технологий за последние двадцать лет). В настоящее время производственники уже практически исчерпали все возможности обычной оптической литографии, т. е. вышли на ее «физические пределы», в результате чего сейчас все шире изучаются возможности использования в промышленности новой нанолитографической технологии высокого разрешения (основанной, например, на электронных или ионных пучках, рентгеновском излучении и др.). В нанолитографии все большее распространение получает позитивный резист PMMA (полиметилметакрилат), шероховатость поверхности которого при осаждении с вращением не превышает 2 нм, несмотря на достаточно высокую молекулярную массу ($\sim 10^6$).

В литературе часто обсуждаются пределы размеров приборов в свете их эволюции, представленной на рис. 1.1. Разумно пред-

положить, что скорость уменьшения этих размеров, определяемая известным законом Мура [2], должна как-то замедлиться в ближайшие годы, поскольку все физические пределы (с точки зрения промышленного и экономического развития) будут достигнуты в следующее десятилетие. Технологические пределы определяются несколькими факторами, из которых мы отметим только два. Во-первых, в любом случае следует учитывать тепловыделение работающего электронного прибора, которое не может быть сведено к нулю, поскольку тепловой баланс системы ограничивается соответствующими коэффициентами теплопроводности используемых материалов и числом молекулярных слоев. Во-вторых, существует так называемый фактор «разброса параметров» в производственном процессе. Например, многие электрические параметры МОП-транзисторов контролируются легированием, однако при очень малых размерах легированной области (около $0,1 \text{ мкм}^3$) число легирующих атомов является столь малым (около 10), что точное управление разбросом этого параметра становится невозможным.

Помимо указанных технологических пределов существуют и более серьезные ограничения, связанные с фундаментальными законами природы, которые могут быть названы физическими пределами. Хотя существующие устройства еще не вышли на эти пределы, имеет смысл перечислить их.

1) Тепловой предел. Энергия, необходимая для записи одного бита информации должна составлять несколько kT , средней энергией тепловых флуктуаций. Например, в КМОП-транзисторах низшее значение энергии записи одного бита должно быть не ниже 2 эВ, т. е. примерно $100 kT$ при комнатной температуре или $\sim 3 \times 10^{-19}$ Дж.

2) Релятивистский предел. Очевидно, что скорость распространения сигнала не может превышать скорости света. Предполагая, что микропроцессор имеет размеры в несколько сантиметров, легко вычислить, что время прохождения сигнала в нем будет составлять 10^{-10} с, чему соответствует частота 10 ГГц.

3) Принцип неопределенности. В соответствии с принципом неопределенности Гейзенберга энергия и время процесса записи или считывания бита информации связаны между собой соотношением $\Delta E \Delta t \geq h$. Для надежной работы системы мы можем, например, потребовать, чтобы произведение $\Delta E \cdot \Delta t$ составляло $100 h$, вследствие чего для энергии ΔE порядка 10^{-19} Дж в будущей цепи с ростом частоты будет легко достигаться квантовый предел.

Современная наноэлектроника развивается одновременно в нескольких направлениях, одним из которых выступает наноэлектроника твердого тела, которой и посвящена большая часть предлагаемой читателю книги. Основой такой электроники выступают гетероструктуры хорошо изученных материалов (Si , SiO_2 , соединения $\text{A}_{\text{III}}\text{B}_{\text{V}}$), а также некоторые типы транзисторов: с гетеропереходами, одноэлектронные, с резонансным туннелированием, баллистические и др. Однако некоторые из задач современной электроники (например, распознавание речи и визуальных образов) требуют столь больших вычислительных мощностей, что исследователи занимаются интенсивным поиском радикально новых приборов и материалов. Некоторые из альтернативных решений, такие, например, как сверхпроводящая электроника и спинтроника, могут использовать технологию изготовления элементов, принципиально не отличающуюся от используемых в настоящее время промышленностью при изготовлении интегральных схем. Сверхпроводящая электроника, идеи которой восходят к началу 70-х годов (а первые прототипы получены в 80-х), основана на сверхпроводниковых переходах джозефсоновского типа, представляющих собой два сверхпроводящих слоя, разделенных очень тонким слоем окисной изолирующей пленки, через которую может осуществляться туннелирование сверхпроводящей электронной пары. Преимущества электронных устройств такого типа основаны на том, что джозефсоновские переходы позволяют работать при очень высоких скоростях (время переключения от 1 до 10 пикосек), из-за чего рассеиваемая энергия очень мала, и электрическим сопротивлением всех соединительных линий сверхпроводящей схемы можно практически пренебречь.

В качестве другой альтернативной технологии предлагается так называемая спинтроника, в которой для записи информации используется ориентация спина электрона. Так называемые электрон-спиновые (спинтронные, спиновые) транзисторы представляют собой трехслойную (сандвич) структуру, в которой слой полупроводника (база) заключен между двумя слоями ферромагнетика (эмиттер и коллектор). Электроны, магнитное состояние которых определяется эмиттером, поступают в коллектор через базу только в том случае, если их спины параллельны направлению намагниченности коллектора. Такие разработки проводятся параллельно с работами в магнетоэлектронике для создания магнитных оперативных запоминающих

устройств (МОЗУ, MRAM), основанных на эффекте гигантского магнетосопротивления и на магнитных туннельных переходах, полученных в 1995 г. Электрон-спиновые транзисторы могут иметь огромные перспективы, если исследователям удастся интегрировать их в широко распространенные КМОП-схемы.

Существуют и другие радикальные альтернативы развития нанoeлектроники, из которых следует упомянуть, прежде всего, молекулярную электронику, основанную на использовании различных состояний или конфигураций молекул (например, на использовании *цис*- и *транс* форм или параллельном либо антипараллельном упорядочении спинов неспаренных электронов). При этом изменение состояний может быть быстрым и не требовать значительных расходов энергии, а также стимулироваться внешними сигналами, регистрироваться зондами и т. п. Если основанную на таких принципах электронику действительно удастся создать, то это приведет к удивительному прогрессу в миниатюризации, поскольку характерные размеры молекул на порядки меньше, чем размеры элементов существующих электронных схем. Огромным преимуществом молекул выступает присущая им способность к самоорганизации в трехмерные супрамолекулярные структуры, не говоря уже о том, что развитие сканирующей атомно-силовой микроскопии позволяет манипулировать отдельными молекулами. В качестве примера объектов молекулярной электроники можно указать молекулы азобензола, гидразобензола и т. д. Большим преимуществом органических соединений (по сравнению с неорганическими) является то, что их значительно легче изолировать, т. е. получать и исследовать в виде отдельных молекул. Возможно, что в будущем молекулы удастся соединять при помощи молекулярных проволок или нанотрубок, однако в настоящее время внимание исследователей сосредоточено на металлических и полупроводниковых контактах. Даже с учетом этих возможностей одной из основных проблем молекулярной электроники является задача обеспечения интерфейса или создания коммуникационных связей «вычислительных молекул» с внешним миром, т. е. с гораздо более крупными системами.

Заканчивая перечисление возможностей, следует упомянуть и о так называемой биоэлектронике, принципы которой связаны с функционированием биологических систем. В попытках копирования природных процессов мы не всегда должны осно-

вываться только на соображениях размера. Например, нейроны нервной системы слишком велики (!) для стандартов нанотехнологии, но наноэлектроника может извлечь много полезной и важной информации из исследований самой способности параллельной обработки информации в таких системах, не говоря уже о возможности имитации трехмерной архитектуры и топологии связей в вычислительных системах живых организмов. Помимо этого, изучение нейронных сетей позволит нам гораздо лучше понять принципы объединения вычислительных и запоминающих функций в живых организмах (создаваемые человеком КМОП-схемы осуществляют эти процессы раздельно).

В завершение раздела можно кратко описать ситуацию в оптоэлектронике. Оптоэлектронные приборы (основанные главным образом на прямозонных полупроводниковых соединениях $A_{III}B_V$) получили большое распространение в связи со значительным прогрессом в оптоволоконной технике связи. Кроме того, в настоящее время развивается тенденция замены, где это возможно, электронных элементов фотонными. Наиболее часто используемые полупроводниковые структуры на основе соединений $A_{III}B_V$ используют гетероструктуры $AlGaAs-GaAs$ и $GaNAsP-InP$, которые позволяют изготавливать приборы в оптическом диапазоне 0,8–1,6 мкм. Для более коротковолнового диапазона примерно десять лет назад были разработаны голубые лазеры на основе GaN. В течение последних двух десятилетий развиваются и внедряются полупроводниковые лазеры на квантовых ямах с очень низким значением порогового тока, а также фотодетекторы, которые постепенно заменяют обычные, особенно в оптической связи на больших расстояниях. Особое внимание привлекают полупроводниковые лазеры на гетероструктурах с «напряженными» слоями, в которых пороговый ток остается достаточно слабым.

В настоящее время лазерные диоды формируются в чипах (вместе с транзисторами и оптическими соединениями) при стандартных методах производства интегральных схем. Такие схемы получили название оптоэлектронных интегральных схем (OEIC). Во всех этих случаях оптоэлектроника позволяет значительно повысить степень интеграции и уменьшить размеры элементов изготавливаемых схем, однако следует особо отметить, что характерные размеры должны оставаться в микронном диапазоне. В этой области также существуют обширные направле-

ние исследований, связанных с совершенствованием архитектуры интегральных схем.

По-видимому, основанная на квантовых полупроводниковых гетероструктурах оптоэлектроника может найти наиболее эффективное применение в системах электрооптического модулирования сигналов. Действительно, рассматриваемые ниже в разделе 8.4 модуляторы, действие которых основано на использовании квантово-размерного эффекта Штарка, на несколько порядков величины эффективнее стандартных объемных аналогов. В разделе 10.8 будет показано, что это связано с тем фактом, что энергия ионизации экситона в квантовых ямах значительно выше соответствующих значений для объемных материалов, вследствие чего такие структуры и могут работать в очень сильных электрических полях.

1.3. Характеристические длины в мезоскопических системах

Мезоскопической физикой называют физику явлений в структурах, размеры которых являются промежуточными между макроскопическими и микроскопическими объектами (к последним можно отнести объекты, размеры которых сопоставимы с атомами). Иногда такие структуры называют просто мезоскопическими системами или наноструктурами, поскольку обычно они имеют размеры в диапазоне от нескольких до примерно 100 нм. В таких мезоскопических системах отчетливо проявляются волновые свойства электронов, вследствие чего поведение последних начинает очень сильно зависеть от конкретной геометрии исследуемых образцов. В этих условиях состояние электронов определяется волновыми свойствами и скорее напоминает поведение электромагнитного излучения в волноводах.

При описании поведения электронов в таких твердотельных структурах очень удобно ввести (и определить) с самого начала набор нескольких характеристических размеров или длин, так как именно в тех ситуациях, когда размеры твердого тела, в котором находится электрон, сопоставимы или меньше характеристических длин материала, начинают проявляться некоторые новые, специфические свойства, отличные от привычных характеристик макроскопических образцов. Физическая сущность новых явлений и свойств материала определяется зако-

нами квантовой механики, но с другой стороны, очевидно, что с возрастанием размеров мезоскопическая структура по всем своим свойствам должна непрерывно переходить в привычное, макроскопическое состояние вещества при размерах, в несколько раз превышающих характеристические длины.

Ниже рассматриваются некоторые наиболее часто встречающиеся и используемые характеристические длины мезоскопических систем.

1. Длина волны де Бройля.

Из квантовой механики известно, что электрону с импульсом p соответствует волна с длиной λ_B , называемой длиной волны де Бройля:

$$\lambda_B = \frac{h}{p} = \frac{h}{m^* v}, \quad (1.1)$$

где в уравнении (1.1) импульс p в квазиклассическом приближении выражен как произведение m^*v (m^* — эффективная масса электрона). Из физики твердого тела известно, что динамическое поведение электронов в полупроводнике может быть описано при условии, что им приписывается именно масса m^* (а не реальная масса электрона в вакууме m_0). Последнее замечание имеет особое значение для рассматриваемой нами теории, так как для многих интересных полупроводников значения m^* намного меньше m_0 . Например, в широко используемых соединениях GaAs и InSb эффективная масса m^* составляет всего 0,067 m_0 и 0,014 m_0 соответственно. Представление об эффективной массе будет дано в разделе 2.6.2, однако можно сразу отметить, что при более низких значениях m^* размерные квантовые эффекты в наноструктурах наблюдаются легче и проявляются сильнее. В этом случае поведение полупроводников резко отличается от большинства металлов, в которых электроны проводимости почти всегда могут считаться квазисвободными. Современная литографическая техника позволяет сравнительно легко создавать полупроводниковые наноструктуры, размеры которых (в одном или двух измерениях) меньше введенной выше величины λ_B .

2. Средний свободный пробег электрона.

Электрон,двигающийся в твердом теле, обычно испытывает рассеяние при взаимодействии с кристаллическими дефектами, например, примесями, протяженными дефектами, колебаниями

решетки (фононами) и др. Обычно такие процессы рассеяния или «столкновения» являются неупругими, в результате чего изменяются значения энергии и импульса. Расстояние, проходимое электроном между двумя такими неупругими процессами взаимодействия, принято называть средним свободным пробегом электрона ℓ_e в данном твердом теле, который равен

$$\ell_e = v\tau_e, \quad (1.2)$$

где через v обозначена скорость электрона, а величина τ_e обозначает так называемое время релаксации.

3. Диффузионная длина.

В мезоскопических системах (с характерным размером L) электроны могут двигаться либо по привычному диффузионному механизму, либо баллистически. Баллистическим механизмом переноса носителей заряда называют движение в системах, где определенная выше средняя длина свободного пробега ℓ_e значительно больше характерного размера структуры L , в результате чего движение происходит фактически без рассеяния и основным фактором рассеяния выступают поверхности самой структуры. В транзисторах с горячими электронами (см. раздел 9.5) перенос электронов происходит именно по баллистическому механизму, в результате чего они и могут приобретать энергию, значительно превышающую kT , которая соответствует тепловой энергии решетки. В обратном случае (т. е. когда $\ell_e \ll L$) движение электронов в системе описывается коэффициентом диффузии D , который связан с так называемой диффузионной длиной L_e соотношением (см. также раздел 3.5.2)

$$L_e = (D\tau_e)^{1/2}, \quad (1.3)$$

где τ_e — время релаксации. В теории полупроводников понятие диффузионной длины используется очень широко и часто. Например, при диффузии электронов в полупроводнике p -типа их концентрация экспоненциально уменьшается с расстоянием, причем коэффициентом затухания в экспоненте выступает именно параметр L_e .

При диффузионном режиме перенос электронов в мезоскопических системах обычно описывается уравнением Больцмана, т. е. так же, как и в случае объемных систем. Для бал-

листического механизма уравнение Больцмана, естественно, неприменимо, поскольку движение электронов через структуру происходит практически без столкновений.

4. Длина экранирования.

В несобственных полупроводниках легирующие примеси обычно ионизированы и выступают в качестве основного фактора процессов рассеяния электронов. Однако в общем случае нельзя утверждать, что электрический потенциал этих примесей спадает пропорционально величине $1/r$. Дело в том, что такие заряженные центры рассеяния часто «экранируются» свободными зарядами обратного знака, в результате чего воздействие примесей оказывается ослабленным. Можно показать (см., например, работу [3]), что изменение потенциала модулируется членом вида $\exp(-r/\lambda_s)$, в котором параметр λ_s (называемый длиной экранирования) определяется уравнением

$$\lambda_s = \left(\frac{\epsilon k T}{e^2 n} \right)^{1/2}, \quad (1.4)$$

где e — заряд электрона; ϵ — диэлектрическая постоянная полупроводника; n — средняя концентрация носителей заряда. Следует быть внимательным в использовании приводимого термина, так как многие авторы часто называют величину λ_s дебаевским расстоянием, радиусом Дебая, длиной Ферми — Томаса и т. д. В обычных полупроводниках величина λ_s составляет от 10 до 100 нм, а ее значение характеризует степень подавления флуктуаций заряда в полупроводнике. Из уравнения (1.4) легко заключить, что значение λ_s в металлах должно быть значительно ниже, чем в полупроводниках.

На рис. 1.2 схематически представлены обычный кулоновский и соответствующий ему экранированный потенциал, причем последний описывается формулой

$$\Phi_{sp} = \gamma \frac{e^{-r/\lambda_s}}{r}, \quad (1.5)$$

где $\gamma = 1/4\pi\epsilon_0$. Как видно из (1.5), при $\lambda_s \rightarrow \infty$ эффект экранирования исчезает, и экранируемый потенциал превращается в обычный кулоновский. Из сравнения кривых на рис. 1.2 можно видеть, что при расстояниях от примеси, превышающих $2\lambda_s$, происходит

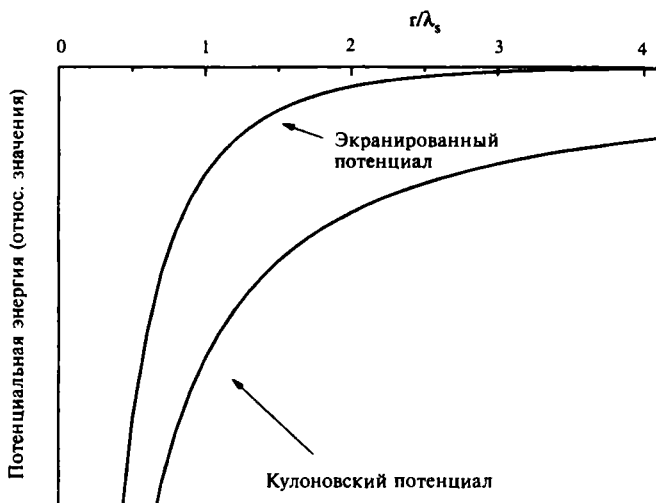


Рис. 1.2. Сравнение экранированного и неэкранированного кулоновских потенциалов

почти полное экранирование. В вышеприведенном обсуждении источником потенциала служил заряженный атом примеси, однако в общем случае неоднородность потенциала может возникнуть вследствие любого нарушения однородности распределения концентрации зарядов. Стоит также отметить, что фигурирующая в приведенных формулах диэлектрическая постоянная ϵ может зависеть от расстояния, т. е. представлять функцию вида $\epsilon(r)$.

5. Длина локализации.

Представление о длине локализации можно пояснить, рассматривая процессы переноса в неупорядоченной среде, где (как известно из курса физики твердого тела) помимо блоховских протяженных (делокализованных) состояний могут существовать также локализованные состояния (см. раздел 7.7.3). В неупорядоченной среде электроны перемещаются в результате «перескоков» между локализованными состояниями (или между локализованными и связанными состояниями).

Для описания параметров «прыжкового» переноса и других мезоскопических характеристик локализованных состояний удобно пользоваться волновой функцией электрона в виде

$$\psi = \exp(-r/\lambda_{loc}), \quad (1.6)$$

где коэффициент λ_{loc} называют длиной локализации. Очевидно, что электропроводность любого материала должна быть пропорциональна (помимо прочих зависимостей) перекрыванию волновых функций. Если размеры образца имеют порядок λ_{loc} , то мы можем считать систему мезоскопической. Мы будем пользоваться представлением о длине локализации ниже, при рассмотрении квантового эффекта Холла в главе 7. Упомянутый тип локализации, кстати, напоминает так называемую локализацию Андерсона, используемую для объяснения переходов металл — изолятор в физике твердого тела [4].

1.4. Квантово-механическая когерентность

В мезоскопических структурах, размеры которых соизмеримы с длиной волны де Бройля $\lambda_{\text{в}}$ для электронов, поведение последних должно описываться в соответствии с законами квантовой механики, т. е. с использованием уравнения Шрёдингера. При неупругом взаимодействии электрона с атомом примеси или другим дефектом происходит изменение энергии и импульса электрона, а также фазы его волновой функции. Длина фазовой когерентности L_{ϕ} определяется как расстояние, проходимое электроном без изменения фазы несущей волны. Понятно, что интерференционные эффекты волн электронов могут наблюдаться только при движении частиц на расстояния порядка или меньше чем L_{ϕ} .

Поскольку нас будут интересовать мезоскопические системы с квазибаллистическим режимом переноса электронов (при котором рассеяние электронов практически отсутствует), то величина L_{ϕ} должна примерно соответствовать определенной выше длине среднего свободного пробега $\ell_{\text{с}}$ (т. е. пробега без неупругого рассеяния). Очевидно, что когерентные состояния могут демонстрировать интерференционные эффекты. С другой стороны, как только когерентные состояния теряют когерентность (например, вследствие неупругого рассеяния), подавляется суперпозиция соответствующих волн и связанная с этим интерференция (можно считать, что при этом волны материи превращаются в частицы). В мезоскопической физике потерю когерентности называют обычно дефазировкой. Естественно, что для мезоскопических систем характерны именно когерентные процессы.

Из приведенного выше определения L_{ϕ} ясно, что интерференционные эффекты для электронов должны проявляться на

расстояниях меньше L_ϕ . Из волновой теории известно, что при интерференции электронов с фазами ϕ_1 и ϕ_2 амплитуда результирующей волны будет меняться по закону $\cos(\phi_1 - \phi_2)$, вследствие чего амплитуды интерферирующих могут складываться или вычитаться, в зависимости от разности фаз. Более подробно интерференционные эффекты в мезоскопических системах будут рассмотрены ниже (глава 7). В разделе 7.5 будет показано, например, что при эффекте Аронова — Бома магнитное поле (в действительности векторный потенциал) позволяет модулировать разницу фаз между двумя электронными токами при их прохождении по различным каналам. Мы покажем, что варьирование магнитного потока позволяет модулировать проводимость (или ее обратную величину — электрическое сопротивление) мезоскопических систем в квантовых единицах.

1.5. Квантовые ямы, проволоки и точки

В разделе 1.3 мы уже ввели характеристические длины λ (соответствующие тем физическим характеристикам электронов, которые зависят от свойств системы) и показали, что если размеры твердого тела сопоставимы с λ или меньше этого значения, то частицы начинают вести себя подобно волнам, и их поведение должно описываться квантовой механикой.

Рассмотрим электрон, помещенный в ящик с размерами L_x , L_y и L_z . При характеристической длине системы λ мы формально можем столкнуться с четырьмя следующими ситуациями:

- 1) $\lambda \ll L_x, L_y$ и L_z .

В этом случае поведение электрона в системе ничем не отличается от поведения в обычных трехмерных (3D) полупроводниках.

- 2) $\lambda > L_x$ и одновременно $L_x \ll L_y$ и L_z .

Этой ситуации соответствует двумерный (2D) полупроводник, перпендикулярный оси x . Такие мезоскопические системы иногда называют квантовыми ямами (см. рис. 4.1 в главе 4).

- 3) $\lambda > L_x, L_y$ и одновременно $L_x, L_y \ll L_z$.

Такая система соответствует одномерному (1D) полупроводнику или квантовой проволоке, направленной вдоль оси z (см. раздел 4.5).

- 4) $\lambda \gg L_x, L_y$ и L_z .

В этом случае говорят о нульмерном (0D) полупроводнике, который называют также квантовой точкой (см. раздел 4.6).

В мезоскопической физике описанные выше твердые тела (обычно полупроводники) обычно называют объектами с пониженной размерностью, если хотя бы одно из их измерений L_i оказывается меньше характеристической длины λ . Например, если L_x и L_y меньше λ , мы говорим о полупроводнике с размерностью 1 и т. д. В некоторых случаях, например, если λ сопоставимо или чуть больше одного из измерений системы (но при этом значительно меньше двух оставшихся размеров), мы имеем квазидвумерную систему, что на практике означает очень тонкий слой, толщина которого, однако, еще не позволяет проявляться квантовым эффектам.

1.6. Плотность состояний и размерность системы

Плотность квантовых состояний будет введена и обсуждена в главе 4, но в данном разделе мы рассмотрим влияние размерности интересующих нас систем на функцию плотности состояний с математической точки зрения. Из курса физики твердого тела известно, что большинство физических характеристик системы существенным образом зависят от вида функции плотности состояний ρ (сокращенно ФПС). При заданном значении энергии E эта функция определена таким образом, что произведение $\rho(E)\Delta E$ равно числу состояний системы, т. е. числу решений уравнения Шрёдингера для заданного интервала значения энергии) в окрестности E . С одной стороны, известно, что если все размеры L_i (L_x , L_y и L_z) являются макроскопическими, то при соответствующем подборе граничных условий все энергетические уровни могут рассматриваться в качестве квазинепрерывных. С другой стороны, очевидно, что при достаточно малых значениях некоторых из этих измерений L_i интересующая нас функция ФПС становится разрывной. Рассмотрение проще всего начать с вычисления ФПС для нескольких типов твердых тел с низкой размерностью. Прежде всего, читатель должен вспомнить, что в твердых телах зависимость $\rho(E)$ в общем случае имеет вид \sqrt{E} (см. раздел 2.3).

Если каждое квантовое или блоховское состояние (состояние с блоховской волновой функцией, см. разделы 2.4 и 2.5) описывается квантовым числом k , то ФПС имеет вид

$$\rho(E) = \sum_k \delta(E - E_k), \quad (1.7)$$

где через E_k обозначены квантовые состояния, $\delta(E)$ — дельта-функция Дирака. Кроме этого (с учетом вырождения электронных состояний по спину), в полученное выражение необходимо ввести дополнительно множитель 2. Рассмотрим простейший случай макроскопического кристалла с размером ребра $L = Na$, где через a обозначена постоянная решетки, а N соответствует числу узлов решетки вдоль одного измерения (поскольку образец является макроскопическим, подразумевается, что это число достаточно велико). В этом случае энергетический спектр собственных состояний может считаться квазинепрерывным, в результате чего суммирование по k в (1.7) может быть заменено интегрированием, что дает (как показано в разделе 2.3) для кубического образца с длиной грани L и объемом $V = L^3$ значение

$$\sum_k \rightarrow \frac{L}{(2\pi)^3} \int dk. \quad (1.8)$$

Из основ теории твердого тела известно также (см. раздел 2.3), что если энергия E_k зависит только от квадрата импульса $\hbar k$ (произведения k на постоянную \hbar), т. е. имеет вид $E_k = (\hbar k)^2/2m^*$, то имеет место соотношение

$$\rho_{3D}(E) = \frac{V}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E}. \quad (1.9)$$

Повторяя в точности этот вывод для 2D- и 1D-полупроводников (с площадью a и длиной L соответственно), легко получить следующие выражения для функций плотности состояний $\rho_{2D}(E)$ и $\rho_{1D}(E)$ в виде

$$\rho_{2D}(E) = \frac{A}{\pi} \left(\frac{m^*}{\hbar^2} \right), \quad (1.10)$$

$$\rho_{1D}(E) = \frac{L}{2\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \frac{1}{\sqrt{E}}. \quad (1.11)$$

Это позволяет сделать важное заключение, что плотность состояний в 3D-полупроводниках всегда пропорциональна \sqrt{E} , в 2D-полупроводниках — остается постоянной, а в 1D-полупроводниках — обратно пропорциональна \sqrt{E} . Последнее об-

стоятельство показывает, что на дне энергетической зоны для 1D-полупроводников плотность состояний имеет особое значение, поскольку при $E = 0$ мы сталкиваемся с математической сингулярностью.

Уравнения (1.10) и (1.11) для одномерных и двумерных твердых тел с идеальной структурой могут быть выведены совершенно строго, однако ясно, что любые реальные кристаллы в действительности являются трехмерными (а двумерность мы приписываем им сами, пренебрегая очень малыми размерами в измерении, перпендикулярном плоскости кристалла). Ниже, в разделе 4.2, будет показано, что поскольку электроны в таких структурах двигаются в плоскости (x, y) практически свободно, уравнение (1.10) для квазидвумерного случая может быть переписано в виде

$$\rho_{2D}(E) = \frac{A}{\pi} \left(\frac{m^*}{\hbar^2} \right) \sum_{n_z} \theta(E - E_{n_z}), \quad (1.12)$$

где через n_z обозначено квантовое число, возникающее при квантовании с учетом ограниченности движения вдоль оси z , а θ — ступенчатая функция. Результирующая функция показана на рис. 4.3.

Аналогично для квазиодномерного твердого тела (квантовой проволоки вдоль оси z) можно получить уравнение

$$\rho_{1D}(E) = \frac{L}{2\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \sum \frac{1}{\sqrt{E - E_{n_x, n_y}}}, \quad (1.13)$$

где квантовые числа n_x и n_y относятся к пространственной локализации вдоль осей x и y соответственно. Полученная функция имеет вид набора пиков, типа представленного на рис. 4.7 (который соответствует одному из значений E_{n_x, n_y}).

Очевидно, что для квази-0D-твердого тела (квантовой точки) мы не сможем получить непрерывную функцию плотности состояний (поскольку квантовые ограничения будут относиться ко всем трем направлениям), в результате чего она примет вид набора пиков

$$\rho_{0D}(E) = \sum_i \delta(E - E_i), \quad (1.14)$$

где $i = (n_x, n_y, n_z)$ (см. рис. 4.8). Реальные пики квантовых точек,

конечно, также не представляют собой идеальные математические δ -функции из-за того, что в системе существуют рассеяние.

1.7. Полупроводниковые гетероструктуры

Хотя существует очень большое число (порядка 100) разнообразных полупроводниковых электронных приборов, все они могут быть в принципе сведены к очень небольшому числу следующих фундаментальных структур: 1) *p-n* гомопереходы (главным образом на основе кристаллического кремния); 2) поверхности раздела металл — полупроводник; 3) структуры типа металл — диэлектрик — полупроводник (в частности, так называемые МОП-структуры или структуры Si — SiO₂-металл) и 4) полупроводниковые гетероструктуры, т. е. поверхности раздела с двух полупроводников с различной шириной запрещенной зоны. Большинство электронных и оптоэлектронных приборов, рассматриваемых в конце книги (главы 9 и 10), основаны именно на использовании гетеропереходов. Кроме того, гетеропереходы весьма удобны для фундаментальных исследований в области мезоскопической физики, особенно двумерных (2D) электронных систем, наглядным доказательством чего может служить обнаружение квантового эффекта Холла К. фон Клитцингом в коммерчески выпускаемых МОП-структурах (раздел 7.7) [5].

В настоящее время квантовые гетероструктуры наиболее высокого качества реализованы на полупроводниках A_{III}B_V. Наиболее распространенными из них являются подробно описанные в разделе 5.3 гетероструктуры в системах AlGaAs — GaAs (изделия на их основе промышленно производятся уже более двадцати лет). Зонная диаграмма простейшего гетероперехода в системе AlGaAs — GaAs показана на рис. 1.3. Как будет показано ниже (в разделе 5.3) наиболее интересный эффект в таких гетероструктурах проявляется вследствие большой концентрации электронов, возникающих из-за легирования AlGaAs *n*-типа проводимости, которые стекают в квантовую яму вблизи границы раздела, в результате чего они оказываются пространственно разделенными с атомами легирующей примеси. Такие структуры называют также модулированно-легированными гетероструктурами, поскольку легирование в них носит избирательный характер. Во внешнем электрическом поле (например, как это имеет место в MOSFET-транзисторах) электроны в таких структурах

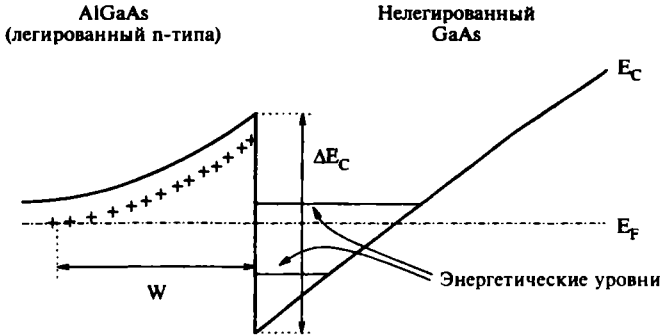


Рис. 1.3. Диаграмма зон проводимости в модулированно-легированном гетеропереходе (сравните ее с рис. 5.3 в разделе 5.3.10)

движутся очень быстро, прежде всего, из-за малости эффективной массы электронов в GaAs, а также (что, кстати, важнее) из-за отсутствия рассеяния на примесях внутри квантовых ям. Кроме этого, электроны в таких структурах двигаются внутри квантовых ям (типичная ширина которых обычно значительно меньше параметра λ_D), вследствие чего их движение квантуется по отношению к перпендикулярному направлению к поверхности раздела. Электронная система в квантовых ямах является двумерной, а функция плотности состояний в ней определяется уравнением (1.12), которое, как будет показано в разделе 5.3, обеспечивает гораздо более высокие значения электронных и оптических характеристик материала.

Производство высококачественных модулированно-легированных гетероструктур стало возможным лишь после разработки процессов молекулярно-лучевой эпитаксии и осаждения металлоорганических соединений из газовой фазы, что позволило получать очень тонкие пленки, наращивая их атомарными слоями на высококачественные подложки при вакууме порядка 10^{-12} мбар. Это позволило получить весьма регулярные (без шероховатостей) границы раздела в латеральном направлении и очень резкий градиент состава в перпендикулярном направлении. Качество таких поверхностей значительно выше, чем у поверхностей типа Si—SiO₂ в обычных МОП-структурах (см. раздел 5.2), вследствие чего квантовые эффекты в них проявляются значительно сильнее.

Полупроводниковые квантовые гетеропереходы в настоящее время являются основой многих коммерчески производимых электронных и оптоэлектронных приборов, и поэтому можно ска-

зять, что некоторые из наиболее современных электронных и оптоэлектронных приборов уже могут быть отнесены к объектам нанотехнологии. Как будет показано в главе 9, диоды с резонансным туннелированием состоят из двух гетеропереходов (между которыми располагается квантовая яма), а транзисторы с высокой подвижностью электронов (раздел 9.2) содержат один гетеропереход. Аналогично на гетероструктурах с квантовыми ямами созданы многие самые современные оптоэлектронные приборы, такие, как лазеры для оптической связи, CD-плееры и высокоскоростные электрооптические модуляторы, описанные в главе 10.

1.8. Квантовые процессы переноса

В течение последних двадцати лет изучение процессов переноса в наноструктурах было связано с исследованием различных эффектов квантовой интерференции (раздел 1.4), а также с изучением одноэлектронных процессов переноса (раздел 6.4.3). Причина такого интереса заключается в том, что средний свободный пробег электронов в полупроводниках обычно значительно больше, чем в металлах, вследствие чего и интерференционные эффекты в полупроводниковых структурах проявляются намного сильнее. В большинстве ранних исследований макроскопических процессов переноса различали два механизма переноса (диффузионный и баллистический), что уже отмечалось в разделе 1.3 [1]. Диффузионный перенос практически не связан с формой системы, поскольку при нем электроны участвуют в многочисленных процессах неупругого рассеяния, подобно переносу в объемных телах. С другой стороны, в квантовых гетероструктурах (где $\ell \gg L$) электроны двигаются баллистически, рассеиваясь лишь на границах системы. В этом случае (когда дополнительно параметр λ_B сравним или превышает значение L) квантование энергии электрона в ямах становится весьма существенным. Баллистические электроны, проходящие через структуру без рассеяния, могут демонстрировать заметные интерференционные эффекты. Особенно часто такие эффекты (описываемые ниже в главе 7) проявляются при наложении внешних магнитных полей.

В металлических наноструктурах величина ℓ составляет около 100 \AA , и перенос в них осуществляется обычно по диф-

фузионному механизму. С другой стороны, в полупроводниковых гетероструктурах величина ℓ доходит до нескольких микрон, в результате чего эффекты пространственной локализации становятся более существенными. В этом случае мы вообще не можем описать процессы переноса, пользуясь макроскопическими понятиями и представлениями (типа общеизвестного коэффициента электропроводности), и вынуждены как-то оперировать волновыми характеристиками функции, определяемой уравнением Шрёдингера. И действительно, в таких системах следует пользоваться новыми определениями, сформулированными Ландауэром и Бутиккером, позволяющими объяснить наблюдаемое квантование проводимости (см. раздел 6.4). Понятие о квантовом переносе относится и к туннелированию электронов через потенциальные барьеры регулируемой высоты и ширины. Примером практического использования квантового переноса могут служить диоды с резонансным туннелированием, которые демонстрируют очень высокую вероятность квантового туннелирования (достигающую при некоторых энергиях почти 100%), а также область отрицательного дифференциального сопротивления (см. раздел 9.4) на вольт-амперных характеристиках.

Литература

1. *Mitin V.V., Kochelap V.A., Strosio M.A.* (1999). *Quantum Heterostructures*. Cambridge. Cambridge University Press.
2. *Goser K., Glösekötter P., Dienstuhl j.* (2004). *Nanoelectronics and Nanosystems*. Berlin. Springer.
3. *Kittel Ch.* (2005). *Introduction to Solid State Physics*, 8th edition. New York. Wiley. Есть русское издание: *Куммель Ч.* Введение в физику твердого тела. Наука, 1962.
4. *Anderson P.W.* (1958). *Phys. Rev.*, **109**, 1492.
5. *von Klitzing K., Dorda G., Pepper M.* (1980)/ *Phys. Rev. Lett.*, **45**, 494.

Дополнительная литература

- Goser K., Glösekötter P., Dienstuhl j.* (2004). *Nanoelectronics and Nanosystems*. Berlin. Springer.
- Imry Y.* (1997). *Introduction to Mesoscopic Physics*. Oxford. Oxford University Press.

Mesoscopic Physics and Electronics (1998). Eds. *Ando T., Arakawa Y., Furuya K., Kotiyama S., Nakasima H.* Berlin. Springer.
Murayama Y. (2001). Mesoscopic Systems. Weinheim. Wiley-VHC.

Задачи и упражнения

1. Уровни Ландау в объемных твердых телах (упражнение предлагается, чтобы напомнить читателю о возможности возникновения эффектов пониженной размерности даже в объемных твердых телах под воздействием внешнего магнитного поля). Покажите, что энергетические уровни трехмерного твердого тела в магнитном поле B (направленном вдоль оси z) описываются уравнением

$$E(n, k_z) = E(n) + E(k_z) = \left(n + \frac{1}{2}\right) \hbar \omega_c + \frac{\hbar^2 k_z^2}{2m^*}, \quad n = 0, 1, 2, K,$$

где ω_c — так называемая циклотронная частота, определяемая соотношением $\omega_c = eB/m^*$ (предполагается, что эквивалентные поверхности в k -пространстве сферически симметричны, т. е. эффективная масса представляет собой скаляр).

2. Магнитная длина. Характеристическая магнитная длина ℓ_m электрона в магнитном поле B определяется уравнением

$$\ell_m = \left(\frac{\hbar}{eB}\right)^{1/2}$$

Покажите, что длина ℓ_m равна радиусу циклотронной орбиты электрона, соответствующему значению квантования энергии в плоскости, перпендикулярной B (см. задачу 1) при $n = 0$. (Подсказка: можно воспользоваться полуклассическим приближением и записать $E = m^*v^2/2$, где $v = r\omega_c$, а ω_c — циклотронная угловая частота.)

3. Длина когерентности в сверхпроводниках. Сравните длину когерентности ξ в сверхпроводнике с характерными длинами, введенными в разделе 1.3. (Подсказка: определение ξ следует из феноменологической теории Гинзбурга — Ландау для сверхпроводимости. Средний свободный пробег ℓ в очень чистых сверхпроводниках очень велик, и можно показать, что $\xi = \hbar v_F / \pi E_g$ (где v_F — скорость Ферми; E_g —

сверхпроводящая щель). Однако в образцах с большим количеством примесей длина ℓ мала, в результате чего имеем $\xi = (2 \hbar v_F \ell / \pi E_g)^{1/2}$.

4. Плотность состояний для твердых тел в магнитном поле. Исходя из выражения для $E(n, k_y)$ (полученного в задаче 1), покажите, что плотность состояний твердого тела в магнитном поле выражается бесконечным рядом односторонних гиперболических ступенчатых функций, подобных показанной на рис. 4.7. Докажите также, что аналитически решение может быть записано в виде

$$n_{1D} \propto \sum_{n=0}^{\infty} (E - E_n)^{-1/2},$$

где E_n определяется квантованными значениями энергии, полученными в задаче 1. Представьте величину n_{1D} в виде функции от E и покажите, что она имеет сингулярности ($n_{1D} \rightarrow \infty$) при значениях энергии $E_n = \hbar \omega_c / 2, 3\hbar \omega_c / 2$ и т. д.

5. Тензорные коэффициенты сопротивления и проводимости в двумерных структурах. Выражения для тензорных коэффициентов сопротивления σ и проводимости ρ в двумерных структурах полезны для понимания процессов переноса с учетом магнитного поля и используются ниже (в главе 7) при объяснении квантового эффекта Холла. Покажите, что в двумерных системах (считая, что ток направлен вдоль оси x , а внешнее магнитное поле — вдоль оси z) тензорные коэффициенты сопротивления и проводимости связаны следующими соотношениями:

$$\rho_{xx} \equiv \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2},$$

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2}.$$

(Подсказка: предположим, что компоненты тока и электрического поля связаны соотношениями

$$J_x = \sigma_{xx} E_x + \sigma_{xy} E_y,$$

$$J_y = \sigma_{yx} E_x + \sigma_{yy} E_y,$$

а образец является гомогенным и изотропным, т. е. $\sigma_{xx} = \sigma_{yy}$ и $\sigma_{xy} = -\sigma_{yx}$.)

6. Коэффициент Холла в двумерных структурах. Решение этой задачи позволяет рассматривать поведение двумерных электронных систем в магнитном поле. Геометрия совпадает с условиями предыдущей задачи (магнитное поле B направлено вдоль оси z). Покажите, что: а) компоненты скорости электронов в плоскости (x, y) описываются уравнениями

$$v_x = -\frac{e\tau}{m^*} E_x - w_c \tau v_y,$$

$$v_y = -\frac{e\tau}{m^*} E_y - w_c \tau v_x$$

(подсказка: воспользуйтесь представлением о приближенном времени релаксации в твердых телах, вводимым ниже в разделе 3.5.1); б) что если при измерениях эффекта Холла выполняется условие $v_y = 0$, то коэффициент Холла R_H ($\equiv E_y/J_x B$) имеет вид

$$R_H = \frac{E_y}{J_x B} = \frac{v_y}{IB} = -\frac{1}{N'_s e},$$

где параметр N'_s относится к концентрации электронов (на единицу площади), а $J_x = I/a$ (a — ширина образца).

ГЛАВА 2

ВВЕДЕНИЕ В ФИЗИКУ ТВЕРДОГО ТЕЛА

2.1. Введение

В этой главе приводится сводка наиболее общих понятий и концепций физики твердого тела, необходимых для дальнейшего рассмотрения поведения и особенностей квантовых наноструктур. Например, поведение электронов в квантовых ямах значительно отличается от поведения в объемных твердых телах лишь в тех случаях, когда электроны преодолевают ограничивающие эти ямы потенциальные барьеры, однако при движении вдоль поверхностей раздела поведение электронов практически не изменяется. В разделе 2.2 мы рассмотрим также некоторые основные понятия квантовой механики, особенно относящиеся к методам теории возмущений для постоянных и меняющихся во времени возмущений, что приводит к так называемому «золотому правилу» для переходов между электронными состояниями. В разделе 2.3 рассмотрена модель для поведения электронов в твердых телах (восходящая еще к Зоммерфельду), основанная на том, что квантовая механика позволяет получить только некоторые определенные значения для энергии частицы в потенциальном ящике. Эта модель игнорирует особенности периодических кристаллических решеток в конкретных твердых телах и связана лишь с представлением о некотором усредненном потенциале V_0 , однако она позволяет качественно объяснить многие характерные свойства твердых тел (в частности, величину и вид температурной зависимости электронной теплоемкости, соотношение Видемана — Франца между коэффициентами теплопроводности и электропроводности металлов и т. п.). В силу своей простоты, естественно, эта модель не способна объяснить механизм более тонких эф-

фектов, типа наличия коэффициентов Холла с положительным знаком в некоторых металлах, или просто продемонстрировать, почему некоторые вещества являются металлами, а другие полупроводниками.

В разделе 2.4 рассматривается теорема Блоха, являющаяся основой формального описания поведения электронов в периодических потенциалах. Затем формализм Блоха используется в разделе 2.5 для описания двух предельных случаев — очень слабого периодического потенциала (при котором поведение электронов может рассматриваться почти как свободное) и настолько сильного потенциала, что в первом приближении мы можем считать атомные потенциалы основным фактором общей энергии (приближение сильной связи). Читателю может показаться несколько удивительным, что (с качественной точки зрения) оба предельных случая приводят к похожим результатам, а именно к предсказанию существования электронной зонной структуры, т. е. электроны в твердых телах могут двигаться лишь внутри некоторых энергетических зон, разделенных запрещенными зонами.

В разделе 2.6 рассмотрена динамика движения электронов в твердых телах и показано, что определение Блоха для электронной волновой функции позволяет сразу ввести понятие скорости электрона (определив эту величину как групповую скорость суперпозиции блоховских состояний). При описании динамики электронов в системах с приложенными внешними полями очень полезным представляется использование представления о тензоре эффективной массы электрона, так как это позволяет описывать движение электронов в зонах, связывая его с «искривлением» зонных структур при заданных энергиях. Глава заканчивается разделами 2.7 и 2.8, посвященными нормальным модам атомных колебаний решетки. Подобно электронам в периодических структурах, эти колебания могут иметь только характерные частоты колебаний, зависящими от волнового вектора. Затем, для придания большей «физичности» описанию, вводится представление о фононах. Теоретически фононы представляют собой квазичастицы, которые могут взаимодействовать или претерпевают столкновения с электронами, и это понятие оказалось исключительно полезным для описания и понимания процессов переноса в твердых телах.

2.2. Краткие сведения из квантовой механики

В этом разделе рассматриваются общие представления квантовой механики, необходимые для понимания свойств объемных полупроводников или полупроводников с пониженной размерностью, а также функционирования современных электронных и оптоэлектронных приборов.

2.2.1. Корпускулярно-волновой дуализм и принцип Гейзенберга

В классической физике мы обычно сталкиваемся с двумя типами объектов, а именно с частицами (типа точечных малых масс, которые фигурируют в уравнениях Ньютона) и волнами (типа электромагнитных или световых волн, поведение которых описывается уравнениями Максвелла). Однако при описании очень малых объектов (типа атомов) оба предложенных выше варианта оказываются несостоятельными, и мы вынуждены пользоваться понятиями квантовой механики, в основе которой лежит представление о дуальности «волна — частица». Например, при описании взаимодействия света с веществом очень удобно рассматривать свет не в виде волн, а в качестве частиц определенного типа (давно получивших название фотонов). С другой стороны, электроны (которые всегда рассматривались в качестве частиц) внутри твердых тел с нанометровыми ($1 \text{ нм} = 10^{-9} \text{ м}$) размерами ведут себя подобно волнам.

При изучении обычных оптических явлений (типа известных интерференции и дифракции) мы исходим из представления о световых волнах длиной λ и частотой ν , связанных соотношением $c = \lambda\nu$, где c — скорость света (обычно в вакууме или в прозрачных средах). Еще в 1901 г. Планк предположил (для объяснения кривой излучения черного тела), что свет может излучаться или поглощаться лишь в виде некоторых «порций» энергии, получивших название квантов или фотонов. Энергия фотона равна

$$E = h\nu = \hbar\omega, \quad (2.1)$$

где h — постоянная Планка ($h = 6,62 \times 10^{-34} \text{ Дж} \cdot \text{с}$, $\hbar = h/2\pi$); $\omega = 2\pi\nu$ — угловая частота.

Представление о световых квантах (фотонах) полностью утвердилось в физике после того, как Эйнштейн использовал его для объяснения фотоэлектрического эффекта. В соответствии с предложенным им объяснением электроны испускаются с поверхности вещества, если падающие на поверхность фотоны обладают достаточной энергией. Очевидно, что энергия падающих фотонов должна быть выше энергетического барьера, удерживающего электроны внутри твердого тела. Эйнштейн также показал, что фотон с импульсом p обладает энергией E , связанной с импульсом соотношением

$$E = cp, \quad (2.2)$$

где c — скорость света. Полезно также запомнить, что фотоны движутся со скоростью света, а их масса покоя равна нулю.

Традиционно электроны рассматривались в качестве частиц (что очень наглядно демонстрировалось, например, столкновениями ускоренных электронов с флюоресцирующим экраном и т. п.), однако в 1927 г. Дэвидсон и Джермер обнаружили явление дифракции электронов при отражении от поверхности кристалла никеля, как если бы они были волнами и подчинялись законам дифракции Брэгга. Поэтому можно утверждать, что концепция дуализма основана на экспериментальных результатах. Действительно, как и предсказал де Бройль в 1924 г., каждой частице с импульсом p можно сопоставить волну с длиной

$$\lambda = \frac{h}{p}. \quad (2.3)$$

Это уравнение может быть переписано в виде

$$p = \frac{h}{\lambda} = \frac{h}{2\pi} \frac{2\pi}{\lambda} = \hbar k, \quad (2.4)$$

где k — так называемое волновое число.

Еще один, чрезвычайно важный квантово-механический закон был открыт в 1927 г., когда Гейзенберг сформулировал так называемый принцип неопределенности, в соответствии с которым в любом эксперименте произведение ошибок изме-

рения импульса частицы Δp_x и ее координаты Δx всегда должно превышать $\hbar/2$, т. е.

$$\Delta p_x \Delta x \geq \hbar/2 \quad (2.5)$$

(разумеется, такие же соотношения справедливы и для измерений по остальным координатам, и их можно выписать для произведений $\Delta p_y \Delta y$ и $\Delta p_z \Delta z$). При этом следует особо подчеркнуть, что принцип неопределенности является фундаментальным свойством природы, а вовсе не связан с конкретными типами приборов, при помощи которых экспериментатор пытается одновременно измерить величины p_x и x . В другой формулировке этот принцип устанавливает связь между ошибками измерений энергии частицы E и временным промежутком Δt , требуемым для измерения:

$$\Delta E \Delta t \geq \frac{\hbar}{2}. \quad (2.6)$$

2.2.2. Уравнение Шрёдингера

Далее, как показал Шрёдингер в 1926 г., дуальность «волна — частица» (т. е. наличие у вещества одновременно волновых и корпускулярных свойств) очень удобно описывать волновой функцией $\Phi(\vec{r}, t)$, которая не только является непрерывной, но и имеет непрерывные производные. Такая волновая функция удовлетворяет линейному дифференциальному уравнению второго порядка

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\vec{r}, t) \right] \Phi = i\hbar \frac{\partial \Phi}{\partial t}, \quad (2.7)$$

которое называется уравнением Шрёдингера. Через ∇^2 обозначен оператор $\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$, а V обозначает потенциальную энергию, которая в общем случае является функцией координат (и, возможно, времени). Хотя сама функция Φ не имеет физического смысла, ее произведение на собственное комплексное сопряжение представляет реальную величину, а именно вероятность dP обнаружения частицы в элементе объема dV :

$$dP = |\Phi|^2 dV. \quad (2.8)$$

Если потенциальная энергия не зависит от времени, то мы можем искать решение уравнения (2.7) в виде

$$\Phi(\vec{r}, t) = \psi(\vec{r}) e^{-i\omega t}. \quad (2.9)$$

Подставляя (2.9) в (2.7) и используя соотношение $E = \hbar\omega$, мы получим независящее от времени уравнение Шрёдингера

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\vec{r}) \right] \psi(\vec{r}) = E \psi(\vec{r}) \quad (2.10)$$

для независящей от времени волновой функции $\psi(\vec{r})$.

Уравнение Шрёдингера решается точно лишь для нескольких систем, простейшей из которых является, по-видимому, свободно двигающаяся частица, например, электрон с энергией E и массой m . В этом случае имеем $V(\vec{r}) = 0$, вследствие чего уравнение (2.7) легко преобразуется в

$$\Phi = A e^{i(kx - \omega t)} + B e^{i(-kx - \omega t)}, \quad (2.11)$$

где

$$k = \left(\frac{2mE}{\hbar^2} \right)^{1/2} \quad (2.12)$$

Таким образом, движение свободного электрона описывается некоторой волной, которая (в соответствии с идеей де Бройля) обладает энергией и импульсом, связанными соотношениями

$$p = \hbar k, \quad E = \frac{p^2}{2m}. \quad (2.13)$$

В общем случае предполагается, что электрон движется по одной из осей координат (например, вдоль оси x слева направо), вследствие чего коэффициент B в уравнении (2.11) можно считать равным нулю. При этом волновое уравнение для свободного электрона приобретает простой вид

$$\Phi = A e^{i(kx - \omega t)}. \quad (2.14)$$

Еще одним примером точного решения уравнения Шрёдингера может служить описание атома водорода с кулоновским потенциалом взаимодействия (т. е. при потенциале V , обратно пропорциональном расстоянию r между протоном и электроном, $V(r) \sim 1/r$). Решая уравнение Шрёдингера при этих условиях, легко получить известное выражение для энергетических уровней электрона

$$E_n = -\frac{m_r e^4}{2(4\pi\epsilon_0)^2 \hbar^2 n^2} = -\frac{13,6}{n^2} (\text{eV}), \quad n = 1, 2, \dots, \quad (2.15)$$

где через m_r обозначена приведенная масса системы протон—электрон. Эта математическая модель атома водорода очень широко применяется в физике твердого тела, например, для описания поведения примесей или экситонов в полупроводниках. Используемые при этом уравнения очень похожи на (2.15), хотя получаемые значения энергии значительно ниже из-за того, что диэлектрическая постоянная для кристаллов значительно выше соответствующего значения ϵ_0 для вакуума (например, значение диэлектрической постоянной в кремнии составляет около $12 \epsilon_0$).

Столь же часто в физике твердого тела используется модель гармонического осциллятора, как, например, при изучении динамики решетки (см. разделы 2.7 и 2.8). Вообще говоря, следует отметить, что модель гармонического осциллятора пригодна для описания практически любой системы, в которой происходят колебания с малыми отклонениями от равновесного состояния. Разрешенные при этом уровни энергии могут быть получены из уравнения Шрёдингера и имеют вид

$$E_n = \left(n - \frac{1}{2} \right) \hbar\omega, \quad n = 1, 2, 3, \dots \quad (2.16)$$

Точное решение уравнения Шрёдингера может быть получено еще для одного широко используемого в теории потенциала, а именно прямоугольной потенциальной ямы с бесконечной глубиной, соответствующей «плоскому» потенциалу шириной a с бесконечно нарастающим потенциалом на границе. Такой потенциал подробно рассматривается в разделах 4.2 и 5.4.1 в связи с тем, что он представляет собой очень хорошее приближение

для описания так называемых модулированно-легированных квантовых ям, которые являются основными структурами в транзисторах и лазерах, действие которых основано на квантовых ямах.

2.2.3. Распределения Ферми — Дирака и Бозе — Эйнштейна

В этом разделе мы рассмотрим наиболее часто используемые функции распределения для частиц разных типов (например, электронов, фотонов и т. п.). Обычно все виды частиц разделяют на две большие группы. Электроны и другие частицы, обладающие полуцелым (нецелочисленным) спином, обычно называют фермионами. Поведение таких частиц определяется принципом исключения Паули и статистикой Ферми — Дирака, к другому классу относят так называемые бозоны, т. е. частицы с нулевым (или вообще целочисленным) спином, которые подчиняются статистике Бозе — Эйнштейна.

В равновесном состоянии среднее число фермионов, занимающих некое состояние с энергией E , описывается функцией распределения Ферми — Дирака $f_{FD}(E, T)$, имеющей вид

$$f_{FD}(E, T) = \frac{1}{\exp[(E - E_F)/kT] + 1} \quad (2.17)$$

и определяющей вероятность заполнения данного состояния. Энергия E_F называется энергией Ферми или уровнем Ферми в полупроводниках. Из уравнения (2.17) следует, что при $E = E_F$ значение функции f_{FD} примерно равно 0,5. На рис. 2.1 представлена функция распределения Ферми — Дирака при нескольких значениях температуры T . Легко заметить, что при $E = E_F$ значения всех кривых действительно совпадают с 0,5. При $T = 0$ к функция f_{FD} превращается в известную из математики ступенчатую функцию, что соответствует полному заполнению всех состояний с энергией ниже E_F , и отсутствию частиц в любых состояниях с энергией выше E_F . Для всех температур переход от 1 к 0 при E_F осуществляется в относительно узком диапазоне энергий, ширина которого составляет около $4 kT$.

Системы бозонов описываются функцией распределения Бозе — Эйнштейна $f_{BE}(E, T)$, имеющей вид

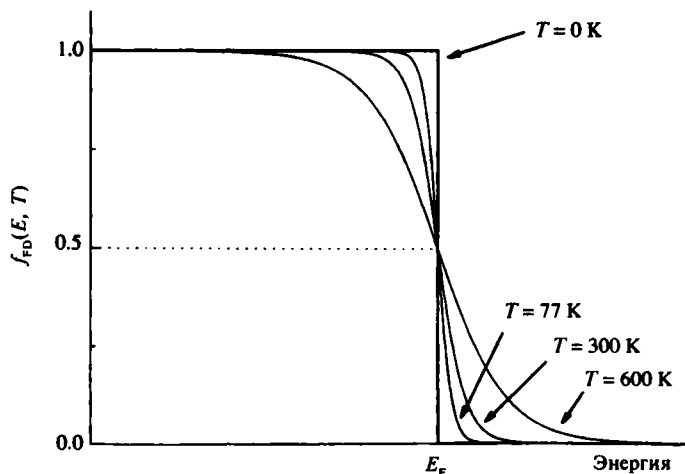


Рис. 2.1. Зависимость функции распределения Ферми — Дирака от энергии при нескольких значениях температуры

$$f_{\text{FD}}(E, T) = \frac{1}{\exp(E/kT) - 1}, \quad (2.18)$$

где привлекает внимание отрицательный знак при единице в знаменателе, в отличие от положительного знака для $f_{\text{FD}}(E, T)$. Это различие имеет важное значение, поскольку при предельном переходе $E \rightarrow 0$ значение функции f_{BE} становится бесконечно большим, вследствие чего бозоны (например, атомы ${}^4\text{He}$) при снижении температуры претерпевают так называемую бозе-конденсацию, при которой все атомы «конденсируются» в основном состоянии. В противоположность этому фермионы стремятся отделиться друг от друга, поскольку в каждом энергетическом состоянии может находиться лишь одна такая частица.

Можно легко заметить, что при высоких (по сравнению с kT) энергиях разница между выражениями (2.17) и (2.18) практически исчезает вследствие незначительности роли единицы в знаменателе, в результате чего обе функции распределения переходят в классическое распределение Больцмана. Таким образом, например, функция распределения электронов в полупроводнике при $(E - E_F) \gg kT$ может быть записана в виде

$$f(E, T) = \exp\left(-\frac{E - E_F}{kT}\right). \quad (2.19)$$

Стоит запомнить, что при энергиях, значительно превышающих kT , оба квантовых распределения превращаются в единую функцию распределения, известную из классической физики.

2.2.4. Методы теории возмущений

В разделе 2.2.2 мы уже рассмотрели несколько важных квантово-механических систем (типа гармонического осциллятора), для которых может быть получено точное решение уравнения Шрёдингера. Однако для большинства интересных с физической точки зрения потенциалов точное решение получить нельзя, вследствие чего и были разработаны разнообразные приближенные методы, важнейшим из которых является метод возмущений. Его применяют в тех случаях, когда гамильтониан H рассматриваемой системы лишь незначительно отличается от гамильтониана H_0 некоторой системы, собственные функции которой известны совершенно точно. В этом случае мы можем положить

$$H = H_0 + H', \quad (2.20)$$

где вспомогательный член H' называется возмущенным гамильтонианом.

Цель задачи состоит в нахождении собственных значений E_n и собственных функций ψ_n гамильтониана H , т. е. решении уравнения

$$H\psi_n = E_n\psi_n = (H_0 + H')\psi_n, \quad (2.21)$$

при условии, что уравнение

$$H_0\phi_n = \xi_n\phi_n \quad (2.22)$$

может быть решено точно, т. е. для него могут быть получены точные собственные значения ξ_n и собственные функции ϕ_n . Прежде всего, неизвестные функции ψ_n следует представить в виде суперпозиции из полного набора невозмущенных волновых функций, т. е. в форме

$$\psi_n = \sum_i c_n \phi_i. \quad (2.23)$$

Во многих учебниках квантовой механики [1] описана стандартная процедура, позволяющая получать несколько последовательных приближений (в нулевом приближении мы имеем $E_n = \xi_n$, причем очень часто вполне достаточным оказывается нахождение приближения первого порядка, т. е. набора

$$E_n \cong \xi_n + H'_{nn} = \xi_n + \langle \phi_n | H' | \phi_n \rangle, \quad (2.24)$$

в который входят только волновые функции приближения нулевого порядка.

В большинстве случаев невозмущенный гамильтониан H_0 представляет собой симметричную функцию, а возмущенный гамильтониан H' обладает какой-либо специфической симметрией (четной или нечетной). Можно достаточно легко доказать, что при четной симметрии H' все матричные элементы, соответствующие состояниям с противоположной четностью, будут равны нулю, и наоборот — при нечетной симметрии гамильтониана H' будут автоматически равны нулю все матричные элементы для состояний с одинаковой четностью. Эти результаты являются прямой аналогией правил отбора в спектроскопии.

До сих пор мы говорили лишь о теории возмущений применительно к стационарным, т. е. независящим от времени состояниям. Для развития теории зависящих от времени возмущений, предположим, что одно из начальных состояний подвергается воздействию внешних сил, зависящих от времени (например, на атом может влиять падающая волна света). Далее, как и раньше, мы можем предположить, что возмущающий член гамильтониана $H'(t)$ является малым. В рамках такого предположения собственные значения и собственные функции для зависящего от времени уравнения Шрёдингера можно искать в виде (2.7)

$$H_\psi = [H_0 + H'(t)]\psi = i\hbar \frac{\partial \psi}{\partial t}. \quad (2.25)$$

Для чего мы можем вновь выразить ψ через суперпозицию невозмущенных волновых функций ϕ_n , а затем искать последующие приближения по стандартному методу.

Очень часто в задачах этого типа мы сталкиваемся с необходимостью вычислить вероятность перехода из некоторого начального состояния в группу конечных состояний с очень

близкими энергиями. В таких случаях удобно описывать конечные состояния функцией плотности состояний $\rho(E)$ (см. раздел 1.6). Можно показать [1], что в этом случае скорость перехода W описывается уравнением

$$W = \frac{2\pi}{\hbar} \rho(E) |H'_{nk}|^2, \quad (2.26)$$

которое часто называют правилом Ферми или (из-за его исключительной полезности и применимости) даже «золотым правилом» квантовой механики. Смысл правила заключается в том, что скорость перехода из ординарного состояния m в группу конечных состояний k представляет собой произведение коэффициента $2\pi/\hbar$ на функцию плотности конечных состояний и квадрат матричного элемента, связывающего начальное и конечное состояния.

Особый интерес представляет случай, когда внешнее возмущение вызывается силой, меняющейся по гармоническому закону, т. е. когда

$$H'(t) = H'_0 e^{i\omega t}, \quad (2.27)$$

где $H'(t)$ соответствует, например, возмущающему электромагнитному полю. В этой ситуации переходы из одних состояний в другие становятся избирательными, т. е. значительно более вероятными для квантов с энергией $\hbar\omega$, поскольку они значительно легче поглощаются (резонансное поглощение) или излучаются (стимулированное излучение) рассматриваемой системой.

2.3. Модель свободных электронов в твердых телах. Функция плотности состояний

Простейшая модель для объяснения поведения электронов в твердых телах (особенно в металлах) основана на предположении, что валентные электроны атомов могут свободно двигаться во всем объеме вещества. Модель фактически предполагает, что свободные электроны могут рассматриваться в качестве своеобразного газа, заполняющего образец вещества. Модель также предполагает, что все свободные электроны находятся под воздействием постоянного электрического потенциала, за

исключением поверхностей твердого тела, где действует энергетический барьер с высотой H , не позволяющий электронам покидать объем образца. С квантово-механической точки зрения эта модель сводится к поведению частиц (в данном случае фермионов) в трехмерной прямоугольной потенциальной яме. В свое время эта относительно простая модель была предложена Зоммерфельдом на основе классической модели Друде.

Поскольку в этой модели потенциальная энергия внутри ямы не изменяется, мы можем решить для электронов зависящее от времени уравнение Шрёдингера (2.10), подставив в него для упрощения потенциал $V = 0$. В разделе 2.2.2 уже было показано, что решением являются волновые функции, описывающие плоские волны с формой:

$$\psi(\vec{r}) = \frac{1}{\sqrt{V}} e^{i\vec{k} \cdot \vec{r}}, \quad (2.28)$$

где трехмерная амплитуда волновой функции равна $1/\sqrt{V}$ (V — объем потенциальной ямы или ящика). Это позволяет нормировать волновые функции естественным образом, т. е. так, чтобы полная вероятность обнаружения электрона в объеме V равнялась 1. Записав это условие в виде

$$\int \psi^* \psi dV = A^2 \int \exp(-i\vec{k} \cdot \vec{r}) \exp(i\vec{k} \cdot \vec{r}) dV = 1, \quad (2.29)$$

легко получить для амплитуды $A = 1/\sqrt{V}$.

Для дальнейшего рассмотрения и использования полученных решений нам необходимо определить для рассматриваемой задачи граничные условия. В физике твердого тела чаще всего рассматриваются и применяются два следующих типа граничных условий. Проще всего выглядят фиксированные граничные условия, когда волновая функция электронов обращается на границе в нуль, что соответствует невозможности для электрона покинуть пределы квантового ящика (предполагается, что энергетический барьер на границе тела имеет бесконечную высоту). Такие условия приводят к нахождению решения в виде стоячих волн. Другим распространенным типом граничных условий выступают так называемые периодические граничные условия (Борна — фон Кармана), при использовании которых мы получаем для электрона решения в виде бегущих волн,

что значительно удобнее при изучении поведения электронов в твердых телах. Эти условия предполагают только, что волновые функции имеют одинаковое значение на всей границе тела. Рассмотрим, например, очень простую ситуацию, когда отдельный электрон движется внутри куба с длиной ребра L . Периодические граничные условия в этом случае записываются в виде системы уравнений:

$$\begin{aligned}\psi(x, y, z) &= \psi(x + L, y, z), \\ \psi(x, y, z) &= \psi(x, y + L, z), \\ \psi(x, y, z) &= \psi(x, y, z + L).\end{aligned}\tag{2.30}$$

Налагая такие граничные условия на определенную уравнением (2.28) волновую функцию, легко получить уравнения

$$e^{ik_x L} = 1, \quad e^{ik_y L} = 1, \quad e^{ik_z L} = 1\tag{2.31}$$

и соответствующие им разрешенные значения для чисел k_x , k_y и k_z в виде

$$k_x = \frac{2\pi}{L} n_x, \quad k_y = \frac{2\pi}{L} n_y, \quad k_z = \frac{2\pi}{L} n_z,\tag{2.32}$$

где $n_x, n_y, n_z = 0, \pm 1, \pm 2, \dots$

Такое квантование значений волновых векторов (2.32) влечет за собой и соответствующее квантование энергетических состояний:

$$E(n_x, n_y, n_z) = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) = \frac{\hbar^2}{2m} \left(\frac{2\pi}{L} \right)^2 (n_x^2 + n_y^2 + n_z^2),\tag{2.33}$$

где по-прежнему должны выполняться условия $n_x, n_y, n_z = 0, \pm 1, \pm 2, \dots$

Выше уже отмечалось, что каждому набору квантовых чисел (n_x, n_y, n_z) соответствует некоторое состояние электрона, волновое число и энергия которого определяются уравнениями (2.32) и (2.33) соответственно. В идеале было бы желательно получить все значения энергий и все волновые функции состояний электрона в твердом теле, однако для практических целей вполне

достаточно определить функцию плотности состояний $\rho(E)$, определенную выше (см. раздел 1.6). Как уже неоднократно отмечалось, важнейшие свойства электронных систем в твердых телах (включая электрические и транспортные характеристики) могут быть легко определены с использованием этой функции.

Каждое электронное состояние определяется набором квантовых чисел (k_x, k_y, k_z) , причем, в соответствии с принципом исключения Паули, в каждом из таких состояний могут находиться лишь два электрона, различающиеся спином (направленный вверх и направленный вниз). Энергия электрона пропорциональна k^2 , вследствие чего точки, соответствующие такому значению энергии в показанной на рис. 2.2 обратной решетке (k -пространстве) при всех возможных комбинациях k_x, k_y, k_z , должны располагаться в узком слое сферы с радиусом $k = k_{max}$. С другой стороны, разность между двумя последовательными значениями каждой из компонент k_i ($i = x, y, z$) всегда составляет величину $2\pi/L$, вследствие чего каждое из разрешенных состояний с набором k (k_x, k_y, k_z) будет занимать в k -пространстве объем:

$$\left(\frac{2\pi}{L}\right)^3 = \frac{(2\pi)^3}{V}, \quad (2.34)$$

где V — объем кристалла. Таким образом, полное число электронных состояний в промежутке значений $(k, k + dk)$ описываемого k -пространства (рис. 2.2) равно

$$2 \frac{4\pi k^2 dk}{(2\pi)^3 / V} = \frac{Vk^2 dk}{\pi^2}, \quad (2.35)$$

где коэффициент 2 возникает из-за необходимости учесть спин электронов. Поскольку из уравнения (2.33) следует $E = E(\vec{k})$, мы можем сразу выписать функцию распределения плотности состояний системы:

$$\rho(E) = \frac{4\pi}{\hbar^3} (2m)^{3/2} E^{1/2}. \quad (2.36)$$

В соответствии со сказанным выше (см. раздел 1.6), функция $\rho(E)$ пропорциональна квадратному корню из значения

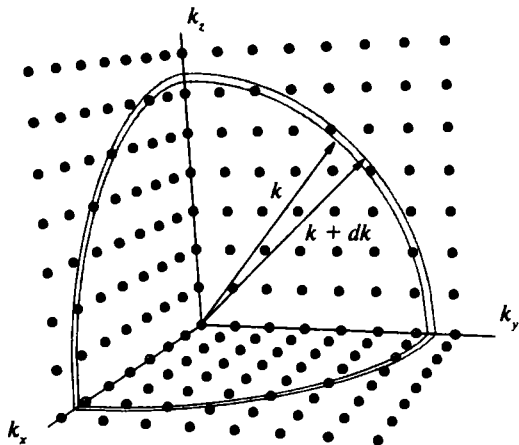


Рис. 2.2. Сферический слой между значениями радиуса k и $k + dk$ в k -пространстве, соответствующий энергиям частиц в диапазоне $(E, E + dE)$

энергии. Умножив $\rho(E)$ на функцию (2.17), соответствующую вероятности заполнения состояния, мы получим функцию распределения электронов по энергии в окончательном виде (рис. 2.3). Очевидно, что при 0 К заняты лишь все состояния до E_F , а при $T > 0$ К некоторые из электронов начинают заполнять уровни выше E_F .

2.4. Теорема Блоха

Теорема Блоха (1928) относится к волновым функциям электронов внутри кристалла и построена на очевидном предположении, что кулоновский потенциал в твердых телах носит периодический характер, т. е. потенциальная энергия $V(\vec{r})$ для электронов внутри кристалла имеет вид

$$V(\vec{r}) = V(\vec{r} + \vec{R}_n), \quad (2.37)$$

где через \vec{R}_n обозначен вектор произвольной трансляции (пространственного переноса) вдоль кристаллографической решетки, т. е. $\vec{R}_n = n_1 \vec{a}_1 + n_2 \vec{a}_2 + n_3 \vec{a}_3$ ($\vec{a}_1, \vec{a}_2, \vec{a}_3$ — единичные векторы (орты) кристаллической решетки).

Теорема Блоха устанавливает, что волновая функция электронов в кристалле $\psi_{\vec{k}}(\vec{r})$, определенная уравнением Шрёдингера

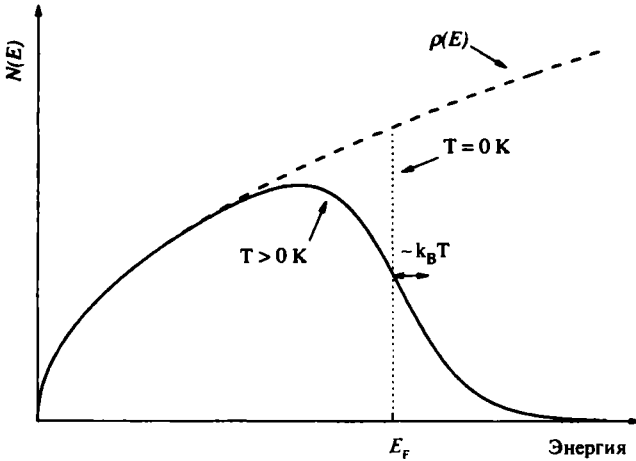


Рис. 2.3. Плотность заполненных электронных состояний как функция энергии

(2.10), может быть представлена в виде произведения плоской волны на некоторую функцию $u_{\vec{k}}(\vec{r})$, периодичность которой совпадает с периодичностью описываемой кристаллической решетки, т. е.

$$\psi_{\vec{k}}(\vec{r}) = e^{i\vec{k} \cdot \vec{r}} u_{\vec{k}}(\vec{r}), \quad (2.38)$$

где

$$u_{\vec{k}}(\vec{r}) = u_{\vec{k}}(\vec{r} + \vec{R}_n). \quad (2.39)$$

Электронные волновые функции в виде (2.38) называют функциями Блоха. При этом следует особо отметить, что сами функции Блоха не являются периодическими функциями, поскольку другой входящий в это произведение член описывает плоскую волну. Легко показать, что периодичностью решетки обладает лишь функция плотности вероятности $|\psi_{\vec{k}}|^2$. Еще одним интересным свойством волновых функций, связанных с теоремой Блоха, является соотношение

$$\psi_{\vec{k}}(\vec{r} + \vec{R}_n) = e^{i\vec{k} \cdot \vec{R}_n} \psi_{\vec{k}}(\vec{r}), \quad (2.40)$$

которое может быть сразу получено из уравнений (2.38) и (2.39).

2.5. Электроны в кристаллических твердых телах

2.5.1. Модель почти свободных электронов

Рассмотренная в предыдущем разделе модель свободных электронов в твердых телах очень широко и эффективно используется при описании многих электрических характеристик металлов, однако она все же слишком проста для описания более сложных свойств твердых тел (например, электронных или оптических характеристик полупроводников). Естественным развитием этой теории стала модель почти свободных электронов, используемая в основном для описания кристаллов, поскольку в ней предполагается, что внутри твердого тела электроны двигаются под воздействием периодического потенциала, как показано на рис. 2.4 для одномерного случая. На рис. 2.4, *a* представлено изменение внутреннего потенциала вдоль линии, соединяющей ядра атомов, объединенных в трехмерный кристалл, где (как следует ожидать) значения кулоновского потенциала стремятся к $-\infty$ в точках, соответствующих положению ядер. Разумеется, большая часть электронов в кристалле двигается в других условиях или под некоторым углом к кристаллографическим направлениям. Например, для движения между двумя плоскостями, образованными заряженными атомами, потенциал имеет вид на рис. 2.4, *б*.

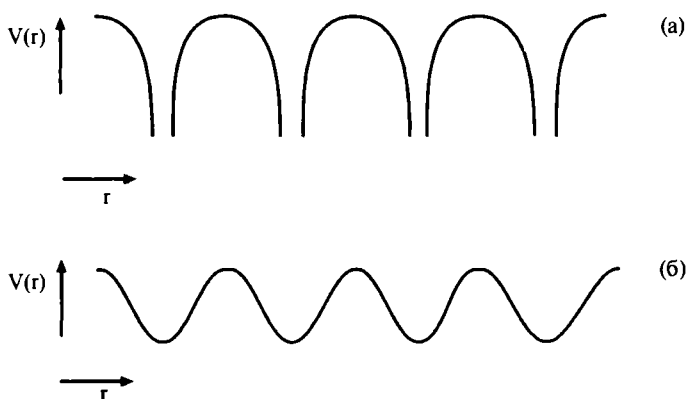


Рис. 2.4. Электростатический потенциал для простейшей кубической решетки. *a* — потенциал вдоль линии, соединяющей ядра атомов; *б* — потенциал вдоль направления между плоскостями ионов

Наличие периодичности в строении кристаллов приводит к тому, что разрешенные значения энергии для электронов также образуют некоторые области (или зоны), в то время как целые диапазоны значений оказываются для них запрещенными. В модели почти свободных электронов периодический потенциал рассматривается в качестве малого возмущения к гамильтониану, соответствующему свободному движению электронов. Приближение является грубым, однако эта простая модель позволяет получать очень важные качественные закономерности для описываемых систем. Мы не будем вдаваться в подробное описание модели (читатель может найти его во многих известных учебниках физики твердого тела [2]), однако приведем ее основные результаты. Прежде всего, по-прежнему вследствие периодичности решетки волновая функция электрона ψ_k должна иметь вид функции Блоха

$$\psi_k = u_k(x) e^{ikx}, \quad (2.41)$$

где функции $u_k(r)$ связаны с этой периодичностью (для простоты приведенная формула относится к одномерной системе с постоянной решеткой a).

Поскольку периодический потенциал внутри кристалла рассматривается в качестве возмущения, мы можем использовать упоминавшуюся выше квантово-механическую теорию возмущений (см. раздел 2.2.4). Применяя разработанный в теории возмущений математический аппарат, легко получить, что разрывы в значениях энергии соответствуют значениям

$$k = \frac{\pi}{a} n, \quad n = \pm 1, \pm 2, \dots, \quad (2.42)$$

причем ширина разрыва (так называемая энергетическая щель или запрещенная зона) пропорциональна коэффициентам $|V_n|^2$ разложения потенциала в ряд Фурье, т. е.

$$V(x) = \sum_{n=-\infty}^{+\infty} V_n e^{i \frac{2\pi}{a} nx}. \quad (2.43)$$

Энергетические щели или запрещенные зоны для одномерной модели показаны на рис. 2.5. Рисунок представляет собой образец широко используемого представления в так называемых

мой схеме расширенных зон, в которой проявляется подобие получаемых зон $E = E(k)$ параболам для свободного электрона, за исключением значений k , определяемых решениями уравнения (2.42), при которых и возникают запрещенные зоны. На рис. 2.6, а приводится представление в схеме так называемых приведенных зон, которое вытекает из периодичности волновой функции в обратном пространстве. Это представление является наиболее распространенным, и оно может быть получено из рис. 2.5 переносом вдоль оси k на величины, кратные $2\pi/a$. На рис. 2.6, б представлена и схема повторяющихся зон, которая применяется очень редко, поскольку дает мало информации.

2.5.2. Приближение сильной связи

Описанная выше модель почти свободных электронов очевидно неприменима к описанию непроводников (т. е. изоляторов) типа алмаза и т. п. Для таких систем было разработано приближение сильной связи, когда (в отличие от модели почти свободных электронов) сразу предполагается, что основную долю общей энергии создает потенциальная энергия электронов в атоме, т. е. предполагается что волновые функции электронов даже двух соседних атомов перекрываются незначительно. Приближе-

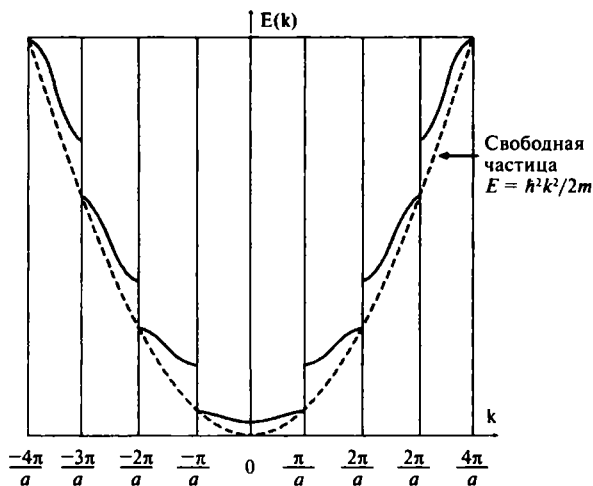


Рис. 2.5. Структура энергетических полос в модели почти свободных электронов

ние сильной связи (предложенное, кстати, тоже Блохом) очень напоминает метод линейной комбинации атомных орбиталей (МЛКАО), используемый для описания молекулярных связей, и является достаточно эффективным для расчета движения электронов в непроводящих веществах или на внутренних электронных оболочках металлических атомов. В методе предполагается,

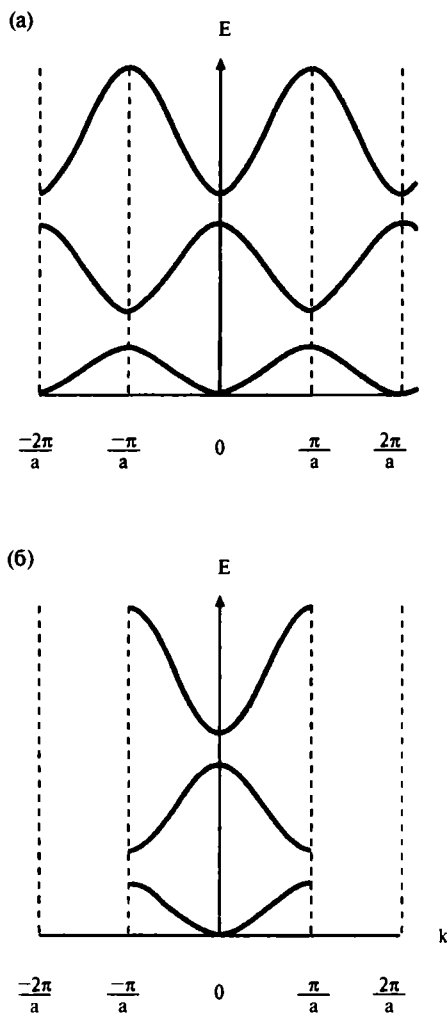


Рис. 2.6. Зависимость энергии от k в схеме повторяющихся зон (а) и в схеме приведенных зон (б)

что с самого начала известны электронные волновые функции $\phi_0(\vec{r} - \vec{R}_n)$, где через \vec{R}_n обозначен главный вектор трансляции рассматриваемой кристаллической решетки. Затем предполагается, что электронная волновая функция в кристалле представляет собой линейную комбинацию атомных орбиталей, т. е.

$$\phi_{\vec{k}}(\vec{r}) = \sum_n c_n \phi_0(\vec{r} - \vec{R}_n), \quad (2.44)$$

где суммирование ведется по всем точкам решетки. Такая волновая функция должна удовлетворять условиям теоремы Блоха и, следовательно, коэффициенты C_n должны иметь вид $e^{i\vec{k} \cdot \vec{R}_n}$, из чего следует

$$\phi_{\vec{k}}(\vec{r}) = \sqrt{N} \sum_n e^{i\vec{k} \cdot \vec{R}_n} \phi_0(\vec{r} - \vec{R}_n), \quad (2.45)$$

где N — число атомов в решетке, а коэффициент $N^{1/2}$, как легко показать, вводится для нормировки.

Использование приближения сильной связи предполагает, что реальный потенциал $V(\vec{r})$, действующий на электрон в твердом теле, очень близок к потенциалу $V_0(\vec{r})$, создаваемому изолированным атомом, вследствие чего гамильтониан для электрона в кристалле может быть записан в виде

$$H = -\frac{\hbar^2}{2m_e} \nabla^2 + V_0(\vec{r} - \vec{R}) + V'(\vec{r} - \vec{R}). \quad (2.46)$$

Можно показать [3], что решение так называемой «возмущенной задачи» для гамильтониана (2.46) с волновыми функциями вида (2.45) может быть записано в виде

$$E(\vec{k}) = E_0 - \alpha - \beta \sum_m e^{i\vec{k} \cdot (\vec{R}_m - \vec{R}_e)}, \quad (2.47)$$

где суммирование ведется по всем ближайшим состояниям. Коэффициенты α и β в уравнении (2.47) представляют собой так называемые интегралы перекрытия, которые часто используются в качестве подгоночных параметров.

В качестве примера мы используем уравнение (2.47) для описания простой кубической решетки с постоянной a и рассмотрим систему из шести ближайших атомов, для которых справедливы соотношения

$$\vec{R}_m - \vec{R}_n = (\pm a, 0, 0), (0, \pm a, 0), (0, 0, \pm a), \quad (2.48)$$

и, следовательно,

$$\vec{E}(\vec{k}) = E_0 - \alpha - 2\beta(\cos k_x a + \cos k_y a + \cos k_z a). \quad (2.49)$$

Стоит сразу отметить, что ширина запрещенной зоны равна 12β , т. е. зависит от интеграла перекрытия, который учитывает лишь перекрытие волновых функций ближайших соседей, вследствие чего очень резко уменьшается с ростом межатомных расстояний. Минимальная энергия соответствует случаю $k = 0$, а максимальная — граничным условиям ($k = \pi/a$, $k_y = \pi/a$, $k_z = \pi/a$) зоны Бриллюэна. Решения уравнения (2.49) могут быть разложены в ряд в окрестности точки $\vec{k} = (k_x, k_y, k_z) = 0$, что дает

$$E(k) = E_0 - \alpha - 6\beta + \beta k^2 a^2. \quad (2.50)$$

В окрестности точки $k = 0$ энергия зависит от k по квадратичному закону, а эквипотенциальная поверхность представляет собой сферу. Уравнение (2.50) позволяет вычислить эффективную массу электрона (см. раздел 2.6.2) в соответствии с соотношением $m^* = \hbar^2/2\beta a^2$ (отметим, что эффективная масса m^* увеличивается с уменьшением величины интеграла перекрытия β). Для изолированного атома значение m^* , как и следует из определения, стремится к бесконечности, поскольку создаваемые электрическими полями внешние силы просто не могут обеспечивать перенос или передачу электрона от одного узла кристаллической решетки к другому.

2.6. Динамика электронов в энергетических зонах

2.6.1. Уравнение движения

Выше уже упоминалось, что волновая функция свободного электрона представляет собой плоскую бегущую волну, что соответствует просто движению свободной частицы с точно определенным импульсом $p = \hbar k$. Из принципа неопределенности Гейзенберга сразу следует, что волна, определяемая (2.14), не может содержать информации относительно пространственной

локализации электрона. Поэтому при любой попытке описать положение и импульс электрона в кристалле мы должны пользоваться представлением о волновом пакете. Обычно волновые пакеты формируются линейной суперпозицией плоских бегущих волн (с соответствующими волновыми векторами) в некотором малом интервале Δk около среднего значения k , т. е. полная волновая функция имеет вид

$$\psi(x, t) \propto \int_{k-\Delta k/2}^{k+\Delta k/2} c(k) e^{i(kx - \omega t)} dk, \quad (2.51)$$

где в общем случае с учетом рассеяния среды $\omega = \omega(k)$. Волновой пакет движется с групповой скоростью v_g , определяемой уравнением

$$v_g = \frac{\partial \omega}{\partial k}, \quad (2.52)$$

которая в принципе отличается от фазовой скорости ($v = \omega/k$) плоских волн.

Очевидно, что в модели почти свободных электронов, описанной в предыдущем разделе 2.5.1, волновая функция электрона, движущегося внутри трехмерного кристалла, должна разлагаться в ряд Фурье:

$$\psi = \sum_k c(k, t) e^{i(\vec{k} \cdot \vec{r} - \omega t)}, \quad (2.53)$$

где значения вектора \vec{k} определяются периодическими граничными условиями. Обобщая уравнение (2.52) для трехмерного случая, легко получить для групповой скорости выражение

$$\vec{v}_g = \vec{\nabla}_{\vec{k}} \omega(\vec{k}) = \frac{1}{\hbar} \vec{\nabla}_{\vec{k}} E(\vec{k}), \quad (2.54)$$

где функция $E = E(\vec{k})$ выражает зависимость между энергией и импульсом в рассматриваемой энергетической зоне.

Предположим далее, что на электрон внутри кристалла действует некоторая внешняя сила (например, электрическое поле \vec{F}). В простейшем одномерном случае элементарная работа δE , производимая при воздействии на электрон с зарядом e поля \vec{F} за промежуток времени δt , равна

$$\delta E = -eF v_g \delta t, \quad (2.55)$$

что с учетом уравнения (2.54) может быть переписано в виде

$$\delta E = \left(\frac{dE}{dk} \right) \delta k = \hbar v_g \delta k. \quad (2.56)$$

В трехмерной системе для действующей на электрон силы можно выписать соотношение

$$-e\vec{F} = \hbar \frac{d\vec{k}}{dt}, \quad (2.57)$$

называемое уравнением движения для электрона. Отметим, что оно весьма похоже на формулировку второго закона Ньютона для свободного электрона, если принять, что импульс электрона \vec{p} имеет вид

$$\vec{p} = \hbar \vec{k}. \quad (2.58)$$

На самом деле момент, определяемый уравнением (2.58), не может считаться истинным, поскольку рассматриваемый в задаче электрон подвергается воздействию не только внешнего электрического поля, что предполагается при выводе из уравнения (2.57), но и внутренних кулоновских сил ионов кристаллической решетки. С точки зрения физики это означает, что мы никогда не можем рассматривать полностью изолированный, отдельный электрон внутри кристалла, без учета окружения. Поэтому импульс, определенный уравнением (2.58), обычно и называют квазиимпульсом (или *crystal momentum*), а его величина свидетельствует именно о взаимодействии электрона со всеми остальными частицами системы, включая электроны и фононы (см. раздел 2.8). Смысл уравнения (2.58) заключается в сохранении энергии и квазиимпульса при всех таких взаимодействиях.

2.6.2. Эффективная масса

Для того чтобы определить величину, которую, собственно, следует подставить в выражение для массы частицы, мы можем воспользоваться соотношениями между силой и ускорением

для предложенной выше ситуации с электроном в кристаллической решетке, на который действует внешнее электрическое поле F . В соответствии с уравнениями (2.52) и (2.57), выражение для ускорения в одномерной задаче имеет вид

$$\frac{d^2v}{dt^2} = \frac{1}{\hbar} \frac{d^2E}{dk^2} = \frac{1}{\hbar} \frac{d^2E}{dk^2} \frac{dk}{dt} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} (-eF), \quad (2.59)$$

из чего можно легко определить эффективную массу m^* электрона в кристалле:

$$\frac{1}{m^*} \equiv \frac{1}{\hbar^2} \frac{d^2E}{dk^2}. \quad (2.60)$$

Отметим, что величина m^* всегда пропорциональна искривленности энергетической зоны («гладкости» зоны в k -пространстве), т. е. меньшей кривизне соответствует большая эффективная масса, а в трехмерном кристалле массе электрона соответствует тензор эффективной массы, имеющий компоненты

$$\left(\frac{1}{m^*} \right)_{ij} \equiv \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \partial k_j}. \quad (2.61)$$

Особо следует подчеркнуть, что в кристалле сила и ускорение, вообще говоря, могут не совпадать по направлению. В простейшем случае, когда совпадают все три эффективные массы, соответствующие главным осям тензора в уравнении (2.61), в результате чего выражение для m^* становится эквивалентным (2.60) для одномерного кристалла, т. е. m^* представляет собой скаляр. Это произойдет, например, при параболической зависимости E от k по каждой из осей в k -пространстве, описываемой уравнением

$$E(\vec{k}) = E_0 + \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2 + k_z^2). \quad (2.62)$$

В этом случае, очевидно, что эквипотенциальные поверхности $E = E(\vec{k})$ представляют собой сферы.

В соответствии с определением эффективной массы легко убедиться, что для показанной на рис. 2.7, *a* зависимости E от

волнового вектора эффективная масса m^* будет описываться кривыми (рис. 2.7, б). Отметим, что значение m^* положительно на дне энергетической зоны, но отрицательно (кривизна противоположного знака) вблизи границ зоны, когда $k = \pm \pi/a$. Физически это означает, что если электрон приближается к границе зоны $k = \pi/a$ (например, слева), то он перестает ускоряться электрическим полем или, говоря иными словами, он начинает отдавать решетке больший импульс, чем получает от внешнего поля. В предельном случае на самой границе зоны (при $k = \pm \pi/a$) электрон просто отражается по брэгговскому механизму.

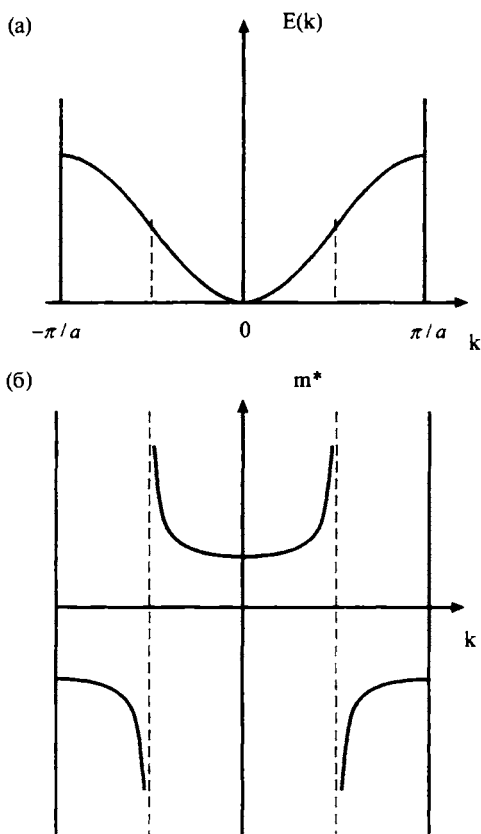


Рис. 2.7. Энергетическая зона для одномерной атомной цепочки (а); энергетическая масса, соответствующая такой энергетической зоне (б)

2.6.3. Дырки

В полупроводниках очень часто приходится иметь дело с задачами, связанными с описанием поведения электронов в почти заполненных зонах, когда очень удобно пользоваться понятием «дырка». Для простоты давайте рассмотрим показанную на рис. 2.8 низшую зону, в которой заполнены все состояния, исключая одно-единственное, соответствующее электрону с волновым вектором \vec{k}_e . Полный волновой вектор \vec{k}_T такой зоны может быть записан в виде волнового вектора полностью заполненной зоны, из которого вычитается волновой вектор отсутствующего электрона, т. е. в виде

$$\vec{k}_T = \sum_k \vec{k}_i - (\vec{k}_e). \quad (2.63)$$

Суммирование в (2.63) осуществляется по всей зоне, и сумма должна равняться нулю для любого кристалла, обладающего инверсной симметрией, поскольку в этом случае такой же симметрией должна обладать зона в k -пространстве, т. е. каждому занятому состоянию с \vec{k}_i должен соответствовать вектор $-\vec{k}_i$ с обратным знаком. Поэтому из уравнения (2.63) следует

$$\vec{k}_T = -\vec{k}_e, \quad (2.64)$$

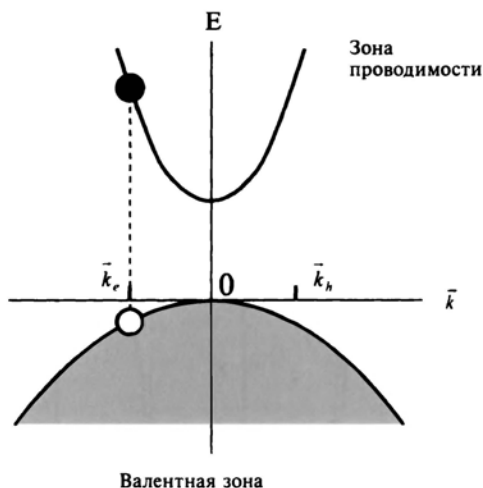


Рис. 2.8. Дырка (отсутствие электрона) в валентной зоне полупроводника с прямыми переходами

и состоянию с отсутствующим электроном в зоне может быть приписана новая частица с названием «дырка», описываемая волновым вектором

$$\vec{k}_h = -\vec{k}_e \quad (2.65)$$

и импульсом

$$\vec{p}_h = -\hbar\vec{k}_e = \hbar\vec{k}_h. \quad (2.66)$$

Поскольку понятие «дырка» вводится для заполненной зоны, в которой отсутствует один электрон, естественно приписать такой дырке положительный заряд. Такая ситуация весьма часто встречается в полупроводниках, когда, например, электрон под воздействием падающего фотона (как показано на рис. 2.8) перескакивает из нижней зоны в более высокую, оставляя свое энергетическое состояние свободным. При повышении энергии фотона могут рождаться дырки с большей энергией, следовательно, энергия дырки возрастает при переходе к более низким состояниям зоны. Кинетическая энергия дырки E_h положительна, и из соотношения $E_h = \hbar^2 k^2 / 2 m^* \hbar$ следует, что ее эффективная масса также положительна. В целом стоит просто отметить, что теория дырок является очень популярной и полезной, поскольку она позволяет очень легко описывать почти заполненные зоны, учитывая в них лишь небольшое число дырок, а не огромное число заполненных электронных состояний.

2.7. Колебания решетки

В этом разделе мы дадим краткое описание колебаний в периодических системах типа кристаллов. В теории твердого тела читатель, возможно, уже сталкивался с «адиабатическим приближением», позволяющим разделить характеристики материала, связанные с электронами (электропроводности и т. п.), от связанных с колебаниями атомов решетки (теплопроводность и т. п.). Рассмотрим механические или звуковые волны, распространяющиеся в объеме твердого тела. При длинах волны λ , значительно превышающих постоянную решетки кристалла, среда распространения может считаться упругой, причем необязательно изотропной. Однако при значениях λ , сравни-

мых или меньших, чем размеры кристаллических ячеек, мы должны как-то учитывать и особенности кристаллической структуры.

Для описания колебаний атомов решетки обычно используют гармоническое приближение. При этом обычно соседние атомы притягиваются друг к другу, образуя так называемые межатомные связи (например, ковалентные, ионные, силы Ван-дер-Ваальса и т. п.). Однако по мере сближения атомов начинается взаимодействие между входящими в их состав электронами, вследствие чего возникают и значительные силы отталкивания (связанные, конечно, с принципом исключения Паули), причем эти силы быстро нарастают с уменьшением расстояния между атомами. Одним из наиболее известных и широко используемых потенциалов межатомного взаимодействия является так называемый потенциал Леннарда — Джонса, который имеет минимум при некотором равновесном межатомном расстоянии r , совпадающем с постоянной решетки a . При близких к a значениях r потенциал $V(r)$ очень близок к параболическому (гармоническому) потенциалу, поэтому предполагается, что гармоническое приближение является особо точным в области $r = a$.

2.7.1. Одномерная решетка

Простейшей моделью для изучения колебаний периодической решетки является, конечно, линейная одноатомная цепочка, представляющая собой просто цепь связанных атомов массой m , находящихся на равновесном расстоянии a друг от друга, взаимодействие между которыми носит гармонический характер, как показано на рис. 2.9, *a*. Обозначив через u_n смещение n -го атома от положения равновесия, мы можем записать для каждого атома цепочки уравнение движения (при гармоническом взаимодействии с двумя ближайшими соседями) в виде

$$m \frac{d^2 u_n}{dx^2} = c [(u_{n+1} - u_n) - (u_n - u_{n-1})] = c (u_{n+1} + u_{n-1} - 2u_n). \quad (2.67)$$

Решение для предлагаемой системы уравнений мы будем искать в виде бегущих плоских волн с амплитудой A , частотой ω и волновым вектором k :

$$u_n = Ae^{i(kx_n - \omega t)}, \quad (2.68)$$

где $x_n = na$ соответствует равносному положению n -го атома. Подставляя выражения (2.68) для смещений u_n , u_{n+1} и u_{n-1} в уравнения движения (2.67), можно после несложных алгебраических преобразований [2] получить выражение для частоты

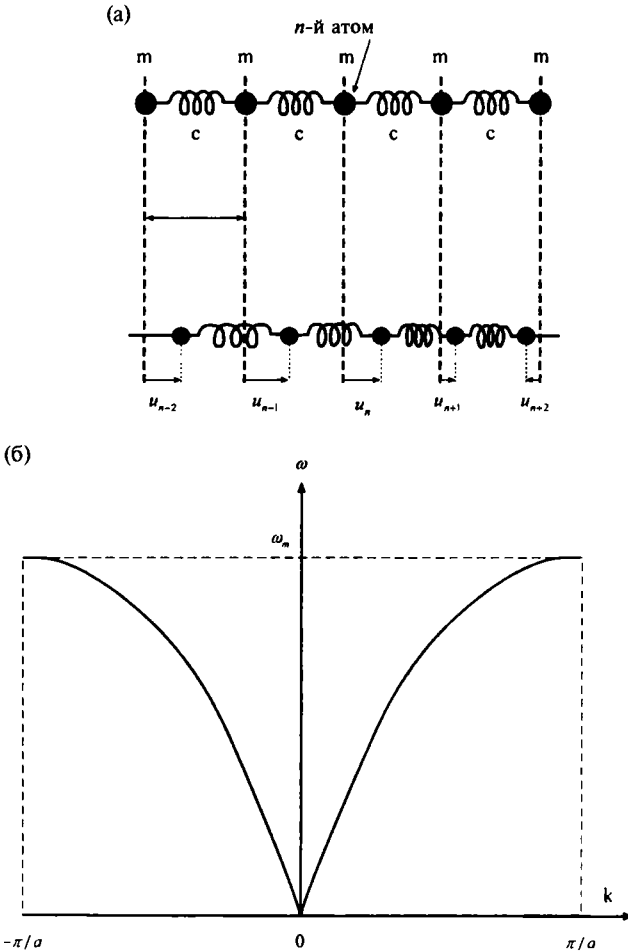


Рис. 2.9. (а) Одномерная цепочка из одноатомных узлов в равновесном состоянии (вверху) и при смещениях атомов относительно равновесного положения (внизу); (б) дисперсионное соотношение для описываемой системы

$$\omega = \left(\frac{4c}{m} \right)^{1/2} \left| \sin \frac{ka}{2} \right|, \quad (2.69)$$

которое называют дисперсионным соотношением для колебаний одномерной решетки (соответствующая кривая приведена на рис. 2.9, б). Одним из важных следствий этого уравнения (которое трудно представить в привычных непрерывных средах) является то, что в цепочке атомов не могут распространяться волны с частотой выше некоторого предельного значения $2(c/m)^{1/2}$. Эта частота в уравнении (2.69) соответствует вектору $k = \pi/a$, т. е. значению $\lambda = 2a$. Ее физический смысл соответствует порогу брэгговского отражения для электронов в периодической структуре (см. раздел 2.6.2), а математически это значение соответствует появлению вместо бегущих волн так называемых стоячих волн, которые не переносят энергию. Этот результат не является неожиданным, поскольку упомянутое условие $k = \pi/a$ соответствует нулевой групповой скорости. Из уравнения (2.69) также сразу следует, что при $ka \rightarrow 0$ частота ω начинает зависеть от k линейно, а групповая скорость совпадает с фазовой скоростью волны, так как обе величины становятся равными $v_s = a(c/m)^{1/2}$. Этот результат тоже является естественным, поскольку при $ka \rightarrow 0$ имеем $a/\lambda \rightarrow 0$, т. е. длина волны становится значительно больше межатомного расстояния, вследствие чего среда может вновь рассматриваться в качестве непрерывной. В этом случае скорость v_s может считаться аналогом скорости звука в веществе.

Используя такой же подход, как и в случае описания электронов в периодических кристаллах (раздел 2.3), можно ввести периодические граничные условия для волновых решений, определяемых уравнением (2.68). Физически введение таких условий соответствует установлению фиксированных связей или ограничивающих сил для первого и последнего атомов в цепочке, с тем чтобы эти атомы двигались одинаково. Получаемые при этом разрешенные значения вектора k (см. уравнения (2.32) имеют вид

$$k = \frac{2\pi}{L} n, \quad n = 0, \pm 1, \pm 2 \dots \pm N, \quad (2.70)$$

где параметры L и N относятся к длине цепочки и полному числу атомов соответственно (т. е. $L = Na$).

Еще одним очень важным следствием дисперсионного соотношения является то, что величина ω не меняется при изменении k на значения, кратные $2\pi/a$, вследствие чего (как и для электронов) можно ограничиться рассмотрением значений k , принадлежащих лишь первой зоне Бриллюэна, т. е.

$$-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}. \quad (2.71)$$

Более сложной является задача о колебаниях решетки, содержащей в каждой из элементарной ячейки больше одного атома. Рассмотрим, например, двухатомную линейную цепочку, показанную на рис. 2.10, *a*, каждый узел которой содержит два атома разных видов (с массами M и m , где $M > m$). Основное усложнение задачи по сравнению с предыдущей обусловлено тем, что амплитуды колебаний атомов M и m не равны. Обозначив одну из них через A , мы получим для другой величину αA , причем коэффициент α представляет собой в общем виде комплексное число, которое зависит как от соотношения амплитуд, так и от разности фаз. Такая задача, конечно, не может быть решена простым развитием метода для одноатомной цепочки, а ее решение становится значительно более сложным. Тем не менее, расчеты могут быть проведены в явном виде, так что читатель может найти решение во многих, даже достаточно простых учебниках физики твердого тела.

Для двухатомной линейной цепочки дисперсионное соотношение принимает вид [2]

$$\omega^2 = \frac{c(m+M)}{Mm} \pm k \left[\left(\frac{M+m}{Mm} \right)^2 - \frac{4}{Mm} \sin^2 \left(\frac{ka}{2} \right) \right]^{1/2}, \quad (2.72)$$

показанный на рис. 2.10, *б*. Дисперсионная зависимость в данном случае описывается двумя ветвями (верхней и нижней), соответствующими знакам \pm в решениях уравнения (2.72). Как и для одноатомной цепочки, стоит проанализировать поведение решений при значениях k вблизи центра зоны ($k = 0$) и на границах ($k = \pi/a$). При $ka \ll 1$ упоминавшийся выше коэффициент α равен 1 для нижней (так называемой акустической ветви) и $\alpha = -M/m$ для верхней (оптической ветви) дисперсионного соотношения. При $\alpha = 1$ соседние атомы M и m колеблются

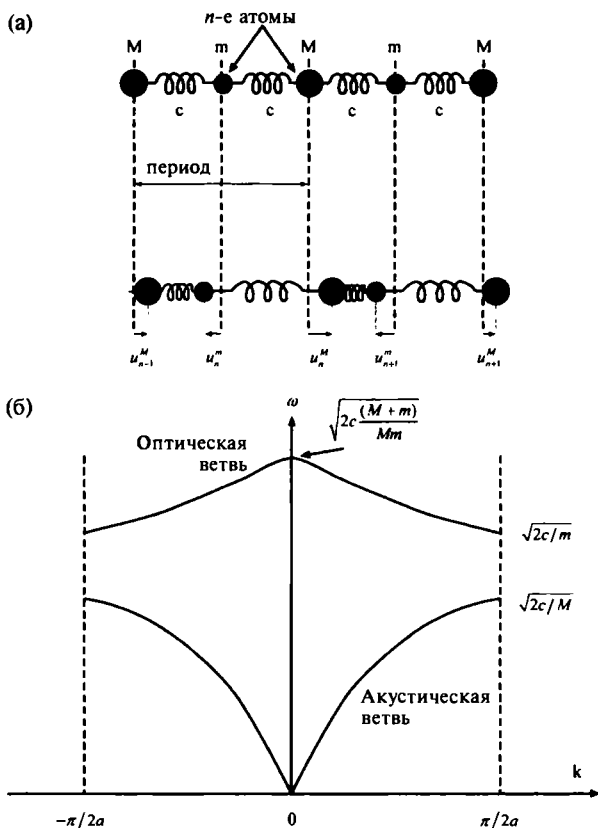


Рис. 2.10. (а) Одномерная цепочка из двухатомных узлов в равновесном состоянии (вверху) и при смещениях атомов относительно равновесного положения (внизу); (б) дисперсионное соотношение для описываемой системы

практически с одинаковой фазой, подобно звуковым волнам в твердых телах при выполнении условия $\lambda \gg a$. При $\alpha = -M/m$ колебания атомов разных видов не совпадают по фазе, и верхняя ветвь имеет некую максимальную частоту, равную

$$\omega_{\max}^+ = \left(\frac{2c(M+m)}{Mm} \right)^{1/2} \quad (2.73)$$

Колебательные моды этой ветви называют оптическими, поскольку соответствующие частоты относятся к инфракрас-

ной области, так что в кристаллах с сильной ионной связью (например, в NaCl) электромагнитным излучением можно возбуждать оптические моды, причем фазы движения отрицательных и положительных ионов при этом не совпадают.

2.7.2. Трехмерная решетка

Полученные в разделе 2.7.1 дисперсионные соотношения для одномерных решеток могут быть обобщены и на случай трехмерных кристаллов [4]. Число акустических ветвей в трехмерных решетках равно трем: одна ветвь соответствует продольным колебаниям (при которых положения атомов смещаются вдоль цепочки), а две других — поперечным (когда колебания атомов происходят перпендикулярно направлению распространения волны). Таким образом, в трехмерной структуре возникает одна продольная акустическая ветвь (LA) и две поперечные акустические ветви (TA), которые часто являются вырожденными. Хотелось бы особо указать, что к приведенному выше замечанию относительно направления колебаний атомов следует относиться очень внимательно, так как во многих случаях (когда вектор \vec{k} не совпадает с направлением высокой симметрии кристалла) атомные смещения не являются строго параллельными или перпендикулярными вектору \vec{k} .

Как и в одномерном случае, наличие нескольких атомов в элементарной ячейке (обозначим их число через p) приводит к появлению оптических ветвей, количество которых в общем случае составляет $(3p - 3)$. При $p = 2$ (например, для галогенидов щелочных металлов) мы получаем три акустические и три оптические ветви, однако необходимо учитывать, что в высокосимметричных направлениях кристалла две поперечные ветви часто имеют вырожденный характер. На рис. 2.11 схематически показаны дисперсионные соотношения для трехмерных структур. Отсутствие вырождения служит свидетельством анизотропности кристалла. Кроме этого, обычно (подобно тому, как это происходит с электронами в кристаллах) дисперсионные кривые пересекают границу зоны Бриллюэна под прямым углом, хотя наблюдаются и исключения из этого правила, в случае зон Бриллюэна очень сложной формы.

2.8. Фононы

При рассмотрении колебаний решетки мы предполагали до сих пор, что взаимодействие между соседними атомами обусловлено гармоническими упругими силами. Именно при этом допущении были получены уравнения движения атомов типа (2.67), которое похоже на уравнение для изолированного гармонического осциллятора при смещении u_n :

$$m \frac{d^2 u_n}{dt^2} = -c u_n. \quad (2.74)$$

Основная разница заключается в том, что правая часть (2.67) соответствует не просто силе, связанной со смещением атома с номером n , а его взаимодействию с двумя ближайшими атомами, обозначенными номерами $(n + 1)$ и $(n - 1)$. Из классической механики известно, что в этом случае мы можем найти так называемые нормальные координаты системы таких частиц, т. е. найти такую линейную комбинацию u_n , при которой уравнения движения принимают простейший вид (2.74). Очевидным преимуществом использования нормальных координат выступает то, что в этом случае гамильтониан системы становится диагональным, т. е. может быть записан в виде суммы отдельных или несвязанных гамильтонианов для гармонических осцилляторов.

Выполнив преобразования, необходимые для перехода к нормальным координатам, легко сформулировать задачу с квантово-механической точки зрения. Действительно, в результате разложения общего гамильтониана на сумму несвязанных отдельных гамильтонианов, можно легко описать каждое квантово-механическое состояние в виде произведения волновых функций гармонических осцилляторов, соответствующих отдельным нормальным модам. Каждая нормальная мода при этом описывается набором собственных частот ω_k , соответствующих разрешенным состояниям с энергией E_k :

$$E_k = \left(n_k + \frac{1}{2} \right) \hbar \omega_k, \quad n_k = 0, 1, 2, \dots \quad (2.75)$$

Такой подход означает, что каждому состоянию k соответствуют n_k квантов с энергией $\hbar \omega_k$. Такие кванты энергии, возникающие

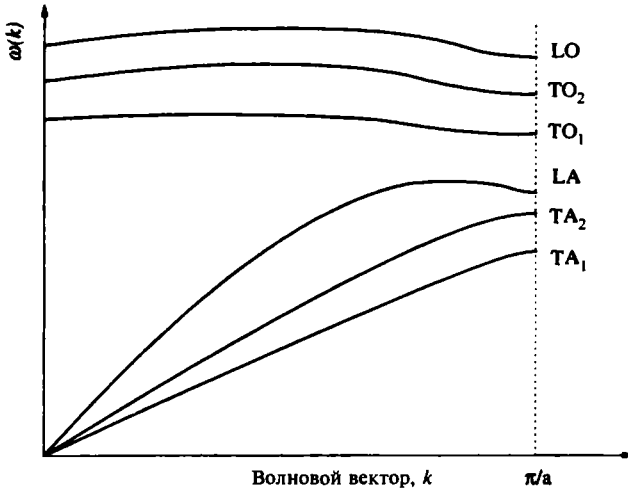


Рис. 2.11. Дисперсионные соотношения для трехмерного кристалла с двумя атомами в элементарной ячейке

в результате колебаний решетки, получили название фононов, вследствие чего обычно при возбуждении моды k от n_k к n_{k+1} говорят о рождении фонона с энергией $\hbar\omega_k$ (аналогично можно говорить и о поглощении или гибели фонона). Использование такой терминологии связано с тем, что фононы являются квазичастицами и их полное число может и не сохраняться. Учитывая сказанное, полную энергию E колебаний решетки можно записать в виде

$$E = \sum_{k,p} \hbar\omega_{k,p} \left(n_{k,p} + \frac{1}{2} \right), \quad (2.76)$$

где суммирование ведется по всем акустическим и оптическим ветвям.

Литература

1. *Merzbacher E.* (1960). *Quantum Mechanics* (Wiley, New York).
2. *Kittel Ch.* (2005). *Introduction to Solid State Physics*, 8th edition (Wiley).
3. *Ibach H., Luth H.* (1991.) *Solid-State Physics* (Springer-Verlag, Berlin).

4. *Elliot S.* (1998). *The Physics and Chemistry of Solids* (Wiley, Chichester).

Дополнительная литература

Some excellent books on solid state physics, at the intermediate level, are the following:

Burns G. (1995). *Solid State Physics* (Academic Press, Boston).

Elliot S. (1998). *The Physics and Chemistry of Solids* (Wiley, Chichester).

Ibach H., Luth H. (1991). *Solid-State Physics* (Springer-Verlag, Berlin).

Kittel Ch. (2005). *Introduction to Solid State Physics*, 8th edition (Wiley).

Singleton J. (2001.) *Band Theory and Electronic Properties of Solids* (Oxford University Press, Oxford).

Задачи и упражнения

1. Энергия Ферми. Предположим, что поведение металлического натрия можно описать в рамках модели свободных электронов. Рассчитайте энергию Ферми (плотность натрия составляет $0,97 \text{ г/см}^3$). Рассчитайте среднюю энергию, приходящуюся на электрон.
2. Энергетические зоны электронов в твердых телах. Покажите, что энергетические зоны в кристаллах, определяемые с использованием модели почти свободных электронов или приближения сильной связи, имеют много общих свойств и особенностей. В частности, покажите, что а) зоны имеют минимумы (для электронных состояний) и максимумы (для дырочных состояний); б) вблизи минимумов и максимумов зависимость зон от k соответствует параболе; в) зоны являются периодическими в обратном пространстве, вследствие чего вся относящаяся к ним информация может быть получена по первой зоне Бриллюэна.
3. Эффективная масса в одномерных твердых телах. В частном случае одномерной решетки энергия электронной зоны может быть записана в виде

$$E(k) = \frac{\hbar^2 k^2}{2m_0^*} - \alpha k^4.$$

Определите величину константы α , эффективную массу, максимальную скорость и соответствующую ей энергию в центре и на границе зоны Бриллюэна (для $k = 0$ и $k = \pi/a$). (Подсказка: начните с вычисления постоянной α , полагая $v_g = 0$ при π/a .)

4. Тензор эффективной массы. Считая известным, что энергетическая зона электронов в простой кубической решетке (с постоянной решетки a) в приближении сильной связи описывается уравнением

$$E = -E_1 (\cos k_x a + \cos k_y a + \cos k_z a),$$

вычислите тензор эффективной массы m_{ij} для центра ячейки в k -пространстве, центра грани и в узле. (Подсказка: начните рассуждения с того, что три точки, в которых вы должны найти эффективную массу, имеют в k -пространстве координаты $k = \pi/a(0, 0, 0)$, $k = \pi/a(0, 0, 1)$ и $k = \pi/a(1, 1, 1)$.)

5. Условие Брэгга для двумерной системы. Покажите, что в квадратной решетке (с одинаковыми атомами и постоянной a) плоская волна удовлетворяет условию брэгговского отражения на границах первой зоны Бриллюэна.
6. Дисперсионные кривые для колебаний решетки. Рассчитайте характеристические частоты для одномерной решетки с двумя молекулами в каждой элементарной ячейке. В качестве примера возьмите полиэтиленовую цепочку ($-\text{CH} = \text{CH} - \text{CH} = \text{CH} -$) с молекулами одинаковой массы M , связанными чередующимися пружинами (характеризующимися упругими постоянными C_1 и C_2) при взаимодействии только с ближайшими ячейками. Постройте дисперсионные кривые для акустических и оптических ветвей. (Подсказка: покажите сначала, что характеристические частоты определяются уравнением:

$$\omega^2 = \frac{C_1 + C_2}{M} \left[1 \pm \left(1 - \frac{4C_1 C_2 \sin^2(ka/2)^{1/2}}{(C_1 + C_2)^2} \right) \right].$$

7. Волновой вектор электрона. Предполагая идеальную, 100%-ю эффективность процесса, оцените изменение волнового вектора при переходе электрона из валентной зоны

в зону проводимости вследствие поглощения фотона. Сравните полученные результаты с типичными размерами первой зоны Бриллюэна для следующих систем: а) голубой лазер CdSe ($E_g = 2,7$ эВ), GaAs ($E_g = 1,6$ эВ) и б) соединение HgCdTe ($E_g = 0,15$ эВ).

ГЛАВА 3

ОБЩИЕ СВЕДЕНИЯ ИЗ ФИЗИКИ ПОЛУПРОВОДНИКОВ

3.1. Введение

В этой главе приводится обзор общих сведений из физики полупроводников, необходимых для понимания поведения полупроводников в мезоскопических системах с низкой размерностью (см. раздел 1.5). Основное внимание будет уделено электронным и оптическим свойствам в различных устройствах (транзисторы, лазеры и т. п.). Глава начинается с описания зонных структур в типичных полупроводниках. Затем описываются методы расчета концентрации электронов и дырок в собственных и примесных полупроводниках, рассматриваются различные механизмы переноса электронов в полупроводниках (как под действием электрического поля, так и вследствие градиента концентрации носителей заряда). Анализ процессов генерации и рекомбинации носителей заряда в электрических полях или при градиенте концентрации приводит к рассмотрению уравнения непрерывности и представлению о времени жизни и диффузионной длине неосновных носителей заряда. Последний раздел главы посвящен оптическим процессам в полупроводниках (особенно поглощению или излучению фотонов и экситонным переходам), поскольку такие процессы лежат в основе действия большинства оптоэлектронных приборов, таких, как лазеры, модуляторы и др.

3.2. Энергетические зоны электронов в типичных полупроводниках

Основным параметром, определяющим электрические, оптоэлектронные и другие свойства полупроводников является ширина энергетической щели или запрещенной зоны, т. е. расстояние между полностью заполненной зоной (называемой валентной) и следующей за ней более высокой энергетической зоной (называемой зоной проводимости), которая может быть пустой или частично заполненной. Для описания интересующих нас процессов, помимо данных по запрещенной зоне, необходимо знать кривизну зон в k -пространстве, определяющую величину эффективной массы носителей заряда (электронов или дырок). Как уже было показано в предыдущем разделе, во многих твердотельных структурах форма зон в k -пространстве может считаться параболической (по крайней мере, в первом приближении), что значительно упрощает описание полупроводников в рамках приближения эффективной массы. Кроме того, многие характеристики полупроводника, особенно относящиеся к возможностям его использования в оптоэлектронике, определяются тем, происходят ли в нем прямые (без изменения квазиимпульса) или непрямые оптические переходы.

Большинство важных в промышленном отношении полупроводников (таких, как соединения $A_{III}B_V$, Si, Ge и т. п.) обладают кубической симметрией, и в них максимум валентной зоны располагается при $\vec{k} = 0$. При этом, однако, минимум зоны проводимости может располагаться при $\vec{k} = 0$ (как в GaAs) или вблизи границы первой зоны Бриллюэна (как в случае Si и Ge). Материалы первого типа называют полупроводниками с прямыми оптическими переходами, так как поглощение или излучение фотонов в них происходит за счет перехода электрона из одной зоны в другую без изменения волнового вектора, т. е. такой переход может быть представлен в виде вертикального переброса в k -пространстве, как показано на рис. 3.1, *a*. В этом случае волновое число (a следовательно, и импульс) электрона при переходе практически не меняются, так как волновое число k фотона ($k = p/h$) пренебрежимо мало по сравнению с соответствующими значениями для электронов. Рассмотрим далее случай, когда электронный переход происходит между максимумом валентной зоны и минимумом зоны проводимости в непрямозонных полупровод-

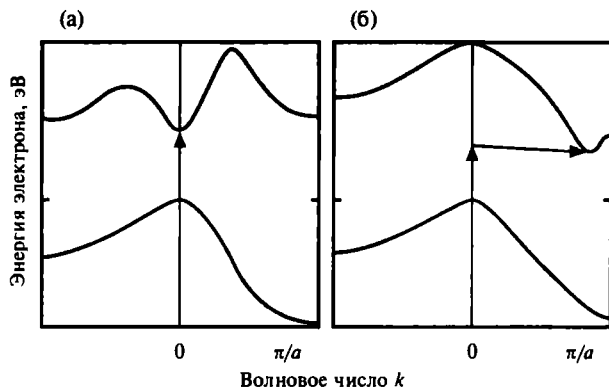


Рис. 3.1. Электронные переходы через запрещенную зону для полупроводников с прямыми (а) и непрямыми (б) оптическими переходами

никах, например в Si и Ge, как показано на рис. 3.1, б. Очевидно, что в этом случае при поглощении или излучении фотона волновой вектор электрона изменяется довольно значительно, почти на величину π/a . Поэтому в таких процессах требуется участие третьих частиц, поглощение или генерация которых позволяет выполнить закон сохранения импульса. Такими третьими частицами являются фононы (см. раздел 2.8), позволяющие обеспечить сохранение импульса. Очевидно, что необходимость третьей частицы в каком-то процессе означает сразу, что вероятность излучения фотона в полупроводниках с непрямыми оптическими переходами намного ниже, чем в полупроводниках с прямыми переходами. Именно поэтому оптоэлектронные приборы, такие, как светоизлучающие диоды или полупроводниковых лазеры обычно создаются на полупроводниках с прямыми оптическим переходами.

Указанные особенности легко прослеживаются для приведенных на рис. 3.2, а и 3.2, б энергетических зон для арсенида галлия и кремния, где энергетическая щель GaAs (полупроводник с прямыми оптическими переходами) составляет $E_G = 1,43$ эВ, в то время как в кремнии (полупроводник с непрямыми оптическими переходами) E_G равно лишь 1,1 эВ. Поскольку соотношение $E = E(\vec{k})$ нельзя отобразить в трехмерном виде, его обычно приводят для одного из направлений высокой симметрии в k -пространстве. Так, например, на рис. 3.2 справа оси ординат относятся к направлению $(1, 0, 0)$, а слева — к направлению $(1, 1, 1)$.

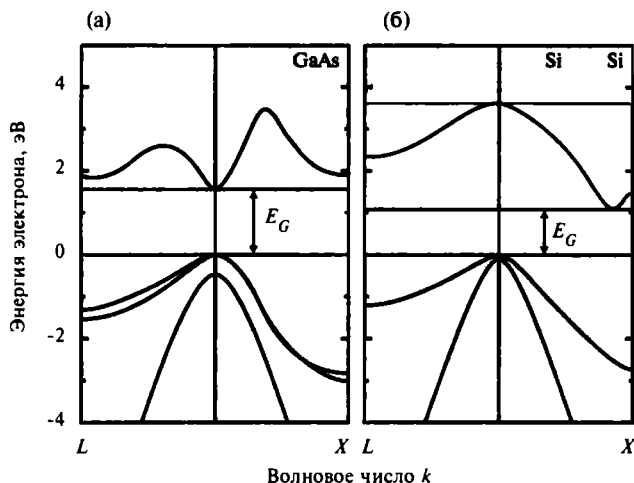


Рис. 3.2. Энергетические зоны для типичных полупроводников: арсенид галлия (слева) и кремний (справа)

Для расчета эффективных масс следует воспользоваться уравнением (2.61) из главы 2, описывающим тензор эффективной массы, что позволяет вычислять ее для движения электрона вдоль любого заданного направления. В случае кристалла GaAs в точке Γ зависимости $E = E(\vec{k})$ для трех основных направлений совпадают, вследствие чего можно воспользоваться следующим приближением:

$$E(k) = \frac{\hbar^2 k^2}{2m_e^*}, \quad (3.1)$$

относящимся к тензору, вырожденному до скаляра. Как показано на рис. 3.2, энергетическая зона в точке Γ обладает большой кривизной, что приводит к малой эффективной массе ($0,066 m_0$). Как видно из уравнения (3.1) для GaAs, поверхности постоянной энергии (эквиэнергетические поверхности, на которых $E(\vec{k}) = \text{const}$) представляют собой сферы.

Аналогичное выражение $E = E(\vec{k})$ для кремния имеет несколько более сложный вид, однако при соответствующем выборе осей координат оно может быть записано в форме

$$E(k) = \frac{\hbar^2}{2} \left(\frac{k_1^2}{m_1^*} + \frac{k_2^2}{m_2^*} \right), \quad (3.2)$$

где k_1 и k_2 относятся к продольной и поперечной компоненте вектора \vec{k} , а величины m_1^* и m_2^* обозначают соответственно эффективную массу электрона в продольном и поперечном направлении. Эти соотношения вытекают из того, что эквивалентные поверхности, описываемые уравнением (3.2), представляют собой эллипсоиды вращения, как показано на рис. 3.3. Из этого же рисунка легко заметить эквивалентность шести направлений ($\pm k_x, \pm k_y, \pm k_z$). Значения эффективной массы электрона в кремнии m_1^* и m_2^* составляют $0,98 m_0$ и $0,066 m_0$ соответственно.

3.3. Собственные и примесные полупроводники

Значения электропроводности полупроводников обычно являются промежуточными между значениями для металлов и изоляционных материалов, и ее характерной особенностью является сильная зависимость от температуры и концентрации примесей. Типичные полупроводники могут представлять собой чистые элементы (кремний, германий), соединения элементов III и V групп периодической таблицы (GaAs, GaP), сочетание элементов II и VI группы (ZnS, CdTe) и т. д. Ширина запрещенной зоны в них может изменяться в пределах от нескольких десятых

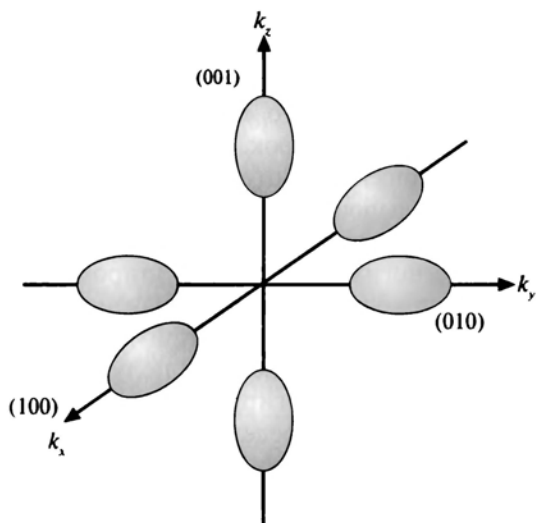


Рис. 3.3. Поверхности постоянной энергии для кремния вблизи минимума зоны проводимости

до примерно 3 эВ. Вещества с шириной зоны от 2 до 3 эВ обычно называют полупроводниками с широкой запрещенной зоной, а с шириной 4 эВ и больше относят к изоляторам.

Собственными или беспримесными полупроводниками называют те, которые отличаются высокой степенью чистоты и совершенства кристаллической структуры. В качестве примера можно указать кремний, являющийся одним из важнейших материалов в электронике. В кристаллическом состоянии атомы кремния расположены в тетраэдрических позициях гранцентрированной кубической решетки, и подобно атомам в решетке алмаза, делят четыре валентных электрона с четырьмя ближайшими соседями, образуя ковалентные связи, как показано на рис. 3.4, а. При температуре 0 К все связи заняты электронами, в результате чего валентная зона оказывается полностью заполненной, в то время как зона проводимости остается совершенно пустой. При повышении температуры выше 0 К некоторые из электронов (как показано на рис. 3.4, б) получают достаточное количество энергии в результате колебаний атомов (тепловой энергии) для разрыва связей и перехода в свободное состояние, причем в этом процессе одновременно возникает и дырка. Понятно, что энергия, необходимая для перехода электрона из валентной зоны в зону проводимости, должна превышать (или по крайней мере равняться) ширине запрещенной зоны. Особо следует отметить, что в собственных полупроводниках каждый переход электрона из валентной зоны в зону проводимости сопровождается обязательным рождением дырки в валентной зоне. В полупроводниках такого типа концентрация электронов всегда равна концентрации дырок. Очевидно также, что фотоны с энергией, превышающей ширину запрещенной зоны, могут также рождать электрон-дырочную пару, что и используется во многих существующих оптоэлектронных структурах.

При создании многих известных и широко распространенных электронных приборов (диодов, транзисторов, солнечных батарей и т. д.) обычно используются примесные полупроводники, которые получают из собственных полупроводников строго контролируемым введением легирующих примесей. Концентрация легирующих примесей определяет электропроводность примесного полупроводника. На рис. 3.5 показана двумерная модель решетки кремния, в которой каждый атом связан с четырьмя ближайшими соседями. Предположим, что в систему вводятся очень небольшие концентрации (речь идет о концентрациях

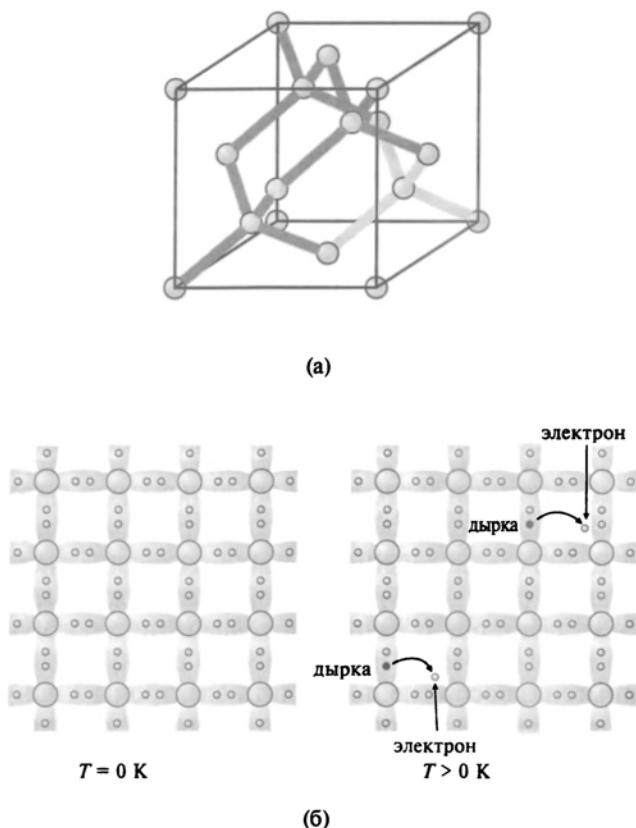


Рис. 3.4. (а) Кристаллическая решетка кремния; (б) при $T > 0$ некоторые связи разрушаются, в результате чего в кристалле появляются одновременно электрон и дырка, которые могут свободно перемещаться в объеме кристалла

порядка одного атома примеси на 10^5 – 10^8 атомов кремния) каких-либо пятивалентных атомов примеси, например, атомов As, Sb и др. На рис. 3.5, а показано поведение примесного атома Sb (имеющего валентность +5) в решетке кремния, если он замещает в ней атом кремния. Четыре электрона атома Sb образуют полагающиеся ковалентные связи, однако пятый электрон в этой ситуации оказывается очень слабо связанным с атомом, поэтому при достаточной температуре такие электроны очень легко покидают атом и начинают свободно передвигаться в кристалле, причем в результате потери электрона атом примеси Sb превращается

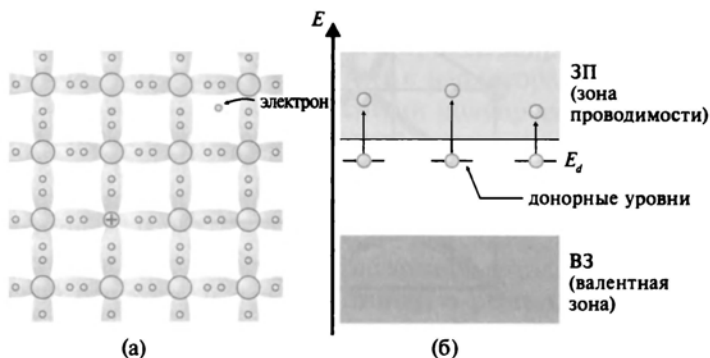


Рис. 3.5. (а) Пятивалентный атом примеси в решетке кремния. (б) Диаграмма энергетических уровней в полупроводнике n -типа (получив энергию E_d , электрон может покинуть атом примеси и свободно перемещаться в объеме кристалла; примесные атомы-доноры при этом остаются неподвижными и положительно заряженными ионами)

в положительно заряженный ион. Как показано на энергетической диаграмме (рис. 3.5, б), такой процесс может быть описан как перенос электрона с некоторого энергетического уровня E_d (лежащего ниже дна зоны проводимости) в зону проводимости. Такие примеси, способные поставлять электроны в зону проводимости, называются донорными, а соответствующий им тип полупроводниковых материалов получил название полупроводников n -типа, поскольку в них концентрация электронов всегда значительно превышает концентрацию дырок ($n \gg p$). Энергия ионизации E_d в данной ситуации может быть оценена по аналогии с атомом водорода, поскольку избыточный пятый электрон притягивается положительно заряженным ионом Sb. Поэтому для оценки E_d может быть использовано уравнение (2.15) из главы 2, относящееся к электронным уровням энергии атома водорода, естественно, после некоторых дополнительных поправок. Прежде всего, необходимо учесть, что рассматриваемая кулоновская система (взаимодействие положительного иона с электроном) не является изолированной, а располагается в некоторой среде, вследствие чего в знаменателе формулы (2.15) следует добавить величину $(\epsilon_r)^2$, где ϵ_r — относительная диэлектрическая проницаемость для конкретного вещества. Кроме этого, массу электрона в формуле следует заменить на введенную выше эффективную массу электрона в данном полупроводнике. С учетом

этих поправок вычисляемые для полупроводников величины E_d составляют около 0,05 эВ, что достаточно близко к экспериментально определяемым значениям.

Рассмотрим далее случай внедрения трехвалентного атома (например, атома бора с валентностью +3) в решетку кремния (рис. 3.6), когда три валентных электрона распределяются среди четырех ближайших атомов Si, но одна из ковалентных связей остается без электрона, в результате чего возникает дефицит одного электрона. Состояние, соответствующее недостатку одного электрона, в точности соответствует описанной выше дырке (см. раздел 2.3.6). При тепловом возбуждении близко расположенные электроны могут смещаться и занять пустующее место, образуя при этом дырку в валентной зоне. Энергия активации такого процесса обычно очень мала, и при комнатных температурах большинство атомов трехвалентных примесей легко получают недостающий электрон и формируют химическую связь, превращаясь при этом в отрицательно заряженные ионы. Естественно, что в таком процессе число возникающих дырок равно числу введенных в структуру атомов примеси. Следовательно, введение в кристалл трехвалентных атомов примеси (их называют просто акцепторами) создает дырки в валентной зоне. На энергетической

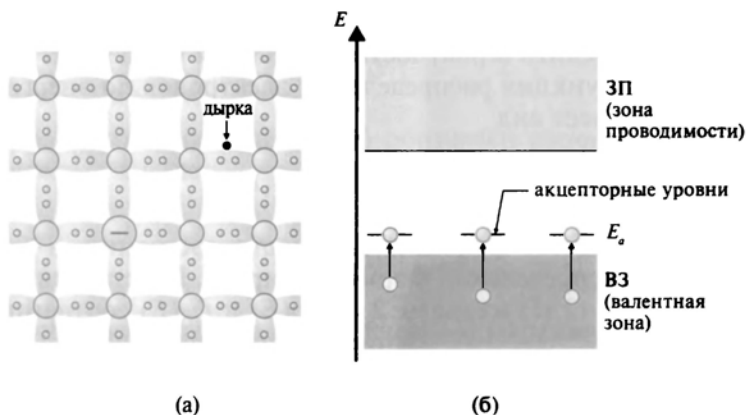


Рис. 3.6. (а) Трехвалентный атом примеси в решетке кремния. (б) Диаграмма энергетических уровней в полупроводнике р-типа (получив энергию E_d , акцепторная примесь может захватить электрон из валентной зоны, превращаясь в отрицательно заряженный неподвижный ион, но одновременно порождая свободную дырку)

диаграмме акцепторные уровни E_a лежат выше потолка валентной зоны, так как процесс подразумевает, что электроны валентной зоны должны получить некоторую энергию для перехода на акцепторный уровень и создания дырки в самой валентной зоне.

Легирование полупроводников $A_{III}B_V$ может обсуждаться в тех же терминах, что и в случае кремния. Например, примесные атомы элементов VI группы (типа Se и S) на позиции мышьяка As (имеющего валентность 5) ведут себя как доноры, в то время как атомы примесей из II группы типа Be и Cd на позиции галлия (имеющего валентность 3) — в качестве акцепторов. Примеси IV группы, введенные в соединения $A_{III}B_V$, могут проявлять себя и как доноры, и как акцепторы, в зависимости от того, какой атом, III или V группы, они замещают. Например, атом кремния может замещать атом галлия в решетке GaAs, проявляя себя при этом типичным донором.

3.4. Концентрации электронов и дырок в полупроводниках

Для расчета концентрации определенного типа носителей (например, электронов) в полупроводниках необходимо определить плотность электронов в единичном интервале энергии в зоне проводимости (т. е. найти функцию плотности состояний) и вычислить вероятность заполнения каждого из этих интервалов. Функция распределения электронов по энергии $n(E)$ при этом имеет вид

$$n(E) = \rho(E) f_{FD}(E), \quad (3.3)$$

где $\rho(E)$ означает функцию плотности состояний, а $f_{FD}(E)$ — функция распределения Ферми — Дирака. Подставляя выражения (2.36) и (2.17) из главы 2 для $\rho(E)$ и $f_{FD}(E)$ соответственно, можно сразу получить распределение электронов по энергии $n(E)$ в зоне проводимости:

$$n(E) = \frac{4\pi}{h^3} (2m_e^*)^{3/2} (E - E_c)^{1/2} \frac{1}{1 + e^{(E - E_F)/kT}}, \quad (3.4)$$

поскольку известно, что минимальная энергия электронов в этой зоне составляет E_c .

Для вычисления полной концентрации n электронов в зоне проводимости необходимо проинтегрировать функцию распределения электронов по энергии $n(E)$, определенную уравнением (3.4), по всем доступным значениям энергии, т. е. от нижнего края зоны E_c до высшего уровня энергии в зоне $E_{c,\max}$. При этом, поскольку уровень Ферми располагается значительно ниже высшего уровня, верхний предел интегрирования можно заменить на ∞ , что существенно облегчает вычисление интеграла:

$$n = \int_{E_c}^{\infty} n(E) dE. \quad (3.5)$$

Подставляя в это уравнение выражение (3.4) для $n(E)$, легко получить

$$n = N_c F_{1/2}(\alpha), \quad (3.6)$$

где параметр α равен

$$\alpha = \frac{E_F - E_c}{kT}, \quad (3.7)$$

а коэффициент N_c представляет собой так называемую эффективную плотность состояний в зоне проводимости:

$$N_c = 2 \left(\frac{m_c^* kT}{2\pi\hbar^2} \right)^{3/2} \quad (3.8)$$

Через $F_{1/2}(\alpha)$ в уравнении (3.6) обозначен интеграл Ферми, определяемый в виде

$$F_{1/2}(\alpha) = \frac{2}{(\pi)^{1/2}} \int_0^{\infty} \frac{y^2 dy}{1 + e^{(y-\alpha)}}, \quad (3.9)$$

который не имеет точного аналитического выражения, вследствие чего подсчет концентрации электронов по уравнению (3.6) обычно осуществляется различными приближенными методами. Например, в случае $\alpha \leq -3$ этот интеграл с очень высокой точностью можно аппроксимировать выражением

$$F_{1/2} \approx \exp\left(\frac{E_F - E_c}{kT}\right). \quad (3.10)$$

Полупроводники, для которых справедливо указанное приближение, т. е. выполняется условие

$$E_c - E_F \geq 3kT, \quad (3.11)$$

называют невырожденными. В этом случае с очень высокой точностью для концентрации электронов в зоне проводимости можно считать, что

$$n = N_c e^{-(E_c - E_F)/(kT)}, \quad (3.12)$$

где коэффициент N_c определяется введенным выше уравнением (3.8).

Аналогичный подход может быть использован для расчета концентрации дырок p в зоне валентности. Функция распределения дырок определяется уравнением, поскольку дырка представляет собой всего лишь некоторое вакантное энергетическое состояние электрона,

$$f_h(E) = 1 - f_{FD}(E). \quad (3.13)$$

Далее можно снова воспользоваться уравнением (3.3), что дает

$$p = \int_{-\infty}^{E_v} \rho_h(E) f_h(E) dE, \quad (3.14)$$

где через E_v обозначен потолок валентной зоны. Пользуясь приведенными выше рассуждениями для электронов, можно показать, что концентрация дырок в невырожденных полупроводниках составляет

$$p = N_v e^{(E_F - E_v)/(kT)}, \quad (3.15)$$

где коэффициент N_v равен

$$N_v \equiv 2 \left(\frac{m_h^* kT}{2\pi\hbar^2} \right)^{3/2} \quad (3.16)$$

и называется соответственно эффективной плотностью состояний в валентной зоне. На рис. 3.17 показаны функция плотности состояний (а), функция распределения Ферми — Дирака

(б) и распределение носителей по энергии (в) для собственных полупроводников и полупроводников n - и p -типа, соответственно. Концентрациям электронов и дырок, определяемым уравнениями (3.12) и (3.15), на рисунке соответствуют заштрихованные участки.

Из приведенных выше уравнений для концентрации электронов и дырок может быть получено очень полезное соотношение, которое иногда называют законом действующих масс для

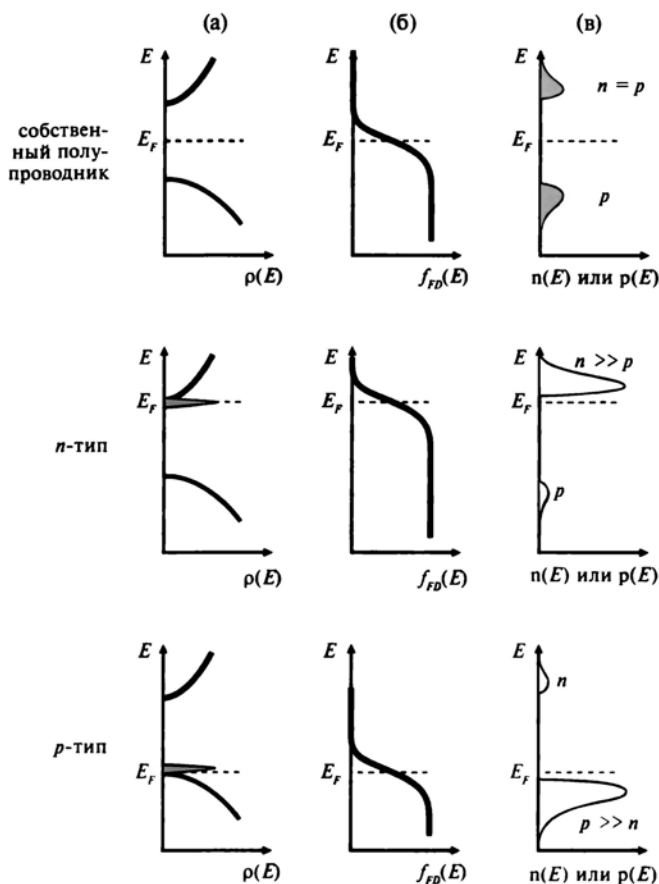


Рис. 3.7. Функция плотности состояний (а), функция распределения Ферми – Дирака (б) и распределение носителей по энергии (в) для беспримесных полупроводников и полупроводников n - и p -типа проводимости

концентрации носителей заряда. Перемножив выражения (3.12) и (3.15), легко найти, что

$$np = N_c N_v e^{-(E_c - E_v)/(kT)} = N_c N_v e^{-E_g/(kT)} = \text{const}(T), \quad (3.17)$$

где E_g — ширина запрещенной зоны полупроводника. Обозначим через n_i концентрацию носителей в собственных полупроводниках. Поскольку из приведенного выше уравнения следует, что в собственных полупроводниках $n = p = n_i$, мы можем получить для n_i уравнение

$$n_i = (np)^{1/2} = (N_c N_v)^{1/2} e^{-E_g/2kT}, \quad (3.18)$$

из которого следует, что в собственных полупроводниках концентрация носителей и концентрация электронно-дырочных пар экспоненциально уменьшаются с возрастанием ширины запрещенной зоны, но очень быстро увеличиваются при повышении температуры. Например, легко показать, что в кремнии при комнатной температуре $n_i \approx 1,5 \times 10^{10} \text{ см}^{-3}$. Такая сверхнизкая концентрация носителей зарядов (читателю можно напомнить, что концентрация носителей в металлах обычно составляет около 10^{22} см^{-3}) позволяет нам понять, почему при комнатных температурах почти все нелегированные полупроводники по своим электрическим характеристикам (т. е. с точки зрения обычной электропроводности) практически являются изоляторами.

Обычно степень легирования данного полупроводника (т. е. концентрация доноров N_D и акцепторов N_A) известна с достаточной точностью, поскольку эти параметры регулируются в процессе изготовления материала. Задачу определения концентраций электронов и дырок можно было бы считать решенной, если бы мы интересовались температурами, при которых ионизированными являются все примесные атомы, что справедливо для большинства используемых полупроводников при комнатных температурах. Например, уже при $T > 100 \text{ К}$ все легирующие примеси в кремнии являются ионизированными вследствие низких значений соответствующих энергий ионизации (см. раздел 3.3). Кроме этого, следует учитывать, что вещество в целом должно оставаться электрически нейтральным, вследствие чего полный положительный заряд (включая мобильные и неподвижные заряды) в полупроводнике должен

равняться полному отрицательному заряду, так что должно выполняться соотношение

$$p + N_D = n + N_A, \quad (3.19)$$

которое обычно называют условием электронейтральности. Из этого уравнения, совместно с упоминавшимся законом действующих масс в форме $np = n_i^2$, можно сразу получить уравнения для концентраций электронов и дырок (предполагается, что все атомы примесей ионизированы):

$$n = \frac{1}{2} \left[(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right], \quad (3.20)$$

$$p = \frac{1}{2} \left[(N_A - N_D) + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right]. \quad (3.21)$$

Стоит отметить, что при условии $N_D = N_A$ такие полупроводники ведут себя подобно собственным (поскольку количества электронов и дырок в них совпадают), из-за чего их обычно называют компенсированными полупроводниками.

В большинстве легированных полупроводников при комнатных температурах одно из значений концентраций носителей (N_D или N_A) значительно больше значения n_i . В частности, в технически важных полупроводниках обычно $N_D \gg N_A$ или $N_A \gg N_D$. В первом случае материал относится к полупроводникам n -типа, так как его проводимость обычно определяется электронами, возникающими при ионизации атомов донорных примесей. Для полупроводников n -типа из уравнений (3.20) и (3.21) следует соотношение

$$n \approx N_D, \quad p \approx \frac{n_i^2}{N_D}. \quad (3.22)$$

Аналогично для полупроводников p -типа можно записать

$$p \approx N_A, \quad n \approx \frac{n_i^2}{N_A}. \quad (3.23)$$

Указанные приближения очевидно неприменимы при высоких температурах, так как n_i квазиэкспоненциально возрастает с увеличением температуры.

Из уравнений (3.12) и (3.15) для концентрации носителей можно непосредственно вычислить значение уровня Ферми в нескольких предельных случаях. Например, для невырожденных полупроводников n -типа в системах, где температура достаточно высока для полной ионизации всех атомов примеси ($n \approx N_D$), но еще недостаточна высока для нарушения соотношения ($N_D \gg n_i$), уровень Ферми определяется из уравнений (3.12) и (3.22) и равен

$$E_c - E_F = kT \frac{N_c}{N_D}. \quad (3.24)$$

Аналогично, полупроводников p -типа можно получить соотношение

$$E_F - E_v = kT \frac{N_v}{N_A}. \quad (3.25)$$

Можно отметить, что уровни Ферми, определяемые уравнениями (3.24) и (3.25), располагаются тем ближе к границам соответствующих зон, чем выше степень легирования.

В собственных полупроводниках $n = p$ и из тех же уравнений (3.12) и (3.15) легко получить выражение для уровня Ферми для этих систем (E_F)_i в виде

$$(E_F)_i = \frac{E_c + E_v}{2} + kT \frac{N_v}{N_c} = \frac{E_g}{2} + \frac{3}{4} kT \frac{m_h^*}{m_c^*}, \quad (3.26)$$

где через E_g обозначена ширина запрещенной зоны. Отметим также, что при относительно небольших значениях T и примерно равенстве эффективных масс m_h^* и m_c^* уровень Ферми собственного полупроводника располагается вблизи середины энергетической щели.

3.5. Элементарные процессы переноса в полупроводниках

Механизм переноса носителей заряда в полупроводниках под воздействием слабых электрических полей весьма похож на механизм переноса в обычных металлах. Хотя электропроводность σ в полупроводниках значительно ниже, чем в металлах, этот

факт можно в основном объяснить очень низкой концентрацией самих носителей, так как в полупроводниках она составляет лишь 10^{14} — 10^{17} см⁻³ (в металлах обычно 10^{22} см⁻³).

С другой стороны, в полупроводниках мы сталкиваемся еще с одним типом проводимости, не имеющим аналога в металлах, а именно с переносом, обусловленным разностью концентрации носителей в различных точках. Это явление наблюдается в полупроводниках довольно часто, например, при их изготовлении с неоднородной концентрацией примесных атомов. Кроме того, во многих случаях мы можем и сами изменять концентрацию носителей в определенных участках, облучая полупроводник фотонами, энергия которых достаточна для образования электронно-дырочных пар. В таких ситуациях электроны и дырки начинают двигаться совершенно независимо, по диффузионному механизму, обусловленному градиентом соответствующих концентраций.

3.5.1. Движение носителей заряда в электрическом поле. Подвижность зарядов

В слабых электрических полях электропроводность в полупроводниках может быть описана в рамках полуклассической модели. В соответствии с уравнением (2.57) раздела 2.6.1 при воздействии на электрон электрического поля \vec{F} волновой вектор \vec{k} электрона (а следовательно, и его квазиимпульс) должны были бы возрастать до бесконечности. В реальных кристаллах, однако,двигающиеся под воздействием поля носители испытывают множество столкновений (с фононами, примесными атомами и т. п.). Такие процессы рассеяния в полуклассической модели рассматриваются в качестве некоторой силы «трения», препятствующей движению, вследствие чего мы можем записать второй закон Ньютона в виде следующего одномерного уравнения:

$$\left(qF - \frac{m_c^*}{\tau_c} \right) v_c = m_c^* \frac{dv_c}{dt}, \quad (3.27)$$

формально описывающего движение электрона с эффективной массой m_c^* . При выводе уравнения (3.27) предполагается, что силы трения пропорциональны дрейфовой скорости электрона v_c , а характерный параметр τ_c (называемый временем релаксации) означает то время, за которое дрейфовая скорость элект-

рона экспоненциально спадет до нуля при «выключении» внешнего поля. Это время определяется уравнением

$$v_c = [v_c]_{t=0} e^{-t/\tau_c}. \quad (3.28)$$

Из уравнения (3.27) следует, что в стационарном состоянии (т. е. при движении, когда сумма внешних сил, действующих на электрон, равна нулю) дрейфовая скорость постоянна и равна:

$$v_c = -\frac{q\tau_c}{m_c^*} F. \quad (3.29)$$

Коэффициент пропорциональности μ_c между величиной электрического поля и дрейфовой скоростью, т. е. величина

$$\mu_c = \left| \frac{F}{v_c} \right|, \quad (3.30)$$

называется подвижностью электрона. Исходя из этого определения, мы можем также считать подвижностью скоростью, приобретаемой электроном под воздействием единичного электрического поля. В соответствии с уравнениями (3.29) и (3.30) мы можем переписать выражение для подвижности электрона в виде

$$\mu_c = \frac{q\tau_c}{m_c^*}, \quad (3.31)$$

а также выписать аналогичное соотношение для подвижности дырок:

$$\mu_h = \frac{q\tau_h}{m_h^*}. \quad (3.32)$$

Плотность тока электронов имеет вид

$$J_c = -qnv_c, \quad (3.33)$$

где q — заряд электрона; n — плотность носителей. Используя уравнение (3.30) для подвижности, его можно переписать также в форме

$$J_c = qn\mu_c F. \quad (3.34)$$

Предположим далее, что величины J_c и E связаны законом Ома в виде

$$J_c = \sigma F, \quad (3.35)$$

где σ — электропроводность. Поскольку ток в полупроводниках (а следовательно, и их проводимость) связаны с переносом двух типов носителей (электронов и дырок), мы можем воспользоваться уравнениями (3.33) и (3.35) для электронов (и аналогичными уравнениями для дырок), что приводит к следующему выражению для электрической проводимости полупроводника:

$$\sigma = q(n\mu_e + p\mu_h), \quad (3.36)$$

где величины n и p обозначают концентрации электронов и дырок соответственно.

Интересно отметить, что такая простая модель позволяет нам даже связать макроскопические параметры (такие, как электропроводность или подвижность) с собственными характеристиками носителей заряда, такими, как их эффективная масса. Например, известно, что в GaAs эффективная масса очень мала ($m_e^* = 0,066 m_0$), и, следовательно, электронная подвижность, определяемая уравнением (3.31), в этом материале должна быть очень высокой.

3.5.2. Диффузионная проводимость

Как уже отмечалось выше, формально диффузионная проводимость в полупроводниках возникает из-за градиента концентрации носителей, но в реальных системах диффузия возникает в результате хаотического движения носителей и никак не связана с их зарядом. Диффузия носителей происходит вследствие нескольких, разных по характеру факторов: градиента концентрации, хаотического теплового движения и процессов рассеяния на несовершенствах кристаллической решетки. Поэтому диффузионные токи электронов и дырок в полупроводниках описываются обобщенными уравнениями диффузии, которые в одномерной системе имеют вид

$$J_c = qD_c \frac{dn}{dx} \quad (3.37)$$

и

$$J_h = -qD_h \frac{dp}{dx}, \quad (3.38)$$

где коэффициенты пропорциональности D_e и D_h называются коэффициентами диффузии электронов и дырок соответственно, а q — заряд электрона. Знак минус в правой части уравнения (3.38) обусловлен отрицательным знаком производной по концентрациям (область высокой концентрации носителей для дырок располагается слева от области с низкой концентрацией) и тем, что под воздействием градиента концентраций движение положительных зарядов (дырок) соответствует положительному x -направлению в рассматриваемой одномерной системе, т. е. создает положительный ток.

Механизм рассеяния при движении носителей в каком-либо полупроводнике не зависит от факторов, вызывающих это движение, т. е. одинаков для потоков, обусловленных градиентом концентраций или воздействием электрического поля. Поэтому коэффициенты подвижности и диффузии не могут быть независимы и их связывают известные соотношения Эйнштейна для электронов

$$\frac{D_e}{\mu_e} = \frac{kT}{q} \quad (3.39)$$

и дырок

$$\frac{D_h}{\mu_h} = \frac{kT}{q}, \quad (3.40)$$

причем для комнатной температуры отношения D/μ равно 0,026 В.

3.5.3. Уравнения непрерывности. Время жизни носителей и длина диффузии

В предыдущих разделах было показано, как возникают потоки носителей в полупроводниках, типичным примером чего можно считать процессы рождения электронно-дырочных пар под воздействием температуры или облучения фотонами. В разделах

3.5.1 и 3.5.2 рассматривалось влияние внешнего поля и градиента концентраций на процессы переноса носителей. Понятно, что возникшие носители могут в дальнейшем рекомбинировать, например, электрон зоны проводимости может просто попасть на незанятое энергетическое состояние (дырку) в валентной зоне и занять его. Обозначив через g и r скорости рождения и рекомбинации соответственно, мы можем записать привычное для электрических процессов уравнение непрерывности для скоростей изменения электронной плотности n . Рассмотрим для простоты одномерную модель, вводя в ней градиент концентраций, представляющий собой функцию координаты и времени. Скорость изменения электронной плотности в малом интервале ($x, x + dx$) определяется уравнением

$$\frac{\partial n}{\partial t} = g - r + \mu_c \frac{\partial(nF)}{\partial x} + D_c \frac{\partial^2 n}{\partial x^2}, \quad (3.41)$$

где два последних члена в правой части связаны с потоками носителей, описываемыми уравнениями (3.34) и (3.37). Естественно, аналогичное уравнение может быть выписано и для концентрации дырок, но (поскольку оба уравнения очевидно связаны друг с другом) решения, вообще говоря, могут быть достаточно сложными. Однако мы можем решить задачу за счет некоторых упрощающих модель предположений. Прежде всего, следует определить так называемые избыточные концентрации носителей заряда. Обозначив, например, равновесную концентрацию электронов через n_0 , а их реальную концентрацию в заданной точке x и в момент времени t через $n(x, t)$, определим избыточную концентрацию электронов просто в виде

$$\Delta n = n(x, t) - n_0. \quad (3.42)$$

Аналогично избыточную концентрацию дырок определим уравнением

$$\Delta p = p(x, t) - p_0. \quad (3.43)$$

Предположим далее, что система в целом подчиняется условию электронейтральности, т. е. избыточная концентрация

электронов компенсируется избыточной концентрацией дырок, что записывается в виде

$$\Delta n = n - n_0 = p - p_0 = \Delta p. \quad (3.44)$$

Последнее упрощающее предположение состоит в том, что избыточная концентрация каждого из носителей остается все же значительно меньшей, чем большая из двух равновесных концентраций n_0 и p_0 . Можно доказать, что (при соблюдении всех перечисленных упрощающих условий) в случае характерных примесных полупроводников описание системы может быть ограничено лишь решениями уравнений непрерывности для неосновных носителей [1]. Например, в случае p -полупроводников мы можем записать уравнение для избыточной концентрации электронов Δn (неосновные носители) в виде

$$\frac{\partial(\Delta n)}{\partial t} = g - r + \mu_e F \frac{\partial(\Delta n)}{\partial x} + D_e \frac{\partial^2(\Delta n)}{\partial x^2}, \quad (3.45)$$

а в случае n -полупроводников — аналогичное уравнение

$$\frac{\partial(\Delta p)}{\partial t} = g - r + \mu_h F \frac{\partial(\Delta p)}{\partial x} + D_h \frac{\partial^2(\Delta p)}{\partial x^2}. \quad (3.46)$$

Для избыточной концентрации дырок Δp , отметим дополнительно, что входящее в уравнения (3.45) и (3.46) электрическое поле F предполагается постоянным.

В качестве примера можно рассмотреть полупроводниковые приборы с p - n переходами. В таких приборах носители определенного типа (например, электроны из полупроводника n -типа), пересекают поверхность раздела и попадают в p -область, где они становятся неосновными носителями. Действительно, характеристики электронных приборов в значительной степени определяются поведением инжектированных неосновных носителей, которое, в свою очередь, зависит от нескольких параметров, главными из которых выступают время жизни носителей и длина диффузии. Предположим, например, что мы облучаем полупроводник n -типа фотонами, создающими в веществе электронно-дырочные пары и описанные выше избыточные концентрации носителей, причем их значения удов-

летворяют упоминавшимся условиям малости $\Delta n = \Delta p \ll n_0$ и $\Delta n = \Delta p \gg p_0$, где n_0 и p_0 — равновесные концентрации соответствующих носителей в полупроводнике n -типа. Если в момент $t = 0$ облучение вещества прекращается, то избыточная концентрация Δp неосновных носителей (в данном случае дырок) начинает уменьшаться за счет рекомбинации, причем скорость уменьшения при этом пропорциональна текущей концентрации дырок, т. е. имеет вид

$$\left[\frac{\partial \Delta p}{\partial t} \right]_{\text{recomb}} = -\frac{\Delta p}{\tau_h}, \quad (3.47)$$

где величина $1/\tau_h$ выступает коэффициентом пропорциональности. Обозначив через $(\Delta p)_0$ избыточную концентрацию дырок в начальный момент $t = 0$, легко получить после интегрирования

$$\Delta p(t) = (\Delta p)_0 e^{-t/\tau_h}. \quad (3.48)$$

Очевидно, что введенный коэффициент τ_h может трактоваться лишь как среднее время жизни неосновного носителя (естественно, мы можем аналогично определить эту величину и для электронов). Время жизни неосновных носителей в легированных полупроводниках составляет обычно около 10^{-7} с, однако оно уменьшается при введении в полупроводник дополнительных примесей или ловушек.

Рассмотрим представленный на рис. 3.8, а гипотетический эксперимент, в котором носители одного типа (например, электроны) с постоянной скоростью инжектируются в полупроводник p -типа через его поверхность, соответствующую значению $x = 0$. Предположим, нам необходимо определить уменьшение концентрации инжектированных носителей (вдоль направления x) за счет их рекомбинации с дырками в таком полупроводнике p -типа. Для этого мы можем использовать выведенное уравнение непрерывности для стационарных граничных условий при нулевом приложенном электрическом поле. Используя в уравнении (3.45) введенное только что определение среднего времени жизни неосновных носителей заряда, получим

$$-\frac{\Delta n(x)}{\tau_e} + D_e \frac{\partial^2 (\Delta n(x))}{\partial x^2} = 0. \quad (3.49)$$

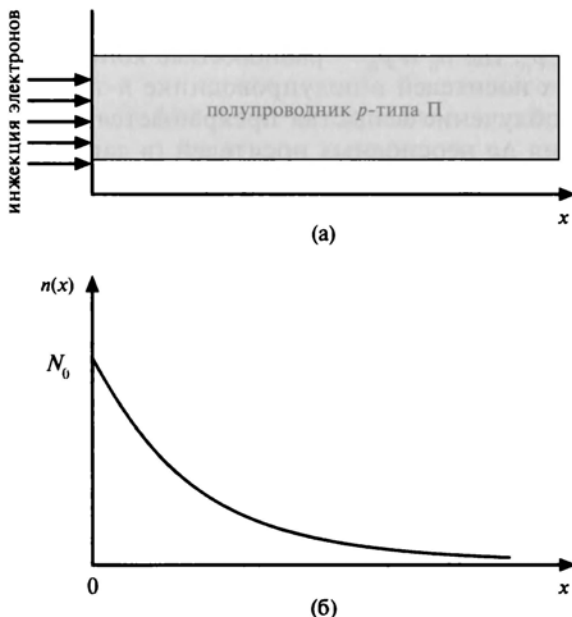


Рис. 3.8. (а) Инжекция электронов в полупроводнике p -типа через поверхность $x = 0$; (б) уменьшение концентрации инжектированных электронов как функция расстояния x

Обозначив затем через N_0 число электронов, непрерывно инжектируемых через поверхность $x = 0$ в единицу времени, можно проинтегрировать это выражение, что дает

$$\Delta n(x) = N_0 e^{-x/L_c}, \quad (3.50)$$

где

$$L_c = (D_c \tau_c)^{1/2}. \quad (3.51)$$

На рис. 3.8, б показан спад концентрации инжектированных электронов с расстоянием в полупроводниках p -типа. Введенный выше параметр L_c называют длиной диффузии неосновных носителей для электронов, а физически эта величина соответствует расстоянию, на котором число инжектированных электронов уменьшается в e раз по сравнению со значением на исходной поверхности $x = 0$.

3.6. Вырожденные полупроводники

В разделе 3.4 невырожденными полупроводниками были названы те из них, в которых уровень Ферми E_F располагается в запрещенной зоне примерно на расстоянии $3kT$ или больше от краев зон. Поскольку к таким полупроводникам можно применять классическую статистику, мы получили там простые выражения (3.12) и (3.15) для концентрации электронов и дырок соответственно. На основе этих же предпосылок мы вывели уравнения (3.24), (3.25) и (3.26) для положения уровня Ферми в полупроводниках n -типа, p -типа и собственных полупроводников.

По мере увеличения концентрации легирующих атомов, положение уровня Ферми изменяется, и, если значения n или p превышают N_c или N_v , определяемые соответственно уравнениями (3.8) и (3.16), уровень Ферми попадает в зону проводимости (в полупроводниках n -типа) или валентную зону (в полупроводниках p -типа). Такие сильнолегированные полупроводники называют вырожденными, что достигается обычно при концентрации легирующих элементов порядка 10^{19} — 10^{20} см $^{-3}$.

В вырожденных полупроводниках волновые функции электронов соседних атомов примеси перекрываются, в результате чего (подобно тому, как это происходит в кристаллах) дискретные примесные уровни образуют как бы собственную узкую примесную зону, показанную на рис. 3.9, а. Такие примесные зоны, соответствующие перекрыванию уровней исходных доноров и акцепторов в зоне проводимости и валентной зоне, становятся их составной частью. Образующиеся при этом энергетические состояния называют «хвостом распределения» плотности состояний, а их наличие приводит к сжатию (уменьшению) ширины запрещенной зоны. Этот эффект наглядно проявляется в лазерных диодах (раздел 10.3) и в спектрах поглощения сильнолегированных полупроводников.

На рис. 3.9, б приведена энергетическая диаграмма для вырожденного полупроводника n -типа. Известно, что в таких материалах все уровни энергии ниже E_{Fn} заполнены почти полностью, поэтому большинство электронов в них имеют энергии в узком диапазоне от E_c до E_{Fn} . Диаграмма уровней несколько напоминает диаграмму для металлов, а уровень Ферми в ней совпадает с высшим энергетическим уровнем электронов в зоне. Однако при очень сильном легировании (порядка 10^{20} см $^{-3}$) из-за

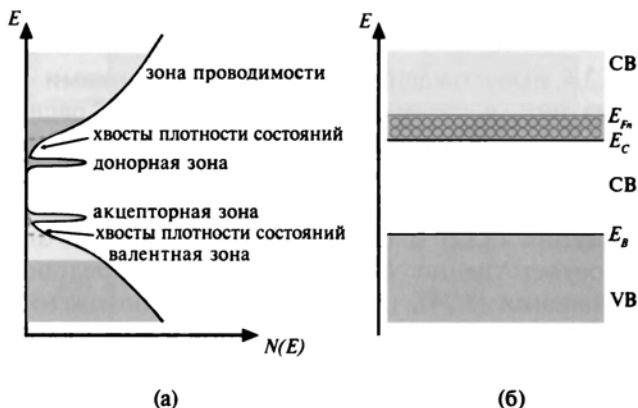


Рис. 3.9. (а) Примесные зоны и хвосты плотности состояний в вырожденном полупроводнике; (б) диаграмма энергетических уровней в вырожденном полупроводнике n -типа.

взаимодействия между атомами примеси начинает проявляться эффект насыщения, т. е. концентрация носителей не увеличивается с концентрацией легирующей примеси. При очень высоких концентрациях становятся неприменимыми уравнения (3.22) и (3.23), а также закон действующих масс в форме (3.17).

3.7. Оптические свойства полупроводников

3.7.1. Оптические процессы в полупроводниках

При воздействии света на полупроводники наблюдается целый ряд оптических явлений (поглощение, пропускание и отражение), вызывающих разнообразные электронные процессы в этих материалах. Естественно, что исследование различных оптических спектров стало важным инструментом для изучения свойств полупроводников и происходящих в них явлений. Например, спектр поглощения типичного полупроводника всегда имеет некоторые характерные особенности, так как в основном поглощаются только те фотоны, энергия которых равна или больше ширины запрещенной зоны, вследствие чего преобладающим процессом становится переход электронов из заполненной валентной зоны в пустую зону проводимости. Такие переходы могут быть прямыми и непрямыми, а их расчет можно осуществить, пользуясь теорией возмущений, зависящей от времени.

На низкоэнергетической границе полосы поглощения при этом можно зарегистрировать так называемое экситонное поглощение в виде серии острых пиков. Экситонами называют электронно-дырочные пары, в которых электрон и дырка соединены кулоновскими силами притяжения (спектры их поглощения обсуждаются ниже, в разделе 3.7.3). Кроме того, поглощение в полупроводниках может быть обусловлено и рядом других процессов, например, электронными переходами с донорных уровней в состоянии зоны проводимости или из состояний в валентной зоне на акцепторные уровни. Соответствующие таким переходам пики поглощения расположены в инфракрасном диапазоне из-за низких значений энергии ионизации доноров и акцепторов (см. раздел 3.3). В сильнолегированных полупроводниках может наблюдаться также значительное оптическое поглощение свободными носителями, так как коэффициент поглощения пропорционален концентрации носителей. Наконец, в ионных кристаллах электромагнитные волны могут и прямо возбуждать оптические фононы из-за сильной дипольной связи между фотонами и поперечными оптическими фононами. Такие пики поглощения, связанные с колебаниями решетки, наблюдаются в инфракрасном диапазоне.

Поглощение света полупроводниками может быть описано макроскопически путем введения соответствующих коэффициентов поглощения. Обозначим через I_0 интенсивность светового потока, падающего на поверхность ($z = 0$) полупроводника. На расстоянии z от поверхности его интенсивность $I(z)$ определяется уравнением

$$I(z) = I_0 e^{-\alpha z}, \quad (3.52)$$

где параметр α , характеризующий конкретное вещество, называется обычно коэффициентом поглощения. Этот параметр зависит от длины волны света и имеет размерность см^{-1} (обратный ему параметр $1/\alpha$ называется глубиной проникновения). Очевидно, что, чем выше коэффициент поглощения, тем меньше глубина, на которую свет может проникнуть в данное вещество. В материалах типа GaAs параметр α обычно резко возрастает для фотонов с энергией выше E_g , так как для таких полупроводников характерны прямые переходы (см. раздел 3.7.2). С другой стороны, для полупроводников с непрямыми переходами (типа Si и Ge) параметр α с ростом энергии фотонов

увеличивается заметно медленнее, так как оптические переходы в таких материалах происходят с участием фононов. Кроме того, в непрямозонных полупроводниках поглощение начинается не точно при энергии фотонов $h\nu = E_g$, а происходит в некотором интервале энергий относительно этого значения, ширина которого соответствует энергии фононов.

3.7.2. Межзонное поглощение

Межзонное поглощение в полупроводниках очень сильно зависит от зонной структуры конкретного твердого тела, особенно в случае запрещенных зон с прямыми и непрямыми оптическими переходами. В этом разделе мы уделим основное внимание процессам межзонного поглощения в полупроводниках с прямыми оптическими переходами, так как именно к этому классу относятся многие материалы, на основе которых создается большинство оптоэлектронных светоизлучающих приборов, описанных в главе 10.

Для расчета коэффициента оптического поглощения мы можем воспользоваться квантово-механической скоростью перехода W_{if} электрона из начального состояния (описываемое волновой функцией ψ_i) в конечное состояние (описываемое волновой функцией ψ_f). Эта скорость определяется «золотым правилом» Ферми для переходов (уравнение (2.26 из главы 2):

$$W_{if} = \frac{2\pi}{\hbar} \rho(E) |H'_{if}|^2, \quad (3.53)$$

В котором матричный элемент H'_{if} соответствует внешнему оптическому возмущению электронов, а $\rho(E)$ — функция плотности состояний для разности энергий между конечным и начальным состояниями при энергии возбуждающего фотона $\hbar\omega$.

Гамильтониан возмущения (см. раздел 2.2.4), связанный с действием электромагнитной волны на электрон в точке \vec{r} , определяется формулой

$$H' = -\vec{p} \cdot \vec{F} = e\vec{r} \cdot \vec{F}, \quad (3.54)$$

и соответствует энергии электрического диполя электрона $-e\vec{r}$ при воздействии электрического поля волны \vec{F} . Для электрон-

ных состояний, описываемых функциями Блоха (раздел 2.4), матричные элементы принимают вид

$$H'_{if} \propto \int_{u.c.} u_i^*(\vec{r}) x_{if}(\vec{r}) d\vec{r}, \quad (3.55)$$

где пределы интегрирования относятся к объему элементарной ячейки, так как интеграл по всему объему кристалла может быть сведен к сумме интегралов по ячейкам. Интегрируемые функции $u_i(\vec{r})$ и $u_f(\vec{r})$ в соответствии с теоремой Блоха должны обладать периодичностью описываемой кристаллической решетки.

Функция плотности состояний $\rho(E)$ в уравнении (3.53) должна быть вычислена для энергии возбуждающего фотона $h\nu$ с достаточной точностью (так как начальное и конечное состояния, лежащие в различных зонах, должны отличаться по энергии именно на величину $h\nu$). По этой причине функцию $\rho(E)$ обычно называют функцией плотности оптически связанных состояний, и ее вычисление требует сведений о структуре зон. Кроме этого, следует помнить, что в полупроводниках с прямыми оптическими переходами волновые векторы электрона в начальном и конечном состоянии должны быть одинаковы, вследствие чего

$$\vec{k}_f = \vec{k}_i, \quad (3.56)$$

так как связанный с фотоном импульс является пренебрежимо малым.

В оптических переходах вблизи значений волнового вектора $\vec{k} = 0$ внутри полупроводников $A_{III}B_V$ (например, в GaAs) вовлекаются атомные орбитали p -типа валентной зоны и состояния зоны проводимости, возникающие из орбиталей s -типа. Известно также, что правила отбора для диэлектрических диполей позволяют переходы из p -состояний в s -состояния, вследствие чего в рассматриваемых системах действительно следует ожидать интенсивное оптическое поглощение.

В качестве примера для расчета оптической функции плотности состояний рассмотрим системы, в которых зона проводимости и три разных типа валентных зон (зоны тяжелых и легких дырок и так называемая отщепленная зона) описываются параболой вблизи точки Γ , как показано на рис. 3.2, *a*. Ширина E_g запрещенной зоны с прямыми оптическими переходами

равна разности энергий между минимумом зоны проводимости и максимумами зон для тяжелых и легких дырок, которые вырождаются в точке Г. Для любых таких переходов из закона сохранения энергии следует условие

$$h\nu = E_g + \frac{\hbar^2 k^2}{2m_c^*} + \frac{\hbar^2 k^2}{2m_h^*} = E_g + \frac{\hbar^2 k^2}{2\mu}, \quad (3.57)$$

где параметр μ соответствует приведенной массе системы (электрон — дырка) и для простоты учитывается лишь одна из двух вырожденных зон дырок, а переход с отщепленной зоны происходит при облучении фотонами более высоких энергий. Исходя из приведенного в разделе 2.3 формулы (2.36) для функции плотности состояний, можно записать коэффициент поглощения α при $h\nu \geq E_g$ в виде

$$\alpha(h\nu) = \frac{1}{2\pi^2} \left(\frac{2\mu}{\hbar^2} \right)^{3/2} (h\nu - E_g)^{1/2}. \quad (3.58)$$

С учетом формулы (3.53) для скорости W_{if} оптических переходов и того факта, что коэффициент поглощения α пропорционален W_{if} , можно утверждать, что при $h\nu \geq E_g$ коэффициент α должен быть также пропорционален квадратному корню из энергии фотона за вычетом ширины запрещенной зоны полупроводника. Следовательно, построив зависимость α^2 от энергии $h\nu$, мы должны получить прямую линию, которая отсекает на горизонтальной оси (при $\alpha = 0$) отрезок, длина которого соответствует ширине запрещенной зоны полупроводника.

В качестве примера на рис. 3.10 приведена зависимость α^2 от энергии падающих фотонов в кристалле PbS, явно демонстрирующая обсуждаемый выше линейный характер [2]. В некоторых других полупроводниках с прямыми оптическими переходами отклонения описываемой зависимости от линейной являются более значительными, что свидетельствует о том, что в таких системах недостаточно строго выполняются все те условия, которые были оговорены при выводе формулы (3.58). Например, в кристаллах GaAs при низких температурах область появления поглощения размывается вследствие поглощения экситонов, как будет показано в следующем разделе. Кроме того, строго говоря, уравнение (3.58) справедливо лишь в ближайшей окрестности

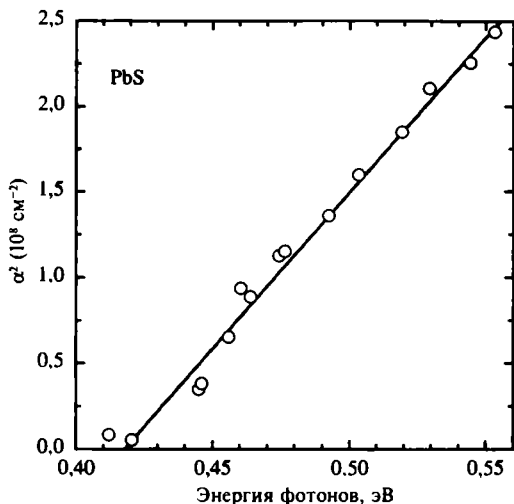


Рис. 3.10. Зависимость квадрата коэффициента поглощения α^2 от энергии фотонов для кристалла PbS (по данным работы [2])

значений вблизи $\vec{k} = 0$, вследствие чего с ростом энергии фотонов это условие почти сразу перестает выполняться.

3.7.3. Экситонные эффекты

Выше отмечалось, что фотоны с энергией, равной или превышающей ширину запрещенной зоны полупроводника, могут создавать электронно-дырочные пары. Обычно возникшие электрон и дырка дальше движутся независимо друг от друга, однако в некоторых случаях благодаря кулоновскому взаимодействию между ними электрон и дырка могут оставаться «вместе», формируя новую электрически нейтральную квазичастицу, которую называют экситоном. Такие частицы не влияют на электропроводность вещества, поскольку не обладают электрическим зарядом. Возникновению экситонов существенно облегчается в квантовых ямах (см. разделы 1.5 и 4.10), так как локализация в ограниченной области усиливает эффекты перекрытия волновых функций электронов и дырок.

Простейшая модель описывает экситон как электрон и дырку, которые вращаются внутри решетки относительно общего центра массы под воздействием кулоновского притяжения, как это показано на рис. 3.11. Существуют два основных типа экситонов: 1) экситоны с незначительным перекрытием волновых

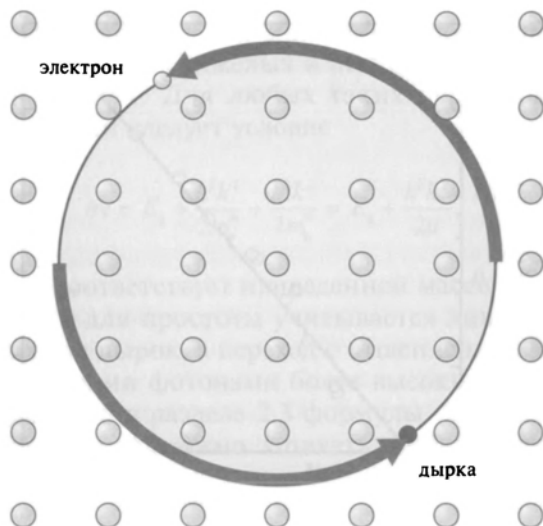


Рис. 3.11. Представление экситона в виде связанного состояния электронно-дырочной пары в полупроводнике

функций электронов и дырок, радиусы которых охватывают множество атомов кристалла. Такие системы называются экситонами Ванье — Мотта и именно они характерны для многих полупроводников; 2) экситоны, характерные в основном для изоляторов, радиус которых имеет порядок постоянной решетки (так называемые экситоны Френкеля).

Экситоны Ванье — Мотта могут быть описаны в рамках водородоподобной модели, вследствие чего энергии связанных состояний для них определяются уравнениями типа (2.15) главы 2 после соответствующих поправок. Примерно такой метод использовался и в разделе 3.3 для определения уровней ионизации донорных и акцепторных примесей в несобственных полупроводниках. Очевидно, что масса в этих уравнениях должна быть заменена на приведенную массу μ системы (электрон + дырка), т. е.

$$\frac{1}{\mu} = \frac{1}{m_e^*} + \frac{1}{m_h^*}. \quad (3.59)$$

Кроме того, используя упомянутые уравнения, мы должны учитывать, что электрон и дырка «погружены» в среду с ди-

электрической постоянной $\epsilon_r \epsilon_0$, где ϵ_r — относительная высокочастотная диэлектрическая постоянная среды. С учетом сказанного, энергии связанных состояний экситонов имеют вид

$$E_n = -\frac{\mu R_H}{m_0 \epsilon_r^2} \frac{1}{n^2} = -\frac{\mu}{m_0 \epsilon_r^2} \frac{13.6 \text{ eV}}{n^2} = -\frac{R_{\text{ex}}}{n^2}, \quad n = 1, 2, 3, \dots, \quad (3.60)$$

где R_H — постоянная Ридберга для атома водорода; R_{ex} — постоянная Ридберга экситона.

На рис. 3.12 представлены связанные состояния экситона, определяемые уравнением (3.60), и энергия ионизации экситона E_I . Дело в том, что энергия фотона, необходимая для создания экситона, меньше энергии, необходимой просто для создания независимой электронно-дырочной пары в полупроводнике, поэтому полный процесс можно описать как создание экситона с его дальнейшим разделением на электрон и дырку. При этом второй процесс требует энергии, равной энергии связи экситона. Поэтому, как показано на рис. 3.12, связанные состояния экситона, определяемые уравнением (3.60), расположены в запрещенной зоне полупроводника, чуть ниже границы зоны проводимости.

На рис. 3.13 показан спектр поглощения фотонов, энергия которых близка к значению ширины энергетической зоны [3]. На рисунке легко выделить три пика, предсказываемые уравне-

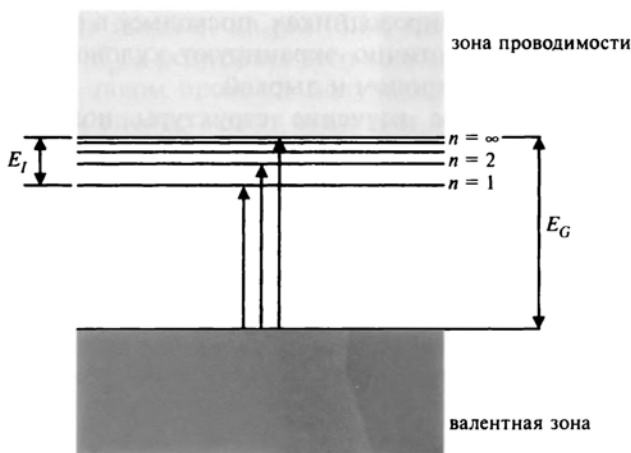


Рис. 3.12. Связанные состояния экситона вблизи границы зоны проводимости

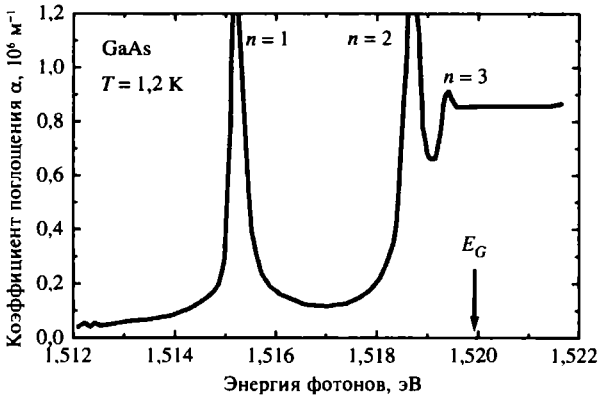


Рис. 3.13. Спектр поглощения экситонов в GaAs при 1,2 К (по данным работы [3])

нием (3.60), с постоянной $R_{\text{ex}} = 4,2$ эВ. Четкость полученного спектра объясняется несколькими причинами: проведением измерений при очень низких температурах, высоким разрешением спектрометра и сверхвысокой чистотой исследуемых образцов. Особое значение для точности таких спектральных измерений имеет низкая температура, при которой экситоны стабильны и не распадаются под воздействием фононов, не говоря уже об отсутствии теплового уширения спектральных линий. Экситоны гораздо легче и чаще наблюдаются в собственных, а не в легированных полупроводниках, поскольку в последних свободные носители частично экранируют кулоновское взаимодействие между электроном и дыркой.

Более тщательное изучение структуры полупроводников позволяет предсказывать, какие именно области k -пространства являются подходящими для образования экситонов. Поскольку экситон, как уже было сказано, представляет собой электронно-дырочную пару, векторы скоростей обеих частиц должны совпадать, вследствие чего, согласно уравнению (2.52) из главы 2, их зона проводимости и валентная зона должны быть параллельны, что особенно отчетливо проявляется в ближайшей окрестности точки $\vec{k} = 0$ в GaAs, т. е. в спектральной области, соответствующей запрещенной зоне с прямыми оптическими переходами.

Если создающее экситоны излучение является достаточно мощным, то плотность возникающих экситонов возрастает настолько, что они начинают взаимодействовать как друг

с другом, так и со свободными носителями. При очень высокой плотности создаваемых частиц могут возникать даже биэкситоны (состоящие из двух экситонов), наличие которых обнаруживается не только в объемных полупроводниках, но и в квантовых ямах и квантовых точках. Биэкситоны из двух электронов и двух дырок можно уподобить (подобно тому, как выше мы сравнивали экситон с атомом водорода) с молекулой водорода. Кроме этого, в настоящее время уже возникла необходимость учета так называемых трионов (состоящих из экситона с добавлением дырки или электрона), наличие которых экспериментально обнаруживается в некоторых наноструктурах, включая квантовые ямы и сверхрешетки в полупроводниках $A_{III}B_V$.

3.7.4. Спектр излучения

В разделе 3.7.2 рассматривались электронные переходы в полупроводниках из валентной зоны в зону проводимости, обусловленные поглощением фотонов. При обратном процессе, т. е. при переходе возбужденного электрона в какое-то из состояний зоны с низшей энергией может происходить излучение света. В этом случае вследствие излучательного перехода наблюдается фотолюминесценция. Частота изучаемого фотона при этом, вообще говоря, может отличаться от частоты поглощенного, и обычно спектр излучения уже, чем спектр поглощения. Механизм процесса показан на рис. 3.14, где поглощается фотон с энергией $h\nu > E_g$, в результате чего возникает электронно-дырочная пара. В таком процессе получаемая электроном и/или дыркой энергия вполне может оказаться выше энергии теплового равновесия в системе. Такие электроны (или дырки) очень быстро теряют избыточную кинетическую энергию за счет излучения фононов (главным образом оптических) и занимают состояния вблизи дна энергетической зоны. Этот процесс является очень коротким ($\sim 10^{-13}$ с) из-за достаточно сильного электрон-фононного взаимодействия.

С другой стороны, известно, что время излучательного перехода, при котором электроны переходят в валентную зону с излучением фотонов с энергией $h\nu \sim E_g$, на несколько порядков величины больше (т. е. составляет несколько наносекунд). Поэтому спектр излучения может превышать E_g на величину порядка kT , поскольку электроны имеют достаточное время для

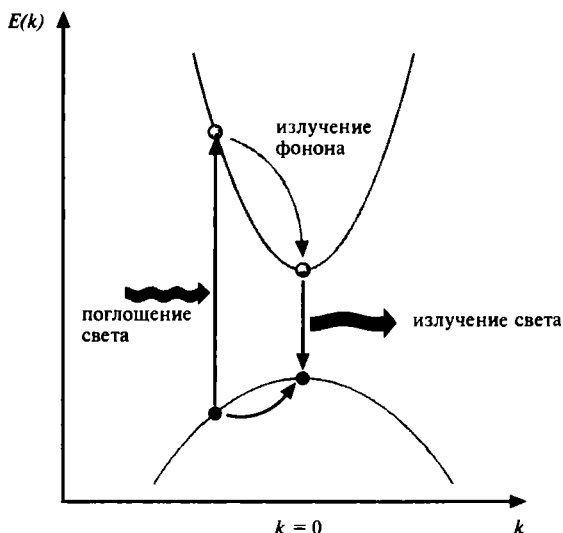


Рис. 3.14. Дезактивация (снятие возбуждения) электрона с излучением оптического фонона и последующим излучением фонона с энергией, примерно равной ширине запрещенной зоны.

термализации, т. е. перехода из состояний с большей энергией на дно зоны проводимости. В настоящее время все эти весьма интересные процессы тщательно изучаются методами фотолюминесцентной спектроскопии с высоким временным разрешением и использованием ультракоротких лазерных импульсов. На рис. 3.15 приведен спектр объемного кристалла GaAs с температурой 77 К, возбуждаемого лазерными импульсами длительностью 14 фемтосекунд [4]. Представлены кривые для трех различных значений концентрации носителей, возрастающие сверху вниз. Четвертая кривая представляет автокорреляционную функцию лазерного излучения [4].

В очень многих полупроводниках рекомбинация электронов и дырок происходит без излучения, т. е. процесс рекомбинации осуществляется посредством центров рекомбинации с энергетическими уровнями внутри запрещенной зоны. В такой ситуации потерянная электроном энергия передается решетке в виде тепла. В разделе 3.3 отмечалось, что состояния легирующих примесей располагаются в полупроводниках внутри запрещенной зоны около границ, однако энергетические уровни других дефектов решетки (например, вакансий или примесей металлов) могут располагаться

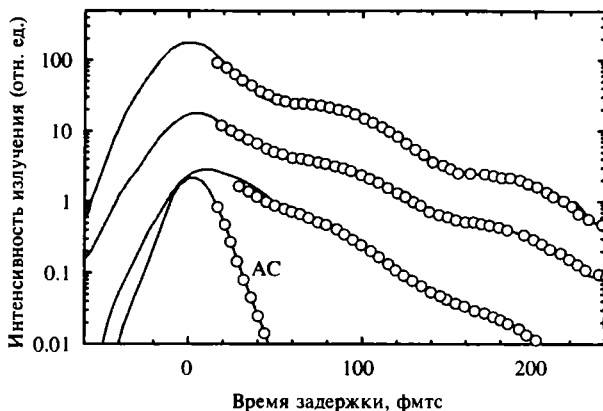


Рис. 3.15. Спектр поглощения экситонов в GaAs при 1,2 К по данным работы [4]

внутри запрещенной зоны гораздо глубже. Лишь для квантовых ям высокой степени совершенства отношение числа испускаемых фотонов к числу возбуждаемых электронно-дырочных пар достигает значений 0,1—1. Даже в объемных полупроводниках с прямыми оптическими переходами квантовый выход люминесценции очень невелик и лежит в пределах от 10^{-3} до 10^{-1} . Атомы примесей (и вообще любые другие дефекты) могут выступать в качестве центров рекомбинации, захватывая электрон или дырку, а затем носитель противоположного знака. Дефекты, на которых захваченные носители освобождаются до того, как происходит захват носителя другого знака, обычно называются ловушками. Центры рекомбинации называют также быстрыми или медленными, в зависимости от времени, которое требуется им для освобождения от одного захваченного носителя и захвата следующего.

3.7.5. Стимулированное излучение

Рассмотрим простейшую электронную систему (рис. 3.16), имеющую лишь два энергетических уровня — E_1 и E_2 ($E_2 > E_1$). В такой системе электроны с основного состояния E_1 могут перейти в возбужденное состояние E_2 , поглотив фотон с энергией ($E_2 - E_1$), а электроны с возбужденного состояния E_2 могут опуститься в состояние E_1 , излучив фотон с такой же энергией ($E_2 - E_1$). Естественно, что интенсивность излучения при переходах с воз-

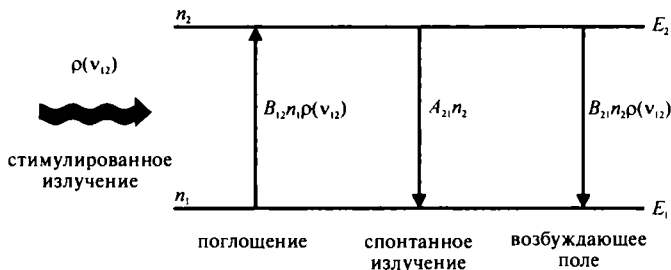


Рис. 3.16. Двухуровневая электронная система, в стационарном состоянии. Интенсивность поглощения (*а*) должна равняться сумме интенсивностей спонтанного (*б*) и стимулированного (*в*) излучений

бужденного состояния E_2 в основное состояние E_1 должна быть пропорциональна электронной заселенности n_2 уровня E_2 . Такой процесс называют спонтанным излучением, а коэффициент пропорциональности — коэффициентом Эйнштейна A_{21} (соответствующий параметр для процесса *поглощения* называется коэффициентом Эйнштейна B_{12}). Названия связаны с тем, что именно Эйнштейн показал, что электроны могут переходить с возбужденного состояния E_2 в основное E_1 под воздействием фотона с энергией $h\nu = E_2 - E_1$. Этот процесс был назван *стимулированным излучением*, а его интенсивность определяется величиной еще одного коэффициента (B_{21}). Поскольку интенсивность стимулированного излучения пропорциональна плотности фотонов $\rho(\nu)$, для его возникновения необходимо облучать вещество потоком фотонов высокой плотности. Все три введенных выше эйнштейновских коэффициента связаны друг с другом, поскольку в стационарном состоянии системы заселенности уровней должны быть постоянными и число переходов сверху вниз должно равняться числу переходов снизу вверх, как показано на рис. 3.16.

Очень интересным и важным аспектом стимулированного излучения является совпадение фаз излученных фотонов с фазой фотонов возбуждающего поля. Строго говоря, работа лазеров основана на процессах стимулированного излучения. Полупроводниковые лазеры, которые будут рассматриваться в главе 10, позволяют получать монохроматическое и когерентное излучение. Интенсивность вынужденного излучения таких структур пропорциональна произведению $n_2\rho(\nu)$, вследствие чего для преобладания процессов излучения над процесса-

ми поглощения необходимо обеспечить выполнение условия $n_2 > n_1$, которое называется инверсной заселенностью, так как в соответствии с распределением Больцмана во всех системах при тепловом равновесии имеет место обратное неравенство ($n_1 > n_2$). Отметим также, что вследствие того, что мощность вынужденного излучения пропорциональна плотности энергии облучения, лазеры снабжаются объемными резонаторами, в которых концентрация фотонов может быть значительно увеличена за счет многократных внутренних оптических отражений.

Инверсная заселенность в полупроводниковых лазерах достигается инжекцией носителей (электронов и дырок) через $p^+ - n^+$ переходы в вырожденных полупроводниках с прямыми оптическими переходами, работающих при прямом смещении. На рис. 3.17, а показан несмещенный $p - n$ переход, а на рис. 3.17, б — тот же переход, к которому приложено прямое смещение. В этой ситуации вблизи контакта материалов типа p^+ и n^+ формируется так называемая активная область, в которой и выполняется условие инверсной заселенности.

Из схемы действия, приведенной на рис. 3.18, а, можно оценить энергетический диапазон требуемых для работы лазера фотонов, т. е. фотонов, облучение которыми способно привести к тому, что интенсивность вынужденного излучения будет выше интенсивности поглощения, что и приводит к оптическому усилению. Исходя из плотности электронных состояний, полученных через функции плотности состояний для зоны

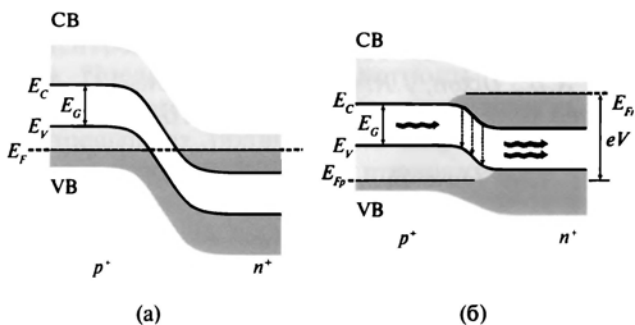


Рис. 3.17. (а) Энергетическая диаграмма для $p - n$ перехода в вырожденных полупроводниках без приложенного напряжения. (б) Та же диаграмма при работе полупроводника с прямым смещением; достаточно высоким для создания инверсной заселенности в активной области

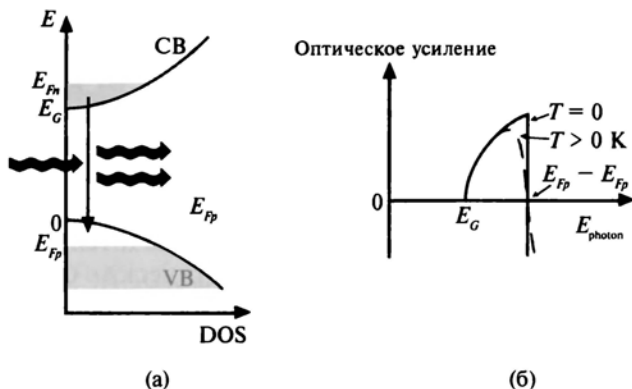


Рис. 3.18. (а) Плотность состояний электронов и дырок в зоне проводимости и валентной зоне соответственно; (б) оптическое усиление в зависимости энергии фотонов

проводимости и валентной зоны, можно качественно оценить параметры оптического усиления в зависимости от энергии фотонов. Очевидно, что фотоны, вызывающие вынужденное излучение, должны обладать энергией выше E_g (рис. 3.18, б), но меньшей разности ($E_{Fn} - E_{Fp}$). При высоких температурах распределение Ферми — Дирака вблизи уровней Ферми расширяется, вследствие чего оптическое усиление понижается.

Литература

1. McKelvey J.R. (1966). Solid-State and Semiconductor Physics (Harper and Row, New York).
2. Schoolar R.B., Dixon, J.R. (1965). Phys. Rev., 137, A667.
3. Fehrenbach G.W., Schäfer W., Ulbrich R.G. (1985). J. Luminescence, 30, 154.
4. Banyai L., Tran Thoai D.B., Reitsamer E., Haug H., Steinbach D., Wehner M.U., Wegener M., Marschner T., Stolz W. (1995). Phys. Rev. Lett., 75, 2188.

Дополнительная литература

- Seeger K. (1999). Semiconductor Physics (Springer, Berlin).
 Singh J. (2003). Electronic and Optoelectronic Properties of Semiconductor Structures (Cambridge University Press, Cambridge).

Yu P., Cardona M. (1996). Fundamentals of Semiconductors (Springer, Berlin).

Задачи и упражнения

1. Средний свободный пробег электрона. Найдите значения среднего свободного пробега электрона в GaAs при комнатной температуре и 77 К, исходя из величин относительной подвижности при этих температурах (3×10^5 и 10^4 см²/В сек, соответственно).
2. Легирование полупроводников. Для создания *p-n* диода образец кремния легируют атомами фосфора (полупроводник *n*-типа) в количестве 5×10^{15} на 1 см³. Кроме этого, часть образца легируется атомами бора (*p*-типа) в концентрации 10^{17} . Рассчитайте положение уровней Ферми при $T = 300$ К на обеих сторонах такого *p-n* перехода. Каково значение контактной разности потенциала?
3. Концентрация носителей в германии. Определите концентрации электронов и дырок в образце германия при комнатной температуре (концентрация доноров и акцепторов $2,5 \times 10^{14}$ /см³ и $3,5 \times 10^{14}$ /см³ соответственно), предполагая, что все примесные атомы ионизированы. Определите тип проводимости (*p* или *n*) такого полупроводника, если концентрация собственных носителей в Ge при комнатной температуре составляет $n_i = 2,5 \times 10^{13}$ /см³.
4. Концентрация носителей в кремнии. Брусек полупроводникового кремния легирован примесями *n*-типа и *p*-типа (с концентрациями 4×10^{14} /см³ и 6×10^{14} /см³, соответственно). Предполагая, что плотность состояний не меняется с ростом температур и подвижность электронов вдвое превышает подвижность дырок, определите концентрацию носителей и тип проводимости такого полупроводника при $T = 300$ и 600 К. Ширина запрещенной зоны в кремнии составляет $E_g = 1,1$ эВ, а $n_i = 1,5 \times 10^{10}$ /см³ (300 К).
5. Диффузионные токи в полупроводниках. Плотность электронов в кристалле GaAs (*n*-типа) изменяется для значений $x > 0$ по закону $n(x) = A \exp(-x/L)$, где $A = 8 \times 10^{15}$ /см³, $L = 900$ нм. Вычислите плотность диффузионного тока при $x = 0$, если коэффициент диффузии электронов равен 190 см²с⁻¹.

6. Диффузионная длина. В образец GaAs (p -типа) через контакт инжектируются электроны. Полагая подвижность неосновных носителей равной $3700 \text{ см}^2\text{В}^{-1}\text{с}^{-1}$, рассчитайте диффузионную длину электрона при комнатной температуре для времени рекомбинации $\tau_n = 0,6 \text{ нс}$.
7. Динамика носителей в полупроводниках. Вблизи вершины валентной зоны энергия электронов хорошо описывается выражением $E(k) = -9 \times 10^{-36} k^2(\text{Дж})$, где k — волновой вектор. Считая, что электрон удаляется из состояния $k = 2 \times 10^9 k_u$ (где k_u — единичный вектор в направлении x), вычислите для образующейся дырки следующие параметры: а) эффективную массу; б) энергию; в) импульс; г) скорость. (Подсказка: предположим, что рассматриваемые состояния находятся очень близко к максимуму валентной зоны или минимуму зоны проводимости, где параболическое дисперсионное соотношение может служить хорошим приближением.)
8. Энергетические зоны в полупроводниках. Известно, что для некоторого полупроводника зона проводимости описывается соотношением $E_{cb}(k) = E_1 - E_2 \cos(ka)$, а валентная зона — соотношением $E_{vb}(k) = E_3 - E_4 \sin^2(ka/2)$, где $E_3 < E_1 - E_2$, $\pi/a \leq k \leq +\pi/a$. Определите для такого полупроводника: а) ширину запрещенной зоны; б) разницу между предельными значениями ($E_{\max} - E_{\min}$) для зон проводимости и валентной зоны; в) эффективные массы электронов и дырок для дна зоны проводимости и потолка валентной зоны соответственно. Нарисуйте схематически зонную структуру.
9. Неравновесные носители заряда. Образец германия (n -типа) освещается таким образом, что концентрация дырок в нем повышается в 5 раз. Определите время, необходимое для того, чтобы концентрация дырок снизилась до значения 10^{11} см^{-3} , при условии $\tau_h = 2,5 \text{ мс}$. Предполагается, что концентрация собственных носителей в полупроводнике составляет 10^{13} см^{-3} , а плотность доноров — $8 \times 10^{15} \text{ см}^{-3}$.
10. Оптическое поглощение в полупроводниках. Образец кремния толщиной 700 нм облучается источником монохроматического красного света ($\lambda = 600 \text{ нм}$) мощностью 40 Вт. Определите: а) поглощаемую полупроводником энергию; б) долю энергии, переходящей в тепло; в) число испускаемых за секунду фотонов, возникающих при таком режиме облучения. Коэффициент поглощения кремния при $\lambda = 600 \text{ нм}$



равен $\alpha = 7 \times 10^4 \text{ см}^{-1}$, а ширина запрещенной зоны в кремнии составляет 1,12 эВ.

11. Экситоны в GaAs. В рассматриваемой задаче можно принять, что приведенная масса экситона равна $\mu = 0,05m_0$, а относительная диэлектрическая проницаемость равна $\epsilon_r = 13$. Рассчитайте для этого случая энергию Ридберга R_H . Оцените предельное значение энергии связи. Оцените радиус экситона r_{ex} в такой системе для основного состояния, исходя из квантовых теорий Бора. Рассчитайте число элементарных ячеек для GaAs ($a_0 = 0,56 \text{ нм}$) в объеме экситона. Предположим, что возникающие экситоны находятся в основном состоянии. До какой температуры эти экситоны можно считать устойчивыми?
12. Экситоны в магнитном поле. Предположим, что кристалл арсенида галлия помещен в магнитное поле B . Оцените величину поля B , при котором циклотронная энергия экситона равна энергии Ридберга для экситона в соответствии с уравнением (3.60), считая диэлектрическую проницаемость среды равной $\epsilon = 13$.

ГЛАВА 4

ФИЗИКА ПОЛУПРОВОДНИКОВ С ПОНИЖЕННОЙ РАЗМЕРНОСТЬЮ

4.1. Введение

Можно сказать, что в течение последних десятилетий основные направления исследований в физике твердого тела почти непрерывно смещались от изучения объемных кристаллов к структурам, размеры которых (хотя бы в одном из трех геометрических измерений) очень малы. В силу относительной легкости изготовления вначале внимание исследователей привлекали очень тонкие твердотельные пленки, толщина которых сопоставима с длиной волны де Бройля λ_b для электронов в твердых телах. Такие работы позволили обнаружить целый ряд новых явлений, среди которых можно отметить квантовый эффект Холла (КЭХ), открытый фон Клитцингом в 1980 г., за что он позднее (1985) и был удостоен Нобелевской премии.

Мы будем говорить, что твердые тела (обычно полупроводниковые материалы) имеют низкую размерность или просто называть их низкоразмерными в тех случаях, когда один из их геометрических размеров имеет порядок λ_b , хотя в некоторых ситуациях удобнее пользоваться и другими характеристическими длинами, введенными в разделе 1.3. Для большинства практически используемых полупроводников величина λ_b лежит в диапазоне от 10 до 100 нм, т. е. мы можем наблюдать в них интересные нас квантовые эффекты (типа квантового эффекта Холла, кулоновской блокады, квантовой проводимости и т. п., описанных в главах 6 и 7) именно в нанометровом диапазоне.

Напомним, что в главе 1 уже отмечалось, что низкоразмерные объекты проще всего классифицировать в соответствии с числом пространственных измерений, имеющих нанометровые параметры. Например, если одно из измерений является нанометровым, то структуру следует считать двумерной (2D) и называть квантовой ямой. Одномерными (1D) являются квантовые проволоки (в которых два геометрических измерения имеют нанометровый характер) и наконец нульмерными (0D) следует называть объекты, у которых все три измерения имеют порядок λ_B .

Как будет показано в этой главе, в современной оптоэлектронике широко применяются структуры с очень тонкими, нанометровыми слоями полупроводников (например, слоистая структура из тонкой пластины GaAs, окруженной с двух сторон слоями полупроводника AlGaAs, обладающего более широкой запрещенной зоной). Другие очень интересные структуры могут быть получены образованием переходов (или, более строго, гетеропереходов) между двумя полупроводниками с различными запрещенными зонами (раздел 1.7). В обоих случаях на границе раздела возникают потенциальные ямы для электронов, похожие на те, которые образуются в МДП-структурах. Если ширина таких ям сопоставима с параметром λ_B , то энергетические уровни электронов в ямах начинают квантоваться. Такие гетеропереходы уже стали основой для изготовления, например, транзисторов типа MODFET с очень высоким быстродействием (см. главу 9).

4.2. Основные характеристики двумерных полупроводниковых наноструктур

Одной из наиболее широко используемых на практике двумерных полупроводниковых структур является трехслойная структура (так называемый сэндвич) из пленки арсенида галлия GaAs нанометровой толщины, окруженной с двух сторон слоями полупроводника, например алюмината арсенида галлия $\text{Al}_x\text{Ga}_{1-x}\text{As}$, с более широкой запрещенной зоной. Например, при ($x \sim 0,3$) запрещенная зона $\text{Al}_x\text{Ga}_{1-x}\text{As}$ близка к 2 эВ, в то время как в GaAs она равна 1,4 эВ. В результате возникает профиль потенциальной энергии, близкий по форме к прямоугольному (рис. 4.1, а) с высотой барьера 0,4 эВ (для электронов) и 0,2 эВ (для дырок). В действительности, конечно, по-

тенциальный барьер имеет более сложную форму, поскольку потенциал зависит и от межатомного расстояния, влияющего на волновые функции. Однако для многих случаев можно рассматривать потенциал, усредненный по нескольким межатомным расстояниям (приближение огибающей функции). Как видно из рис. 4.1, *a*, из-за наличия стенок ямы движение носителей (и электронов, и дырок) не может происходить вдоль оси *z*, перпендикулярной яме, однако в двух остальных направлениях (*x* и *y*, параллельных плоскости границы раздела) электроны остаются свободными и их движение ничем не ограничено.

Поведение электронов в системах, где их движение ограничено вдоль одной из координатных осей, приводит нас к известной квантово-механической задаче о движении частицы в так называемом потенциальном ящике с бесконечно высокими стенками. В разделе 4.3 мы рассмотрим эту задачу для случая барьеров конечной высоты на границах раздела. Из квантовой механики известно, что волновые функции и энергетические уровни связанных электронных состояний в такой ситуации (при наличии бесконечного по высоте потенциального барьера) определяются формулами

$$\psi_n(z) = \left(\frac{2}{a}\right)^{1/2} \sin\left(\frac{\pi n z}{a}\right), \quad (4.1)$$

$$E_n = \frac{\hbar^2 \pi^2}{2m_e^* a^2} n^2, \quad (n = 1, 2, \dots), \quad (4.2)$$

где величина m_e^* соответствует эффективной массе электрона при движении в данном веществе вдоль оси *z*, а *a* — ширина потенциальной ямы. Из уравнения (4.2) можно сразу получить несколько важных характеристик поведения системы. Прежде всего, ясно, что квантовые эффекты должны особенно сильно проявляться в структурах с малыми значениями *a*, причем именно в материалах, где эффективная масса электрона является очень малой. Рассматриваемые нами наноструктуры GaAs очень удобны в этом отношении, так как в этом материале $m_e^* \sim 0,067 m_0$ (m_0 — масса свободного электрона). Другими словами, квантовые эффекты гораздо легче наблюдать в системах, где высоки значения подвижности электронов (раздел 3.5.1) или

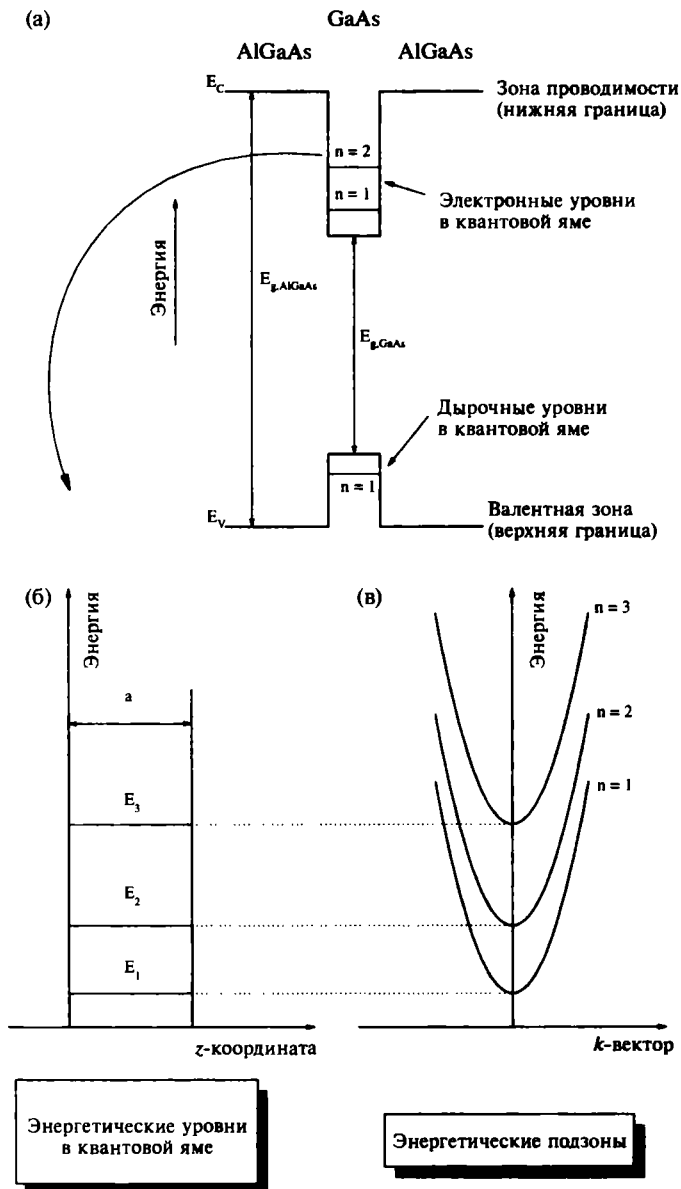


Рис. 4.1. (а) Прямоугольная потенциальная яма в трехслойной системе AlGaAs – GaAs – AlGaAs; (б) энергетические уровни; (в) энергетические подзоны

длина их свободного пробега. Кроме этого, столь же ясно, что размерные квантовые эффекты (связанные с электронными переходами между уровнями) гораздо удобнее наблюдать при низких температурах, поскольку средняя тепловая энергия носителей заряда имеет порядок kT .

В соответствии с предложенным описанием движение электронов в квантовых ямах ограничено лишь в одном направлении (z), а в плоскости (x, y) они остаются свободными и их поведение аналогично наблюдаемому в трехмерных твердых телах. Поэтому представляется полезным записать волновую функцию электрона в виде произведения волновых функций по отдельным осям (ψ_x, ψ_y и ψ_z), т. е. представить ее в виде

$$\Psi = \psi_x \psi_y \psi_z, \quad (4.3)$$

где в соответствии с используемой нами очень простой моделью функции ψ_x и ψ_y должны являться решениями уравнения Шрёдингера для свободного электрона, т. е. описывать бегущую волну. Одновременно третья функция ψ_z должна представлять собой решение того же уравнения Шрёдингера, но для частицы в прямоугольной потенциальной яме $V(z)$, т. е. удовлетворять приведенному выше уравнению (4.1).

С учетом сказанного для полной энергии электронов в потенциальной яме можно выписать соотношение

$$E(k_x, k_y, n) = \frac{\hbar^2}{2m_e^*} (k_x^2 + k_y^2) + E_n = \frac{\hbar^2}{2m_e^*} (k_x^2 + k_y^2) + \frac{\hbar^2 \pi^2}{2m_e^* a^2} n^2, \quad (n = 1, 2, \dots), \quad (4.4)$$

где квазинепрерывные значения k_x и k_y должны определяться в соответствии с периодическими граничными условиями, как и в случае свободных электронов в объемном твердом теле.

На рис. 4.1, б схематически показаны дискретные значения E_n для движения электронов вдоль направления z , а на рис. 4.1, в представлены зависимости $E(p)$ (напомним, что $p = \hbar k$) для значений \vec{p} в плоскости (p_x, p_y). Для каждого фиксированного E_n значения E (как функции p) определяют так называемые энергетические подзоны, представленные на рис. 4.1, в. Особый интерес представляет то, что минимальная энергия электронов

E_1 отлична от нуля. Этот результат явно противоречит принципам классической механики, но согласуется с законами квантовой механики, так как значение $E_1 = 0$ нарушало бы принцип неопределенности. Для описываемых систем значение $E = E_1$ называют нулевой энергией.

Из курсов физики твердого тела известно, что многие физические характеристики материалов (такие, как оптическое поглощение, перенос зарядов и т. п.) зависят одновременно и от энергетического спектра, и от вида функции плотности состояний, определяющей концентрацию электронов для каждого конкретного значения энергии. В разделе 2.3 мы уже показали, что для трехмерной электронной системы эта функция имеет параболический вид. В двумерном случае ситуация кардинально изменяется. Следуя методу, использованному для трехмерной задачи, можно показать, что в двумерных системах (рис. 4.2) разрешенные значения k_x , k_y имеют периодичность $2\pi/L$, где L — размер образца, который без потери общности рассмотрения предполагается квадратным. Число состояний в k -пространстве внутри кругового кольца, заключенного меж-

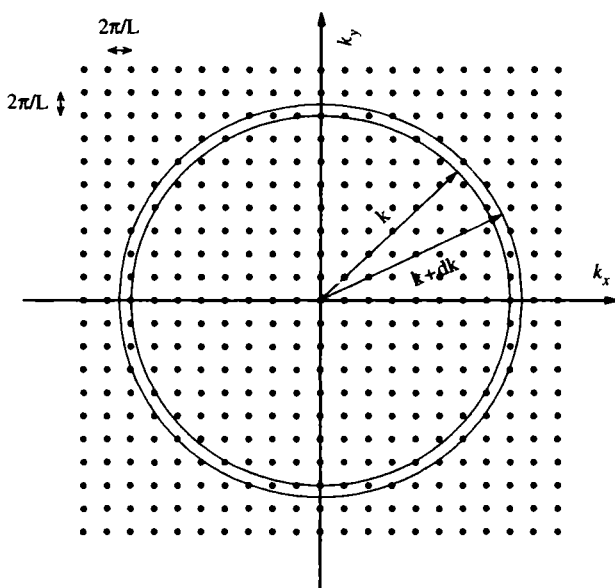


Рис. 4.2. Представление состояний (k_x, k_y) в k -пространстве для двумерной электронной системы

ду k и $k + dk$, равно

$$n_{2D}(k)dk = \frac{2\pi k dk}{(2\pi/L)^2},$$

а число состояний в k -пространстве на единицу площади составляет

$$n_{2D}(k) = \frac{k}{2\pi}. \quad (4.5)$$

Если мы хотим вычислить функцию плотности состояний по энергии в подзоне, нам следует определить функцию $n_{2D}(E)$ таким образом, чтобы величина $n_{2D}(E)\delta E$ соответствовала числу состояний в диапазоне значений δE . Плотности состояний по энергиям и волновые векторы обусловлены соотношением

$$n_{2D}(E)\delta E = n_{2D}(k)\delta k, \quad (4.6)$$

где величины E и k связаны формулой $E = \hbar^2 k^2 / 2m_c^*$. Дифференцируя выражение (4.6) после подстановки в него (4.5) и введения коэффициента 2 (для учета спина электронов), получим окончательно

$$n_{2D}(E) = 2 \frac{k}{2\pi} \frac{\delta k}{\delta E} = 2 \frac{k}{2\pi} \frac{m_c^*}{\hbar^2 k} = \frac{m_c^*}{\pi \hbar^2}. \quad (4.7)$$

Следует сразу отметить, что полученная для двумерного случая функция плотности состояний является постоянной, т. е. не зависит от энергии. Покажем, что для двумерных систем полная функция плотности состояний, график которой приведен на рис. 4.3, является ступенчатой, причем все ступени имеют одинаковую высоту, но располагаются на дискретных значениях E_n , соответствующих решениям уравнения (4.2). Действительно, из рис. 4.1, *в* можно заключить, что значения энергии между 0 и E_1 являются запрещенными. При значениях E в интервале $E_1 < E < E_2$ электроны могут располагаться в подзоне, соответствующей $n = 1$, где функция плотности состояний, как следует из 4.7, равна $m_c^* / \pi \hbar^2$. В интервале

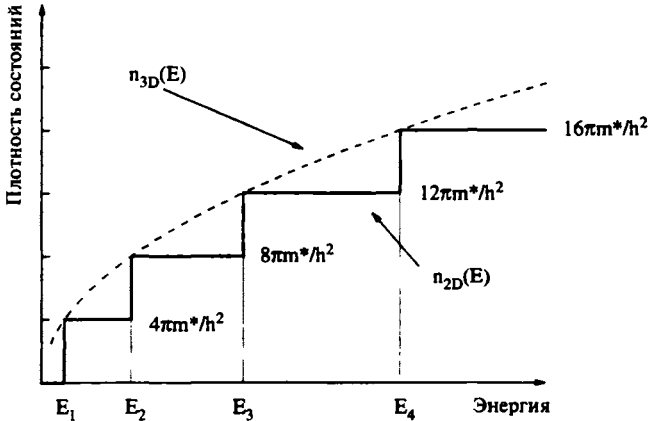


Рис. 4.3. Функция плотности состояний для двумерной электронной системы

между E_2 и E_3 электроны могут располагаться одновременно в двух подзонах (соответствующей $n = 1$ и $n = 2$), вследствие чего функция плотности состояний просто удваивается по значению, т. е. становится равной $2m_e^*/\pi h^2$, и т.д. Ступенчатый характер функции $n_{2D}(E)$ подтверждается прямыми измерениями оптического поглощения, как показано в главе 8. На рис. 4.3 для наглядности представлена и огибающая, т. е. параболическая кривая для трехмерного случая, демонстрирующая, что разница между двумерными и трехмерными системами значительно сильнее проявляется при малых значениях n .

Относительная доля кинетической энергии зависит от номера подзоны, к которой принадлежит данный электрон, и описывается уравнением (4.4). Например, при одном и том же значении энергии в интервале между E_2 и E_3 электроны в подзоне с $n = 2$ обладают большей энергией в z -направлении (E_2), чем электроны подзоны $n = 1$ с энергией E_1 . Поэтому для движения электронов подзоны с $n = 2$ в плоскости (p_x, p_y) остается меньше энергии, чем для подзоны с $n = 1$. Такое разделение энергии по отдельным «компонентам» (напомним, что энергия является скалярной величиной) является прямым следствием использованного выше упрощенного представления для волновой функции ψ и энергии в уравнениях (4.3) и (4.4).

4.3. Прямоугольная потенциальная яма конечной глубины

Решая в предыдущем разделе задачу 4.2 о поведении электронов и дырок в наноструктуре GaAs, покрытой с двух сторон пленками AlGaAs с более широкой запрещенной зоной, мы получили замкнутую систему решений для волновых функций и энергетических уровней, предполагая, что потенциальные ямы имеют бесконечную глубину. В реальных материалах очевидно, что глубина ямы имеет конечную величину и должна совпадать со значением разрыва ΔE_c зоны проводимости на границе раздела сред GaAs и AlGaAs, которое для рассматриваемой задачи составляет несколько десятых эВ. Однако для практических целей гораздо удобнее воспользоваться полученным решением для ямы бесконечной глубины, но не рассматривать решения, относящиеся к очень глубоким уровням энергии электронов.

Сказанное поясняет рис. 4.4. Называя величину V_0 глубиной прямоугольной потенциальной ямы, мы подразумеваем, что при энергиях $E < V_0$ мы имеем так называемые связанные состояния (т. е. электроны, захваченные или удерживаемые внутри ямы шириной a), а при $E > V_0$ мы имеем дело с непрерывными дело-

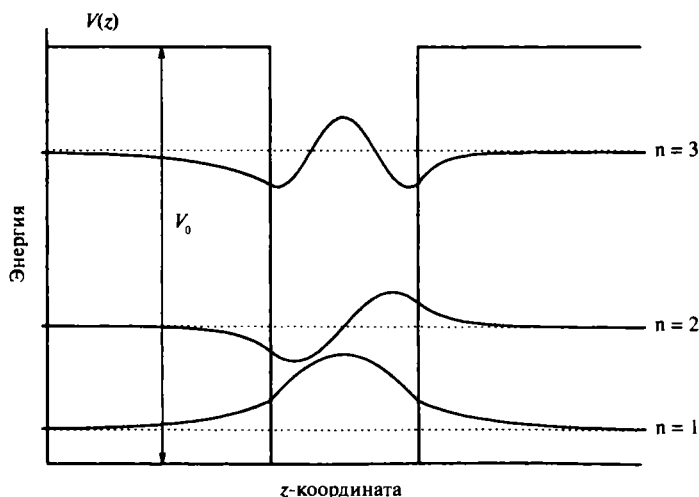


Рис. 4.4. Первые три разрешенных энергетических уровня и соответствующие им волновые функции для прямоугольной потенциальной ямы конечной глубины

кализованными состояниями, в которых электроны могут свободно двигаться вдоль оси z от $z = -\infty$ до $z = +\infty$. Так как эта задача обладает инверсной симметрией относительно центра ямы, удобно выбрать эту точку в качестве начала координатной оси z -координат. Понятно, что волновые функции связанных состояний внутри потенциальной ямы при этом должны иметь тот же вид, что и для ямы бесконечной глубины, т. е. соответствующие им решения могут сохранять свою симметричность или несимметричность, выражаясь через те же функции синусов и косинусов. С другой стороны, из квантовой механики известно, что решения для пространства вне потенциальной ямы (полученные из уравнения Шрёдингера при потенциальной энергии V_0) должны представлять собой экспоненциально убывающие функции. Поэтому естественно искать общие решения для волновых функций в виде линейных комбинаций именно таких функций:

$$\psi_n(z) = \begin{cases} D \exp(kz), & z < -a/2, \\ C \cos(kz), C \sin(kz), & -a/2 < z < a/2, \\ D \exp(-kz), & z > a/2, \end{cases} \quad (4.8)$$

где $k = \left(\frac{2m_c^* E}{\hbar^2} \right)^{1/2}$ внутри потенциальной ямы (4.9a)

и $k = \left(\frac{2m_c^* (V_0 - E)}{\hbar^2} \right)^{1/2}$ вне потенциальной ямы (4.9б)

Отметим, что при выводе уравнения (4.9) предполагается, что эффективная масса электрона одинакова для потенциальной ямы и за ее пределами. Это приближение является достаточно хорошим по двум причинам. Во-первых, вещества барьера и ямы (GaAs и AlGaAs с небольшой примесью алюминия) достаточно схожи по своим физическим характеристикам, а во-вторых, вероятность проникновения волновых функций в барьеры для низших связанных состояний достаточно мала. Рассматривая решения задачи, мы должны учитывать, что функции $\psi(z)$ должны быть непрерывными, вследствие чего все решения уравнений (4.8) должны совпадать при $z = \pm a/2$, причем в приближении одинаковой эффективной массы должны совпадать не только сами функции, но и их производные. Из этого

требования можно получить трансцендентное уравнение, которое может быть легко решено численными методами [1], причем решение приводит к нескольким важным выводам. Например, можно показать, что для одномерной ямы всегда можно найти, по крайней мере, одно связанное состояние, независимо от того, насколько мало значение V_0 . Далее, для слабосвязанных состояний коэффициент убывания k в экспоненте уравнения (4.8) не очень велик, вследствие чего волновые функции, представленные на рис. 4.4, могут достаточно глубоко проникать в область барьера. С другой стороны, можно показать, что для сильносвязанных состояний глубина проникновения в запрещенную зону очень мала.

4.4. Параболическая и треугольная квантовые ямы

4.4.1. Параболическая потенциальная яма

Модель параболической потенциальной ямы широко используется в физике твердого тела, поскольку она очень часто применяется для описания колебаний атомов в кристаллических решетках, при квантовании которых (в первом приближении гармонического осциллятора) и возникает представление о фотонах. Кроме этого, параболический потенциал часто возникает в задачах, где рассматривается воздействие магнитного поля на двумерные электронные системы, а также колебания электронов при так называемых циклотронных частотах. Более того, квантовые ямы с параболическим потенциалом могут быть получены экспериментально с использованием молекулярно-лучевой эпитаксии (МВЕ). Например, таким методом может быть получена слоистая структура из чередующихся пленок GaAs и $\text{Al}_x\text{Ga}_{1-x}\text{As}$, в которой толщина слоев AlGaAs квадратично возрастает с расстоянием, в то время как толщина соответствующих слоев GaAs уменьшается в такой же пропорции.

Известно, что разрешенные энергии в случае наиболее распространенного потенциала одномерного гармонического осциллятора

$$V(z) = \frac{1}{2}k^2z^2 \quad (4.10)$$

(k — постоянная) являются эквидистантными (т. е. отличаются друг от друга на одинаковую величину) и описываются формулой

$$E_n = \left(n - \frac{1}{2} \right) \hbar \omega_0, \quad n = 1, 2, 3, \dots, \quad (4.11)$$

где ω_0 — так называемая собственная угловая частота.

На рис. 4.5 показана параболическая потенциальная яма и волновые функции для трех первых связанных электронных состояний. Математические волновые функции записываются в виде полиномов Эрмита. Стоит отметить, что, как и в случае прямоугольной ямы, волновые функции симметричны или антисимметричны относительно центра ямы и экспоненциально убывают в запрещенной энергетической зоне.

4.4.2. Треугольная потенциальная яма

Треугольная потенциальная яма является одним из наиболее часто используемых приближений, поскольку потенциальный профиль поперек квантовых гетеропереходов (см. главу 5), та-

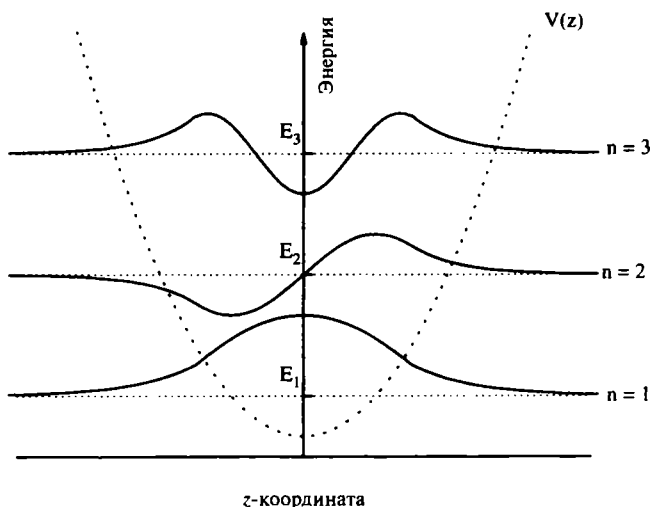


Рис. 4.5. Первые три энергетических уровня и соответствующие им волновые функции для потенциальной ямы $V(z)$ гармонического осциллятора

ких, как широко известных модулированно-легированных гетеропереходах $\text{AlGaAs}-\text{GaAs}$ для электронов в GaAs , имеет форму, близкую к треугольной. Из всех гетероструктур эта является наиболее исследованной и ее свойства будут более подробно рассмотрены ниже, в разделе 5.3.1. Еще один, очень важный случай, когда образуется потенциальная яма почти треугольной формы, реализуется в полупроводниках с МОП-структурой (раздел 5.2).

На рис. 4.6 представлена треугольная потенциальная яма, в которой для простоты считается, что с левой стороны потенциальный барьер является бесконечно высоким, а справа для $z > 0$ он возрастает по линейному закону, т. е. описывается уравнениями

$$V(z) = eFz, \quad \text{для } z > 0, \quad (4.12a)$$

$$V(z) = \infty, \quad \text{для } z \leq 0, \quad (4.12b)$$

где e — заряд электрона; F — однородное электрическое поле. Как и в предыдущих задачах, энергия и волновые функции состояний электронов определяются решениями уравнения Шрё-

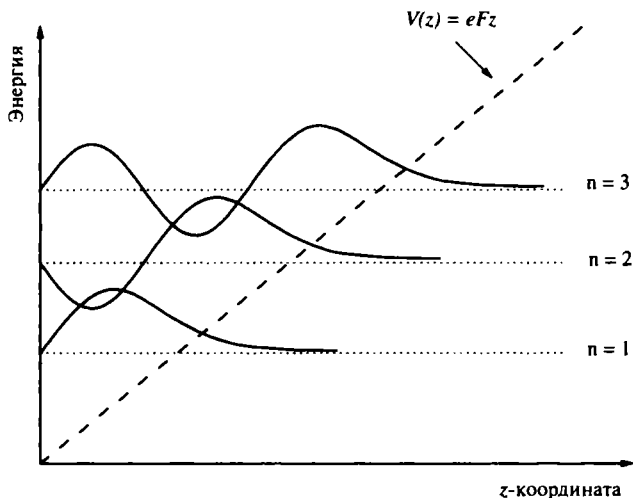


Рис. 4.6. Первые три энергетических уровня и соответствующие им волновые функции для треугольной потенциальной ямы $V(z)$

дингера при граничном условии $\psi(z = 0) = 0$. В этом случае собственные значения определяются известными в математической физике функциями Эйри. Однако, пользуясь известным методом ВКБ (квазиклассическим приближением), можно показать, что при малых значениях n , справедливо соотношение [1]

$$E_n = \left[\frac{3}{2} \pi \left(n - \frac{1}{4} \right) \right]^{2/3} \left(\frac{e^2 F^2 \hbar^2}{2m} \right)^{1/3}, \quad n = 1, 2, \dots \quad (4.13)$$

Показанные на рис. 4.6 промежутки между уровнями энергии несколько уменьшаются с ростом n , в противоположность параболическому потенциалу, при котором эти промежутки строго равны друг другу. На рисунке представлены также волновые функции, для которых можно сразу отметить, что с ростом n их периодичность возрастает на половину периода. Кроме того, в отличие от систем с параболическим потенциалом, эти функции не могут быть отнесены ни к симметричным, ни к антисимметричным, поскольку в данном случае нельзя говорить о симметрии относительно центра потенциальной ямы.

4.5. Квантовые проволоки

Модель двумерного электронного газа, рассмотренная в разделе 4.2, позволяет легко перейти к описанию одномерного электронного газа, в котором движение электронов ограничено сразу в двух измерениях (x и y), вследствие чего они могут свободно передвигаться только в одном направлении z , которое считается перпендикулярным к плоскости (x, y). С формальной точки зрения такое движение описывается аналогично распространению электромагнитной волны.

Предположим, что ограничивающий движение потенциал является функцией $r = (x, y)$, т. е. $V = V(\mathbf{r})$ и что к задаче применим метод разделения переменных, использованный для решения уравнения Шрёдингера в разделе 4.2. Исходя из тех же посылок, можно записать волновые функции в виде

$$\psi(\mathbf{r}) = e^{ik_z z} u(\mathbf{r}), \quad (4.14)$$

где волновая функция $u(\mathbf{r})$ может быть получена из решения следующего двумерного уравнения Шрёдингера:

$$\left[-\frac{\hbar^2}{2m_e} \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} \right) + V(\mathbf{r}) \right] u_{n_1, n_2}(\mathbf{r}) = E_{n_1, n_2} u_{n_1, n_2}(\mathbf{r}), \quad (4.15)$$

где $(n_1, n_2) = 1, 2, 3, \dots$

Выражение для полной энергии электронов в квантовой проволоке может быть записано в виде:

$$E_{n_1, n_2}(k_z) = E_{n_1, n_2} + \frac{\hbar^2 k_z^2}{2m_e^*}, \quad (4.16)$$

где последний член в правой части описывает кинетическую энергию движения электрона вдоль оси z .

В качестве примера вычисления энергии для конкретной системы мы рассмотрим простейший случай двумерного прямоугольного потенциала бесконечной длины с размерами (a_x, a_y) , т. е. потенциал вида

$$\begin{aligned} V(x, y) &= 0, & 0 < x < a_x, & 0 < y < a_y, \\ V(x, y) &= \infty, & x \leq 0, & x \geq a_x, & y \leq 0, & y \geq a_y, \end{aligned} \quad (4.17)$$

для которого, используя полученный в разделе 4.2 результат, можно сразу выписать значения разрешенных уровней энергии:

$$E_{n_1, n_2} = \frac{\hbar^2 \pi^2}{2m_e^*} \left(\frac{n_1^2}{a_x^2} + \frac{n_2^2}{a_y^2} \right), \quad n_1, n_2 = 1, 2, 3, \dots \quad (4.18)$$

Другая достаточно легко решаемая в цилиндрических координатах задача связана с рассмотрением проволоки с круглым сечением, когда решения могут быть найдены в виде функций Бесселя. Вообще говоря, в случае квантовых проволок энергетические уровни, соответствующие поперечному движению, описываются двумя квантовыми числами — n_1 и n_2 , соответствующими конкретным значениям E_{n_1, n_2} для дна параболической одномерной подзоны в пространстве k_z . Отметим также, что значения уровней E_{n_1, n_2} для электронных состояний возрастают при уменьшении толщины квантовых проволок.

Рассмотрим далее задачу о плотности состояний одномерного электронного газа. Концентрация состояний (в зависимости от энергии) связана с волновым числом соотношением

$$n_{1D}(E)\partial E = n_{1D}(E)\frac{dE}{dk}\partial k = 2n_{1D}(k)\partial k, \quad (4.19)$$

где коэффициент 2 вводится для учета двух возможных направлений движения электронов вдоль проволоки. Обобщая рис. 4.2 на одномерный случай, легко заключить, что плотность состояний в k -пространстве равна $1/2\pi$. Подставляя это значение в (4.19) и учитывая, что $E = \hbar^2 k^2 / 2m_c^*$, получим окончательно выражение

$$n_{1D}(E) = \frac{1}{\pi\hbar} \sqrt{2m_c^*/E}, \quad (4.20)$$

которое очевидно расходится при $E = 0$.

Используя введенное ранее представление (2.52) о групповой скорости v_g , можно переписать полученное выражение в форме

$$n_{1D}(E) = \frac{2}{\pi\hbar v_g}, \quad (4.21)$$

из которого следует интересный результат, что ток в одномерной системе является постоянным, а его величина пропорциональна скорости и плотности состояний. Мы еще вернемся к уравнению (4.21) для плотности состояний в квантовых проволоках в главе 5 при рассмотрении задачи о квантовании проводимости в таких системах.

Из уравнения (4.20) можно получить также выражение для полной плотности состояний (в пересчете на единицу длины проволоки) в виде

$$n_{1D}(E) = \sum_{n_1, n_2} \frac{1}{\pi\hbar} \sqrt{\frac{2m_c^*}{E - E_{n_1, n_2}}}. \quad (4.22)$$

Функция плотности состояний для одномерной системы представлена на рис. 4.7, которую можно сравнить с параболической функцией плотности состояний для трехмерной системы. Важнейшее различие состоит в том, что в одномерных системах функция расходится при значениях энергии E_{n_1, n_2} , соответствующих дну подзон, что имеет важные следствия для физических характеристик квантовых проволок.

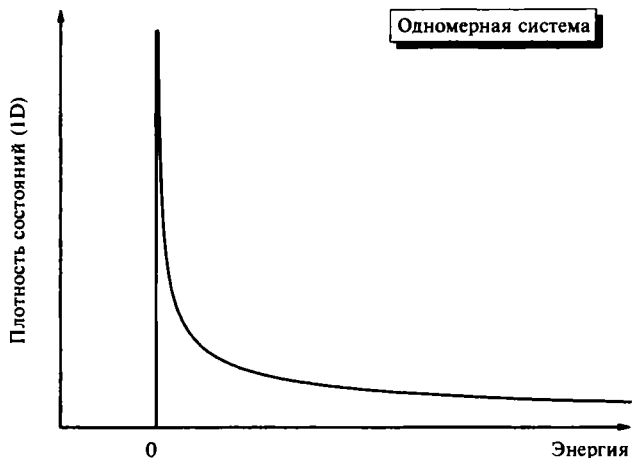


Рис. 4.7. Функция плотности состояний в зависимости от энергии для одномерной электронной системы

4.6. Квантовые точки

Квантовые точки часто представляют собой нанокристаллы, обладающие всеми тремя пространственными измерениями в нанометровом масштабе. Иногда в случае полупроводников типа II—VI (например, в CdSe или CdS) такие нанокристаллы могут быть получены выращиванием из жидкофазных растворов при строго заданных температурах. В других случаях их можно получать из макроскопических материалов, применяя литографию и методы создания микротекстур требуемого типа.

Хотя сам термин «точка» как бы подразумевает бесконечно малые размеры объекта, реальные квантовые точки могут содержать достаточно большое число атомов (до 10^4 — 10^6), сохраняя при этом наномасштабы во всех трех измерениях (последнее утверждение означает, что длина волны де Бройля остается соизмеримой со всеми размерами объекта). Очень часто квантовые точки называют искусственными атомами, поскольку, как будет показано ниже, их энергетические спектры напоминают спектры атомов. По аналогии с атомами мы можем определить для квантовых точек энергию ионизации, т. е. энергию, необходимую для ввода в систему (или удаление из нее) дополнительного электрона. Эту энергию по по-

нятым причинам иногда называют энергией зарядки точки, исходя из обычных представлений об электрической емкости систем, когда добавление или удаление электрических зарядов определяется кулоновским взаимодействием. Многие «атомарные» свойства квантовых точек изучаются именно посредством измерения их электрических характеристик. С этой точки зрения очень важно отметить, что введение или удаление хотя бы одного электрона в квантовую точку может, в отличие от одномерных (1D) или двумерных (2D) систем, весьма существенно изменить электрические характеристики точки, что проявляется в таких известных эффектах, как большие колебания проводимости и эффект кулоновской блокады (см. главу 6).

Рассмотрим энергетический спектр квантовых точек в простейшем случае, когда ограничивающий потенциал равен нулю внутри некоторого ящика (размерами a_x , a_y и a_z) и бесконечен вне этого объема. Решением этой широко известной в квантовой механике задачи являются стоячие волны в качестве волновых функций электронов и энергетические уровни вида

$$E_{n_1, n_2, n_3} = \frac{\hbar^2 \pi^2}{2m_e^*} \left(\frac{n_1^2}{a_x^2} + \frac{n_2^2}{a_y^2} + \frac{n_3^2}{a_z^2} \right), \quad n_1, n_2, n_3 = 1, 2, 3, \dots \quad (4.23)$$

В отличие от одномерных и двумерных систем, энергия в этом случае квантуется полностью (по всем трем измерениям), и, подобно атомарным системам, решения не допускают никаких свободных движений для электронов. Однако стоит отметить, что уровни часто являются вырожденными, например, в тех случаях, когда два или три измерения ящика совпадают по размерам.

Можно получить также точное решение этой задачи для сферической точки, когда потенциал равен нулю внутри сферы и бесконечности вне ее. В такой постановке задача явно напоминает задачу о сферически симметричном атоме, для которой известно, что энергия определяется двумя квантовыми числами, а именно главным квантовым числом n (возникающим при решении одномерного радиального уравнения) и квантовым числом углового момента l .

Поскольку в квантовых точках движение электронов полностью ограничено, энергетический спектр должен быть толь-

ко дискретным, а функция плотности состояний должна (по крайней мере, теоретически) представлять собой набор пиков с бесконечно малой шириной и бесконечной высотой, как показано на рис. 4.8. Понятно, что практически такие пики имеют малую (но конечную) ширину, хотя бы потому, что электроны не являются абсолютно независимыми, и должны как-то взаимодействовать с фононами и примесными атомами решетки.

4.7. Напряженные слои

Вообще говоря, качество границы раздела между двумя материалами определяется соотношением постоянных их кристаллических решеток. В некоторых случаях, когда эти постоянные очень близки по значению (как, например, в гетеропереходах типа $\text{GaAs}-\text{Al}_x\text{Ga}_{1-x}\text{As}$, для которых различия в значениях постоянных решетки не превышают 0,2% во всем диапазоне изменений x), и коэффициенты теплового расширения почти совпадают, возникающие на поверхности контакта напряжения незначительны. Однако во многих других случаях (см. рис. 4.9) гетеропереходы образуются из материалов, постоянные решетки которых отличаются довольно значительно (до 6%), как это имеет место, например, в случае контакта $\text{GaAs}-\text{In}_x\text{Ga}_{1-x}\text{As}$.

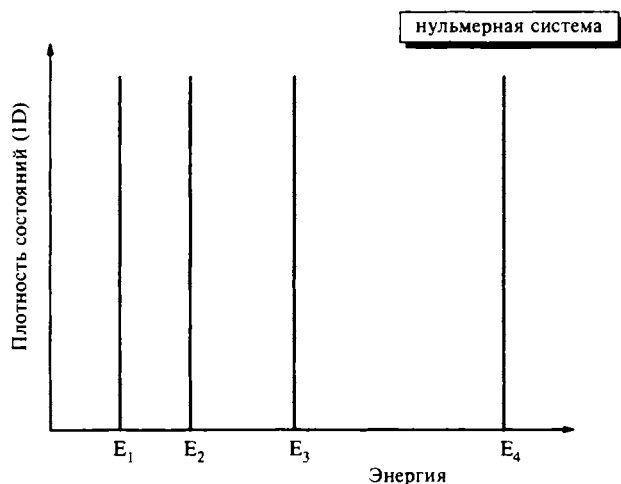


Рис. 4.8. Функция плотности состояний для нульмерной (0D) электронной системы

Такие переходы характеризуются значительными механическими напряжениями, в результате чего на поверхностях подложек удается выращивать только очень тонкие пленки, состоящие буквально из нескольких атомарных слоев. Тем не менее такие напряженные слои проявляют целый ряд новых свойств и эффектов, которые уже используются в разнообразных оптоэлектронных приборах, например в лазерах на квантовых ямах и электрооптических модуляторах (см. главу 10). На этой основе возникло новое направление, получившее название эпитаксии напряженных слоев. Обычно при эпитаксиальном выращивании постоянная решетки выращиваемого на подложке слоя несколько изменяется, «подстраиваясь» под постоянную решетки подложки, что даже получило специальное название псевдоморфного роста. Возникающие при этом в растущем слое напряжения могут приводить к очень важным физическим эффектам, например, к снятию вырождения тяжелых и легких дырок в валентной зоне материала (см. раздел 4.8), к изменению ширины запрещенной зоны и т. д. Причины такого поведения полупроводников очевидны, так как зонная структура твердых тел чрезвычайно чувствительна не только к изменению

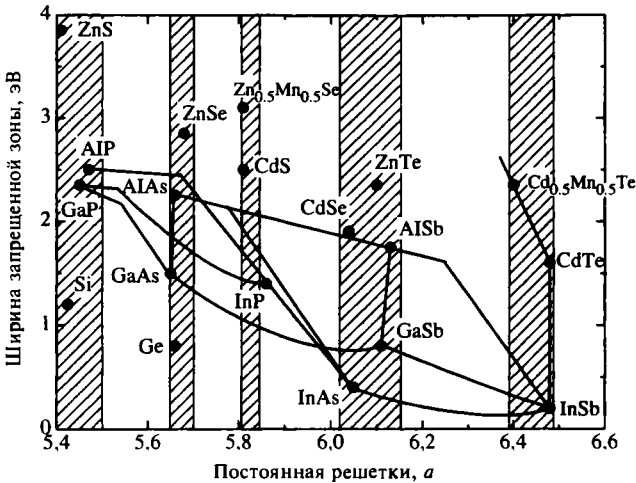


Рис. 4.9. Значения ширины запрещенной зоны и постоянных решетки в различных полупроводниках. Заштрихованные участки относятся к наиболее известным семействам (группам) полупроводников, способным образовывать несколько типов гетеропереходов

размеров элементарных ячеек решетки, но и к изменениям их симметрии.

Рассмотрим ситуацию, когда на подложке (с постоянной решетки a_s) эпитаксиально выращивается слой с иным значением постоянной решетки (a_l). Механическое напряжение ϵ в растущем слое определяется формулой

$$\epsilon = \frac{a_l - a_s}{a_s}. \quad (4.24)$$

В качестве примера такой системы на рис. 4.10 показана пленка $\text{In}_x\text{Ga}_{1-x}\text{As}$, выращенная на подложке GaAs (в такой системе $a_l > a_s$). На рис. 4.10, а проиллюстрирована ситуация отдельного слоя и подложки. Если эпитаксиальный слой является достаточно тонким, то атомы в нем выстраиваются в соответствии с кристаллическим строением подложки, в результате чего внутри такого слоя возникает сжимающее механическое напряжение вдоль плоскости роста, а межплоскостное расстояние в перпендикулярном направлении (направлении роста) возрастает (рис. 4.10, б). Наличие искажений в решетке InGaAs приводит к тому, что в системе накапливается упругая энергия, величина которой пропорциональна толщине пленки. Естественно, при достижении некоторого порогового значения (называемого *критической толщиной* h_c) в системе происходит релаксация, в результате чего становится энергетически выгодным рождение дислокаций вблизи границы раздела (4.10, в). Используя более сложную технику, можно вырастить слои с толщиной больше h_c и без дислокаций, естественно, такие структуры являются метастабильными. Поэтому обычно толщина выращиваемых слоев бывает несколько ниже значения h_c . Помимо уже упоминавшейся двухслойной системы ($\text{In}_x\text{Ga}_{1-x}\text{As}$ на подложке GaAs), в последние годы внимание исследователей в области микроэлектроники все больше привлекает другая перспективная система ($\text{Ge}_x\text{Si}_{1-x}/\text{Si}$).

4.8. Влияние напряжений на валентную зону

Для исследования электронных и оптоэлектронных характеристик многих важных полупроводниковых материалов (например, кремния и соединений A_{III}B_V) очень важно иметь

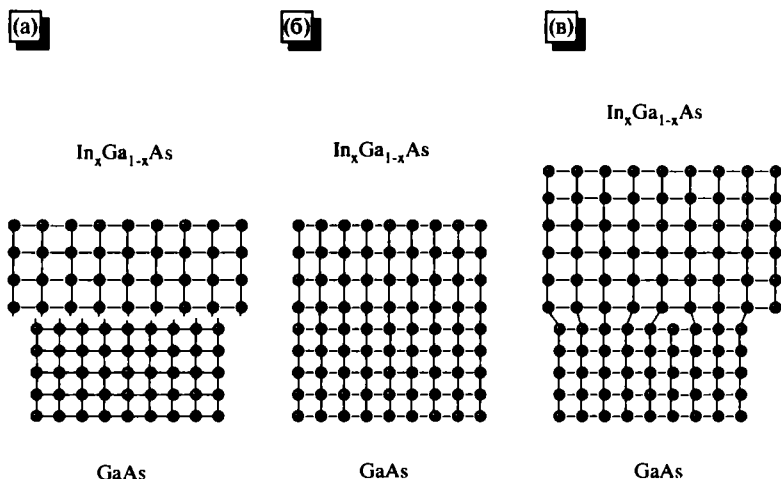


Рис. 4.10. Рост пленки кристалла с постоянной решеткой a_L на подложке с постоянной решеткой a_S ($a_L > a_S$): *a* — слой не связан с подложкой; *б* — выращиваемый слой испытывает сжимающие нагрузки; *в* — при толщине выращиваемого слоя больше критического значения, вблизи границе раздела образуются дислокации

подробную информацию о форме зоны проводимости и валентной зоны в k -пространстве. Для электронов в кристалле можно считать, что состояния в зонах возникают из верхних уровней электронов в атоме. В случае зоны проводимости, поведение электронов также часто может быть описано с использованием непрерывных состояний зон, которые для полупроводников с прямыми оптическими переходами относятся непосредственно к s -типу (используя систему обозначений атомной физики, можно сказать, что волновые функции вида $1s$ описывают связанную с решеткой периодическую часть блоховских функций). Однако в полупроводниках с непрямыми оптическими переходами ситуация выглядит сложнее, и состояния в зонах представляют собой некоторую комбинацию s и p состояний.

Еще сложнее выглядит ситуация в валентной зоне, где существуют три ветви решений, которые очень близки в области вокруг $k = 0$ и возникают из атомных p -состояний. Именно это наблюдается в большинстве важных полупроводников (типа Si, Ge и GaAs), где мы находим вырожденные дырочные состояния при

$k = 0$, т. е. вблизи потолка валентной зоны. На рис. 4.11 показаны две зоны, одна из которых имеет большую кривизну (эта зона соответствует легким дыркам), а другая — меньшую кривизну, что соответствует тяжелым дыркам. Кроме того, на рисунке представлена и третья, так называемая отщепленная зона (*split-off band*), возникающая ниже указанных двух зон, возникновение которой обусловлено релятивистскими эффектами при спин-орбитальных взаимодействиях. Очень часто такое расщепление достаточно велико, и отщепленную зону можно не учитывать, поскольку дырки не заполняют соответствующие состояния. Кроме этого, следует учитывать, что вблизи точки Γ состояния в зонах легких и тяжелых дырок достаточно хорошо описываются параболической зависимостью от k , поэтому при их описании можно пользоваться представлением об эффективных массах для тяжелых и легких дырок (m_{hh}^* и m_{lh}^* соответственно).

Для нахождения зависимости E от k в валентной зоне мы можем воспользоваться формализмом Латтинджера — Кона, предложившими для расчета зонной структуры метод $k \cdot p$, в которой собственными значения выступают угловые моменты p -состояний. Метод Латтинджера — Кона позволяет выразить зависимость энергии от k в виде следующей, достаточно простой формулы:

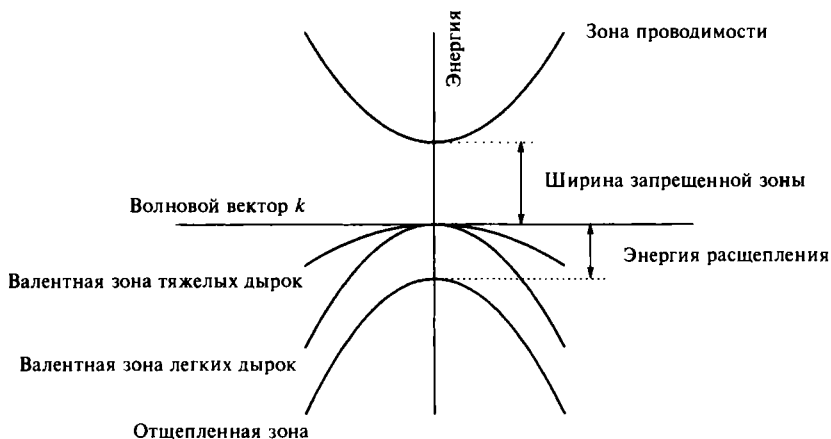


Рис. 4.11. Структура валентной зоны в некоторых типичных полупроводниках (Si, GaAs и т. п.) при $k = 0$. Показаны зоны, соответствующие тяжелым и легким дыркам, а также отщепленная зона

$$E(k) = E_v - \frac{\hbar^2}{2m_0} \left[Ak^2 \pm \sqrt{(Bk^2)^2 + C^2 (k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)} \right], \quad (4.25)$$

где верхний знак — плюс — перед корнем относится к тяжелым, а нижний — минус — к легким дыркам соответственно. Входящие в уравнение безразмерные величины A , B и C связаны с так называемыми параметрами Латтинджера — Кона (γ_1 , γ_2 и γ_3) соотношениями

$$A = \gamma_1, \quad B = 2\gamma_2, \quad C = 12(\gamma_3^2 - \gamma_2^2). \quad (4.26)$$

Для GaAs эти параметры имеют следующие конкретные значения: $\gamma_1 = 6,85$, $\gamma_2 = 2,1$, $\gamma_3 = 2,9$.

Влияние механических напряжений в слое на зонную структуру приобретает особое значение для дырочных валентных зон, что, естественно, вытекает из изменений постоянных кристаллической структуры и связанных с этим изменений симметрии. В результате появляется возможность управления шириной запрещенной зоны и вырождением в точке Г. Смещения энергетических уровней электронов и дырок вследствие механических напряжений вычисляются достаточно просто стандартными методами расчета зонной структуры.

Описываемые эффекты очень четко проявляются в изменении ширины запрещенной зоны и энергетической структуры валентных зон в полупроводниках $A_{III}B_V$. На графиках рис. 4.12 схематически представлена зонная структура слоев, выращенных вдоль оси (001) в отсутствие напряжения и в случае напряжений сжатия или растяжения в плоскости границы раздела. Как уже отмечалось, в первую очередь снимается вырождение в точке Г и зоны легких и тяжелых дырок в этой точке расщепляются. При сжимающих напряжениях более высокая валентная зона (зона тяжелых дырок) соответствует большему значению эффективной массы m_c (НН), хотя для движения в плоскости мы должны исходить из меньшего значения этой массы (рис. 4.12, б). При растягивающих напряжениях (рис. 4.12, в), когда постоянная решетки растущего слоя меньше постоянной подложки, мы сталкиваемся с обратным порядком энергетических зон, так как зона легких дырок оказывается при этом выше зоны тяжелых дырок. Видно, что в этом случае зоны пересекаются при некотором значении $k_{||}$

(k_{\parallel} — произвольный вектор в плоскости k_x, k_y), хотя обычно такие состояния просто не достигаются. Таким образом, наличие механических напряжений приводит, помимо анизотропии, к расщеплению уровней на величину, достигающую 0,05 эВ. Напряженные кристаллические структуры находят применения для изменения ширины запрещенной зоны в гетеропереходах (как в случае Si/SiGe, GaP/GaAsP), а также при производстве сине-зеленых лазеров на основе гетероструктур ZnCdSe/ZnSe.

4.9. Зонная структура в квантовых ямах

Для правильной интерпретации экспериментальных данных по оптическому поглощению в квантовых ямах мы должны более четко представлять себе строение зонной структуры изучаемых материалов. На рис. 8.2 (в главе 8) показаны спектры поглощения периодической структуры GaAs/AlGaAs с множественными квантовыми ямами (40 периодов) (такие системы обозначают аббревиатурой MQW — multiple quantum wells), ширина барьеров в которой составляет около 7,6 нм [3]. Отметим, что спектры в целом соответствуют ступеням в функции плотности

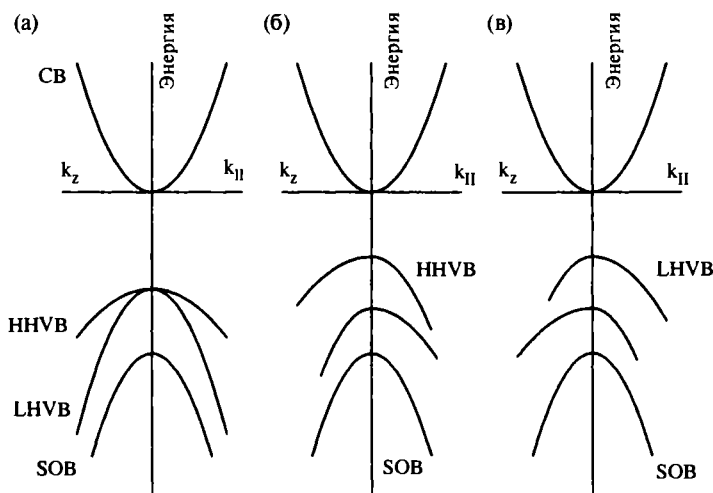


Рис. 4.12. Зонная структура для слоев, осажденных на подложке: а — ненапряженный слой; б — слой со сжимающими напряжениями; в — слой с растягивающими напряжениями

состояний для двумерных полупроводников (см. раздел 4.2). На краю каждой ступеньки проявляется резкий максимум, который в соответствии с изложенной в следующем разделе теорией следует приписать экситонным эффектам. Из рис. 8.2 также видно, что на границе первого электронного перехода между валентной зоной и зоной проводимости (при $n = 1$) возникает пик при 1,59 эВ, соответствующий тяжелым дыркам (НН) валентной зоны и пик при 1,61 эВ для легких дырок (ЛН), расположенный ниже пика для тяжелых (раздел 4.8).

Расщепление вырожденного состояния валентной зоны определяется тем, что одномерный потенциал, связанный с квантовыми ямами, нарушает кубическую симметрию кристаллов, в результате чего и снимается вырождение в дырочных зонах GaAs, подобно тому, которое было описано в предыдущем разделе для случая механических напряжений. Точные расчеты (мы не приводим их, чтобы не усложнять текст) показывают, что наличие потенциала ямы приводит к тому, что состояния для легких дырок смещаются вниз сильнее, чем состояния для тяжелых (см. рис. 4.13). Интересно, что если в расчетах пренебречь очень малыми членами в разложении для возмущенного гамильтониана, то зоны дырок могут даже пересекаться, после чего зона тяжелых дырок начнет смещаться вниз с большей скоростью. Результатом такого пересечения двух зон, обусловленным их различной кривизной, мог бы стать эффект, получивший название обращения массы. Такой эффект не наблюдается, так как более строгие вычисления показывают, что вместо пересечения должен наблюдаться другой эффект, получивший название антикроссинга (антипересечения).

В целом следует отметить, что зонная структура в квантовых ямах имеет очень сложный вид (особенно в отношении дырок), так что во многих случаях ее можно рассчитать только с применением численных методов. Дополнительные сложности возникают при расчете структур в прямоугольных ямах конечной высоты (раздел 4.3). На рис. 4.14 приведены результаты численного расчета для множественных квантовых ям (МКЯ) в структуре AlGaAs—GaAs [4]. Отметим, что форма многих зон сильно отличается от идеальной параболы, а в некоторых случаях наблюдается изменение знака кривизны, что подразумевает существование дырочных состояний с отрицательной эффективной массой.

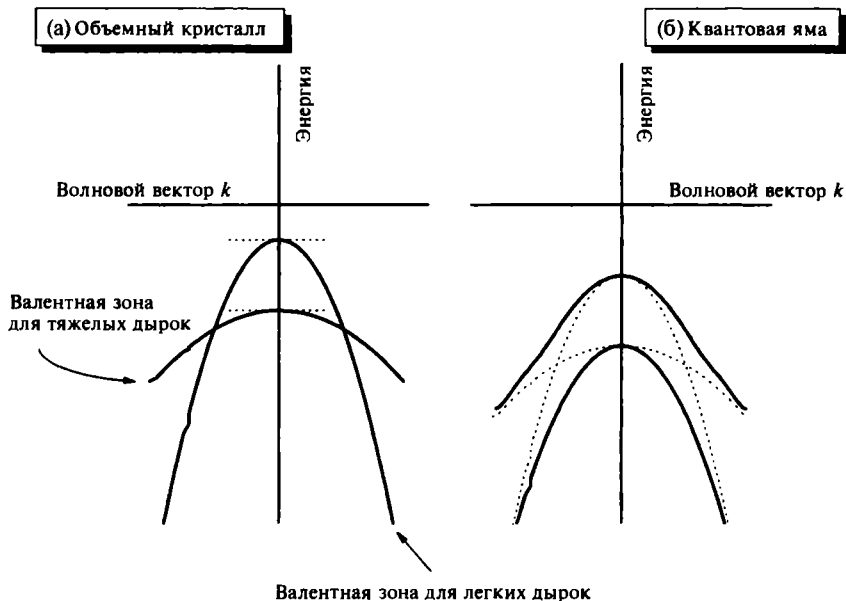


Рис. 4.13. (а) Валентные зоны в объемном кристалле GaAs; (б) расположение валентных зон в квантовой яме внутри кристалла GaAs

Выше было показано, что квантовые ограничения могут приводить к существенным изменениям энергетического спектра как для легких, так и для тяжелых дырок. Кроме того, положение энергетических уровней в зонах может изменяться под действием сжимающих или растягивающих напряжений (раздел 4.8). В некоторых случаях зоны легких дырок могут оказаться выше зон тяжелых дырок и даже быть вырожденными. Результатом такого вырождения может стать очень высокая концентрация дырочных состояний, и этот эффект используется в оптоэлектронных модуляторах.

4.10. Экситонные эффекты в квантовых ямах

В системах рассматриваемого типа (квантовые ямы, проволоки и точки), где движение электронов и дырок существенно ограничивается по сравнению с обычными объемными кристаллами, особую важность приобретают экситонные эффекты. Причиной

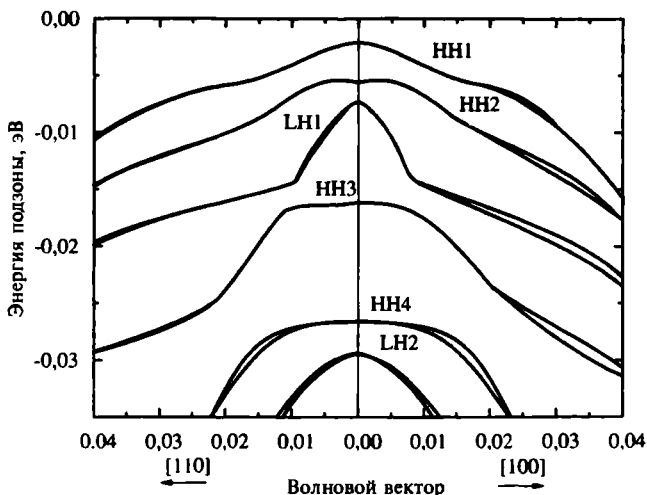


Рис. 4.14. Структура валентных зон для множественных квантовых ям (МКЯ) в структуре AlGaAs—GaAs (по данным работы [4]). Отметим, что некоторым зонам (HH2) соответствуют отрицательные эффективные массы дырок вблизи $k = 0$, а форма некоторых зон (HH3) значительно отличается от параболической

этого является и то, что *энергия связи*, необходимая для формирования экситона в виде электронно-дырочной пары, в системах с квантовыми ограничениями значительно превышает энергию такой же связи в объеме кристалла. Поэтому экситонные переходы в таких системах могут наблюдаться даже при комнатных температурах, в то время как в объемных структурах для их проявления необходимы очень низкие температуры.

Более высокое значение энергии связи в квантовых системах с ограниченными размерами легко объяснить качественно. Например, в объемном кристалле GaAs энергия образования экситона составляет лишь $E_b = 4,2$ мэВ, а соответствующий боровский радиус достаточно велик ($a_b = 150$ Å). На рис. 4.15 показаны две возможные ситуации для экситона в квантовой яме: а) боровский радиус a_b намного меньше ширины квантовой ямы (что соответствует объемному кристаллу) и б) боровский радиус намного превышает ширину ямы. Во втором случае расстояние между электроном и дыркой ограничено шириной ямы, поэтому экситон «сжимается», что, естественно, приводит к усилению кулоновского притяжения. Для двумерного атома

водорода можно легко показать, что величина E_B увеличивается примерно в четыре раза по сравнению с трехмерным случаем. Важным следствием увеличения энергии связи является то, что экситоны в квантовых ямах не разрушаются под воздействием очень сильных электрических полей, что используется, например, для создания оптоэлектронных модуляторов (раздел 10.8). Кроме того, так называемая сила осциллятора для экситонных переходов ограничена более узким диапазоном энергий, в результате чего их интенсивность возрастает.

Более реалистичные (по сравнению с теоретической моделью двумерного атома водорода) численные расчеты позволяют более точно определить энергию связи экситона в зависимости от ширины квантовой ямы. Такая зависимость для квантовой ямы в структуре $\text{Al}_x\text{Ga}_{1-x}\text{As}-\text{GaAs}$ приведена на рис. 4.16 [5]. В тех случаях, когда ширина ямы много больше боровского радиуса экситона, зависимость практически не отличается от той, которая характерна для объемных кристаллов. Однако затем, по мере уменьшения ширины ямы, экситоны в ней постепенно сжимаются, в результате чего возрастает кулоновское взаимодействие, и при ширине ямы около 3–4 нм энергия связи достигает своего максимума, примерно втрое превышающего значение для

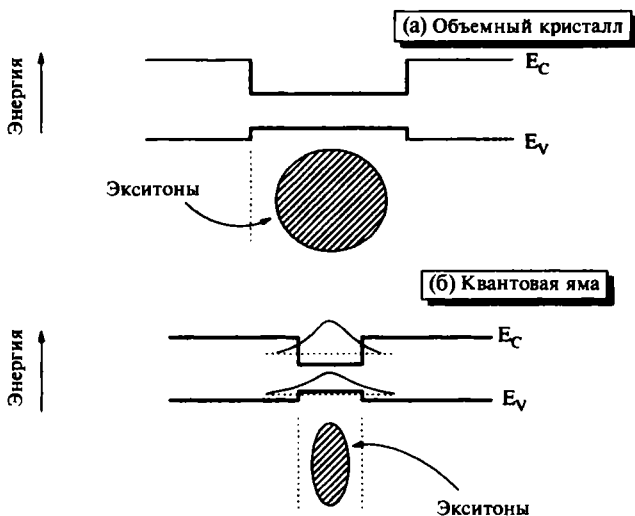


Рис. 4.15. (а) Орбиты экситона в объемном кристалле; (б) сферическая форма экситона вытягивается внутри квантовой ямы, ширина которой меньше радиуса экситона

объемного кристалла. Когда яма конечной глубины становится слишком узкой, энергия локализации экситона возрастает, в результате чего волновые функции электрона (и в меньшей степени дырки) начинают преодолевать энергетический барьер за счет квантового туннелирования. Такой процесс уменьшает эффективную локализацию экситона и его энергию связи, которая постепенно понижается до обычного уровня в объеме вещества барьера (в данном случае $\text{Al}_x\text{Ga}_{1-x}\text{As}$). Теоретически можно показать, что влияние локализации в ограниченной области на энергию связи экситона в одномерных (1D) и нульмерных (0D) системах (квантовые проволоки и точки соответственно) должны быть значительно большими, чем в рассматриваемых двумерных ямах. Однако технология создания таких структур пока не так развита, как для создания двумерных систем, поэтому эффекты изменения энергии связи экситона в одномерных и двумерных системах для решения технических задач пока не используются. Единственным исключением можно считать лазеры на квантовых точках, которые описываются в разделе 10.6.

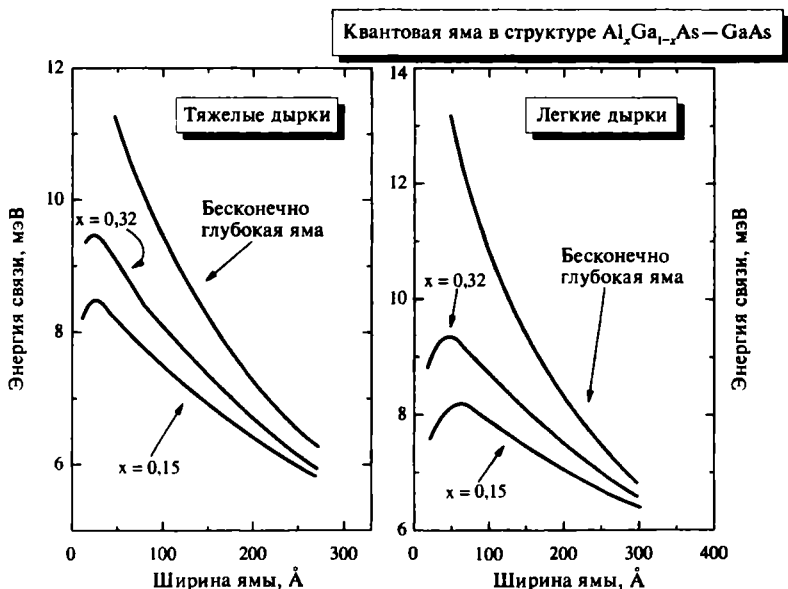


Рис. 4.16. Зависимость энергии связи для дырок в квантовой яме структуры $\text{GaAs} - \text{AlGaAs}$ от ширины ямы (по данным работы [5])

Литература

- [1] Davies, J.H. (1998) *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge).
- [2] Reilly, E.P.O. (1989) *Semiconductor Science and Technology*, 4, 121.
- [3] Фох, А.М. (1996) *Contemporary Physics*, 37, 111.
- [4] Chang, Y.C. & Shulman, J.N. (1985) *Phys. Rev. B*, 31, 2069.
- [5] Greene, R.L., Bajaj, K.K. & Phelps, D.E. (1984) *Phys. Rev. B*, 29, 1807.

Дополнительная литература

В качестве наиболее простого и легкого введения в теорию наноструктур можно рекомендовать книгу М. Джароса. Более сложными и фундаментальными являются две предлагаемые монографии (Вейсбуха и Винтера, Келли). Книга Дэвиса посвящена наиболее фундаментальным проблемам полупроводниковых наноструктур. Многие теоретические и фундаментальные вопросы интересно освещены в книге Митина и др., но, к сожалению, в ней совершенно не упоминаются некоторые важные вопросы (например, поведение наноструктур в магнитных полях или квантовый эффект Холла). Наиболее сложной и современной представляется книга Ферри и Гудника. Хотя она и посвящена специально теории переноса, в этой работе есть также много других интересных данных о фундаментальных проблемах процессов, происходящих в наноструктурах.

Davies J.H. (1998). *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge).

Ferry D.K., Goodnick S.M. (1997). *Transport in Nanostructures* (Cambridge University Press, Cambridge).

Jaros M. (1989). *Physics and Applications of Semiconductor Microstructures* (Clarendon Press, Oxford).

Kelly M.J. (1995). *Low-Dimensional Semiconductors* (Clarendon Press, Oxford).

Mitin V.V., Kochelap V.A., Stroschio M.A. (1999). *Quantum Heterostructures* (Cambridge University Press, Cambridge).

Weisbuch C., Vinter B. (1991). *Quantum Semiconductor Structures* (Academic Press, Boston).

Задачи и упражнения

1. Электронные уровни в квантовой яме. Рассмотрим квантовую яму с шириной a в зоне проводимости с настолько большим разрывом зон, что мы вправе использовать модель квантовой ямы бесконечной глубины. Объясните, каким образом эффективная масса электрона в полупроводнике влияет на положение энергетических уровней внутри ямы. Рассмотрите случаи InSb ($m^* = 0,014m_0$), GaAs ($m^* = 0,067m_0$) и GaN ($m^* = 0,17m_0$).
2. Ширина запрещенной зоны в системах с квантовыми точками. Рассмотрим квантовую точку в системе $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ при значениях $x \approx 0,3$. Рассчитайте для нее ширину запрещенной зоны при ширине ямы $a = 16$ нм. Является ли адекватной модель ямы бесконечной глубины при таком значении a ? Будет ли она адекватной при значительном уменьшении величины a , например, до 2 нм?
3. Длина волны Ферми. Длина волны Ферми λ_F определяется как длина волны электрона при энергии Ферми E_F . Рассчитайте значение λ_F для электронов в квантовой яме типа $\text{AlGaAs}/\text{GaAs}/\text{AlGaAs}$ при $n_{2D} = 10^{16} \text{ м}^{-2}$. Сравните полученный результат с известными значениями λ_F для распространенных металлов.
4. Функция плотности состояний для двумерных систем. Покажите, что функция плотности состояний n_{2D} (рис. 4.3) для двумерных систем в пересчете на единицу площади описывается соотношением

$$n_{2D} = \frac{m^*}{\pi\hbar^2} \sum_n \Theta(E - E_n),$$

где $\Theta(z)$ — ступенчатая функция (т. е. $\Theta(z) = 1$ для $z > 0$ и $\Theta(z) = 0$ для $z < 0$), а суммирование проводится по подзонам.

5. Концентрация электронов в квантовой яме. Используя полученное в предыдущей задаче выражение для плотности состояний и распределение Ферми — Дирака, покажите, что концентрация n_{2D} электронов в двумерных системах определяется выражением

$$n_{2D} = \frac{kTm^*}{\pi\hbar^2} \sum_n I_n \left(1 + e^{(E_F - E_n)/kT}\right),$$

где E_F — энергия Ферми; E_n — энергетические уровни в квантовой яме.

6. Предельные квантовые переходы в двумерных системах. Покажите, что полученное выражение для концентрации n_{2D} электронов в двумерных системах в предельном случае (когда заполнена только первая подзона с $n = 1$)

$$n_{2D} = \frac{m^*}{\pi\hbar^2} E_F,$$

где E_F — энергия Ферми. Покажите также, что волновой вектор Ферми k_F в этом случае имеет вид

$$k_F = (2\pi n_{2D})^{1/2}.$$

7. Плотность состояний в k -пространстве. Покажите, что плотность состояний в k -пространстве $g(k)$ имеет вид $g(k) \propto k^{n-1}$, где n — размерность системы. (Подсказка: начните с предположения, что значения компонент вектора \vec{k} могут быть получены с использованием метода периодических граничных условий, как это было сделано в разделе 2.3 для значения $n = 3$.)
8. Прямоугольная потенциальная яма с конечной высотой. Рассмотрим прямоугольную потенциальную яму с барьером конечной высоты V_0 и шириной a , для которой волновые функции определяются формулами (4.8) и (4.9). Покажите, что для возникновения в такой яме связанных состояний необходимо выполнение следующих условий:

$$\tan(k_w a/2) = \frac{m_w^* k_b}{m_b^* k_w},$$

$$\tan(k_w a/2) = -\frac{m_b^* k_w}{m_w^* k_b},$$

где нижние индексы w и b относятся к материалу ямы и барьера соответственно. (Подсказка: в качестве граничных условий наложите требования непрерывности как самих волновых функций, так и потока частиц $(1/m^*)d\psi/dz$.)

9. Реалистическая модель квантовой ямы в системе AlGaAs/GaAs/AlGaAs. Используя полученные в предыдущей задаче результаты, определите энергию основного состояния для квантовой ямы системы AlGaAs/GaAs/AlGaAs с шириной $a = 12$ нм, высотой потенциальной ямы $V_0 = 0,36$ эВ и эффективными массами $m_w^* = 0,067 m_0$ и $m_b^* = 0,095 m_0$. (Подсказка: используйте результаты предыдущей задачи, переписав значения для волновых векторов в виде функции энергии, определенной уравнением (4.9). Поскольку в этом случае нельзя получить аналитическое выражение для E , решите задачу графически.)
10. Структуры пониженной размерности. Рассмотрим полупроводниковую квантовую яму с шириной a , квантовую проволоку с квадратным сечением ($a \times a$) и квантовую точку с объемом ($a \times a \times a$). Рассчитайте энергию E_{el} основного электронного состояния для перечисленных структур в GaAs, где $a = 6$ нм. Качественно объясните, почему значения E_{el} возрастают с ростом локализации в структуре.
11. Энергия Ферми для электронных систем с низкой размерностью. Покажите, что для трехмерных, двумерных и одномерных (3D, 2D, 1D) электронных систем, где энергия E связана с волновым вектором k и эффективной массой m^* соотношением $E = \hbar^2 k^2 / 2m^*$, энергия Ферми E_F может быть выражена в виде

$$3D: E_F = \frac{\hbar^2}{2m^*} (3\pi^2 n)^{2/3},$$

$$2D: E_F = \frac{\hbar^2}{2m^*} 2\pi n,$$

$$1D: E_F = \frac{\hbar^2}{2m^*} \left(\frac{\pi}{2} n \right)^2,$$

где величина n соответствует концентрации электронов в каждой из этих систем.

ГЛАВА 5

ПОЛУПРОВОДНИКОВЫЕ КВАНТОВЫЕ НАНОСТРУКТУРЫ И СВЕРХРЕШЕТКИ

5.1. Введение

В предыдущей главе рассматривалось физическое поведение электронов в полупроводниках с пониженной размерностью. В этой главе мы опишем ряд наноструктур и приборов, включающих один или несколько гетеропереходов. Изучение и применение таких структур и приборов на их основе является сейчас одним из важных направлений развития микроэлектроники и оптоэлектроники. Особое значение в таких структурах приобретают гетеропереходы (т. е. переходы между двумя полупроводниками с разными запрещенными зонами) и МОП-структуры (структуры металл – окисел – полупроводник). Указанные виды наноструктур позволяют создавать наноразмерные потенциальные ямы для электронов вблизи перехода. МОП-структуры уже давно и очень успешно используются в микроэлектронике (особую важность имеют полевые МОП-транзисторы), поэтому неудивительно, что ученые продолжают использовать их для дальнейших исследований поведения электронов в наноразмерных потенциальных ямах. В частности, именно при изучении характеристик переноса электронов в канале полевого транзистора при одновременном воздействии электрических и магнитных полей К. фон Клитцингу [1] удалось в 1980 г. обнаружить квантовый эффект Холла (КЭХ). Особо можно отметить, что многие квантовые эффекты (КЭХ, эффект Аронова — Бома и колебания Шубникова — де Гааза), описанные

ниже (в главе 7), сильнее всего проявляются в гетеропереходах $A_{III}B_V$, поскольку эффективная масса электрона в структурах на основе GaAs намного меньше, чем в кремнии.

В первом разделе данной главы рассмотрено поведение электронов в двумерных ямах на границе раздела полупроводник — окисел в полевых МОП-транзисторах. Затем будут описаны модулированно-легированные гетеропереходы в структурах $A_{III}B_V$, используемые в высокочастотных транзисторах, а также напряженные гетеропереходы SiGe, имеющие интересные технические приложения. Следующий раздел посвящен модулированно-легированным прямоугольным квантовым ямам. Эта простая структура может использоваться в электронных приборах сама по себе, но чаще всего их применяют в структурах с множественными квантовыми ямами. При малой толщине барьеров между соседними ямами в таких структурах начинается процесс туннелирования электронов, что позволяет создавать сверхрешетки (SL), зонная диаграмма которых напоминает диаграммы некоторых кристаллов. При этом, однако, ширина разрешенных и запрещенных зон соответствует существенно меньшим диапазонам энергии, поскольку пространственная периодичность сверхрешеток (равная сумме толщин ямы и барьера) значительно больше, чем постоянные кристаллических решеток. Зонная структура сверхрешеток может регулироваться подбором толщины ямы и барьера, вследствие чего такие структуры иногда называют искусственными кристаллами, так как они обладают зонными энергетическими диаграммами кристаллов, хотя сами они и не существуют в природе. Идея использования сверхрешеток была предложена Эсаки и Тцу (Esaki, Tsu) в 1970 г. и успешно реализована позднее.

5.2. Структуры полевых МОП-транзисторов (MOSFET)

В современной технологии вообще (и в микроэлектронике в частности) огромное значение имеют полевые МОП-транзисторы (MOSFET). Такие транзисторы являются основными элементами сверхбольших интегральных схем. Объем рынка электронных приборов на основе полевых МОП-транзисторов составляет почти 90% от продукции полупроводниковой промышленности. Транзисторы этого типа образованы МОП-

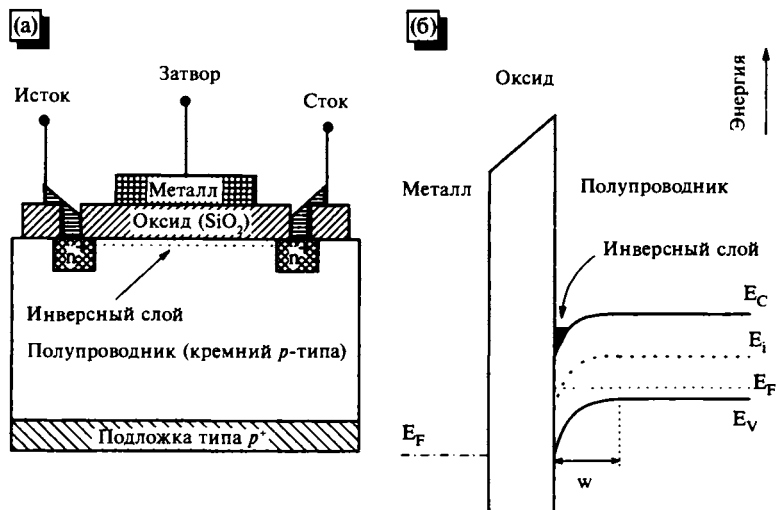


Рис. 5.1. (а) Структура МОП-транзистора; (б) зонная диаграмма МОП-структуры при достаточно сильном положительном смещении на вентиле

структурой и двумя $p-n^+$ переходами, в которых полупроводник n -типа сильно легирован (рис. 5.1, а) и действует в качестве истока и стока в полевом транзисторе. Затвором (управляющим электродом) такого транзистора выступает МОП-структура. В качестве полупроводника обычно используется кремний p -типа, на котором термическим окислением выращивается тонкий слой подзатворного окисла.

Принцип действия описанных приборов изложен ниже. На рис. 5.1, б показана энергетическая диаграмма МОП-структуры на кремний p -типа при достаточно сильном положительном смещении. Когда к затвору прикладывается положительный потенциал, электроны из n^+ области и электроны из объема p -кремния накапливаются в области границы $\text{Si}-\text{SiO}_2$. Такие электроны располагаются в потенциальной наноразмерной яме (почти треугольной формы) и образуют так называемый инверсный слой или канал. Такая форма ямы объясняется наличием пространственного заряда ионизированных акцепторов в p -кремнии, соответствующие которым дырки «выталкиваются» электрическим полем положительно заряженного затвора. В такой системе электроны в канале способны создавать ток

при приложении потенциала между истоком и стоком, причем величина тока может модулироваться потенциалом затвора, так как количество электронов в инверсном слое зависит от напряженности электрического поля, приложенного к слою изолятора. Именно вследствие того, что поведение таких приборов управляется электрическим полем, они получили название транзисторов с полевым эффектом или просто полевых транзисторов.

Само существование очень развитой технологии полевых МОП-транзисторов (и создание на этой основе мощной микроэлектронной промышленности) связано с исключительными физическими характеристиками поверхностей раздела Si — SiO₂ и возможностью уменьшения размеров элементов структур полевых МОП-транзисторов в новейших интегральных схемах до величин, меньших 100 нм. Еще в 1957 г. Шриффер указал на возможность квантования состояний электронов в квантовых ямах, однако прошло более десяти лет, прежде чем этот эффект удалось экспериментально обнаружить именно на границе раздела Si — SiO₂. Для регистрации эффекта понадобилось добиться сочетания, по крайней мере, двух важных обстоятельств: 1) следовало получить аморфный изолятор SiO₂ без высокой концентрации примесных атомов (ионы Na⁺) и без захваченных зарядов; 2) необходимо было получить поверхность раздела Si — SiO₂ с «атомарной гладкостью», поскольку шероховатость поверхности вблизи инверсионных каналов значительно снижала подвижность электронов в инверсном слое.

Для исследования поведения электронов в потенциальных ямах очень важным оказалось осознание, что электроны инверсного слоя следует рассматривать в качестве двумерной электронной системы, «погруженной» в потенциальные ямы треугольной формы (см. раздел 4.4.2), расположенной внутри полупроводника, очень близко к оксидному слою. В МОП-структурах движение электронов в перпендикулярном к поверхности направлении ограничено, однако они могут свободно перемещаться вдоль поверхности раздела. В соответствии с полученным в разделе 4.2 результатом квантованные значения энергии для ограниченного движения имеют вид

$$E = E_n + \frac{\hbar^2}{2m_x^*} k_x^2 + \frac{\hbar^2}{2m_y^*} k_y^2, \quad (5.1)$$

где величины E_n означают разрешенные уровни энергии в треугольной потенциальной яме в соответствии с уравнением (4.13)

$$E_n \approx \left[\frac{3}{2} \pi \left(n - \frac{1}{4} \right) \right]^{2/3} \left(\frac{e^2 F^2 \hbar^2}{2m_z^*} \right)^{1/3}, \quad n = 1, 2, \dots \quad (5.2)$$

Совершенно очевидно, что уравнение (5.1) представляет параболы рис. 4.1, в, в обратном пространстве, где значения E_n соответствуют минимумам кривых. Аналогично функция плотности состояний, соответствующая двумерному случаю, принимает вид

$$g(E) = g_v \frac{m_x^*}{\pi \hbar^2}, \quad (5.3)$$

где дополнительный коэффициент g_v вводится для учета вырождения долины зоны проводимости. Вырождение обусловлено тем, что эквивалентные поверхности зоны проводимости кремния образуются шестью эллипсоидами в направлении $\langle 001 \rangle$ пространства импульсов (рис. 3.3). Продольные оси эллипсоидов соответствуют продольным эффективным массам $m_L^* = 0,91 m_0$, а две одинаковые короткие оси — поперечной эффективной массе $m_T^* = 0,19 m_0$, использованной в уравнении (5.13), вследствие чего мы получаем два минимума зоны проводимости (соответствующие тяжелой эффективной массе) и четыре минимума для легкой массы. Решая уравнение Шрёдингера и пренебрегая взаимодействием электронов в различных минимумах зоны проводимости, можно получить два различных набора значений энергии для подзон, которые называют лестницами подзон для плоскости $\langle 001 \rangle$ кремния. Рассмотрим сначала поведение электронов в минимумах (долинах), расположенных перпендикулярно к границам раздела. В уравнение (5.2) войдет эффективная масса m_z^* . Кроме этого, в уравнении (5.1) $m_x^* = m_y^* = m_T^*$ и значение коэффициента $g_v = 2$, что приводит к подзонам с низшей энергией (при более высоком значении эффективной массы). Очевидно, что лестница второй подзоны должна возникать в тех случаях, когда рассматриваются параллельно расположенные долины. В этом случае в выражение (5.2) для E_n следует подставлять эффективную массу m_T^* и значение коэффициента $g_v = 4$.

На рис. 5.2 приведены результаты расчета энергий для кремниевого полевого МОП-транзистора с изолирующим оксидным слоем, выращенным на поверхности кремния $\langle 001 \rangle$ [2]. Кривые получены численным самосогласованным решением уравнений Шрёдингера и Пуассона для параметров $m_L^* = 0,91 m_0$ и $m_T^* = 0,19 m_0$. Наборы энергий E_0, E_1, E_2 и E_0', E_1', E_2' получены для эффективных масс $m_L^* = 0,91 m_0$ и $m_T^* = 0,19 m_0$ соответственно. На рисунке представлен также профиль потенциала дна зоны проводимости $V(z)$ (сплошная кривая) и квадраты значений волновых функций для наборов уровней двух подзон. Как будет показано в главе 7, существующие эксперименталь-

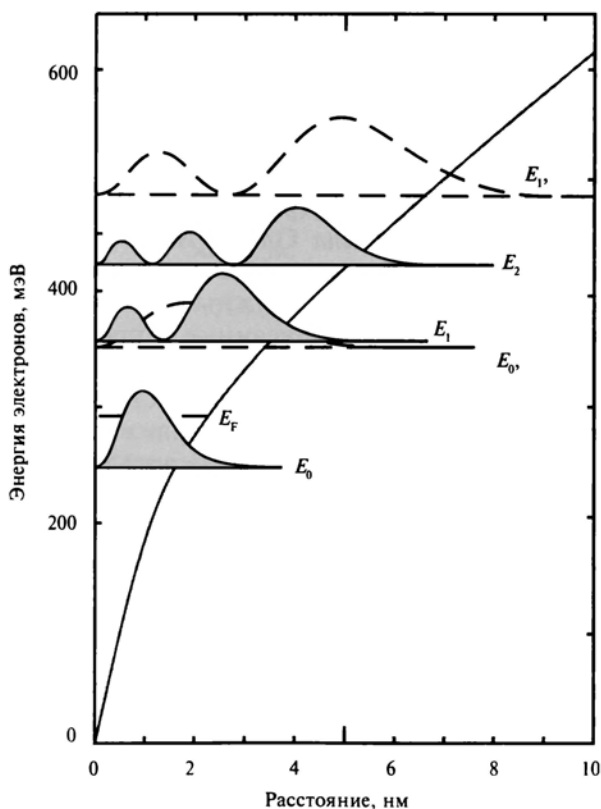


Рис. 5.2. Последовательности подзон, форма зоны проводимости $V(z)$ и его значения квадратов волновых функций для потенциальной ямы в МОП-структуре для Si $\langle 001 \rangle$ (по данным работы [2])

ные данные по многим явлениям (квантовый эффект Холла, эффект Шубникова — де Гааза) свидетельствуют, что двумерная система электронов в инверсном слое полевых МОП-транзисторов квантуется.

5.3. Гетеропереходы

5.3.1. Гетеропереходы с модулированным легированием

Гетеропереходы, представляющие собой поверхности раздела между двумя полупроводниками с различными запрещенными зонами, являются чрезвычайно важными и разнообразными блоками электронных приборов, особенно создаваемых на основе полупроводниковых соединений $A_{III}B_V$. По-видимому, наиболее изученной в настоящее время гетероструктурой является та, которая образована $Al_xGa_{1-x}As$ (n -типа) и почти беспримесным или слаболегированным $GaAs$ p -типа. Так же, как и в случае МОП-структур (раздел 5.2), внутри $GaAs$ очень близко к поверхности раздела $GaAs-AlGaAs$ образуется инверсный слой электронов, вследствие чего физические явления в таких электронных системах должны быть очень похожи на явления в МОП-структурах, описанные в предыдущем разделе. Приборы, создаваемые на основе $AlGaAs$, обычно используются при значительно более высоких частотах, чем приборы на основе кремния, в силу большей подвижности электронов в $GaAs$. Поскольку поверхности раздела оксидов и изоляторов, осаждаемых на поверхности $GaAs$, являются недостаточно качественными, в большинстве важных технических применений используются структуры Шоттки типа металл — $AlGaAs-GaAs$ (рис. 5.3, *а*). В этом разделе мы будем рассматривать в основном, физические свойства и зонные диаграммы для гетероперехода $Al_xGa_{1-x}As-GaAs$, показанного на рис. 5.3, *б*.

Прежде всего, следует обсудить качественную картину возникновения электронной наноразмерной ямы на поверхности раздела $AlGaAs-GaAs$. Предположим, как показано на рис. 5.4, что в рассматриваемом гетеропереходе $AlGaAs-GaAs$ слева располагается полупроводник арсенид галлия, легированный алюминием, а справа — почти беспримесный материал $GaAs$. Такая структура называется гетеропереходом с модулированным

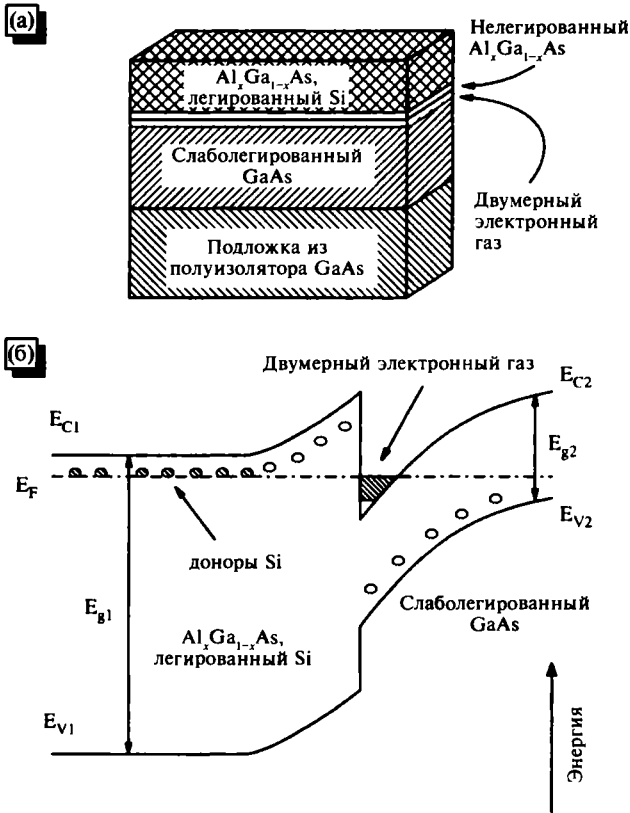


Рис. 5.3. (а) Модулированно-легированный гетеропереход $\text{AlGaAs}-\text{GaAs}$; (б) соответствующая зонная диаграмма

легированием, а метод ее создания модулированным легированием. Рассмотрим сначала гипотетическую ситуацию рис. 5.4, а (левая часть) до момента контакта двух описанных полупроводников. На рисунке для простоты схематически показаны только дно зон проводимости и уровень Ферми, который для n -типа AlGaAs располагается очень близко к зоне проводимости, а для слабелегированного GaAs p -типа расположен почти в середине запрещенной зоны (см. раздел 3.4). Очевидно, что зоны являются плоскими, поскольку материал в целом является электрически нейтральным, а легирование носит однородный характер. Высота барьера между материалами (обозначенная на рисунке через ΔE_c) может быть приблизительно оценена

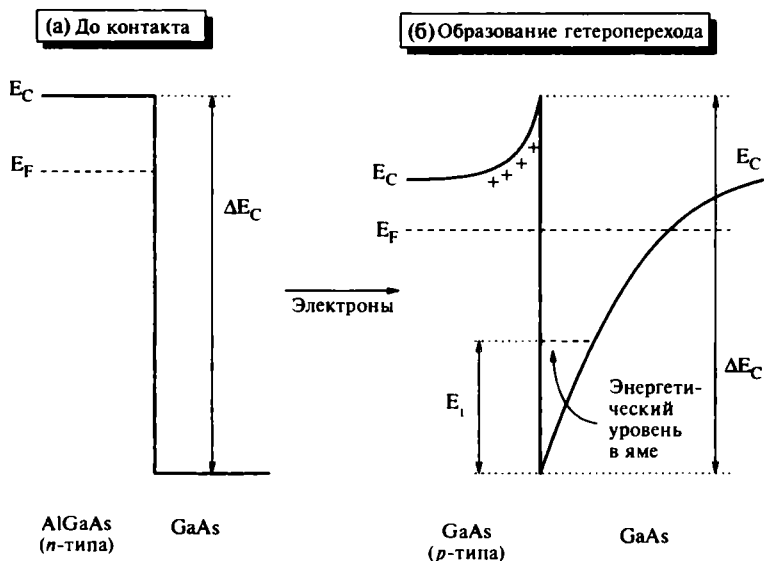


Рис. 5.4. Формирование зонной диаграммы в модулированно-легированном гетеропереходе AlGaAs—GaAs при контакте этих полупроводников

по так называемому правилу Андерсона, в соответствии с которым при соединении двух полупроводниковых материалов их уровни в вакууме должны выравняться. Обозначив через χ_A и χ_B электронное сродство материалов AlGaAs и GaAs соответственно, легко получить $\Delta E_c \equiv \chi_A - \chi_B$, поскольку электронное сродство в полупроводниках определяется как энергия, необходимая для переноса электрона со дна зоны проводимости во внешнее по отношению к твердому телу пространство, т. е. $\chi = E_{\text{vac}} - E_c$. Исходя из правила Андерсона, можно вычислить, что для легированного $\text{Al}_x\text{Ga}_{1-x}\text{As}$ (при значении x около 0,3) значение ΔE_c составляет 0,35 эВ.

Ситуация изменяется, когда описываемые полупроводниковые материалы AlGaAs и GaAs вступают в контакт. При этом часть электронов от доноров материала n -типа пересекают поверхность раздела и попадают в нелегированный GaAs, в результате чего (как и в случае p - n переходах) возникает внутреннее электрическое поле, которое направлено от нескомпенсированных доноров в AlGaAs в сторону GaAs с дополнительным отрицательным зарядом. Такое поле приводит к искривлению (из-

гибу) зоны, как показано рис. 5.3, б. В равновесном состоянии оба уровня Ферми должны совпадать, в результате чего изгиб зоны очень напоминает изгиб для упомянутых p - n переходов, а разница заключается лишь в наличии дополнительного барьера ΔE_c . Отметим также, что по мере удаления от поверхности раздела дно зоны проводимости вновь становится плоским, причем ее расстояние до уровня Ферми E_F совпадает с величиной, показанной на рис. 5.4, а. Поэтому можно довольно легко нарисовать зонную диаграмму, приведенную на рис. 5.4, б, на которой квантовая яма для электронов формируется слева потенциальной стенкой высотой ΔE_c , а справа — кривой потенциальной энергии для электронов в GaAs.

Форма квантовой ямы для электронов, образующейся на поверхности раздела AlGaAs—GaAs, близка к треугольной и напоминает яму для случая МОП-структур. Действительно, если мы обозначим в качестве оси z перпендикулярное к поверхности раздела направление, то инверсный слой будет представлять собой двумерную электронную систему, в которой электроны могут свободно перемещаться в плоскости (x, y) , однако их движение вдоль оси z будет квантоваться в соответствии с ограничениями, создаваемыми условиями квантовой ямы. Наиболее важной физической особенностью описываемых гетеропереходов выступает то, что носители заряда локализуются в некоторой области (в основном в материале GaAs), которая пространственно отделена от полупроводника AlGaAs, в котором и возникают свободные электроны. Электроны в такой яме должны обладать очень высокой подвижностью в плоскости (x, y) , так как они двигаются внутри GaAs, в котором практически отсутствуют легирующие примеси (известно, что именно рассеяние на таких примесях ограничивает подвижность носителей заряда, особенно при низких температурах). Из этого следует, что подвижность электронов в таких переходах должна быть значительно выше, чем в МОП-структурах, которые рассматривались в предыдущем разделе.

Хотя многие приведенные выше доводы выглядят качественными, они позволяют достаточно точно описать многие свойства гетеропереходов AlGaAs—GaAs. При более строгом описании мы, конечно, должны были бы получить самосогласованные решения уравнений Пуассона (для потенциала) и Шрёдингера (для волновых функций электрона). Однако та-

кие достаточно сложные задачи иногда удается решить, используя некоторые приближения (например, считая потенциал ямы строго треугольным по форме). Однако нельзя решать задачу в предположении бесконечной глубины ямы, поскольку в нашем случае должно выполняться условие $\Delta E_c \approx 0,3$ эВ. Точные расчеты позволяют также найти среднюю ширину ямы (40—80 Å), концентрацию электронов на единицу площади ($n_s \approx 10^{12}$ см⁻²) и энергию электронов на первом уровне ($\epsilon_1 \approx 0,04$ эВ).

Представляется очевидным, что, создавая полевой транзистор на основе гетероперехода AlGaAs—GaAs, необходимо изготовить сток и исток так, как это было показано на рис. 5.1 для полевого МОП-транзистора. Прикладывая к затвору разность потенциалов, мы можем регулировать число носителей в канале (а следовательно, и его проводимость) методом, описанным в разделе 5.2 для полевых МОП-транзисторов. Для того чтобы подчеркнуть значение модулированного легирования, такие структуры называют модулированно-легированными полевыми транзисторами (MODFET) или полевыми транзисторами с высокой подвижностью электронов (HEMT). Мы подробно обсудим работу таких транзисторов в разделе 9.2, так как благодаря высокой подвижности электронов в канале они часто используются в высокочастотных приборах.

5.3.2. Напряженные гетероструктуры на основе SiGe

Длительное время гетероструктуры типа SiGe не привлекали особого внимания исследователей, что объясняется значительной разницей в значениях постоянных кристаллических решеток германия и кремния, достигающей до 4% (см. рис. 4.9). Такая разница подразумевала, что выращенные на подложке слои всегда будут напряженными, вследствие чего их толщина будет ограничена определенным критическим значением (см. раздел 4.7) или же образующиеся двухслойные структуры будут непрочными. Более того, поскольку разница в энергии запрещенных зон этих материалов также достаточно мала (энергия E_g составляет для германия и кремния 1,12 и 0,66 эВ соответственно), высота возникающего на контактной поверхности барьера также должна быть малой. Однако, несмотря на эти кажущиеся недостатки и сложности, гетероструктуры типа SiGe

нашли ряд интересных технических применений, связанных, прежде всего, с высокочастотными транзисторами (глава 9) и инфракрасными фотодетекторами (глава 10). Поскольку, как уже отмечалось, такие структуры всегда являются механически напряженными, в них снимается вырождение зон легких и тяжелых дырок (раздел 4.8), а их зонная структура имеет в целом вид, показанный на рис. 4.12.

На рис. 5.5 показаны два типичных примера гетероструктур SiGe. Подложкой структуры рис. 5.5, а является Si с ориентацией поверхности $\langle 001 \rangle$ ($E_g = 1,17$ эВ), на которой образован активный напряженный слой материала $\text{Si}_{0,7}\text{Ge}_{0,3}$ с $E_g = 0,78$ эВ. В такой структуре разрыв зон проводимости очень мал, в отличие от разрыва валентных зон. В такой системе формируется двумерный дырочный газ в сплаве SiGe, в котором подвижность дырок достигает $2 \text{ м}^2\text{В}^{-1}\text{с}^{-1}$, что соответствует примерно половине значения подвижности электронов в типичных полевых МОП-транзисторах (см. раздел 5.2). На рис. 5.5, б представлена как бы обратная ситуация, в которой напряженный слой на подложке образован кремнием. В этой ситуации разрыв между

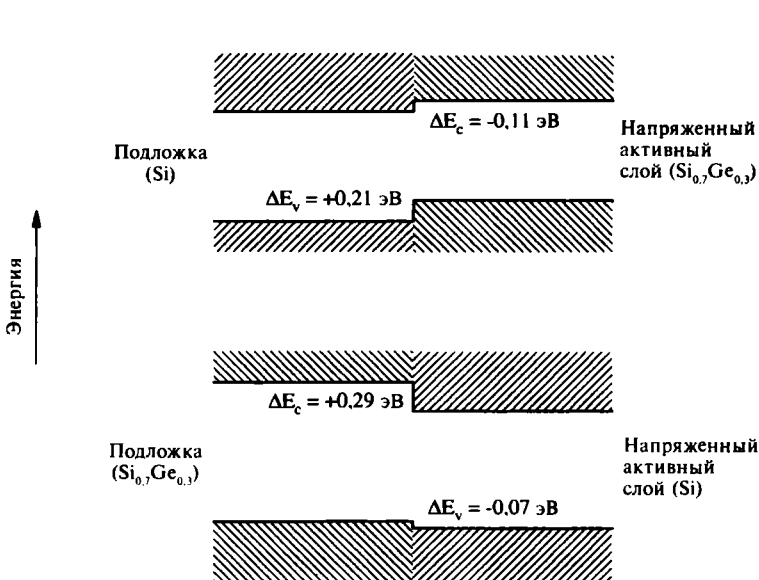


Рис. 5.5. Схематические зонные диаграммы для двух противоположных типов напряженных гетероструктур на основе SiGe

зонами проводимости является достаточно большим, и двумерный газ образуют электроны, которые могут свободно двигаться в плоскости контакта. Такому движению соответствует эффективная масса электрона в кремнии, равная малой поперечной массе ($m_T^* \approx 0,19 m_0$), что обеспечивает высокую подвижность (порядка $20 \text{ м}^2\text{В}^{-1}\text{с}^{-1}$), в несколько раз превышающую подвижность электронов в полевых МОП-транзисторах.

Гетероструктуры на основе SiGe нашли также очень важное применение для создания биполярных кремниевых транзисторов, которые будут подробнее описаны в разделе 9.3. Одним из направлений повышения эффективности работы биполярных транзисторов является использование материалов с узкой запрещенной зоной в области базы, что позволяет значительно повысить эффективность работы кремниевой области эмиттера. Преимущества в данном случае возникают вследствие низких значений запрещенной зоны в сплавах SiGe (по сравнению с чистым кремнием). Важно отметить, что механические напряжения в гетероструктурах также способствуют понижению ширины запрещенной зоны. Кроме этого, большие разрывы зон позволяют использовать высоколегированные материалы с низким сопротивлением, что также дает возможность эффективно использовать кремниевые транзисторы при значительно больших частотах.

5.4. Квантовые ямы

5.4.1. Модулированно-легированные квантовые ямы

Очень часто, особенно в структурах с множественными квантовыми ямами, желательно было бы иметь дело с примерно симметричными и прямоугольными ямами, а не с треугольными, типа тех, которые рассматривались в предыдущем разделе для простых модулированно-легированных гетеропереходов. Пример реализации такой структуры представлен на рис. 5.6, *a*, где показана симметричная квантовая яма, образованная двумя примыкающими друг к другу (но противоположно направленными) гетеропереходами на основе рассмотренной выше системы AlGaAs—GaAs. Полупроводники с широкой запрещенной зоной ($\text{Al}_x\text{Ga}_{1-x}\text{As}$) в этой структуре располагаются на краях, а GaAs — в центральной части. Предположим далее, что расстояние между двумя поверхностями раздела становится до-

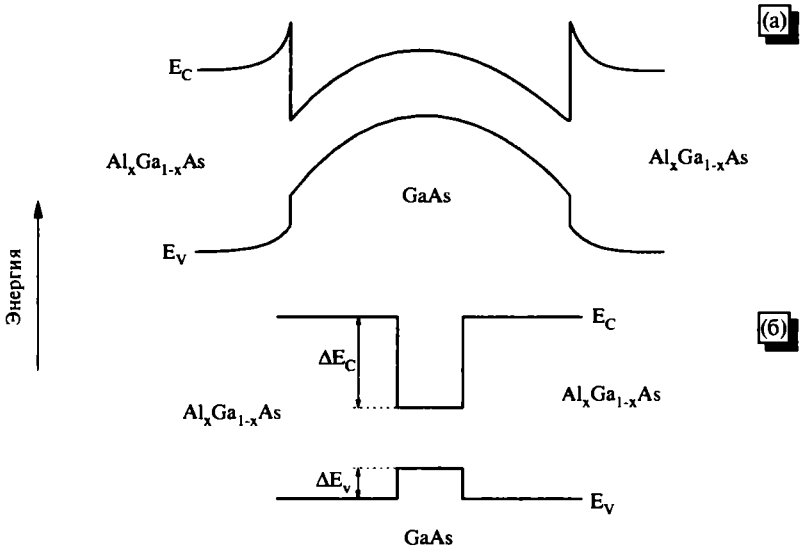


Рис. 5.6. Схематическая зонная диаграмма для квантовой ямы в модулированно-легированном материале AlGaAs—GaAs—AlGaAs. (а) широкая яма; (б) узкая яма

статочны малым. В такой ситуации образуется практически прямоугольная квантовая яма (рис. 5.6, б) для дырок и для электронов, с одинаковой высотой барьера на обеих сторонах, равной показанной на рис. 5.6, а.

Особо следует отметить, что (как и в разделе 5.3.1) в такой яме отсутствуют ионизированные доноры, которые располагаются внутри барьеров AlGaAs, вследствие чего электроны внутри ямы (возникающие в прилегающем донорном материале AlGaAs) могут двигаться в яме GaAs или канала с очень высокой подвижностью. Как и в случае модулированно-легированных гетеропереходов, это обстоятельство может быть использовано для изготовления высокочастотных полевых МОП-транзисторов при создании соответствующих контактов стока и истока.

Структуры из квантовых ям с высокой или низкой подвижностью электронов могут быть созданы введением в материалы регулируемых количеств примесей. Например, на основе двойных квантовых ям (с высокой и низкой подвижностью электронов) могут быть созданы транзисторы с модуляцией скорости, в которых переключение из одного состояния в другое управ-

ляется электрическим полем, приложенным поперечно к слоям полупроводниковых материалов, в результате чего происходит перераспределение потока носителей заряда, а значит, и тока между ямами. Такие транзисторы с модуляцией скорости могут работать при очень высоких частотах.

Очень трудно найти точные решения системы уравнений для волновых функций и энергетических уровней электронов и дырок в описываемых системах квантовых ям, поскольку потенциал дна ямы не является плоским и сильно зависит от неоднородности границы раздела. Обычно такие задачи решают численно, однако для описания поведения систем в общих чертах можно воспользоваться теорией возмущений, рассматривая в качестве возмущения изменения потенциала дна зоны проводимости. Потенциал возмущения должен иметь симметричный вид, т. е. представлять собой четную функцию (по переменной z), а получаемые решения должны соответствовать комбинации состояний с одинаковой четностью (эта проблема будет обсуждена подробнее в главе 8 при рассмотрении оптических переходов в квантовых ямах). Из этого условия, например, сразу следует, что если мы предполагаем существование в яме лишь трех уровней, то внутренние связи и переходы в такой системе могут происходить лишь между первым и третьим уровнями, а поскольку заселенность уровней зависит от концентрации электронов в яме, то чаще всего заполненным оказывается лишь первый уровень.

5.4.2. Множественные квантовые ямы (MQW)

Сигналы, поступающие от отдельной квантовой ямы, обычно настолько слабы, что их трудно использовать для создания каких-либо твердотельных приборов, поэтому обычно в реальных оптоэлектронных приборах (например, в фотодетекторах) применяются структуры, формируемые из целых наборов или массивов отдельных квантовых ям, которые, естественно, получили общее название множественных квантовых ям (MQW). Обычно говорят о двух разных типах таких структур. Если ямы, соответствующие электронам и дыркам, имеют одинаковую пространственную локализацию (т. е. располагаются в одной и той же области, как показано на рис. 5.7, *а*), то соответствующие структуры называют MQW типа I. В том случае, когда эти ямы располагаются поочередно (как показано на рис. 5.7, *б*), структуры MQW относят к типу II.

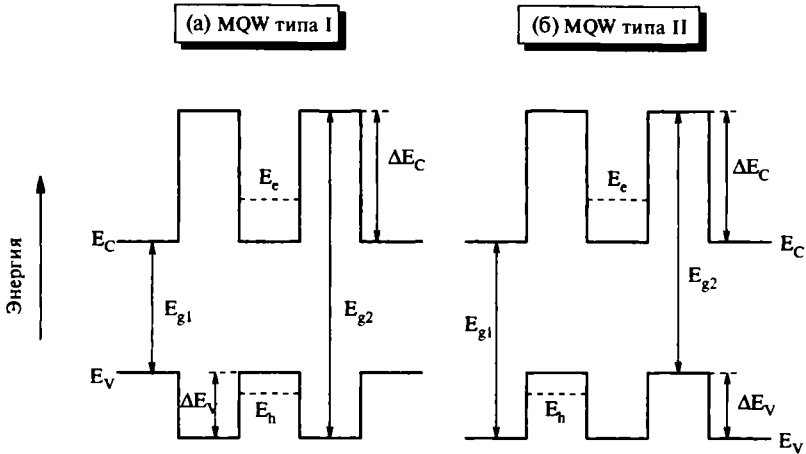


Рис. 5.7. Множественные квантовые ямы (MQW) типа I (а) и типа II (б)

В большинстве случаев предполагается, что в MQW-структурах соседние квантовые ямы не взаимодействуют, поскольку разделяющие их барьеры являются достаточно протяженными (обычно порядка и больше 10 нм). Однако иногда, когда стенки барьеров становятся достаточно тонкими, квантовые ямы оказываются связанными друг с другом процессами туннельных переходов электронов из одной в другую. В следующем разделе будет показано, что при туннельном эффекте дискретные уровни энергии квантовых ям превращаются в энергетические зоны. Такие MQW-структуры (их принято называть сверхрешетками) представляют очень большой практический интерес, так как их энергетические спектры демонстрируют много очень интересных особенностей.

Чаще всего в оптоэлектронике изготавливают и применяют MQW структуры, содержащие примерно 50 отдельных квантовых ям. Периодичность такой структуры (ширина ямы + толщина барьера) составляет около 20 нм, в результате чего полный размер структуры равен примерно 1 мкм. Это значение обычно и считают пределом толщины, поскольку в противном случае любое излучение будет в значительной степени поглощаться и не сможет достигать внутренних квантовых ям. Кроме того, следует учитывать и то обстоятельство, что спектр полного оптического сигнала от системы с такими ямами, конечно, не совпадает точно со спектром отдельной ямы, так как ширина ям не является

абсолютно одинаковой (например, читатель может попробовать вычислить в процентах неточность в ширине ям, обусловленную изменением этой ширины на моноатомный слой). Кстати, ширина энергетических уровней MQW-структур может быть использована именно в качестве показателя однородности их толщины.

На рис. 5.8 приведена зонная структура типичной MQW-структуры, используемой в инфракрасных фотодетекторах или электрооптических переключающих модуляторах, при приложении электрического поля. ИК-фотодетекторы обычно работают на длинах волн в диапазоне 10 мкм, поэтому, если бы мы желали зафиксировать настолько длинноволновое излучение, обычным объемным полупроводником, нам понадобился бы материал с очень малой шириной запрещенной зоны ($\sim 0,1$ эВ), вследствие чего гораздо проще зарегистрировать такие сигналы, возбуждая переходы между подзонами в квантовых ямах. Используя гетеропереходы для создания таких детекторов, мы можем не только менять высоту барьера E_c (варьируя, например, значение x для структуры $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$), но и даже положение энергетических уровней (меняя ширину ямы). Как показано на рис. 5.8, уровни могут быть рассчитаны таким образом, чтобы второй уровень располагался лишь чуть ниже нижней границы зоны проводимости материала с широкой запрещенной

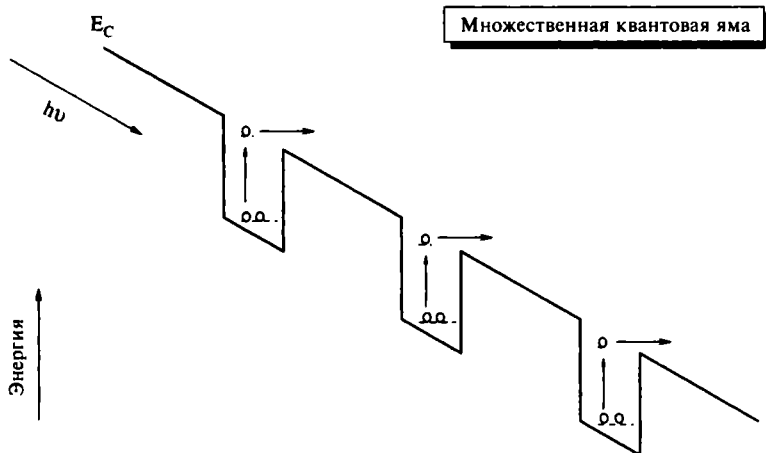


Рис. 5.8. Зонная структура типичной MQW-структуры, используемой в инфракрасных фотодетекторах, при наличии внешнего электрического поля

зоной (в данном случае $\text{Al}_x\text{Ga}_{1-x}\text{As}$). Кроме того, легирование полупроводника может осуществляться таким образом, чтобы уровень $n = 1$ был заполнен, а уровень $n = 2$ — пустым. С учетом всех этих соображений можно создать структуру, где после попадания в нее фотона с энергией $\hbar\omega$ (равной $E_1 - E_2$) высвобождаются электроны, которые начинают двигаться вдоль оси z под воздействием слабого приложенного электрического поля.

5.5. Сверхрешетки

5.5.1. Концепция сверхрешеток

Сама концепция или идея сверхрешеток, предложенная Лео Эсаки и его сотрудниками в конце 60-х годов, была позднее реализована с использованием метода молекулярно-лучевой эпитаксии (МЛЭ). В настоящее время сверхрешетки можно считать одним из важнейших искусственных материалов, созданных человечеством. Они образуются периодическими наборами MQW, в которых расстояние между отдельными ямами достаточно мало. По мере уменьшения ширины барьеров возрастает степень перекрытия волновых функций состояний в ямах из-за туннелирования, вследствие чего дискретные уровни энергии в ямах уширяются и на их месте возникают энергетические зоны. Это процесс очень похож на возникновение энергетических зон из состояний отдельных атомов, помещенных в кристаллическую решетку. Наиболее важным моментом в создании сверхрешеток выступает возникновение новой периодичности в материала с периодом d , равным сумме ширины ямы a и ширины барьера b (обычно значения a и b лежат в диапазоне от 4 до 2 нм соответственно). Строгий контроль таких малых толщин может достигаться только при использовании методов осаждения, как молекулярно-лучевая эпитаксия или газофазное химическое осаждение из паров металлоорганических соединений.

Рассмотрение принципиальных основ возникновения энергетических зон в сверхрешетках можно начать с задачи перекрытия волновых функций электронов в очень простой, двухямной системе. Проблема является классической для квантовой механики, поскольку формально она во многом напоминает задачу об уровнях в двухатомной молекуле. На рис. 5.9, *a* показаны две квантовые ямы, соответствующие так называемой двойной системе связанных квантовых ям, для ко-

торой квантово-механическое решение может быть получено на основе теории возмущений. В соответствии с теорией, при разделении квантовой ямы с заданным уровнем (обозначаемым через E_1) на две разных, с уровнями энергии (как показано на рис. 5.9, б)

$$E = E_1 \pm |V_{12}| \quad (5.4)$$

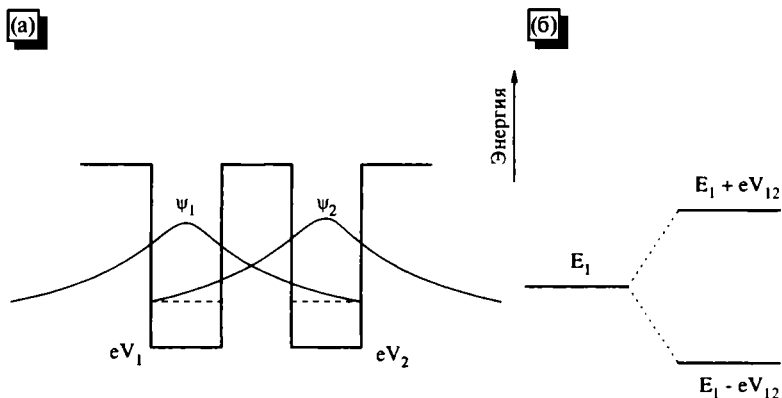
интеграл перекрытия имеет вид

$$V_{12} = \int_{-\infty}^{+\infty} \psi_1^* V(z) \psi_2 dz, \quad (5.5)$$

причем два образующихся уровня разделены интервалом $2|V_{12}|$, где величина V_{12} в уравнении (5.5) является показателем того, насколько одна яма может воздействовать на уровни в соседней, вследствие чего ее называют интегралом перекрытия.

5.5.2. Модель сверхрешетки Кронига — Пенни. Расщепление зон

Для описания структуры электронных зон в сверхрешетках можно воспользоваться аналогиями из курса физики твердо-



Система из двух связанных квантовых ям

Рис. 5.9. Система из двух связанных квантовых ям (а); диаграмма соответствующих энергетических уровней (б)

го тела, т. е. уподобить сверхрешетку кристаллической решетке и считать потенциалы квантовых ям аналогами потенциалов атомов в кристалле. В соответствии с этим мы можем изучать зонную структуру сверхрешеток в рамках двух используемых в физике твердого тела приближений (модели Кронига — Пенни и приближения сильной связи соответственно), описанным в главе 2 для обычных твердых тел.

Интересно отметить, что еще в 1931 г. Крониг и Пенни [3] предложили в теории твердого тела модель, в которой электроны двигались в периодическом потенциале из строго прямоугольных ям (рис. 5.10), поразительно напоминающем представленный выше потенциал для сверхрешеток. Предложенный ими периодический одномерный потенциал описывался следующими параметрами: ширина ямы a , ширина барьера b , высота барьера V_0 . Периодичность такого потенциала составляла $d = a + b$. Даже с учетом явной упрощенности потенциала рис. 5.10 по сравнению с потенциалами реальных систем он позволяет получить много очень интересных результатов, относящихся к структуре уровней, запрещенных зон, размеров зон и др. Для внутренних областей ямы ($0 < z < a$), где $V = 0$, из уравнения (2.11) можно получить решения для волновых функций в виде

$$\psi(z) = Ae^{ik_0z} + Be^{-ik_0z}, \quad (5.6)$$

где

$$k_0^2 = \frac{2mE}{\hbar^2}. \quad (5.7)$$

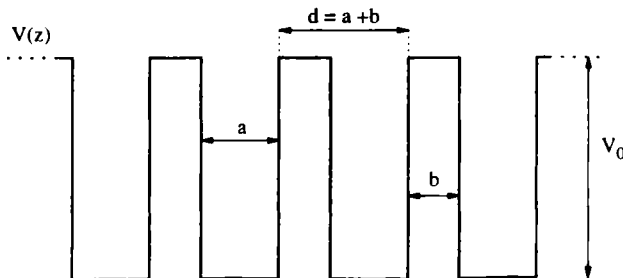


Рис. 5.10. Периодический потенциал в сверхрешетке

Вследствие эффектов туннелирования волновые функции могут быть продолжены и внутрь энергетического барьера с высотой V_0 и шириной b , в результате чего для области $(-b < z < 0)$ справедливо соотношение

$$\psi(z) = Ce^{\alpha z} + De^{-\alpha z}, \quad (5.8)$$

в котором волновой вектор и энергия связаны уравнением

$$V_0 - E = \frac{\hbar^2 q^2}{2m}. \quad (5.9)$$

Из условия непрерывности волновых функций и их производных при $z = 0$ и $z = a$ (переменная по оси z отсчитывается от левого края ямы) после некоторых преобразований можно получить соотношения

$$A + B = C + D, \quad (5.10)$$

$$ik_0(A - B) = q(C - D). \quad (5.11)$$

В соответствии с теоремой Блоха, приводящей к уравнению (2.40), волновые функции в двух разных координатах могут быть связаны как

$$\psi(a) = \psi(-b)e^{ik(a+b)},$$

где k — волновой вектор блоховской волновой функции. Используя это уравнение для волновых функций, соответствующих яме и барьеру, можно переписать уравнения (5.10) и (5.11) в виде

$$Ae^{ik_0 a} + Be^{-ik_0 a} = (Ce^{-qb} + De^{qb})e^{ik_0(a+b)}, \quad (5.12)$$

$$ik_0(Ae^{ik_0 a} - Be^{-ik_0 a}) = q(Ce^{-qb} - De^{qb})e^{ik_0(a+b)}. \quad (5.13)$$

Четыре уравнения (5.10) — (5.13) для четырех амплитуд (A , B , C и D) могут иметь общее решение лишь при условии равенства

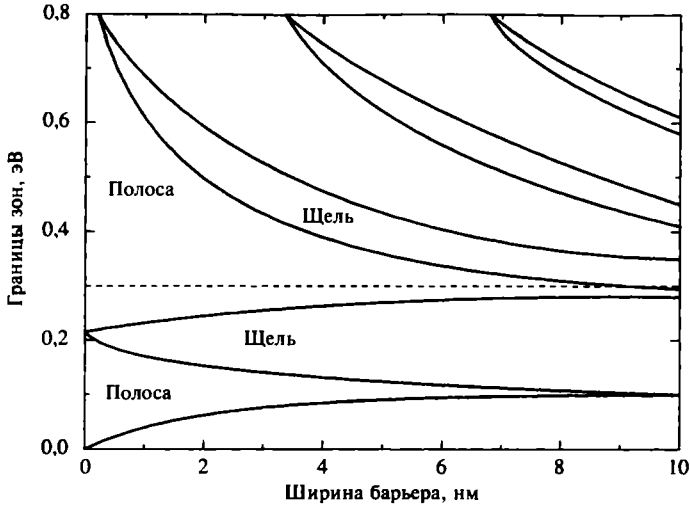


Рис. 5.11. Диаграмма энергетических зон для сверхрешеток, в которых ширина барьера совпадает с шириной ямы (по данным работы [5])

нулю детерминанта коэффициентов, что, после некоторых вычислений приводит к важному соотношению [4]:

$$\frac{q - k_0^2}{2qk_0} \sin k_0 a \sinh qb + \cos qa \cosh qb = \cos q(a + b). \quad (5.14)$$

Полученное уравнение достаточно легко решается численными методами. Рассмотрим, например, достаточно простой случай с $a = b$ и одинаковой эффективной массой электронов в яме и вещества барьера. На рис. 5.11 приведены вычисленные значения разрешенных и запрещенных энергий в сверхрешетке с разрывом зон 0,3 эВ и эффективной массой $m_e^* = 0,067 m_0$, что соответствует системе AlGaAs/GaAs при $E < V_0$ [5]. Из представленных зависимостей легко заметить, что при значениях a выше 10 нм энергии электронов четко определены и соответствуют отдельным квантовым ямам. Однако как только ширина барьера доходит до или становится ниже 6 нм, ситуация изменяется и в структуре возникают зоны, включая запрещенные.

На рис. 5.12 представлены наиболее характерные особенности зависимостей $E(k)$ для сверхрешеток в рамках модели

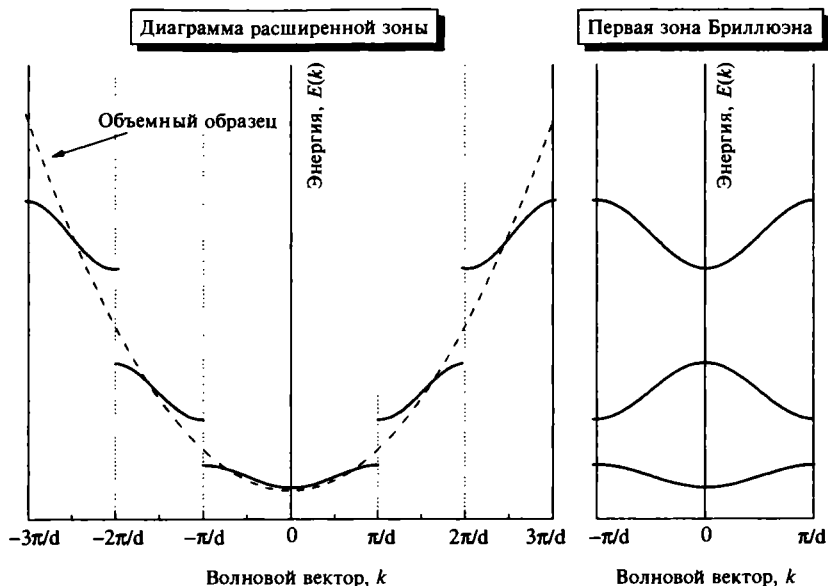


Рис. 5.12. Диаграмма расширенной зоны сверхрешетки в обратном пространстве (слева). Справа приведена диаграмма для приведенной первой зоны Бриллюэна

Кронига — Пенни, которые (после замены постоянных периода сверхрешеток на привычные постоянные решетки) очень похожи на кривые для электронов в объемных твердых телах, рассмотренные в разделе 2.5. Параболическая кривая для свободных электронов разрывается на несколько зон и щелей на границах зон Бриллюэна при значениях $k = \pm\pi n/d$, как показано на рис. 5.12 для диаграммы расширенных зон. Часть зон может быть перенесена в приведенную зону ($-\pi/d \leq z \leq +\pi/d$). Особо следует отметить, что все происходит таким образом, как будто потенциал сверхрешетки расщепляет зону квазисвободных электронов твердого тела в центре приведенной зоны. Поскольку обычно $a \ll d$, ширина таких разрешенных и запрещенных зон в сверхрешетках всегда намного меньше, чем ширина зоны полупроводника, вследствие чего их часто даже называют мини-зонами и мини-щелями. Такой эффект получил название расщепления зон, так как в результате отдельные участки зон в расширенной зоне расщепляются на меньшие зоны со значениями k меньше $2\pi/a$.

Эффект расщепления зон может оказывать принципиальное влияние на характер прямых и непрямых переходов в полупроводниковых структурах. На рис. 5.13 представлена энергетическая диаграмма для типичного полупроводника с непрямыми переходами с минимумом зоны проводимости на краю зоны. Предположим, что мы создаем сверхрешетку из чередующихся монослоев двух типов полупроводников, обладающих похожими электронными характеристиками и хорошо согласованными значениями постоянных решеток, и зона проводимости одного из них имеет вид, показанный на рис. 5.13. В этой ситуации, как показано на рисунке штрихованной линией, минимум зоны проводимости вследствие описанного расщепления зоны смещается к значению $k = 0$. Поэтому рассматриваемая полупроводниковая структура будет обладать запрещенной зоной с прямыми оптическими переходами в области $k = 0$ такой же ширины, что зона с непрямым переходом.

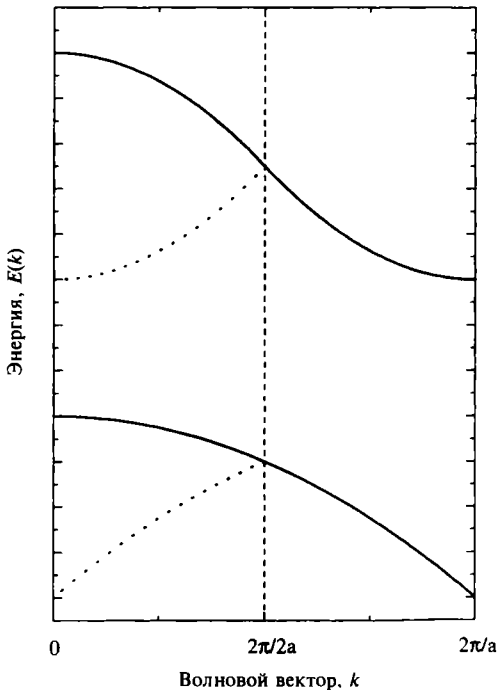


Рис. 5.13. Влияние расщепления зон на энергетическую диаграмму полупроводника с непрямым переходом

Пользуясь представлением об эффекте расщепления зон, можно создавать сверхрешетки из GaAs и AlAs, обладающих запрещенными зонами с прямыми и с непрямыми оптическими переходами, которые будут демонстрировать квазипрямые оптические переходы. Еще одним путем развития является создание сверхрешеток из Si_nGe_m на основе гетероструктур SiGe (см. раздел 5.3.2). В таких сверхрешетках (обычно их создают симметричными, $n = m$) минимум зоны проводимости кремния вблизи края зоны может быть приведен в положение k , близким к $k = 0$ последовательным расщеплением зон. Очевидно, что при возрастании числа монослоев n поведение такой сверхрешетки будет все больше напоминать полупроводниковую структуру, обладающую запрещенной зоной с прямыми оптическими переходами.

Особо следует подчеркнуть, что сверхрешетки находят множество применений в инфракрасной оптике благодаря малой ширине энергетических зон и щелей, а также квазипрямым оптическим переходам. Кроме того, они могут демонстрировать такие интересные свойства, как локализация Ванье — Штарка и блоховские осцилляции (см. раздел 8.5).

5.5.3. Приближение сильной связи в теории сверхрешеток

В этом разделе мы получим ряд свойств зонных диаграмм сверхрешеток, используя описанное в разделе 2.5.2 приближение сильной связи из физики твердого тела, подобно тому, как в предыдущем разделе мы применили для описания сверхрешеток модель Кронига — Пенни. Мы будем рассматривать сверхрешетку в качестве набора из N слабосвязанных квантовых ям, расположенных вдоль оси z (по аналогии с потенциалами, воздействующими на электроны в твердых телах). Например, блоховская волновая функция основного состояния сверхрешетки (обозначаемая ниже через $\psi_{g.s.}$) должна представлять собой линейную комбинацию волновых функций для каждой из квантовых ям $\psi(z - nd)$ с потенциальной энергией $V(z - nd)$, где предполагается, что центр каждой из ям располагается точно при координате $z = nd$. Такая линейная комбинация записывается в виде

$$\psi_{g.s.} = \frac{1}{\sqrt{N}} \sum_n e^{iqnd} \psi(z - nd). \quad (5.15)$$

В соответствии с методами приближения сильной связи необходимо записать гамильтониан в уравнении Шрёдингера для волновой функции в виде суммы $H_0 + H_1$, где член H_0 относится к одной изолированной квантовой яме, а H_1 описывает возмущение (в данном случае потенциал, создаваемый всеми остальными ямами). Решая так называемую возмущенную задачу с учетом воздействия ближайших соседей по методу, использованному ранее для вывода уравнения (2.49), получаем формулу

$$E(q) = E_0 + s + 2t \cos qd, \quad (5.16)$$

из которой следует, что форма зоны описывается синусоидой (рис. 5.14, а) напоминает аналогичную зависимость для зон в твердых телах (рис. 2.6) [6]. Отметим также, что из уравнения (5.16) сразу следует, что ширина зоны равна $\Delta E = 4|t|$, т. е. зависит от интеграла переноса t , который вводится для учета взаимодействия между ближайшими соседями и сам зависит

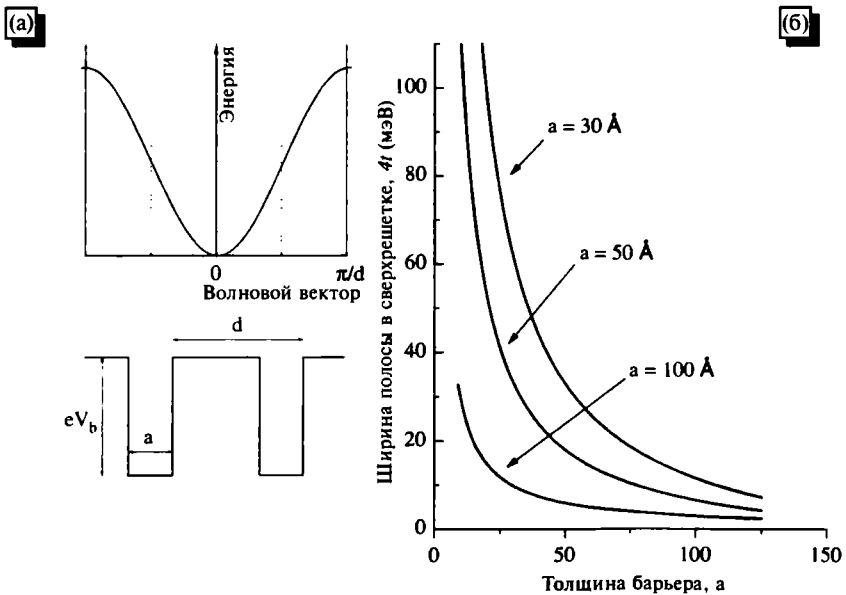


Рис. 5.14. (а) Энергетическая диаграмма для сверхрешетки в приближении сильной связи; (б) зависимость ширины энергетических зон от толщины барьера между квантовыми ямами (по данным работы [6])

от параметров сверхрешетки. На рис. 5.14, б приведена зависимость ширины зоны от ширины барьера между ямами. Как и следовало ожидать, при значительном увеличении ширины барьера (т. е. при переходе к $t \approx 0$) мы получаем решение для системы из многих отдельных квантовых ям с отдельными уровнями энергии.

Необходимо подчеркнуть еще одно важное обстоятельство, связанное с периодичностью сверхрешетки в z -направлении, поскольку реальные сверхрешетки являются все же трехмерными структурами. Поэтому понятно, что выражение для полной энергии должно учитывать и кинетическую энергию движения электронов в плоскостях (x , y). С учетом этого полная энергия электронов в i -й подзоне должна складываться из кинетической энергии и энергии, соответствующей движению поперек плоскости сверхрешетки. Полное выражение имеет вид

$$E(k, q) = \frac{\hbar^2 k_{\parallel}^2}{2m_{\parallel}} + E_0 + s = 2t \cos qd, \quad (5.17)$$

где нижние индексы k и m относятся к двумерному движению через поверхность раздела. Функция плотности состояний может быть вычислена методом, описанным в разделе 4.2 для отдельной квантовой ямы, где был получен множитель $m_{\parallel}/\pi\hbar^2$ для каждой зоны. Используя полученное выражение и интегрируя по q от $-\pi/d$ до $+\pi/d$, после очевидных вычислений [4] можно получить выражение для функции плотности состояний в явном виде

$$n_{sl} = N \frac{m_{\parallel}}{\pi^2 \hbar^2} \cos^{-1} \left(\frac{E_i - E_{oi} - s_i}{2t_i} \right). \quad (5.18)$$

На рис. 5.15 представлена функция плотности состояний для сверхрешетки и для отдельной квантовой ямы [5]. Для сравнения приведена также соответствующая функция плотности состояний для трехмерной структуры. Отметим, что при $E = E_n$ (т. е. при значениях энергии, совпадающих с энергией подзон) функция плотности состояний для сверхрешетки равна половине энергии для отдельной квантовой ямы. Стоит также подчеркнуть, что по мере возрастания значения интеграла переноса t отклонение от спектра квантовой ямы также возрастает.

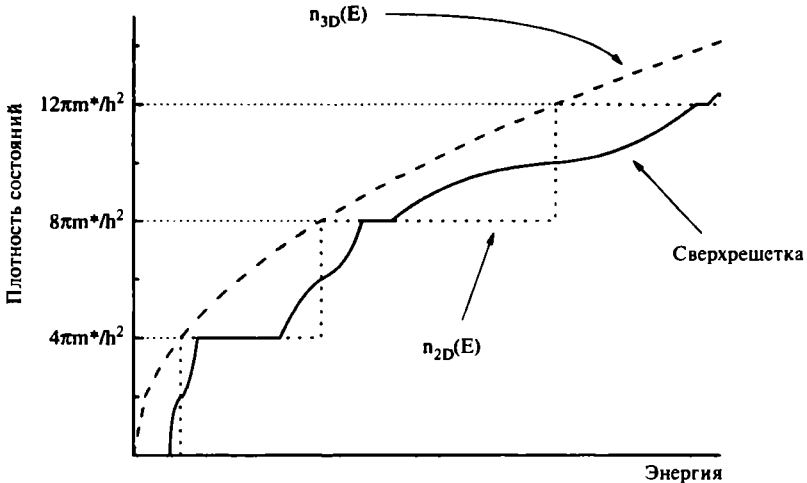


Рис. 5.15. Функция плотности состояний для сверхрешетки

5.5.4. Сверхрешетки типа *ipi*

Рассматриваемые до сих пор сверхрешетки представляли собой лишь периодические структуры из отдельных квантовых ям (рис. 5.10), формируемые чередованием двух полупроводниковых материалов с различными запрещенными зонами. Как предлагали еще Эсаки и Тцу, существует альтернативный метод получения сверхрешеток из периодически упорядоченных *p-n* гомопереходов, когда разрыв зон отсутствует, так как *p-* или *n-*тип материалов (например, кремния) не приводит к изменению ширины запрещенной зоны. Такой тип структур называется легированными или *ipi* сверхрешетками. В аббревиатуре *ipi* буква *i* означает собственный полупроводник (intrinsic), который в этом случае всегда (даже при очень малой толщине слоя) разделяет области *p-* и *n-*типа проводимости. В некоторых случаях в системы специально вводят и дополнительные слои собственного полупроводника. В отличие от сверхрешеток квантовых ям (где пространственное модулирование электрического потенциала вдоль направления роста достигается за счет разрывов в зоне проводимости и валентной зоне), структуры *ipi* позволяют модулировать потенциал вводом ионизированных доноров или акцепторов, которые и создают заряженные *p-n* области вблизи поверхностей контакта.

На рис. 5.16, а представлена гипотетическая ситуация для p - и n - областей, в которых доноры и акцепторы ранее не были ионизированы. При контакте таких областей доноры и акцепторы получают соответственно положительные и отрицательные заряды, вследствие чего возникает разность потенциалов, обусловленная пространственными зарядами, возникающими из-за искривления E_c и E_v зон, как показано на рис. 5.16, б. Учитывая электростатику p - n переходов, представляется очевидным, что амплитуда модуляции E_M зависит от концентрации легирующих атомов (N_D , N_A). Как легко понять из рис. 5.16, б, ширина запрещенной зоны для сверхрешетки (т. е. разность энергий E_c и E_v) равна разности ширины запрещенной зоны основного вещества E_G (формирующего гомопереход) и энергии E_M , т. е.

$$E_{SL} = E_G - E_M. \quad (5.19)$$

Из сказанного ясно, что ширину запрещенной зоны для сверхрешетки можно регулировать, меняя концентрацию примесей. Из рис. 5.16, б также следует, что потенциальные ямы для электронов располагаются в n -областях. Вблизи минимума ямы ее потенци-

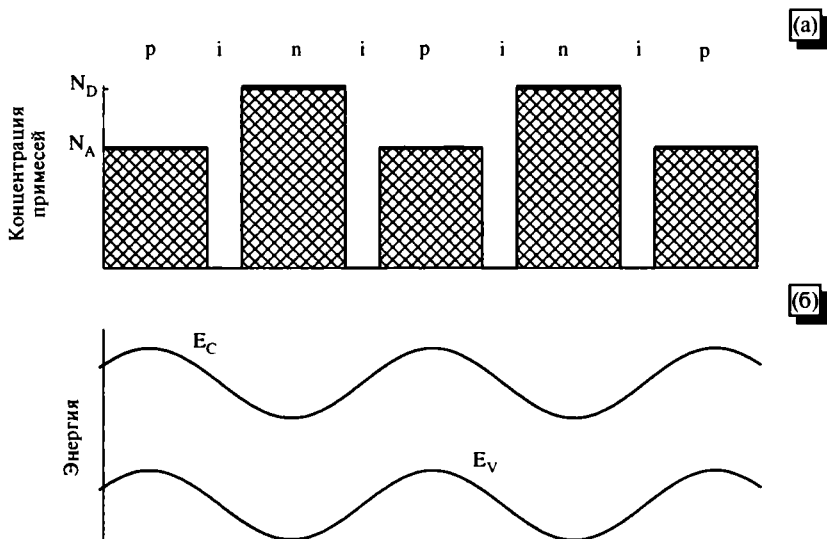


Рис. 5.16. (а) Концентрация примесей в сверхрешетках типа ip_i ; (б) изменения дна зоны проводимости и потолка валентной зоны в сверхрешетках типа ip_i

ал можно аппроксимировать параболой (раздел 4.4.1). Квантование энергии электронов в зоне проводимости можно рассчитать в рамках модели гармонического осциллятора, пользуясь уравнением (4.11). Из решения уравнения Пуассона для p - n перехода легко найти (см. ниже задачу 5.5) значения E_n для квантовых уровней энергии:

$$E_n = \hbar e \left(\frac{N_D}{\epsilon m_c^*} \right)^{1/2} \left(n + \frac{1}{2} \right), \quad n = 0, 1, 2, \dots \quad (5.20)$$

которые зависят от основных характеристик полупроводникового материала (эффективной массы электронов, диэлектрической постоянной и концентрации примесей). Из других особенностей сверхрешеток типа *nipi* следует выделить следующие: 1. Подвижность электронов в них обычно ниже, чем в композиционных сверхрешетках, поскольку носители заряда в них пространственно не разделены с атомами примесей. 2. Оптическое поглощение в них обычно слабее, чем в композиционных, из-за того, что переходы из валентной зоны в зону проводимости происходят с меньшей вероятностью, так как электроны и дырки располагаются пространственно разделены. 3. По этой же причине, т. е. вследствие пространственного разделения неравновесных носителей заряда, их время жизни в сверхрешетках типа *nipi* существенно превышают привычные значения. 4. Сверхрешетки типа *nipi* находят интересное применение при создании оптических вычислительных устройств на основе квантово-размерного эффекта Штарка (см. раздел 8.4).

Литература

1. von Klitzing K., Dorda G., Pepper M. (1980). Phys. Rev. Lett, 45, 494.
2. Hamaguchi Ch. (2001). Basic Semiconductor Physics (Springer, Berlin).
3. Kronig R. de L., Penney W.G. (1931). Proc. Roy. Soc. A, 130, 499.
4. Basu P.K. (1997). Theory of Optical Processes in Semiconductors (Clarendon Press, Oxford).
5. Esaki L. (1983). Recent Topics in Semiconductor Physics, Eds. Kamimura H. and Toyozawa Y., 1-71 (World Scientific, Singapore).
6. Bastard G. (1993). Acta Electron., 25, 147.

Дополнительная литература

- Davies J.H.* (1998). *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge).
- Ferry D.K., Goodnick S.M.* (1997). *Transport in Nanostructures* (Cambridge University Press, Cambridge).
- Hamaguchi Ch.* (2001). *Basic Semiconductor Physics* (Springer, Berlin).
- Jaros M.* (1989). *Physics and Applications of Semiconductor Microstructures* (Clarendon Press, Oxford).
- Kelly M.J.* (1995). *Low-Dimensional Semiconductors* (Clarendon Press, Oxford).
- Mitin V.V., Kochelap V.A., Strosio M.A.* (1999). *Quantum Heterostructures* (Cambridge University Press, Cambridge).
- Murayama Y.* (2001). *Mesoscopic Systems* (Wiley-VCH, Weinheim).
- Schäfer W., Wegener M.* (2002). *Semiconductor Optics and Transport Phenomena* (Springer, Berlin).
- Weisbuch C., Vinter B.* (1991). *Quantum Semiconductor Structures* (Academic Press, Boston).

Задачи и упражнения

1. Полевой МОП-транзистор. Оцените напряженность электрического поля в n -канале полевого Si МОП-транзистора при напряжении затвора $V_g = 4$ В и толщине оксидного слоя 50 нм. Диэлектрические постоянные равны $\epsilon_{ox} = 3,6$ и $\epsilon_{Si} = 12$. (Подсказка: исходя из значения электрического поля в оксиде, рассчитайте поле на границе раздела, используя граничные условия для вектора смещения.) Рассчитайте энергию основного состояния. (Подсказка: воспользуйтесь треугольной формой для потенциала квантовых ям.)
2. Модулированно-легированная гетероструктура. Рассмотрите характеристики модулированно-легированной гетероструктуры вида AlGaAs/GaAs с плотностью носителей $n_{2D} = 10^{12}$ см⁻². Рассчитайте положение уровня Ферми при нормальных температурах, предполагая, что все электроны локализованы в первой подзоне. (Подсказка: начните с доказательства того, что если энергия первой подзоны равна E_1 , то

$$n_{2D} = \frac{m_c^* kT}{\pi \hbar^2} \ln \left[1 + \exp \left(\frac{E_F - E_1}{kT} \right) \right].$$

3. Минизоны в сверхрешетках. Рассмотрим сверхрешетку $\text{AlGa}_{1-x}\text{As}/\text{GaAs}$ при значении $x \approx 0,3$, ширине ямы $a = 10$ нм, ширине барьера $b = 2,2$ нм и высоте барьера $V_0 = 0,25$ эВ. Исходя из модели Кронига — Пенни, описанной в разделе 5.5.2, проведите следующие расчеты: а) найдите ширину минизон и минищелей, которые могут существовать внутри барьера; б) исходя из заданного значения ширины ямы ($a = 10$ нм), нарисуйте зависимость энергии первой ($n = 1$) минизоны проводимости и первой минизоны для тяжелых дырок от ширины барьера b , при ее изменении в диапазоне 0—5 нм.
4. Сверхрешетки Si/SiGe . Рассчитайте период сверхрешетки Si/SiGe , при котором минимум зоны проводимости Si (который соответствует волновому числу $k \approx 0, 8\pi/a_0$, где a_0 — постоянная решетки) может быть приведен к $k = 0$. (Подсказка: в соответствии с концепцией расщепления зон, предложенной в разделе 5.5.2, рассматриваемая сверхрешетка может вести себя подобно объемному полупроводнику с прямыми оптическими переходами.)
5. Сверхрешетки типа *nipi*. Рассмотрим сверхрешетку арсенида галлия типа *nipi* при концентрации доноров и акцепторов $N = N_D = N_A$, составляющей около $5 \times 10^{17} \text{ см}^{-3}$. Рассчитайте расстояния между уровнями в зоне проводимости и валентной зоне. Покажите, что эффективная ширина запрещенной зоны равна

$$E_g = E_{\text{bulk}} + E_{\text{el}} + E_{\text{hl}} - 2V_0,$$

где V_0 — амплитуда периодического потенциала. Покажите также, что при $N = N_D = N_A$ амплитуда периодического потенциала для сверхрешетки определяется уравнением

$$V_0 = \frac{e^2}{2\epsilon} N z_0^2,$$

где z_0 — ширина обедненного слоя. Рассчитайте величину Латинжера и эффективную ширину запрещенной зоны, исходя из заданных концентраций и считая $z_0 = 20$ нм.

ГЛАВА 6

ПРОЦЕССЫ ПЕРЕНОСА В НАНОСТРУКТУРАХ В ЭЛЕКТРИЧЕСКИХ ПОЛЯХ

6.1. Введение

В предыдущих главах было описано образование квантовых ям на поверхностях раздела полупроводников с различными запрещенными зонами. Электроны в зоне проводимости таких квазидвумерных ям движутся практически свободно в плоскости, параллельной границам ямы. Для описания такого процесса переноса, обычно называемого продольным, вполне обоснован полуклассический подход, в некоторых чертах подобный используемому в трехмерном случае. Основные различия в теории переноса между этими ситуациями связаны, естественно, со свойствами функции плотности состояний и с особенностями механизма рассеяния электронов в физических структурах с пониженной размерностью. Однако процессы переноса через потенциальные барьеры в области границы раздела, которые принято называть поперечным переносом, совершенно отличаются от переноса в обычных трехмерных объектах, поскольку определяются в основном квантовым туннелированием.

В этой главе мы будем рассматривать процессы переноса, обусловленные только наличием электрического поля, а в следующей главе обобщим рассмотрение на системы, в которых одновременно действуют и электрические, и магнитные поля. Читатель сможет убедиться, что в полупроводниках с пониженной размерностью под влиянием магнитного поля проявляются совершенно новые квантовые эффекты, относящиеся к процессам переноса (например, квантовый эффект Холла и эффект Аронова — Бома). Однако хотелось бы отметить, что даже под влиянием

только электрического поля в наноструктурах могут наблюдаться неожиданные и интересные явления, связанные с переносом (квантование проводимости, кулоновская блокада и др.).

6.2. Продольный перенос

Электронный перенос в двумерных квантовых гетероструктурах, направленный параллельно потенциальным барьерам на поверхности раздела, может рассматриваться в рамках полуклассического подхода, подобно тому, который используется для описания объемных объектов. Разумеется, мы должны учесть дополнительные механизмы рассеяния электронов (например, рассеяние на «шероховатостях» поверхности раздела), а также особенности низкоразмерных систем. Изучение продольного переноса в наноструктурах началось с измерений электронной проводимости вдоль канала полевых МОП-структур. Эти исследования завершились большим успехом, и по их результатам в 70-х годах было начато промышленное производство полевых МОП-транзисторов, основу которых составляют модулированно-легированные квантовые гетероструктуры. Электроны в таких структурах двигаются в области, свободной от заряженных атомов примесей, вследствие чего их подвижность значительно повышается.

6.2.1. Механизмы рассеяния электронов

Основные механизмы рассеяния электронов при продольном переносе в полупроводниковых наноструктурах связаны, как и в объемных образцах, с фононами и атомами примесей (заряженными или нейтральными). Кроме того, возникают и дополнительные механизмы, специфические именно для наноструктур (например, упомянутое выше рассеяние на «шероховатостях» поверхности раздела). Ниже все эти механизмы рассматриваются отдельно.

1. Электрон-фононное рассеяние.

Расчеты механизмов электрон-фононного рассеяния в низкоразмерных полупроводниковых структурах показывают, что они во многом схожи с процессами в объемных полупроводниках, например, такое рассеяние является преобладающим при температурах выше 50 К и т.п. Однако существует и значительное различие по сравнению с трехмерными структурами, обусловленное тем, что при очень малой ширине квантовых ям

a возрастает роль акустических фононов. Это различие обусловлено отсутствием инвариантности при движении в перпендикулярном направлении, например, для двумерных квантовых ям, где неопределенность в перпендикулярной компоненте момента должно быть $\geq \hbar/a$. Поэтому значение момента акустических фононов в очень узких квантовых ямах не сохраняется, в отличие от объемных систем, где они обладают хорошо определенным импульсом. Возрастание неопределенности в значении импульса приводит к увеличению числа разнообразных механизмов электрон-фононного рассеяния, а затем и к возрастанию роли таких процессов в низкоразмерных полупроводниках.

Процессы рассеяния на оптических фононах в низкоразмерных структурах также существенно отличаются от аналогичных процессов в трехмерных полупроводниках, особенно в случае сильно полярных материалов типа соединения $A_{III}B_V$. Взаимодействие проявляется с особой силой в квантовых ямах (раздел 5.4), где нет перекрытия энергетических зон оптических фононов полупроводниковой ямы (например, GaAs) и полупроводникового барьера (например, AlGaAs). В таких системах вклад в фоннное рассеяние локальных оптических мод и мод, связанных с поверхностями раздела, становится гораздо более существенным, чем вклад от обычных объемных оптических фононов.

2. Рассеяние на примесных атомах.

При низких температурах в полупроводниках с пониженной размерностью основной вклад в процессы рассеяния (как и в объемных системах) возникает из-за рассеяния на ионизированных или нейтральных примесных атомах. Основное различие между процессами рассеяния в дву- и трехмерных системах возникает из-за наличия продольного переноса, при котором рассеивающие атомы примеси часто пространственно разделены с двумерной плоскостью, в которой движутся электроны. В модулированно-легированных полупроводниках (рис. 5.4, 5.6) заряженные доноры располагаются в барьере AlGaAs, а движение самих электронов происходит в яме GaAs, параллельно поверхности раздела полупроводников. Аналогично в МОП-структуре (рис. 5.1) электроны двигаются внутри инверсного канала, отделенного от атомов примеси, расположенных в тонком слое подзатворного окисла.

Для расчета рассеяния на атомах примесей в квантовой гетероструктуре полевого МОП-транзистора (раздел 5.3.1) необ-

ходимо использовать некоторые упрощающие предположения, например, использовать представление о так называемом δ -легировании, при котором предполагается, что все ионизированные примесные атомы лежат в двумерной области (плоскости) на расстоянии d от электронного канала, а энергия всех электронов, участвующих в процессах рассеяния, близка к значению уровня Ферми. Далее следует предположить, что концентрация примесных атомов не очень велика, т. е. все заряженные примеси взаимодействуют с носителями независимо друг от друга. Как показано, например, в работе [1], исходя из этих предположений, можно легко показать, что подвижность носителей возрастает как $\sim d^3$. С другой стороны, при очень больших значениях d концентрация электронов в канале должна стремительно уменьшаться из-за уменьшения напряженности электрического поля, что приводит к уменьшению крутизны полевого МОП-транзистора. Поэтому для каждой такой структуры должно существовать некое оптимальное значение d .

3. Рассеяние на шероховатостях границы раздела.

На теоретической, абсолютно гладкой границе раздела процессы рассеяния электронов должны быть только упругими, однако реальные поверхности всегда имеют несовершенства на атомарном уровне, вследствие чего отражения носителей перестают быть «зеркальными», а потеря импульса приводит к различным релаксационным явлениям. Собственно говоря, различные процессы рассеяния на поверхностях раздела изучались физиками уже долгие годы, так как они играют важную роль при продольном переносе носителей заряда в тонких пленках. Однако современное, основанное на квантовой механике описание этих явлений, применительно к системам с пониженной размерностью, стало развиваться относительно недавно (в 80-е годы).

Значение таких процессов часто зависит и от конкретного вида системы. Например, они играют не столь значительную роль в модулированно-легированных гетероструктурах, с высокосоввершенными границами раздела, полученными с использованием таких методов роста, как молекулярно-лучевая эпитаксия. В этом случае поверхности получают практически плоские, с небольшим числом моноатомных ступенек. С другой стороны, рассеяние на поверхностях раздела значительно возрастает в МОП-структурах, где слой оксида выращивается термически, вследствие чего его поверхность является не такой

идеальной. Кроме этого, относительный вклад рассеяния на границах раздела зависит от ширины квантовых ям, так как по мере уменьшения их ширины волновые функции электронов глубже проникают в потенциальный барьер на границе оксид — полупроводник, т. е. электроны становятся более «чувствительными» к шероховатости поверхности и вероятность соответствующего рассеяния возрастает. Это обстоятельство объясняет, кстати, некоторое снижение подвижности при увеличении напряжения на затворе. В любом случае рассеяние на неоднородностях поверхности, как и рассеяние на примесях, заметно проявляется лишь при низких температурах, когда фонным рассеянием можно пренебрегать. Наконец, следует отметить, что для узких квантовых проволок вклад рассеяния на поверхностях раздела почти на порядок превосходит вклад аналогичных процессов в двумерных системах. Это особенно заметно и важно в тех случаях, когда проволоки изготавливаются с использованием литографии, так как при этом именно шероховатость границ проволоки становится фактором, определяющим подвижность электронов даже при комнатных температурах.

4. Межподзонное рассеяние.

Рассмотрим двумерную электронную систему, локализованную в потенциальной яме, входящей в состав модулированно-легированной гетероструктуры или полевого МОП-транзистора (см. раздел 5.3.1). Очевидно, что при достаточно высокой концентрации электронов в яме начнут заполняться и уровни, расположенные выше основного состояния E_1 . Представим себе далее, что концентрация электронов в системе настолько высока, что уровень Ферми E_F попадает на квантовый уровень, соответствующий состоянию с $n = 2$. В этой ситуации электроны с энергиями, близкими E_F после процессов рассеяния, могут осуществлять как внутрizonный переход (внутри подзоны с $n = 2$), так и межзонный (между подзонами с $n = 1$ и $n = 2$). Наличие двух каналов рассеяния увеличивает полную вероятность рассеяния электронов, вследствие чего их подвижность уменьшается. Описываемый эффект может быть, естественно, обобщен и на другие подзоны, и в целом можно констатировать, что по мере увеличения концентрации электронов в квантовой яме возникают новые каналы рассеяния, повышающие эффективность общего рассеяния и, следовательно, снижающие подвижность такого двумерного электронного газа. Эффект еще

более заметно проявляется в одномерных системах, где плотность состояний расходится при значениях энергии, совпадающих с энергией квантовых уровней (раздел 4.5).

Влияние подзонного рассеяния на подвижность электронов было исследовано Штёрмером и другими [2] в 1982 г. в модулированно-легированных гетероструктурах на основе AlGaAs/GaAs, к которым был добавлен третий вывод (затвор полупроводника), типа используемого в полевых МОП-транзисторах, для управления концентрацией электронов в ямах. Как легко заметить из рис. 6.1, введение второго канала позволяет обеспечить рассеяние между подзонами с $n = 1$ и $n = 2$. Действительно, при заданном напряжении затвора полевого транзистора уровень Ферми доходит до уровня $n = 2$, в результате чего соответствующая подзона может «принимать» рассеянные электроны, снижая тем самым подвижность в соответствующем интервале напряжений.

6.2.2. Экспериментальные данные по продольному переносу

На рис. 6.2 представлены данные, иллюстрирующие прогресс, достигнутый в области повышения подвижности электро-

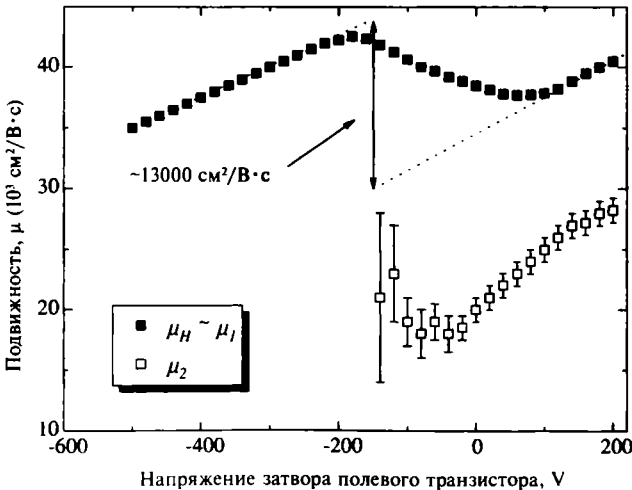


Рис. 6.1. Понижение подвижности электронов в результате межподзонного рассеяния между подзонами с $n = 1$ и $n = 2$ (по данным работы [2])

нов при продольном переносе за последние двенадцать лет [1] в наноструктурах на основе GaAs, типа описываемых выше полевых МОП-транзисторов. Очевидный успех был достигнут за счет нескольких факторов, прежде всего, как указывалось выше, за счет физического разделения легирующих примесей и носителей в модулированно-легированных гетероструктурах. Для большей эффективности такого разделения, в структурах начали создавать вспомогательный из «полуизолятора» (*spacer*), играющий роль прослойки между слоем доноров и двумерным слоем электронов в проводящем канале. Такой слой особенно эффективен при низких температурах, когда преобладают механизмы рассеяния электронов на атомах примеси. Другим фактором, способствующим увеличению подвижности электронов, стала высокая чистота используемых объемных материалов, связанная с развитием методов выращивания пленок на основе полупроводников $A_{III}B_V$, что специально показано на рис. 6.2 отдельной кривой для «чистых объемных» материалов (стоит отметить прекрасные показатели в области температур $T \geq 100$ K). Повышение чистоты наращиваемых пленок, связанное с использованием сверхвысокого вакуума и чистых газов, что характерно для таких методов роста, как молекулярно-лу-

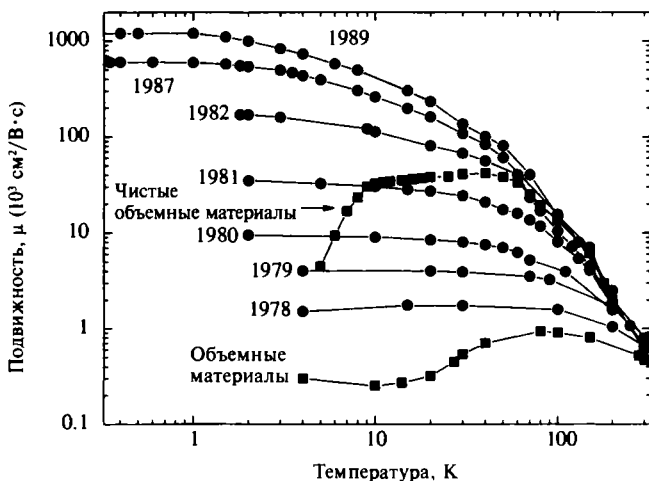


Рис. 6.2. Прогресс в повышении подвижности электронов при продольном переносе в гетеропереходах соединений $A_{III}B_V$ (по данным работы [1]). Для сравнения представлены аналогичные данные для «чистых объемных» материалов

чевая эпитаксия, также явилось важным фактором повышения подвижности. При температурах 100 К и выше, вплоть до комнатной, основными механизмами рассеяния является рассеяние на фононах, особенно связанное с оптическими фононами в случае полярных материалов типа GaAs, что будет рассматриваться ниже в разделе 6.2.1 [1].

Как и следовало ожидать, подвижность электронов в кремниевых полевых МОП-транзисторах оказывается значительно ниже, чем в полевых транзисторах с модулированным легированием MODFET. Как показано на рис. 6.3, подвижность в модулированно-легированных структурах AlGaAs/GaAs может достигать 10^7 см²/В·с, что почти на три порядка выше, чем в полевых МОП-транзисторах на основе кремния [3], что может быть объяснено рядом причин. Во-первых, эффективная масса электронов в кремнии намного больше, чем в GaAs. Во-вторых, влияние рассеяния на примесях в кремниевых полевых МОП-транзисторах, вызываемые зарядами и примесями в окисле и границе раздела, значительно сильнее, чем в AlGaAs/GaAs, где заряженные примеси и носители пространственно разделены достаточно эффективно. В-третьих, как показано на рис. 6.3, при низких температурах доминирующим становится рассеяние на неровностях границы раздела. Этого следовало ожидать вследствие того, как отмечалось в разделе 6.2.1 [3], при термическом окислении граница раздела кремний — окисел не такая совершенная, как в структурах AlGaAs/GaAs, изготавливаемых с использованием гораздо более сложных и точных технологий, таких, как молекулярно-лучевая эпитаксия.

При воздействии очень сильных электрических полей (\geq МВ см⁻¹) рассеяние, обусловленное рассеянием на шероховатости границы раздела становится доминирующим, и именно оно становится главным фактором ограничения подвижности электронов. В этой связи интересно упомянуть, что экспериментальные значения подвижности электронов в квантовых проволоках, определяемые шероховатостью поверхности, вносимой в процессе современных методов микротекстурирования, много меньше значений, предсказываемых теорией.

В последние годы проводились очень серьезные исследования подвижности в напряженных Si—Ge гетероструктурах (см. раздел 5.3.2), которые представляются перспективными

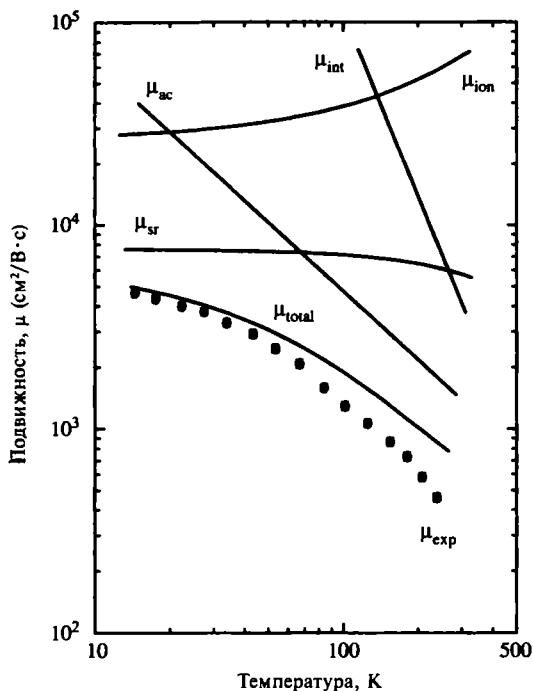


Рис. 6.3. Температурная зависимость подвижности электронов в кремниевых полевых МОП-транзисторах (по данным работы [3])

для производства высокочастотных биполярных транзисторов на гетероструктурах (см. раздел 9.3) и транзисторов с высокой подвижностью электронов, на основе кремниевой технологии. Как известно, в гетероструктурах на основе соединений $A_{III}B_V$ дырки имеют очень большую эффективную массу, что и привлекает внимание исследователей к процессам двумерного переноса дырок в Si—Ge гетероструктурах с достаточно большим разрывом валентной зоны. Считается, что в таких структурах можно будет достичь подвижности дырок вплоть до значений $10^5 \text{ см}^2/\text{В}\cdot\text{с}$.

6.2.3. Продольный перенос горячих электронов

В некоторых типах полевых транзисторов (раздел 9.5) и наноструктур кинетическая энергия электронов, ускоряемых электрическим полем, может становиться очень высокой и значи-

тельно превышать равновесную тепловую энергию, имеющую порядок kT . Естественно, что эффективная температура, соответствующая распределению по энергии таких ускоренных электрическим полем электронов, будет намного выше температуры кристаллической решетки. В этих случаях принято говорить, что распределение электронов «отрывается» от распределения решетки, а сами такие электроны получили название горячих электронов. Следуя квазиклассическому подходу, эффективную температуру электронов для распределения со средней энергией \bar{E} можно определить из соотношения

$$\bar{E} = \frac{3}{2} kT_c. \quad (6.1)$$

Перенос горячих электронов хорошо изучен в объемных полупроводниках, а с начала 90-х годов это явление стали исследовать и в различных наноструктурах. Изучение продольного переноса в гетероструктурах AlGaAs/GaAs [4] показало, что под воздействием электрического поля скорость электронов в них действительно значительно превышает значения для обычных, объемных кристаллов GaAs, причем разница возрастает с уменьшением температуры, как показано на рис. 6.4. Увеличение скорости приписывали квантованию энергии электронов в квантовых ямах. Значения скорости особенно высоки для низшей подзоны ($E = E_1$) по сравнению со второй подзоной ($E = E_2$), в которой электронные волновые функции могут простираться достаточно далеко в область барьера и как следствие носители располагаются гораздо ближе к заряженным донорам, повышая эффективность рассеяния на примесных атомах.

Очень интересный эффект, названный пространственным переносом горячих электронов (RST), возникает при продольном движении горячих электронов в квантовых гетероструктурах, и он уже стал основой нового типа высокочастотных устройств. Этот эффект заключается в том, что при достаточно высокой энергии электронов некоторые из них могут просто «выскочить» из ямы, подобно тому как это показано на рис. 6.5 [5] для квантовых ям в структурах типа AlGaAs/GaAs/AlGaAs, где электроны переходят из нелегированного слоя GaAs в легированный барьер AlGaAs. В электронных приборах на основе структур с пониженной размерностью, подобных показанному

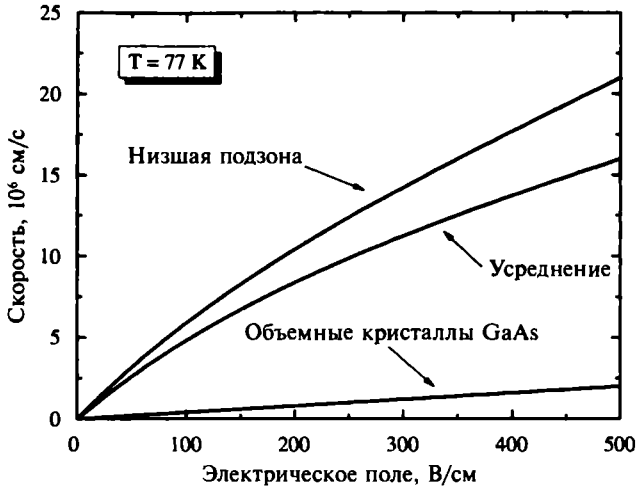


Рис. 6.4. Дрейфовая скорость электронов при продольном переносе в модулированно-легированных гетероструктурах AlGaAs/GaAs (по данным работы [4]). Для сравнения приводится и кривая для объемного GaAs

на рис. 6.5, б, при повышении напряженности между источником и стоком электроны могут переходить из материала с высокой подвижностью электронов (GaAs) в материал с низкой подвижностью (AlGaAs). В результате этого процесса на вольт-амперной характеристике, как показано на рис. 6.5, в, возникает область с отрицательным дифференциальным сопротивлением (ОДС). Ниже, в главе 9 будет показано, как эффект отрицательного дифференциального сопротивления может быть использован для создания нового класса устройств, называемых резонансными туннельными транзисторами.

В объемных материалах движение носителей в электрическом поле обычно изучалось при размерах образцов, значительно превышающих свободный пробег электронов. В современных электронных приборах, основанных на полевом эффекте (например, в полевых МОП-транзисторах), расстояние исток — сток и длина затвора становятся очень короткими (порядка нескольких сотен нм). Такое уменьшение размеров приводит к тому, что электроны в канале ускоряются электрическим полем практически без столкновений. Такие электроны получили название *баллистических*, и достигаемые ими

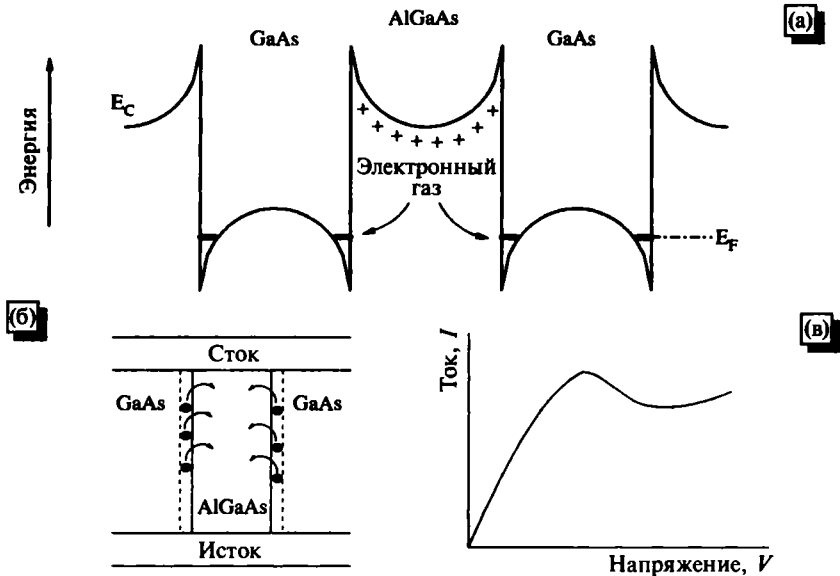


Рис. 6.5. (а) Механизм возникновения отрицательного дифференциального сопротивления (ОДС); (б) структура прибора, работающего с использованием эффекта ОДС; (в) вольт-амперная характеристика (по данным работы [5])

дрейфовые скорости могут достигать значений порядка 10^7 см/с, что вдвое превышает дрейфовую скорость насыщения для объемных полупроводников. Это явление называют эффектом всплеска дрейфовой скорости (velocity overshoot effect), и он уже используется в полевых транзисторах для сокращения времени пролета электронов между истоком и стоком, что позволяет повысить высокочастотные характеристики приборов (см. раздел 9).

6.3. Поперечный перенос

В этом разделе мы рассмотрим движение носителей в направлении, перпендикулярном плоскостям потенциальных барьеров, разделяющих квантовые гетероструктуры. Такой вид переноса часто ассоциируется с квантовым переносом или туннелированием, поскольку при этом энергия носителей может быть меньше энергии, требуемой для преодоления потенциальных

барьеров. Для преодоления частицей потенциального барьера ее волновая функция и ее производная должны быть непрерывными (в указанном перпендикулярном направлении), что сразу приводит нас к задаче о прохождении и отражении на границах раздела. Как мы увидим дальше, туннелирование через потенциальный барьер приводит нас также к концепции отрицательного дифференциального сопротивления на вольт-амперной характеристике, явлению, обнаруженному Эсаки в 1957 г. Через шестнадцать лет после этого, работая в известной фирме ИБМ, Эсаки (в соавторстве с Тцу) первым объявил о регистрации отрицательного дифференциального сопротивления в сверхрешетках AlGaAs/GaAs вследствие резонансного туннелирования сквозь барьеры. Однако лишь в начале 80-х годов удалось изготовить достаточно высококачественные гетеропереходы, чтобы их можно было применить в диодах и транзисторах резонансным туннелированием (см. глава 9).

6.3.1. Резонансное туннелирование

Резонансное туннелирование (РТ) сквозь двойной потенциальный барьер является одним из явлений вертикального квантового переноса, уже нашедший широкое практическое применение в создании диодов и транзисторов, некоторые из которых описаны в главе 9. На рис. 6.6, *а* представлены схематические энергетические диаграммы наноструктуры с двойным барьером, изготовленной из нелегированного GaAs, покрытого с двух сторон слоями AlGaAs, а на рис. 6.6, *б* и 6.6, *в* представлена аналогичная структура, при приложении возрастающего по величине внешнего электрического поля [6]. Резонансное туннелирование происходит при напряжении $V_1 = 2E/e$, где E совпадает с квантовым энергетическим уровнем E_1 . При этом уровень Ферми E_F для металлического контакта слева совпадает с уровнем $n = 1$ ямы и коэффициент туннельного пропускания приближается единице, в результате чего ток через структуру возрастает. Когда величина приложенного поля становится выше $2E_1/e$ и уровень E_F превышает E_1 , ток через структуру уменьшается, как показано на рис. 6.6, *в*. На рис. 6.6, *г* представлена схематически зависимость тока от напряжения (вольт-амперная характеристика) для такой структуры. Очевидно, что при дальнейшем возрастании напряжения V барьеры, которые

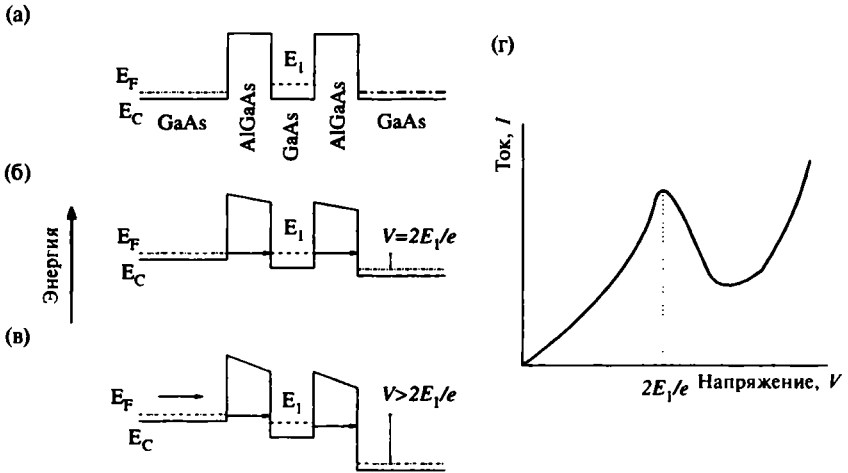


Рис. 6.6. Общая схема, описываемого в тексте, резонансного туннельного эффекта (по данным работы [6])

приходится преодолевать электронам, становятся меньшими по величине, и ток через структуру должен вновь нарастать. Это качественное описание было подтверждено количественными данными экспериментов Эсаки и Тцу как для диодов, так и для сверхрешеток из квантовых гетероструктур, выращенных методом молекулярно-лучевой эпитаксии. Наиболее важной особенностью получаемых вольт-амперных характеристик (типа представленной на рис. 6.6, *з*) является то, что после максимума наклон кривой становится отрицательным, т. е. появляется область отрицательного дифференциального сопротивления. Для понимания характеристик пропускания двойного барьера удобно воспользоваться расчетами, относящимися к одиночным барьерам, для которых вероятность пропускания $T(E)$ (иногда этот коэффициент называют просто прозрачностью барьера) непрерывно возрастает с ростом энергии E электронов в диапазоне энергий $E/e < V_0$. Ситуация кардинально изменяется в случае двойного барьера, когда сама функция $T(E)$ приобретает более сложный вид и представляет собой произведение двух величин, а именно T_E (для первого барьера или эмиттера) и T_F (для второго барьера или коллектора), что дает

$$T(E) = T_E T_C. \quad (6.2)$$

При этом нас вновь будут интересовать лишь ситуация, когда E меньше высоты барьеров. Для нахождения $T(E)$ можно воспользоваться так называемым методом обращения матриц, хорошо известным из учебников по квантовой механике и оптике. Метод связывает коэффициенты падающих и отраженных волновых функций от двух соседних барьеров при помощи матрицы 2×2 , называемой матрицей переноса. Задача наиболее просто решается в случае идентичности барьеров. Коэффициент пропускания такой двухбарьерной структуры определяется соотношением [3]

$$T(E) = \frac{T_0^2}{T_0^2 + 4R_0 \cos^2(ka - \theta)}, \quad (6.3)$$

где величины T_0 и R_0 представляют собой коэффициенты пропускания и отражения для одиночного барьера, a — толщина ямы, параметр k является волновым числом электрона для волновой функции внутри ямы, а θ — фазовый угол.

На рис. 6.7 представлена зависимость $T(E)$ от E для некоторой структуры с резонансным туннелированием (RT-структуры), описываемой тремя энергетическими уровнями в квантовой яме. Отметим, что коэффициент пропускания становится равным единице при трех значениях энергии, совпадающих с энергией каждого из уровней, т. е. когда энергия падающего электрона точно равна энергии одного из уровней. При этом ширина резонансного пика возрастает с энергией, что может быть качественно объяснено на основе принципа неопределенности Гейзенберга (в соответствии с этим принципом величина ΔE должна быть обратно пропорциональна времени жизни τ состояний внутри ямы). Туннелирование электронов на более высоких уровнях проходит через более низкие барьеры, вследствие чего им и соответствуют меньшие значения τ .

6.3.2. Влияние поперечных электрических полей на свойства сверхрешеток

В главе 5 уже указывалось, что электронные состояния в сверхрешетках образуют электронные зоны или подзоны, которые гораздо уже, чем соответствующие зоны в обычных кристаллах (см. рис. 5.11, 5.12). Малая ширина зон и энергетических щелей

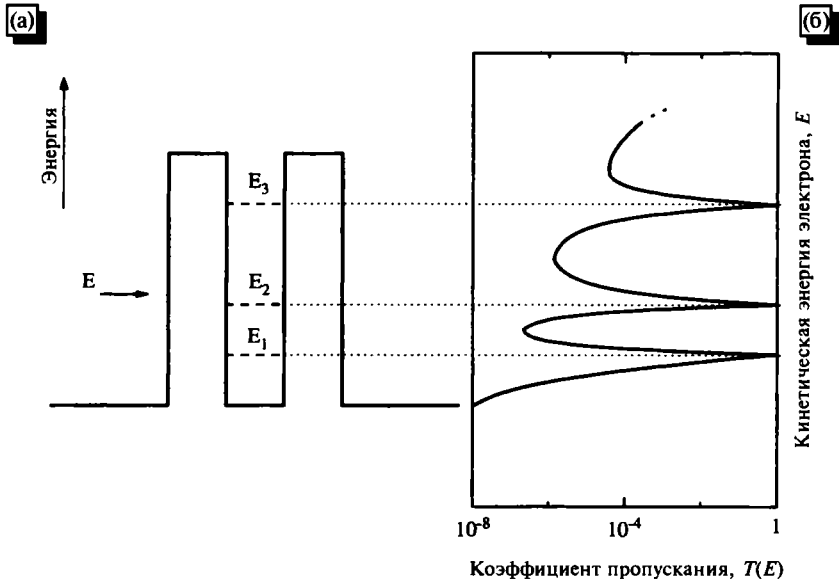


Рис. 6.7. (а) Схема преодоления электроном с энергией E двойного резонансного барьера. Квантовая яма имеет три энергетических уровня (E_1 , E_2 и E_3); (б) зависимость коэффициента пропускания от энергии падающего электрона

является следствием того, что период сверхрешетки d обычно много больше постоянной a решетки кристалла. Ниже в этом разделе будет показано, что при воздействии электрических полей электроны в таких узких зонах проявляют необычные свойства, демонстрируя существование некоторых физических эффектов (типа осцилляций Блоха), которые были, кстати, теоретически предсказаны еще десятки лет назад. Кроме того, выяснилось, что под воздействием поля энергетические уровни ямы (шириной a) в сверхрешетке образуют так называемую штарковскую лестницу из «ступеней» высотой eFa , где F — приложенное электрическое поле.

Рассмотрим электронную зону в k -пространстве, показанную на рис. 6.8, которая похожа на первую подзону в сверхрешетке (см. рис. 5.12). Поле F прилагается в заданном направлении (мы обозначим его осью z , считая его направленным перпендикулярно к плоскости расположения квантовых ям), так что рассматриваемая задача сводится

к одномерной. Движение электрона в такой зоне под влиянием электрического поля описывается введенным в главе 2 уравнением (2.57)

$$\hbar \frac{dk}{dt} = -eF, \quad (6.4)$$

решением которого в случае постоянного поля является значение волнового числа в виде

$$k(t) = k(0) - \frac{eF}{\hbar} t. \quad (6.5)$$

В соответствии с решением (6.5) волновой вектор должен возрастать линейно по времени. Будем считать, что электрон первоначально покоился в точке O начала координатной оси, как показано на рис. 6.8, а направление электрического поля противоположно направлению вектора k . При этом электрон начинает двигаться из точки O по направлению к точке A , и это движение продолжается до тех пор, пока он не достигнет точки B , соответствующей границе зоны Бриллюэна ($k = \pi/d$). В точке B его скорость уменьшается до нуля, в соответствии с нулевым углом наклона кривой, как следует из уравнения (2.52). После этого электрон переносится в точку C (по вектору обратной решетки G), которой соответствует значение волнового вектора $k = -\pi/d$, что просто означает результат брэгговского отражения (см. раздел 2.5.1). Из точки C электрон в k -пространстве под воздействием поля смещается через точку D в точку O , завершая тем самым цикл движения. Скорость электрона при таком периодическом движении определяется уравнением

$$v = \frac{1}{\hbar} \frac{dE}{dk} \quad (6.6)$$

и также меняется периодически, если энергия зоны имеет вид, представленный на рис. 6.8. Иными словами, движение электрона является периодическим одновременно и в реальном, и в k -пространстве.

Период таких колебаний в k -пространстве T_B определяется временем, необходимым для «прохождения» зоны Бриллюэна ($= 2\pi/d$) и равен

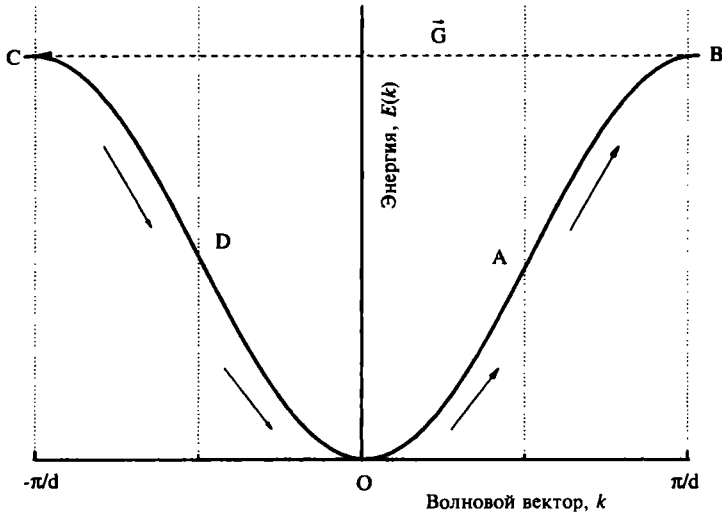


Рис. 6.8. Движение электрона внутри энергетической зоны в k -пространстве под воздействием приложенного электрического поля (процессы рассеяния не учитываются)

$$T_B = \frac{2\pi}{\omega_B} = \frac{2\pi\hbar}{eFd}. \quad (6.7)$$

Следует отметить, что величины T_B и ω_B зависят лишь от периодичности сверхрешетки и напряженности приложенного поля, однако совершенно не зависят от ширины энергетической подзоны. Представляется очевидным, что для экспериментального наблюдения блоховских осцилляций необходимо, чтобы период T_B был меньше времени релаксации, связанного с процессами рассеяния. Ранее блоховские осцилляции экспериментально не могли быть зарегистрированы в объемных кристаллах в силу того, что характерные значения T_B ($\sim 10^{-11}$ с) значительно превышали соответствующие значения в сверхрешетках, поскольку величина d обычно на два порядка превосходит стандартные значения постоянной решетки в привычных полупроводниковых кристаллах. Эти ограничения приводили к тому, что электроны, расположенные близко к точке O на диаграмме рис. 6.8, просто не могли получить достаточно энергии для достижения точки B на границе зоны Бриллюэна ($k = \pi/d$), так как волновой вектор, определяемый уравнением (6.5), не мог возрасти до требуемого значения из-за процессов рассея-

ния, «отбрасывающих» электроны назад к точке O . С другой стороны, следует заметить, что в практических экспериментах значение T_B нельзя уменьшить просто за счет усиления напряженности прилагаемого поля F , поскольку при таком усилении возникает так называемое зеннеровское туннелирование (при котором электроны из наклонной подзоны, как показано на рис. 6.9, б, могут туннелировать через запрещенную зону в соседнюю подзону), в результате чего блоховские осцилляции вообще не возникают. Таким образом, для регистрации блоховских осцилляций необходимо иметь очень узкие подзоны и, наоборот, широкие мини-щели.

На рис. 6.9, а схематически приведена энергетическая структура всего лишь двух подзон в сверхрешетке. При наложенном постоянном по величине электрическом поле F (направленном по оси z) зоны «наклоняются» под углом $-eF$, в результате чего выражение для потенциальной энергии приобретает вид

$$E(z) = E_0 - eFz, \tag{6.8}$$

где E_0 — энергия исходного состояния.

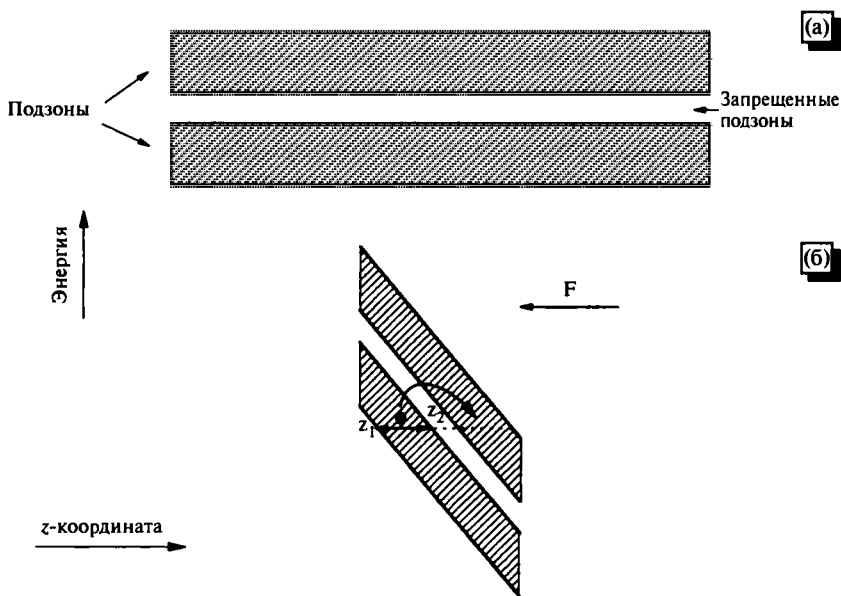


Рис. 6.9. (а) Подзоны в сверхрешетке; (б) наклон подзон в сверхрешетке под воздействием приложенного электрического поля

Вследствие показанного на рис. 6.9, б наклона зон электрон с полной энергией E_T может колебаться в пространстве между положениями с координатами z_1 и z_2 . При возрастании значения F наклон зон увеличивается, вследствие чего электрон пространственно локализуется в меньшем объеме. Очевидно, что при очень высоких напряжениях приложенного поля электрон может быть локализован в пределах одной квантовой ямы, для чего требуется, чтобы разность энергетических уровней ΔE в двух соседних ямах превышала ширину подзон Δ , т. е. должно выполняться условие $\Delta E = eFd > \Delta$. При этом квантовые ямы могут считаться несвязанными (рис. 6.10, а). Таким образом, при значениях электрического поля больше Δ/ed электроны локализуются в квантовых ямах, чьи собственные энергетические состояния существенно различаются, в результате чего понятие подзон становится неприменимым. Вместо этого в системах возникают новые структуры квантовых энергетических состояний, получившие название штарковских лестниц. Такая

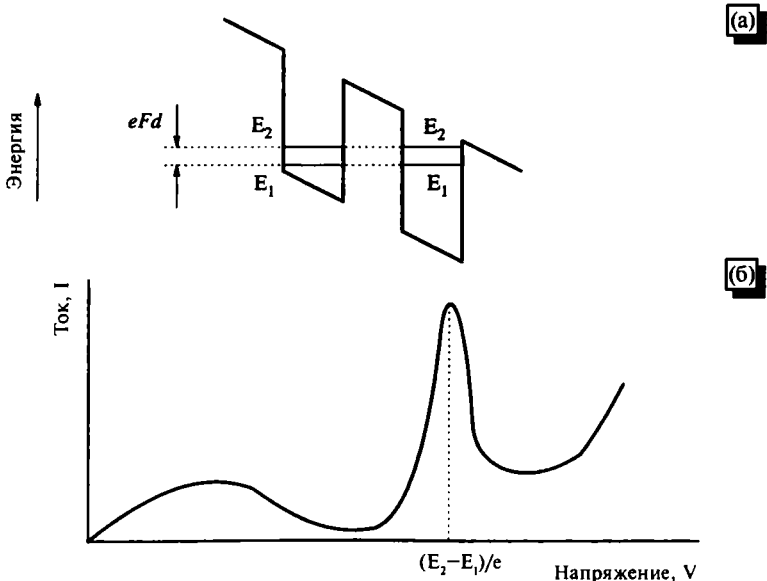


Рис. 6.10. (а) При наложении сильных электрических полей спектр подзон сверхрешетки разрушается, и на его месте возникает система, которую можно рассматривать в качестве множественных квантовых ям (MQW) с разностью энергий $= eFd$; (б) вольт-амперная характеристика ($I - V$) сверхрешетки

штарковская локализация в сверхрешетках типа AlGaAs/GaAs была впервые зарегистрирована в эксперименте Мендесом [6], а затем нашла широкое применение в разнообразных электрооптических приборах (глава 10).

Подобно диодам с резонансным туннелированием, сверхрешетки также имеют участки с отрицательным дифференциальным сопротивлением (NDR) на вольт-амперных характеристиках, что может быть использовано в целом ряде электронных приборов. Такие участки возникают в наноструктурах при наложении настолько сильных электрических полей, что энергия соседних квантовых ям начинает различаться на величину, сравнимую с произведением eFd . В разделе 6.3.1 было показано, что резонансное туннелирование происходит при условии

$$E_2 - E_1 = eFd, \quad (6.9)$$

где E_1 и E_2 означают энергии уровней при наложенном поле F , которые не должны обязательно совпадать с значениями при отсутствии поля ($F = 0$). Как показано на рис. 6.10, б, область проявления эффекта отрицательного дифференциального сопротивления (NDR) располагается на вольт-амперной характеристике сверхрешетки сразу после резонансного пика, что позволяет использовать такие структуры, подобно диодам с резонансным туннелированием, в качестве высокочастотных осцилляторов и усилителей.

6.4. Квантовый перенос в наноструктурах

Рассмотрим далее процессы квантового переноса, происходящие при протекании через наноструктуры тока от присоединенных к ним внешних источников. Такие процессы можно также назвать мезоскопическим переносом, исходя из предложенного в разделе 1.3 термина «мезоскопический», относящегося к системам, промежуточным между макроскопическими и микроскопическими (или атомарными, которые описываются квантовой механикой). В электронике или, точнее, в микроэлектронике такие системы известны как приборы с размерами в субмикронном или нанометровом диапазоне. Очень интересным явлением, проявляющимся при мезоскопическом переносе, является квантование проводимости в единицах $2e^2/h$ (см.

ниже разделы 6.4.1 и 6.4.2). Другое не менее интересное явление называется кулоновской блокадой и может наблюдаться в очень малых наноструктурах (типа квантовых точек) (разделе 6.4.3).

Для наблюдения квантовых эффектов в полупроводниковых наноструктурах должен быть удовлетворен ряд условий. Из наиболее общих требований стоит отметить, прежде всего, то, что при заданной температуре квантовый перенос сильнее проявляется в тех наноструктурах, эффективная масса электрона в которых меньше, поскольку это обычно подразумевает и более высокую подвижность. Кроме того, как уже отмечалось в разделе 4.2, уменьшение эффективной массы способствует повышению энергетических уровней электрона в квантовой яме. В целом можно утверждать, что, чем меньше эффективная масса, тем при более высокой температуре может наблюдаться квантовый перенос.

Перенос в мезоскопических системах обычно происходит в баллистическом режиме (см. раздел 6.2.3), так как их размеры обычно меньше, чем средний свободный пробег электронов, который в гетероструктурах AlGaAs/GaAs при низких температурах обычно составляет несколько микрон. Помимо отсутствия процессов рассеяния, баллистический перенос отличается еще и тем, что при нем электроны не теряют фазовую когерентность, поскольку не участвуют в неупругих столкновениях. Благодаря этой особенности электроны в мезоскопических системах могут демонстрировать фазовые интерференционные эффекты.

6.4.1. Квантовая проводимость.

Формула Ландауэра

Для самого простого описания эффектов квантовой проводимости удобно рассмотреть одномерную мезоскопическую полупроводниковую структуру, типа квантовой проволоки. Если такая проволока является достаточно короткой (т. е. ее длина меньше среднего свободного пробега электрона в рассматриваемом веществе), то движение электронов будет происходить без рассеяния, и перенос будет носить баллистический характер. Предположим, что, как показано на рис. 6.11, такая квантовая проволока идеальными контактами (т. е. такими, в которых полностью отсутствуют процессы рассеяния) соединена с двумя резервуарами, характеризующимися уровнями Ферми E_{F1} и E_{F2} , между которыми приложено слабое напряжение V

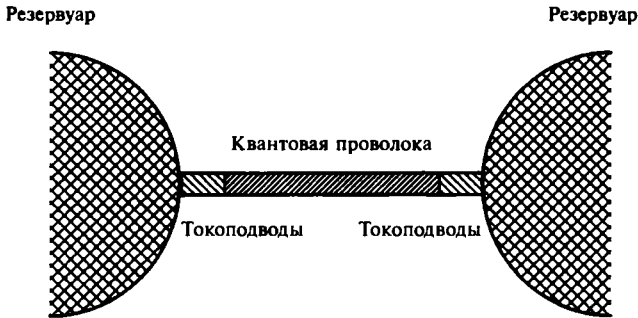


Рис. 6.11. Схематическое представление одномерной мезоскопической системы, используемое для вывода формулы Ландауэра

для обеспечения протекания тока через проволоку. В результате между резервуарами возникает разность потенциалов eV , равная $(E_{F1} - E_{F2})$. Величина протекающего при этом по проволоке тока I равна произведению концентрации электронов (которую можно определить по функции плотности состояний $n_{1D}(E)$ в интервале энергий eV) на скорость электронов $v(E)$ и единичный заряд:

$$I = en_{1D}(E)v(E)eV. \quad (6.10)$$

Подставляя в это выражение формулу (4.21) для плотности состояний $n_{1D}(E)$ (из формулы выбрасывается только коэффициент 2, поскольку в рассматриваемой системе электроны могут двигаться лишь в одном направлении), можно легко получить для тока выражение

$$I = \frac{2e^2}{h}V, \quad (6.11)$$

которое, что довольно интересно, оказывается не зависящим от скорости носителей. Проводимость $G \equiv (I/V)$ при этом равна

$$G = \frac{2e^2}{h}. \quad (6.12)$$

Стоит отметить также, что (в отличие от классической проводимости, обратно пропорциональной длине проводника)

проводимость квантовой проволоки вообще никак не зависит от ее длины. Отношение

$$G_0 = \frac{e^2}{h} \quad (6.13)$$

называется *квантовой единицей проводимости*, а соответствующее обратное отношение

$$R_0 = \frac{h}{e^2} = 25.812807 \text{ к}\Omega \quad (6.14)$$

называется квантовым сопротивлением и может быть измерено экспериментально. Поскольку отношение $2e^2/h$ используется в теории очень часто, его иногда называют также фундаментальной проводимостью.

Все приведенные формулы для квантовой проводимости и сопротивления были получены на основе чрезвычайно простой, одномерной мезоскопической модели, однако сам факт квантования классических физических параметров (типа проводимости и сопротивления) в физике мезоскопических систем имеет фундаментальное значение. Для рассмотрения более сложных систем мы постараемся обобщить полученные результаты. Один из вариантов такого обобщения, предложенный в следующем разделе, состоит в использовании наноструктур с большим числом соединений (а не двух, как в случае одномерной системы). Еще вариант обобщения результатов связан с учетом энергетических подзон в рассматриваемых низкоразмерных полупроводниках (см., например, раздел 4.2). Если концентрация электронов или их энергия достаточно велики, в перенос могут вовлекаться электроны подзон, лежащих выше первого уровня квантования.

Для квантовых проволок такие подзоны (каналы, по терминологии квантового переноса) возникают из поперечных состояний (раздел 4.5). Предполагая наличие нескольких каналов, можно представить, что электроны могут инжектироваться из контактов в любой канал (или моду) m , поступать в мезоскопическую структуру, а затем, после взаимодействия с рассеивающим центром, возникать в другом канале — n . Такие электроны будут вносить свой вклад в полную или общую проводимость системы, равный произведению кванта проводимости $2e^2/h$ на квантово-мехни-

ческую вероятность перехода $|t_{nm}|^2$, соответствующую инжекции электронов в канал m и их переходу в другой канал n (отметим, что в такой формулировке вероятность перехода выражается через амплитуды или вероятности пропускания t_{nm} волновых функций электрона). Полная проводимость в этом случае может быть получена суммированием процессов по всем каналам, т. е.

$$G = \frac{2e^2}{h} \sum_{n,m}^N |t_{nm}|^2, \quad (6.15)$$

где N — полное число каналов, участвующих в рассматриваемых процессах проводимости. Уравнение (6.15), называемое формулой Ландауэра, может рассматриваться как обобщение уравнения (6.12) для мезоскопической системы с двумя контактами и большим числом каналов.

При изучении процессов квантового переноса часто используются наноструктуры, состоящие из сужений внутри двумерной системы. В качестве примера можно привести показанную на рис. 6.12 структуру, в которой движение электронов в двумерной гетероструктуре управляется расщепленным затвором. Использование электрода с такой специальной формой позволяет при приложении напряжения вследствие формируемого распределения потенциала ограничить движение электронов в плоскости двумерной системы и заставить их двигаться в очень малой квазиодномерной области. Такие структуры называют квантовым точечным контактом (QPC) или даже электронным волноводом, по аналогии с привычными волноводами в радиофизике.

На рис. 6.12 представлены результаты первого эксперимента по обнаружению квантовой проводимости, проведенного Визом и другими [7] в 1988 г. на квантовом точечном контакте (форма которого приведена на врезке), образованном в квантовой гетероструктуре AlGaAs/GaAs. Легко заметить, что с ростом приложенного напряжения экспериментально измеренная квантовая проводимость меняется скачками (квантуется) с шагом, равным упомянутой выше фундаментальной проводимости $2e^2/h$. Квантование явно следует из уравнения (6.15), в котором коэффициенты пропускания приближаются к единице вследствие очень низких скоростей процессов рассеяния, что заведомо справедливо для квантовых точечных контактов. При этом экспериментальное наблюдение горизонтальных участков вольт-амперной характери-

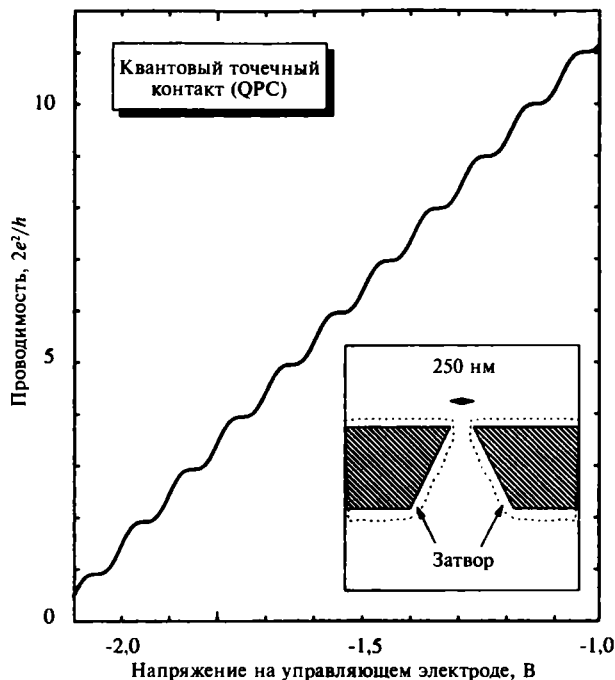


Рис. 6.12. Зависимость квантовой проводимости от напряжения на управляющем электроде (форма которого приведена на врезке) при 0,6 К для квантовых точечных контактов, создаваемых в гетероструктуре AlGaAs/GaAs (по данным работы [7])

ки представляет собой часто сложную задачу, так как эта ломаная линия «сглаживается» в результате многих побочных процессов: влияния неупругого рассеяния, конечного сопротивления контактов, наличия примесных атомов, шероховатости поверхности и т. д. В результате указанных факторов неточность определения экспериментально измеряемых значений ступенек на кривой проводимости может достигать нескольких процентов, что и показано на рисунке. С другой стороны, как будет показано в следующем разделе книги, при наложении сильных магнитных полей, в силу столь же объективных причин, точность измерения высоты ступенек на кривой проводимости повышается на несколько порядков и возрастает до 10^6 раз! Именно по этой причине квантовый эффект Холла, описываемый в разделе 7.7, находит множество применений в метрологии и технике точных измерений.

6.4.2. Формула Ландауэра — Бюттикера для квантового переноса в многозондовых структурах

Полученное в предыдущем разделе выражение (6.15), описывающее квантовый перенос в наноструктуре с двумя контактами, может быть обобщено на случай систем с большим числом контактов. Рассмотрим, например, наноструктуру типа представленной на рис. 6.13, похожую на те, которые часто используются в различных экспериментах, связанных с квантовым эффектом Холла (раздел 7.7), с двумя токовыми контактами, соединенные с соответствующими резервуарами и несколькими потенциальными контактами. Резервуары в данном случае выступают в качестве бесконечных источников и стоков для электронов, причем их температура остается постоянной, даже когда они поставляют электроны в наноструктуру или поглощают их. Мы можем, как и выше, вычислить зависимость тока в каждом подводящем проводе i , соединенном с резервуаром μ_i , предполагая, что каждому из контактов соответствует лишь один канал. Аналогично мы можем построить матрицу рассеяния или прохождения из коэффициентов пропускания T_{ij} , относящихся ко всем комбинациями индексов i и j . Поскольку электроны, попадающие в структуру от любого контакта, могут отражаться, мы должны ввести соответствующие коэффициенты отражения R_i . Кроме этого, для нахождения величины тока I_i (в контакте i) мы должны учитывать и следующие факторы: 1) величину тока, инжектированного через контакт i из резервуара μ_i , равную произведению $(2e/h)\mu_i$; 2) частичное отражение тока обратно в контакт, описываемое коэффициентом отражения R_i ; 3) все токи, поступающие в данный контакт i от других контактов. Сумма таких вкладов, с учетом знака, позволяет записать для тока I_i (в контакте i) выражение

$$I_i = \frac{2e^2}{h} \left[(1 - R_i)V_i - \sum_{j \neq i} T_{ij}V_j \right], \quad (6.16)$$

где через V_i обозначено напряжение, соответствующее μ_i , т. е. $\mu_i = eV_i$. При этом следует отметить, что использованное выше обозначение V_i определяется относительно общего напряжения $V_0 = \mu_0/e$, где μ_0 соответствует низшему уровню распределения Ферми в резервуарах, ниже которого все энергетические состо-

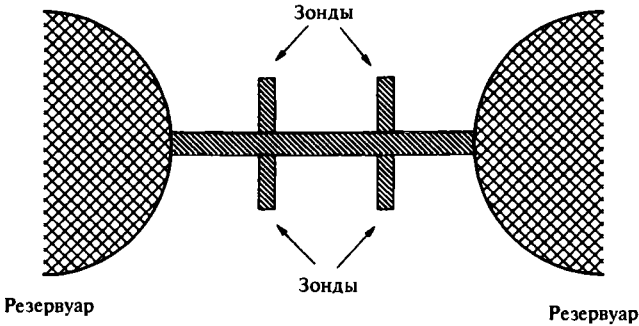


Рис. 6.13. Диаграмма типичной наноструктуры, используемой в экспериментальных измерениях, связанных с квантовым эффектом Холла

яния заполнены и поэтому не могут никак участвовать в процессах переноса носителей заряда. Очевидно, что при близких к $T = 0$ К температурах величина μ_0 должна совпадать с минимальным из значений уровней Ферми для всех μ_i .

Приведенное уравнение получено для контактов с одним каналом. Многозондовое обобщение предполагает, что в каждом контакте i существует N_i каналов распространения, вследствие чего мы должны ввести обобщенные коэффициенты пропускания $T_{ij,\alpha\beta}$, соответствующие вероятности носителя в контакте j и канале β перейти в контакт i канала α . Аналогично должны быть введены и обобщенные коэффициенты отражения $R_{i,\alpha\beta}$, соответствующие вероятности отражения носителя из канала β в канал α для одного и того же контакта i . Учитывая полные вклады в ток через контакт i , можно получить выражение

$$I_i = \frac{2e^2}{h} \left[(N_i - R_i) V_i - \sum_{j \neq i} T_{ij} V_j \right], \quad (6.17)$$

где V_i — напряжение на резервуаре i , а T_{ij} и R_i — приведенные коэффициенты пропускания и отражения, определяемые уравнениями

$$T_{ij} = \sum_{\alpha,\beta} T_{ij,\alpha\beta} \quad \text{и} \quad R_i = \sum_{\alpha,\beta} R_{i,\alpha\beta}. \quad (6.18)$$

Уравнение (6.17) называется формулой Ландауэра — Бюттикера квантового переноса в многозондовых системах для

контактов с большим числом каналов. Стоит подчеркнуть, что в этом уравнении вновь, как и выше в выражениях для квантовой проводимости в разделе 6.4.1, появляется характерный множитель фундаментальной проводимости $2e^2/h$.

Полученное уравнение может быть упрощено с учетом условия сохранения тока в мезоскопической структуре. Полный ток, протекающий через структуру от контакта i , равный разности значений $(2e^2/h)N_iV_i$ и $(2e^2/h)R_iV_i$ (отраженная часть), должен совпадать с суммой всех компонент тока, возникающих от I_i и вытекающих из структуры по другим контактам, т. е. с величиной $(2e^2/h) \sum_{j \neq i} T_{ij} V_{ij}$.

Исходя из этого, можно выписать соотношение

$$N_i - R_i = \sum_{j \neq i} T_{ij}, \quad (6.19)$$

с помощью которого можно переписать уравнение (6.17) в виде

$$I_i = \frac{2e^2}{h} \sum_{j \neq i} T_{ij} (V_i - V_j), \quad (6.20)$$

представляющим собой просто иную запись формулы Ландауэра — Бюттикера.

6.4.3. Кулоновская блокада

Известно, что в микроэлектронных приборах (типа полевых МОП-транзисторов) величина протекающего тока снижается по мере уменьшения размеров самого прибора. Естественно, возникает вопрос о возможных пределах такой зависимости, о том, что произойдет после того, как мы начнем оперировать токами, создаваемыми одним-единственным электроном. Давайте представим себе полупроводник, имеющий наноразмеры во всех трех измерениях, типа квантовой точки (раздел 4.6). В этом разделе мы покажем, что в таких сверхмалых системах или точках наличие или отсутствие даже единичного заряда может приводить к заметным и измеряемым изменениям в характеристиках электрического переноса. Простейшее объяснение такого эффекта (называемого кулоновской блокадой) заключается в следующем.

Представьте себе полупроводниковую квантовую структуру (точку), связанную с двумя резервуарами электронов через по-

тенциальные барьеры или туннельные переходы, как показано на рис. 6.14, а. Барьеры должны быть достаточно тонкими, чтобы электроны могли поступать в резервуары (или покидать их) с помощью туннелирования. На рис. 6.14, б показана энергетическая диаграмма такой квантовой точки, число электронов N в которой меняется поштучно. Естественно, в такой системе может быть создана разность потенциалов за счет подключения внешнего источника напряжения.

Предположим, что мы собираемся увеличить число электронов N в точке на единицу, например, за счет туннелирования в точку одного электрона из левого резервуара. Для этого нам необходимо передать электрону потенциальную энергию eV от внешнего источника. Обозначим исходный заряд квантовой точки через Q , а ее емкость через C . Потенциальная энергия при этом составляет $Q^2/2C$, и, следовательно, электрону для попадания в точку необходимо придать потенциальную энергию, по меньшей мере равную $e^2/2C$. Необходимое для этого напряжение составляет $e/2C$. Так как электрон может либо входить в точку, либо покидать ее (эти процессы эквивалентны, выход электрона означает поступление в точку дырки), туннелирование невозможно, если

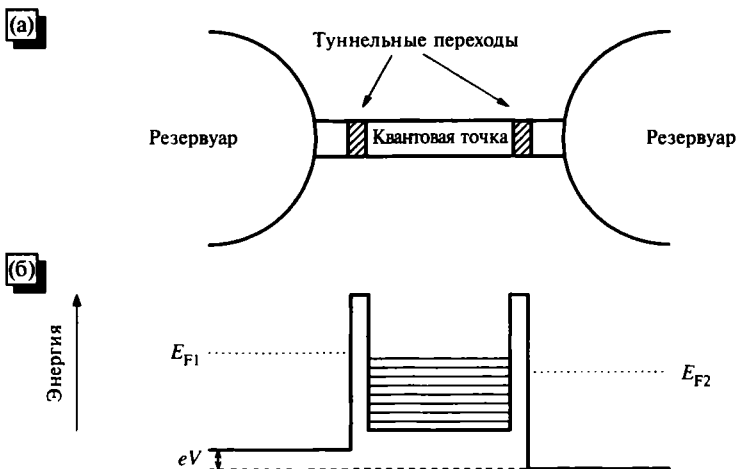


Рис. 6.14. (а) Схематическое изображение квантовой системы для обнаружения эффектов кулоновской блокады; (б) энергетические уровни при приложенном внешнем напряжении. Указаны уровни для N электронов в квантовой точке

$$|V| < \frac{e}{2C}. \quad (6.21)$$

Полученное неравенство означает, что на вольт-амперной характеристике структуры существует диапазон напряжений (от $-e/2C$ до $e/2C$), в котором ток через точку не может протекать ни при каких условиях, как показано на рис. 6.15. Именно поэтому описываемый эффект получил название кулоновской блокады.

Очевидно продолжая этот процесс, т. е. вводя в точку новые электроны, мы получим ситуацию, показанную на рис. 6.16, когда значения протекающего через точку тока будут меняться скачком при значениях приложенного напряжения:

$$V = \left(\frac{1}{2C}\right)(2n+1)e, \quad n = 0, 1, 2, \dots \quad (6.22)$$

Отметим дополнительно, что на рис. 6.16 данные представлены в нормированных координатах (по обеим осям) для того, чтобы выделить и подчеркнуть эффекты квантования тока и напряжения в описываемых структурах.

Особо интересно отметить, что из полученных уравнений вытекает следующее нетривиальное условие: по мере уменьше-

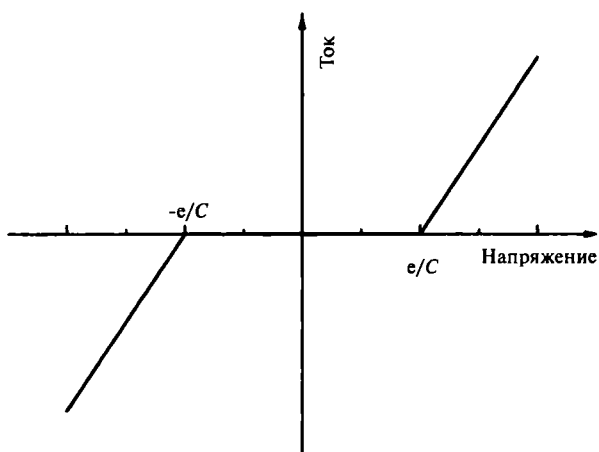


Рис. 6.15. Вольт-амперная характеристика квантовой точки, поясняющая эффект кулоновской блокады

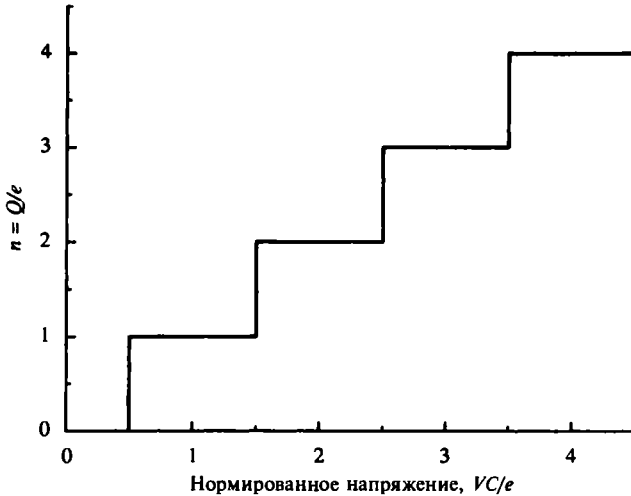


Рис. 6.16. Изменение электрической емкости квантовой точки в зависимости от напряжения в нормированных координатах (пояснения в тексте)

ния размеров квантовой точки (и соответствующего уменьшения величины C) возрастает значение энергии, необходимой для изменения числа электронов в квантовой яме. Это обстоятельство, кстати, облегчает экспериментальное наблюдение кулоновской блокады, поскольку возрастают также значения напряжения и электрической энергии, необходимой для ввода электронов в точку. При этом изменение электрической энергии должно значительно превышать тепловую энергию kT рабочих температур, при которых осуществляются измерения параметров эффекта кулоновской блокады. Поэтому для значений емкости должно выполняться соотношение

$$C \ll \frac{e^2}{kT}, \quad (6.23)$$

которое справедливо либо при очень малых значениях емкости точки (эту величину очень трудно сделать меньше 10^{-16} Ф), либо при очень низких температурах (обычно ниже 1 К).

Кроме того, для наблюдения эффектов, связанных с поведением отдельных электронов, естественно, необходимо, чтобы их число в квантовой яме не флуктуировало в равновесном состоянии. Для оценки этого обстоятельства предположим, что

время перехода электрона в точку (или выхода из нее) имеет порядок $R_T C$, где R_T — эквивалентное сопротивление барьера, а C — емкость точки. Флуктуации числа электронов внутри точки будут приводить к изменениям потенциальной энергии порядка e^2/C , вследствие чего на основе принципа неопределенности можно выписать соотношение

$$\Delta E \Delta t = \frac{e^2}{C} R_T C > \hbar, \quad (6.24)$$

из которого и следует условие

$$R_T \gg \frac{\hbar}{e^2} = 25,8 \text{ k}\Omega \quad (6.25)$$

возможности надежного наблюдения и регистрации эффектов кулоновской блокады в квантовых ямах.

В экспериментах с переносом отдельного электрона обычно измеряют ток, который пропорционален проводимости G , поэтому имеет смысл переписать полученное выше условие, используя именно параметр G , что дает

$$G \ll \frac{e^2}{\hbar}. \quad (6.26)$$

Возможность использования переключающих устройств, оперирующих с одним-единственным электроном, открывает перед электроникой очень интересные перспективы, и в разделе 9.7 мы рассмотрим принципы работы одноэлектронного транзистора (SET), в котором используются именно эти эффекты.

Литература

1. *Davies J.H.* (1998). *The Physics of Low-Dimensional Semiconductors* (Cambridge University Press, Cambridge).
2. *Störmer H.L., Gossard A.C., Wiegmann W.* (1982). *Solid State Commun.*, 41, 707.
3. *Masaki K., Taniguchi K., Hamaguchi C., Iwase M.* (1991). *Jpn. / Appl. Phys.*, 30, 2734.

4. *Schubert E.E., Ploog K.* (1984). *Appl Phys.*, A33, 183.
5. *Luryi S., Zaslavsky A.* (1998). In *Modern Semiconductor Device Physics*, Ed. *Sze S.M.* (Wiley, New York).
6. *Méndez E.E.* (1987). In *Physics and Applications of Quantum Well and Superlattices*, Eds. *Mendez E.E., von Klitzing K.*, NATO ASI Series B, Physics, 170, 159-188 (Plenum, New York).
7. *van Wees B.J.* (1988). *Phys. Rev. Lett.*, 60, 848.

Дополнительная литература

Ferry D.K., Goodnick S.M. (1997). *Transport in Nanostructures* (Cambridge University Press, Cambridge).

Hamaguchi Ch. (2001). *Basic Semiconductor Physics* (Springer, Berlin).

Mitin V.V., Kochelap V.A., Stroschio M.A. (1999). *Quantum Heterostructures* (Cambridge University Press, Cambridge).

Schafer W., Wegener M. (2002). *Semiconductor Optics and Transport Phenomena* (Springer, Berlin).

Задачи и упражнения

1. Рассеяние на примесях при продольном переносе. Подвижность электронов двумерных структур при продольном переносе в области низких температур ограничена (подобно тому, как это показано на рис. 5.4) рассеянием на ионизированных примесных атомах. С использованием некоторых упрощающих приближений она может быть записана в виде [1]

$$\mu \approx 16|d|^3 \frac{e}{\hbar} (2\pi n_{2D})^{1/2},$$

где d — расстояние от плоскости движения электронов до плоскости расположения примесных атомов; n_{2D} — плотность этих атомов. Рассчитайте величину μ для значений параметров $d = 10$ нм и $d = 20$ нм при $n = 10^{16}$ м⁻². Сравните полученные значения μ с данными рис. 6.2 в области низких температур. Объясните качественно сильную зависимость μ от d , а также необходимость согласования условий для больших и малых значений d .

2. Сопротивление отдельного канала. Используя формулу Ландауэра (6.15) для полного сопротивления системы, пока-

жите, что сопротивление одного (отдельного) канала может быть определено в виде

$$R = \frac{h}{2e^2} + \frac{h}{2e^2} \frac{R}{T},$$

где R и T — коэффициенты отражения и прохождения соответственно. Покажите, что если материал канала представляет собой идеальный проводник, то полученное выражение переходит в

$$R = \frac{h}{2e^2}.$$

Объясните, почему сопротивление в этом случае отлично от нуля.

3. Потенциальные контакты. Рассмотрим систему, образованную одномерным проводником с двумя контактами (1 и 2), по которой протекает ток I под воздействием приложенного напряжения V . Потенциальный контакт (контакт 3) присоединяется к середине проводника, между контактами 1 и 2. Покажите, что напряжение на этом контакте V_3 равно $V/2$, т. е. совпадает со значением, предсказываемым классической теорией. (Подсказка: ток в датчике 3 равен $I = 0$. Используя это значение, получите из формулы Ландауэра — Бюттикера (6.17) выражение для V_3 и используйте соображения симметрии $t_{3,1} = t_{3,2}$).
4. Некогерентный переход через двойной барьер. Рассмотрим два туннельных барьера, расположенных последовательно на близком расстоянии. Покажите, что коэффициент пропускания T такого двойного барьера равен

$$T = \frac{|t_1|^2 |t_2|^2}{1 - |r_1|^2 |r_2|^2},$$

где t_1 и t_2 — амплитуды прохождения через отдельные барьеры; r_1 и r_2 — соответствующие амплитуды отражения. (Подсказка: обозначив через α амплитуду от барьера 1 (попадающую в барьер 2), через β — амплитуду от барьера 2 (попадающую в барьер 1) и через γ — выходную амплитуду барьера 2, можно получить соотношения

$$\begin{aligned}
 |\alpha|^2 &= |r_1|^2 + |r_1|^2 |\beta|^2, \\
 |\beta|^2 &= |\alpha|^2 |r_2|^2, \\
 |\gamma|^2 &= |\alpha|^2 |r_2|^2.
 \end{aligned}$$

5. Последовательное соединение двух сопротивлений. Рассмотрим систему из двух последовательно соединенных туннельных барьеров с амплитудами прохождения и отражения через отдельные барьеры, определенными в предыдущей задаче. Покажите, что полное сопротивление такой системы равно

$$R = \frac{h}{2e^2} \left(1 + \frac{|r_1|^2}{|t_1|^2} + \frac{|r_2|^2}{|t_2|^2} \right).$$

Сравните полученный результат с классическим выражением, получаемым по закону Ома.

6. Сопротивления с двумя и четырьмя выводами. Ток величиной I протекает между контактами 1 и 2, расположенными на концах удлиненной структуры. Потенциальные контакты 3 и 4 присоединяются в точках, расположенных между контактами 1 и 2. Определите величину сопротивления между двумя контактами $R_{12,12}$ в виде отношения напряжения V_{12} (между контактами 1 и 2) к току I_{12} между этими контактами. Определите также четырехконтактное сопротивление $R_{12,34}$ в виде отношения разности потенциалов между контактами 3 и 4 к току I_{12} . Пользуясь описанным методом Ландауэра — Бюттикера, получите соотношения

$$\begin{aligned}
 R_{12,12} &= \frac{h}{2e^2} \frac{1}{t_{12}}, \\
 R_{12,34} &= \frac{h}{2e^2} \frac{1}{t_{12}} \frac{t_{31}t_{42} - t_{32}t_{41}}{(t_{31} + t_{32})(t_{41} + t_{42})}.
 \end{aligned}$$

(Возможно, читателя удивит сам факт, что полученные результаты не совпадают, поскольку потенциальные контакты не дают вклад в ток. Эта проблема вызывает много споров.)

7. Регистрация одноэлектронных эффектов. Рассмотрим металлическую квантовую точку в виде плоского круглого диска радиуса R , параллельного бесконечно протяженной металлической плоскости и расположенного на расстоянии

L от нее. Покажите, что для наблюдения одноэлектронных эффектов при комнатных температурах радиус такой квантовой точки R должен составлять несколько нанометров. Рассмотрите аналогичную ситуацию при очень низких температурах. (Подсказка: из электростатики можно предложить, что при $R \ll L$ емкость рассматриваемой квантовой точки составляет $C = 8\epsilon_0\epsilon_r R$. Подставьте в это выражение вместо ϵ_r значение относительной диэлектрической постоянной для кремния.)

8. Одноэлектронный «турникет». Одноэлектронный «турникет» или пропускное устройство состоит из квантовой точки, которая под воздействием приложенного переменного напряжения способна пропускать через себя по одному электрону за период. Два сигнала переменного тока (противоположные по фазе, т. е. различающиеся на 180°), подаваемые на каждый из барьеров, окружающих точку, изменяют глубину ямы на одинаковые значения с противоположным знаком. Рассмотрите работу такого одноэлектронного «турникета» на примере структуры квантовой ямы (рис. 6.14, б). Покажите, что такое устройство генерирует постоянный ток силой $I = ef$, где e — заряд электрона, f — частота приложенного сигнала переменного тока.

ГЛАВА 7

ПЕРЕНОС В МАГНИТНЫХ ПОЛЯХ И КВАНТОВЫЙ ЭФФЕКТ ХОЛЛА

7.1. Введение

В этой главе рассматриваются в основном характеристики переноса двумерных электронных систем в магнитных полях, что уже привело к открытию квантового эффекта Холла, ставшего одним из важнейших достижений современной физики твердого тела. Значение и ценность квантового эффекта Холла для физики особо подчеркивает тот факт, что за открытие этого эффекта Нобелевская премия была присуждена дважды: в 1985 г. фон Клитцингу (за открытие целочисленного эффекта Холла) и Тсуи, Штёрмеру и Лафлину в 1999 г. (за открытие дробного эффекта Холла). В разделе 7.2 дается краткое описание эффектов воздействия магнитных полей на характеристики объемных металлов и полупроводников, в частности, о квантовании орбитального движения электронов на уровнях Ландау. В разделах 7.3 и 7.4 описывается воздействие магнитных полей на двумерные электронные системы в квантовых гетеропереходах. В этих системах постоянная по величине функция плотности состояний «схлопывается» в δ -функцию (в реальных системах ее ширина определяется разбросом из-за рассеяния электронов) вследствие вырождения уровней Ландау.

Воздействие магнитных полей на двумерные электронные системы имеет очень важные практические приложения. Например, в разделе 7.5 будет показано, что токи в мезоскопических системах проявляют интерференционные эффекты, проявляющиеся в модулировании проводимости под воздействием магнитного векторного потенциала \vec{A} (эффект Аронова — Бома). Еще одним важным и интересным явлением являются колеба-

ния Шубникова — де Гааза (раздел 7.4), наблюдаемые в двумерных системах при наложении магнитного поля и проявляющиеся намного сильнее, чем в объемных полупроводниках.

После этих вводных разделов общего характера читателю в разделе 7.7 предлагается описание квантового эффекта Холла. В соответствии с теорией целочисленного квантового эффекта Холла (IQHE, раздел 7.7.1), напряжение Холла в мезоскопических двумерных электронных системах квантуется в виде набора плоских участков, в соответствии с чем, естественно, квантуются и значения проводимости и обратной ей величины — сопротивления. Весьма яркой характеристикой такого квантования является то, что оно не зависит ни от материала, формирующего гетеропереход, ни от геометрической формы изучаемого образца. При этом высота ступеней в холловском сопротивлении сохраняется с такой точностью, что это позволило даже ввести новую фундаментальную постоянную, получившую название константы фон Клитцинга ($R_K = h/e^2$), точность измерения которой может быть доведена до 10^{-10} . Начиная с 1990 г., его даже используют для определения стандарта электрического сопротивления (единицы Ом) в системе СИ. Как будет показано в разделе 7.7.2, связь между формализмом теории целочисленного квантового эффекта Холла (IQHE) и предложенной выше (в главе 6) формулой Ландауэра — Бюттикера для квантового переноса может быть установлена с помощью новых квантово-механических концепций, открытых орбит и краевых состояний. С физической точки зрения наиболее сложными в теории IQHE представлениями являются вводимые в разделе 7.7.3 концепции локализованных состояний, позволяющих объяснить ширину участков «плато» при наблюдении эффекта Холла.

Экспериментальные данные по дробному квантовому эффекту Холла (FQHE) несколько напоминают данные по целочисленному эффекту, хотя следует отметить, что теоретические описания этих эффектов совершенно непохожи. Дробный эффект Холла объясняют на основе так называемого многочастичного взаимодействия электронов при очень низких температурах под воздействием магнитного поля. Полная теория этого явления, объясняющая все его особенности, пока не создана, поэтому ниже, в разделе 7.7.5, мы представим некоторые положения двух основных теорий, кажущихся нам наиболее перспективными. Прежде всего, следует отметить теорию Лафлина (1983), основанную на упомянутом выше многочастичном взаимодействии электронов, что позволяет

автору объяснить дробные значения коэффициентов заполнения для уровней Ландау и даже постулировать наличие возбужденных состояний (квазичастиц), обладающих дробными значениями заряда (например, $1/3$ заряда электрона). Теория составных фермионов (CF), предложенная Джейном (Jain, 1989 и 1990), постулирует существование CF в качестве новой физической единицы, образуемой электроном при соединении с двумя так называемыми флюксонами $\Phi_0 (= e/h)$. В последние годы именно эта теория, предсказывающая существование квазичастиц с дробными электрическими зарядами, получает все больше подтверждений.

7.2. Воздействие магнитного поля на кристаллы

Прежде всего, рассмотрим влияние магнитного поля на электроны проводимости в твердых телах. Из физики твердого тела известно, что очень высокие магнитные поля могут вызывать в кристаллах целый ряд поразительных эффектов, среди которых стоит отметить коллапс (схлопывание) энергетических состояний электронов в уровни Ландау, осцилляции намагниченности M при изменении плотности магнитного потока \vec{B} (эффект де Гааза — ван Альфена), осцилляции электрического сопротивления при изменении магнитного поля (эффект Шубникова — де Гааза) и т. д. Все эти явления связаны с квантованием энергии электронов проводимости под влиянием магнитного поля B_z (направленного по оси z). Квантование описывается формулой

$$E_n = \left(n + \frac{1}{2}\right) \hbar \omega_c + \frac{\hbar^2 k_z^2}{2m_c^*}, \quad (n = 0, 1, 2, \dots), \quad (7.1)$$

где ω_c — циклотронная частота, определяемая уравнением

$$\omega_c = \frac{eB_z}{m_c^*}, \quad (7.2)$$

и соответствующая частоте циклотронных орбит электрона в плоскости (x, y) . Квантовые числа n в уравнении (7.1) соответствуют различным уровням Ландау. С другой стороны, понятно, что магнитное поле, направленное по оси z , не может менять характер движения электронов вдоль этой оси. Как показывает уравнение

(7.1) и данные на рис. 7.1, вдоль оси z движение электронов не отличается от свободного, что и следовало ожидать, так как магнитное поле не создает сил, действующих на электроны, если они движутся параллельно полю. С другой стороны, движение электронов в плоскости (x, y) квантуется и может быть описано в рамках известной модели квантового гармонического осциллятора.

Рассмотрим далее влияние магнитного поля на трехмерную плотность состояний. Из курса физики твердого тела известно, что под воздействием магнитного поля B_z трехмерные разрешенные состояния в k -пространстве «схлопываются» в набор концентрических трубок или цилиндров, параллельных \vec{B} , в результате чего каждая из трубок описывает вырождение состояний, соответствующее числу $g_n = eV/\pi\hbar$. Кроме этого, поскольку каждый из уровней Ландау связан с некоторым одномерным движением свободных электронов вдоль направления \vec{B} , зависимость функции плотности состояний n_{2D} от энергии должна, в соответствии с ранее введенным уравнением (4.20), иметь вид $g(E) \propto 1/\sqrt{E}$. Из

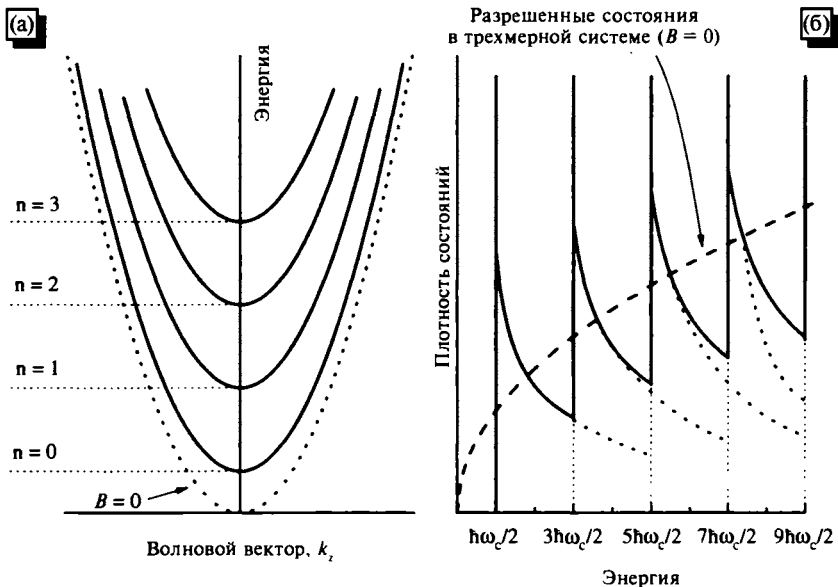


Рис. 7.1. Зависимость энергетических зон электронов в трехмерном твердом теле от волнового вектора в z -направлении для различных уровней Ландау ($n = 1, 2, \dots$); (б) функция плотности состояний для уровней Ландау в сравнении с аналогичной функцией для газа свободных электронов при $B = 0$

этого вытекает наличие сингулярностей в минимумах каждой из подзон (см. рис. 7.1, б), соответствующих каждому из уровней Ландау. Естественно, что в реальной ситуации такие сингулярности размываются на каждом из уровней Ландау ($n = 1, 2, 3, \dots$) из-за разнообразных процессов электронного рассеяния.

7.3. Поведение систем пониженной размерности в магнитных полях

В отличие от объемных кристаллов, где под воздействием магнитного поля \vec{B} движение электронов квантуется только в плоскости, перпендикулярной \vec{B} , в двумерных электронных системах энергетический спектр квантуется полностью, что легко показать, выписав уравнение Шрёдингера для электронов в такой системе, когда к ней приложено магнитное поле, направленное по оси z , перпендикулярной к плоскости системы. В этом подходе можно использовать представление Ландау, при котором вектор-потенциал \vec{A} имеет только одну компоненту (например, A_y), т. е. $A_y = Bx$. Во многих учебниках квантовой механики указывается, что уравнение Шрёдингера в системах с электромагнитным полем может быть получено при замене в обычном выражении канонического момента \vec{p} на разность $\vec{p} - q\vec{A}$ (подстановка Пайерлса). Поэтому мы можем сразу выписать уравнение Шрёдингера для волновой функции $\psi(\vec{r}) = \psi(x, y)$ двумерной системы в виде

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2m} \left(i\hbar \frac{\partial}{\partial y} + eBx \right)^2 \right] \psi(x, y) = E\psi(x, y), \quad (7.3)$$

что после преобразования приводит к выражению

$$\left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) - \frac{i\hbar eBx}{m} + \frac{(eBx)^2}{2m} \right] \psi(x, y) = E\psi(x, y). \quad (7.4)$$

Будем искать решение в виде

$$\psi(x, y) = \varphi(x) e^{iky}, \quad (7.5)$$

где плоская волна соответствует координате y (движение свободных электронов) и вводится для того, чтобы подчеркнуть независимость \vec{A} от y . Подставляя выражение (7.5) для волно-

вой функции в уравнение (7.4), можно сразу получить уравнение для функциональной зависимости от x в виде

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2} m \omega_c^2 (x - x_0)^2 \right] \varphi(x) = E_n \varphi(x), \quad (7.6)$$

где частота ω_c определяется уравнением (7.2), а величина x_0 равна

$$x_0 = \frac{\hbar k}{eB}. \quad (7.7)$$

Легко заметить, что соотношение (7.6) представляет собой уравнение Шрёдингера для одномерного гармонического осциллятора, поскольку добавление члена x_0 к x означает смещение центра параболического потенциала на величину x_0 (поэтому параметр x_0 называют центром координат). Таким образом, мы получили достаточно важный результат, демонстрирующий, что собственные состояния двумерной системы в магнитном поле определяются уравнением

$$E_n = \left(n + \frac{1}{2} \right) \hbar \omega_c, \quad n = 0, 1, 2, 3, \dots \quad (7.8)$$

Стоит отметить, что полученные значения энергии зависят от квантового числа n и величины магнитного поля B (через циклотронную частоту ω_c).

7.4. Плотность состояний двумерных систем в магнитных полях

В предыдущем разделе мы показали, что в сильных магнитных полях, направленных перпендикулярно плоскости квазидвумерной электронной системы, электроны двигаются по циклотронным орбиталам с частотой ω_c , определяемой уравнением (7.2), и что их энергия при этом квантуется в соответствии с правилами для одномерного гармонического квантового осциллятора. Из этого следует, что функция плотности состояний двумерного электронного газа (определенная при $v = 0$) после наложения поля B для каждого из уровней Ландау «схлопывается» в δ -функцию.

На рис. 7.2 показано, как такая функция n_{2D} (которая, как было показано в разделе 4.2, при $B = 0$ является постоянной, т. е. не за-

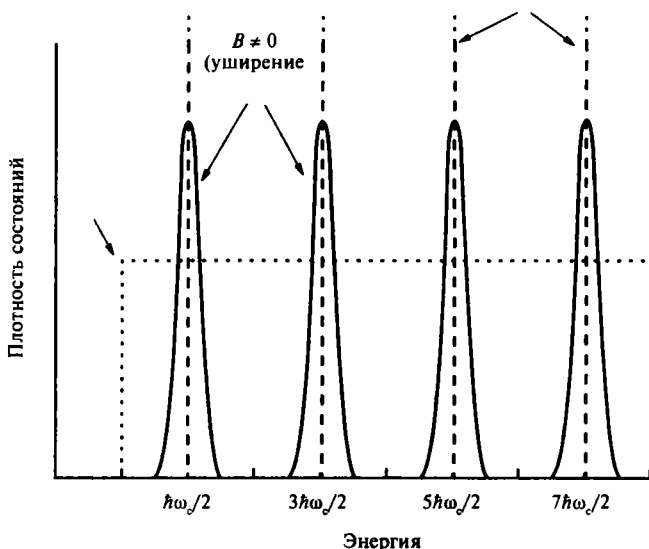


Рис. 7.2. Зависимость функции плотности состояний двумерного электронного газа от энергии в магнитном поле (для сравнения приведена аналогичная зависимость для двумерной системы при $B = 0$)

висит от энергии) для каждого из уровней Ландау превращается в δ -функцию после наложения магнитного поля, направленного перпендикулярно плоскости двумерной системы. Из рис. 7.2 можно видеть, что низший из уровней Ландау соответствует энергии $\hbar\omega_c/2$, которая лежит выше дна параболической подзоны. Вследствие рассеивания электронов на примесях происходит естественное уширение линий, «размывающее» δ -функции, как показано на рисунке. Такие эффекты (вызываемые рассеянием на примесях или взаимодействием с колебаниями решетки, т. е. фононами), могут быть, как обычно, оценены введением характеристической ширины разброса по энергиям Γ . Понятно, что уровни Ландау могут быть надежно выявлены при выполнении условия $\hbar\omega_c > \Gamma$, что можно заменить эквивалентным требованием, чтобы характерное время рассеяния τ_i было достаточно большим (или подвижность электронов системы была достаточно велика).

Поскольку после наложения поля B все уровни в интервале энергий $\hbar\omega_c$ «схлопываются» в один общий уровень Ландау, *вырождение D* уровней Ландау описывается соотношением

$$D = \frac{m_c^*}{2\pi\hbar^2} \hbar\omega_c = \frac{eB}{2\pi\hbar}, \quad (7.9)$$

где использована формула (4.7) из главы 4 для функции плотности состояний n_{2D} , но без учета вырождения по спину. Отметим сразу, что в формуле (7.9) вырождение уровней Ландау линейно возрастает с величиной приложенного магнитного поля, что имеет особо важное значение для объяснения квантового эффекта Холла, которое будет представлено в разделе 7.7.

7.5. Эффект Аронова — Бома

В твердых телах магнитные поля могут приводить к различным интерференционным эффектам между волновыми функциями электронов. Очевидно, что для наблюдения таких интерференционных эффектов фаза электронных волн должна сохраняться. Напомним, что в разделе 1.4 была определена длина фазовой когерентности L_ϕ как среднее расстояние, проходимого электроном без изменения фазы волны. Обычно фаза волновой функции изменяется или «разрушается» при неупругих столкновениях с дефектами решетки, вследствие чего именно электроны, двигающиеся в соответствии с описанным выше баллистическом режиме (раздел 1.8) должны проявлять заметные интерференционные эффекты. Они проходят через решетку практически без столкновений благодаря тому, что средняя длина их пробега l по определению значительно превосходит размеры образца L ($l \gg L$).

В 1959 г. Аронов и Бом предположили, что волновая функция электрона в твердом теле имеет фазовый множитель, величиной которого можно управлять с помощью магнитного поля. Этот эффект был действительно обнаружен в 1985 г. сотрудником фирмы ИБМ Веббом [1] в структуре, типа представленной на рис. 7.3, *a* (состоящей из металлического кольца диаметром 800 нм, изготовленного из проволоки диаметром около 50 нм). Амплитуда волновых функций электронов, попадающих в такую структуру через канал P , разделяется на две равные части, каждая из которых проходит по разным ветвям кольцевой структуры, в результате чего на выходе Q эти амплитуды могут интерферировать. Предположив далее, что создаваемый соле-ноидом магнитный поток Φ проходит внутри кольца, мы полу-

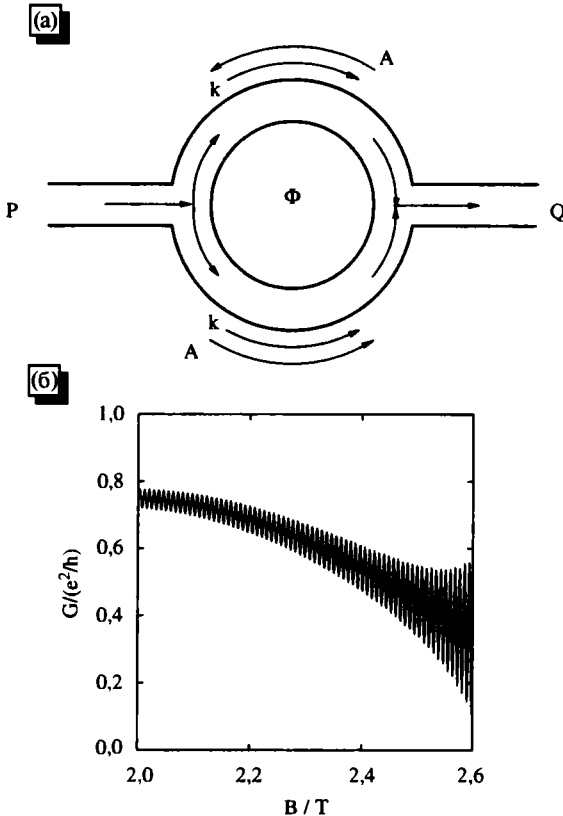


Рис. 7.3. (а) Схема мезоскопической системы для измерения интерференционных явлений при эффекте Аронова — Бома; (б) осцилляции проводимости в зависимости от напряженности приложенного магнитного поля при эффекте Аронова — Бома (по данным работы [2])

чаем очень симметричную структуру, удобную для математического описания и расчета.

Импульс электрона $\vec{p} = \hbar\vec{k}$ в магнитном поле \vec{B} следует заменить величиной $\vec{p} + e\vec{A}$, где \vec{A} — вектор-потенциал поля ($\vec{B} = \text{rot } \vec{A}$). Из квантовой механики известно, что изменение фазы волны электрона при движении от P к Q на рис. 7.3, а определяется соотношением

$$\vartheta(\vec{r}) = \frac{e}{\hbar} \int_P^Q \vec{A} \cdot d\vec{s}, \quad (7.10)$$

где интегрирование ведется вдоль линии, соединяющей точки P и Q . Поэтому разность фаз, возникающая при движении электронов по верхней или нижней траектории на рис. 7.3, a , равна

$$\Delta\vartheta = \vartheta_1 - \vartheta_2 = \frac{e}{\hbar} \left[\int_{\text{нижний маршрут}} \vec{A} \cdot d\vec{s} - \int_{\text{верхний маршрут}} \vec{A} \cdot d\vec{s} \right] = \frac{e}{\hbar} \int_{\text{по контуру}} \vec{A} \cdot d\vec{s}, \quad (7.11)$$

так как электронные волны распространяются по этим траекториям в противоположных направлениях по отношению к полю \vec{A} . Применяя к выражению (7.11) теорему Стокса из векторного анализа, получаем

$$\Delta\vartheta = \frac{e}{\hbar} \Phi, \quad (7.12)$$

где

$$\Phi = \int_{\text{по площади}} (\text{curl} \vec{A}) \cdot d\vec{S} = \int_{\text{по контуру}} \vec{B} \cdot d\vec{S}. \quad (7.13)$$

Используя величину Φ_0

$$\Phi_0 = \frac{h}{e} \quad (7.14)$$

в качестве *кванта магнитного потока*, из уравнения (7.12) можно получить для разности фаз $\Delta\vartheta$ выражение

$$\Delta\vartheta = \vartheta_1 - \vartheta_2 = \frac{2\pi\Phi}{\Phi_0}. \quad (7.15)$$

Интенсивность интерференции таких волн $\psi_i \propto \exp(i\vartheta_i)$ пропорциональна амплитуде вероятности, определяемой уравнением

$$P = (\psi_1 + \psi_2)^2 \propto \cos(\vartheta_1 - \vartheta_2) = \cos 2\pi \frac{\Phi}{\Phi_0}, \quad (7.16)$$

что и позволяет получать и регистрировать *интерференционные эффекты* за счет изменения величины потока Φ .

Из сказанного выше следует, что в описываемой системе можно наблюдать полное колебание в тех случаях, когда изменение магнитного потока Φ внутри структуры совпадает с од-



ним квантом магнитного потока Φ_0 . Поскольку сечение потока остается фиксированным, мы можем заключить из выражения (7.13), что изменение \vec{B} будет приводить к колебаниям экспериментально измеряемых физических величин, например проводимости. В качестве примера можно указать осцилляции проводимости, полученные еще в 1994 г. в работе Форда и соавт. [2], которые приводятся на рис. 7.3, б.

Другой очень интересный аспект описываемого эффекта Аронова и Бома связан с тем, что изменения фазы могут быть вызваны изменением величины поля \vec{B} даже в тех случаях, когда электронные волны не подвергаются непосредственному воздействию поля. Такая возможность долгое время даже отвергалась некоторыми исследователями в силу внутренней противоречивости, но позднее было установлено, что такой эффект действительно существует. Дело в том, что (в отличие от магнитного поля) вектор-потенциал A существует также вне показанного на рис. 7.3, а кольца и может участвовать в процессах изменения фазы в соответствии с уравнением (7.11). Другое важное обстоятельство, связанное с квантовым эффектом Аронова — Бома, заключается в том, что этот эффект часто наблюдается даже в образцах микрометрового размера.

7.6. Эффект Шубникова — де Гааза

В этом разделе мы продолжим изучение воздействия магнитного поля на электронные и транспортные характеристики двумерных систем. В предыдущем разделе было показано, что изменение напряженности \vec{B} приводит к изменениям энергии и вырождения уровней Ландау, что, в свою очередь, может существенно менять характеристики процессов переноса в материалах вообще и в двумерных системах в частности. В большинстве случаев эксперименты проводятся в условиях постоянной концентрации двумерных электронов n_{2D} при изменении напряженности магнитного поля. С ростом напряженности поля в энергия уровней Ландау повышается, поскольку увеличивается расстояние $\hbar\omega_c$ между ними. Одновременно возрастает и степень вырождения уровней D , определяемая уравнением (7.9). Представляется удобным ввести коэффициент заполнения ν уровня, определяемый как

$$\nu = \frac{n_{2D}}{D} = \frac{2\pi\hbar n_{2D}}{eB}, \quad (7.17)$$

т. е. как отношение плотности электронов к степени вырождения каждого из уровней. Обычно коэффициент ν является дробным, но при $T = 0$ максимальное целое число, меньшее ν , равно количеству N полностью заполненных уровней Ландау. Очевидно, что в общем случае верхний уровень Ландау заполнен частично, но, если коэффициент заполнения равен целому числу, все уровни Ландау полностью заполнены. Поэтому значения магнитного поля B_N , при котором обеспечивается полное заполнение уровней Ландау ($\nu = N$), могут быть определены из уравнения (7.17) в виде соотношения

$$B_N = \frac{1}{N} \frac{2\pi\hbar n_{2D}}{e}, \quad (N = 1, 2, \dots). \quad (7.18)$$

Рассмотрим далее зависимость уровня Ферми E_F от B при $T = 0$. Качественно можно сразу сказать, что при целочисленных коэффициентах заполнения (т. е. при $\nu = N$) энергия E_F должна попадать в запрещенную зону между уровнями Ландау (см. рис. 7.2) в области, где $n_{2D} \approx 0$, а напряженность поля B_N при этом определяется уравнением (7.18). Более того, поскольку условие $n_{2D} \approx 0$ выполняется в достаточно большом интервале, небольшие изменения в энергии не должны никак влиять на величину n_{2D} . В таких системах электрическая проводимость образца должна быть очень малой, поскольку в них вообще отсутствуют носители в окрестности уровня Ферми E_F . В противоположность этому, если ν отличается от целого числа на плюс или минус $1/2$, значения n_{2D} достигают максимумов (см. рис. 7.2), значения E_F должны совпадать с пиками, и очень небольшие изменения в энергии должны приводить к весьма существенным изменениям n_{2D} и электропроводности образца в целом. На рис. 7.4 представлена осциллирующая зависимость разности потенциалов U_{pp} между двумя зондами, расположенными вдоль образца, как показано на врезке рис. 7.4, от напряжения на затворе. Очевидно, что это падение напряжения пропорционально сопротивлению ρ_{xx} вдоль образца. Используемая геометрия образца позволяла также измерять сопротивление ρ_{xy} в поперечном направлении, которое пропорционально напряжению Холла U_H . Представленные на рис. 7.4 данные относятся к измерениям в магнитном поле $B = 18$ Тл, направленном поперечно двумерной структуре при температуре $T = 1,5$ К. Этот рисунок относится к экспериментам, проведен-

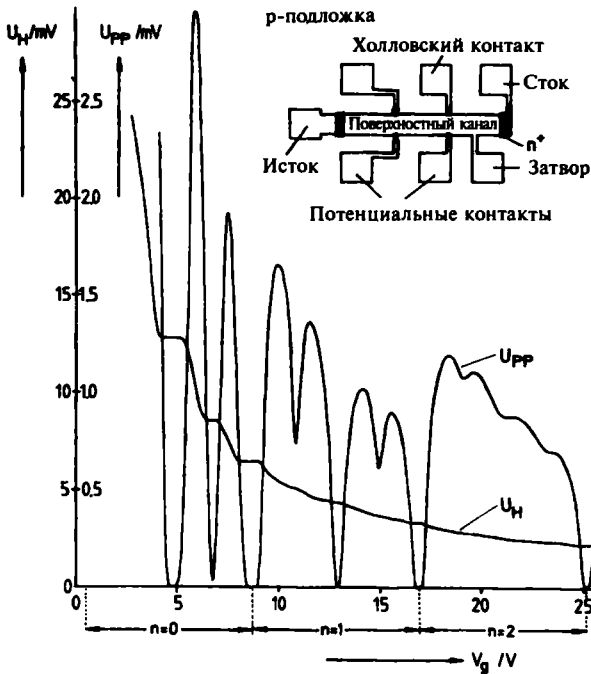


Рис. 7.4. Зависимость напряжения Холла U_H и напряжения в продольном направлении образца (V_{pp}) от напряжения на затворе для двумерной системы в канале полевого МОП-транзистора (при $T = 1,5$ К) в магнитном поле напряженностью 18 Тл. Эти результаты работы [3] в свое время привели к открытию квантового эффекта Холла

ным фон Клитцингом и его сотрудниками [3] в 1980 г., которые привели к обнаружению квантового эффекта Холла.

Давайте рассмотрим более подробно поведение зависимостей V_{pp} или ρ_{xx} на рис. 7.4. Из сказанного ясно, что колебания напряжения являются следствием описываемого эффекта Шубникова — де Гааза, который возникает вследствие образования уровней Ландау для двумерных систем электронов, подвергаемых воздействию магнитного поля. Очевидно, что существуют и осцилляции проводимости, связанные с осцилляциями продольного сопротивления. В экспериментах описываемого типа коэффициент заполнения уровней Ландау изменяется при положительном напряжении на затворе, которое управляет концентрацией электронов на поверхности раздела Si—SiO₂. Стоит

отметить также, что осцилляции Шубникова — де Гааза в двумерных системах зависят лишь от компоненты \vec{B} , перпендикулярной поверхности раздела. Ранее такие колебания были зарегистрированы и в объемных проводниках, но они были очень слабыми и зависели одновременно от обоих компонент \vec{B} , лежащих в плоскости и перпендикулярных ей.

7.7. Квантовый эффект Холла

7.7.1. Экспериментальные данные и элементарная теория целочисленного квантового эффекта Холла (IQHE)

Одним из наиболее неожиданных и труднообъяснимых результатов измерений, представленных на рис. 7.4, является квантование напряжения Холла V_H (на рисунке обозначено через U_H). Обычно, описывая эти эксперименты, авторы говорят о холловском сопротивлении R_H , однако другие авторы предпочитают термины «удельное поперечное сопротивление» или «удельное сопротивление Холла» ρ_{xy} . Поскольку «ступеньки» в изменении напряжения не зависят от формы образца, терминологическая разница между сопротивлением и удельным сопротивлением не столь важна. Нижние индексы в выражениях демонстрируют тот факт, что разница в напряжениях измеряется в направлении (y), перпендикулярном направлению электрического тока (x), а продольное удельное сопротивление ρ_{xx} получается при измерении напряжения в направлении вдоль тока.

Квантование значений сопротивления Холла с очень высокой точностью определяется уравнением

$$R_H = \frac{h}{e^2} \frac{1}{n} = 25812.807 \Omega \left(\frac{1}{n} \right), \quad n = 1, 2, \dots, \quad (7.19)$$

причем точность регистрации значений R_H связана с тем, что экспериментально получаемые кривые на рис. 7.4 имеют широкие «плато» и остаются постоянными в широком диапазоне напряжений на затворе, т. е. не зависят от концентрации электронов в описываемых двумерных электронных системах (ниже будет показано, что подобные «плато» появляются и при измерении электрических характеристик в зависимости от напряженности магнитного поля). Интересно отметить, что эти «плато» на зави-

симости ρ_{xy} проявляются точно при значениях ρ_{xx} , равных нулю, что легко заметить на рис. 7.4. Особо следует подчеркнуть, что сопротивление Холла R_H нельзя путать с коэффициентом Холла, для которого используется такая же терминология. В настоящее время R_H обычно называют постоянной фон Клитцинга (при $n = 1$ в уравнении (7.19) и обозначают как R_K . Значение R_K может быть сейчас измерено с очень высокой точностью (до 10^{-9} и выше), вследствие чего эту величину уже стали использовать в качестве нового стандарта в метрологии (см. раздел 7.7.4).

Наиболее простая теория, объясняющая квантование значений энергии сопротивления Холла R_H в уравнении (7.19), основана на представлениях о положении уровней Ландау и их вырождении в сочетании с классической теорией процессов переноса для заряженных частиц. Действительно, в определении R_H мы можем заменить напряжение Холла V_H на величину bVv (где b — ширина образца, v — дрейфовая скорость носителей), а значение тока I на величину $bn_{2D}ev$. После такой подстановки выражение для квантованных значений сопротивления приобретает вид

$$R_H = \frac{V_H}{I} = \frac{B}{en_{2D}}, \quad (7.20)$$

из которого следует, что при заданном распределении и полностью заполненном последнем уровне Ландау и следующем абсолютно пустом, никакого рассеяния электронов в системе быть не может. Подставляя в уравнение (7.20) выражения (7.18) для магнитного поля B_N при полностью заполненных уровнях, легко получить для R_H формулу

$$R_H = \frac{1}{n} \frac{h}{e^2}, \quad n = 1, 2, 3, \dots \quad (7.21)$$

Отметим, обратная к R_H величина

$$(R_H)^{-1} = \frac{e^2}{h} n \quad (7.22)$$

соответствует значениям, полученным для квантованной проводимости (раздел 6.4.1).

Интересное замечание можно сделать относительно того, что квантованные значения R_H соответствуют нулевым значе-

ниям продольной проводимости, показанным на рис. 7.4. Такое соответствие связано с тем, что эффект Шубникова — де Гааза в разделе 7.6 также объяснялся заполнением уровней Ландау. Результаты, рассмотренные в данном разделе иллюстрируют целочисленный квантовый эффект Холла (IQHE).

7.7.2. Краевые состояния и IQHE

Целочисленный квантовый эффект Холла (IQHE) был также объяснен Бюттикером в конце 80-х годов в рамках формализма теории Ландауэра — Бюттикера [4] для многозондового квантового переноса, описанного в разделе 6.4.2. Мы можем вновь рассмотреть различные уровни Ландау для двумерных электронных систем, соответствующих классическим циклотронным орбитам, возникающим при наложении перпендикулярно направленного внешнего магнитного поля. Для простоты представим удлиненный образец с электрическими контактами со всех сторон. Движение по возникающим при этом циклотронным орбитам, как показано на рис. 7.5, *а*, происходит против часовой стрелки. Читатель должен учитывать, что предлагаемые теории являются очень качественными, а более точные описания явлений можно найти лишь в более специальных работах [5]. Замкнутые циклотронные орбиты, которые в среднем вообще не соответствуют протеканию тока, не могут существовать вблизи краев рассматриваемого образца. Поэтому, как показано на рис. 7.5, *а*, электроны на таких краях двигаются по так называемым *открытым орбитам*, обладающим некоторой эффективной дрейфовой скоростью и, следовательно, создающих некоторый «краевой» электрический ток. В квантовой механике состояния, связанные с такими орбитами, называют краевыми состояниями. Верхним краевым состоянием на рис. 7.5, *а* соответствует положительное направление скорости, а нижним — отрицательное. Такие краевые состояния, распространяющиеся в противоположных направлениях, обычно представляют собой компактные линии краевых токов, как бы локализованных в некотором квантовом волноводе (рис. 7.5, *б*).

В разделе 6.4 было показано, что квантование проводимости (а следовательно, и сопротивления) может быть получено непосредственно и просто из формализма теории Ландауэра — Бюттикера для процессов квантового переноса. Аналогично может быть построена и теория краевых каналов проводимости

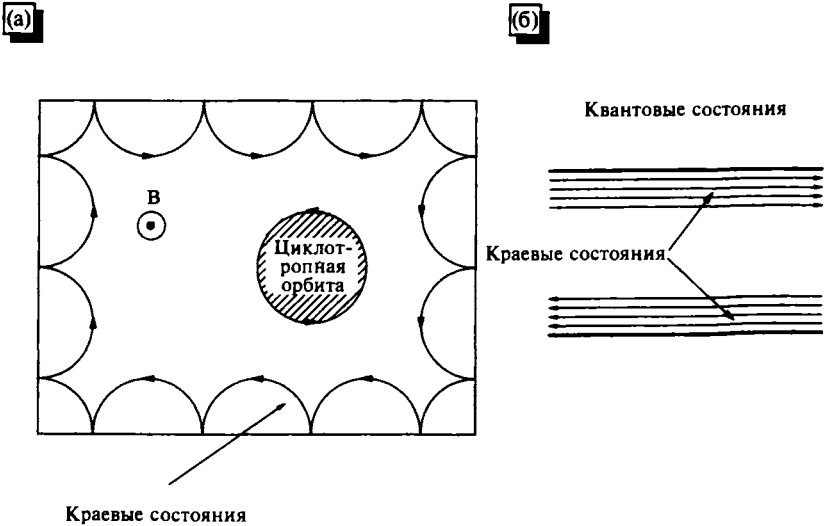


Рис. 7.5. (а) Открытые орбиты, соответствующие краевым состояниям и циклотронные орбиты для 2D электронной системы под воздействием магнитного поля; (б) схематичное представление краевых токов

в рассматриваемых системах. Для оценки величины тока, протекающего через краевой или пограничный канал, мы можем рассмотреть образец в виде стержня, типа тех, которые использовались ранее при обсуждении целочисленного квантового эффекта Холла (IQHE) и выводе общего уравнения (6.17) в главе 6. Для удобства обсуждения мы выпишем это уравнение еще раз:

$$I_i = \frac{2e^2}{h} \left[(N_i - R_i) V_i - \sum_{j \neq i} T_{ij} V_j \right], \quad (7.23)$$

а затем будем просто следовать тому подходу, который был применен Ландауэром и Бюттикером. Предположим, что краевые токи представлены большим числом каналов N (хотя на рис. 7.6 показаны только два канала). Ток может втекать в образец или вытекать из него лишь через контакты 1 и 4, при этом возникает напряжение Холла между контактами 6 и 2 (или между контактами 5 и 3). Продольное электрическое сопротивление образца (см. раздел 7.7.1) может измеряться между контактами 5 и 6 или 3 и 2. Как показано для краевых каналов на рис. 7.6, ток

из контакта 1 поступает в контакт 6, однако, поскольку последний — потенциальный контакт, ток через него течь не может, т. е. $I_6 = 0$. Аналогичные рассуждения применимы к другим потенциальным контактам (2, 3 и 5). Применим уравнение (7.23) для описания токов и предположим, что контакты идеальны, т. е. в них отсутствует отражение токов ($R_i = 0$). На рис. 7.6 показано, что N состояний распространяются от контакта 1 к контакту 2 (т. е. $T_{21} = N$), однако никакие состояния не передаются из контакта 2 в контакт 1 (т. е. $T_{12} = 0$). Очевидно, что соответствующие соображения относятся и к остальным близлежащим контактам, вследствие чего имеем $T_{12} = T_{43} = T_{54} = T_{65} = T_{16} = N$ и $T_{23} = T_{34} = T_{45} = T_{56} = T_{61} = 0$. Все остальные величины T_{ij} равны нулю, поскольку токи не могут «перескакивать» контакты, как и показано на рис. 7.6. С учетом сказанного уравнение (7.23) можно переписать в более компактной форме, используя матричную запись:

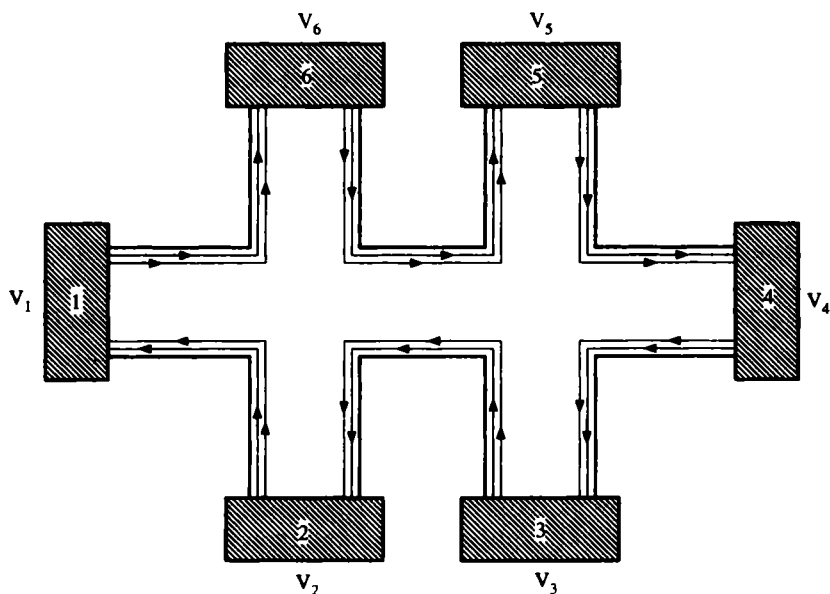


Рис. 7.6. Двумерный образец для измерений квантового эффекта Холла. Ток протекает от контакта 1 к контакту 4. Напряжение Холла измеряется контактами 6 и 2 (или 5 и 3). Падение напряжения в направлении вдоль тока измеряется с помощью контактов 5 и 6 (или 3 и 2). На рисунке стрелками показаны также два из «краевых» токов

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \end{bmatrix} = \begin{bmatrix} I_1 \\ 0 \\ 0 \\ I \\ 0 \\ 0 \end{bmatrix} = \frac{Ne^2}{h} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \end{bmatrix}. \quad (7.24)$$

Мы должны написать, что $I_1 = -I$ и $I_4 = I$, поскольку стрелки на рисунке указывают направление движения электронов. Из выписанной системы уравнений можно сделать ряд интересных заключений. Из уравнений для I_2 и I_3 можно получить равенства $V_2 = V_3 = V_4$ и $V_1 = V_5 = V_6$, соответственно, т. е. все контакты вдоль верхней части системы имеют потенциал истока, а все контакты вдоль нижней части — потенциал стока.

Для расчета сопротивления Холла мы должны разделить напряжение между контактами 6 и 2 на значение тока между контактами 1 и 4, что дает уравнение

$$R_{14,62} = \frac{V_6 - V_2}{I}, \quad (7.25)$$

подставляя в которое значения V_6 и V_2 из уравнения (7.24) можно получить искомое сопротивление Холла в виде

$$R_H = \frac{h}{e^2} \frac{1}{N}. \quad (7.26)$$

Используя этот же метод для расчета продольного сопротивления, можно найти разность напряжений между датчиками 5 и 6, которая, как и следует ожидать, равна нулю. Полученные результаты показывают, что квантование поперечного сопротивления (а следовательно, и поперечной проводимости) согласуется с квантованием проводимости, рассмотренным выше в разделе 6.4.1.

7.7.3. Протяженные и локализованные состояния

Предложенные модели целочисленного квантового эффекта Холла (IQHE) позволяют достаточно точно предсказывать квантование значений сопротивления Холла (параметр ρ_{xy}), а также нулевые значения продольного сопротивления ρ_{xx} , однако они

не дают объяснения наличию упоминавшихся выше «плато» Холла, т. е. постоянства напряжения или сопротивления Холла на некоторых, достаточно больших диапазонах изменения магнитного поля. Существование «плато» Холла конечных размеров может быть объяснено в терминах локализации электронных состояний как следствие некоторой неупорядоченности. В разделе 7.4 мы уже упоминали нарушения в структуре в качестве фактора, вызывающего уширение уровней Ландау. Из курсов физики твердого тела известно, что помимо так называемых блоховских *протяженных состояний* существуют и *локализованные состояния*, возникающие в неупорядоченных структурах, что было объяснено позднее *теорией локализации Андерсона* в 1958 г. [6]. Такие состояния могут быть охарактеризованы длиной локализации α , введенной в разделе 1.3. В соответствии с теорией, при достаточной степени разупорядочения полупроводника, все состояния в трехмерной системе могут стать локализованными. Частичное разупорядочение кристалла может возникать из-за точечных дефектов (типа атомов примеси), из-за структурной неоднородности, приводящей к флуктуациям электростатических ионных потенциалов или из-за процессов рассеяния на шероховатых поверхностях раздела (см. раздел 6.2.1) в случае гетеропереходов.

На рис. 7.7 показана функция плотности состояний, для которой δ -функции несколько уширились за счет процессов, описанных в разделе 7.4 и на рис. 7.2. На рисунке мы видим блоховские протяженные состояния вблизи пиков с энергией $E = (n + S)\hbar\omega_c$, соответствующих подвижным электронам и локализованным непроводящим состояниям. На рисунке указаны также «границы подвижности», разделяющие протяженные и локализованные состояния. В таких системах, как было показано в 1981 г. Лафлином [7], уровни Ландау при заданной концентрации не заполняются. Если уровень Ферми материала попадает в область локализованных состояний, то эти состояния не принимают участия в проводимости и проводимость осуществляется лишь за счет протяженных состояний. Теоретические расчеты показывают, что протяженные состояния должны компенсировать эти потери и обеспечить постоянство значений сопротивления, как показано на рис. 7.4. Некоторый дополнительный ток при этом возникает вследствие ускорения, связанного с рассеивающими потенциалами в неупорядоченных системах. С этой точки зрения интересно отметить, что,

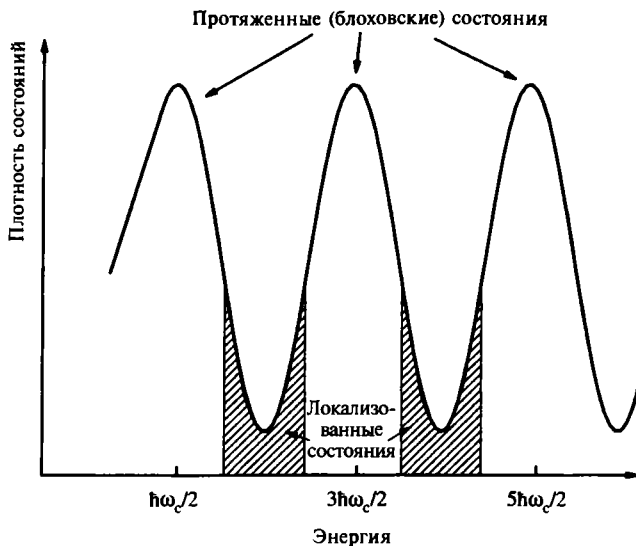


Рис. 7.7. Схематическое представление протяженных и локализованных состояний для функции плотности состояний двумерной электронной системы в магнитном поле

хотя целочисленный квантовый эффект Холла (IQHE) и был обнаружен более двадцати лет назад, его теоретическое описание все еще остается областью интенсивных разработок.

7.7.4. Использование квантового эффекта Холла (IQHE) в метрологии

Главной особенностью и преимуществом современной Международной системы единиц (СИ) является то, что она основана на фундаментальных константах, т. е. некоторых постоянных, не зависящих от условий измерения и воспроизводимых при любых обстоятельствах. В качестве примера можно привести единицу времени (секунду), которая может быть определена в единицах периода излучения радиоактивных атомов цезия.

Одним из наиболее важных применений квантового эффекта Холла стала метрология — наука о единицах измерения. В частности, начиная с 1990-х гг. в качестве эталона или стандарта электрического сопротивления стали использоваться измерения, относящиеся к описанным «плато» Холла для попе-

речного сопротивления ρ_{xy} . Как указывалось в разделе 7.7.1, эта величина, получившая название постоянной фон Клитцинга, может быть определена из уравнения (7.19) при $n = 1$ и равна

$$R_K = \frac{h}{e^2} = 25812.807572\Omega. \quad (7.27)$$

Именно эта строго заданная уравнением (7.19) величина сопротивления была рекомендована в качестве единицы сопротивления Международным консультативным комитетом после рассмотрения наилучших экспериментальных данных к 1998 г. Эта величина была измерена на три порядка с большей точностью по сравнению с ее первым использованием в качестве стандарта в 1990 г. Точность измерений соответствует нескольким долям на 10^{10} (т. е. достигает порядка 10^{-9}), что по меньшей мере, на два порядка выше точности, получаемой при определении значения единицы *ом* на стандартных эталонах электрического сопротивления.

Такая впечатляющая точность определения значения R_K связана с несколькими обстоятельствами. Во-первых, она основана на существовании самих «плато» Холла, величина которых остается поразительно постоянной при изменениях как магнитного поля, так и напряжения затвора в сравнительно широких диапазонах (что можно сразу отметить по данным рис. 7.4, полученным для кремниевого полевого МОП-транзистора). Во-вторых, видно, что значение R_K ($\equiv h/e^2$) не зависит от свойств конкретного вещества, хотя сама величина была реально измерена лишь для двух типов образцов (кремниевые полевые МОП-структуры и гетеропереходы на основе AlGaAs/GaAs). В-третьих, значение R_K не зависит от геометрии образца (например, его ширины и т. п.). Уже проверено, что всеми эффектами, связанными с размерами используемых в метрологии эталонов, в рассматриваемом случае можно пренебречь. С другой стороны, при любых измерениях (включая метрологические) необходимо тщательно следить за качеством электрических контактов и плотностью используемых токов (поскольку при превышении критического значения плотности тока квантовый эффект Холла перестает проявляться вообще). Изучение условий подавления целочисленного квантового эффекта Холла при больших плотностях тока представляет собой область специальных теоретических и экспериментальных исследований.

Одной из самых известных констант физики выступает так называемая постоянная тонкой структуры α , определяемая соотношением

$$\alpha = \frac{e^2}{2\varepsilon_0 ch}, \quad (7.28)$$

представляющая собой уникальный набор фундаментальных констант, в который входят заряд электрона, скорость света, постоянная Планка и диэлектрическая проницаемость вакуума. Уравнения (7.27) и (7.28) позволяют связать постоянную фон Клитцинга R_k с постоянной тонкой структуры соотношением

$$R_k = \frac{h}{e^2} = \frac{1}{2\varepsilon_0 c \alpha}, \quad (7.29)$$

в результате чего столь важная (для физики вообще) постоянная тонкой структуры α может быть определена в экспериментах с использованием целочисленного квантового эффекта Холла. Более того, значения α , получаемые с использованием измерений R_k , по точности сравнимы или даже превосходят теоретически вычисляемое (на основе квантовой электродинамики) значение α , точность которого составляет около 10^{-9} . Стоит также упомянуть, что точность определения α в экспериментах с целочисленным квантовым эффектом Холла оказывается выше, чем в экспериментах по аномальному магнитному моменту электронов.

7.7.5. Дробный квантовый эффект Холла (FQHE)

Из материала предыдущего раздела ясно, что для наблюдений целочисленного квантового эффекта Холла (IQHE) вовсе не требуется, чтобы кристаллические образцы обладали сверхвысокой чистотой, и более того, наличие некоторых нарушений кристаллической структуры даже представляется необходимым, например, для возникновения «локализованных состояний» (раздел 7.7.3) или объяснения «плато» в холловском напряжении. Через два года после обнаружения целочисленного квантового эффекта Холла в 1982 г. Тсуи, Штермер и Госсард [9] открыли так называемый дробный квантовый эффект Холла

(FQHE). В этом эффекте, как показано на рис. 7.8, обнаруживаются новые «плато», связанные с ρ_{xy} , а также с нулевыми значениями ρ_{xx} , причем они соответствуют нецелочисленным (дробным) значениям коэффициента заполнения ν , определяемого уравнением (7.17). Наблюдаемые дробные значения имеют вид $\nu = p/q$, где p и q — целые числа, причем q — нечетное. Дробный квантовый эффект наблюдался на двумерных системах, образованных гетеропереходами в очень чистых полупроводниковых образцах (т. е. при очень высокой подвижности электронов) и при низких концентрациях двумерных электронов. Кроме этого, для наблюдения эффекта требуются очень низкие температуры (близкие к 0 К), а прилагаемое магнитное поле должно быть достаточно сильным, чтобы в системе существовал только один уровень Ландау. В этих условиях коэффициент заполнения равен или меньше единицы и все электроны имеют

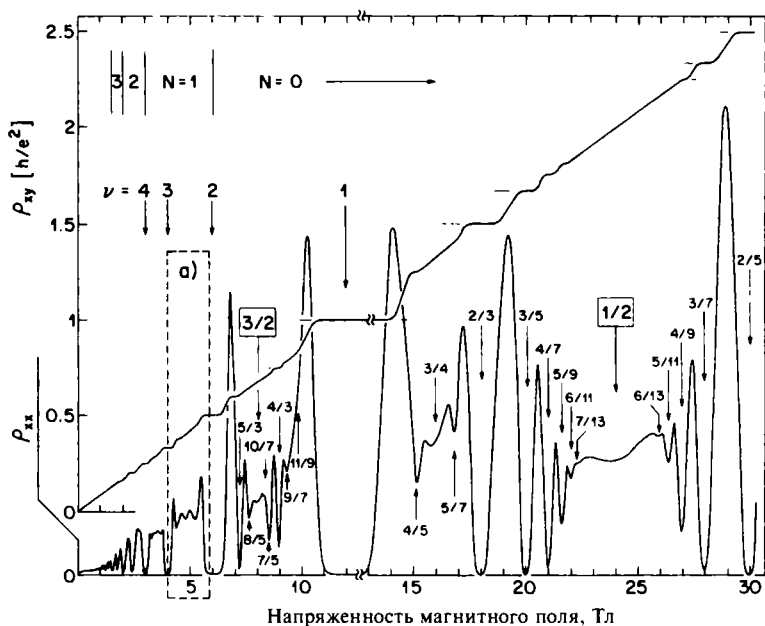


Рис. 7.8. Зависимость продольного сопротивления ρ_{xx} и дробного квантового сопротивления Холла ρ_{xy} от напряженности магнитного поля для двумерной электронной системы в модулированно-легированных гетеропереходах AlGaAs/GaAs (при $T = 0,085$ К).

одинаковую энергию, совпадающую с минимальным значением, определяемым уравнением (7.8) и равным $h\omega_c/2$. Как видно из рис. 7.8, наиболее сильно эффект проявляется при значениях коэффициента заполнения $\nu = 1/3, 2/3, 2/5, 3/5$ и т. д.

Из приведенных экспериментальных данных следует, что дробный квантовый эффект Холла (FQHE) существенно отличается от целочисленного эффекта Холла. Поскольку последний хорошо описывался моделью на основе представления о невзаимодействующих электронах, разумно было бы предположить, что описание дробного эффекта должно каким-то образом учитывать взаимодействие электронов, т. е. сводиться к некоторому многочастичному взаимодействию. Действительно, вскоре после обнаружения FQHE возникло предположение, что при этом эффекте электронный газ конденсируется в систему, называемую *вигнеровским кристаллом*, в котором минимизируется энергия, обусловленная кулоновским отталкиванием. Однако вскоре в результате теоретических вычислений выяснилось, что значения полной энергии электронной системы с небольшим числом электронов в сильных магнитных полях может быть ниже энергии вигнеровского кристалла. Действительно, выше было показано, что при очень высоких значениях в уровни Ландау в двумерной электронной системе «схлопываются» в одно, основное состояние.

В 1983 г., вскоре после обнаружения FQHE, Лафлин [10] выдвинул объясняющую этот эффект теорию, основанную на учете многочастичного кулоновского взаимодействия между электронами при очень высоких значениях напряженности магнитного поля и низких температурах. В соответствии с этой теорией электронный газ конденсируется в коррелированную квантовую жидкость, в которой кулоновское взаимодействие разделяет электроны друг от друга. Лафлин вычислил волновую функцию основного состояния такой многочастичной электронной системы и нашел, что основное состояние в ней отделено от возбужденных состояний энергетической щелью. Теория Лафлина позволила также предсказать, что коррелированное основное состояние формируется именно тогда, когда уровни Ландау заполнены лишь частично (например, с коэффициентом $\nu = 1/3$). Кроме этого, из теории следовало, что элементарные возбуждения в такой системе соответствуют частицам с дробным электрическим зарядом, равным $1/3$ заряда электрона. Следует особо отметить, что самые современные экспери-

менты, основанные на измерении электрических шумов в токах, протекающих через точечные квантовые контакты, похоже, подтверждают возможность существования *дробных значений электронного заряда*.

Одной из наиболее привлекательных теорий, выдвинутых для объяснения дробного квантового эффекта, является модель так называемых *составных фермионов (CF)*, введения в 1989 г. Джейном [11], а позднее развитая Гальпериным и соавт. [12], в которой считается, что каждый электрон связан с двумя трубками магнитного потока, равного кванту Φ_0 . Эти трубки читатель может представить себе, например, в виде бесконечно малых соленоидов, несущих квант потока Φ_0 , определяемый уравнением (7.14). Кстати, сам термин «составной фермион» и связанные с этим представления возникли именно из-за того, что такое образование в целом ведет себя подобно фермионам. Более подробное изучение данных рис. 7.8 позволяет заметить, что вблизи значений коэффициента заполнения $\nu = S$ продольное сопротивление ρ_{xx} можно назвать «нормальным», т. е. система ведет себя так, как будто внешнее поле B отсутствует. Показано, что при $\nu = S$ и соответствующем значении $B = B_0$ поведение составного фермиона (CF) действительно соответствует отсутствию поля, хотя при значениях B в чуть больших или меньших B_0 составной фермион реагирует на прикладываемое поле совершенно обычным образом. Основываясь на описанной модели, Джейн сумел теоретически описать большинство наблюдений, связанных с наборами дробных значений $\nu = p/q$ при регистрации нецелочисленного квантового эффекта Холла.

Литература

1. *Webb R.A., Washburn S., Umbach C.R., Laibowitz R.G.* (1965). Phys. Rev. Lett, 54, 2696.
2. *Ford C.J.B., Simpson P.J., Zeiber L., Franklin J.D.F., Barnes C.H.W., Frost J.E.F., Ritchie D.A., Pepper M.* (1994). Journal of Physics: Condensed Matter, 6, L725.
3. *von Klitzing K., Dorda G., Pepper M.* (1980). Phys. Rev. Lett., 45, 494.
4. *Büttiker M.* (1988). Phys. Rev. B, 38, 9375.
5. The Quantum Hall Effect (1990). Eds. *Prange R.E., Girvin S.M.* (Springer-Verlag, New York).

6. *Anderson P.W.* (1958). *Phys. Rev.*, 109, 1492.
7. *Laughlin R.B.* (1981). *Phys. Rev. B*, 23, 5632.
8. *Jeckelmann B., Jeanneret B.* (2001). *Rep. Prog. Phys.*, 64, 1603.
9. *Willet R., Eisenstein J.R., Stormer H.L., Tsui D.C., Hwang J.C.M., Gossard A.C.* (1987). *Phys. Rev. Lett.*, 59, 1776.
10. *Laughlin R.B.* (1983). *Phys. Rev. Lett.*, 50, 1395.
11. *Jain J.K.* (1990). *Phys. Rev. B*, 41, 7653.
12. *Halperin B.I., Lee P.A., Read N.* (1993). *Phys. Rev. B*, 47, 7312.
13. *von Klitzing K.* (1984). *Physica*, 126, B-C, 242.

Дополнительная литература

- Ferry D.K., Goodnick S.M.* (1997). *Transport in Nanostructures* (Cambridge University Press, Cambridge).
- Hamaguchi Ch.* (2001). *Basic Semiconductor Physics* (Springer, Berlin).
- Imry Y.* (1977). *Introduction to Mesoscopic Physics* (Oxford University Press, Oxford).
- Murayama Y.* (2001). *Mesoscopic Systems* (Wiley-VCH, Weinheim).
- Schäfer W., Wegener M.* (2002). *Semiconductor Optics and Transport Phenomena* (Springer, Berlin).

Задачи и упражнения

1. Сходство эффекта Аронова — Бома с интерференционными явлениями в СКВИДАХ (полупроводниковых квантовых интерференционных устройствах). Обсудите аналогии в работе квантовых интерференционных устройств (СКВИД) и эффекта Аронова — Бома (раздел 7.5). (Подсказка: напомним, что критическое значение тока в СКВИДах модулируется квантовым потоком Φ_0 , создающим интерференционные картины (интерферограммы) Фраунгофера.) Исходя из значения Φ_0 и заданной площади поверхности СКВИД ($\sim 1 \text{ см}^2$), сравните точность определения изменений напряженности поля B (в единицах тесла, Тл) по числу циклов тока с точностью измерений сопротивления, описанных в разделе 7.7.4. Обсудите аналогии между когерентностью фаз макроскопической волновой функции в полупроводнике и когерентностью фаз, рассмотренной в разделах 7.5 и 1.4.

- Осцилляции Шубникова — де Гааза. Покажите, что период $1/B$ осцилляций Шубникова — де Гааза в двумерных электронных системах определяется соотношением $\Delta(1/B) = 2e/hN_s$, где N_s — концентрация двумерных электронов. Исходя из представленных на рис. 7.4 данных по целочисленному квантовому эффекту Холла, оцените приблизительно двумерную плотность носителей в системе.
- Целочисленный квантовый эффект Холла (IQHE). Рассмотрите значения «плато» ρ_T на рис. 7.4 и покажите, что они согласуются с уравнением (7.19), т. е.

$$R_H = \frac{h}{e^2} \frac{1}{n} = 25812.807 \Omega \left(\frac{1}{n} \right), \quad n = 1, 2, \dots$$

Воспользовавшись зависимостью ρ_L от B , полученной в работе [13], вычислите двумерную плотность электронов. (Подсказка: рассмотрите область малых значений B , поскольку при больших значениях заселенными могут оказаться многие подзоны.)

- Зависимость уровня Ферми от B и целочисленный квантовый эффект Холла (IQHE). Магнитное поле B приложено перпендикулярно плоскому гетеропереходу, подобно схеме измерения IQHE. Предположим, что концентрация электронов n_{2D} составляет 10^{12} см^{-2} , а температура настолько мала, что выполняется сильное неравенство $kT \ll \hbar\omega_c$, где $\omega_c = eB/m_e^*$ — циклотронная частота. Найдите зависимость энергии Ферми E_F от B (обозначим через E_F^0 уровень Ферми при $v = 0$). (Подсказка: при очень низких температурах, когда полностью заполнены n уровней Ландау, уровень E_F лежит в «пустой» области энергий, между уровнями n и $n + 1$ или на уровне $n + 1$, если это состояние заполнено частично. Значение уровня E_F линейно возрастает с ростом напряженности поля до тех пор, пока оно не достигнет величины, при которой незаполненным оказывается предыдущий уровень (вырождение уровней возрастает с увеличением напряженности поля, а коэффициент заполнения падает), в результате чего величина E_F попадает на предыдущий по порядку уровень, т. е. происходят осцилляции E_F относительно значения E_F^0 . Отметим, что такое поведение E_F действительно приводит к колебаниям продольного сопротивления ρ_{xx} . Оцените максимальную температуру, при которой невозможно будет регистрировать «ступени» в значениях ρ_T на рис. 7.4.

5. Двумерная электронная система в очень сильных магнитных полях. Используя формулы, полученные при решении задачи 1.5 для тензорных коэффициентов сопротивления и проводимости, а также для компонент скорости электронов и R_H , полученные при решении задачи 1.6, покажите, что при очень высоких значениях магнитного поля (необходимых для наблюдения целочисленного квантового эффекта Холла) должны выполняться условия: $\rho_{xx} \approx \sigma_{xx}/\sigma_{xy}^2$ и $\rho_{xy} \approx 1/\sigma_{xy} = R_H/B$. Отметим, что в результате величина ρ_{xx} оказывается пропорциональной σ_{xx} . Попробуйте объяснить этот несколько неожиданный результат.
6. Спиновое расщепление уровней Ландау в двумерной электронной системе. Покажите, что для каждой из энергий уровней Ландау, определенных в разделе 7.2, необходимо ввести множитель, связанный с расщеплением по спину, в результате чего уровни Ландау реально должны описываться уравнением

$$E_n = \frac{\hbar e B}{m_e^*} \left(n + \frac{1}{2} \right) \pm \frac{1}{2} g^* \mu_B B, \quad n = 0, 1, 2, \dots,$$

где коэффициент g^* представляет собой эффективный g -фактор, а коэффициент μ_B — магнетон Бора (отметим, что, хотя сказанное подразумевает удвоение числа уровней Ландау, это не изменяет характер подхода разделов 7.6 и 7.7, поскольку степень вырождения также уменьшается в два раза). Рассчитайте значения g^* и $\mu_B B$ и сравните их с разницей значений энергии для уровней Ландау.

7. Эффекты заполнения подзон в целочисленном квантовом эффекте Холла (IQHE). Нарисуйте схематическую графическую зависимость ρ_L от B для подзон с $n = 1$ и $n = 2$. (Подсказка: используйте формулу $\Delta(1/B) = 2e/\hbar n_{2D}$ для двух разных значений n_{2D} .) Сравните механизмы возникновения осцилляций в зависимости ρ_L от B при очень больших и малых значениях магнитного поля B .

ГЛАВА 8

ОПТИЧЕСКИЕ И ЭЛЕКТРООПТИЧЕСКИЕ ПРОЦЕССЫ В КВАНТОВЫХ ГЕТЕРОСТРУКТУРАХ

8.1. Введение

Определение оптических характеристик полупроводниковых структур пониженной размерности является очень простым и ясным методом проверки справедливости физических теорий, которые уже были изложены в предыдущих главах для квантовых ям, сверхрешеток и квантовых точек. Хорошее совпадение теоретических и экспериментальных данных является доказательством того, что предсказания, основанные на зависимости функции плотности состояний от энергии, а также поведение экситонов в двумерных и нульмерных системах являются достаточно корректными. В этой главе рассматриваются оптические характеристики квантовых гетероструктур малой размерности, которые позволили уже создать целое поколение новых электронных приборов разного назначения (лазеры на квантовых ямах и точках, инфракрасные фотодетекторы, электрооптические модуляторы и т. д.). Принципы работы таких приборов, многие из которых уже выпускаются в промышленном масштабе, будут рассмотрены в главе 10.

В разделе 8.2 рассматриваются внутризонные и межзонные оптические переходы в квантовых ямах и влияние экситонов на эти переходы. Кроме этого, в разделе 8.2 предлагается обобщение теории оптических характеристик квантовых ям на сверхрешетки. В разделе 8.3 будут рассмотрены оптические свойства квантовых точек и нанокристаллов и показано, что ограниче-

ния по всем трем геометрическим измерениям в случае квантовых точек приводят к возникновению новых явлений, которые могут быть использованы для создания лазеров и инфракрасных фотодетекторов. Раздел начинается с описания методов выращивания квантовых точек и нанокристаллов, поскольку именно технология выращивания во многом определяет их оптические свойства.

Раздел 8.4 посвящен описанию воздействия электрических полей на квантовые ямы. Причина повышенного интереса к этим явлениям связана с тем, что влияние электрического поля на основные оптические свойства материалов (коэффициент преломления, спектр пропускания и т. п.) резко усиливается для объектов с низкой размерностью. Например, в современных электрооптических модуляторах, основанных на квантово-размерном эффекте Штарка, используется существенное изменение оптического поглощения экситонов под воздействием внешнего электрического поля. В разделе 8.5 рассмотрено влияние электрического поля на сверхрешетки, где оно приводит к локализации носителей в ямах и возникновению системы эквидистантных энергетических уровней (лестницы Штарка). В этом же разделе приводятся и обсуждаются новые данные по осцилляциям Блоха и микроволновому излучению.

8.2. Оптические свойства квантовых ям и сверхрешеток

Оптические свойства, связанные с межзонными переходами в квантовых ямах, совершенно непохожи на соответствующие характеристики объемных материалов, что легко объяснить особенностями двумерной оптической плотности и тем, что экситонное поглощение в двумерных системах намного сильнее (см. главу 4). Значительная разница между трех- и двумерными системами наблюдается и в случае внутризонных переходов. Такие переходы в двумерных системах могут происходить и без участия фононов, за счет электронов (или дырок) в локализованных состояниях ям, а также между локализованными состояниями и непрерывным спектром. Более того, в двумерных системах такие переходы могут специально созданы для излучения света (квантовые каскадные лазеры) или его регистрации (инфракрасные фотодетекторы).

На рис. 8.1 представлена квантовая яма (как в реальном пространстве, так и в пространстве волновых векторов), в которой происходят межзонные переходы из начального состояния в валентной зоне в конечное, расположенное в зоне проводимости. Отметим сразу, что поглощение проявляется при более высоких энергиях, чем в трехмерных системах, поскольку разность энергий между состояниями больше ширины запрещенной зоны полупроводника. Для расчета скоростей перехода можно воспользоваться введенным в главе 2 «золотым правилом» Ферми (2.26) для процессов, происходящих под воздействием зависящего от времени возмущения:

$$W = \frac{2\pi}{\hbar} \rho(E) |H'_{nk}|^2, \tag{8.1}$$

считая, что возмущение обусловлено взаимодействием фотон — электрон и имеет вид

$$H' = -e\vec{r} \cdot \vec{E}_0, \tag{8.2}$$

где вектор \vec{r} относится к положению электрона в плоскости раздела, т. е. $\vec{r} = (r_x, r_y)$, а \vec{E}_0 — амплитуда электрического поля падающей световой волны. Матричные элементы для пе-

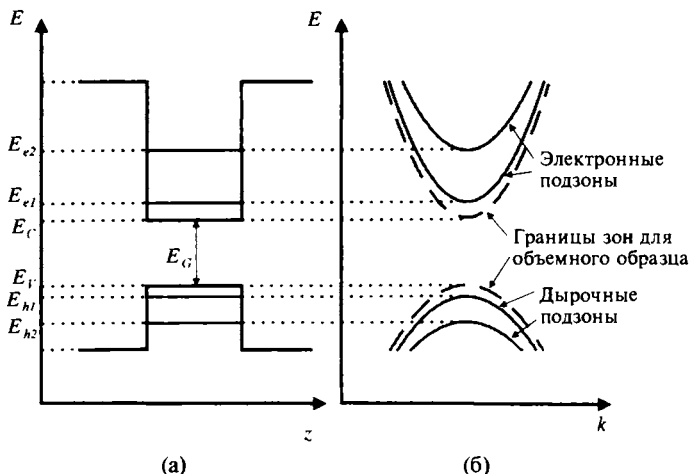


Рис. 8.1. Энергетическая диаграмма квантовой ямы: *a* — в реальном пространстве; *b* — в пространстве волновых векторов (k -пространстве)

перехода электрона от исходного состояния ψ_i в валентной зоне в конечное состояние ψ_f в зоне проводимости определяются уравнением

$$M = \int \Psi_f^* r_x \Psi_i dr. \quad (8.3)$$

Необходимо отметить, что, поскольку введенный выше вектор $\vec{r} = (r_x, r_y)$ является планарным, в уравнении 8.2 можно рассматривать одну из его компонент — r_x или r_y , которые являются координатами электрона в плоскости. Предположим далее, что рассматриваемая кристаллическая структура имеет кубическую (в данном случае квадратную) симметрию, тогда направления x и y эквивалентны. В кристаллической двумерной структуре электронные состояния описываются функциями Блоха:

$$\Psi_i \propto u_v(\vec{r}) \psi_{vn}(z) e^{i\vec{k} \cdot \vec{r}}, \quad (8.4)$$

$$\Psi_f \propto u_c(\vec{r}) \psi_{cn'}(z) e^{i\vec{k}' \cdot \vec{r}}, \quad (8.5)$$

где векторы \vec{r} и \vec{k} соответствуют электронам в двумерной системе. Мы пользуемся одинаковыми векторами в обоих уравнениях (8.4) — (8.5), поскольку при рассматриваемых переходах сохраняется общий момент, а импульсом фотона можно пренебречь по сравнению с импульсом электрона. Подставляя (8.4) и (8.5) в уравнение (8.3), можно получить выражение для матричного элемента M в виде произведения двух величин

$$M = M_{vc} M_{nn'}, \quad (8.6)$$

где

$$M_{vc} = \int u_c^*(\vec{r}) r_x u_v(\vec{r}) d\vec{r} \quad (8.7)$$

есть матричный элемент, соответствующий дипольному моменту переходов между блоховскими состояниями в валентной зоне и зоне проводимости, а

$$M_{nn'} = \int_{\text{well}} \psi_{cn'}^*(z) \psi_{vn}(z) dz \quad (8.8)$$

соответствует перекрытию электронных и дырочных волновых функций в квантовых ямах.

Таким образом, в соответствии с уравнением (8.6) матричные элементы межзонных переходов распадаются на два сомножителя, для которых могут быть сформулированы следующие условия или правила отбора:

1. Перекрытие электронных и дырочных огибающих функций, определенное уравнением (8.8) должно быть отлично от нуля. Поэтому в случае прямоугольной ямы бесконечной глубины должно выполняться (из-за ортогональности волновых функций) правило

$$\Delta n = n' - n = \delta_{nn'}, \quad (8.9)$$

где квантовые числа n и n' относятся к электронным и дырочным квантовым ямам соответственно.

2. Матричные элементы между блоховскими функциями, соответствующие электрическим дипольным переходам и определенные уравнением (8.7), также должны быть отличны от нуля. Для квантовых ям в структурах AlGaAs/GaAs выражение (8.9), относящееся к огибающим функциям, на практике может рассматриваться как правило отбора.

На рис. 8.2, взятом из работы [1], представлен спектр поглощения множественных квантовых ям (MQW) в GaAs/AlAs при температуре 6 К, из которого следует, что поглощение растет с энергией фотонов и в общих чертах соответствует ступенчатой функции плотности состояний для двумерной электронной системы (см. раздел 4.2). Кроме этого, на приведенном графике отчетливо выделяются экситонные пики в начале каждой «ступеньки». В двумерных системах энергия связи экситонов (как и связанное с ними поглощение) очень сильно возрастает вследствие пространственной локализации, в результате чего экситоны регистрируются гораздо легче, чем в трехмерных полупроводниках. Каждый пик на рис. 8.2 связан с дублетом спин-орбитального взаимодействия, соответствующим характеристикам валентных зон тяжелых и легких дырок в соединениях класса $A_{III}B_V$ (см. раздел 4.9). На рис. 8.2 можно четко выделить переходы из состояний с числами $n = 1, 2, 3$ из зон тяжелых (НН) и легких дырок (ЛН) в электронные состояния.

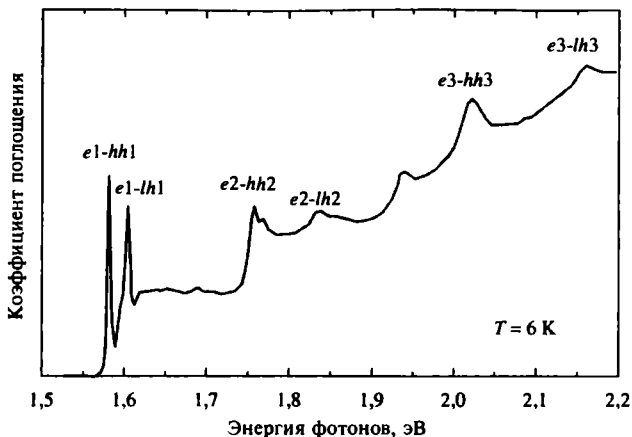


Рис. 8.2. Спектры поглощения множественных квантовых ям (MQW) с шириной около 7,6 нм в структурах GaAs/AlAs. Спектр описывает переходы в электронные состояния из зон тяжелых (НН) и легких дырок (ЛН). Кривая взята из работы [1]

Процесс излучения или поглощения фотонов для внутризонных переходов схематически показан на рис. 8.3, а. «Свободные» носители в зоне проводимости (или валентной зоне) занимают подзоны, соответствующие уровням каждого электрона (или дырки), расположенных в данной яме. Отметим, что межподзонные переходы происходят только между квантовыми состояниями внутри одной ямы. Дополнительно, межзонные переходы могут происходить между состояниями в квантовых ямах и протяженными электронными состояниями, как показано на рис. 8.3, б. Исходя из соображений типа предложенных выше для межзонных переходов, можно предложить еще одно правило отбора, связывающее состояния с противоположной четностью (поскольку в рассматриваемом случае матричные элементы должны включать в себя и координаты), что дает условие

$$\Delta n_z = n_{z,f} - n_{z,i} = \pm 1. \quad (8.10)$$

В идеальном случае это правило отбора означает, что величина Δn_z должна равняться нечетному числу, однако уже при $\Delta n_z = \pm 3$ интенсивность таких переходов настолько ослабевает, что их уже можно не учитывать формулой (8.10). Еще одно требование связано с тем, что направление вектора \vec{E} долж-

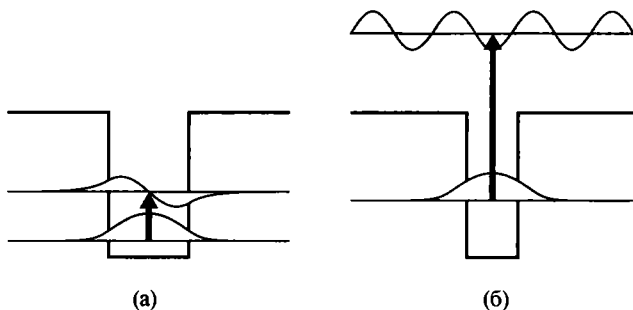


Рис. 8.3. Внутризонные оптические переходы в квантовой яме по данным работы [2]: *a* — внутризонные переходы; *b* — переходы между связанными состояниями в яме и протяженными состояниями

но быть перпендикулярно к яме, т. е. свет должен быть поляризован вдоль z -направления. Практически это означает, что свет не должен падать вертикально (или, с другой стороны, на поверхности образца должна располагаться какая-либо дифракционная решетка). Оба условия предполагают некоторое ослабление интенсивности используемого светового потока, вследствие чего, естественно, следует стремиться к созданию и использованию MQW с большим числом ям (около сотни).

Изменения, вызываемые электрическим полем в оптических свойствах квантовых ям и сверхрешеток, весьма различаются по своей природе, и мы еще вернемся к этому вопросу в разделах 8.4 и 8.5 соответственно. В настоящее время исследования электрооптических эффектов в сверхрешетках связаны в основном с блоховскими осцилляциями и возможностью излучения в терагерцовом диапазоне (раздел 8.5). С другой стороны, некоторые оптические свойства сверхрешеток похожи на свойства квантовых ям. В отличие от множественных квантовых ям, сверхрешетки и их оптические особенности уже нашли некоторые интересные применения в оптике, вследствие чего сейчас исследователи проявляют все больший интерес к некоторым смежным вопросам материаловедения, например, к методам выращивания квантовых точек, созданию и «проектированию» материалов с требуемыми энергетическими зонами и т. д.

Из различий в оптических свойствах квантовых ям и сверхрешеток можно выделить следующие, представляющие практический интерес. В сверхрешетках край полосы поглощения сдвинут

в голубую сторону спектра из-за расщепления уровней в мини-зоны. Кроме этого, все переходы являются размытыми, что объясняется отличием функции плотности состояний для сверхрешеток (рис. 5.15) от соответствующей, ступенчатой функции для квантовых ям. Вследствие туннелирования электронов через барьеры волновая функция экситонов может простираться сразу на несколько ям внутри сверхрешетки с сильной связью. Таким образом, в сверхрешетках экситоны оказываются значительно менее локализованными и имеют меньшую энергию связи, чем в квантовых ямах. Слабые экситонные особенности проявляются также вблизи переходов из дна мини-зоны дырок в верхнюю часть мини-зоны электронов (экситоны «седловой» точки). На рис. 8.4 представлен спектр поглощения сверхрешетки в структурах AlGaAs/GaAs (полученный в работе [2]), напоминающий спектр для отдельной квантовой ямы. Поведение экситонов на рисунке показано переходами с $n = 1$, соответствующими переходам от тяжелых и легких дырок в первое электронное состояние. На графике выделены пики S_{hh} и S_{lh} , относящиеся к упомянутым выше переходам вблизи седловой точки.

8.3. Оптические характеристики квантовых точек и нанокристаллов

8.3.1. Методы выращивания кристаллов. Самоорганизация квантовых точек

Развитие методов выращивания квантовых точек приобрело в последнее время особое значение, поскольку выяснилось, что именно технология изготовления определяет важнейшие характеристики квантовых точек, такие, как их структура, форма и распределение по размерам, стехиометрия, структура границ и поверхностей раздела и т. п. Поэтому, предваряя описание оптических свойств квантовых точек (следующий раздел), мы хотим очень кратко описать особенности методов выращивания таких структур, обращая особое внимание на новейшее направление исследований, связанное с самоорганизацией таких систем.

Физические методы создания наноструктур сводятся в основном к литографии и травлению, что не позволяет создавать квантовые точки достаточно малых размеров (т. е. настолько

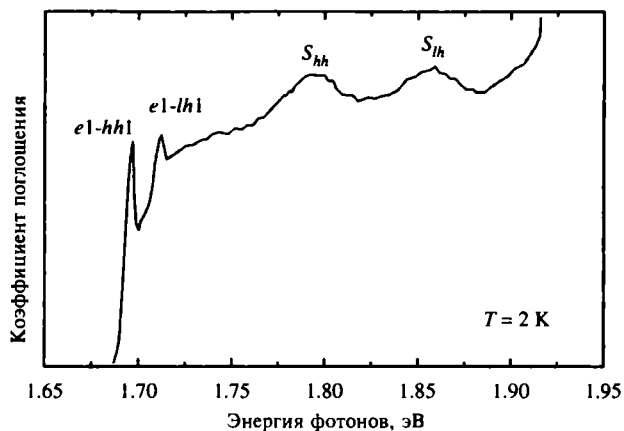


Рис. 8.4. Спектр поглощения сверхрешетки в AlGaAs/GaAs (по данным работы [2])

малых, чтобы в них проявлялись интересные для нас квантовые эффекты), даже при использовании новейших методов, использующих электронные пучки. Преимуществом физических методов является то, что они хорошо согласованы с микроэлектронной технологией. Из новейших и быстро прогрессирующих методов этого типа можно выделить сканирующую туннельную микроскопию (СТМ) и связанную с ней атомно-силовую микроскопию (АСМ). Возможно, именно эти технологии позволят нам в близком будущем изготавливать изолированные квантовые точки или массивы таких точек.

Гораздо чаще для изготовления квантовых точек применяются химические методы, однако они настолько разнообразны, что мы просто перечислим их, упомянув основные особенности.

1. Нанокристаллы в стеклянных матрицах.

Технология выращивания нанокристаллов в стеклянных матрицах является весьма развитой и распространенной, частично вследствие того, что уже давно существует промышленное производство светофильтров и фотохромных стекол на основе галидов меди (таких, как CuCl, CuBr, CuI и т. п.). Нанокристаллы соединений класса II—VI (CdS, CdSe, ZnSe и т. п.), получаемые обычно из пересыщенных вязких растворов, очень широко используются в стеклянных матрицах при изготовлении оптических фильтров, прежде всего, из-за легкости регулирования размеров точек.

2. Синтез в коллоидных системах.

Для производства металлических нанокластеров (например, из золота) очень широко используется коллоидный синтез, основанный на восстановлении солей металлов органическими лигандами. Например, такими методами производят наночастицы многих полупроводниковых соединений $A_{II}B_{VI}$ (CdS , $CdSe$ и т. п.), используя реагенты, содержащие нанокристаллы. Один из реагентов при этом содержит ион металла (например, Cd^{2+}), а другой — ион халкогенида (например, Se^{2-}), а размер образующихся нанокристаллов может регулироваться изменением температуры раствора и концентрации реагентов или стабилизаторов системы.

3. Газофазный метод.

Газофазные методы получения наночастиц основаны на формировании кластеров в газовой фазе, который затем осаждаются на подложке или адсорбируются на ней. Газофазные методы включают также напыления и лазерное испарение, которые широко применялись для получения кластеров металлов. Однако в последние годы была обнаружена видимая фотолюминесценция пористого кремния, после чего многие исследователи занялись проблемой осаждения кремниевых кластеров, осаждаемых путем стимулированного лазером разложения SiH_4 или магнетронного напыления для получения новых типов источников излучения света.

4. Самоорганизация квантовых точек.

Начиная с 90-х годов прошлого века все большее внимание исследователей привлекают технологии так называемой самоорганизации квантовых точек, которые совместимы с существующими методами микроэлектроники и оптоэлектроники. Одна из таких технологий основана на модификации метода Странского — Крастанова ($S - K$), при котором материал выращивается послойно на подложке, с близкой постоянной решетки. При малой (до нескольких процентов) разнице в постоянных решетки сначала осуществляется послойное наращивание, но на некоторой стадии слои начинают расти островками, поскольку при этом минимизируется упругая энергия (режим Странского — Крастанова). При очень большей разнице постоянных решетки такой островковый рост начинается с самого начала процесса (режим Фольмера — Вебера).

Описанные три режима послойного выращивания кристаллических слоев на подложке представлены на рис. 8.5, где схематически показаны (а) послойный рост, (б) метод Стран-

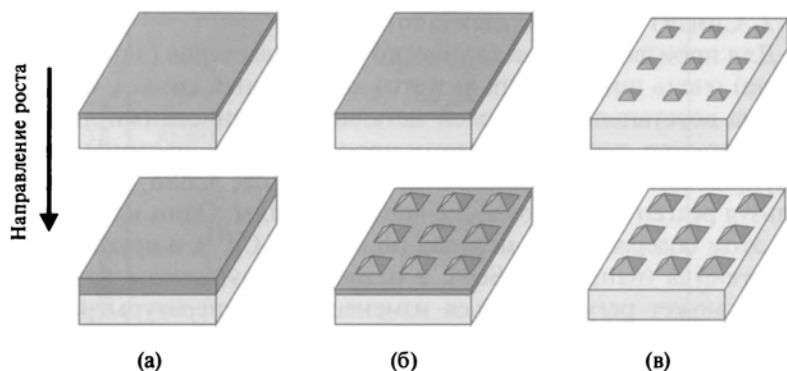


Рис. 8.5. Схематические диаграммы различных режимов выращивания кристаллических пленок на подложках: *а* — послойный рост; *б* — метод Странского — Крастанова (С — К); *в* — метод Фольмера — Вебера (Ф — В)

ского — Крастанова (С — К) и (в) метод Фольмера — Вебера (Ф — В). Наличие определенного режима роста связано не только с различием постоянных решетки подложки и кристалла, но и со значениями поверхностной энергии. Например, как будет показано в разделе 10.6, методом С — К выращиваются квантовые точки для лазеров в материалах типа InGaAs/GaAs, однако метод может применяться и в системах SiGe/Si или CdSe/ZnSe. Преимуществом метода С — К перед известными литографическими методами является то, что он позволяет получать квантовые точки меньших размеров.

8.3.2. Оптические свойства

В квантовых точках (иногда их называют квантовыми ящичками) потенциал ограничивает движение электронов во всех трех пространственных измерениях, естественно имеющих нанометрические размеры, в результате чего энергетические уровни образуют ряд дискретных связанных состояний (раздел 4.6), несколько напоминающих электронные уровни для изолированного атома. При этом пространственная локализация волновых функций электронов очень сильно влияет на оптические свойства вещества, поэтому мы начнем изложение с перечисления некоторых общих *оптических свойств нульмерных ограниченных систем*.

1. Уширение запрещенной зоны.

Одной из важнейших особенностей оптических свойств квантовых точек является уширение запрещенной зоны, которое можно сразу заметить на рис. 8.1, *a*. В зависимости от размера R квантовой точки (для простоты предполагается, что она имеет просто сферическую форму) можно рассмотреть несколько ситуаций. Чтобы разделить эти ситуации, сравним радиус квантовой точки R с размерами экситонов, определяемыми соответствующим боровским радиусом a_B :

$$a_B = \frac{\hbar^2}{4\pi\epsilon_0\epsilon \mu e^2}, \quad (8.11)$$

где параметр μ означает приведенную массу экситона (см. раздел 3.7.3). В так называемом режиме «сильной локализации» (область значений $R \leq a_B$) энергия квантовой локализации (конфаймента) превышает энергию кулоновского взаимодействия, вследствие чего можно пренебречь экситонными эффектами, связанными с взаимодействием электронов и дырок. При таком подходе мы можем рассматривать электроны и дырки, расположенные рядом в одних и тех же квантовых точках. Все сказанное вытекает из того, что кулоновское взаимодействие обратно пропорционально величине R , в то время как ограничивающий потенциал (т. е. энергия) возрастает обратно пропорционально квадрату R . Например, боровский радиус экситона в CdS составляет 29 \AA , вследствие чего все нанокристаллы с размерами меньше 50 \AA ведут себя как квантовые точки. Однако при значениях R , значительно меньших, чем a_B , уже нельзя применять приближение эффективной массы, вследствие чего квантовая точка должна рассматриваться в качестве просто большой молекулы, которую мы должны описывать соответствующими молекулярными орбиталями.

В так называемом режиме «слабой локализации» ($R \leq 4a_B$) огибающие функции практически не изменяются, хотя кинетическая энергия экситонов (связанная с движением центра масс) возрастает, вследствие чего должна уменьшаться энергия связи. Это наблюдается, например, в CdSe при размерах квантовых точек $R > 100 \text{ \AA}$. В кристаллах галидов меди (CuCl, CuBr) размер экситонов a_B настолько мал ($\sim 1 \text{ нм}$), что квантовые точки практически всегда существуют в режиме «слабого ограничения».

Наиболее трудными для теоретического описания являются, естественно, системы со средней или промежуточной степенью локализации, для которых мы не можем использовать предложенные выше упрощенные модели.

2. Повышение силы осцилляторов.

Из приведенных выше данных легко заметить некоторую общую закономерность, а именно по мере уменьшения размерности системы ($3D \rightarrow 2D \rightarrow 1D \rightarrow 0D$) электронные состояния становятся все более «концентрированными» по энергии, что и отражают соответствующие функции плотности состояний. В предельном случае (при размерности $0D$) квантовые точки ведут себя подобно атомам, обладающим хорошо выраженными энергетическими уровнями, в результате чего возрастают силы осцилляторов, соответствующих оптическим переходам. Это обстоятельство чрезвычайно важно для создания новых оптоэлектрических устройств, и мы рассмотрим его подробнее в разделах 10.3 и 10.6, где будет показано, каким образом концентрация энергетических состояний может быть использована для повышения коэффициентов усиления лазерных устройств. Из тех же общих соображений, по мере уменьшения размерности систем, в них сильнее проявляются электрооптические эффекты (разделы 8.4 и 8.5), используемые для создания, например, оптоэлектронных модуляторов на квантовых ямах.

3. Оптические переходы.

В разделе 8.2 уже отмечалось, что межподзонные оптические переходы в двумерных системах разрешены лишь тогда, когда свет распространяется в плоскости квантовой ямы, т. е. электрическое поле фотона направлено перпендикулярно поверхностям раздела. Напротив, квантовые точки способны поглощать падающий свет любого направления вследствие того, что локализация по всем трем оптическим осям означает, что и волновые функции электронов также квантуются по всем трем пространственным координатам.

4. Уширение спектров.

Еще одной важной особенностью оптических характеристик квантовых точек является отсутствие температурной зависимости для ширины спектральных линий, что объясняется просто тем, что в них нет непрерывных областей электронных состояний. Напротив, в одномерных и двумерных системах существуют одно и два направления для непрерывных k -состояний соответс-

твенно. Разумеется, говоря о дискретности спектров квантовых точек, не следует считать, что они представляют собой совершенно правильные, теоретические δ -функции, так как некий реальный разброс по ширине линий (~ 10 мэВ) всегда существует из-за неоднородностей в распределении размеров квантовых точек. При очень высокой однородности размеров квантовых точек ширина линий может быть снижена и до значений < 10 мэВ, однако, с другой стороны, при большом разбросе значений ширина линий может возрасти до $0,1$ эВ. Небольшие изменения состава также являются одной из причин уширения запрещенной зоны в системах квантовых точек. Другие причины связаны с примесями, поверхностными или межфазными состояниями и т. п. Для изучения собственных характеристик квантовых точек и уменьшения описанных эффектов уширения спектральных линий уже возник новый раздел спектроскопии, называемый спектроскопией изолированных квантовых точек (см. ниже).

Описав основные особенности оптических свойств квантовых точек (КТ), мы должны рассмотреть и некоторых оптические свойства различных систем КТ.

1. Квантовые точки в полупроводниковых соединениях.

Ниже приводятся лишь некоторые примеры из огромного разнообразия результатов исследований спектров квантовых точек в полупроводниковых соединениях. Наиболее тщательно изучены квантовые точки в нанокристаллах соединений $A_{II}B_{VI}$, поскольку именно такие полупроводники могут быть получены достаточно легко и имеют важные технологические применения. Во многих случаях соединения $A_{II}B_{VI}$ вводятся в качестве легирующих примесей в легированные стекла в процессе выращивания из расплавленной стеклянной массы (при этом обычно введение этих соединений осуществляется для регулирования цвета образующихся стекол за счет образующихся нанокристаллов). На рис. 8.6 приводятся спектры поглощения квантовых точек CdS ($E_G = 2,6$ эВ) в стекле, полученные в работе [3] при различных размерах нанокристаллов. Обращает внимание «синий сдвиг» окраски, связанный с наноразмерными квантовыми эффектами. Используя смеси CdS с другими соединениями того же класса, например с CdSe ($E_G = 1,75$ эВ), можно покрыть практически весь видимый диапазон. Очень интересные результаты по нанокристаллам CdS были получены в 1996 г. в работе Эмпедокла и соавт. [4],

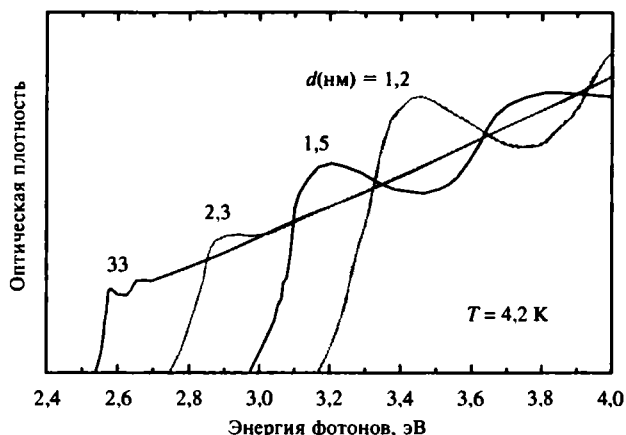


Рис. 8.6. Спектр поглощения нанокристаллов CdS различного размера в стекляннй матрице (по данным работы [3])

которые применили спектроскопию изолированных квантовых точек, напоминающую методику известной спектроскопии изолированных молекул. Кривые рис. 8.7 демонстрируют явные отличия спектра изолированного нанокристалла и набора или ансамбля таких нанокристаллов.

2. Самоорганизация квантовых точек.

Еще в 1995 г. в работе Гольдштейна и соавт. [5] была обнаружена возможность образования трехмерных островковых структур при выращивании сверхрешеток InAs/GaAs. В работе было показано, что при наращивании нескольких монослоев (в среднем $\sim 2,5$) происходит образование квантовых точек из InAs, что может быть надежно зарегистрировано методами просвечивающей электронной микроскопии высокого разрешения. Линейный спектр фотолюминесценции при этом оказался достаточно широким, благодаря размытости распределения точек по размерам, и пики фотолюминесценции были смещены в сторону меньших энергий из-за того, что точки имеют больший размер, чем периоды сверхрешеток. Примерно с 1994 г. исследователи стали все больше интересоваться процессами создания квантовых точек для использования в лазерах на основе систем типа InGaAs/GaAs. На рис. 8.8 приведены спектры фотолюминесценции и электролюминесценции для квантовых точек InAs/GaAs, выращенных методами химического осаждения из газовой фазы с использованием металлоорганических

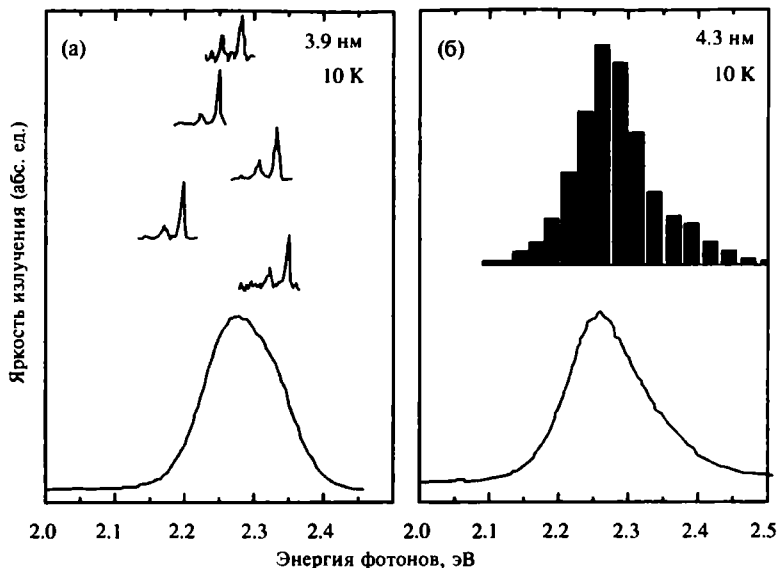


Рис. 8.7. (а) Сравнение спектра изолированных нанокристаллов CdS со средним размером 3,9 нм (верхняя часть рисунка) со спектром ансамбля таких нанокристаллов (нижняя часть); (б) гистограмма энергий системы из примерно 500 квантовых точек со средним размером 4,3 нм (верхняя часть рисунка) и спектр излучения этой системы (нижняя часть)

соединений (MOCVD), похожими на те, которые используются для получения лазеров на квантовых точках [6].

В работе 1995 г. Грандмана и соавт. [7] для исследования изолированных квантовых точек InAs/GaAs использовались методы локальной катодолюминесценции (рис. 8.9). Авторам удалось зарегистрировать несколько очень важных особенностей изучаемых систем: а) уширение спектральных линий не наблюдается вплоть до 70 К (это является доказательством того, что нанокристаллы, как отмечалось, имеют нулевую размерность); б) ширина отдельных линий не превышает 0,15 мэВ; в) спектроскопическое разрешение метода достаточно высоко, чтобы при относительно небольшом числе молекул в каждой точке (около 2000) регистрировать добавление или исключение даже отдельных молекул InAs.

3. Нанокристаллы полупроводников, обладающих запрещенной зоной с непрямыми оптическими переходами.

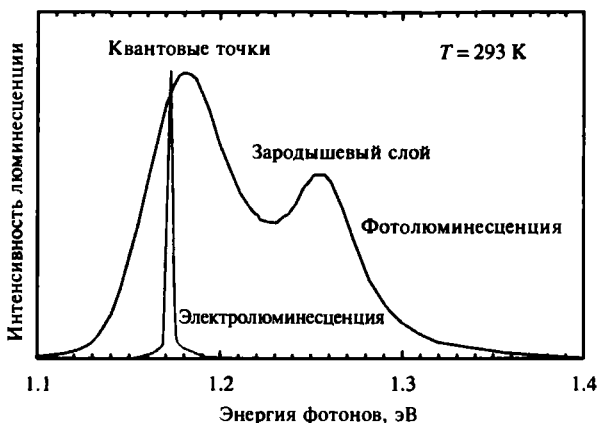


Рис. 8.8. Спектры фотолюминесценции (PL) и электролюминесценции (EL) для квантовых точек InAs/GaAs, выращенных методами химического осаждения из газовой фазы с использованием металлоорганических соединений (MOCVD) (по данным работы [6])

Начиная с открытия Канхамом [8] в 1990 г. явления сильной люминесценции в кристаллах кремния, исследования кристаллических кластеров Si и пористого кремния представляют особый интерес для специалистов в этой области. Этот интерес связан с тем, что, хотя из физики полупроводников считалось известным, что эффективность излучения света объемными полупроводниками, обладающими запрещенной зоной с непрямыми оптическими переходами, всегда является незначительной, ситуация резко меняется при использовании нанокристаллических образцов. Например, оказалось, что излучение нанокристаллов кремния лежит не в инфракрасном диапазоне ($E_G = 1,1 \text{ эВ}$), а видимой области ($E_G \sim 2 \text{ эВ}$), что объясняется именно уширением запрещенной зоны из-за квантовых эффектов (см. рис. 8.1, а). Кроме того, многочисленные эксперименты показали связь эффекта «синего сдвига» линий с уменьшением размера кристаллов. Что касается эффективной люминесценции, то она может объясняться, как обычно, релаксацией закона сохранения волнового вектора \vec{k} (или сохранения момента $\vec{p} = \hbar\vec{k}$) в условиях низкоразмерной локализации (см. ниже).

Сильная люминесценция в видимом диапазоне особенно заметна для пористого кремния (ПК), получаемого из монокристаллов кремния анодным электролитическим травлением

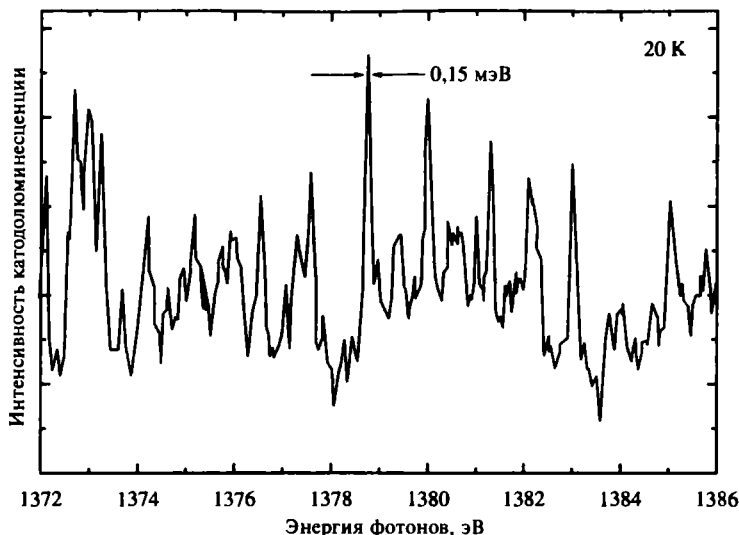


Рис. 8.9. Спектр локальной катодолюминесценции для квантовых точек InAs/GaAs (по данным работы [7])

в электролите, состоящем из раствора HF — этанол. Образующийся пористый слой формируется из нульмерных и одномерных наноразмерных кристаллов, а наблюдаемый при этом очень широкий спектр эмиссии объясняется квазигансовским распределением нанокристаллов по размерам. Сначала многие исследователи считали, что наблюдаемое сильное излучение в видимой области связано с образованием разнообразных соединений, на поверхности нанокристаллов в процессе электролитического травления. К числу таких примесей, возникающих в процессе формирования пористого кремния, относили силосен ($\text{Si}_6\text{O}_3\text{H}_6$), гидриды, оксиды и оксигидриды кремния, однако, хотя в некоторых случаях такое объяснение полностью исключить нельзя, в настоящее время считается, что объяснение связано с учетом квантовых эффектов вследствие нанометровых размеров структур. Действительно, такая люминесценция наблюдалась и в пористом кремнии без поверхностных соединений, что неоднократно проверялось с использованием самых современных технологий анализа поверхности, структуры и состава пленок и т. п., таких, как оже-электронная спектроскопия и спектроскопия для химического анализа, спектро-

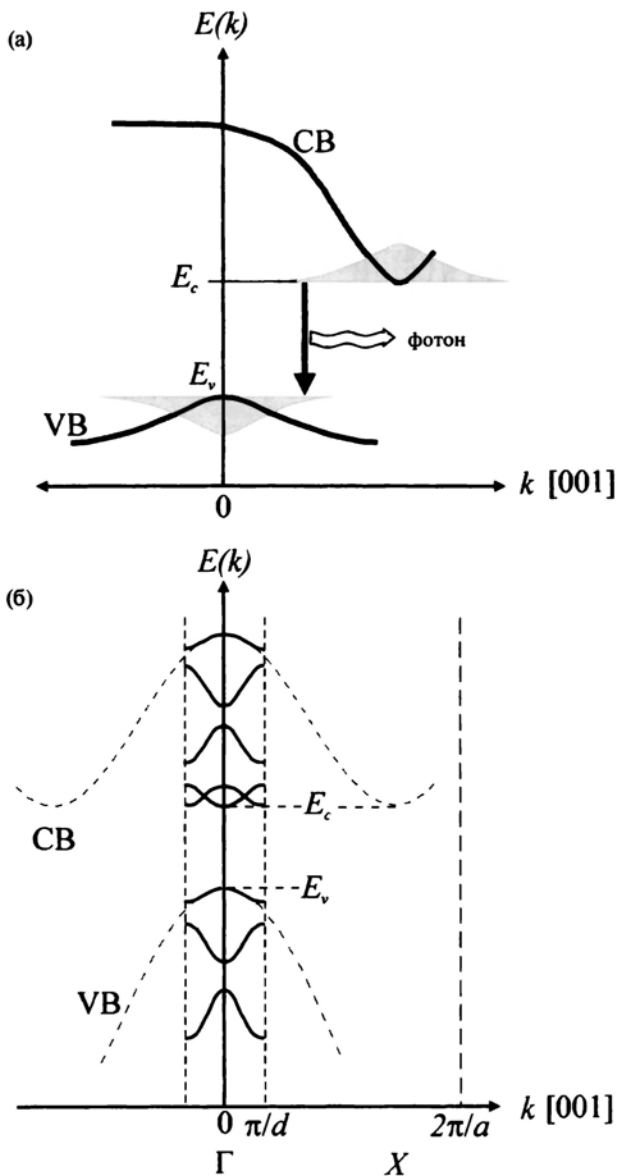


Рис. 8.10. (а) Повышение вероятности оптических переходов в нанокристаллах на основе кремния из-за принципа неопределенности Гейзенберга; (б) схематическое описание эффекта «расщепления» зон Бриллюэна в сверхрешетках SiGe

метрии резерфордовского обратного рассеяния, инфракрасной фурье-спектроскопии и т. д.

Квантовые эффекты, связанные с размерами структур, наблюдались и в других соединениях кремния, например, в сверхрешетках и квантовых ямах на основе SiGe. Повышенный интерес к гетероструктурам в SiGe объясняется их успешным применением в производстве биполярных транзисторов на гетероструктурах, о которых рассказывается в разделе 9.3. Были созданы напряженные гетероструктуры в SiGe, использующие различие ширины запрещенной зоны (1,1 эВ в кремнии и 0,74 эВ в германии соответственно), а также соединения SiGe с промежуточными значениями ширины запрещенной зоны.

Интенсивная люминесценция, наблюдаемая в нанокристаллах кремния и сверхрешетках состава кремний — германий, может быть объяснена на основе принципа неопределенности Гейзенберга или концепции расщепления зон Бриллюэна. Предположим, например, что размеры всех нанокристаллов Si в системе не превышают 30 нм. В этой ситуации распределение носителей заряда по импульсам имеет вид, показанный на рис. 8.10, а, что приводит к повышению вероятности прямых переходов. В сверхрешетках SiGe мини-зоны и запрещенные мини-зоны формируются на энергетической диаграмме вблизи значения $k = 0$ (как показано на рис. 8.10, б и объяснено в разделе 5.5), в результате чего сверхрешетки должны вести себя подобно полупроводникам с квазипрямыми оптическими переходами, причем вероятность оптических переходов при этом является достаточно высокой.

8.4. Электрооптические эффекты в квантовых точках.

Эффект квантово-размерный Штарка

Выше было показано, что внешнее электрическое поле может существенно изменять оптические характеристики рассматриваемых материалов. Такие изменения получили название электрооптических эффектов, и они могут влиять как на величину коэффициента преломления (эффект электропреломления), так и коэффициента поглощения (эффект электропоглощения). Если энергия фотонов не превышает ширины запрещенной зоны, то основным фактором является электропреломление, так как пог-

лощением в этом случае можно пренебречь из-за его малости. Электрооптические эффекты широко используются во многих оптоэлектронных устройствах для управления светом с помощью внешних электронных схем. Такие устройства являются промежуточными и могут быть отнесены одновременно и к фотонике, и к микроэлектронике. Вследствие квантовой локализации электронов, высокой энергии связи экситонов и возможности «подгонки» характеристик запрещенной зоны многие полупроводниковые квантовые ямы (и вообще полупроводниковые наноструктуры) часто обладают очень интересными электрооптическими свойствами. Например, перпендикулярное к плоскости ямы электрическое поле, позволяет уменьшить энергию края полосы поглощения («красный сдвиг») и повысить коэффициент преломления материала ниже этого края. В окрестности границ запрещенной зоны эти изменения становятся существенными и очень сильно зависят от энергии фотонов.

Простейшим вариантом регулирования интенсивности светового луча является использование электропоглощения. Из физики твердого тела известно, что внешнее электрическое поле, приложенное к трехмерному образцу полупроводника, уменьшает энергию края поглощения и создает колебания при энергиях выше ширины запрещенной зоны (так называемый эффект Франца — Келдыша). Таким образом, для фотонов с энергией чуть ниже края поглощения полупроводник является «прозрачным» в отсутствие внешнего поля, но поглощающим при наложении поля. В объемных полупроводниках этот эффект проявляется слишком слабо для практического использования, однако в квантовых точках пространственная локализация приводит к новым эффектам электропоглощения, намного более сильным, чем в объемных материалах (см. раздел 10.8), что позволяет на основе структур с квантовыми ямами создать оптические модуляторы на эффекте электропоглощения. В тех случаях, когда вектор электрического поля параллелен плоскости квантовой ямы, ситуация очень похожа на случай обычного трехмерного полупроводника, поскольку электроны могут свободно двигаться в этом направлении, вследствие чего электрооптический эффект напоминает эффект Франца — Келдыша и, подобно последнему, не представляет практического интереса. Значительно более интересная ситуация возникает в системах, где электрическое поле приложено перпендикулярно слоям, когда возникающий эффект

подобен эффекту Штарка в атомах (сдвиг энергетических уровней электронов под воздействием поля), вследствие чего это явление называют *квантоворазмерным эффектом Штарка (QCSE)*.

На рис. 8.11 показаны диаграммы энергетических зон для квантовой ямы до и после наложения постоянного внешнего поля F , направленного перпендикулярно слоям структуры. Связанная с таким полем потенциальная электростатическая энергия eFz для электронов «накладывается» на потенциал внутрикристаллического поля, определяемый профилем минимума зоны проводимости. Вследствие этого профиль наклоняется под действием внешнего поля (как показано на рисунке), подобные явления происходят и для дырок. Следствием такого воздействия становится сильная деформация волновых функций электронных состояний и понижение энергии локализации носителей заряда. Понятно, что волновые функции частиц с отрицательным и положительным зарядом (электронов и дырок соответственно) деформируются разным образом: волновые функции электронов смещаются против направления поля, а дырок — в обратном направлении. Понижение энергии квантовой локализации для электронов

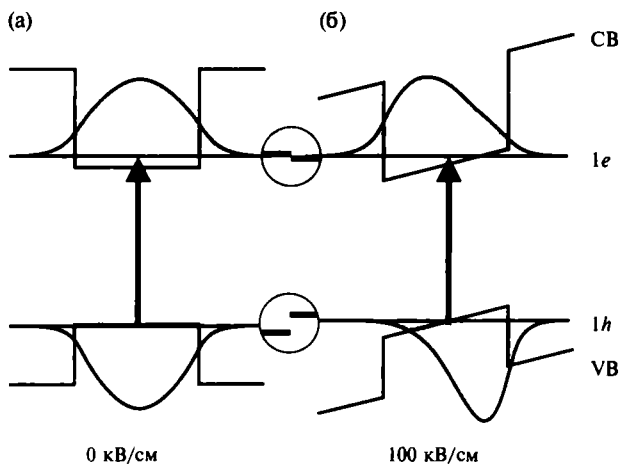


Рис. 8.11. Зонная структура и волновые функции основных состояний электронов и тяжелых дырок для квантовых ям толщиной 100 Å в структуре $\text{GaAs}/\text{Al}_{0,35}\text{Ga}_{0,65}\text{As}$ при различных значениях напряженности прилагаемого поля: *a* — отсутствие поля; *b* — поле напряженностью 100 кВ/см. На круговых врезках показаны штарковские сдвиги энергетических уровней (с тройным увеличением в масштабе)

и дырок приводит к сдвигу края полосы поглощения квантовых ям в сторону более низких энергий («красный сдвиг»).

На рис. 8.12 показаны взятые из работы [9] спектры поглощения слоя квантовых ям (толщиной 9,4 нм) в GaAs при различных значениях внешнего электрического поля, направленного перпендикулярно слою. Можно отметить сразу, что, в отличие от объемных кристаллов, экситонные эффекты в слое проявляются при напряженностях поля вплоть до 100 кВ/см. Объяснение этого факта основано на том, что в рассматриваемой системе не происходит ионизация экситонов, так как потенциал ям удерживает электроны и дырки на расстоянии, меньшем толщины слоя. На рисунке показан также «красный сдвиг» границы зоны поглощения и экситонного пика. Для проведения простого расчета мы можем рассмотреть прямоугольную квантовую яму с бесконечным барьером и шириной L . При умеренных значениях напряженности внешнего поля из уравнения Шрёдингера можно легко получить аналитическое выражение для штарковского сдвига ΔE , рассматривая прикладываемый потенциал как возмущение второго порядка. Для основного уровня сдвиг равен (см. задачу 7 в конце главы)

$$\Delta E_1 \approx -\frac{2.19 \times 10^{-3} e^2}{\hbar^2} m^* F^2 L^4, \quad (8.12)$$

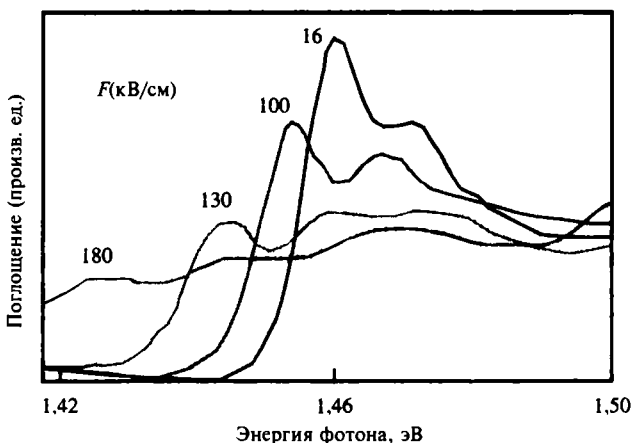


Рис. 8.12. Спектры поглощения квантовой ямы толщиной 9,4 нм в GaAs при различных значениях электрического поля, приложенного к слоям (по данным работы [9])

где m^* — эффективная масса в направлении перпендикулярном слоям; e — заряд электрона. Отметим, что из уравнения следует очень сильная зависимость величины сдвига от ширины ямы (четвертая степень) и его квадратичная зависимость от напряженности электрического поля, при этом, естественно, не зависит от знака электрического поля (хотя необходимо отметить, что это утверждение справедливо лишь для квантовых ям симметричной формы). Кроме этого, благодаря линейной зависимости от эффективной массы величина сдвига для тяжелых дырок оказывается значительно большей, чем для уровней электронов и легких дырок. Результатом такого воздействия становится то, что вызываемое электрическим полем уменьшение ширины запрещенной зоны квантовой ямы оказывается связанным, главным образом, с вкладом тяжелых дырок. На рис. 8.13 показана зависимость штарковского сдвига от напряженности приложенного поля для квантовых ям различной ширины в GaAs [10]. Например, для квантовой ямы толщиной 100 \AA в GaAs/Al_{0,35}Ga_{0,65}As уравнение (8.12) предсказывает для уровней тяжелых дырок и электронов сдвиги -10 и -2 мэВ соответственно. Методы численного расчета, позволяющие учитывать конечную высоту барьеров и отличие эффективной массы в барьерах, приводят к значениям сдвига -15 и -6 мэВ соответственно. Для фотонов с энергией чуть

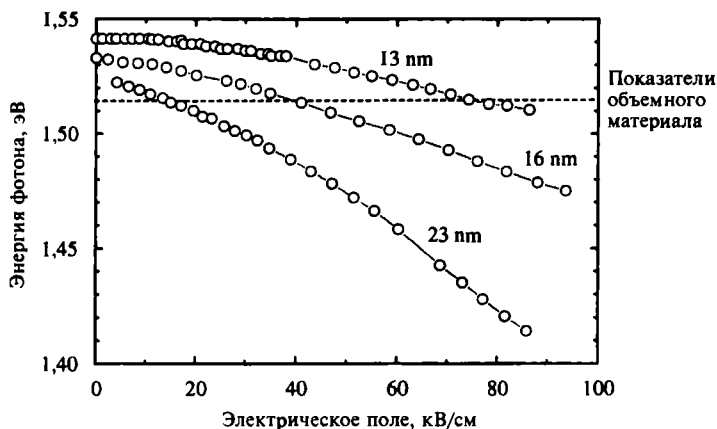


Рис. 8.13. Зависимость ширины запрещенной зоны для экситонов от напряженности электрического поля, приложенного к квантовым ямам различной ширины в GaAs/Al_{0,35}Ga_{0,65}As. Данные получены в экспериментах по измерению фотолюминесценции при температуре 5 К в работе [10]

меньше энергии запрещенной зоны в отсутствие поля можно считать, что яма является прозрачной, однако поглощение света резко повышается с ростом напряженности поля.

Обобщая сказанное, можно отметить, что квантовые ямы представляют большой интерес для прямой модуляции интенсивности пропускаемого света, так как при этом эффективность модуляции оказывается значительно выше, чем в приборах, основанных на объемных полупроводниковых или других электрооптических материалах, применяемых в настоящее время (типа ниобата лития и т. п.). Более подробно электрооптические модуляторы на квантовых ямах рассматриваются в разделе 10.8.

8.5. Электрооптические эффекты в сверхрешетках. Лестницы Штарка и осцилляции Блоха

В связанных квантовых ямах приложенное электрическое поле очень сильно влияет на взаимодействие между ямами. Прежде всего, это связано с тем, что электростатическая энергия электронов, находящихся в однородном электрическом поле, линейно растет с расстоянием, что приводит к сдвигу энергетических уровней в соседних ямах на величину, равную $\Delta E = eFd$, где d — расстояние между ямами. Резонанс между двумя такими состояниями в грубом приближении должен быть обратно пропорционален разнице энергий уровней, вследствие чего изменение напряженности поля должно позволять легко осуществлять «подгонку» между состояниями в различных ямах.

При слабой связи между ямами электронное состояние локализуется внутри ямы, так что оптические переходы между электронными и дырочными состояниями внутри одной квантовой ямы могут называться переходами внутри ямы, а переходы между состояниями разным ям — переходами между ямами (рис. 8.14). Электрооптические эффекты для переходов внутри ямы во многом напоминают переходы для изолированных квантовых ям. Более сложной выглядит ситуация в случае переходов между ямами, для которых внешнее электрическое поле, как отмечалось, создает сдвиг по энергиям $\Delta E = eFd$, знак которого определяется направлением поля. Если оно совпадает с направлением от электронных состояний к дырочным, то при переходах должно наблюдаться красное смещение. Кроме того, с ростом напря-

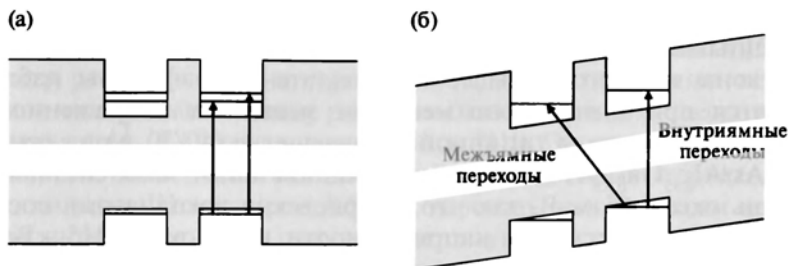


Рис. 8.14. Схематическое изображение симметричной двойной связанной ямы, демонстрирующее механизм оптических переходов из дырочного состояния в правой квантовой яме: а — в отсутствие электрического поля ямы связаны, и электронные состояния простираются (растягиваются) на обе ямы; б — при наложении конечного внешнего поля состояния локализируются в основном в одной из ям, вследствие чего возникает возможность классификации переходов как переходов внутри ям и между ямами

женности приложенного поля интенсивность переходов внутри ям уменьшается и они постепенно перестают проявляться вообще. Использование линейного сдвига в связанных квантовых ямах предлагает богатые возможности «подгонки» переходов при достаточно умеренных значениях внешнего поля.

В сверхрешетках наиболее интенсивный экситонный пик соответствует переходу с вершины дырочной мини-зоны на дно электронной мини-зоны, как показано на рис. 8.15, а. В электрическом поле, направленном нормально к слоям, связь между квантовыми ямами ослабляется, вследствие чего сами мини-зоны расщепляются на ряд эквидистантно расположенных энергетических уровней (расстояние между уровнями при этом составляет $\Delta E = eFd$, где d — период сверхрешетки). Образующаяся система уровней получила название *лестницы Штарка*, каждый уровень которой соответствует электронным состояниям, локализованным вблизи разных квантовых ям (рис. 8.15, б). Уже при значениях напряженности поля порядка $F \approx \Delta / ed$ (где Δ и d — ширина мини-зоны и постоянная сверхрешетки соответственно) состояния лестницы Штарка становятся полностью локализованными в соответствующей квантовой яме (см. рис. 8.15, в). Такой эффект трудно уловить в объемных кристаллах, поскольку ширина зоны в них обычно составляет несколько эВ, вследствие чего для проявления описываемой

локализации необходимо воздействовать на кристаллы очень мощными электрическими полями. В сверхрешетках ширина зоны намного меньше, в результате чего эффекты наблюдаются при значительно меньших значениях напряженности поля. Например, в типичной сверхрешетке ($40/20 \text{ \AA}$) на основе $\text{GaAs}/\text{Al}_{0,35}\text{Ga}_{0,65}\text{As}$ ширина электронной мини-зоны составляет лишь около 65 мэВ, так что штарковская локализация состояний достигается при напряженности поля около 115 кВ/см. Ширина мини-зоны тяжелых дырок еще меньше и составляет всего 6 мэВ, так что локализация состояний тяжелых дырок достигается уже при полях порядка 10 кВ/см.

Приведенный спектр поглощения сверхрешетки, как и некоторые другие спектры систем этого типа (например, показанный на рис. 8.16 спектр фототока, взятый из работы [11]), отражает очевидный эффект локализации состояний, вызываемых внешним электрическим полем. В отсутствие электрического поля на таких спектрах наблюдается один экситонный переход (как отмечалось выше, с вершины дырочной мини-зоны на дно электронной мини-зоны), что соответствует переходам между

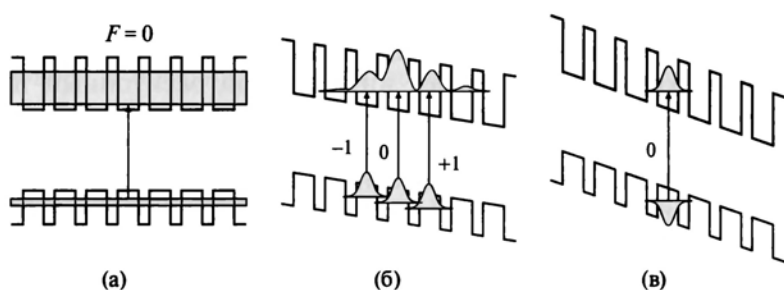


Рис. 8.15. Воздействие электрического поля на состояния в сверхрешетке: *a* — в отсутствие электрического поля экситонные переходы могут происходить между краями минизон; *b* — в электрических полях умеренной напряженности состояния дырок локализируются, а электронная минизона расщепляется, образуя лестницу Штарка, благодаря чему в системе могут осуществляться показанные переходы как внутри ям, так и между ямами; *в* — в сильных электрических полях электронные состояния локализируются в одиночных ямах, вследствие чего переходы между ямами в структуре перестают происходить. Во всех случаях *a* — *в* синий сдвиг края полосы поглощения составляет примерно половину ширины мини-зоны. Для простоты на рисунке указаны лишь переходы электронных состояний центральной квантовой ямы

делокализованными состояниями, как показано на рис. 8.15, а. В умеренных электрических полях, как показано на рис. 8.15, б, спектр расщепляется на серию пиков, соответствующих переходам между ямами между различными состояниями лестницы Штарка (рис. 8.17, по данным работы [11]). Затем, когда напряженность поля достигает значений, при которых происходит локализация Штарка (рис. 8.15, в), переходы между ямами исчезают, и остается лишь пик, соответствующий переходу внутри

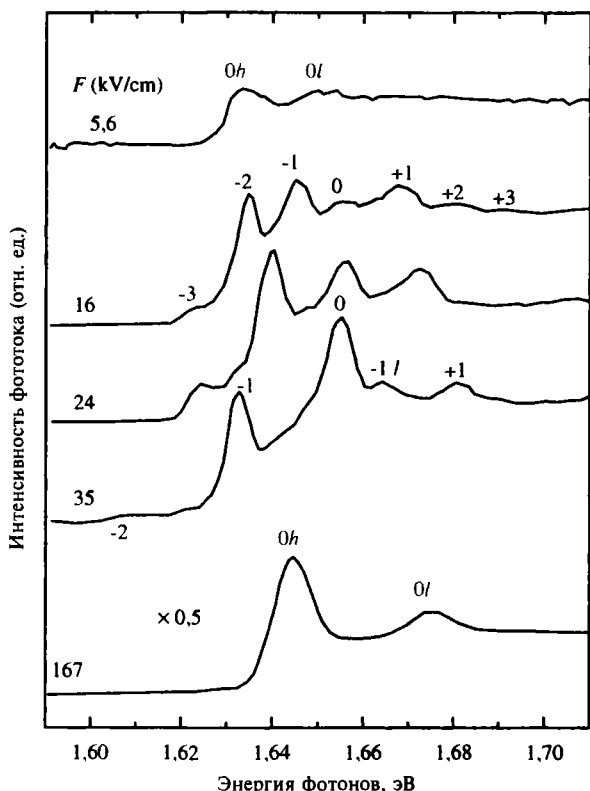


Рис. 8.16. Спектры фототока в сверхрешетке (40/20 Å) на основе GaAs/Al_{0,35}Ga_{0,65}As при температуре 5 К для различных значений приложенного электрического поля (по данным работы [11]) (кривые смещены для удобства рассмотрения). Переходы между ямами обозначаются цифрами, соответствующими числу периодов решетки между электронными и дырочными состояниями. Особо стоит отметить «синий сдвиг» края зоны поглощения от 0 до 167 кВ/см

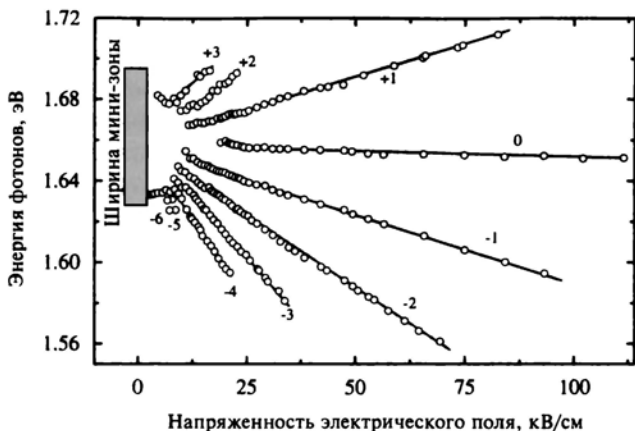


Рис. 8.17. Зависимость энергии переходов тяжелых дырок от напряженности приложенного электрического поля, измеренные в тех же условиях, что и данные на рис. 8.16 (по данным работы [11]). Для наглядности экспериментальные точки соединены прямыми линиями

ямы. Поскольку уровень этой единственной ямы располагается близко к центру мини-зоны, на краю зоны поглощения сверхрешетки при переходе к полям высокой напряженности наблюдается «синий сдвиг» (от нулевого значения). Такой «синий сдвиг» (резко отличающийся от обычного красного, наблюдаемого для достаточно толстых одиночных ям) равен примерно половине ширины электронной мини-зоны, что уже позволяет создавать на этой основе некоторые электрооптические приборы.

В предыдущем разделе мы рассматривали только статические режимы поведения электронов в условиях существования лестницы Штарка. Однако динамические характеристики также представляют большой интерес для исследований, так как их использование позволяет создавать очень быстрые электромагнитные излучатели, получившие название осцилляторов Блоха. Излучение с терагерцовой частотой в таких устройствах было впервые зарегистрировано в 1993 г., однако оно было предсказано еще в самой первой публикации Эсаки и Чанга по сверхрешеткам 1970 года. Для понимания механизма мы можем вновь рассмотреть эффект воздействия приложенного постоянного электрического поля на электронные состояния в сверхрешетках, как показано на рис. 8.18. В отсутствие электрического поля состояния кван-

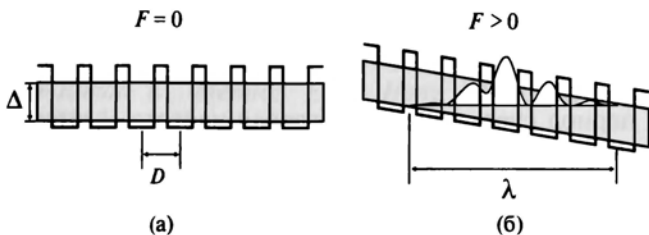


Рис. 8.18. Схематическое представление воздействия электрического поля на электронные характеристики полупроводниковой сверхрешетки. Поля направлено перпендикулярно слоям: а) поле отсутствует, б) изменение картины при наложении поля

товых ям являются полностью связанными, так что электронные состояния в сверхрешетках имеют бесконечную протяженность, а энергетические уровни образуют непрерывный спектр между потолком и дном минизон, как показано на рис. 8.18, а. В электрическом поле конечной напряженности связь между ямами ослабевает, а сами мини-зоны расщепляются, образуя описанную выше лестницу Штарка для уровней (рис. 8.18, б). Пространственную протяженность возникающих состояний можно оценить, рассматривая наличие мини-зон в качестве удобного приближения. Краевые значения таких мини-зон должны «изгибаться» в электрическом поле (из-за наличия в выражении для потенциальной энергии электростатического взаимодействия члена eFz). В квазиклассическом приближении можно считать, что кинетическая энергия электрона может лежать только в диапазоне энергий мини-зоны, что сразу определяет пространственные ограничения области расположения энергетических уровней данного электрона. Границы этой области определяются теми значениями z , при которых уровни пересекаются с границами мини-зоны. Расстояние между двумя такими точками пересечения равно

$$\lambda = \frac{\Delta E}{eF}. \quad (8.13)$$

В отсутствие рассеяния любой электронный волновой пакет должен осциллировать от одного конца промежутка до другого. По аналогии с классической задачей о гармоническом осцилляторе частота таких колебаний определяется расстоянием между уровнями, т. е. величиной

$$\omega_B = \frac{\Delta E}{\hbar} = \frac{eFd}{\hbar}, \quad (8.14)$$

где d — период сверхрешетки. Угловая частота ω_B при этом называется частотой Блоха, а ее значения могут достигать терагерцового диапазона. Сам факт существования блоховских осцилляций в полупроводниковых сверхрешетках уже подтвержден различными оптическими измерениями.

Обобщая материал этой главы, можно сказать, что осцилляции Блоха в сверхрешетках позволяют генерировать микроволновое излучение под воздействием электрического поля. Излучение при этом создается переходами между уровнями, которыми образуют лестницу Штарка. Частота такого излучения лежит в области 1 ТГц, что практически недоступно для существующих полупроводниковых источников микроволнового излучения. Стоит отметить, однако, что эффективность эмиссии приборов на основе сверхрешеток пока является недостаточной, так что такие устройства нуждаются в серьезной доработке для внедрения в коммерческое производство.

Литература

1. Fox A.M. (1996). Contemporary Physics, 37, 111.
2. Cingolani R., Tapfer L., Ploog K. (1990). Appl. Phys. Lett, 56, 1233.
3. Ekimov A.I., Efros A.L. (1991). Acta Physica Polonica A, 79, no. 1, 5.
4. Empedocles S.A., Norris D.J., Bawendi M.G. (1996). Phys. Rev. Lett, 11, 3873.
5. Goldstein L., Glas R., Marzin J.Y., Charasse M.N., Le Roux G. (1985). Appl Phys. Lett, 47, 1099.
6. Bimberg D., Grundman M., Ledentsov N.N. (1998). MRS Bulletin, 23, no. 2, 31-34.
7. Grundmann M., Christen J., Lendestov N.N., Böhrer J., Bimberg D., Ruvimov S.S., Werner P., Richter V., Gösele V., Heydenreich J., Ustinov V.M., Egorov A. Yu., Zhukov A.Z., Kop'ev P.S., Alferov Zh.I. (1995). Phys. Rev. Lett, 74, 4043.
8. Canham L.T. (1990). Appl. Phys. Lett., 51, 1046.
9. Miller D.A.B., Chemla D.S., Damen T.C., Gossard A.C., Wiegmann W., Wood T.H., Burrus C.A. (1985). Phys. Rev. B, 32, 1043.

10. *Vina L., Mendez E.E., Wang W.I., Chang L.L., Esaki L.* (1987). *J. Phys. C: Solid State Phys.*, 20, 2803.
11. *Agulló-Rueda E., Mendez E.E., Hong J.M.* (1989). *Phys. Rev. B*, 40, 1357.

Дополнительная литература

- Basu P.K.* (1997). *Theory of Optical Processes in Semiconductors* (Clarendon Press, Oxford).
- Bimberg D., Grundmann M., Ledentsov N.N.* (2001). *Quantum Dot Heterostructures* (Wiley, Chichester).
- Bryant G.W., Solomon G.S.* (2005). *Optics of Quantum Dots and Wires* (Artech House, Boston).
- Klingshirn C.* (2005). *Semiconductor Optics* (Springer, Berlin).
- Schäfer W., Wegener M.* (2002). *Semiconductor Optics and Transport Phenomena* (Springer, Berlin).
- Weisbuch C. Vinter B.* (1991) *Quantum Semiconductor Structures* (Academic Press, Boston).
- Yu P., Cardona M.* (1996) *Fundamentals of Semiconductors* (Springer, Berlin).

Задачи и упражнения

1. Оптические переходы в квантовых ямах. Рассчитайте вероятности переходов для $n = 1$ и 2 в квантовых ямах систем AlGaAs/GaAs (полагая ширину ямы равной $a = 10$ нм), исходя из правила отбора, предложенного в разделе 8.2. Нарисуйте схематически зависимость коэффициента оптического поглощения от энергии фотонов. (Подсказка: учтите, что поглощение начинает сильно возрастать при переходе ТД (тяжелая дырка) $\rightarrow e$ для $n = 1$, затем уменьшается и опять возрастает для перехода ЛД (легкая дырка) $\rightarrow e$ при $n = 1$. Почти одновременно начинается рост из-за функции плотности состояний n_{2D} . Примените этот же подход к рассмотрению переходов при $n = 2$.)
2. Уширение линий в квантовых ямах. Одной из причин уширения линий в квантовых ямах являются флуктуации ширины ям (параметр a), для описания которых обычно вводится некоторый инкремент Γ (коэффициент увеличения) ширины зоны. Покажите, что в данном случае уширение

линий поглощения или спектра поглощения описывается выражением

$$\Gamma = \frac{\hbar^2 \pi^2}{2 a^3} \left(\frac{1}{m_c^*} + \frac{1}{m_h^*} \right) \Delta a.$$

Оцените величину Γ для типичной квантовой ямы с шириной $\Delta a = a_0/2$, где a_0 — постоянная решетки материала ямы. Отметим, что такая ситуация встречается также, если, например, в получаемых по методу молекулярно-лучевой эпитаксии квантовых ямах рост останавливается и последний слой останется незавершенным.

3. Экситоны в квантовых ямах. По представленным на рис. 4.16 данным оцените качественно зависимость экситонных энергий тяжелых и легких дырок от ширины ямы. Объясните разницу в значениях энергии связи для тяжелых и легких дырок, а также разницу в энергиях для значений параметра $x = 0,32$ и $0,15$ (x является показателем содержания атомов Al в структуре $\text{Al}_x \text{Ga}_{1-x} \text{As/GaAs}$).
4. Квантовые точки. Рассмотрите поведение квантовой точки из GaAs в форме куба (с размерами 5 нм) с энергией квантовой локализации 0,8 эВ. Рассчитайте значение кулоновской энергии и покажите, что она намного меньше энергии квантовой локализации.
5. Сверхрешетки Si/SiGe. Рассчитайте период кубической сверхрешетки Si/SiGe (с размером 5 нм), в которой минимум зоны Si (соответствующий волновому числу $k \approx 0,8\pi a_0$, где a_0 — постоянная решетки) сводится к значению $k = 0$. (Подсказка: воспользуйтесь концепцией расщепления зон, описанной в разделе 5.5.2, в соответствии с которой оптические характеристики сверхрешетки подобны характеристикам запрещенной зоны с прямыми оптическими переходами в объемном полупроводнике.)
6. Сверхрешетки типа II в электрических полях. В сверхрешетках типа II (см. рис. 5.7, б) волновые функции электронов и дырок смещены на половину периода, т. е. на величину $d/2$. Рассчитайте для сверхрешетки типа II из $80/30 \text{ \AA} \text{ AlAs/GaAs}$ энергетический сдвиг, возникающий при наложении электрического поля напряженностью $F = 10^4 \text{ В/см}$. Вычислите значение штарковского сдвига и отметьте, что оно

меньше полученного перед этим значения энергетического сдвига. Примечание: известно, что рассматриваемые сверхрешетки при толщине слоя GaAs менее 35 \AA относятся именно к типу II.

7. Квантово-размерный эффект Штарка. Выведите уравнение (8.12) для квантово-размерного эффекта Штарка в квантовой яме. (Подсказка: используйте в качестве гамильтониана возмущений выражение $H' = eFz$, где F — электрическое поле вдоль направления z (ось роста кристалла). Поскольку первое приближение теории возмущений приводит к нулевому результату, попробуйте использовать приближение второго порядка, используя волновые функции для квантовых ям бесконечной глубины.) Примечание: данный эффект похож на квадратичный эффект Штарка в атомах водорода.
8. Осцилляции Блоха. Электронный транспорт в сверхрешетках может быть описан как электронный транспорт в системах со слабым периодическим потенциалом (обозначим период потенциала через a). Пренебрегая рассеянием электронов, рассчитайте частоту осцилляций Блоха при значениях $a = 4 \text{ нм}$ и $F = 3 \times 10^6 \text{ В/м}$. Считая концентрацию электронов малой, покажите, что поляризация P , связанная с электронами, т. е. продемонстрируйте, что $P = ex$, где x — координата электрона, а $v = (1/\hbar)(dE/dk)$ — групповая скорость, осциллирующая в зависимости от времени.

ГЛАВА 9

ЭЛЕКТРОННЫЕ ПРИБОРЫ НА НАНОСТРУКТУРАХ

9.1. Введение

Высокая степень интеграции, характерная для современной кремниевой технологии, не может быть достигнута при использовании полупроводниковых соединений $A_{III}B_V$, однако эти соединения обеспечивают большее быстродействие, прежде всего, за счет высокой подвижности μ носителей и меньших значений эффективной массы электронов в таких соединениях. Как было показано в главе 3 (раздел 3.5.1), подвижность носителей в GaAs примерно на порядок превышает соответствующее значение для чистого кремния. А скорость электронов в полупроводниковых материалах под влиянием внешнего электрического поля является основным параметром при проектировании новых высокоскоростных электронных приборов. Как будет показано в этой главе, MODFET на основе модулированно-легированных квантовых гетероструктур (раздел 5.3.1) могут обеспечить очень высокое быстродействие благодаря очень высоким значениям μ при продольном транспорте электронов. Граничная частота таких устройств обычно превышает соответствующие значения для полевых МОП-транзисторов на кремниевой основе, а также полевые транзисторы с барьером Шоттки в качестве затвора (MESFET) на основе GaAs. Необходимо также упомянуть, что высокая подвижность электронов в этих структурах является следствием квантования электронных состояний в образующихся двумерных системах, а также высокого совершенства изготовляемых поверхностей раздела AlGaAs—GaAs.

На рис. 9.1 представлена взятая из работы [1] зависимость рабочей частоты (в ГГц) различных типов модулированно-легированных полевых транзисторов от длины затвора (в микронах).

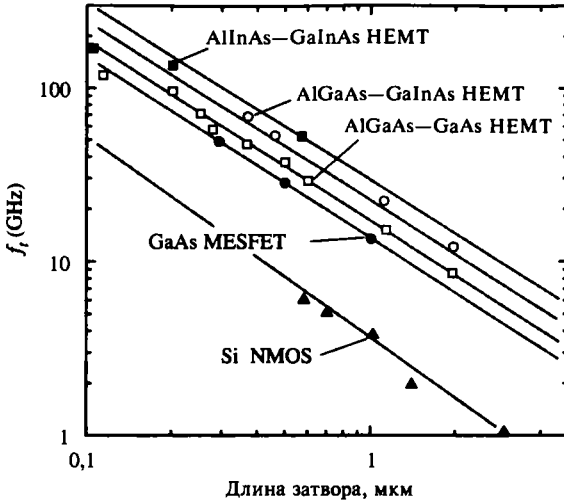


Рис. 9.1. Зависимость максимальной рабочей частоты различных типов транзисторов (MODFET, MESFET и полевых МОП-транзисторов) от длины затвора (по данным работы [1])

Благодаря своим характеристикам такие устройства получили также название полевых транзисторов с высокой подвижностью электронов (HEMT). И они будут рассмотрены ниже в разделе 9.2. Для сравнения на рис. 9.1 приведены также типичные характеристики кремниевых полевых МОП-транзисторов и полевых транзисторов с барьером Шоттки на основе GaAs. Значения частот приводятся для комнатной температуры (300 К), хотя стоит отметить, что они гораздо выше при температуре около 0 К вследствие роста подвижности при низких температурах (см. рис. 6.2 и 6.3). В настоящее время уже существуют модулированно-легированные полевые транзисторы с длиной затвора около 100 нм и рабочей частотой при комнатной температуре порядка нескольких сотен гигагерц (ГГц).

Использование квантовых гетероструктур не ограничивается полевыми транзисторами, в которых транспорт электронов осуществляется лишь параллельно поверхности квантовой ямы, а включает также транзисторы, в которых транспорт происходит перпендикулярно поверхности раздела гетероструктуры. Работа таких транзисторов основана на приложении разности потенциалов к эмиттеру, базе и коллектору, что напоминает механизм действия биполярных транзисторов. Максимальная

рабочая частота биполярных транзисторов ограничивается временем пролета носителей заряда через базу. Как будет показано ниже (раздел 9.3), гетеропереходные биполярные транзисторы (НБТ) на основе переходов в $\text{AlGaAs} - \text{GaAs}$ или $\text{Si} - \text{Ge}$ позволяют значительно повысить ряд важных параметров, таких, как предельная частота (частота отсечки), β -фактор, сопротивление базы и т. п. по сравнению с обычными кремниевыми биполярными транзисторами.

Обсуждая возможности создания транзисторов новых типов, следует особо отметить еще один очень интересный квантовый эффект, а именно так называемое резонансное туннелирование (раздел 9.4). При самом простом описании диоды с резонансным туннелированием (*RTD*), действующие на основе этого эффекта, представляют собой квантовую яму, окруженную двумя потенциальными барьерами, достаточно тонкими для того, чтобы через них могло осуществляться туннелирование электронов. Время пролета электронов через такую гетероструктуру чрезвычайно мало, вследствие чего устройства на основе *RTD* обладают исключительным быстродействием и могут работать при частотах порядка 1 ТГц. Объединяя *RTD* с биполярным или полевым транзистором, можно создать так называемые транзисторы с резонансным туннелированием (*RTT*), описанные в разделе 9.6. В этих транзисторах структура с резонансным туннелированием инжектирует горячие электроны (т. е. электроны с высокой кинетической энергией) в активную область транзистора, что позволяет создавать так называемые транзисторы на горячих электронах (*HET*), рассматриваемые в разделе 9.5.

Снижение характерных размеров приборов в нанометровый диапазон приводит к заметному уменьшению числа электронов, соответствующих прохождению электрического сигнала через прибор. Эта тенденция неизбежно подводит к созданию так называемых одноэлектронных транзисторов (*SET*). Как будет показано в разделе 9.7, характеристики одноэлектронных транзисторов определяются эффектом кулоновской блокады, проявляющимся в нульразмерных полупроводниковых структурах, типа квантовых точек (раздел 4.6). Электронный ток через квантовую точку в одноэлектронном транзисторе, соединенном с выводами посредством туннельных переходов, позволяет контролировать поток электронов «поштучно» подачей сигнала на

электрод, который в данном случае ведет себя подобно вентилю (затвору) транзистора.

9.2. Модуляционно-легированные полевые транзисторы (MODFET)

Как было показано в разделе 5.3.1, наличие потенциальной ямы очень малых размеров, сформированной в гетеропереходах структуры AlGaAs—GaAs, может приводить к квантованию уровней энергии, соответствующих движению электронов в направлении, перпендикулярном поверхности раздела, хотя движение электронов в плоскости, параллельной поверхности раздела, практически не отличается от движения свободных частиц. При этом отмечалось, что подвижность электронов в этой плоскости может быть исключительно высокой, поскольку возникающие в слое AlGaAs электроны поступают в нелегированный слой GaAs, где отсутствует рассеяние на примесных атомах и они могут двигаться параллельно поверхности раздела совершенно свободно под воздействием электрического поля. Именно на этом принципе с начала 1980-х г. начали создаваться полевые транзисторы с высокой подвижностью носителей (HEMT), которые иногда называют полевыми транзисторами с модулированным легированием (MODFET), так как в них используются модулированно-легированные гетеропереходы, а их действие основано на возможности регулирования движения электронов вдоль канала воздействием электрического поля. Модулированно-легированные полевые транзисторы (MODFET), которые уже нашли много полезных применений в высокочастотной технике, могут служить наглядным примером приборов, в которых высокие технические характеристики достигаются за счет использования квантового поведения электронов, локализованных в нанометровых потенциальных ямах с размерами меньше длины волны де Бройля электронов (раздел 1.3).

Полевые транзисторы на гетероструктурах имеют слоистую структуру, позволяющую создавать двумерный электронный газ с высокой подвижностью. На рис. 9.2, а схематически представлено сечение типичного модулированно-легированного полевого транзистора (MODFET), включающего в себя все привычные электроды транзистора (сток, исток и затвор). Энергетическая

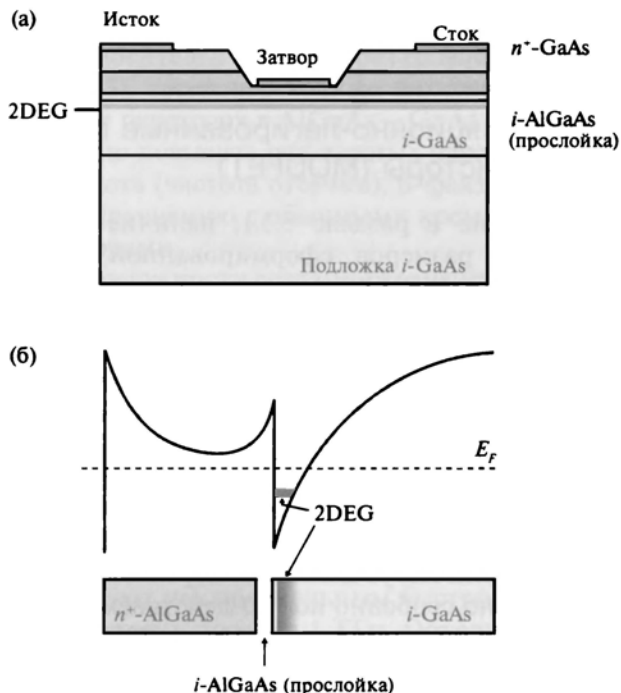


Рис. 9.2. (а) Схема сечения типичного модулированно-легированного полевого транзистора (MODFET или HEMT); (б) схематическое строение зоны проводимости в направлении, перпендикулярном структуре

диаграмма (а точнее, структура зоны проводимости в направлении, перпендикулярном структуре) приведена на рис. 9.2, б. Наиболее характерной особенностью транзисторов этого типа выступает квантовая яма для электронов, образующаяся между n -легированным слоем полупроводника AlGaAs и слоем обычного, нелегированного GaAs. Читатель может вспомнить (раздел 5.3.1), что в квантовая потенциальная яма в гетероструктурах AlGaAs—GaAs формируется на поверхности раздела из-за того, что ширина запрещенной зоны AlGaAs ($E_g \approx 2$ эВ) значительно превышает ширину зоны в GaAs ($E_g = 1,41$ эВ). Обычно ширина такой квантовой ямы (приблизительно треугольной формы) составляет около 8 нм, т. е. является настолько тонкой, что электронный газ действительно может образовывать двумерную систему. На рис. 9.2, б показан только один энергетический

кий уровень. Прослойка из нелегированного AlGaAs вводится в структуру для того, чтобы еще больше удалить проводящий канал от слоя AlGaAs *n*-типа (где генерируются носители) и тем самым повысить подвижность электронов вследствие ослабления взаимодействия с ионизированными донорами. Типичная ширина такой прослойки составляет около 50 Å.

Легко заметить, что показанная на рис. 9.2 структура MODFET или HEMT очень похожа на полевые МОП-транзисторы, которые рассматривались в разделе 5.2, у которых потенциальная яма для электронного канала также располагалась на поверхности раздела структуры Si—SiO₂. Обычный режим работы полевых транзисторов с высокой подвижностью носителей (HEMT) похож на режим полевых МОП-транзисторов, в которых поток электронов движется от истока к стоку под воздействием приложенного напряжения. Такой ток может модулироваться сигналом напряжения, подаваемым на затвор. Аналитическое выражение для зависимости тока насыщения $I_{D,sat}$ от напряжения на затворе V_G имеет вид: $I_{D,sat} \sim (V_G - V_T)^2$, где величина V_T — величина порогового напряжения. Поэтому вольт-амперные характеристики модулированно-легированных полевых транзисторов очень похожи на характеристики полевых МОП-транзисторов. Скорость переключения и высокочастотные характеристики таких транзисторов могут быть повышены за счет уменьшения времени пролета электронов t_f , для чего конструкторы таких устройств стремятся максимально сократить длину затвора L (которая обычно составляет около 100 нм), одновременно стараясь увеличить ширину затвора, поскольку это позволяет повысить величину сигнала и так называемую крутизну транзистора. Известно, что для изготовления полевых транзисторов с барьером Шоттки для достижения высокой крутизны необходимо использовать высоколегированные материалы (со уровнем легирования порядка 10^{18} – 10^{19} см⁻³), что ограничивает дрейфовую скорость электронов из-за рассеяния на большом числе примесных атомов. Таким образом, использование модулированно-легированных полевых транзисторов представляет конструкторам приборов дополнительные преимущества, так как в таких структурах транспорт носителей осуществляется в нелегированном слое (GaAs).

В настоящее время модулированно-легированные полевые транзисторы доминируют на рынке малошумящих приборов,

так как они способны работать в очень широком диапазоне частот: от микроволновых до частот около 100 ГГц (см. рис. 9.1). В новейших гетероструктурах систем AlGaAs—InGaAs—GaAs не только квантовая локализация электронов в ямах более эффективна, чем в гетеропереходах на основе AlGaAs—GaAs, но и электроны двигаются в слое InGaAs с более высокой дрейфовой скоростью насыщения, чем в GaAs. Крутизна такого транзистора достигает значений 100 мС/мм, частота отсечки составляет около 100 ГГц, а уровень шумов составляет лишь 2 дВ. Такие высокие характеристики достигаются за счет уменьшения расстояния затвор—канал (из-за более резких барьеров) и снижения паразитных емкостей системы. По всем этим причинам модулированно-легированные полевые транзисторы превосходят другие приборы при усилении сигналов в микроволновом диапазоне, вплоть до частот 300 ГГц, т. е. примерно в шесть раз превышают по быстродействию лучшие из транзисторов, изготовленных на основе МОП-технологий при заданном уровне литографического разрешения (см. рис. 9.1). Модулированно-легированные полевые транзисторы могут также изготавливаться на основе структур SiGe, однако такие устройства не выпускаются промышленно (подобно гетеропереходным биполярным транзисторам, описываемым в разделе 9.3) из-за относительно высоких значений токов утечки.

9.3. Биполярные транзисторы на гетеропереходах

Основной целью конструкторов гетеропереходных биполярных транзисторов является обеспечение максимального значения коэффициента усиления β при возможно более высоких рабочих частотах. Максимальная рабочая частота зависит от многих факторов, в число которых входят геометрические размеры и степень легирования областей эмиттера, базы и коллектора. Для повышения значений β необходимо, чтобы значения двух важных параметров системы (α именно коэффициент усиления по току α и коэффициент инжекции эмиттера γ) были максимально близки к единице (упомянутые параметры являются стандартными при описании биполярных транзисторов). Из этих требований сразу вытекает, что степень легирования эмиттера должна быть намного выше, чем базы. При этом, однако, следует учитывать,

что очень высокая степень легирования полупроводника уменьшает в нем ширину запрещенной зоны (см. раздел 3.6), например, при степени легирования 10^{20} см^{-3} ширина запрещенной зоны уменьшается на 14%, что приводит к уменьшению коэффициента инжекции носителей из области эмиттера в область базы. Поэтому почти сразу после изобретения биполярных транзисторов с однородными переходами Шокли предложил изготавливать эмиттер транзистора на основе полупроводников с более широкой запрещенной зоной, что должно было уменьшить число носителей, инжектируемых из базовой области в область эмиттера, и тем самым повысить общий коэффициент инжекции эмитора. Позднее, в 70-х годах началось коммерческое производство биполярных транзисторов на гетеропереходах (НВР).

На рис. 9.3, а показана разница, возникающая в зонной структуре при-транзисторов с гетеро- и гомопереходами. Следует особо отметить, что в последнем случае (рис. 9.3, б) ширина запрещенной зоны эмиттера превышает ширину зоны базы, вследствие чего барьер для инжекции электронов из эмиттера в базу (eV_n) оказывается ниже соответствующего значения для дырок (eV_p), что и проявляется в значительном повышении коэффициента β . Даже небольшое изменение высоты барьера может очень сильно влиять на процесс инжекции, который описывается квазиэкспоненциальной зависимостью от высоты барьера. Действительно, коэффициент β пропорционален отношению концен-

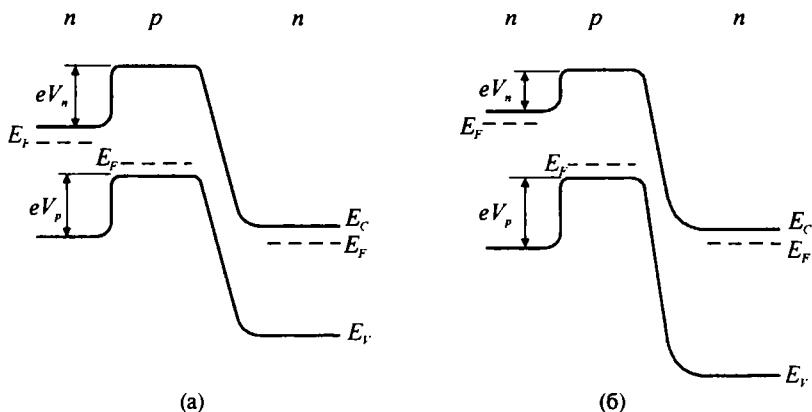


Рис. 9.3. Зонная структура при поляризации в активной зоне (а) транзистора на гомопереходе и (б) гетеропереходного биполярного транзистора (НВТ)

трации легирующей примеси в эмиттере и базе, а также члену $\exp(\Delta E_g/kT)$, где ΔE_g — разность между большей шириной запрещенной зоны в эмиттере и меньшей — в базы. При комнатных температурах (когда $kT \sim 0,026$ эВ) небольшая разница в значениях ΔE_g позволяет значительно изменить величину коэффициента β . Сказанное позволяет считать, что гетеропереходные биполярные транзисторы предоставляют богатые возможности для создания транзисторов с высокой степенью легирования базы, малым сопротивлением базы и малым временем пролета электронов через базовую область. Кроме того, можно даже уменьшать степень легирования базы, вследствие чего должна уменьшаться паразитная емкость, связанная с переходом эмиттер—база. Одновременное уменьшение сопротивления базы и емкости перехода эмиттер—база очень важно для повышения высокочастотных рабочих характеристик приборов на основе описываемых гетеропереходных биполярных транзисторов.

Другой важной особенностью гетеропереходов является возможность создания гетеропереходных биполярных транзисторов с базой переменного состава, в которых ширина запрещенной зоны постепенно уменьшается от эмиттера к коллектору (рис. 9.4, а). В такой системе создается внутреннее электрическое поле, позволяющее ускорять электроны при прохождении базовой области и тем самым дополнительно повышать быстродействие транзисторов. В предельном случае, когда область коллектора такого транзистора также изготовлена из полупроводника с широкой запрещенной зоной (как показано на рис. 9.4, б), пробивное напряжение на переходе база—коллектор может быть значительно увеличено. Кроме того, такие структу-

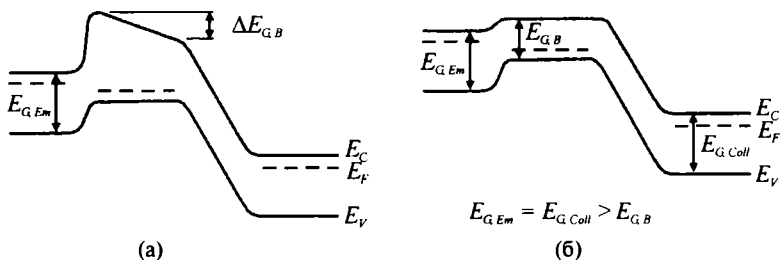


Рис. 9.4. (а) Гетеропереходный биполярный транзистор (НВТ) с базой переменной ширины запрещенной зоны; (б) двойной гетеропереходный биполярный транзистор (ДНВТ) с широкой запрещенной зоной полупроводника в области эмиттера и коллектора

ры (называемые двойными гетеропереходными биполярными транзисторами, ДНБТ) позволяют менять местами эмиттер и коллектор, что значительно расширяет возможности конструирования различных интегральных схем.

Гетеропереходные биполярные транзисторы (НБТ) обычно создаются на основе полупроводниковых соединений $A_{III}B_V$, что обусловлено хорошими характеристиками гетеропереходов в структурах $AlGaAs-GaAs$ и высокой подвижностью электронов. Типичные НБТ обычно имеют длину базы около 50 нм и являются высоколегированными (порядка 10^{19} см^{-3}). Такие транзисторы обычно имеют рабочую частоту около 100 ГГц, что значительно выше соответствующих параметров для кремниевых биполярных транзисторов. Дальнейшее повышение высокочастотных характеристик связано с использованием гетеропереходов в системах $InGaAs-InAlAs$ и $InGaAs-InP$, что позволяет получать рабочие частоты до ~ 200 ГГц. Очень ценной особенностью НБТ на основе полупроводников класса $A_{III}B_V$ является то, что они легко интегрируются в одну схему, включающую как электронные, так и оптоэлектронные приборы. На этой основе уже началось производство так называемых оптоэлектронных интегральных схем (ОEIC), включающих в себя полупроводниковые лазеры (см. главу 10), что представлялось невозможным в рамках привычных, кремниевых технологий.

Некоторые исследовательские проекты нацелены на создание гетеропереходных биполярных транзисторов на основе кремниевой технологии, что позволит использовать в них кремниевые полупроводниковые соединения с широкой запрещенной зоной. Одним из таких соединений является карбид кремния SiC (для которого ширина запрещенной зоны изменяется от 2,3 эВ для кубической модификации до более чем 3 эВ для гексагональных модификаций), а другим, весьма интересным для проектировщиков веществом выступает аморфный гидрогенизированный кремний (ширина зоны — 1,6 эВ). Техническая проблема при использовании этих материалов связана с высоким сопротивлением эмиттера, обусловленным либо свойствами самих материалов, либо металлическими контактами. По-видимому, наиболее перспективными кремниевыми материалами для получения НБТ являются сплавы на основе $SiGe$, в которых гетеропереходы могут быть сформированы вследствие того, что ширина запрещенной зоной в кремнии равна 1,12 эВ, а в гер-

мании — 0,66 эВ. Приборы с гетероструктурами Si—SiGe были созданы лишь в 1998 г. (т. е. значительно позднее приборов на GaAs и других соединениях класса $A_{III}B_V$), что легко объясняется недостаточным развитием методов эпитаксиального роста SiGe. Для изготовления НВТ на основе Si или SiGe необходимо, чтобы в создаваемой структуре сразу после кремниевой области эмиттера располагалась область базы SiGe, в которой ширина запрещенной зоны намного меньше, чем в Si, поскольку именно такая разница в ширине запрещенной зоны позволяет создавать в области базы относительно высокую концентрацию легирующих примесей, что и обеспечивает высокую рабочую частоту структуры, сравнимую с частотой приборов на основе соединений $A_{III}B_V$. Частота отсечки промышленно выпускаемых гетеропереходных биполярных транзисторов в настоящее время превышает 100 ГГц, а в опытных образцах — даже 400 ГГц. Такие высокие значения частоты отсечки частично связаны с использованием структур со сжимающими механическими напряжениями, что позволяет менять энергетическую структуру в напряженных слоях, в результате чего происходит уменьшение эффективной массы носителей. Повышение подвижности носителей при этом может достигать 60%.

Конечным результатом описанных приемов стало создание базовых областей с плавным изменением состава x в соединениях типа Ge_xSi_{1-x} . Наклон энергетической зоны, возникающий вследствие изменений ширины запрещенной зоны вдоль базовой области, обеспечивает очень высокие значения (вплоть до ~ 10 кВ/см) встроенного электрического поля, что и позволяет резко уменьшить время прохождения электронами базовой зоны. Такие гетеропереходные биполярные транзисторы (НВТ), конечно, потребляют и рассеивают гораздо больше энергии, чем полевые МОП-транзисторы, однако позволяют работать при гораздо более высоких частотах и при меньшем уровне шумов. Эти преимущества и позволяют рассматривать гетеропереходные биполярные транзисторы на основе SiGe в качестве весьма перспективных приборов.

9.4. Резонансный туннельный эффект

В предыдущем разделе было показано, что в гетеропереходах и квантовых ямах отклик электронов на приложенное элект-

рическое поле, направленное параллельно поверхности раздела, соответствует очень высокой подвижности. В этом разделе рассматривается отклик электронов на электрические поля, направленные перпендикулярно потенциальным барьерам на поверхностях раздела. В этом случае электроны могут, при соблюдении определенных условий, просто туннелировать через потенциальные барьеры, осуществляя так называемый перпендикулярный транспорт (см. раздел 6.3). Туннельные токи через гетеропереходы могут приводить к формированию областей с отрицательным дифференциальным сопротивлением (NDR) на вольт-амперной характеристике, для которых величина протекающего тока уменьшается с ростом прикладываемого напряжения. Этот эффект впервые был обнаружен Лео Эсаки еще в 1957 г. при изучении туннельных диодов с *p-n* переходами. В 1970 г. он (вместе с Тцу) предположил, что такой же эффект может наблюдаться в токах, протекающих через квантовые ямы, однако лишь к середине 1980-х г. развитие методов осаждения позволило вырастить структуры с гетеропереходами и квантовыми ямами, на основе которых удалось создать реальные устройства, в которых используется обсуждаемый эффект.

Действие электронных устройств на основе эффекта отрицательного дифференциального сопротивления (NDR) квантовых ям связано с так называемым эффектом резонансного туннелирования (RTE), наблюдаемым при прохождении электрического тока через структуру из двух тонких барьеров, между которыми располагается квантовая яма. Вольт-амперная характеристика (зависимость $I - V$) этих устройств похожа на характеристики туннельных диодов Эсаки. На рис. 9.5, *a* схематически представлена зона проводимости для двойного гетероперехода с квантовой ямой между переходами. Предполагается, что ширина квантовой ямы настолько мала (5–10 нм), что яма может содержать лишь один электронный уровень с энергией E_1 (резонансный уровень). Область ямы состоит из слаболегированного GaAs, окруженного слоями AlGaAs с большей шириной запрещенной зоны. Внешние слои состоят из сильнолегированного GaAs *n*-типа (n^+ GaAs), который обеспечивает электрические контакты. Уровень Ферми для n^+ GaAs располагается в зоне проводимости, поскольку этот материал может рассматриваться как вырожденный полупроводник (см. раздел 3.6).

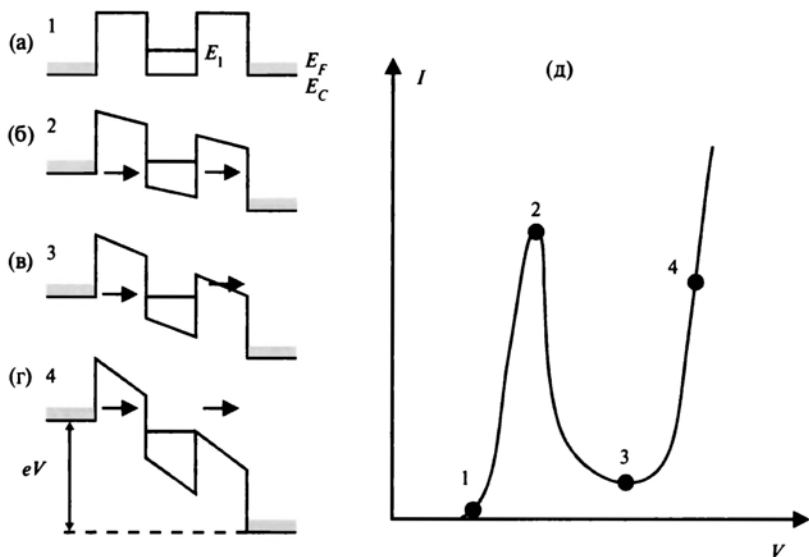


Рис. 9.5. Схематическое представление зоны проводимости резонансного туннельного диода: *а* — в отсутствие внешнего напряжения; *б* — *г* — при повышении приложенного напряжения; *д* — вольт-амперная характеристика системы

Рассмотрим поведение описываемой системы при повышении приложенного электрического напряжения V , начиная с напряжения 0 В. Можно ожидать, что при небольшом приложенном напряжении электроны будут туннелировать из зоны проводимости n^+ GaAs через потенциальный барьер, в результате чего увеличение напряжения должно приводить к возрастанию тока, что и демонстрирует участок 1—2 в области малых напряжений на вольт-амперной характеристике (рис. 9.5, *д*). При дальнейшем росте напряжения до значения $2E_1/e$ энергия электронов в n^+ GaAs в окрестности уровня Ферми совпадает с резонансным уровнем E_1 электронов внутри квантовой ямы, как показано на рис. 9.5, *б*). Такое совпадение соответствует резонансу, при котором коэффициент квантовой проницаемости барьера резко возрастает. Резонанс объясняется тем, что при этих условиях волновая функция электрона в яме когерентно отражается между двумя барьерами (этот эффект аналогичен оптическому отражению в резонаторах Фабри — Перо). При этом электронная волна, попадающая в структуру слева воз-

буждает резонансный уровень электронов в яме, повышая тем самым коэффициент прохождения электронов (а следовательно, и величину тока) через потенциальный барьер, что соответствует области точки 2 на вольт-амперной характеристике рис. 9.5, *д*. Возникающую при этом ситуацию можно сравнить с впрыскиванием электронов слева в квантовую яму и их дальнейшим освобождением через второй барьер. При дальнейшем повышении напряжения (рис. 9.5, *е*) резонансный уровень энергии в яме расположен ниже уровня Ферми в катоде, и ток начинает уменьшаться (область 3, рис. 9.5, *д*), в результате чего и возникает эффект отрицательного дифференциального сопротивления (NDR), соответствующий участку на вольт-амперной характеристике рис. 9.5, *д* в промежутке между точками 2 и 3. Затем, при дальнейшем повышении напряжения, ток через структуру начинает возрастать благодаря термоионной эмиссии через барьер (область 4 на рис. 9.5, *г* и промежуток между точками 3 и 4 на характеристике).

Именно на этом эффекте основано действие многих промышленно выпускаемых диодов с резонансным туннелированием (RTD), широко применяемых в микроволновой технике. Основной характеристикой, используемой для оценки рабочих параметров, выступает отношение токов пикового тока к минимальному току (PVCR) на вольт-амперной характеристике, т. е. отношение максимального тока (точка 2) к минимальному току на впадине (точка 3). Для обычных структур AlGaAs—GaAs при комнатных температурах это отношение составляет около 5, однако в структурах из напряженных слоев InAs, окруженных барьерами из материала AlAs, работающих при температуре жидкого азота, это отношение может быть доведено до 10.

Диод с резонансным туннелированием (RTD) можно представить в виде отрицательного сопротивления, соединенного с параллельной емкостью диода C и последовательным сопротивлением R_s (так же, как и в случае обычных диодов). Эта схема позволяет довольно легко продемонстрировать, что максимум рабочей частоты повышается при уменьшении C . Диод с резонансным туннелированием обычно изготавливается из низколегированных полупроводников, в результате чего возникает достаточно широкая область пространственного заряда между барьерами и областью коллектора, которой соответствует малая эквивалентная емкость. Вследствие этого рабочие частоты RTD

и могут достигать нескольких терагерц (ТГц), что значительно выше рабочих частот туннельных диодов Эсаки (порядка 100 ГГц, с временем отклика до 10^{-13} с). Низкие значения отрицательного дифференциального сопротивления, т. е. очень резкий спад после максимума на вольт-амперной характеристике, позволяет обеспечивать высокую частоту работы устройства, вследствие чего RTD являются единственными электронными приборами, способными функционировать на частотах порядка 1 ТГц, т. е. являются приборами с минимальным временем пролета электронов.

Вообще говоря, передаваемая от транзисторов RTD на внешнюю нагрузку мощность, достаточно мала, и их выходной импеданс также достаточно мал, вследствие чего такие транзисторы очень трудно согласовать в схемах с волноводами или антеннами. Выходной сигнал таких транзисторов обычно составляет лишь несколько милливатт, поскольку их выходное напряжение обычно меньше 0,3 В, что обусловлено значениями высоты барьеров и энергетических уровней в квантовых ямах. В настоящее время диоды с резонансным туннелированием очень часто используются для демонстрации различных возможностей их применения в разнообразных устройствах, включая статические запоминающие устройства с произвольным доступом (статические ОЗУ), генераторы импульсов, многозначные запоминающие устройства, многозначные и переключающиеся логические устройства, аналогово-цифровые преобразователи, осцилляторы, сдвиговые регистры, усилители с низким уровнем шумов, логические схемы типа MOBILE или нечеткой логики, умножители частоты, нейронные сети и т. п. В частности, особый интерес создателей различных логических схем привлекают устройства со значениями коэффициента PVCR (отношение максимального тока к току в долине) порядка 3 или выше, особенно в сочетании с высокими значениями плотности пикового тока J_p . Структуры со значениями PVCR порядка 3 и величиной J_p около нескольких A/m^2 представляются почти идеальными для создания многих типов запоминающих устройств, а высокие значения J_p и коэффициенты PVCR ~ 2 очень удобны для создания новых типов высокочастотных осцилляторов. В табл. 9.1 приводятся характерные значения параметров устройств такого вида, создаваемых на основе существующих полупроводниковых систем.

Табл. 9.1. Сравнительные характеристики транзисторов RTD, изготовленных из различных полупроводниковых систем

Материал	InGaAs	InAs	Si/SiGeя	GaAs	Si (диод Эсаки)
J_p , кА/см ²	460	370	282	250	151
PVCR	4	3,2	2,4	1,8	2,0
$\Delta I \Delta V$	5,4	9,4	43,0	4,0	1,1
R_D (Ω)	1,5	14,0	12,5	31,8	79,5
Площадь, мкм ²	16	1	25	5	2,2

В таблице приведены значения плотности пикового тока J_p , коэффициента PVCR (отношение максимального тока к току в долине); предельная мощность устройства, т. е. максимум произведения $\Delta I \Delta V$ в области отрицательного дифференциального сопротивления NRD (в предположении 100%-й эффективности) и значение отрицательного сопротивления диода RD в области NRD (по данным работы: *Paul D.J.*, *Semicond. Sci. Technol.*, 19, 2004, R75 – R108).

9.5. Транзисторы на горячих электронах

Энергия электронов, ускоренных сильным электрическим полем, может значительно превосходить энергию, соответствующую тепловому равновесию системы. В трехмерной системе энергию электронов в зоне проводимости легко описать некоторой электронной температурой T_e , воспользовавшись очень простым соотношением $E_k = (3/2)kT_e$, связывающим температуру со средней кинетической энергией электронов. Очевидно, что для двумерного электронного газа численный коэффициент в этом выражении должен равняться единице. В полупроводнике при тепловом равновесии электронная температура T_e должна совпадать с температурой кристалла T , однако в неравновесных системах (например, во внешних полях, способных ускорять электроны до очень высоких энергий) кинетическая энергия электронов может очень высокой, а следовательно, и электронная температура T_e может значительно превышать температуру кристаллической решетки. Такие электроны, далекие от состояния термодинамического равновесия с кристаллом, получили название горячих электронов.

Гетеропереходы между полупроводниками с различной шириной запрещенной зоны позволяют генерировать горячие электроны, приобретающие значительную кинетическую энергию из-за «разрыва» энергии ΔE_c зоны проводимости при переходе от полупроводника с широкой зоной к полупроводнику с узкой зоной. В частном случае гетеропереходов в системах AlGaAs—GaAs величина ΔE_c варьируется в диапазоне от 0,2 до 0,3 эВ, что примерно в десять раз выше значения $kT = 0,026$ эВ при комнатных температурах, и соответствует скорости носителей порядка 10^8 см/с и выше. Помимо этого, следует учитывать, что пересекающий область перехода пучок электронов ускоряется под действием электрического поля и коллимируется в направлении, перпендикулярном поверхности раздела. Этот эффект, называемый инжекцией электронов из гетероперехода, приводит к тому, что пучок электронов на выходе концентрируется в конусе с угловой апертурой около 10° .

Одним из методов отбора наиболее высокоэнергетических электронов из заданного распределения состоит в организации процесса пересечения ими потенциального барьера. Очевидно, что только наиболее энергетические электроны смогут преодолевать достаточно широкий барьер по механизму, подобному термоионному эффекту. Гораздо более эффективный метод инжекции горячих электронов заключается в формировании тонких потенциальных барьеров в зоне проводимости полупроводниковых структур, допускающих достаточно эффективное туннелирование. В этом случае можно организовать поток почти монохроматических (моноэнергетических) электронов.

Идея создания на этой основе транзисторов на горячих электронах (НЕТ) была предложена еще в начале 60-х годов Мидом (который проектировал транзисторы с высоким быстродействием), однако ее удалось осуществить лишь через несколько десятилетий, когда достаточно развились методы выращивания полупроводниковых структур с использованием молекулярно-пучковой эпитаксии (МВЕ), позволяющей создавать высококачественные гетеропереходы в системах AlGaAs—GaAs. На рис. 9.6, а, схематически показана типичная структура транзистора на горячих электронах [2], содержащая эмиттер из n^+ GaAs, очень тонкий барьер AlGaAs (~ 50 Å), базовую область GaAs (~ 1000 Å), еще один тонкий барьер AlGaAs (толщиной

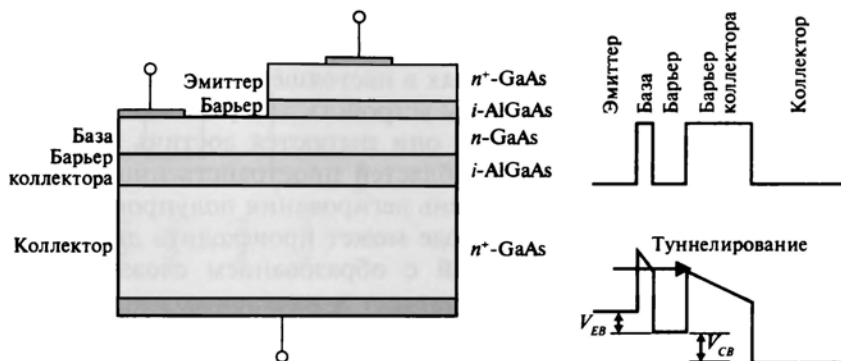


Рис. 9.6. Транзистор на горячих электронах по данным работы [2]: а — схема устройства; б — энергетическая диаграмма (для зоны проводимости) при положительном напряжении на коллекторе

около 3000 \AA) и коллектор из $n^+ \text{ GaAs}$. При подаче на коллектор положительного напряжения начинается инжекция горячих электронов из эмиттера путем туннелирования через тонкий барьер AlGaAs , поскольку на базе формируется положительное смещение относительно эмиттера (рис. 9.6, б). Следует отметить, что эффективная толщина барьера может модулироваться изменением разности напряжений между эмиттером и базой V_{BE} . Скорость инжектируемых электронов в этом конкретном случае составляет около $5 \times 10^8 \text{ см/с}$, что значительно выше, чем в любых других типах транзисторов, не говоря о том, что электронный пучок является коллимированным в конусе с углом около 6° . Усиление по току через базу (коэффициент α) при этом может быть доведено почти до единицы, если удастся одновременно максимально понизить рассеяние в области базы (которая обычно является очень узкой) и отражение от барьера коллектора. Время пролета электронов через базу для транзистора с приложенным напряжением может быть доведено до величины порядка нескольких десятков фемтосекунд, однако относительно большим остается время преодоления барьера коллектора. Сейчас ведутся поиски вариантов уменьшения времени преодоления барьера коллектора, однако его высота не может быть снижена, так как это приведет к возрастанию токов утечки. Практически можно предсказать, что в близком будущем полное время пролета будет составлять около 1 пикосекунды.

Как и в случае описанных ранее полевых транзисторов с высокой скоростью электронов (HEMT), разработчики транзисторов на горячих электронах в настоящее время настойчиво стараются уменьшить размеры устройств для уменьшения времени пролета. Прежде всего, они пытаются достичь этого за счет уменьшения толщины областей пространственного заряда, повышая, например, уровень легирования полупроводника. К сожалению, при этом подходе может происходить диффузия атомов легирующих примесей с образованием сложных химических соединений, что приведет к изменению химического состава материалов. Для преодоления таких проблем было предложено заменить вещество базы полупроводника на материал, подобный металлу, не загрязняющему структуру и не подверженному электромиграции. В результате были созданы так называемые транзисторы с металлической базой (МВТ), где базовая область сформирована из материалов типа силицида кобальта CoSi_2 (этот силицид обладает очень высокой проводимостью, сравнимой с проводимости металлов, и химически совместим с уже существующей кремниевой технологией). Общеизвестно, что рабочая скорость биполярных транзисторов ограничивается низкой подвижностью дырок, поэтому очевидным преимуществом МВТ является то, что такие транзисторы являются униполярными устройствами и могут работать при более высоких частотах.

На рис. 9.7 схематически представлены особенности двух наиболее известных структур МВТ. Слева (рис. 9.7, а) показана энергетическая структура устройства с гетероструктурой типа металл — оксид — металл — оксид — металл при положительном смещении между электродами эмиттер — база и база — коллектор. В этой ситуации туннелирующие электроны инжектируются через тонкий барьер в эмиттерном переходе. Справа (рис. 9.7, б) показана более простая схема транзистора с металлической базой, образованная гетероструктурой $\text{Si} - \text{CoSi}_2 - \text{Si}$, в которой первый переход Шоттки смещен в прямом направлении, вследствие чего электроны могут преодолевать эмиттерный барьер за счет термоионного эффекта. В качестве материала базы используется CoSi_2 , который хорошо согласуется с параметрами решетки кремния (что обеспечивает высокое качество поверхностей раздела) и одновременно обладает высокой стойкостью к электромиграции. В таких транзисторах горячие электроны,

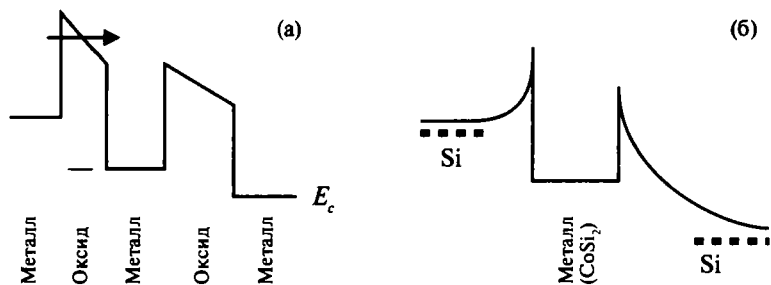


Рис. 9.7. Энергетическая диаграмма баллистических транзисторов: *a* – гетероструктура типа металл – оксид – металл – оксид – металл; *б* – гетероструктура Si – CoSi₂ – Si с металлической базой (силицид кобальта)

достигшие области базы, ведут себя подобно описанным выше баллистическим электронам, т. е. двигаются практически без рассеяния, поскольку длина их свободного пробега значительно превышает ширину базовой области.

9.6. Транзисторы с резонансным туннелированием

Описанные в разделе 9.4 диоды с резонансным туннельным эффектом (RTD, RTE) могут быть легко объединены со стандартными биполярными транзисторами, полевыми транзисторами и транзисторами на горячих электронах, что позволяет создавать более сложные устройства, получившие название транзисторов с резонансным туннелированием (RTT). Рассмотрение устройства таких транзисторов можно начать с биполярных транзисторов, в которых диоды с резонансным туннелированием добавлены к эмиттерному переходу. Так как туннельный резонансный ток в этом случае определяется напряжением, приложенным V_{EB} между эмиттером и базой, ток коллектора таких устройств очень похож на соответствующие характеристики RTD (рис. 9.8, *a*). На рис. 9.8, *б* представлена зависимость тока коллектора от напряжения V_{CE} , из чего видно, что выходная вольт-амперная характеристика ($V - I$) представляет собой чередование областей с положительной и отрицательной крутизной, переходом между которыми можно управлять путем изменения напряжения V_{EB} .

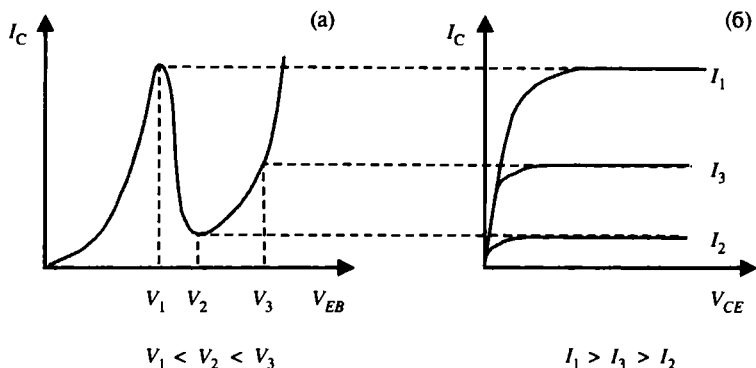


Рис. 9.8. Качественная зависимость тока коллектора в транзисторе с резонансным туннелированием от (а) напряжения база—эмиттер; (б) напряжения коллектор—эмиттер

На рис. 9.9 представлена диаграмма энергетических уровней так называемых транзисторов на горячих электронах с резонансным туннелированием со смещением в активной области [3]. В таких транзисторах между эмиттерной и базовой областью располагается гетероструктура с резонансным туннелированием, способная инжектировать значительные токи при выполнении условий резонанса. Положение резонансного уровня относительно эмиттера регулируется напряжением V_{BE} , подаваемым на базовую область, которое повышается до значений, соответствующих резонансному туннелированию, когда наблюдается максимум тока на выходе I_C . При дальнейшем повышении напряжения V_{BE} ток начинает уменьшаться и доходит до минимума при некотором значении V_2 , подобно тому как это происходит на вольт-амперной характеристике рис. 9.5 в разделе 9.4. Таким образом, выходные характеристики описываемого транзистора также имеют области отрицательного дифференциального сопротивления, но, в отличие от обычных транзисторов на горячих электронах (НЕМ), резонансные структуры позволяют инжектировать электроны в очень узком энергетическом диапазоне (~ 1 мэВ), что примерно в десять раз превышает соответствующие параметры для транзисторов с высокой подвижностью электронов (НЕМТ), описанных в разделе 9.2. Как упоминалось выше, диоды с резонансным туннелированием могут быть объединены различным образом с биполярными транзисторами. На рис. 9.10 представлена общая схема

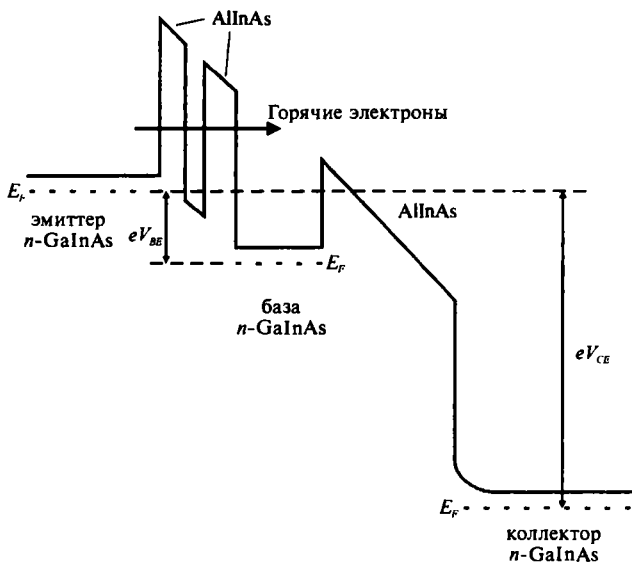


Рис. 9.9. Схематическая диаграмма энергетических уровней транзистора с резонансным туннелированием электронов со смещением в активной области по данным работы [3]

биполярного транзистора AlGaAs—GaAs, к выводу базы которого присоединен диод с резонансным туннелированием [4]. Квантовая яма между двумя потенциальными барьерами диода играет в данном случае особую роль, так как выше уже было показано, что она может иметь несколько квантовых уровней. Соответственно в коллекторном токе такого устройства могут проявляться несколько пиков, в отличие от единственного пика для структуры, представленной на рис. 9.5.

Транзисторы с резонансным туннелированием (РТТ) позволяют создавать много новых, практически важных устройств (особенно в области цифровой электроники), что представляется очевидным в свете описанных выше зависимостей тока коллектора I_C от напряжения. Например, можно представить схему из нескольких последовательно соединенных приборов с резонансным туннелированием, подключенных к сети с напряжением V и сопротивлением R . Точки пересечения кривой нагрузки с характеристической кривой туннельного прибора позволяют выбрать области стабильных состояний. Если таких областей две, можно образовать элемент бинарной логической

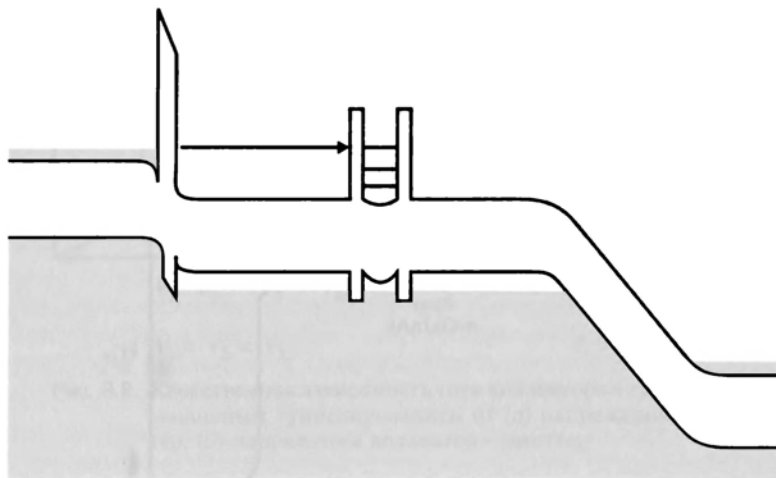


Рис. 9.10. Схема транзистора с резонансным туннелированием (RTT) и квантовой ямой в области базы (из работы [4])

схемы. Очевидно, что в случае квантовой ямы с несколькими энергетическими уровнями число стабильных точек будет равно числу пиков тока. Такие «дискретные» усилители позволяют создавать логические вентили, используя меньшее, чем обычно, число транзисторов описываемого типа. Например, на основе одного биполярного транзистора с резонансным туннелированием и двух обычных транзисторов можно создать схему суммирования, для построения которой обычно используется примерно 40 транзисторов. Это позволяет сразу резко повысить плотность монтажа и быстродействие создаваемых схем.

9.7. Одноэлектронные транзисторы

Одним из самых интересных направлений развития электроники является возможность создания устройств, в которых заряд в очень небольшой области может регулироваться с предельной точностью, т. е. путем добавления или выведения из области одного-единственного электрона. Эта область исследований непосредственно связана с созданием цифровых и аналоговых устройств, установлением метрологических стандартов, обработкой квантовой информации и т. п. Концепция *одноэлектронных транзисторов (SET)* связана, прежде всего, с поведением нуль-

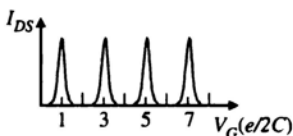
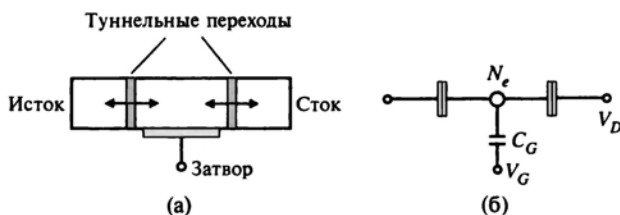
мерных (0D) нанометрических структур, типа квантовых точек, в которых электроны распределены по дискретной системе энергетических уровней. Одной из наиболее интересных особенностей таких структур, обусловленной квантованием энергетических уровней, является так называемый эффект кулоновской блокады, рассмотренный в разделе 6.4.3. Эффект связан с тем, что в очень малых областях проводящего материала (т. е. в так называемых «островковых» структурах) электростатический потенциал может существенно изменяться при добавлении или удалении из системы даже отдельных электронов.

В разделе 6.4.3 было показано, что правильная работа одноэлектронных транзисторов обеспечивается двумя факторами. Прежде всего, электрическая энергия, связанная с попаданием электрона в квантовую яму или вылетом из нее (эту величину можно называть энергией зарядки), должно значительно превышать значение kT системы, что в терминах емкости может быть записано как в уравнении (6.23): $C \ll e^2/kT$. Кроме этого, сопротивление R_T туннельного перехода должно быть достаточно большим по сравнению с определенным уравнением (6.25) квантовым сопротивлением $R_Q = h/e^2$, равным примерно $\sim 25,8$ кОм, что необходимо для исключения флуктуаций числа электронов в квантовой точке по принципу неопределенности Гейзенберга.

Прежде всего, следует учитывать, что до сих пор мы рассматривали квантовые точки в качестве двухполюсных устройств, однако для создания транзисторов на основе эффекта кулоновской блокады, очевидно, необходимо три контакта. Один из них должен использоваться в качестве затвора для управления потоком электронов через квантовую точку. Поэтому одноэлектронный транзистор должен содержать квантовую точку, связанную со стоком и истоком посредством туннельных переходов. Электрод затвора при этом должен быть отделен от квантовой точкой изолирующим материалом, чтобы электроны не могли туннелировать через этот барьер. Поскольку электрический ток в стоке и истоке структуры регулируется затвором, описанное трехполюсное устройство может работать в качестве транзистора, но не может использоваться для усиления сигналов. Пользуясь терминологией, используемой для контактов в полевых МОП-транзисторах, можно сказать, что квантовая точка в такой структуре играет роль области канала в полевом МОП-транзисторе.

На рис. 9.11, *а* представлена общая схема одноэлектронного транзистора, а на рис. 9.11, *б* — соответствующая ей эквивалентная электрическая цепь трехполюсного устройства, где квантовая точка с полным электрическим зарядом N_e (ее обычно называют кулоновским островом) подсоединена к стоку и истоку двумя туннельными барьерами. Число электронов в кулоновском острове при этом определяется внешним напряжением V_G через эквивалентную емкость затвора полупроводниковой структуры, добавленной для создания транзистора. В отличие от потенциальных барьеров стока и истока через этот контакт, туннельный ток отсутствует.

Вольт-амперная характеристика одноэлектронного транзистора может быть получена при измерениях, осуществляемых при непрерывно возрастающем напряжении V_G , подаваемом на электрод затвора. Это напряжение создает заряд GV_G на противоположной обкладке конденсатора, что компенсируется туннелированием одного электрона в квантовую точку, в результате чего возникает конкурентный механизм между индуцированным зарядом и дискретным зарядом, туннелирующим через барьеры. В результате такого процесса возникают так называемые кулоновские осцилляции, обусловленные прохождением дискретных зарядов через туннельный барьер, которые можно зарегистрировать, измеряя колебания тока I_{CB} между



(в)

Рис. 9.11. (а) Общая схема одноэлектронного транзистора (SET); (б) эквивалентная трехполюсная электрическая схема; (в) зависимость тока от напряжения на затворе

стоком и истоком (в виде функции от напряжения затвора), как показано на рис. 9.11, в. Отсутствие тока между двумя последовательными пиками означает, что число электронов в квантовой точке не изменяется, вследствие чего в системе и не возникает никаких потоков. Периодичность пиков напряжения ΔV на представленных зависимостях определяется изменением числа электронов, уже содержащихся в данной квантовой яме на единицу (т. е. $\Delta V = e/C$), поэтому емкость квантовой ямы может быть определена просто измерением разности напряжений между двумя соседними пиками.

С точки зрения специфических применений данного эффекта, можно сразу указать на возможность создания логических схем на основе одноэлектронных транзисторов, работающих в качестве инвертора и заменяющих КМОП-транзисторы. Схема инвертора на транзисторе описываемого типа представлена на рис. 9.12, а [5], где каждый туннельный барьер представлен эквивалентной схемой, состоящей из сопротивления с параллельно включенным конденсатором. Легко заметить, что выходное напряжение V_0 такого устройства представляет собой периодическую функцию от входного напряжения V_1 , как и показано на рис. 9.12, б. Периодичность выходного напряжения определяется отношением e/C_G , а амплитуда — величиной $e/(C_S + C_D + C_G)$. Усиление по напряжению такого транзистора $A_V = C_G/C_S$ может быть доведено примерно до 10, поскольку $C_G \sim 1$ fF, в то время как $C_S = C_D \sim 0,1$ fF. Практически коэффициент усиления в реальных устройствах может быть в настоящее время доведен примерно до 3 при низких температурах, но не близких к абсолютному нулю (теоретическое значение A_V достигается только при $T = 0$ К). Описанные выше инверторы уже практически используются в разнообразных логических схемах в качестве элементов памяти, составленных из двух инверторов, объединенных с полевым МОП-транзистором.

На основе получаемых в эксперименте характеристик проводимости разнообразных нанотрубок (напоминающих кулоновскую блокаду в металлических и полупроводниковых квантовых проволоках и точках) уже создаются транзисторы на полевом эффекте в углеродных нанотрубках, которые могут использоваться для производства запоминающих схем с накоплением заряда на основе одно- или несколько электронных систем. Полевые транзисторы с нанотрубками (*nanotube FETs*) отличаются

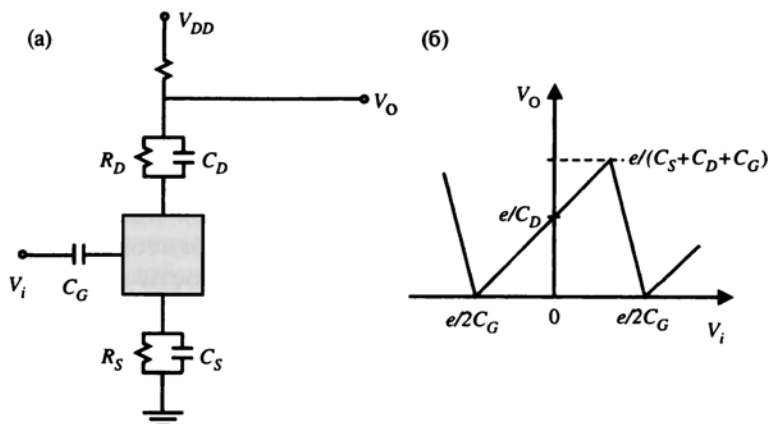


Рис. 9.12. (а) Схема одноэлектронного транзистора в виде инвертора; (б) идеальная характеристика переходного процесса по данным работы [5]

исключительно высокими характеристиками по подвижности носителей (около $10\,000\text{ см}^2/\text{В с}$), большой геометрической емкостью и большой крутизной. Кроме этого, последние исследования показали, что полевые транзисторы, создаваемые на базе так называемых одностенных углеродных нанотрубок, могут применяться и в логических схемах, обеспечивая при обычных температурах очень высокие значения важнейших параметров, а именно коэффициента усиления (выше 10) и отношения уровней тока в положениях «включено—выключено» (выше 10^5). Транзисторные схемы на углеродных нанотрубках могут не только выполнять дискретные логические операции, но и применяться в статических запоминающих схемах с произвольной выборкой, в логических устройствах «ИЛИ—НЕ» (NOR), кольцевых осцилляторах и т. п.

В заключение можно провести очень краткое сравнение одноэлектронных и полевых МОП-транзисторов. Именно полевые МОП-транзисторы чаще всего используются в современных полупроводниковых чипах, однако, как отмечалось в разделе 1.2, размеры этих устройств практически достигли своего предела. В принципе одноэлектронные транзисторы позволяют нам продолжить дальнейшую миниатюризацию чипов, не говоря уже о том, что они рассеивают меньше энергии, так как энергопотребление пропорционально числу электронов в выходном



токе устройства. Значительным недостатком одноэлектронных транзисторов является их высокий выходной импеданс, обусловленный сопротивлением туннельных барьеров. Еще один недостаток одноэлектронных транзисторов связан с размером квантовых точек, поскольку для работы при обычных температурах их емкость должна быть значительно снижена.

Литература

1. *Sze S.M.* (1991). *High Speed Devices* (Wiley New York)
2. *Capasso F.* (1986). In *High-Speed Electronics*, Eds. *Kallback B., Beneking H.* (Springer, Verlag, Berlin), 22, 50-61.
3. *Yokoyama N., Imamura K., Ohnishi H., Mori T., Muto S., Shibatomi A.* (1988). *Solid State Electronics*, 31, 577.
4. *Capasso F.* (1989). *JJJ rra*«5. *Electron Devices*, 36, 2065.
5. *Timp G., Howard R.E., Mankiewich P.M.* (1999). In *Nanotechnology*, Ed. *Timp, G.* (Springer, New York).

Дополнительная литература

Goser K., Glösekötter P., Dienstuhl J. (2004). *Nanoelectronics and Nanosystems* (Springer, Berlin).

Mitin V.V., Kochelap V.A., Stroschio M.A. (1999). *Quantum Heterostructures* (Cambridge University Press, Cambridge).

Modern Semiconductor Device Physics (1998). Ed. *Sze S.M.* (Wiley, New York).

Nanotechnology (1999). Ed. *Timp G.* (Springer, New York).

Задачи и упражнения

1. Полевые транзисторы с модулированным легированием (MODFET). Сравните значения подвижности носителей заряда, времен их пролета под затвором и максимальной рабочей частоты (см. рис. 9.1) для полевых транзисторов с модулированным легированием и полевых МОП-транзисторов. Объясните, почему в MODFET на структурах AlGaAs/GaAs предельная частота выше, чем у полевых транзисторов на GaAs с барьером Шоттки в качестве затвора. Исходя из данных, представленных на рис. 9.1, объясните более высокие значения характеристик транзисторов

на основе AlInAs/GaInAs по сравнению с транзисторами на основе AlGaAs/GaAs.

2. Биполярные транзисторы на гетероструктурах. Покажите, что коэффициент усиления по току β гетеропереходного биполярного транзистора (НВТ) пропорционален $\exp(\Delta E_g/kT)$ и имеет вид

$$\beta \propto \frac{N_d}{N_a} e^{\Delta E_g/kT},$$

где ΔE_g — разница между шириной запрещенных зон материалов эмиттера и базы; N_d и N_a — степени легирования эмиттера и базы соответственно. Рассчитайте приращение коэффициента β при комнатной температуре для стандартных значений $\Delta E_g \approx 0,2$ эВ. Объясните, почему даже в транзисторах с однородными переходами величина ΔE_g не равна нулю (Подсказка: степень легирования эмиттера обычно более чем на порядок величины превосходит степень легирования базы). Объясните, почему в гетеропереходном биполярном транзисторе уровень легирования эмиттера может быть снижен до более реалистичных значений одновременно с понижением сопротивления базы.

3. Транзисторы на горячих электронах. Рассмотрим транзистор с горячими электронами на основе структуры AlGaAs/GaAs, в котором электроны инжектируются из эмиттера в базу, как показано на рис. 9.6. Покажите, что скорость v_B попадающих в базу электронов определяется соотношением $v_B \approx (2eV_B/m^*)^{1/2}$, где eV_B — высота потенциального барьера эмиттера. Покажите также, что при этом электроны коллимируются, образуя узкий конус с характеристическим углом θ , равным $(kT/eV_B)^{1/2}$. Рассчитайте значения θ при низких температурах, полагая $eV_B \approx 0,3$ эВ.
4. Резонансный транзистор на горячих электронах. Покажите, как резонансный транзистор на горячих электронах (RHET) может быть использован для выполнения логической операции «исключающее ИЕ—ИЛИ». (Подсказка: рассмотрите еще раз представленную на рис. 9.9 общую схему RHET и представьте, как она будет работать в схеме с общим эмиттером, подобно биполярным транзисторам. Два бинарных входных сигнала (A и B) параллельно подаются на базу,

а результирующий сигнал логической операции C снимается непосредственно с коллектора.)

5. Перенос горячих электронов. Предположим, что представленная на рис. 6.5 структура образована квантовой ямой, окруженной двумя барьерами, материалы которых характеризуются подвижностью μ_w и μ_b и концентрациями n_w и n_b соответственно. Покажите, что при $\mu_w \gg \mu_b$ на вольт-амперной характеристике ($I - V$) наблюдаются области отрицательного дифференциального сопротивления (NDR) и колебания при высоких напряжениях электрического поля. (Подсказка: рассмотрите процесс выхода из ямы электронов, получивших от электрического поля достаточно энергии, чтобы считаться горячими.) Предположим, что в таком устройстве контакт стока разделен на два разных контакта, один из которых присоединен к яме, а второй — к барьеру. Покажите, как такое трехполюсное устройство может работать в качестве транзистора и что ток, протекающий в яме, демонстрирует заметный NDR и высокий коэффициент PVR (отношение максимального тока к току в долине).
6. Одноэлектронные транзисторы. Покажите, что переходная характеристика инвертора на основе одноэлектронного транзистора (рис. 9.12, *a*) должна иметь пилообразную форму (показанную на рис. 9.12, *b*), где параметр C_G представляет емкость между точкой и затвором, а параметр $C_\Sigma = C_S + C_D + C_G$ — сумму емкостей между истоком, стоком и затвором (предположите: $C_G = 8 C_S$, $C_S = C_D$ и $R_S = R_D \equiv R$). Объясните, почему период осцилляций напряжения затвора равен q/C_G , а максимальное усиление напряжения проявляется на спадающей части кривой и равно $A_v = -|C_S/C_G| = 8$, в то время как на восходящей части характеристики это значение составляет $C_G/(C_D + C_G)$.

ГЛАВА 10

ОПТОЭЛЕКТРОННЫЕ УСТРОЙСТВА НА ОСНОВЕ НАНОСТРУКТУР

10.1. Введение

В этой главе описываются оптоэлектронные и фотонные приборы, создаваемые на основе наноструктур, которые являются объектом интенсивного изучения и внедрения в течение последних двадцати лет, особенно в связи с развитием оптоволоконных средств связи. В настоящее время наблюдается сильная тенденция к замене электронных устройств на оптические, а также к объединению электронных и оптических устройств в единых структурах, получивших название оптоэлектронных интегральных схем (ОЕИС).

Вскоре после появления в начале 60-х годов полупроводниковых лазеров на гомопереходах стало ясно, что лазеры на двойных гетеропереходах (раздел 10.2) могут обеспечивать значительно более эффективную пространственную локализацию носителей заряда и фотонов, что приводит к меньшим значениям порогового тока и лучшей эффективности. Появление в конце 80-х годов полупроводниковых лазеров более эффективных и сложных типов (например, описанных в разделе 10.3 лазеров на квантовых ямах) позволило снизить плотность порогового тока до $\sim 50 \text{ А} \cdot \text{см}^2$, особенно в механически напряженных системах с квантовыми ямами (раздел 10.5). В разделе 10.4 описаны поверхностно излучающие лазеры с вертикальным резонатором (VCSEL) на квантовых ямах. Миллионы таких лазеров, излучающих с поверхности, могут быть объединены на одном чипе, что, естественно, создает множество возможностей их практического использования в дисплеях и системах обработки оптических сигналов.

В разделе 10.6 рассматриваются чрезвычайно перспективные лазеры (которые еще только начинают внедряться в промышленное производство), основанные на системах с размерностью меньше 2D, типа квантовых точек. Такие лазеры на системах с нулевой размерностью значительно превосходят лазеры на квантовых ямах по низким значениям порогового тока и другим эксплуатационным характеристикам. Проблема их производства состоит в том, что существующая литографическая технология не позволяет пока создавать в производственных условиях квантовые точки достаточно малого размера с достаточно однородным распределением. Большие надежды в этом направлении исследователи связывают с развитием новых технологий, типа самоорганизации нульмерных и одномерных систем. Два последних раздела главы (10.7 и 10.8) посвящены фотодетекторам и модуляторам на квантовых ямах, которые уже стали привычными устройствами для детектирования сигналов в длинноволновой области и высокоскоростной модуляции оптических сигналов соответственно.

10.2. Лазеры на полупроводниковых гетероструктурах

Прежде чем перейти к рассмотрению квантовых гетероструктур с низкой размерностью, представляется полезным очень кратко описать существующие полупроводниковые лазерные устройства. В разделе 3.7.5 уже отмечалось, что оптическое усиление в вырожденных полупроводниках может быть получено в результате индуцированного (вырожденного) излучения на переходах $p^+ - n^+$ в GaAs при прямом смещении. В таких структурах можно создавать активную зону с инверсной заселенностью, поскольку квазиуровни Ферми в вырожденных p^+ и n^+ материалах располагаются внутри зоны проводимости и валентной зоны соответственно. Непрерывность работы лазера обеспечивается постоянной инъекцией носителей в переход при прямом смещении. Лазеры, основанные на $p-n$ переходах из одного материала (например, в GaAs), имеют несколько недостатков, часть которых связана с недостаточной определенностью активной зоны излучения света, размеры которой примерно соответствуют диффузионной длине LD, т. е. составляют несколько микрометров (разделы 1.3 и 3.5.1). Кроме этого, следует учитывать, что в рассматриваемых системах достаточно велик поро-

говый ток, т. е. минимальный ток, требуемый для обеспечения работы лазера.

В 70-х годах стало ясно, что лазеры на двойных гетероструктурах (DHL), в которых обеспечивается и пространственная локализация носителей заряда и световых волн, не только значительно превосходят лазеры на гомопереходах по эффективности, но и имеют, по крайней мере на порядок, более низкую величину плотности порогового тока ($\sim 1000 \text{ А см}^{-2}$). Благодаря этим улучшенным свойствам лазеры на двойных гетероструктурах нашли множество применений в различных устройствах оптической связи. На рис. 10.1 приведены для сравнения общий вид структур полупроводниковых лазеров на гомопереходах и двойных гетероструктурах [1]. Как показано на рис. 10.1, а, гетеропереходы не только позволяют формировать потенциальные ямы для электронов и дырок, повышая концентрацию носителей, но и, что более важно, увеличивая инверсную заселенность электронов и дырок (см. раздел 3.7.5). Размеры активной области в лазерах на двойных гетероструктурах пока имеют размер порядка 0,1 мкм, что недостаточно мало для квантования энергии в потенциальных ямах. Лазеры на квантовых ямах будут рассмотрены в следующем разделе.

Еще один интересный аспект работы лазеров на двойных гетероструктурах связан с тем, что показатель преломления в GaAs несколько выше (примерно на 5%), чем в окружающем материале AlGaAs. Эта разница оказывается достаточной для обеспечения весьма эффективной оптической локализации. Коэффициент оптической локализации Γ , соответствующий доле плотности фотонов, локализованных в активной области лазера, определяется выражением

$$\Gamma = \frac{\int_{\text{act. region}} |E(z)|^2 dz}{\int_{-\infty}^{+\infty} |E(z)|^2 dz}, \quad (10.1)$$

где функция $E(z)$ описывает распределение амплитуды электромагнитной волны в направлении, перпендикулярном к поверхности раздела. В лазерах на двойных гетероструктурах значения коэффициента Γ могут достигать единицы. Описываемый эффект оптической локализации для лазеров на двойных гетероструктурах представлен на рис. 10.1, б.

Для повышения эффективности лазеров на двойных гетероструктурах, в большинстве случаев используется конфигурация

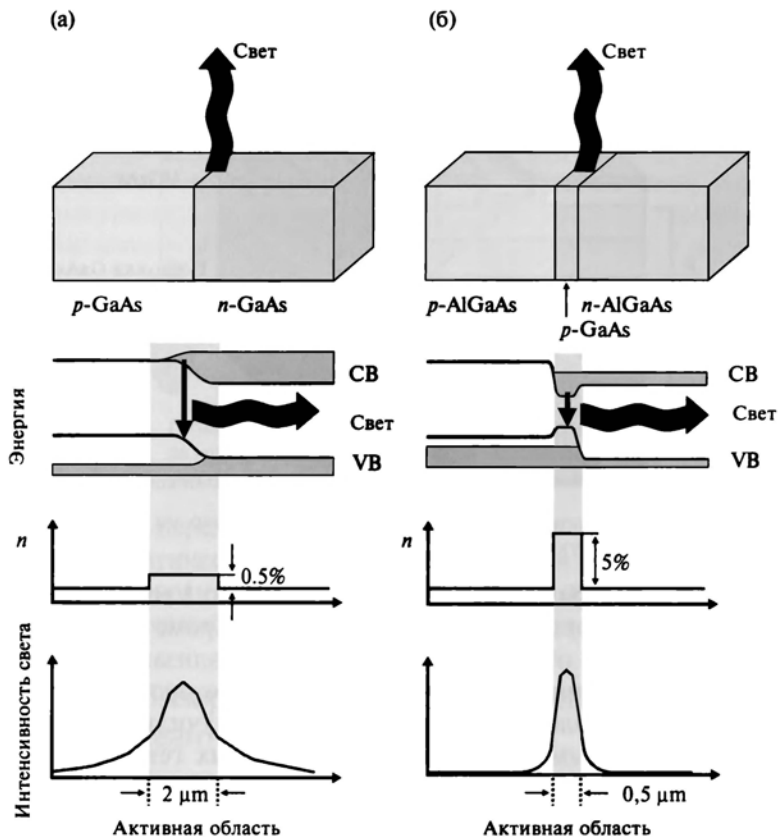


Рис. 10.1. Сравнение структуры и характеристик полупроводниковых лазеров на гомопереходах (а) и двойных гетероструктурах (б) (по данным работы [1]). Снизу вверх представлены: изображения полупроводниковых структур, формирующих лазер; энергетические диаграммы с указанием потенциальных ям для электронов и дырок; изменения показателей преломления в структурах; пространственная локализация оптических волн ограничения в активной зоне

с полосковой геометрией (рис. 10.2), в которой размеры активной зоны в поперечном или горизонтальном измерении (а следовательно, и значения порогового тока) значительно уменьшены. Благодаря особенностям формы активной зоны такие полосковые лазеры могут быть очень легко присоединены к другим устройствам, типа волокон, волноводов и т. п. Ширина активной области лазера

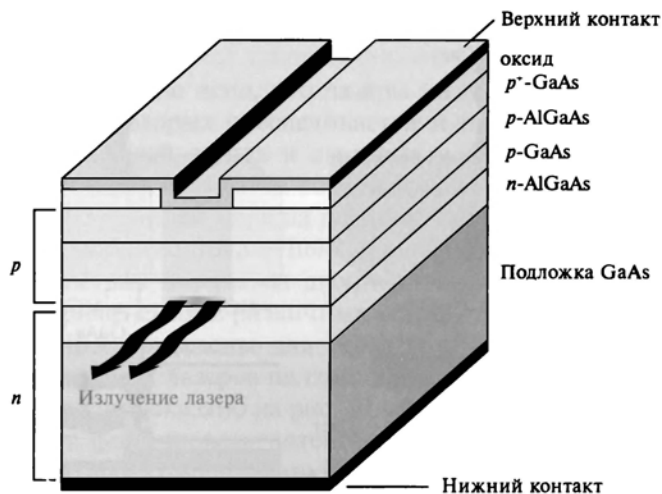


Рис. 10.2. Полосковый полупроводниковый лазер на двойной гетероструктуре

при такой геометрии может быть уменьшена до 1 мкм, вследствие чего пороговые токи могут достигать 10^{-2} А. Кроме этого, в такой геометрии можно обеспечить оптическую локализацию и в поперечном направлении, аналогично тому, как это происходит в вертикальном направлении. Такие структуры получили название лазеров с раздельным ограничением на двойных гетероструктурах (рис. 10.3). Ширина w оптического резонатора Фабри — Перо таких устройств может быть сделана достаточно малой, чтобы в ней возбуждались лишь низшие поперечные моды поля излучения, однако следует учитывать, что в действительности поперечные моды (например, TEM_{00}) включают несколько продольных мод, чье разделение по частоте зависит от длины резонатора.

Как показано на рис. 10.1, б, двойные гетероструктуры типа n -AlGaAs—GaAs (активная область)— p -AlGaAs—GaAs позволяют создавать потенциальные ямы, обеспечивающие очень высокую степень удержания носителей и оптического поля. Особо следует отметить, что потенциальные ямы для электронов и дырок возникают вследствие различия в ширине запрещенных зон полупроводников, образующих гетеропереходы, и в положении уровней Ферми. Несмотря на сужение запрещенной зоны в вырожденных полупроводниках (см. раздел 3.6), длина волны излучения лазера на гетеропереходах AlGaAs—GaAs все еще остается слишком

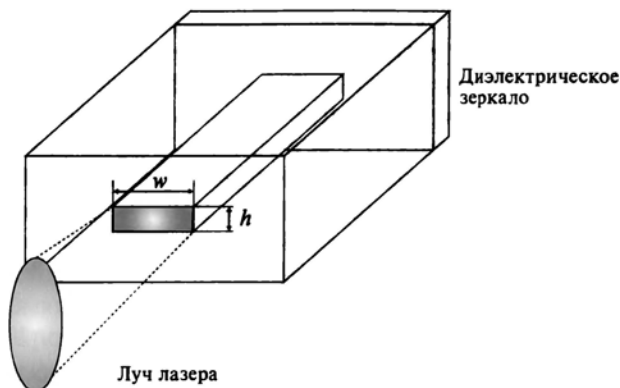


Рис. 10.3. Геометрия лазера с раздельным ограничением (в отличие от лазера с полосковой геометрией, структура позволяет регулировать оптическую локализацию в поперечном направлении)

короткой для передачи по оптическим волокнам, вследствие чего для систем оптической связи рассматриваются четырехкомпонентные полупроводники InGaAsP на подложке из InP (см. рис. 4.9) для использования «оптических окон» волокна (1,3 и 1,55 мкм).

10.3. Лазеры на полупроводниковых квантовых ямах

Описанные в предыдущем разделе лазеры на двойных гетероструктурах являются весьма перспективными устройствами в области оптических средств связи благодаря высокой эффективности, низкому значению порогового тока и возможности модуляции в широком диапазоне частот. В поисках возможностей дальнейшего улучшения характеристик (особенно получения очень узкого спектра излучения и возможности перестройки длины волны излучения) в конце 70-х годов наметился переход к созданию *лазеров на квантовых ямах (QW)*. Улучшения характеристик таких лазеров, как будет показано в этом разделе, удалось добиться за счет особенностей функции плотности состояний в двумерных системах и характеристик квантовых ям. Одним из недостатков описанных выше лазеров на двойных гетероструктурах является то, что в них удержание носителей и так называемое волноводное распространение света происходят в одной и той же области полупроводника. На рис. 10.4, а и 10.4, б представлена широко исполь-

зуемая структура, позволяющая пространственно разделить эти области. Такие структуры могут быть получены, например, постепенным изменением состава полупроводникового соединения $\text{Al}_x\text{Ga}_{1-x}\text{As}$ от значения $x = 0$ до 0,3, что позволяет изменить ширину запрещенной зоны от 1,41 до 2 эВ. В структурах с разделением областей локализации (типа показанной на рис. 10.4, а) квантовые ямы с шириной d порядка 5–10 нм погружены в более широкий резонатор с шириной w , оптимизированный для повышения эффективности волноводного эффекта. Волноводный эффект может быть усилен путем пространственного изменения коэффициента преломления (как показано в нижней части рис. 10.4, б), созданием так называемых гетероструктур с разделенным ограничением с изменяющимся коэффициентом преломления (GRIN-SCH). Очень часто, для усиления лазерного сигнала, вместо одной квантовой ямы используются структуры, содержащие множественные квантовые ямы, как показано на рис. 10.4, в.

Рассмотрим систему с одной квантовой ямой. В разделе 3.7.5 для объемного полупроводника уже было получено условие ге-

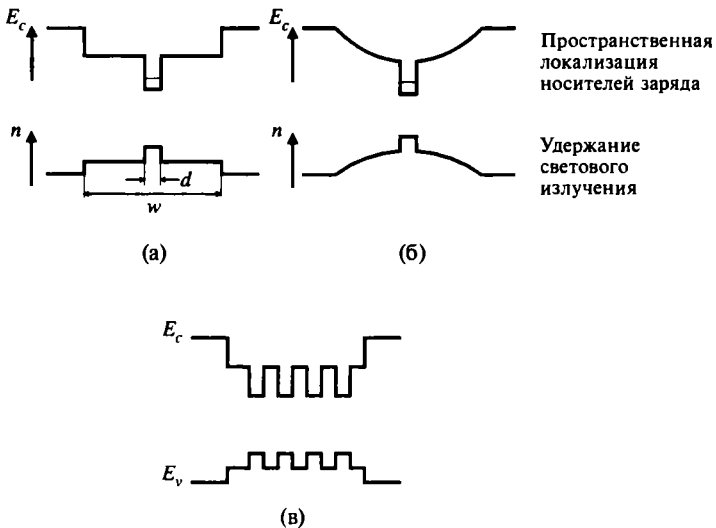


Рис. 10.4. Структуры с пространственно разделенными областями локализации и квантовыми ямами внутри оптических резонаторов: а — профиль зоны проводимости и коэффициента преломления; б — структура GRIN-SCH; в — гетероструктура с пространственно разделенными областями локализации с множественными квантовыми ямами

нерации когерентного излучения в форме $E_g < h\nu < E_{Fc} - E_{Fh}$, демонстрирующее, что энергия излучаемых фотонов должна лежать в некотором интервале, нижняя граница которого соответствует ширине запрещенной зоны, а верхняя — разности между квазиуровнями Ферми для двух разных вырожденных полупроводников, образующих данный гетеропереход. Зависимость коэффициента усиления от энергии $h\nu$ для описываемой системы представлена на рис. 3.18, б.

Рассмотрим далее коэффициент усиления в двумерной системе и сравним его со значением для обычного объемного полупроводника, предполагая, что генерация излучения происходит при переходах между уровнями $n = 1$ для электронов и дырок. Представляется очевидным, что усиление в двумерной системе должно начинаться при энергии $E_{g1} = E_g + E_{cl} + E_{vl}$ (как показано на рис. 10.5), в отличие от трехмерного случая, где оно начинается при меньших энергиях и нарастает медленнее. Такое поведение объясняется тем, что плотность электронных состояний в объемном кристалле при значениях энергии выше E_g растет пропорционально квадратному корню из значения энергии, в то время как в двумерной системе плотность состояний n_{2D} представляет собой ступенчатую функцию, показанную на рис. 10.5, а. Рассматривается только первая «ступенька», поскольку вторая соответствует очень высоким энергиям. Важно отметить, что в двумерных системах скачок функции происходит почти сразу (т. е. почти с бесконечной крутизной), в отличие от постепенного возрастания в 3D-системах. Спад зависимости коэффициента усиления от энергии (рис. рис. 10.5, в) связан с поведением распределения Ферми, представленного на рис. 10.5, б. Очевидно, что в предельном случае $T = 0$ кривая на рис. 10.5, б должна скачком переходить от $+1$ к -1 при значении энергии $(E_{Fn} - E_{Fp})$ и оставаться постоянной в диапазоне от $(E_g + E_{cl} + E_{vl})$ до $(E_{Fn} - E_{Fp})$. Сам факт такой «концентрации» коэффициента усиления в столь узком диапазоне энергий (по сравнению с трехмерными системами) уже подразумевает возможность создания в лазерах на квантовых ямах инверсионной заселенности при очень малых значениях тока инжекции (по сравнению с лазерами на двойных гетеропереходах).

Давайте наметим последовательность расчета коэффициента как функции от энергии фотонов при различных концентрациях носителей, при условии, что заполненными являются только одна из подзон зоны проводимости и одна из подзон валентной зоны.

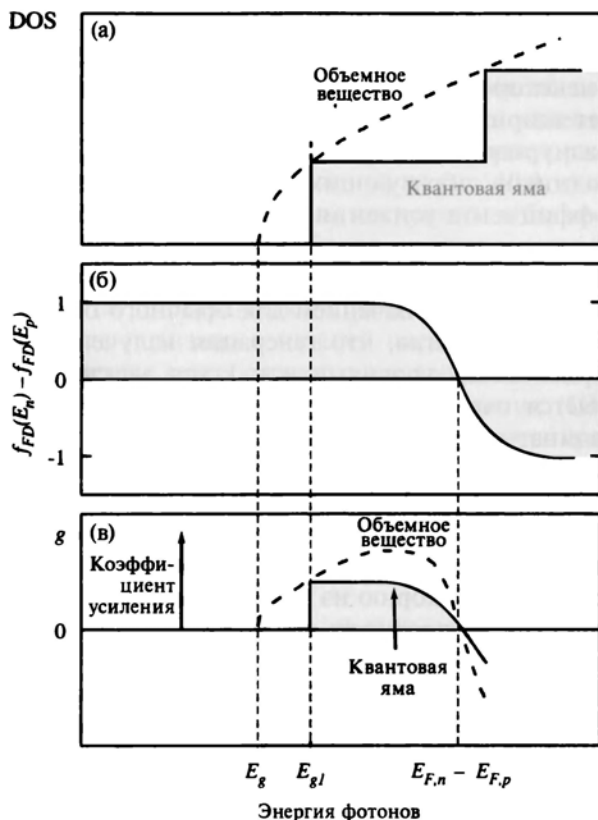


Рис. 10.5. Зависимость коэффициентов усиления для двумерных систем (2D) на квантовых ямах и объемных (3D) полупроводников: *a* — Функция плотности состояний; *б* — коэффициенты вероятности заполнения состояний; *в* — коэффициенты усиления

В этом случае квазиуровни Ферми могут быть вычислены как по концентрации инжектированных электронов и дырок, так и по матричным элементам межзонных переходов. Далее, коэффициент усиления может быть связан с плотностью тока J введением времени рекомбинации τ , поскольку плотность тока $J = end \tau^{-1}$, где d — толщина активной зоны, а параметр τ может быть определен из измерений скорости излучательной рекомбинации. При этом должна быть учтена скорость безызлучательной рекомбинации, а также все остальные потери, такие, как ток утечки. На рис. 10.6,

a представлены значения коэффициента усиления, полученные в работе [2]. Для структур с множественными квантовыми ямами (MQW) с количеством квантовых ям n_w и коэффициентом усиления в каждой яме g_w , полный коэффициент усиления, равный $n_w g_w$, представлен на рис. 10.6, *б* в зависимости от полного тока инжекции nJ для $n_w = 1, 2, 3$ и 4. Можно отметить, что коэффициенты усиления становятся положительными практически через равные промежутки (в зависимости от плотности тока), пропорциональные числам n . Интересно также сравнить зависимости для одиночных и множественных квантовых ям (SQW и MQW соответственно). Понятно, что при необходимости работать с очень низкими значениями порогового тока J_{th} следует применять структуры с оди-

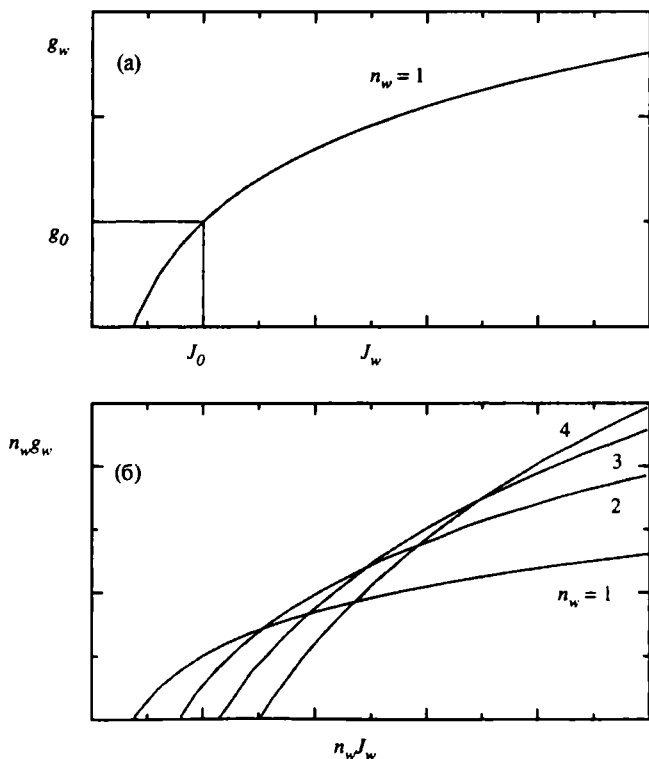


Рис. 10.6. Зависимость коэффициента усиления от плотности тока инжекции для (а) одиночной квантовой ямы и (б) структуры с множественными квантовыми ямами с числом ям $n = 1, 2, 3$ и 4 (по данным работы [2])

ночными квантовыми ямами, а при необходимости обеспечения высокого дифференциального усиления — с множественными.

Лазеры на квантовых ямах очень надежны в работе и (в отличие от лазеров на двойных гетеропереходах) не подвержены деградации зеркальных граней, вызываемой высокой скоростью рекомбинации в активной зоне. По сравнению с лазерами на двойных гетеропереходах, лазеры на множественных квантовых ямах обладают высокой эффективностью и имеют меньшие внутренние потери. По-видимому, их основным преимуществом, при использовании в высокоскоростных устройствах оптической связи, является их высокое дифференциальное усиление (определяемого, как упоминалось выше, производной $G = dg/dJ$), которое в системах с множественными квантовыми ямами может достигать величины $10 \text{ см}^{-1} \text{ мА}^{-1}$, что примерно на один порядок превышает соответствующие значения для лазеров на двойных гетеропереходах. Естественно, обеспечение надежной высокоскоростной работы таких устройств требует соответствующего проектирования гетероструктур, обеспечивающих оптимальное разделение областей локализации. Одним из решений проблемы может служить создание встроенных электрических полей (см. раздел 9.3, посвященный транзисторам на гетероструктурах), что позволяет значительно снизить эффективное время захвата и освобождения носителей заряда внутри областей, обеспечивающих пространственную локализацию носителей заряда и светового излучения. Граница модуляции таких лазеров за счет инжекции токов может быть доведена до 30 ГГц.

10.4. Поверхностные лазеры с вертикальным резонатором (VCSEL)

Наиболее важной особенностью поверхностных лазеров с вертикальным резонатором (VCSEL) является то, что они излучают свет в направлении, перпендикулярном к гетеропереходу. Такая геометрия сразу создает несколько очевидных преимуществ для использования этих структур в практических устройствах, поскольку она позволяет легко проверять качество элементов на пластинах перед компоновкой, создавать большие массивы светоизлучающих структур (вплоть до миллиона на одном чипе), подсоединять устройства к волоконным выводам и обеспечивать оптическую связь между чипами.

На рис. 10.7 представлена геометрия поверхностного лазера с вертикальным резонатором (VCSEL), состоящая из вертикального резонатора (расположенного вдоль направления протекания тока, а не перпендикулярно к направлению тока!), в котором поверхность активной области имеет очень небольшие размеры, вследствие чего свет излучается в основном с торца резонатора, а не с его боковых поверхностей. Сверху и снизу активный слой закрыт очень эффективными отражателями, в качестве которых используются диэлектрические зеркала, изготовленные из чередующихся слоев с высоким и низким значениями коэффициента преломления (например, слоев GaAs и AlGaAs) толщиной в четверть длины волны. Такие диэлектрические зеркала состоят из *распределенных брэгговских отражателей (DBR)*, имеющими очень высокую селективную отражательную способность для волн с длиной λ . При этом возникает так называемая конструктивная интерференция, определяемая соотношением

$$n_1 d_1 + n_2 d_2 = \frac{\lambda}{2}. \quad (10.2)$$

Отражательная способность (коэффициент отражения) брэгговских зеркал может быть доведен почти до 99% при использовании достаточно большого числа слоев (порядка 30). Высокий коэффициент отражения позволяет компенсировать низкое оп-

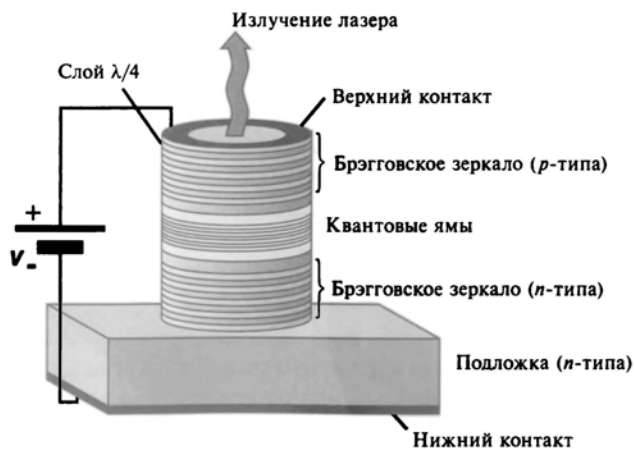


Рис. 10.7. Схема строения поверхностного лазера с вертикальным резонатором (VCSEL)

тическое усиление активной зоны, обусловленное малыми размерами резонатора. Латеральная локализация может быть получена путем создания мезоструктуры, показанной на рис. 10.7, с помощью травления верхней поверхности брэгговских зеркал DBR (в некоторых случаях травление может доходить до активной области). Очевидно, что инжекция зарядов через брэгговские зеркала может происходить лишь в случае, когда они изготовлены из полупроводниковых материалов. В случае диэлектрических материалов (например, оксиды титана или кремния) для инжекции зарядов в активную область можно использовать кольцевые контакты. Активная область обычно очень тонкая (порядка 100 нм) и содержит несколько квантовых ям, расположенных в максимуме стоячей волны, образованной двумя наборами зеркал. Область с квантовыми ямами окружена с двух сторон разделительными слоями (spacer) так, что возникает лазер с вертикальным резонатором длиной $L \approx \lambda/n$. При этом в активной области структуры возникает одна полная волна с постоянной амплитудой и затухающие волны в зеркалах, поскольку из-за большого коэффициента отражения в этой области распространение волн запрещено.

Важным отличием VCSEL от обычных «продольных» лазерных структур является малая длина резонатора, что приводит к существенному разделению оптических мод. Частотное разделение мод составляет около 10^{13} Гц, что сравнимо с частотой, соответствующей максимальному отражению DBR, и позволя-

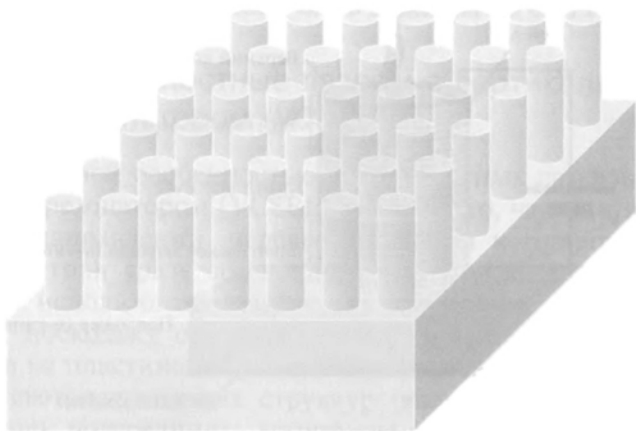


Рис. 10.8. Массив инжекционных микролазеров VCSEL

ет получать одномодовую генерацию (т. е. работать с одним типом колебаний). Напомним, что обычные излучающие с торца лазеры обычно на два порядка величины превосходят VCSEL по длине, что соответственно весьма ограничивает возможности разделения мод. Из-за малых размеров резонатора устройства VCSEL часто называют микролазерами. Массивы VCSEL, как показано на рис. 10.8, могут содержать более миллиона инжекционных микролазеров на одном микрочипе, что позволяет создавать мощные источники оптического излучения, которые уже нашли много интересных применений в вычислительной технике и устройствах связи.

10.5. Лазеры на напряженных структурах с квантовыми ямами

Самые первые лазеры на квантовых точках были созданы на основе гетероструктур с согласованными параметрами кристаллических решеток (например, на основе соединений AlGaAs и InGaAsP), и лишь через несколько лет удалось получить лазеры на механически напряженных кристаллических структурах, которые обладают рядом существенных преимуществ, особенно связанных с более низким значением порогового тока. Кроме того, оказалось, что механические напряжения кристаллической решетки создают новую возможность перестройки длины волны лазера, дополняющую возможности регулирования за счет изменения ширины квантовой ямы и высоты ее потенциального барьера. Лазеры на механически напряженных структурах с квантовыми ямами были предложены на рубеже 80—90-х годов и быстро начали развиваться. Как было показано в разделе 4.7, напряженность структур определяется различием параметров решеток используемых материалов. В частности, постоянная решетки α в тройных соединениях типа $\text{In}_{1-x}\text{Ga}_x\text{As}$, часто используемых в лазерах на квантовых точках, может быть вычислена линейной интерполяцией постоянных для решеток структур GaAs и InAs (см. рис. 4.9).

Одной из самых распространенных и изученных систем для лазеров на напряженных структурах с квантовыми ямами является GaAs—InGaAs—GaAs, в которой внутренний слой InGaAs испытывает сжимающие напряжения, что, как было показано в работе [3], приводит к важным изменениям в зонной струк-

туре материала (рис. 10.9). Помимо описанного в разделе 4.7 снятия вырождения между зонами для легких и тяжелых дырок при $\vec{k} = 0$, наличие напряжений значительно изменяет эффективную массу дырок и увеличивает ширину запрещенной зоны. При этом существенно уменьшается параллельная эффективная масса, что приводит к уменьшению порогового тока. Например, сжимающие механические нагрузки могут примерно втрое уменьшить параллельную эффективную массу дырок (для дырок,двигающихся в плоскости), в результате чего эта масса становится близкой к эффективной массе электронов, т. е. кривизна энергетических зон $E = E(k)$ становится почти

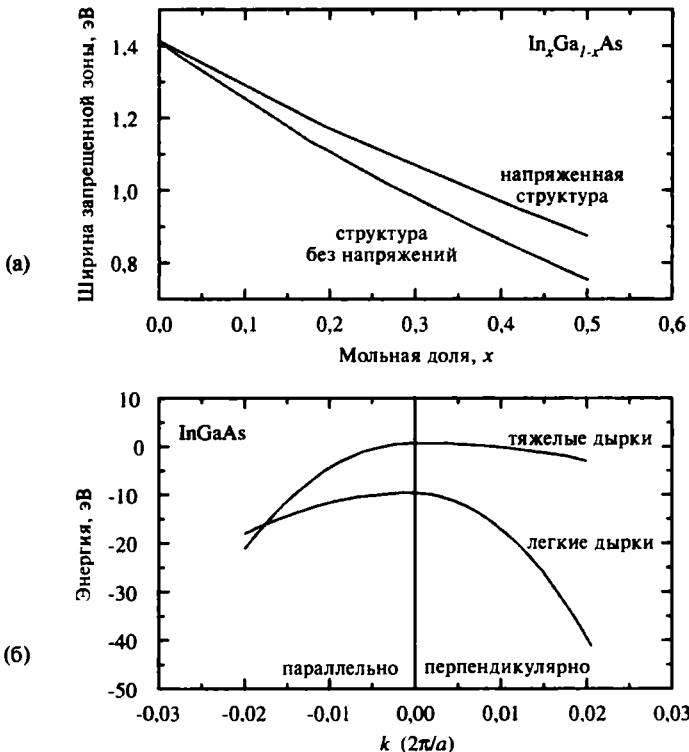


Рис. 10.9. Характеристики напряженных слоев InGaAs , помещенных между слоями GaAs : a — Зависимость ширины запрещенной зоны $\text{In}_x\text{Ga}_{1-x}\text{As}$ от состава; b — валентные зоны тяжелых и легких дырок в InGaAs при сжимающих нагрузках (по данным работы [3])

одинаковой для валентной зоны и зоны проводимости. Можно показать [2], что при этом повышается инверсия заселенности, т. е. коэффициент усиления возрастает с концентрацией инжектируемых носителей и генерация стимулированного излучения осуществляется при меньшем значении порогового тока.

Для создания лазеров на механически напряженных слоях с квантовыми ямами уже были использованы разнообразные гетероструктуры (InGaAs-GaAs , InGaAs-InP , InGaAsP-InGaAs , InGaAlP-GaInP и т. п.), что позволило перестраивать длину волны излучения в широких пределах (0,9—1,55 мкм). Дополнительным преимуществом лазеров с квантовыми ямами в напряженных системах (помимо отмеченного выше низкого значения порогового тока) выступает их высокая надежность, которую можно объяснить тем, что в напряженных полупроводниковых слоях распространение дефектов происходит медленнее.

Особенно тщательно роль напряженных слоев изучалась для структур с квантовыми ямами в системе InGaAs-InP , обладающей малой шириной запрещенной зоны и представляющей большой интерес для устройств волоконной оптики. Значения J_{th} в экспериментальных устройствах долгое время были значительно выше предсказываемых (даже в случае структур GRINSCH, описанных в разделе 10.3), однако эту проблему удалось решить, используя в таких системах напряженные активные слои, что позволило довести значения порогового тока до $J_{\text{th}} \approx 200 \text{ А см}^{-2}$. В этом случае влияние напряжений было связано с симметризацией зоны проводимости и валентной зоны. Для точной оценки величины напряжений в слое тройного соединения $\text{In}_{1-x}\text{Ga}_x\text{As}$ (на подложке InP) постоянная решетки $a(x)$ материала может быть представлена в виде линейной интерполяции постоянных решетки (в ангстремах) для материалов GaAs и InAs , т. е. в виде соотношения $a(x) = 5,6533x - 6,0584(1 - x)$. Если слой InGaAs выращивается на подложке InP (с постоянной решетки $5,8688 \text{ \AA}$), то изменением параметра x можно обеспечить в слоистой структуре как сжимающие, так и растягивающие напряжения. Например, при $x \approx 0,47$ гетеропереход не будет подвергаться никаким механическим напряжениям. В настоящее время физические особенности и возможные применения лазеров с квантовыми ямами в напряженных системах остаются объектом интенсивных исследований.

10.6. Лазеры на квантовых точках

Одним из наиболее перспективных применений полупроводниковых квантовых точек и проволок является создание лазерных диодов. Еще в 1976 г. было обнаружено, что повышение степени локализации носителей заряда значительно улучшает характеристики лазерных диодов по сравнению с их аналогами в объемных материалах. Из общих теоретических соображений ясно, что лазеры на идеальных объектах (квантовые точки и проволоки) должны иметь более высокие коэффициенты усиления, малые значения порогового тока, высокую стабильность работы и очень узкие низкие линии излучения, если сравнивать их с лазерами на двойных гетероструктурах или квантовых ямах. Естественно, реально создаваемые квантовые точки и проволоки не могут быть избавлены от дисперсии размеров (хотя бы и очень малых), дефектов, ограниченного захвата носителей и т. д., вследствие чего характеристики экспериментально создаваемых лазеров на квантовых точках и проволоках пока далеки от теоретических предсказаний. Однако технология выращивания квантовых структур непрерывно совершенствуется (особенно в отношении качества поверхностей раздела, однородности проволок, разработки оптических резонаторов с высокими коэффициентами локализации и т. п.), вследствие чего эта область исследований продолжает оставаться одной из самых интересных и перспективных.

Положение носителей в квантовых точках ограничено по всем трем измерениям в очень небольших пространственных областях, что приводит к квантовым эффектам в электронных свойствах вещества. Для применения в оптических приборах квантовые точки должны локализовать электроны в зоны проводимости и дырки валентной зоны в одной пространственной области. Функция плотности состояний для полупроводниковых квантовых точек характеризуется очень острыми пиками, соответствующими переходам между дискретными энергетическими состояниями электронов и дырок, причем вне этих уровней плотность становится совершенно незначительной. Во многих случаях электронная структура квантовых точек очень напоминает структуру для отдельных атомов, вследствие чего лазеры на квантовых точках во многом похожи на обычные ионные газовые лазеры, однако явным преимуществом лазеров на квантовых точках является то, что их электронная структура

может управляться путем проектирования структуры, подбора материала, а также размеров и формы точек.

Рассмотрим квантовую точку, размеры которой достаточно малы для того, чтобы разность между первыми двумя энергетическими уровнями (и для электронов, и для дырок) была намного больше тепловой энергии kT (т. е. $E_{2c} - E_{1c} \gg kT$ и $E_{2hh} - E_{1hh} \gg kT$). В такой системе все инжектируемые электроны и дырки будут занимать лишь низшие уровни, вследствие чего все такие инжектируемые электроны будут принимать участие в стимулированных переходах между уровнями E_{1c} и E_{1hh} , что позволяет снизить пороговый ток по отношению к любой системе с меньшей степенью локализации. На рис. 10.10 представлены значения плотности порогового тока для разнообразных лазерных структур за последние десятилетия по данным работы [4], из которого видно, что на лазерах с квантовыми точками уже достигнуты самые низкие значения порогового тока. Если тепловая энергия системы ниже разности первых и вторых уровней энергии, то линия излучения идеального лазера на квантовой точке должна быть исключительно острой и не зависеть от температуры, т. е. квантовые точки обеспечивают прекрасную температурную стабильность и не требуют охлаждения.

На рис. 10.11 на основе данных работы [5] представлена рассчитанная зависимость коэффициента усиления для идеальных систем разной размерности с квантовой локализацией. Кван-

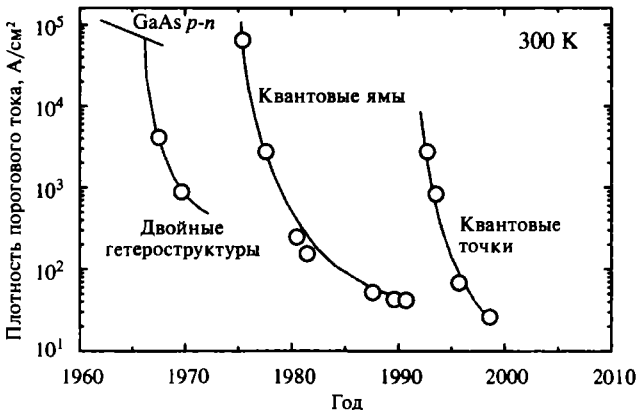


Рис. 10.10. Плотность порогового тока для лазерных структур с различной локализацией, достигнутая за последние десятилетия по данным работы [4]

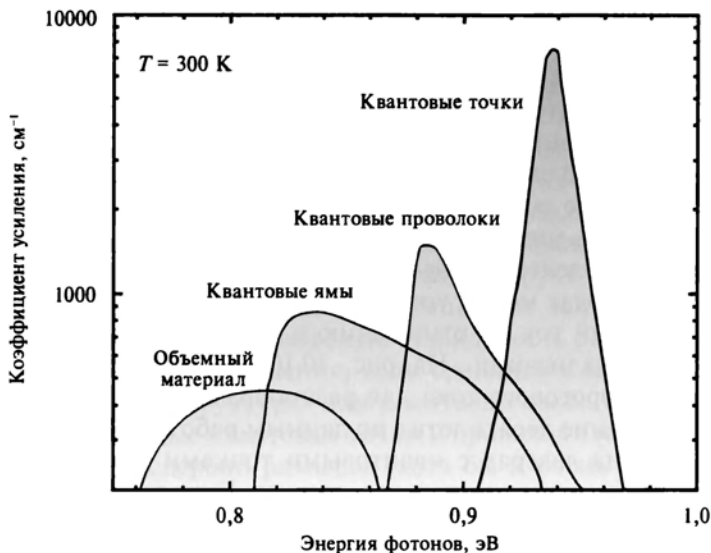


Рис. 10.11. Спектр коэффициентов усиления для лазеров на идеальных объемных полупроводниках, квантовых ямах, квантовых проволоках и квантовых точках (по данным работы [5])

Квантовые точки позволяют добиться максимальной остроты пиков спектра при самых высоких значениях коэффициента усиления. Следует также отметить, что, помимо этого, спектры квантовых точек отличаются симметричностью, что позволяет избежать нестабильности в случаях, когда длина волны лазерного излучения соответствует центру спектра. В системах реальных квантовых точек, конечно, существует некоторое распределение по размерам, приводящее к уширению линий спектра и снижению коэффициентов усиления, однако этот недостаток компенсируется возможностью регулирования спектральных характеристик за счет изменения размеров и т. д., что позволяет создавать лазеры на квантовых точках, перестраиваемые в широком диапазоне. Заинтересовавшийся читатель может детально ознакомиться с проблемами создания лазеров на квантовых точках по работе [6], содержащей обширную библиографию по данной тематике.

Несмотря на многообещающие перспективы лазеров на квантовых точках, их развитие пока сдерживается большими техническими сложностями создания требуемых массивов однородных по размеру и бездефектных квантовых точек. Тра-

диционные методы получения квантовых точек основаны на осаждении полупроводниковых соединений в стеклянных матрицах или на травлении предварительно сформированных эпитаксиальных слоев кристалла. Оба метода не позволяют создавать квантовые точки с достаточной плотностью и достаточным контролем их размера и формы. Более того, при этих методах часто возникают дополнительные дефекты в точках и множественные поверхностные состояния, приводящие к безызлучательной рекомбинации. Поэтому, возможно, настоящий прогресс в создании лазеров на квантовых точках будет достигнут лишь после того, как будут развиты методы создания квантовых точек на основе так называемой самоорганизации.

Наиболее известным из них является метод самоорганизации квантовых точек на поверхности раздела двух материалов с разными параметрами кристаллической решетки. При этом материал, такой, как InAs, выращивается химическим осаждением паров из газовой фазы, эпитаксией металлоорганических соединений из газовой фазы и молекулярно-пучковой эпитаксией на подложке из кристалла с большим значением постоянной кристаллической решетки и большей шириной запрещенной зоны (например, на GaAs). Несколько первых монослоев кристалла вырастают в планарном режиме, создавая достаточно большое механическое напряжение в кристаллической структуре и формируя так называемый зародышевый слой, однако после достижения некоторой критической толщины энергетически более выгодным становится формирование отровковых структур (так называемый режим Странского — Крастанова), показанных на рис. 8.5 (см. раздел 8.3.1). Такой режим приводит к появлению когерентных массивов пирамидальных квантовых точек на поверхности зародышевого слоя, в результате чего этот слой начинает эпитаксиально разрастаться на вершинах квантовых точек, создавая гетероструктуры с прекрасными характеристиками между двумя описанными монокристаллическими материалами, т. е. точками и окружающей их матрицей.

На рис. 10.12 схематически представлено устройство лазера с краевой эмиссией на самоорганизующихся квантовых точках. Структура состоит из нескольких слоев материалов, образующих рп-диод. Эти слои включают (снизу вверх на рисунке): подложку n -GaAs, слой n -AlGaAs, слой GaAs с собственной проводимостью, содержащий квантовые точки, слой p -AlGaAs и верхний

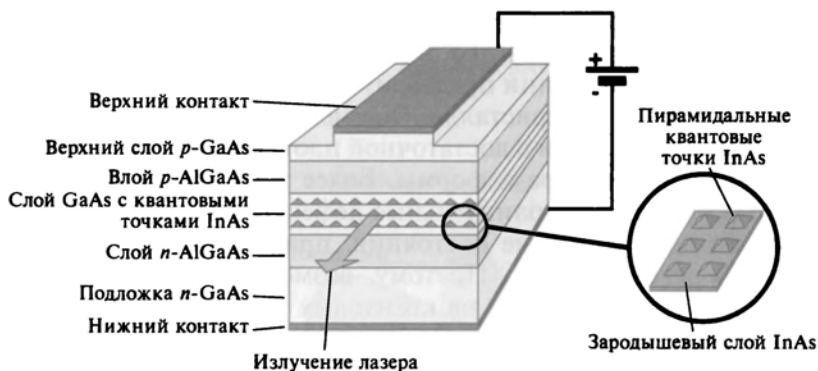


Рис. 10.12. Схема устройства лазера с краевым излучением на самоорганизованных квантовых точках. На вставке показан зародышевый слой с пирамидальными квантовыми точками

слой p -GaAs. Металлические контакты на подложке и верхнем слое соединяют структуру с внешней электрической цепью. При подаче прямого смещения электроны и дырки инжектируются во внутренний слой GaAs с собственной проводимостью или в активный слой, где они попадают в квантовые точки с меньшей запрещенной зоной, где и происходит рекомбинация. Длина волны излучения при этом соответствует межзонным переходам в квантовых точках InAs. Слой GaAs, заключенный между слоями AlGaAs с меньшим коэффициентом преломления, локализует излучение, что приводит к более эффективному его взаимодействию с носителями заряда. Зародышевый слой InAs повышает эффективность диффузии носителей в квантовые точки, поскольку его ширина запрещенной зоны меньше, чем в GaAs. Поскольку зародышевый слой является очень тонким, ширина запрещенной зоны для него превышает соответствующую ширину для квантовых точек, вследствие чего носители, попадающие в него из GaAs, быстро диффундируют внутрь квантовых точек. Для увеличения поверхностной плотности квантовых точек можно в принципе использовать несколько зародышевых слоев с пирамидальными квантовыми точками, выращиваемыми (или вырастающими) последовательно друг на друге внутри слоя GaAs, формируя массивы квантовых точек.

Самый первый лазер с резонатором Фабри — Перо с самоорганизующимися квантовыми точками был создан в 1994 г. на основе точек из InGaAs в матрице GaAs [7]. В настоящее время

лазеры на квантовых точках (с излучением в видимом и инфракрасном диапазонах) уже являются коммерческим продуктом.

Несмотря на богатые возможности улучшения параметров, описываемый класс лазеров уже продемонстрировал явные преимущества перед лазерами на квантовых ямах (например, более широкий спектр усиления позволяет использовать их в телекоммуникационных усилителях и процессах перестройки частоты излучения), не говоря уже об упоминавшейся выше температурной стабильности режима работы. И наконец, длина волны излучения существующих лазеров GaAs на квантовых ямах не соответствует «окнам прозрачности» оптоволокна. В требуемом диапазоне работают приборы на основе полупроводников InP, однако технология производства таких материалов пока недостаточно развита. С другой стороны, выращивание обогащенных индием квантовых ям в GaAs пока представляется сложной технической задачей вследствие большой разницы в постоянных кристаллической решетки. Преимуществом создаваемых сейчас лазеров на квантовых точках в структурах InGaAs/GaAs является то, что в них можно использовать широко развитую технологию применения подложек GaAs и в то же время длина волны их излучения попадает в диапазон пропускания оптоволокна, используемого в локальных сетях (1,3 мкм).

10.7. Фотодетекторы на квантовых ямах и сверхрешетках

10.7.1. Фотодетекторы на подзонах квантовых ям

В принципе квантовые ямы могут применяться для регистрации света в любой области частот, что связано с их оптическими характеристиками (см. раздел 8.2), однако именно в инфракрасной области (от 2 до 20 мкм) фотодетекторы на квантовых ямах представляют особый интерес, поскольку на них могут быть организованы системы ночного и теплового видения.

Основная проблема использования фотодиодов, основанных на межзонных переходах через запрещенную зону E_g в полупроводниках с p - n переходами, связана с тем, что малые значения E_g диктуют необходимость работы при очень низких температурах. Например, для полупроводников класса $A_{III}B_V$ (см. рис. 4.9) мы ограничены соединениями $InAs_{1-x}Sb_x$ с $x \approx 0,5$. Для детекторов

инфракрасного излучения можно использовать и полупроводники $A_{II}B_{VI}$ (например, $HgCdTe$), однако эти вещества являются очень мягкими, плохо поддаются обработке и имеют высокие значения темновых токов. Кроме этого, квантовые ямы представляются весьма удобными объектами для регистрации инфракрасного излучения, поскольку для них (как было показано в разделе 8.2) дипольные матричные элементы, соответствующие межподзонным оптическим переходам, очень велики. Кроме этого, очень важно, что длина волны излучения может быть легко перестроена за счет изменений параметров ямы в процессе изготовления (например, ширины). Особенную ценность для практических применений может иметь создание ИК-детекторов в области 8—12 мкм, поскольку именно в этом диапазоне лежит так называемое атмосферное окно, т. е. область пропускания волн, позволяющая осуществлять оптическую связь, связь со спутниками, использовать лазеры на CO_2 (10,6 мкм) и т. п.

На рис. 10.13 показаны разрешенные переходы с поглощением, удобные для детектирования ИК-излучения изолированными квантовыми ямами при воздействии внешнего электрического поля (естественно, что в практических устройствах используются не отдельные, а множественные квантовые ямы). В верхней части (рис. 10.13, а) представлены два энергетических уровня в каждой яме, причем второй уровень располагается близко к вершине барьеров. Расстояние между уровнями должно составлять 0,1—0,2 эВ, что для полупроводников $A_{III}B_V$ соответствует ширине ям около 10 нм. Напомним, что из-за правил отбора (см. раздел 8.2) поляризация падающего излучения должна быть параллельна направлению квантовой локализации. При облучении светом такие фотодетекторы генерируют ток благодаря туннелированию носителей из квантовых ям. В некоторых случаях наиболее эффективным методом регистрации представляется использование в таких устройствах поглощательных переходов между единственным энергетическим уровнем ямы и первой узкой непрерывной зоной вне ее (рис. 10.13, б). Для систем $AlGaAs - GaAs - AlGaAs$ эта разность энергий составляет около 0,12 эВ, вследствие чего область спектральной чувствительности расположена в районе 10 мкм. Преимущество использования описываемых структур связано с тем, что темновой ток фотодетектора оказывается меньше, чем в ранее описанных структурах, где носители покидают ямы благодаря туннелированию.

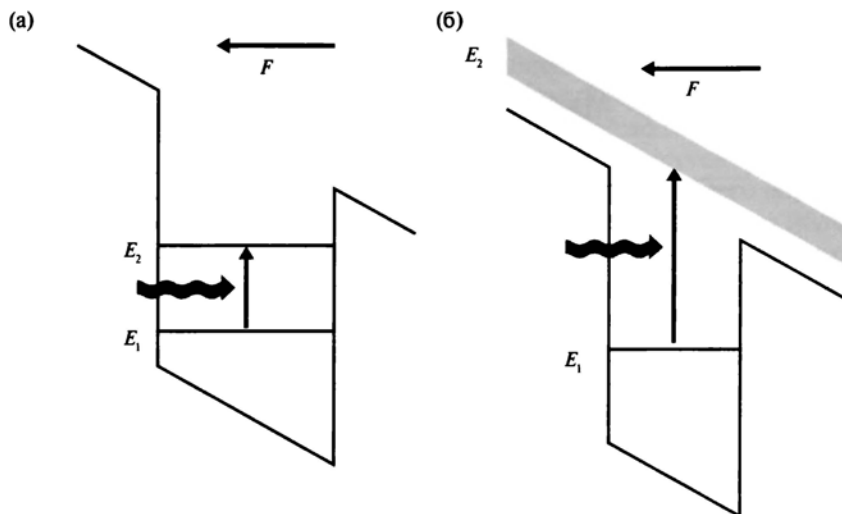


Рис. 10.13. Оптические переходы с поглощением при регистрации ИК-излучения квантовых ям: *а* — межподзонные переходы; *б* — переход между связанным состоянием и состоянием внутри узкой непрерывной зоны состояний вне потенциальных ям (F — приложенное электрическое поле)

10.7.2. Лавинные фотодетекторы на сверхрешетках

Известно, что при работе так называемых полупроводниковых лавинных фотодетекторов (APD) возникают очень сильные шумы, требующие применения специальных методов для их подавления. Уровень этих шумов может быть значительно снижен, если значение коэффициента лавинного умножения α для одного типа носителей (например, электронов) намного выше, чем для другого (дырок). В этом отношении кремний является очень важным полупроводниковым материалом для создания лавинных фотодетекторов (APD), так как в нем соотношение коэффициентов α_e/α_h равно примерно 30. Для любого конкретного типа полупроводников это отношение определяется энергетической зонной структурой.

С другой стороны, процесс создания квантовых ям позволяет варьировать соотношение коэффициентов α_e/α_h . Например, сверхрешетка или структура из множественных квантовых ям MQW может быть спроектирована таким образом, что в ней разрыв зоны проводимости ΔE_c будет значительно выше значений ΔE_v , соответствующих разрыву валентной зоны. В такой струк-

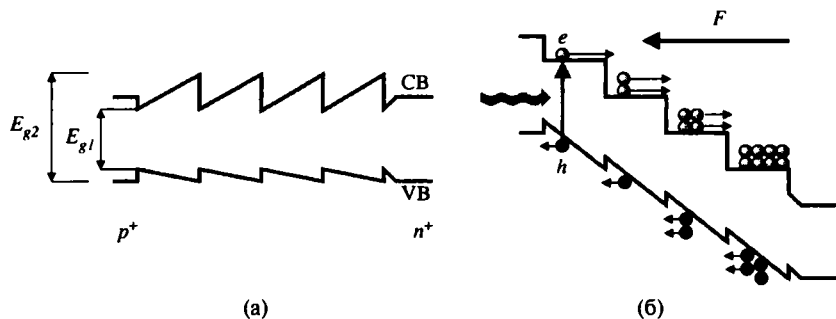


Рис. 10.14. Общая схема лавинных фотодетекторов на сверхрешетках: *a* — энергетическая диаграмма сверхрешетки ступенчатого типа; *б* — процесс нарастания электронной лавины в детекторе со смещением при попадании света

туре (при пересечении гетероперехода) электроны приобретают значительно большую кинетическую энергию, чем дырки. Такая же цель может быть достигнута при создании сверхрешеток в виде «ступенек» (рис. 10.14, *a*), где ширина запрещенной зоны меняется для каждой ямы (в этом случае электроны получают дополнительную кинетическую энергию ΔE_c перехода в каждую следующую яму). Наличие этой избыточной энергии приводит к весьма эффективным механизмам ударной ионизации, поэтому лавинное умножение легко достигается при наложении электрического поля F , как показано на рис. 10.14, *б*. В противоположность этому дырки приобретают в поле лишь незначительную энергию ΔE_v , недостаточную для начала ударной ионизации. Большинство сверхрешеток, используемых в лавинных фотодетекторах (APD), создано на основе полупроводниковых соединений класса $A_{III}B_V$, типа GaAs или InP. Стоит отметить, что внедрение упомянутых выше ступенчатых сверхрешеток осложняется трудностью их изготовления, поскольку производство требует тщательного контроля состава четверных соединений $A_{III}B_V$.

10.8. Модуляторы на квантовых ямах

Квантовые ямы обычно используются для непосредственной модуляции светового излучения, поскольку в них электрооптические эффекты проявляются значительно сильнее, чем в обычных объемных полупроводниковых материалах (в сущности, из-за слабости электрооптических эффектов в полупроводниках в мо-

дуляторах используются такие материалы, как ниобат лития). При рассмотрении экситонного поглощения (см. раздел 8.4) уже было показано, что из-за квантово-размерного эффекта Штарка (QCSE) при наложении внешнего электрического поля могут возникать очень значительные изменения оптического спектра поглощения квантовых ям. Из-за высокого энергетического барьера экситоны в таких наноструктурах ионизируются внешним полем гораздо слабее, чем в объемных структурах, т. е. «выдерживают» поля примерно до 10^5 В/см. Очень важным преимуществом модуляторов на квантовых ямах является то, что они легко совмещаются с разнообразными микроэлектронными устройствами.

Действие модуляторов электропоглощения основано на изменении коэффициентов оптического поглощения квантовых ям под воздействием внешнего электрического поля (см. раздел 8.4). На рис. 10.15 показана мезоструктура модулятора, работающего на подобном эффекте, сформированная путем травления. Очевидно, что эффективность действия таких устройств может быть значительно повышена при использовании множественных квантовых ям (MQW), т. е. структур или массивов из нескольких ям (с размерами от 5 до 10 нм) в материалах типа AlGaAs—GaAs—AlGaAs. Структура размещается между

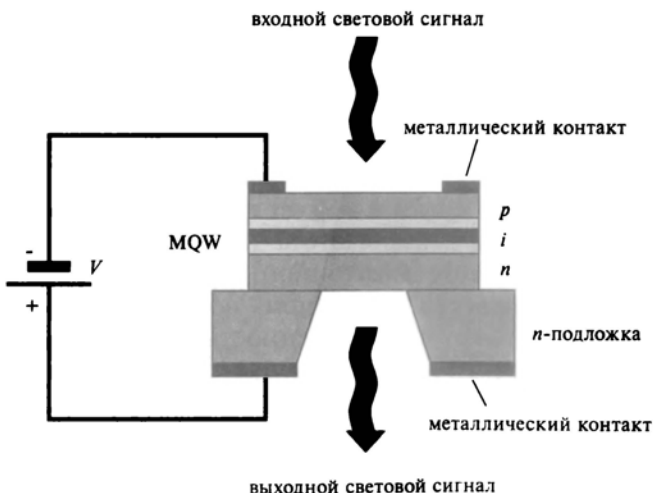


Рис. 10.15. Модулятор электропоглощения, получаемый мезоразмерным травлением на основе квантово-размерного эффекта Штарка

$p+$ и $n+$ областями обратно смещенного перехода. Поскольку толщина структур с множественными квантовыми ямами составляет около 0,5 мкм, достаточно малые обратные напряжения могут создавать электрические поля с напряженностью до 10^4 — 10^5 В/см, способные изменять край зоны экситонного поглощения на 0,01—0,05 эВ (см. рис. 8.12).

Модуляторы электропоглощения описанного типа позволяют модулировать световое излучение с очень большой скоростью. При этом так называемый коэффициент контрастности может быть доведен почти до значений ~ 100 за счет использования режима отражения (вместо режима пропускания). Технически это достигается введением дополнительного металлического слоя в подложке, что заставляет луч света проходить структуру два раза. Коэффициент модуляции может быть повышен при понижении температуры. Модуляторы электропоглощения обычно работают в диапазоне частот десятки гигагерц, но при высоких приложенных напряжениях частота может быть доведена до 100 ГГц. Верхняя граница рабочей частоты прибора определяется механизмом выведения носителей из квантовых ям. При небольшом внешнем поле, возникающие (вследствие поглощения) электрон-дырочные пары не могут покинуть пределы ямы, но при достаточно сильных внешних полях электроны и дырки туннелируют из ям с характерными временами порядка нескольких пикосекунд.

Еще один тип модуляторов на квантовых ямах также связан с использованием квантово-размерного эффекта Штарка (QCSE), рассмотренного в разделе 8.4, но он использует фотоны с энергией ниже края экситонного поглощения. В этом случае электрическое поле воздействует главным образом на коэффициент преломления среды и меняет фазу падающего излучения. Для повышения эффекта частота падающего фотона должна быть близка к частоте экситонного резонанса, однако не так близка, чтобы привести к заметному поглощению света. В симметричных квантовых ямах зависимость коэффициента преломления от напряженности приложенного поля является квадратичной, что напоминает известный электрооптический эффект (или эффект Керра) в объемных полупроводниковых материалах с объемно-центрированной кристаллической решеткой. Однако в структурах с квантовыми ямами значения соответствующего коэффициента примерно на два порядка выше, что позволяет существенно сократить путь, проходимый светом внутри

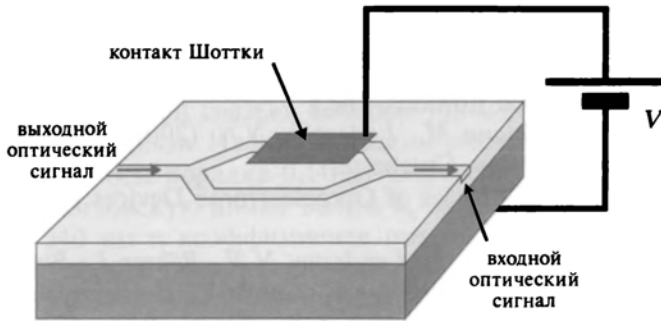


Рис. 10.16. Общая схема интерферометра Маха — Зендлера

материала. Устройства с длиной порядка сотен микрон могут быть при этом объединены в так называемые оптоэлектронные интегральные схемы (ОEIC), примером которых может служить показанный на рис. 10.16 интерферометр Маха — Зендлера, в котором входной сигнал из оптического волновода расщепляется на два луча одинаковой интенсивности, которые проходят в материале по разным каналам, а затем воссоединяются. При наложении электрического поля на одном из маршрутов меняется фаза проходящего сигнала, в результате чего при соединении лучей наблюдается интерференционная картина.

Литература

1. *Panish M.B., Hayashi I., Sumski S.* (1970) *Appl Phys. Lett*, 16, 326
2. *Chuang S.L.* (1995). *Physics of Optoelectronic Devices* (Wiley, New York).
3. *Zory P.S. Jr.* (1993). *Quantum Well Lasers* (Academic, Boston).
4. *Hatcher M.* (2000). *Opto-Laser Europe*, 74, 33—34.
5. *Asada M., Miyamoto Y., Suematsu Y.* (1986). *IEEE J. Quantum Electron*, QE-22, 1915.
6. *Bimberg D., Grundmann M., Ledentsov N.N.* (2001). *Quantum Dot Heterostructures* (Wiley, Chichester).
7. *Kirstaedter N., Ledentsov N.N., Grundmann M., Bimberg D., Ustinov V.M., Ruvimov S.S., Maximov M.V., Kop'ev P.S., Alferov Zh.L., Richter U., Werner P., Gösele U., Heydenreich J.* (1994). *Electron Lett*, 30, 1416.
8. *Singh J.* (1995). *Semiconductor Optoelectronics* (McGraw-Hill, International Editions).

Дополнительная литература

- Bhattacharya P.* (1994). *Semiconductor Optoelectronic Devices* (Prentice Hall International Editions, Inc.).
- Bimberg D., Grundmann M., Ledentsov N.N.* (2001). *Quantum Dot Heterostructures* (Wiley, Chichester).
- Chuang S.L.* (1995). *Physics of Optoelectronic Devices* (Wiley, New York).
- Grundmann M., Christen J., Ledentsov N.N., Böhrer J., Bimberg D., Ruvimov S.S., Werner P., Richter V., Gösele V., Heydenreich J., Ustinov V.M., Egorov A.Yu., Zhukov A.Z., Kop'ev P.S., Alferov Zh.I.* (1995). *Phys. Rev. Lett.*, 74, 4043.
- Mitin V.V., Kochelap V.A., Stroschio M.A.* (1999). *Quantum Heterostructures* (Cambridge University Press, Cambridge).
- Yariv A.* (1997). *Optical Electronics in Modern Communications* (Oxford University Press, Oxford).
- Zory P.S.Jr.* (1993). *Quantum Well Lasers* (Academic, Boston).

Задачи и упражнения

1. Лазеры на двойных гетероструктурах. Предположим, что мы имеем лазер с гетероструктурой на основе $\text{AlGaAs}-\text{GaAs}-\text{AlGaAs}$, излучающий на частоте 885 нм при длине оптического резонатора 150 мкм. Определите порядок n основной моды оптического излучения такого лазера (считая, что коэффициент преломления GaAs равен 3,8). Определите различие между модами излучения по длинам волн и по волновым векторам. Вычислите изменение длины волны излучения и пересчитайте порядок основной моды при увеличении температуры на 12°C (коэффициент преломления GaAs меняется примерно на $1,5 \times 10^{-4}$ /град).
2. Оптическое излучение в квантовых ямах. Для обычного лазера на квантовых ямах в системе $\text{AlGaAs}-\text{GaAs}-\text{AlGaAs}$ (с шириной ямы 8 нм) определите энергию и длины волны излучения. Сравните полученные величины со значениями для обычного полупроводникового лазера на GaAs . Рассмотрите поведение решений при бесконечной и конечной глубине ям.
3. Распределенные брэгговские рефлекторы (DBR) и одномодовые лазеры. Структура распределенного брэгговского рефлектора представляет гребенчатую диэлектрическую структуру,

образующую периодическую решетку и сконструированную подобноотражательным дифракционным решеткам. Такие структуры уже сейчас используются в лазерах с распределенной обратной связью, которые применяются в системах оптической связи. Их можно рассматривать как лазеры с шириной полосы порядка 0,1 нм. Определите для рефлектора DBR брэгговскую длину волны λ_B при периоде «гофрирования» 240 нм и коэффициенте преломления 3. (Подсказка: покажите, что при рассмотрении структуры DBR в качестве решетки в первом приближении можно получить $\lambda_B = 2l_c n$, где l_c — период гофрирования структуры; n — коэффициент преломления). Вычислите длину волны излучения λ лазера на основе структуры DBR для оптического резонатора длиной 300 мкм. (Подсказка: можно показать, что в лазерах на основе DBR частоты разрешенных мод излучения определяются выражениями $\lambda = \lambda_B + \lambda_B^2/2nL$ или $\lambda = \lambda_B - \lambda_B^2/2nL$, где L — длина оптической полости; λ_B — брэгговская длина волны, определенная полученным ранее решением.

4. Коэффициент усиления лазера на квантовых ямах. Покажите, что коэффициент усиления $\gamma(\hbar\omega_0)$ лазера на квантовых ямах пропорционален функции плотности состояний n_{2D} и коэффициенту инверсной заселенности $f_c(\hbar\omega_0) - f_v(\hbar\omega_0)$, где функции f_c и f_v соответствуют заселенностям распределения Ферми — Дирака в состояниях с $n = 1$ для зоны проводимости и валентной зоны квантовых ям соответственно. (Подсказка: предположите, что граница зон вблизи $k = 0$ (в k -пространстве) имеет параболическую форму и что существенны только оптические переходы, соответствующие $n = 1$. Используйте этот подход для объемного полупроводника, что должно привести к результату

$$\gamma(\hbar\omega_0) = \frac{m_r^* \lambda_0^2}{4\pi\hbar a n^2 \tau} [f_c(\hbar\omega_0) - f_v(\hbar\omega_0)],$$

где m_r^* — эффективная приведенная масса системы электрон—дырка; a — ширина ямы; n — коэффициент преломления; τ — время релаксации.

5. Модуляторы на квантовых ямах. Рассмотрим модулятор электропоглощения на множественных квантовых ямах в системе $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ с шириной $a = 8$ нм и общей толщиной активного слоя $L = 640$ нм. Полуширина экситонного пика со-

ставляет 25 мэВ. Рассчитайте коэффициент модуляции, т. е. отношение интенсивностей передаваемого света при наличии и отсутствии внешнего электрического поля $F = 10^5$ В/см. (Подсказка: для описания квантовой ямы с $x = 0,3$ и соответствующим содержанием Al в структуре можно использовать формулу для экситонного поглощения из работы [8]:

$$\alpha(\hbar\omega_0) \approx \frac{(2.9)(10^3)}{a\sigma} \exp\left(-\frac{E_{ex} - \hbar\omega}{(2)^{1/2} \sigma^2}\right),$$

где a должно быть выражено в ангстремах; σ — в эВ; а α — в см^{-1} . Отметим, что при отсутствии внешнего поля мы имеем $E_{ex} = \hbar\omega$, а в поле необходимо пользоваться формулой (8.12) для ΔE_{ex} .)

6. Модуляторы электропоглощения. Модуляторы электропоглощения (типа, показанного на рис. 10.15) представляют собой устройства поперечной передачи сигнала, в которых коэффициент поглощения является функцией приложенного обратного напряжения, т. е. $\alpha = \alpha(V)$, и определяется как усредненный коэффициент поглощения для области множественных квантовых ям. Такие модуляторы работают при энергии фотонов $\hbar\omega_0$, близкой к частоте экситонного пика, а изменения интенсивности самого пика (в зависимости от V) могут быть найдены из набора зависимостей типа рис. 8.12 для различных значений напряженности поля. Покажите, что соотношение контрастности R (определенное как отношение коэффициентов пропускания при $V = 0$ и $V = V$, может быть выражено соотношением

$$R(\text{db}) = 4.34[\alpha(V) - \alpha(0)]W,$$

где через W обозначена ширина структуры множественных ям (из формулы видно, что соотношение контрастности возрастает с ростом W). Поскольку понятно, что ширина W не может быть слишком большой (в противном случае, на выходе просто не будет никаких сигналов), попробуйте оценить оптимальное значение W , при котором отношение P_{out}/P_{in} еще не становится слишком малым.

ДОПОЛНЕНИЕ

МЕТРОЛОГИЧЕСКИЙ И СТАНДАРТИЗАЦИОННЫЙ БАЗИС НАНОТЕХНОЛОГИЙ

П. А. ТОДУА

*Научно-исследовательский центр по изучению свойств
поверхности и вакуума (НИЦПВ), Москва; Московский физико-
технический институт (МФТИ), Москва
e-mail: fgupnicpv@mail.ru*

История развития науки и техники неразрывно связана с развитием системы, методов и средств измерений. Нанотехнологии поставили ряд новых специфических задач, обусловленных малыми размерами элементов и структур, с которыми приходится иметь дело в данной области. Здесь, как нигде, актуален тезис: «Если нельзя правильно измерить, то невозможно создать». Все страны, вступившие в нанотехнологический прорыв, прекрасно представляют необходимость опережающего развития метрологии в этой бурно развивающейся области знания, поскольку именно уровень точности и достоверности измерений способен либо стимулировать развитие соответствующих отраслей экономики, либо служить сдерживающим фактором. Особо это подчеркивается тем обстоятельством, что в нанотехнологиях приборно-аналитическая и технологическая составляющие работают на пределе своих возможностей, что увеличивает вероятность ошибки, тем более связанной с человеческим фактором.

Одна из первоочередных задач стандартизации в нанотехнологиях — стандартизация параметров и свойств материалов, объектов, элементов и структур нанотехнологий, подлежащих измерениям. При межотраслевом и междисци-

плинарном характере нанотехнологий, различной терминологии и разных исследовательских и измерительных приемах и методах — это непростая, последовательно решаемая задача, несущая в себе объединяющее начало. К ней вплотную примыкает другая задача — необходимость стандартизации терминов и определений в нанотехнологиях, направленная на решение проблем общения и взаимопонимания различных групп исследователей не только внутри одной отдельно взятой страны, но и в рамках междисциплинарного обмена информацией между странами. Отсюда закономерное следствие — необходимость аттестованных и стандартизованных методик выполнения измерений, методик калибровки и проверки средств измерений, применяемых в нанотехнологиях, и многое другое, что определяется потребностями развития инфраструктуры nanoиндустрии.

Особый аспект стандартизации — решение задач обеспечения здоровья и безопасности операторов технологических процессов и лиц, взаимодействующих с продукцией нанотехнологий на всех этапах ее производства, испытаний, исследований и применений, а также экологической безопасности окружающей среды.

Логически следует, что «наибольший статистический вес» приходится на метрологию, поскольку именно она является количественным базисом стандартизации и сертификации. Специфика нанотехнологий привела к развитию нового направления в метрологии — нанометрологии, с которой связаны все теоретические и практические аспекты метрологического обеспечения единства измерений в нанотехнологиях. Из самого определения нанотехнологии, оперирующей с объектами нанометровой протяженности, естественным образом следует первоочередная задача измерений геометрических параметров объекта, что, в свою очередь, обуславливает необходимость обеспечения единства линейных измерений в нанометровом диапазоне. Но этим обстоятельством роль нанометрологии линейных измерений не исчерпывается. Она присутствует в неявном виде в подавляющем большинстве методов и средств обеспечения единства измерений физико-химических параметров и свойств объектов нанотехнологий, таких как механические, оптические, электрические, магнитные, акустические и т. д. Часто необходимо

осуществлять прецизионное пространственное позиционирование зонда измерительного устройства в место требуемого съема измерительной информации. При этом диапазон линейного сканирования по каждой координате может простирается от единиц нанометров до сотен и более микрометров, а требуемая точность выставления координаты составлять десятые доли нанометра.

Почему в нанометрологии столь большое внимание уделяют проблеме реализации линейной шкалы в нанометровом и прилегающем к нему диапазонах? Во-первых, потому что решение первоочередной задачи метрологии в нанотехнологиях — обеспечение единства измерений геометрических параметров нанообъекта — опирается на метрологию линейных измерений. Во-вторых, как указано выше, измерения механических, электрических, магнитных, оптических и многих других параметров и свойств объектов нанотехнологии связаны с необходимостью позиционирования зонда измерительного устройства в заданное место с наивысшей точностью [1, 2].

Обеспечение единства измерений физико-химических параметров и свойств объекта измерения требует привязки соответствующего средства измерений к эталону, воспроизводящему единицу данной физической величины (например, проводимости — к эталонному сопротивлению), а в нанотехнологиях в большинстве случаев — еще и обязательной привязки к базисному эталону единицы длины (рис. Д.1) для «точности попадания в цель». Этим дуализмом не ограничивается уникальность базисного эталона. Если обратить внимание на параметры, то видно, что диапазон измерений длины от единиц нанометров до сотен и более микрометров перекрывает более пяти порядков значений измеряемой величины при точности измерений в десятые-единицы нанометра во всем диапазоне.

Спектр объектов нанотехнологий и собственно наноиндустрии чрезвычайно широк, он простирается от ультрадисперсных сред до наноструктурированных многослойных материалов и кристаллов, включая квантоворазмерные структуры с размерностями локализации: один — так называемые квантовые ямы (сверхтонкие слои), два — квантовые проволоки или нити, три — квантовые точки. Проявляющиеся особенности физических эффектов и протекающих при этом процессов, в том



Рис. Д.1.

числе оптических, люминесцентных, электрических, магнитных, механических и многих других, определяются характерным размером, причем в одном и том же материале различные эффекты, связанные с размером, проявляются по-разному, например, особенность оптических свойств материала в ультрадисперсном виде может проявляться при одних размерах наночастиц, а теплофизических — при других.

Большинство методов исследований и измерений свойств нанообъектов — просвечивающая и растровая электронная микроскопия, сканирующая зондовая микроскопия, ионнополевая микроскопия, фотоэмиссионная и рентгеновская спектрометрия и рентгеновская дифрактометрия и другие — широко применяющиеся в технике измерений свойств материалов и объектов наноиндустрии, требуют калибровки средств измерений по стандартным образцам состава, структуры, свойств с известными размерными (то есть, геометрическими) характеристиками. Например, один из известных способов определения размеров ультрадисперсных частиц заключается в изучении рассеяния света на них. Рассеяние за-

висит от соотношения размеров частиц, длины волны падающего излучения и поляризации. При определении размеров частиц, как правило, используется лазерное излучение, но для калибровки такого средства измерений необходим набор ультрадисперсных частиц с дискретным рядом точно заданных размеров.

При доведении широкозонных полупроводниковых соединений группы A_2B_6 до ультрадисперсного состояния происходит «голубое смещение» полосы люминесценции, по которому можно судить о размерах ультрадисперсных частиц люминофора. Но в каждом конкретном случае используемого полупроводникового материала для калибровки необходим набор стандартных образцов из того же материала с целым рядом размеров.

При контроле технологических процессов создания многослойных тонкопленочных структур, в том числе и многослойных гетероструктур, необходимо привлечение рентгенодиагностических методов контроля скрытых слоев и, соответственно, наличие многослойных стандартных образцов состава и структуры для калибровки соответствующих средств измерений.

Фундаментальные исследования, связанные с прямыми измерениями физико-химических параметров веществ и материалов нанотехнологии, элементов и устройств нанотехники, требуют понимания закономерностей взаимодействия зонда измерительного средства с объектом измерения. Особую важность приобретают вопросы метрологии и стандартизации таких измерений, метрологического обеспечения, передачи размера единицы физической величины в нанометровый диапазон [3], характеризующийся специфическими особенностями.

Первостепенной задачей опережающего развития нанометрологии считают необходимость реализации наношкалы в нанометровом и прилегающем к нему диапазонах. Именно этой первостепеннейшей задаче нанометрологии посвящаются многие конференции и многочисленные публикации. Здесь нельзя не отметить существенный вклад России в решение этой фундаментальной измерительной проблемы. Достижение предельных возможностей при измерениях длины в нанометровом диапазоне связано с использованием высокоразрешающих методов

растровой электронной и сканирующей зондовой микроскопии в сочетании с лазерной интерферометрией и рентгеновской дифрактометрией при сохранении абсолютной привязки к первичному эталону метра.

В результате длительных исследований в России концептуально решена задача создания основ метрологического обеспечения измерений длины в диапазоне 1—1000 нм. При этом разработаны: методология обеспечения единства измерений в диапазоне длин 1—1000 нм, основанная на принципах зондовой микроскопии и лазерной интерферометрии и рентгеновской дифрактометрии; эталонный комплекс средств измерений, обеспечивающий воспроизведение и передачу размера единицы длины в диапазоне 1—1000 нм вещественным мерам длины с погрешностью 0,5 нм; поколение мер малой длины для калибровки средств измерений в диапазоне 1—1000 нм, в том числе меры нанорельефа поверхности; методология и алгоритмы измерения параметров профиля элементов микро- и наноструктур и пакет компьютерных программ для автоматизации таких измерений.

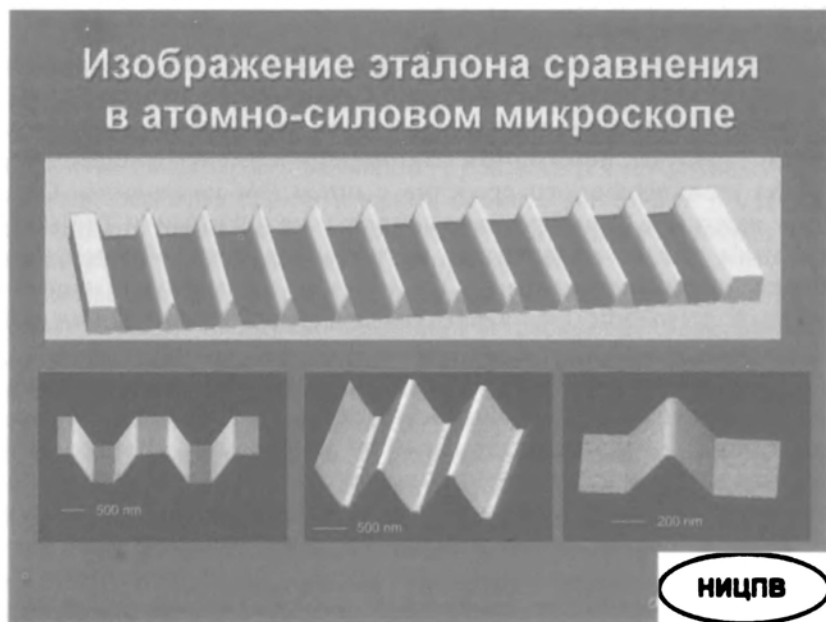


Рис. Д.2.



Рис. Д.3.

Важнейшим этапом в решении задач метрологического обеспечения линейных измерений в нанометровом диапазоне явилось создание вещественных носителей размера — мер с программируемым нанорельефом поверхности, обеспечивающих калибровку средств измерений с наивысшей точностью (рис. Д.2, Д.3, Д.4).

Именно такие трехмерные меры малой длины, или эталоны сравнения, — материальные носители размера, позволяющие осуществлять комплексную калибровку и контроль основных параметров растровых электронных и сканирующих зондовых микроскопов, предназначены для перевода этих сложных устройств из разряда устройств для визуализации исследуемого объекта в разряд средств измерений, т. е. в разряд приборов для измерений линейных размеров объектов исследования, обеспечивающих привязку измеряемых величин в нанометровой области к первичному эталону единицы длины — метру [4—9].

Аттестация эталонов сравнения осуществляется с использованием эталонной трехмерной интерферометрической системы измерений наноперемещений.

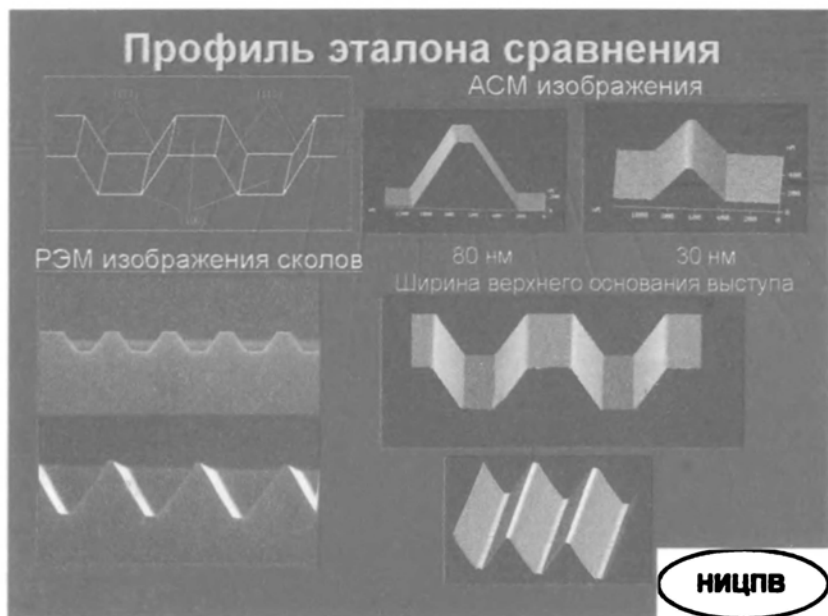


Рис. Д.4.

Аттестуются шаг меры и размеры верхних и нижних оснований выступов и канавок (ширина линии), а также высота (глубина) рельефа. При одном и том же шаге структуры возможно изготовление эталонов сравнения с шириной линий в диапазоне 10—1500 нм и высотой рельефа 100—1500 нм.

Мера позволяет по одному ее изображению в растровом электронном микроскопе (даже по одному сигналу), что очень важно для контроля технологических процессов, выполнить калибровку микроскопа (рис. Д.5), определить увеличение микроскопа, линейность его шкал и диаметр электронного зонда [10—15]. Слева вверху на рис. Д.5 приведены характерные параметры эталонной меры, в центральной части — параметры видеосигнала, в правой части — реальное РЭМ-изображение.

Кроме того, при необходимости подтверждения правильности измерений можно контролировать параметры растрового электронного микроскопа непосредственно в процессе проведения измерений размеров исследуемого объекта, что является дополнительной гарантией высокого качества измерений. Мера позволяет легко автоматизировать линейные измерения

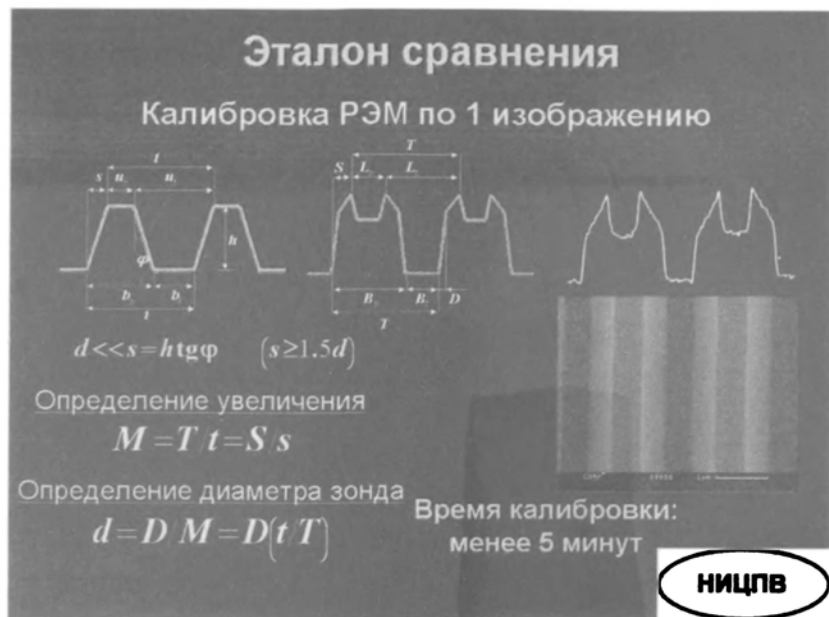


Рис. Д.5.

и создавать на основе растровых электронных микроскопов автоматизированные измерительные комплексы. Ряд подобных комплексов уже существует. В частности, в НИЦПВ создан автоматизированный комплекс для линейных измерений в области размеров от 1 нм до 100 мкм на основе растрового электронного микроскопа JSM-6460LV.

Аналогичным образом по заданным параметрам меры проводятся калибровка и контроль [16—21] таких характеристик атомно-силовых микроскопов, как цена деления и линейность шкал по всем трем координатам, ортогональность систем сканирования, радиус острия зонда (кантилевера), настройка параметров и выход микроскопа в рабочий режим (рис. Д.6). На рис. Д.6 слева сверху характерные параметры эталонной меры и ее АСМ-изображение (справа). Системы калибровки и аттестации атомно-силовых микроскопов успешно внедряются на предприятиях, специализирующихся на создании оборудования для нанотехнологий.

Развитие нанотехнологий ужесточает требования к измерительным системам, погрешности измерений которых должны



Рис. Д.6.

быть сравнимы с межатомными расстояниями. Все это требует серьезного отношения к вопросу обеспечения единства линейных измерений в нанометровом диапазоне.

Растровый электронный и сканирующий зондовый микроскопы только тогда могут считаться средствами измерений, когда их параметры будут соответствующим образом аттестовываться, калиброваться и контролироваться, причем последнее непосредственно в процессе измерений. Трехмерные меры или эталоны сравнения (материальные носители размера) — своеобразный мост между объектом измерений и эталоном метра — являются идеальным средством для осуществления таких операций. Непреложно одно: культура измерений требует, чтобы любой растровый электронный или сканирующий зондовый микроскоп, независимо от того, где он находится — в научной или промышленной лаборатории, учебном заведении или участвует в технологическом процессе, — должен быть укомплектован мерами, обеспечивающими калибровку и контроль параметров этого сложного устройства. Только тогда измерения, производимые на нем, могут претендовать на достоверность.

Кроме того, использование методов и средств калибровки и аттестации растровых электронных и сканирующих зондовых микроскопов производителями соответствующих приборов позволит им создавать новые приборы с лучшими характеристиками, которые, в свою очередь, позволят продвинуться дальше на пути развития нанотехнологий.

Для обеспечения нормативной базы нанометрологии, разработаны и введены в действие семь российских стандартов [22]:

- ГОСТ Р 8.628—2007 «Государственная система обеспечения единства измерений. Меры рельефные нанометрового диапазона из монокристаллического кремния. Требования к геометрическим формам, линейным размерам и выбору материала для изготовления»;
- ГОСТ Р 8.629—2007 «Государственная система обеспечения единства измерений. Меры рельефные нанометрового диапазона с трапецеидальным профилем элементов. Методика поверки»;
- ГОСТ Р 8.630—2007 «Государственная система обеспечения единства измерений. Микроскопы сканирующие зондовые атомно-силовые измерительные. Методика поверки»;
- ГОСТ Р 8.631—2007 «Государственная система обеспечения единства измерений. Микроскопы электронные растровые измерительные. Методика поверки»;
- ГОСТ Р 8.635—2007 «Государственная система обеспечения единства измерений. Микроскопы сканирующие зондовые атомно-силовые. Методика калибровки»;
- ГОСТ Р 8.636—2007 «Государственная система обеспечения единства измерений. Микроскопы электронные растровые. Методика калибровки»;
- ГОСТ Р 8.644—2008 «Государственная система обеспечения единства измерений. Меры рельефные нанометрового диапазона с трапецеидальным профилем элементов. Методика калибровки».

Для решения научно-технической проблемы обеспечения единства измерений в нанотехнологиях необходимо осуществить ряд научно-методических, технических и организационных мероприятий. В первую очередь, это создание новой структурной схемы передачи размера единиц физических величин от первичных эталонов рабочим средствам измерений, исключающей многоступенчатость передачи (см. рис. Д.1).

В этот комплекс мероприятий входят также: фундаментальные исследования механизмов взаимодействия зондов измерительных систем с объектом измерения; разработка новых алгоритмов измерений и соответствующего им математического обеспечения, учитывающего влияние взаимодействия рабочего средства измерений с измеряемым объектом; создание новых мер — материальных носителей размера, обладающих свойствами, аналогичными свойствам вторичного эталона и измеряемого объекта; разработка и создание стандартных образцов состава, структуры и рельефа поверхности и стандартизованных методик измерений в нанометрии, обеспечивающих прослеживаемость передачи размера единицы физической величины от эталона рабочим средствам измерений в нанометровый диапазон без существенной потери точности для аттестации, калибровки и поверки средств измерений.

Достижение данной цели вполне реально, поскольку фундамент решения указанной проблемы основан на концепции базисного эталона (см. рис. Д.1), в котором реализована нанoshкала. Этот эталон — основа для передачи единиц физических величин в нанометровый диапазон. Дело за немногим — необходима гармонизированная система стандартных образцов состава, структуры и свойств, служащих потребностям нанотехнологий. Все это создает предпосылки и закладывает основы ускоренного развития высоких технологий в России, и особенно главной из них — нанотехнологии.

ПО ВОПРОСАМ ПРИОБРЕТЕНИЯ ЭТАЛОНОВ СРАВНЕНИЯ — МЕТРОЛОГИЧЕСКИ АТТЕСТОВАННЫХ МЕР ДЛЯ КАЛИБРОВКИ РАСТРОВЫХ ЭЛЕКТРОННЫХ И АТОМНОСИЛОВЫХ МИКРОСКОПОВ В НАНОМЕТРОВОМ ДИАПАЗОНЕ ОБРАЩАТЬСЯ:

119421, РОССИЯ, МОСКВА, УЛ. НОВАТОРОВ,
ДОМ 40, КОРП. 1, НИЦПВ
ТЕЛ.: (+7) 495-935-9777
ТЕЛ/ФАКС: (+7) 495-935-5911
E-mail: fgupnicpv@mail.ru
www.nicpv.ru

Литература

1. M. T. Postek Nanometer — Scale Metrology//Proceedings of SPIE.2002, vol.4608, p.84—96.
2. П. А. Тодуа. Метрология в нанотехнологии//Российские нанотехнологии. 2007, том 2, № 1—2, с. 61—69.
3. П. А. Тодуа, В. А. Быков, Ч. П. Волк, Е. С. Горнев, Ж. Желкобаев, Л. М. Зыкин, А. Б. Ишанов, В. В. Календин, Ю. А. Новиков, Ю. В. Озерин, Ю. И. Плотников, А. М. Прохоров, А. В. Раков, С. А. Саунин, В. Н. Черняков. Метрологическое обеспечение измерений длины в микрометровом и нанометровом диапазоне и их внедрение в микроэлектронику и нанотехнологию// Микросистемная техника, 2004, № 1, с. 38—44; № 2, с. 24—39; № 3, с. 25—32.
4. Yu. A. Novikov, A. V. Rakov, P. A. Todua. Metrology in linear measurements of nanoobject elements//Proceedings of SPIE, 2006, vol. 6260, p. 626013-1—626013-8.
5. Yu. A. Novikov, V. P. Gavrilenko, Yu. V. Ozerin, A. V. Rakov, P. A. Todua. Silicon test object of linewidth of nanometer range for SEM and AFM//Proceedings of SPIE, 2007, vol. 6648, p. 66480R-1—66480R-11.
6. Yu. A. Novikov, V. P. Gavrilenko, A. V. Rakov, P. A. Todua. Test objects with right-angled and trapezoidal profiles of the relief elements//Proceedings of SPIE, 2008, vol. 7042, p. 704208-1—704208-12.
7. P. A. Todua, V. P. Gavrilenko, Yu. A. Novikov, A. V. Rakov. Check of the quality of fabrication of test objects with a trapezoidal profile// Proceedings of SPIE, 2008, vol. 7042, p. 704209-1—704209-8.
8. М. А. Данилова, В. Б. Митюхляев, Ю. А. Новиков, Ю. В. Озерин, А. В. Раков, П. А. Тодуа. Тест-объект с шириной линии менее 10 нм для растровой электронной микроскопии//Измерительная техника, 2008, № 8, с. 20—23.
9. М. А. Данилова, В. Б. Митюхляев, Ю. А. Новиков, Ю. В. Озерин, А. В. Раков, П. А. Тодуа. Тест-объект с тремя аттестованными размерами ширины линии для растровой электронной микроскопии//Измерительная техника, 2008, № 9, с. 49—51.
10. Yu. A. Novikov, A. V. Rakov, P. A. Todua. Linear sizes measurements of relief elements with the width less 100 nm on

- a SEM// Proceedings of SPIE, 2006, vol. 6260, p. 626015-1—626015-6.
11. V. P. Gavrilenko, M. N. Filippov, Yu. A. Novikov, A. V. Rakov, P. A. Todua. Measurements of linear sizes of relief elements in the nanometer range using a scanning electron microscopy// Proceedings of SPIE, 2007, vol. 6648, p. 66480T-1—66480T-12
 12. Yu. A. Novikov, S. A. Darznek, M. N. Filippov, V. B. Mityukhlyayev, A. V. Rakov, P. A. Todua. Nanorelief elements in reference measures for scanning electron microscopy// Proceedings of SPIE, 2008, vol. 7025, p. 702511-1—702511-10.
 13. V. P. Gavrilenko, Yu. A. Novikov, A. V. Rakov, P. A. Todua. Measurements of the parameters of the electron beam of a scanning electron microscopy// Proceedings of SPIE, 2008, vol. 7042, p. 70420C-1—70420C-12.
 14. Ю. А. Новиков, А. В. Раков, П. А. Тодуа. Точность измерения линейных размеров на растровых электронных микроскопах в микро- и нанотехнологиях//Измерительная техника, 2008, № 6, с. 15—18.
 15. Ч. П. Волк, Ю. А. Новиков, А. В. Раков, П. А. Тодуа. Калибровка растрового электронного микроскопа по двум координатам с использованием одного аттестованного размера//Измерительная техника, 2008, № 6, с. 18—20.
 16. P. A. Todua, M. N. Filippov, V. P. Gavrilenko, Yu. A. Novikov, A. V. Rakov. Measurement of linear sizes of relief elements in the nanometer range using an atomic force microscopy// Proceedings of SPIE, 2007, vol. 6648, p. 66480S-1—66480S-12.
 17. Yu. A. Novikov, M. N. Filippov, I. D. Lysov, A. V. Rakov, V. A. Sharonov, P. A. Todua. Direct measurement of the linewidth of relief elements of AFM in nanometer range// Proceedings of SPIE, 2008, vol. 7025, p. 702510-1—702510-10.
 18. Ю. А. Новиков, А. В. Раков, П. А. Тодуа. Геометрия формирования изображения в сканирующей зондовой микроскопии// Микроэлектроника, 2008, т. 37, №6, с. 448—469.
 19. А. В. Раков, Ю. А. Новиков, П. А. Тодуа. Калибровка АСМ по трем координатам с использованием одного аттестованного размера// Измерительная техника, 2008, № 5, с. 13—15.
 20. А. В. Раков, П. А. Тодуа. Измерение линейности сканирования в атомно-силовом микроскопе// Измерительная техника, 2008, № 6, с. 12—14.



21. Ю. А. Новиков, А. В. Раков, П. А. Тодуа. Прямое измерение ширины линии на атомно-силовом микроскопе//Измерительная техника, 2008, № 5, с. 10—12.
22. В. П. Гавриленко, Е. Н. Лесновский, Ю. А. Новиков, А. В. Раков, П. А. Тодуа, М. Н. Филиппов. Первые российские стандарты в нанотехнологиях//Известия РАН, сер. физич., 2009, т. 73, № 4, с. 454—462.

Заявки на книги присылайте по адресу:
125319 Москва, а/я 91
Издательство «Техносфера»
e-mail: knigi@technosfera.ru
sales@technosfera.ru
факс: (495) 956 33 46

В заявке обязательно указывайте
свой почтовый адрес!

Подробная информация о книгах на сайте
<http://www.technosfera.ru>

Мартинес-Дуарт Дж.М., Мартин-Палма Р.Дж., Агулло-Руеда Ф.

**Нанотехнологии для микро- и оптоэлектроники
Издание 2-е, дополненное**

Компьютерная верстка – С.С. Бегунов
Корректор – М.Е. Козлова
Дизайн – И.А. Куколева
Дизайн книжных серий – С.Ю. Биричев
Выпускающий редактор – О.Н. Кулешова
Ответственный за выпуск – В.М. Макарец

Формат 60x90/16. Печать офсетная.
Гарнитура Ньютон
Печ. л. 23. Тираж 1500 экз. Зак. № 219
Бумага офсет №1, плотность 65г/м², 80г/м².

Издательство «Техносфера»
Москва, ул. Краснопролетарская, дом 16, стр.2

Диалозитивы изготовлены ООО «Европолиграфик»

Отпечатано в ООО «Чебоксарская типография №1»
г. Чебоксары, пр. И. Яковлева, 15