Государственное образовательное учреждение высшего
профессионального образования
Уральский государственный университет им. А. М. Горького

В.Ф. Раданцев

Электронные свойства полупроводниковых наноструктур

Учебное пособие

Екатеринбург
2008

Оглавление

1 Введение **7**

2 Статистика в d –мерном электронном газе **27**

2.1 Квантование спектра в пространстве импульсов 28

2.2 Плотность состояний 35

2.2.1 Обобщенное выражение для плотности состояний при степенном законе дисперсии 42

2.3 Эффективные массы зонных носителей 46

2.3.1 Эффективная масса плотности состояний 49

2.3.2 Анизотропный квадратичный закон дисперсии . . . 50

2.3.3 Изотропный закон дисперсии 51

2.3.4 Динамическая эффективная масса 53

2.4 Концентрация электронов в d – мерном электронном газе . 56

2.4.1 Вырождение электронного газа 57

2.4.2 Обобщенная формула для d -мерной концентрации электронов со степенным законом дисперсии 62

2.4.3 Аппроксимации для сильно вырожденного и классического электронного газа 66

2.4.4 Статистика зонных и примесных состояний в трехмерном электронном газе 74

2.4.5 Зонная теория и многоэлектронные эффекты 82

2.5 Задачи 91

2.6 Вопросы для самопроверки 94

3 Теплоемкость и экранирование в d –мерном электронном газе **100**

3.1 Теплоемкость d –мерного газа 100
3.1.1 Классический газ 101
3.1.2 Вырожденный газ 102
3.1.3 Электронная и решеточная теплоемкость. 108
3.2 Экранирование статических полей 119
3.2.1 Статистика электронов в присутствии электростатического потенциала 121
3.2.2 Уравнение Пуассона 122
3.2.3 Режим линейного экранирования 126
3.2.4 Экранирование в двумерном электронном газе . . . 135
3.2.5 Экранирование кулоновского потенциала 139
3.3 Обобщенное соотношение Эйнштейна 145
3.4 Задачи 147
3.5 Вопросы для самопроверки 149

4 Классическая теория полупроводниковых гетероструктур 155
4.1 Одиночный гетеропереход 155
4.1.1 Зонная диаграмма одиночного гетероперехода (модель Шокли–Андерсона) 160
4.1.2 Методы определения параметров гетероструктур . . 170
4.2 МДП-структуры 172
4.2.1 Пространственное распределение потенциала 173
4.2.2 Поверхностная плотность заряда и дифференциальная емкость ОПЗ 174
4.2.3 Режимы обогащения, обеднения и инверсии 176
4.2.4 Классический предел (статистика Больцмана) . . . 177
4.2.5 ОПЗ с вырожденным электронным газом. 182
4.2.6 Квантовый характер приповерхностных потенциальных ям в МДП – структурах 187
4.3 Задачи 189
4.4 Вопросы для самопроверки 195

5 Размерное квантование в структурах с пространственным ограничением 201

5.1 Общая формулировка проблемы. Основные уравнения . . 203
5.2 Прямоугольная потенциальная яма в композиционной гетероструктуре 208
5.2.1 Приближение бесконечно высоких барьеров 208
5.3 Потенциальная яма произвольной формы. 212
5.4 Одиночный гетеропереход 215
5.4.1 Модель треугольной потенциальной ямы 217
5.5 Структуры с δ-легированными слоями 220
5.6 Приповерхностная квантовая яма в МДП-структурах . . . 221
5.6.1 Модель треугольной потенциальной ямы 222
5.7 Задачи 224
5.8 Вопросы для самопроверки 226

6 Полуклассическая и квантовая механика электрона проводимости в электрическом и магнитном поле 230
6.1 Классическая динамика частицы с произвольным законом дисперсии 231
6.2 Движение в постоянном и однородном электрическом поле 232
6.3 Классическое движение электрона проводимости в постоянном однородном магнитном поле 236
6.3.1 Циклотронная частота и циклотронная масса 238
6.3.2 Связь двумерного движения в координатном и импульсном пространствах 240
6.4 Квантовая механика электрона в магнитном поле 243
6.4.1 Квантование в магнитном поле (квадратичный закон дисперсии 245
6.4.2 Квантование в магнитном поле (произвольный закон дисперсии) 250
6.5 Кратность вырождения уровней Ландау 253
6.6 Особенности циклотронного движения в 2D системах . . . 257
6.7 Движение в наклонном магнитном поле 262
6.8 Спиновое расщепление 265
6.9 Задачи 268
6.10 Вопросы для самопроверки 270

7 Двумерный электронный газ в скрещенных электрическом и магнитном полях 275

7.1 Граничные состояния 275

7.2 Динамика электрона проводимости в скрещенных полях . 283

7.2.1 Классическая динамика 285

7.2.2 Квантование спектра в скрещенных электрическом и магнитном полях 291

7.3 Калибровочная инвариантность и эффект Бома–Ааронова 295

7.4 Задачи 306

7.5 Вопросы для самопроверки 307

8 Явления переноса 310

8.1 Кондактанс и удельная проводимость в магнитном поле . . 311

8.2 Тензоры двумерной удельной проводимости и сопротивления в магнитном поле 317

8.3 Уравнение Больцмана 325

8.4 Электропроводность d – мерного электронного газа 331

8.5 Задачи 348

8.6 Вопросы для самопроверки 349

9 Квантовый перенос в магнитном поле 352

9.1 Квантовая диффузия в магнитном поле 353

9.2 Квантовый эффект Холла 359

9.2.1 Перколяционная модель 360

9.3 Кондактанс квантовой нити 372

9.3.1 Формула Ландауэра 372

9.3.2 Проводимость в конечной длины квантовой проволоке 375

9.4 Краевая модель квантового эффекта Холла 379

9.5 Задача 382

9.6 Вопросы для самопроверки 383

10 Эффекты релятивистского типа 386

10.1 Теоретическая модель 388

10.2 Квазиультрарелятивистское приближение 394

10.3 Параметры двумерных подзон в пренебрежении спиновыми эффектами (приближение Клейна - Гордона 395
10.4 Эффекты спинорного типа 403
10.4.1 Резонансный сдвиг 404
10.4.2 Снятие спинового вырождения поверхностным потенциалом 408
10.4.3 Заселенности подзон и циклотронные массы 409
10.4.4 Зависимость спинорных эффектов от величины кейновской щели и легирования 411

Глава 1

Введение

Еще из школьного курса физики известно, что состояние частицы в классической механике полностью определяется двумя векторами - координатой (радиус вектором $\mathbf{r}$) и скоростью $\mathbf{v}$, (импульсом $\mathbf{p}$). Если на частицу не действуют никакие силы, т.е. в отсутствии внешних полей (мы имеем дело исключительно с полями электромагнитной природы, поэтому под частицей в дальнейшем понимаем заряженную частицу), то скорость частицы и ее импульс, согласно законам Ньютона, сохраняются во времени (являются интегралами движения), причем как координата, так и импульс могут принимать любые, сколь угодно большие или малые значения (в релятивистской механике скорость ограничена скоростью света $v < c$, в отличие от импульса $\mathbf{p} = m\mathbf{v}/\sqrt{1 - v^2/c^2}$), разделенные сколь угодно малыми интервалами $\Delta\mathbf{r}$ и $\Delta\mathbf{p}$. Отсутствие полей означает эквивалентность всех точек пространства, т.е его однородность. Сказанное выше обычно формулируется более высоким научным стилем: закон сохранения импульса является следствием однородности пространства. Поскольку однородно и время, то имеется еще один интеграл свободного движения - кинетическая энергия, для которой также допустимы любые (положительные) значения [1]. Сказанное выше остается справедливым и в квантовой механике - импульс и энергия являются хорошими квантовыми числами для свободного движения.

Выше мы неявно предполагали бесконечность доступного для части-

[1] В действительности дело обстоит с точностью до наоборот: - из эмпирического факта - неукоснительного выполнения законов сохранения в всех известных явлениях природы мы делаем вывод об однородности пространства и времени.

цы пространства, поскольку присутствие границ несомненно нарушает его однородность. В ограниченном пространстве, но в отсутствии полей внутри его (на границе, конечно, существует поле, не дающее проникнуть частице за нее) импульс уже не является сохраняющейся величиной - при взаимодействии с границей он может изменить как направление, так и величину (в случае упругого столкновения с границей, сохраняется, однако, модуль импульса и энергия). Что касается допустимых значений $\mathbf{p}$, то при классическом рассмотрении ограничения, накладываемые на $\mathbf{r}$, не ограничивают допустимых значений импульса и энергии. Совершенно иная ситуация имеет место для ограниченного движения в квантовой механике. При учете волновой природы электрона, ему ставится в соответствие волна де Бройля с волновым вектором $\mathbf{k} = \mathbf{p}/\hbar$ (длиной волны $\lambda = 2\pi/k$). При движении в ограниченном пространстве, подобно тому как это имеет место для волн в резонаторах или волноводах, "выживают" только те состояния, для которых амплитуда волны (волновой функции) в результате интерференции волн при их последовательном отражении от границ не обращается в ноль (конструктивная интерференция). В результате, величина импульса оказывается ограниченной со стороны малых значений (при непроницаемой границе длина волны не может быть больше удвоенного расстояния между границами L - на длине L должно укладываться полуцелое число волн). Со стороны больших значений импульс по прежнему не ограничен, т.е. число состояний остается бесконечным, однако это бесконечность совершенно другого качества - счетная бесконечность. В отличие от движения в неограниченном пространстве (или в ограниченном при классическом рассмотрении) значения импульса, а, следовательно, и энергии, для соседних состояний разделены пусть и малыми, но конечными интервалами. Сказанное относится к каждой компоненте импульса, так что направления импульса также квантованы, тогда как в классической механике состояния с данной абсолютной величиной импульса в пространстве размерности $d > 1$ бесконечно вырождены по его направлению.

Таким образом, учет электронной природы электрона кардинально меняет его свойства при пространственном ограничении - энергетиче-

ский спектр становится дискретным. Однако для ограниченных областей пространства макроскопических размеров расстояние между соседними квантовыми уровнями энергии крайне мало (квазинепрерывный спектр) и не может проявиться в каких-либо физических явлениях. Для оценки размеров, начиная с которых квантовый характер спектра начнет играть роль, можно использовать тот факт, что квантовые явления начинают ярко проявляются на атомном и субатомном уровне. Удобно провести сопоставление с простейшим природным объектом такого рода - атомом водорода. Отметим, однако, что если в атоме водорода область локализации электрона обусловлена и однозначно определяется силой его электростатического притяжения к ядру (протону), действующей на любых расстояниях, то при пространственном ограничении силами отталкивания (также электростатическими), действующими только на границах, выбор расстояния между которыми, в принципе, полностью в наших руках. Характерный размер области локализации электрона в атоме водорода a_B может быть легко оценен, используя соотношение неопределенностей $\Delta p \Delta x \geq \hbar/2$ (это выражение полностью согласуется с указанным выше фактом, что длина волны электрона $2\pi\hbar/p$ не может быть больше размеров области, в которой он локализован). Согласно этому соотношению, величина импульса электрона в атоме есть величина порядка $\hbar/a_B$, тогда кинетическая энергия $p^2/2m_0$ (m_0 – масса свободного электрона) соответственно порядка $\hbar^2/m_0 a_B^2$. Поскольку мы имеем дело с кулоновским потенциалом, то согласно известной теореме механики (теорема вириала) кинетическая и потенциальная e^2/a_B энергии по абсолютной величине одного порядка ($U = -2E_k$)

$$\frac{\hbar^2}{2m_0 a_B^2} \simeq \frac{e^2}{a_B},$$

откуда для размера области локализации имеем

$$a_B = \hbar^2/m_0 e^2.$$

Мы получили формулу, в точности совпадающую с выражением для боровского радиуса, который, как известно, есть величина порядка 0.5 Å=0.05 нм. Совершенно очевидно, что практически невозможно искус-

ственно осуществить пространственное ограничение такого масштаба для свободного электрона (даже в одном направлении, например "поселив" его между отрицательно заряженными пластинами конденсатора) [2].

Меняется ли что-нибудь в отношении сказанного выше для электрона в кристалле (поскольку предметом нашего изучения являются свойства, связанные со свободными, коллективизированными носителями заряда, то имеются в виду полупроводники и металлы)? Периодический потенциал нарушает, конечно, однородность пространства. Однако, в определенном смысле однородность пространства имеет место и в присутствии периодического потенциала. А именно, на масштабах длин, значительно превышающих постоянную решетки, различные участки идеального кристалла не могут существенно отличаться друг от друга. На основании представленной выше качественной картины, можно ожидать, что и в кристалле должен существовать некоторый интеграл движения, родственный (и совпадающий по физической размерности) импульсу свободного электрона, описывающий свойства электрона проводимости (и других свободных в смысле свободы передвижения в кристалле квазичастицы) на достаточно больших масштабах длин. Как мы знаем, зонная теория полностью подтверждает этот качественный вывод - свободный электрон проводимости можно полностью охарактеризовать квазиимпульсом, если не рассматривать деталей его движения на масштабах длин, меньших или сравнимых с размерами элементарной ячейки. Именно на этом основана возможность описания движения (как классического так и квантового) носителя заряда, в том числе во внешнем поле, как свободного электрона в рамках метода эффективной массы. В наиболее общем виде этот метод сводится к утверждению, что не масштабах, больших по сравнению с постоянной решетки, движение может быть описано теми же уравнениями (уравнениями движения при классическом рас-

[2]Для наблюдения квантования в таком пространственном "ящике" , его границы должны быть достаточно жесткими и гладкими (их шероховатость должна быть много меньше приведенной величины 0.05 нм., т.е. по крайней мере не больше 0.005 нм. Первое из этих условий предполагает, что реализация такого одномерного искусственного "атома" возможна только использовании твердого тела, например кристалла. Однако, поскольку постоянная решетки в кристаллах не менее нескольких ангстрем, то второе условие заведомо не может быть реализовано (вообще-то это сразу очевидно, поскольку расстояние между атомами вещества в любом его агрегатном состоянии превышает размеры составляющих его атомов).

смотрении и уравнением Шредингера при квантовом), что и движение свободного электрона, с перенормированными, однако, параметрами (в общем случае и их числом) закона дисперсии $\varepsilon(p)$. Периодический потенциал при таком описании не входит в эти уравнения, но неявно учтен в зависимости $\varepsilon(p)$ (в простейшем случае зоны с параболическим спектром перенормируется масса электрона, откуда и проистекает название этого подхода).

Наиболее существенная особенность перенормировки закона дисперсии для квазичастицы (электрона зоны проводимости, дырки валентной зоны и др.) в кристалле - периодический характер $\varepsilon(p)$. В отличие от свободного электрона, энергия которого неограниченно растет с импульсом, значения энергии электрона проводимости при неограниченном возрастании квазиимпульса не могут выйти за пределы энергетической зоны [3]. В виду периодического характера зависимости $\varepsilon(p)$ состояния с различными значениями $\mathbf{p}$, соответствующие одной энергии (отличающиеся на вектор обратной решетки), физически эквивалентны и, в принципе, можно ограничиться рассмотрением интервала квазиимпульсов в пределах первой зоны Бриллюэна. В этом смысле можно сказать, что физически осмысленные значения квазиимпульса также ограничены со стороны высоких значений. Фактически это соответствует тому, что состояния, для которых длина волны меньше постоянной решетки, не могут характеризоваться квазиимпульсом, описывающим электрон на масштабах, превышающих ее величину.

Рассмотрим, к каким последствиям приводит уменьшение области, доступной для электрона в координатном пространстве в случае кристалла. Нетрудно видеть, что в рамках метода эффективной массы рассмотренные выше особенности движения для свободного электрона (в неограниченном и ограниченном пространстве) непосредственно переносятся на случай электронов в периодическом потенциале, за исключением того, что вместо квантования импульса квантуется квазиимпульс, а

[3]Электрон может, в принципе, перейти в более высокую по энергии зону, однако для такого перехода ему требуется сообщить энергию, превышающую энергетическое расстояние до этой зоны, что возможно, как правило, только при достаточно сильных внешних воздействиях, например при поглощении фотона или в электрическом поле, достаточном для межзонного зинеровского пробоя.

энергетическое положение соответствующих квантовых уровней определяется перенормированным законом дисперсии. В реальных кристаллах имеется, однако, ряд особенностей, отличающих их от идеальной периодической решетки.

Во первых, неизбежное присутствие дефектов решетки (в виде примеси посторонних атомов, вакансий, дислокаций и т.п.) и тепловое движение атомов решетки приводит к нарушению строгой периодичности потенциала, и как следствие, к изменению квазиимпульса при взаимодействии с потенциалом дефекта (время взаимодействия крайне мало по сравнению с временем движения электрона между дефектами и взаимодействие носит характер столкновения). Пространственное распределение дефектов и электронов, как и распределение электронов по скоростям, носит случайный характер, т.е. мы имеем дело с неупорядоченной, стохатизированной системой, описание которой не может быть основано только на динамических законах, а требует статистического подхода. Как следствие, в качестве характеристик при описании явлений, связанных с рассеянием на дефектах, выступают средние значения соответствующих физических параметров. С точки зрения рассматриваемых нами эффектов, связанных с размерами системы (образца), для нас существенно появление в неупорядоченной системе такого параметра, имеющего размерность длины, как длина свободного пробега - средней величины расстояния, проходимого электроном между столкновениями. В зависимости от того, сохраняется ли при столкновениях энергия или нет, следует отличать длину свободного пробега по отношению к упругим столкновениям ℓ_p (в дальнейшем просто длина свободного пробега), и к неупругим ℓ_φ, которую принято называть длиной когерентности. Введение последнего термина связано с тем, что в отличие от упругих столкновений, которые не изменяют длину волны электрона $2\pi\hbar/p$ и закон изменения фазы волновой функции со временем $\exp(i\varepsilon t/\hbar)$ (электрон "помнит" о своей фазе до столкновения), неупругие акты столкновений приводят к изменению длины волны де Бройля и нарушают когерентность, в результате чего интерференция электронных волн в состояниях до и после неупругого столкновения становится невозможной.

Во вторых, в кристалле мы имеем дело не одним электроном, а с их ансамблем достаточно высокой концентрации. Даже в пренебрежении непосредственным ("силовым") взаимодействием между отдельными электронами, присутствие других электронов существенно сказывается на свойствах электронного газа в кристаллах. В силу принципа Паули электроны в не могут находиться в одном энергетическом состоянии, а поскольку такие величины, как среднее время между столкновения, длина свободного пробега и др., как правило, зависят от энергии, то следует говорить о их средних значениях при какой-то характерной энергии. В случае классического распределения электронов по состояниям это средняя тепловая энергия k_BT, при сильном вырождении электронного газа (всегда имеет место в металлах, а при низких температурах и в сильнолегированных полупроводниках)– энергия Ферми.

Теперь вернемся к вопросу о зависимости свойств электронной подсистемы кристалла от его размеров. Длина свободного пробега и длина когерентности непосредственно связаны с важнейшей характеристикой полупроводников и металлов - их удельной проводимостью. Разумно полагать, что удельная проводимость кристалла не должна зависеть от его размеров, если последние намного превышают длины свободного пробега ℓ_p и ℓ_φ, как это и полагается в традиционной трехмерной теории явлений переноса, основанной на диффузионной теории. Однако, если размер кристалла становится меньше длины свободного пробега, диффузионная теория, основанная на предположении о многократных актах рассеяния, не может быть использована, поскольку столкновения при движении электрона от контакта на одной стороне кристалла до контакта на другой отсутствуют. Движение электрона меняет характер, оно становится баллистическим (от подобия с движением снаряда) и теория явлений переноса должна быть модифицирована (а возможно и полностью пересмотрена).

В зависимости от формы образца (со сторонами длиной L_x, L_y, L_z), условие $L_i >> \ell_p$ может оказаться обрушенным только для одного из трех направлений. Тогда перенос в в остальных двух направлениях (в плоскости образованного при таком ограничении слоя) будет по прежне-

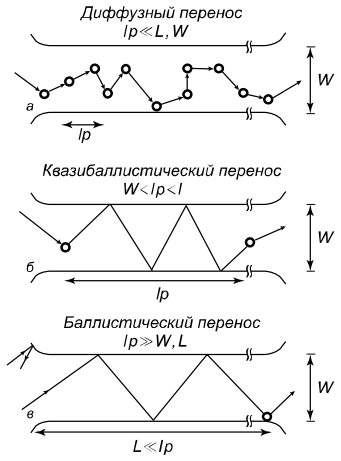

Рис. 1.1:

му подчиняться закономерностям, характерным для диффузного переноса [4]. Системы, для которых выполнено это условие, часто называют двумерными, однако как мы увидим ниже, этот термин более оправдано применять к системам, в которых разница в размерах системы по различным направлениям носит более фундаментальный характер. Чтобы не вносить неоднозначности в терминологию (в физике хватает и неоднозначности в обозначениях), лучше говорить о проводимости двумерной природы, или просто о двумерной проводимости и отличать ее от проводимости двумерной системы (определенной ниже). По-видимому не требует особого разъяснения представление об одномерной проводимости, когда условие $L_i >> \ell_p$ выполнено только по одной координатной оси.

Рассмотрим теперь вопрос о пространственном квантовании электронного спектра в кристаллах различного размера и формы. Очевидно, что необходимым условием образования стоячей волны (квантового состояния), скажем по оси x, является неравенство $L_x << \ell_p$, поскольку изменение направления импульса при столкновениях, а тем более энергии, а, как следствие, и длины волны де Бройля, нарушит конструктивную интерференцию. Однако это условие является необходимым, но недостаточным для пространственного квантования. Как мы видели выше, при больших размерах области локализации (даже если они мень-

[4]Необходимо, однако, учесть дополнительное рассеяние носителей на боковых границах образца, не дающее заметного вклада для больших (в указанном смысле) во всех трех направлениях образцов.

ше длины свободного пробега) спектр электрона квазинепрерывен, т.е близок к классическому. Рассмотренные выше условие для проявления квантовых эффектов фактически сводится к требованию, чтобы длина волны де Бройля была больше или порядка размеров образца (теорему вириала мы использовали только с целью сделать численные оценки критического для эффективного квантования размера, сведя ее к известной величине радиуса электронной орбиты в атоме водорода). В случае кристалла выражение для a_B нужно модифицировать с учетом перенормировки массы, ее отличия от массы свободного электрона m_0. Как правило, эффективная масса электрона в полупроводниках m значительно меньше m_0. Для большого класса важных полупроводниковых соединений отношение масс $m_0/m \sim 10$, а для узкощелевых полупроводников достигает нескольких тысяч. В кристалле из-за поляризации решетки перенормируется также и заряд электрона $e \to e/\sqrt{\epsilon}$. В большинстве полупроводников величина диэлектрической проницаемости $\epsilon \gtrsim 10$. Записывая выражение для модифицированного таким образом боровского радиуса в виде

$$a_B = \frac{\hbar^2}{m_0 e^2}\frac{m_0}{m}\epsilon,$$

видим, что его величина в полупроводниках при типичных значениях m и ϵ по крайней мере на два порядка выше, чем для свободного электрона, т.е. приходится на диапазон от нескольких до десятков нанометров. Такие значения критического с точки зрения проявления квантовых эффектов размера пространственного ограничения, намного превышают постоянную решетки (это обеспечивает справедливость наших оценок, основанных не методе эффективной массы) и при использовании современных нанотехнологий вполне могут быть реализованы в полупроводниковых гетероструктурах. Оценим масштаб энергий, соответствующий уровням размерного квантования. Модифицированное выражение для боровской энергии получается совершенно аналогично формуле для боровского радиуса

$$E_B \sim \frac{\hbar^2}{2ma_b^2} \sim \frac{m_0 e^4}{2\hbar^2}\frac{m}{m_0}\frac{1}{\epsilon^2} = E_{B0}\frac{m}{m_0}\frac{1}{\epsilon^2},$$

где E_{B0} - боровская энергия для свободного электрона. Благодаря пе-

ренормировке эффективной массы и поляризации решетки E_B в полупроводниках уменьшается почти на три порядка или даже больше. Поскольку $E_{B0} \sim 13$ эВ , то типичные наибольшие расстояния между квантовыми уровнями при пространственном ограничении в полупроводниках приходятся на интервал энергий милли-электронвольтового диапазона. Отсюда получаем оценку для интервала температур, при которых можно ожидать проявления квантовых эффектов. Отчетливое наблюдение последних (например в спектрах поглощения) возможно, если ширина фемиевской ступеньки $k_B T \sim 0.1T$ мЭв, определяющая заполнение энергетических уровней электронами, не превышала расстояния между ними. Таким образом, экспериментальное исследование низкоразмерных систем в полупроводниковых наноструктурах требует использования низких температур, зачастую гелиевого диапазона или ниже.

Если пространственное (размерное) квантование реализовано только в одном направлении и отсутствует в двух других, такая система называется двумерной (квазидвумерной), или квантовой плоскостью. Движение носителей в двумерной плоскости при этом остается свободным и может носить как диффузный, так и баллистический характер в зависимости от соотношения меду длиной свободного свободного пробега и размерами двумерного слоя. В отличие от случая упомянутых выше слоев с двумерной проводимостью, для которых все три компоненты квазиимпульса остаются хорошими квантовыми числами (в промежутках между столкновениями) и энергетический спектр в отсутствии внешних полей характеризуется трехмерным законом дисперсии $\varepsilon(p_x, p_y, p_z)$, свободное движение электрона в двумерной системе описывается энергетическим положением уровня размерного квантования и двумя компонентами квазиимпульса (в общем случае двумерное дисперсионное соотношение $\varepsilon(p_x, p_y)$ может отличаться от закона дисперсии в трехмерном кристалле). В результате размерного квантования может возникнуть несколько квантовых уровней E_i, для каждого из которых имеется непрерывный спектр энергетических состояний, отличающихся квантовыми числами p_x, p_y. Такую совокупность состояний для каждого значения квантового числа i, нумерующего уровни размерного квантования, называют подзоной

размерного квантования или двумерной подзоной. Таким, образом вместо квантовых чисел p_x, p_y, p_z, характеризующих орбитальное движение в трехмерной систем, в квантово-размерных структурах с двумерным газом каждое такое состояние описывается квантовыми числами i, p_x, p_y (при учете спиновой степени свободы также проекцией спина), т.е. спектр носит дискретно-непрерывный характер.

Если имеется только один уровень размерного квантования, или все заполненные электронами состояния в нижней подзоне имеют энергию меньшую уровня размерного квантования следующей по номеру подзоны, то никакие упругие процессы (рассеяние на примесях или акустических фононах) равно как и рассеяние электронов друг на друге не могут перевести электрон в состояния другой подзоны, поскольку такой процесс требует увеличения энергии. В этой ситуации электроны при рассеянии могут изменять свое состояние только в пределах одной основной подзоны. Часто понятие двумерного электронного газа относят только к таким структурам с одной заполненной подзоной. Следует однако заметить, что в силу закона сохранения импульса, вероятность межподзонных переходов в упругих процессах при достаточно большом энергетическом расстоянии между уровнями размерного квантования (небольшом числе уровней) мала, т.е время жизни носителя в подзоне τ_i велико, так что подзонные состояния почти всегда являются хорошо определенными (неопределенность в энергии $\sim \hbar/\tau_i$ много меньше расстояния между уровнями размерного квантования). По этой причине использование термина i–ой двумерной подзоны оправдано и при заполнении нескольких подзон размерного квантования.

Если система пространственно ограничена по двум степеням свободы, например при дополнительном (латеральном) ограничении в наноструктуре с двумерным электронным газом, то носители могут двигаться только вдоль одного направления, описываясь однокомпонентным волновым вектором. В поперечном направлении спектр размерно-квантован и полный спектр $\varepsilon_{i,j}(p)$ характеризуется теперь двумя, по числу степеней свободы, дискретными квантовыми числами и одномерным квазиимпульсом p. Структуры с таким спектром принято называть структу-

рами с одномерным электронным газом или квантовыми нитями (проволоками). И, наконец, если система пространственно ограничена по трем степеням свободы и условие $L_i \lesssim \lambda_d$ выполнено по всем пространственным направлениям, спектр будет полностью дискретным, характеризуясь тремя дискретными квантовыми числами $\varepsilon_{i,j,k}$, а соответствующий объект называют квантовой точкой. По существу это есть искусственный атом, в котором роль кулоновского притяжения к ядру играет отталкивание от потенциальных стенок, ограничивающих движение электрона в квантово-размерных наноструктурах.

Таким образом, если электронная система, характеризуемая при свободном движении тремя квазинепрерывными квантовыми числами - компонентами квазиимпульса p_x, p_y, p_z, заключена в области пространства, размеры которой по одному или нескольким направлениям порядка или меньше характерной длины волны де Бройля, энергетический спектр вырождается в двумерную $\varepsilon_i(p_x, p_y)$ или одномерную $\varepsilon_{i,j}(p_x)$ подзону размерного квантования (в общем случае в систему подзон), а при полном ограничении в систему дискретных уровней. Двумерный и одномерный газ носителей заряда обладают целым рядом уникальных и интересных (зачастую неожиданных) особенностей, кардинально отличающих их от трехмерного газа в неограниченном полупроводнике. Поскольку число, энергетическое положение и расстояние между квантовыми уровнями (подзонами) полностью определяется масштабом пространственного ограничения, которое может варьироваться в широких пределах, полупроводниковые наноструктуры являются мощным средством для создания материалов с желаемым энергетическим спектром - инженерии энергетического спектра.

Вообще говоря, история изучения электронного газа в размерно - квантованных структурах насчитывает без малого полвека (и более того в отношении теоретических исследований). Для современного этапа развития науки это весьма солидный срок и, на первый взгляд, можно было бы ожидать уменьшения интереса к исследованиям в этой области. Однако характерное для размерно-квантованных полупроводниковых структур богатство физических явлений, зачастую неожиданных и

весьма интригующих, связывающих их с другими, такими вроде бы далекими разделами теоретической физики, как квантовая электродинамика, физика плазмы и сверхтекучести, обеспечивает высокую актуальность исследований в этой области. Более того, физика низкоразмерных систем является в настоящее время несомненно магистральным путем в области полупроводниковой физики и электроники. Чтобы убедиться в этом, достаточно обратиться к оглавлениям ведущих физических журналов или программам наиболее представительных научных конференций. Тесная взаимосвязь физики, технологии и приборных применений, и ранее характерная для исследований в области полупроводников, особенно ярко проявляется на современном этапе. Успехи технологии изготовления полупроводниковых наноструктур позволяют создавать все новые объекты для физических исследований, результаты которых, в свою очередь, ставят, с одной стороны, новые задачи для технологов, а, с другой, служат основой для создания новых перспективных типов приборов электроники, формируя такие новые ее направления, как наноэлектроника, одноэлектроника, спинтроника и др.

Как это ни странно, но следует признать, что особенностью предмета "Электронные свойства полупроводниковых наноструктур" , несмотря на огромный, но разрозненный массив литературы в этой области, является отсутствие достаточно полноценного учебника на русском языке. Хотя и имеется несколько учебных пособий по физике низкоразмерных систем, однако они зачастую малодоступны для массового читателя, или рассматриваемый в них круг вопросов в значительной мере обусловлен научными интересами написавших их специалистов, с изложением деталей по частным проблемам, но при отсутствии достаточно общего взгляда на предмет. В настоящем пособии сделана попытка дать доступный для начинающего читателя общий взгляд на электронные свойства полупроводниковых структур с электронным газом различной размерности, оставляя в стороне многие из важных, но носящих (конечно на взгляд автора) частный характер, проблем. Мы не рассматриваем вопросов технологии изготовления полупроводниковых наноструктур (здесь имеется достаточно обширная и доступная литература) и, за редким исключени-

ем, не приводим конкретных экспериментальных результатов. Пособие ни в коей мере не претендует на роль полноценного учебника и может рассматриваться только как минимальный плацдарм для дальнейшего изучения предмета, обращаясь к обзорам и оригинальной литературе.

Из сделанного краткого предисловия по-видимому ясен тот минимальный круг вопросов, которые должны быть освещены при изучении этого раздела современной физики полупроводников и полупроводниковой электроники.

Во первых, необходимо рассмотреть типы полупроводниковых структур, используемых для создания электронного газа пониженной размерности, описать виды ограничивающего движение электронов (дырок) потенциальных ям и подходы к расчету пространственного распределения потенциала в них. Этому целиком посвящена глава 4 пособия.

Далее требуется рассмотреть общий подход к расчету энергетического спектра носителей в ограничивающем потенциале наноструктур, наиболее часто используемые при расчетах приближения и особенности моделей, используемых при описании различным видов структур с пространственным ограничением (глава 5).

Энергетический спектр, т.е. структура уровней размерного квантования и подзонные законы дисперсии, полностью определяют характеристики как свободного электронного газа пониженной размерности, так и его поведение во внешних полях. При этом требует обобщения на случай произвольной размерности такое фундаментальное понятие электронной системы, как плотность состояний, лежащее в основе расчета практически всех электронных свойств, в том, числе концентрации носителей и энергии Ферми. Логически этот раздел должен бы следовать после раздела, посвященного расчету спектра, как это и удобно сделать при чтении лекций студентам. Однако, чтобы избежать повторения материала, относящегося к трехмерным системам (используемого в главах 4 и 5), этот материал рассматривается в первой главе пособия.

Вторая глава посвящена сравнительному изучению в системах различной размерности простейших равновесных характеристик электронного газа - экранированию статических электрических полей и тепло-

емкости. Там же рассматривается теплоемкость фононной подсистемы. Хотя с практической точки зрения рассмотрение теплоемкости двумерной или одномерной решетки представляет скорее академический интерес, сопоставление свойств бозонного и фермионного газа различной размерности представляется в значительной мере полезным с методической точки зрения.

Информация о свойствах любой системы, лежащих в их основе явлений и механизмов, может быть получена только при изучении реакции системы на внешние воздействия. В главе 6 рассмотрены особенности поведения электронных систем различной размерности во внешних электрическом и магнитном полях. Предельно выраженная пространственная анизотропия двумерного и одномерного электронного газа приводит к ярко выраженной специфике их поведения во внешних полях различной ориентации относительно одномерной нити или двумерной плоскости по сравнению с трехмерной системой. Это имеет место как при классическом, так и при квантовом описании движения носителей заряда.

Глава 7 по-существу является продолжением предыдущей. Однако из-за ряда специфических особенностей в поведении электронов в системах пониженной размерности при одновременном воздействии магнитного и электрического полей (в том числе поля, обусловленного скачком потенциала на границах), автор счел полезным вынести рассмотрение этих вопросов в отдельный раздел.

В главе 8 на примере электропроводности и эффекта Холла рассмотрены особенности явлений переноса в системах различной размерности в рамках феноменологического подхода и квазиклассического описания, основанного на уравнении Больцмана.

В главе 9 рассмотрена электропроводность и эффект Холла в двумерной системе (ясно, что в идеальном одномерном газе магнитное поле воздействует только на спиновую степень свободы, не влияя на орбитальное движение носителей тока) при учете квантования энергетического спектра в магнитном поле. Автор не дает подробного изложения весьма обширного круга вопросов, относящихся к квантованию холловской проводимости, ограничиваясь рассмотрением основных физических

механизмов, приводящих к этому необычному макроскопическому квантовому эффекту, тем более что по этому вопросу имеется достаточно обширная и вполне доступная литература, включая обзоры и монографии.

Общей отличительной особенностью настоящего пособия является то, что на протяжении всего изложения проводится мысль об определяющей роли в формировании свойств электронной системы как размерности пространства, так и законов дисперсии. Дело в том, что стандартное изложение, принятое практически во всех пособиях и в большинстве монографий, основано, как правило, на предположении о квадратичности дисперсионного соотношения в полупроводниках, образующих квантоворазмерную структуру (и, как следствие, квадратичности спектра подзон размерного квантования). При этом может возникнуть впечатление об общности результатов, полученных в действительности в рамках простейшего приближения (зачастую предполагаемого по умолчанию). Однако для широкого класса полупроводников, используемых при формировании наноструктур, спектр отличается (в некоторых случаях весьма существенно) от квадратичного, тем более, что в случае квантовых ям характерные энергии отстоят значительно дальше от дна соответствующих зон, чем это имеет место при изучении большинства трехмерных свойств, определяемых состояниями вблизи дна разрешенных зон, где дисперсия действительно хорошо аппроксимируется квадратичным законом. Между тем, закон дисперсии является фундаментальной характеристикой, определяющей практически все основные свойства полупроводников и металлов или структур на их основе. На материале отдельных глав показывается, что вид дисперсионного соотношения играет определяющую роль в пространственном распределении ограничивающего потенциала в полупроводниковых наноструктурах, энергетическом положении уровней размерного квантования, в формировании фундаментальных физических характеристик трехмерного, двумерного и одномерного электронного газа, начиная с плотности состояний и концентрации и, как следствие, критически сказывается как на равновесных свойствах электронного газа (экранирование, теплоемкость), так и на явлениях переноса

(за важным исключением квантового переноса в квантующих магнитных полях).

Для иллюстрации этого на протяжении всего курса проводится сопоставление для каждой размерности системы результатов для двух простейших законов дисперсии - стандартного (квадратичного) и линейного. Последний (аналогичен, например, законам дисперсии нейтрино, фотонов, ультрарелятивистских электронов и т.п.) описывает спектры в кейновских полупроводниках при энергиях $\varepsilon >> \varepsilon_g$ (при всех энергиях в полупроводниках с нулевой кейновской щелью ε_g). Рассмотрение более общего дисперсионного соотношения кейновского типа (см. ниже) по большей части предлагается в виде задач. Там, где это возможно, выводятся обобщенные выражения для соответствующих характеристик электронного газа, в которых в качестве параметров фигурируют размерность пространства d и показатель степени в дисперсионном соотношении r (формально последнему можно приписать любое значение). При таком представлении отчетливо видна критическая роль соотношения между этими параметрами на ряде важнейших характеристик (например видом энергетической зависимости плотности состояний) и свойств электронной подсистемы.

Вторая особенность курса заключается в использовании единого подхода для определения собственных значений уравнения Шредингера при расчете энергетического спектра в полупроводниковых наноструктурах, в том числе во внешних полях, а именно квазиклассического приближения. Определение уровней размерного квантования, подзонных законов дисперсии и спектра в магнитном поле при таком подходе сводится к тривиальному вычислению определенных интегралов, т.е кардинально упрощает математику, делая не нужным обращение к учебникам по дифференциальным уравнениям и справочникам по специальным функциям. С методической точки зрения наиболее важно то, что решения могут быть зачастую представлены в аналитическом виде, обеспечивая ясное понимание качественных особенностей спектра и характера его зависимости от основных, определяющих параметров системы. С другой стороны, хотя формально условия квазиклассичности выполняются

только для больших квантовых чисел, квазиклассическое приближение в действительности дает хорошее описание спектра даже для основного состояния (параметром разложения является отношение $1/\pi^2 i^2$), а для ряда важных физических задач квазиклассический спектр совпадает с точным или весьма близок к нему.

Изложение в Гл. 1-9 основано фактически на однозонном приближении в предположении, что известен закон дисперсии $\varepsilon(p)$ (само дисперсионное соотношение, устанавливаемое экспериментально, учитывает, конечно, взаимодействие с другими энергетическими зонами). Однако в структурах на основе полупроводников с неквадратичным спектром возникает ряд новых эффектов, отсутствующих в системах со стандартными зонами, рассмотрение которых требует многозонного рассмотрения, т.е. основано на матричном уравнении Шредингера. Наиболее ярко эти явления проявляются в наноструктурах на основе кейновских полупроводников с узкой запрещенной зоной, дисперсионное соотношение для которых аналогично закону дисперсии релятивистского электрона $\varepsilon = \sqrt{p^2 s^2 + m^2 s^4} - ms^2$ с точностью до замены массы свободного электрона на эффективную кейновскую массу (в отличие от массы свободного электрона эта величина может меняться в широких пределах, включая $m = 0$) и скорости света c на кейновскую скорость $s \sim c/300$, практически одинаковую для всех полупроводников этого класса. В связи с указанной аналогией, специфические эффекты, обусловленные таким видом дисперсии, могут быть названы эффектами (квази)релятивистского типа (очевидно, что они имели бы место и при пространственном ограничении релятивистского электрона, если бы его удалось реализовать на достаточно малых масштабах). Некоторые из них рассмотрены в заключительной главе, материал которой, в значительной мере отражая научные интересы автора, потребует от читателя несколько большего внимания, нежели при чтении предыдущих глав. Поскольку часть квазирелятивистских эффектов, связанных со спином, имеет место только в наноструктурах с асимметричным ограничивающим потенциалом, рассматривается наиболее асимметричная система - двумерный электронный газ в приповерхностных квантовых ямах (в обогащенных или ин-

версионных слоях) МДП-структур.

Электрическое поле в приповерхностных квантовых ямах МДП структур на основе кейновского полупроводника играет двоякую роль. С одной стороны оно приводит к размерному квантованию энергетического спектра в направлении конфайнмента и, как следствие, к двумеризации электронного газа. С другой, будучи перпендикулярным скорости (двумерному квазиимпульсу), может приводить к спин-орбитальному взаимодействию ($U_{so} \propto \mathbf{E} \times \mathbf{v}$), подобно тому как это имеет место для электрона в атоме, "орбита "которого перпендикулярна кулоновскому полю ядра. Как следствие, это приводит к такому необычному явлению, как снятию спинового вырождения подзон в отсутствии магнитного поля! В определенной мере наблюдение этого эффекта, позволяющего управлять спиновой степенью свободы носителей тока с помощью электрического поля (подобно тому, как, прикладывая к затвору напряжение, управляют потоком заряда носителей в полевом транзисторе), внесло существенный вклад в формирование новому направлению в современной электронике - спинтронике. Это явление, наряду с некоторыми другими, связанными со спином эффектами, рассматриваются в главе 10. Для теоретического анализа используется имеющий ясную физическую интерпретацию подход (развитый ранее для описания "родственной" системы - вакуумного конденсата дираковских электронов вблизи сверхкритических ядер), в рамках которого многозонное матричное уравнение сводится к уравнению шредингеровского типа с некоторым эффективным потенциалом (подобно центробежному потенциалу в модифицированном уравнении Шредингера для атома водорода). В результирующем уравнении легко выделяются ответственный за эффекты непараболичности и не зависящий от спина потенциал типа Клейна-Гордона, и различные для случаев $E_g > 0$ и $E_g < 0$ потенциалы спинорного типа, ответственные за "резонансный сдвиг"и спин-орбитальное расщепление подзонных спектров. В рамках этого подхода легко видны качественное подобие и количественная разница в параметрах подзон для материалов с $E_g > 0$ и $E_g < 0$. Особое внимание уделено описанию двумерного спектра в квазиультрарелятивистской аппроксимации $E_g = 0$, в рамках которой, используя,

как обычно, квазиклассическое квантование, удается получить аналитические выражения для ряда важных в экспериментальном отношении параметров.

Глава 2

Статистика в d –мерном электронном газе

Концентрация частиц (атомов газа, фотонов, электронов, фононов и т.д.) т. е. число частиц на единицу объема определяется числом возможных состояний в единице объема, в которых может находиться частица и вероятностью того, что состояние ею занято. Для свободной частицы в пространстве размерности d ее состояние характеризуется d компонентами импульса (орбитальное состояние) и проекцией спина [1]. В рамках классического рассмотрения компоненты импульса могут принимать любые сколь-угодно мало отличающиеся друг от друга значения. Поэтому в сколь-угодно малом интервале импульсов (и соответственно энергии $\varepsilon = \varepsilon(\mathbf{p})$) имеется бесконечное число состояний, так что задача подсчета числа состояний в классической механике при любом объеме и размерности пространства фактически беспредметна. Конечно, понятие концентрации частиц используется практически во всех разделах классической физики, например концентрация носителей в теории кинетических явлений Друде–Лоренца, однако в плане теоретического описания оно по существу является феноменологическим параметром (который, конечно, может быть сопоставлен с результатами экспериментальных измере-

[1] С точки зрения рассматриваемого нами круга вопросов электронных явлений в кристаллах реальный физический интерес представляют, конечно, только системы с $d \leq 3$. Однако, имея целью описание электронных систем с $d = 1 \div 3$ в общем виде (не конкретизируя пространственную размерность системы), мы в большинстве случаев получаем соотношения, справедливые для произвольной размерности. И хотя системы с $d > 3$ не имеют реального воплощения, общий, пусть даже абстрактный взгляд на зависимость тех или иных физических свойств системы от ее размерности позволяет зачастую получить более полную физическую картину.

ний), не поддающимся теоретическому расчету "из первых принципов" в рамках классического рассмотрения. Квантовое происхождение таких на первый взгляд вполне классических понятий как плотность состояний и концентрация частиц очевидно уже из того факта, что выражения для этих величин, вычислением которых мы займемся ниже, всегда содержат постоянную Планка (или волновой вектор $k = p/\hbar$).

При квантово-механическом описании ситуация совершенно другая. В этом случае частицы описываются волнами де Бройля, волновые векторы которых сопоставляются с механическим импульсом частицы $k_{x_i} = p_{x_i}/\hbar$, и в любом конечном объеме, например в d – мерном кубе, могут существовать стационарные состояния (стоячие или бегущие волны) только с вполне определенными длинами волн де Бройля $\lambda\!\!\!^{-} = 2\pi/k_{x_i}$. Волны с другими квазиимпульсами при последовательных отражениях от границ при интерференции волн гасятся. При постулировании волновой природы частиц подсчет числа квантовых состояний в механической системе фактически полностью аналогичен вычислению числа собственных колебаний (мод) волнового поля.

2.1 Квантование спектра в пространстве импульсов

Для системы макроскопических размеров мы предполагаем ящик, в котором заключена система, конечным, но произвольно большого размера, по крайне мере значительно превышающим характерные физические длины в системе - постоянную решетки, длину волны электрона и т. п. Очевидно, что в этих условиях всегда можно также полагать выполненными условия применимости метода эффективной массы, так что квантование в ящике для свободного электрона и электронов проводимости в периодическом потенциале решетки эквивалентны (с точностью до замены импульса на квазиимпульс! - см. ниже). Более того, движение может рассматриваться квазиклассичным. Разрешенные значения волновых векторов определяются, помимо размеров ящика, граничными условиями, накладываемыми на волновую функцию. Здесь возможны

два наиболее простых варианта (для простоты выкладок мы начнем со случая одного измерения, который затем тривиальным путем расширяется на случай произвольной размерности пространства).

1. Фиксированные граничные условия в ящике, когда на волновую функцию накладываются нулевые граничные условия на стенках ящика
$$\Psi(0) = \Psi(L) = 0. \tag{2.1}$$

2. Периодические граничные условия (условия Борна–Кармана), которые в случае одного измерения могут быть интерпретированы как соединение двух концов одномерной нити длиной L. Преимущество такого подхода в том что устраняется влияние границ системы. В этом случае необходимо удовлетворить условию непрерывности волновой функции и ее производной на "границе".

$$\Psi(x) = \Psi(x+L), \tag{2.2a}$$
$$\left.\frac{\partial\Psi}{\partial x}\right|_x = \left.\frac{\partial\Psi}{\partial x}\right|_{x+L}. \tag{2.2b}$$

Условия (2.1) предполагают существование на границах бесконечно высокой потенциальной стенки $U(0) = \infty, U(L) = \infty$ – только в этом случае квантовая частица не может проникнуть за стенку, и, следовательно, волновая функция обращается в ноль [2]. Хотя уравнение Шредингера
$$\left[\frac{\hat{p}^2}{2m} + U(x)\right]\Psi(x) = \varepsilon\Psi(x) \tag{2.3}$$
для области внутри ящика в этом случае не содержит координаты [3] ($U(x) = 0$) и аналогично уравнению для свободной частицы, волновая функция не может быть плоской волной свободного электрона $\Psi_p(x) =$

[2] С формальной точки зрения, если $\Psi(0) \neq 0,\ \Psi(L) \neq 0$, то член $U(x)\Psi(x)$ в уравнении Шредингера в точках $x = 0$ и $x = L$ перестает быть конечным, а уравнение Шредингера хорошо определенным.

[3] При учете границ потенциальная энергия является, конечно, функцией координат ($U(0 < x < L) = 0$, $U(x \leq 0) = U(x \geq L) = \infty$)!

$const \cdot \exp\left(\frac{i}{\hbar}px\right)$, поскольку из-за наличия стенок движение в ящике финитно. Удовлетворяющим граничному условию (2.1) при $x = 0$ нормированным решением уравнения (2.3) при $U(x) = 0$ является

$$\Psi_p(x) = \frac{\sqrt{2}}{L} \cdot \sin(\frac{px}{\hbar}). \tag{2.4}$$

Граничное условие $\Psi_p(L) = 0$ приводит к уравнению $\sin(\frac{pL}{\hbar}) = 0$, которое и определяет разрешенные значения импульса [4]

$$p_m = \frac{\pi\hbar}{L}m, \quad m = 1\ ,2\ ,3\ ,\ \ldots \tag{2.5}$$

Квантовые числа m и разрешенные значения импульса здесь положительные величины. Волновые функции соответствуют стоячим волнам, не переносящим тока.

В случае циклических условий частица не ограничена в пространстве какими либо барьерами, потенциальная энергия всюду $U(x) \equiv const$ и система в этом случае действительно описывается уравнением Шредингера для свободной частицы, которое, как легко видеть, удовлетворяется волновыми функциями типа бегущих волн

$$\Psi_p(x) = const \cdot \exp\left(\frac{i}{\hbar}px\right). \tag{2.6}$$

Здесь однако мы имеем дело с особым видом свободного движения, своеобразие которого заключается в том, что, хотя движение не ограничено, частица все же локализована в конечной области пространства протяженностью L. Граничные условия (2.2) для такой неограниченной, но замкнутой, циклически повторяющейся системы, по существу, вытекают из требования однозначности волновой функции. Разрешенные значения импульса даются условием цикличности (2.2а) для волновой функции (2.6), т. е. уравнением

$$\exp\frac{ipx}{\hbar} = \exp\frac{ip(x+L)}{\hbar}, \tag{2.7}$$

[4]Тот же результат может быть получен из того соображения, что, поскольку на границах волновая функция равна нулю, на длине L должно укладывается полуцелое число длин волн де Бройля $\lambda\!\!\!^{-} = 2\pi\hbar/p$.

из которого следует $\exp\left(\frac{i}{\hbar}pL\right) = 1$, так что $pL/\hbar = 2\pi n$, откуда

$$p_n = \frac{2\pi\hbar}{L}n, \quad n = 0,\ \pm1,\ \pm2,\ \pm3,\ \ldots \tag{2.8}$$

Нетрудно видеть, что при этих значениях p_n условие (2.2b) также удовлетворено. В отличие от случая (2.5) энергетические уровни $\varepsilon_n(|p_n|)$, соответствующие разрешенным значениям (2.8), двукратно вырождены (за исключением уровня $n = 0$), что соответствует двум противоположным направлениям распространения бегущей волны (двум знакам p)

$$\Psi_n = \sqrt{1/L}\exp\left(\frac{i}{\hbar}p_n x\right). \tag{2.9}$$

Для свободного электрона число разрешенных квантовых состояний (2.5), (2.8) остается, как это имеет место и при классическом рассмотрении, бесконечным при любом размере системы L. Это является следствием неограниченности спектра свободного электрона $\varepsilon = p^2/2m_0$, допускающего произвольно большие значения энергии, а, следовательно, и импульса (это остается справедливым и в релятивистской области $\varepsilon = \sqrt{p^2c^2 + (m_0c^2)^2} - m_0c^2$). Поскольку значения p не ограничены сверху, то допустимы состояния с произвольно большими значениями m или n (волны де Бройля со сколь угодно малой длиной волны). Однако в отличие от случая классической частицы, эта бесконечность меньшей мощности – дискретная (счетная) бесконечность. Квантование (дискретизация) спектра в ящике приводит к конечному числу состояний в любом конечном интервале пространства импульсов (а, следовательно, и энергии). Действительно, согласно (2.8), разрешенные значения p_n разделены интервалами

$$\Delta p = p_n - p_{n-1} = \frac{2\pi\hbar}{L},$$

одинаковыми для всех квантовых чисел n (спектр в пространстве импульсов эквидистантен) и не зависящими от параметров закона дисперсии. Таким образом, на одно квантовое состояние приходится объем $2\pi\hbar/L$ пространства импульсов, или объем

$$L\Delta p = 2\pi\hbar \tag{2.10}$$

фазового пространства. Формула (2.10) есть не что иное, как соотношение неопределенностей для локализованной в ящике L частицы, которое можно интерпретировать как несовместимость с квантовой механикой деления фазового пространства на клетки, меньшие $2\pi\hbar$ (дискретность фазового пространства). Хотя величина Δp уменьшатся с ростом размеров системы (в пределе больших L спектр становится квазинепрерывным), число состояний в любом интервале δp импульсного пространства

$$N(\delta p) = \frac{\delta p}{\Delta p} = \frac{L\delta p}{2\pi\hbar}$$

в расчете на единицу объема координатного пространства L

$$n(\delta p) = \frac{\delta p}{2\pi\hbar}$$

от размеров ящика не зависит. Число же состояний в единице объема фазового пространства

$$\frac{N(\delta p)}{L\delta p} = \frac{1}{2\pi\hbar}$$

есть универсальная величина, вообще не зависящая от каких либо параметров системы.

Следует обратить внимание, что при использовании нулевых граничных условий (2.1) величина интервала Δp, согласно (2.5), вдвое меньше. Но при этом учитываются только положительные значения квантовых чисел m. С другой стороны, в случае циклических условий, для заданного квантового числа $|n|$ имеются две волны (2.9), отличающиеся знаком n (знаком p), которые однако эквивалентны одной стоячей волне, поскольку

$$\Psi_n + \Psi_{-n} = \sqrt{1/L}\exp\frac{ip_n x}{\hbar} + \sqrt{1/L}\exp\frac{-ip_n x}{\hbar} = \sqrt{\frac{2}{L}}\cos\frac{p_n x}{\hbar}$$

В результате, число состояний на один и тот же интервал импульсов $\delta|p|$ в обоих случаях одинаков. Хотя оба подхода к подсчету числа квантовых состояний приводят к одинаковым результатам, использование циклических условий представляется более последовательным, не только с точки зрения устранения влияния границ, но и по той причине, что волновые

функции, описывающие двигающуюся в пространстве (и во времени – с учетом временного множителя, соответствующего стационарной энергии $\varepsilon_n(p_n)$ волновая функция $\Psi_n(x,t) = \sqrt{1/L}\exp\left[\frac{i}{\hbar}(p_n x - \varepsilon_n t)\right])$ частицу, наиболее естественны для свободного электрона, в том числе для электрона в кристалле.

Для электронов в периодическом потенциале решетки [5] верхние значения квазиимпульса ограничены первой зоной Бриллюэна (в одномерном случае $|p| \leq \pi\hbar/a$, где a – период потенциала), поскольку состояния $\Psi_n = \sqrt{1/L}\exp\left(\frac{i}{\hbar}p_n x\right)$ и $\Psi_n = \sqrt{1/L}\exp\left[\frac{i}{\hbar}(p_n \pm 2\pi\hbar/a)x\right]$ физически эквивалентны. В результате, число квантовых состояний в присутствии периодического потенциала в каждой энергетической зоне в точности равно числу периодов решетки в блоке L. Ограниченность спектра импульсов со стороны больших значений p не накладывает каких-либо ограничений на применимость полученных результатов, поскольку в полупроводниках актуальный в физическом отношении диапазон импульсов находится далеко от границ зоны Бриллюэна.

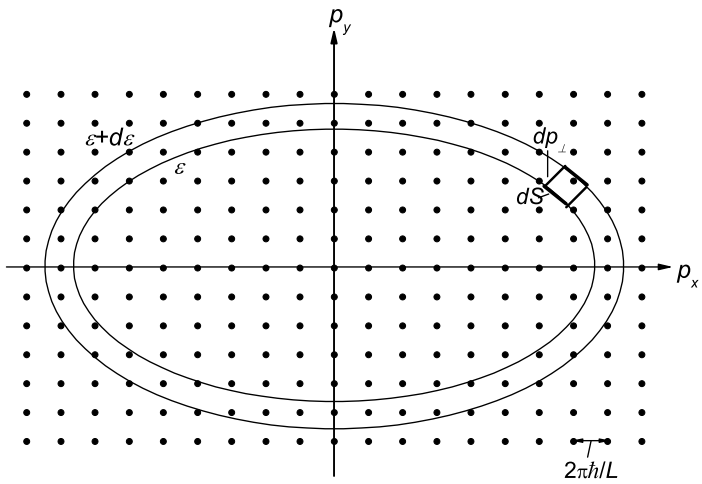

Рис. 2.1: Разрешенные состояния в пространстве импульсов, обозначенные точками, для циклических граничных условий и эквипотенциальные поверхности для анизотропного закона дисперсии (на примере двух измерений). Все пространство будет заполнено, если каждую разрешенную точку окружить d – кубической ячейкой со стороной $2\pi\hbar/L$.

[5]Результаты (2.5), (2.8) могут быть получены и вне рамок метода эффективной массы, поскольку условия (2.1) или (2.2) могут быть наложены и непосредственно на блоховские волновые функции $u_k(x)\exp\frac{i}{\hbar}px$. Действительно, при больших размерах системы можно всегда выбрать L кратным периоду решетки, так что $u_k(x) = u_k(x+L)$ и мы приходим к тем же соотношениям (2.7), (2.8).

Распространение полученного результата на случай пространства произвольной размерности d по сути тривиально. Поскольку в отсутствии внешних полей [6] гамильтониан частицы не зависит от ее координат, то уравнение Шредингера допускает полное разделение переменных и волновая функция в d - мерном ящике (не умаляя общности можно рассмотреть d - мерный "куб" линейного размера L) есть произведение свободно распространяющихся в каждом направлении независимых волн, т. е. плоская волна

$$\Psi_{\mathbf{p}}(\mathbf{R}) = \frac{1}{\sqrt{L^d}} \cdot \exp \frac{i\mathbf{P}_d \cdot \mathbf{R}_d}{\hbar},$$

где в показателе экспоненты стоит скалярное произведение d -мерного импульса на d -мерный радиус вектор. Наложение циклических условий (2.2) на каждую пространственную компоненту $\Psi_{\mathbf{p}}(\mathbf{R})$ приводит, в полной аналогии со случаем одного измерения [7], к дискретному и эквидистантному по каждой координате пространства импульсов спектру (рис. 2.1)

$$\mathbf{P}_d = \left(\frac{2\pi\hbar n_1}{L}, \frac{2\pi\hbar n_2}{L}, \ \ldots \frac{2\pi\hbar n_d}{L}\right), \quad n_i = 0, \ \pm 1, \ \pm 2, \ \pm 3, \ \ldots \qquad (2.11)$$

Объем фазового пространства, приходящийся на одно орбитальное квантовое состояние равен

$$(L\Delta p)^d = (2\pi\hbar)^d. \qquad (2.12)$$

Число квантовых состояний, содержащихся в объеме фазового пространства $L^d\Delta\tau_p$, соответствующего d – мерному интервалу импульсов $\Delta\tau_p = \prod\limits_1^d \Delta p_i$ равно $L^d\Delta\tau_p/(2\pi\hbar)^d$. В расчете на единицу объема координатного и импульсного пространства это число $D_p = 1/(2\pi\hbar)^d$ (плотность состояний в фазовом пространстве) не зависит от положения в импульсном пространстве выбранного интервала $\Delta\tau_p$ (плотность состояний D_p одно-

[6] Согласно теореме Лиувилля о несжимаемости объема фазового пространства для систем, описываемых гамильтоновой функцией, число состояний не изменяется и при наложении внешнего электромагнитного поля.

[7] В двумерном случае цикличные условия в обоих направлениях могут быть реализованы в топологии поверхности тора. Для трехмерной системы выполнение всех трех условий невозможно при любой топологии, однако, как математический прием, эти условия прекрасно работают.

родна и изотропна), как и от размеров использованного для квантования ящика, так что выбор последнего совершенно произволен и в случае произвольной размерности пространства. Увеличение размеров системы просто увеличивает плотность состояний в пространстве импульсов.

2.2 Плотность состояний

При известном квантованном спектре частицы в пространстве импульсов, определяющем плотность состояний в фазовом пространстве, поставленная в начале настоящей главы задача вычисления концентрации сводится к суммированию по всем квантовым состояниям в занимаемой системой фазовом пространстве с весовым множителем, равным вероятности заполнения $f(\mathbf{r}, \mathbf{p})$

$$n_d = \sum_{\mathbf{p}} f(\mathbf{r}, \mathbf{p}).$$

Для элемента объема макроскопического размера спектр в импульсном пространстве можно рассматривать как непрерывный и суммирование можно заменить интегралом по фазовому пространству. Поскольку, согласно (2.12), объем фазового пространства, приходящийся на одно орбитальное квантовое состояние, равен $(2\pi\hbar)^d$, концентрация электронов дается интегралом

$$n_d = \frac{g}{V(2\pi\hbar)^d} \int_{\tau_p} \int_{\tau_r} f(\mathbf{r}, \mathbf{p})) d\tau_{\mathbf{r}} d\tau_{\mathbf{p}} \tag{2.13}$$

где $d\tau_{\mathbf{r}} = \prod\limits_1^d dx_i$ и $d\tau_{\mathbf{p}} = \prod\limits_1^d dp_i$ – элементы объема в d –мерных координатном и импульсном пространствах. Помимо орбитальных квантовых чисел p_i, состояние электрона характеризуется также проекцией спина, а в случае нескольких эквивалентных экстремумов в зоне Бриллюэна и номером долины. В связи с этим в (2.13) введен фактор вырождения $g = g_s g_v$, где g_s – фактор спинового вырождения, равный для электронов 2, и g_v – кратность долинного вырождения. Интегрирование в (2.13) ведется по объему кристалла и первой зоны Бриллюэна. Для однородной

системы вероятность заполнения не зависит от пространственных координат, интегрирование по $d\tau_{\mathbf{r}}$ в (2.13) дает объем системы и выражение для концентрации принимает вид [8]

$$n_d = \frac{g}{(2\pi\hbar)^d} \int_{\tau_p} f(\mathbf{p})) d\tau_{\mathbf{p}}. \tag{2.14}$$

В условиях термодинамического равновесия вероятность заполнения определяется функцией распределения

$$f(\mathbf{p}) = f_0(E(\mathbf{p}), E_F, T) = \frac{1}{\exp\left(\frac{E(\mathbf{p})-E_F}{k_B T}\right) \pm 1} \tag{2.15}$$

где знаки $+$ и $-$ относятся соответственно к распределению Ферми-Дирака (для фермионов) и Бозе-Эйнштейна (для бозонов). Поскольку речь идет о заполнении зонных состояний, характеризуемых квазиимпульсом (законом дисперсии $\varepsilon(\mathbf{p})$), а последние приходятся на область энергий, лежащих выше дна соответствующей энергетической зоны E_c (для дырок ниже потолка зоны), где импульс равен нулю $E(p=0) = E_c$, то в функции распределения должна фигурировать положительная кинетическая энергия зонных носителей $\varepsilon(p) = E(p) - E_c$. Такое описание достигается введением химического потенциала $\mu = E_F - E_c$,

$$f_0(\varepsilon(\mathbf{p}), \mu, T) = \frac{1}{\exp\left(\frac{E(\mathbf{p})-E_c-(E_F-E_c)}{k_B T}\right) \pm 1} = \frac{1}{\exp\left(\frac{\varepsilon(\mathbf{p})-\mu}{k_B T}\right) \pm 1} \tag{2.16}$$

Будучи безразмерной функцией скалярных термодинамических величин (энергии, химического потенциала μ и температуры T) функция распределения одинакова для пространства любой размерности. Поскольку $f_0(\varepsilon(\mathbf{p}), \mu, T)$ зависит от импульса только через энергию, удобно перейти в (2.14) к интегрированию по энергии и ввести плотность состояний в энергетическом пространстве $D_d(\varepsilon)$ (в дальнейшем, просто плотность состояний), которая определяется как число состояний dN в единице объема, содержащихся в интервале энергии между ε и $\varepsilon + d\varepsilon$, отнесенное

[8] Если образец неоднороден, выражение (2.14) можно рассматривать как определение локальной концентрации. Однако такое определение справедливо, если концентрация не изменяется заметным образом на масштабе порядка постоянной решетки и обуславливающие такую неоднородность внешние поля не меняют локального спектра.

к величине этого интервала

$$D_d(\varepsilon) = \frac{g}{(2\pi\hbar)^d}\frac{d}{d\varepsilon}\int d\tau_p. \tag{2.17}$$

В отличие от плотности состояний в импульсном и фазовом пространстве, плотность состояний $D_d(\varepsilon)$ в общем случае зависит как от величины, так и направления импульса. Интегрирование ведется по объему слоя в d –мерном пространстве квазиимпульсов, ограниченного поверхностями постоянной энергии $\varepsilon(\mathbf{p}) = \varepsilon$ и $\varepsilon(\mathbf{p}) = \varepsilon + d\varepsilon$, откуда ясно, что величина $D(\varepsilon)$ тесным образом связана с формой изоэнергетических поверхностей. Задача вычисления $D_d(\varepsilon)$, в сущности, сводится к вычислению объема этого слоя. Будучи выраженной через плотность состояний, концентрация записывается в виде

$$n_d = \int\limits_{(E)} D_d(\varepsilon) f_0(\varepsilon, \mu, T) d\varepsilon. \tag{2.18}$$

В этом легко убедиться, подставляя $D_d(\varepsilon)$ из (2.17) в выражение (2.18), что приводит к формуле (2.14).

Имея в виду вывести общее выражение для плотности состояний, справедливое для произвольной размерности пространства и произвольной формы поверхностей постоянной энергии (произвольного закона дисперсии), при вычислении этого объема удобно перейти от декартовых координат к другой ортогональной системе, в которой координатные поверхности совпадают с изоэнергетическими. В этой системе орбитальное состояние электрона характеризуется его положением на поверхности $\varepsilon(\mathbf{p}) = \varepsilon$. Пусть dS_d - элемент площади на изоэнергетической поверхности d –мерного импульсного пространства, тогда элементом объема слоя является прямой "цилиндр" с основанием dS_d и высотой $dp_\perp$, равной расстоянию между поверхностями $\varepsilon(\mathbf{p}) = \varepsilon$ и $\varepsilon(\mathbf{p}) = \varepsilon + d\varepsilon$ (рис. 2.1; в общем случае значение $dp_\perp$ при переходе от одной точки поверхности к другой может изменяться). Следовательно

$$\int\limits_{(E)} d\tau_p = \int\limits_{S} dS_d dp_\perp. \tag{2.19}$$

Градиент энергии в пространстве квазиимпульсов также направлен вдоль нормали к поверхности, и поэтому величина $|\nabla_{\mathbf{p}}\varepsilon(\mathbf{p})|\, dp_\perp$ есть разность энергий в точках, где нормаль пересекает изоэнергетические поверхности

$$d\varepsilon = |\nabla_{\mathbf{p}}\varepsilon(\mathbf{p})|\, dp_\perp. \tag{2.20}$$

Подставляя отсюда $dp_\perp$ в (2.19) получаем для элемента объема в рассматриваемой ортогональной системе координат

$$dS_d dp_\perp = dS_d \frac{d\varepsilon}{|\nabla_{\mathbf{p}}\varepsilon|} \tag{2.21}$$

(величина $|\nabla_{\mathbf{p}}\varepsilon(\mathbf{p})|^{-1}$ является якобианом перехода от d переменных p_x, p_y, p_z... к переменным ε, ξ, η ..., где ξ, η... – любая ортогональная система координат на поверхности $\varepsilon(\mathbf{p}) = \varepsilon$; $dS_d = d\xi d\eta$)...) и окончательно для плотности состояний имеем

$$D_d(\varepsilon) = \frac{g}{(2\pi\hbar)^d} \int\limits_S \frac{dS_d}{|\nabla_{\mathbf{p}}\varepsilon|} = \frac{g}{(2\pi\hbar)^d} \int\limits_S \frac{dS_d}{v_g}. \tag{2.22}$$

В случае изотропного закона дисперсии $\varepsilon = \varepsilon\left(|\mathbf{p}|\right)$ групповая скорость электрона $v_g = |\nabla_{\mathbf{p}}\varepsilon(\mathbf{p})| = \partial\varepsilon/\partial p$ в (2.22) не зависит от направления импульса (от положения на изоэнергетической поверхности) и может быть вынесена за знак интеграла. Выражение для плотности состояний в этом случае приобретает особенно простой вид

$$D_d(\varepsilon) = \frac{g}{(2\pi\hbar)^d} \frac{1}{\partial\varepsilon/\partial p} \int\limits_S dS_d = \frac{g}{(2\pi\hbar)^d} \frac{S_d}{v(\varepsilon)}. \tag{2.23}$$

Фигурирующая в этой формуле площадь $S_d = S_d(\varepsilon)$ изоэнергетической поверхности $\varepsilon = \varepsilon\left(|\mathbf{p}|\right)$ зависит как от размерности пространства, так и от закона дисперсии, тогда как $v = \partial\varepsilon/\partial p$ определяется только законом дисперсии.

Суммирование по состояниям в пространстве импульсов требуется при вычислении практически всех свойств многочастичной системы. В связи с эти интегралы интегралы вида

$$\int\limits_{\tau_p} f(\varepsilon, \mathbf{p}) \frac{g}{(2\pi\hbar)^d} d\tau_p$$

фигурируют во многих разделах физики конденсированного состояния (и неоднократно встретятся нам в дальнейшем). Согласно изложенному они могут быть приведены к удобному для вычислений виду интегралов по изоэнергетической поверхности и энергии

$$\int\limits_{\tau_p} f(\varepsilon, \mathbf{p}) \frac{g}{(2\pi\hbar)^d} d\tau_p = \int\limits_{\varepsilon} f(\varepsilon, \mathbf{p}) d\varepsilon \int\limits_{S} \frac{g}{(2\pi\hbar)^d} \frac{dS_d}{|\nabla_{\mathbf{p}}\varepsilon|}. \tag{2.24}$$

Если изоэнергетическая поверхность - сфера (изотропный закон дисперсии), интегрирование по поверхности можно свести к интегралу по d – мерному телесному углу

$$dS_d = p^{d-1} d\Omega_d = S_d \frac{d\Omega_d}{\Omega_d},$$

где Ω_d - полный телесный угол и учтено, что $S_d = p^{d-1}\Omega_d$. В результате интеграл (2.24) может быть выражен через плотность состояний (2.23)

$$\int\limits_{\tau_p} f(\varepsilon, \mathbf{p}) \frac{g}{(2\pi\hbar)^d} d\tau_p = \int\limits_{\varepsilon} \oint\limits_{\Omega} f(\varepsilon, \mathbf{p}) D_d(\varepsilon) d\varepsilon \frac{d\Omega_d}{\Omega_d}. \tag{2.25}$$

Если функция $f(\varepsilon, \mathbf{p})$ не зависит от направления волнового вектора, то интегрирование по телесному углу дает Ω и интеграл сводится к интегралу по энергии

$$\int\limits_{\tau_p} f(\varepsilon, p) \frac{g}{(2\pi\hbar)^d} d\tau_p = \int\limits_{\varepsilon} f(\varepsilon, p) D_d(\varepsilon) d\varepsilon. \tag{2.26}$$

В виду широкого использования интегралов по состояниям в пространстве импульсов, плотность состояний, как видим, является чрезвычайно полезным физическим понятием не только с точки зрения вычисления концентраций частиц в квантовой системе.

Вернемся к ее вычислению. Фигурирующая в (2.23) площадь равна $S_3 = 4\pi p^2$ в случае трех измерений и $S_2 = 2\pi p$ в двумерном случае . В одномерном пространстве (очевидно, что одномерная система всегда изотропна) изоэнергетическая поверхность вырождается в две точки, соответствующие двум возможным направлениям импульса $+p$ и $-p$ и ее

площадь $S_1 = 2$. Последнее формально следует из выражения для площади d- мерной сферы $S_d = dV_d/dR$ (объем сферы $V_d = \int dS_d dR$), поскольку для одномерной "сферы" $V_1 = 2R$). Полученные на основании (2.23) выражения для плотности состояний в случае квадратичного и линейного законов дисперсии представлены в табл. 2.1, а характер их зависимости от энергии проиллюстрирован на рис. 2.2.

Таблица 2.1: Плотность состояний как функция квазиимпульса и энергии для изотропных квадратичного и линейного законов дисперсии в случае трех, двух и одного измерений ($g = 2$)

d	S_d	$D(p,\varepsilon)$	$D(p)$		$D(\varepsilon)$	
			$(\varepsilon = \frac{1}{2m}p^2)$	$(\varepsilon = ps)$	$(\varepsilon = \frac{1}{2m}p^2)$	$(\varepsilon = ps)$
$1D$	2	$\frac{2}{\pi\hbar\lvert\nabla_p\varepsilon\rvert}$	$\frac{2m}{\pi\hbar}p^{-1}$	$\frac{2}{\pi s\hbar}$	$\frac{(2m)^{1/2}}{\pi\hbar}\varepsilon^{-1/2}$	$\frac{2}{\pi s\hbar}$
$2D$	$2\pi p$	$\frac{1}{\pi\hbar^2\lvert\nabla_p\varepsilon\rvert}p$	$\frac{m}{\pi\hbar^2}$	$\frac{1}{\pi s^2\hbar^2}p$	$\frac{(2m)}{2\pi\hbar^2}$	$\frac{1}{\pi s^2\hbar^2}\varepsilon$
$3D$	$4\pi p^2$	$\frac{1}{\pi^2\hbar^3\lvert\nabla_p\varepsilon\rvert}p^2$	$\frac{m}{\pi^2\hbar^3}p$	$\frac{1}{\pi^2 s^3\hbar^3}p^2$	$\frac{(2m)^{3/2}}{2\pi^2\hbar^3}\varepsilon^{1/2}$	$\frac{1}{\pi^2 s^3\hbar^3}\varepsilon^2$

В частном случае изотропной двумерной системы плотность состояний согласно (2.23) равна (далее всюду полагается $g = 2$)

$$D_{2D}(\varepsilon) = \frac{1}{\pi\hbar^2}\frac{p}{d\varepsilon/dp} \tag{2.27}$$

и для электронного газа со стандартным спектром $\varepsilon = p^2/2m$ не зависит от импульса, а, следовательно, и от энергии

$$D_{2D}(\varepsilon) = \frac{m}{\pi\hbar^2}. \tag{2.28}$$

На первый взгляд, этот необычный (для трехмерной физики) результат определяет особое место двумерных систем. Действительно, поскольку $S_d \propto p^{d-1}$, а групповая скорость $v_g = d\varepsilon/dp = p/m$, то, согласно (2.23)

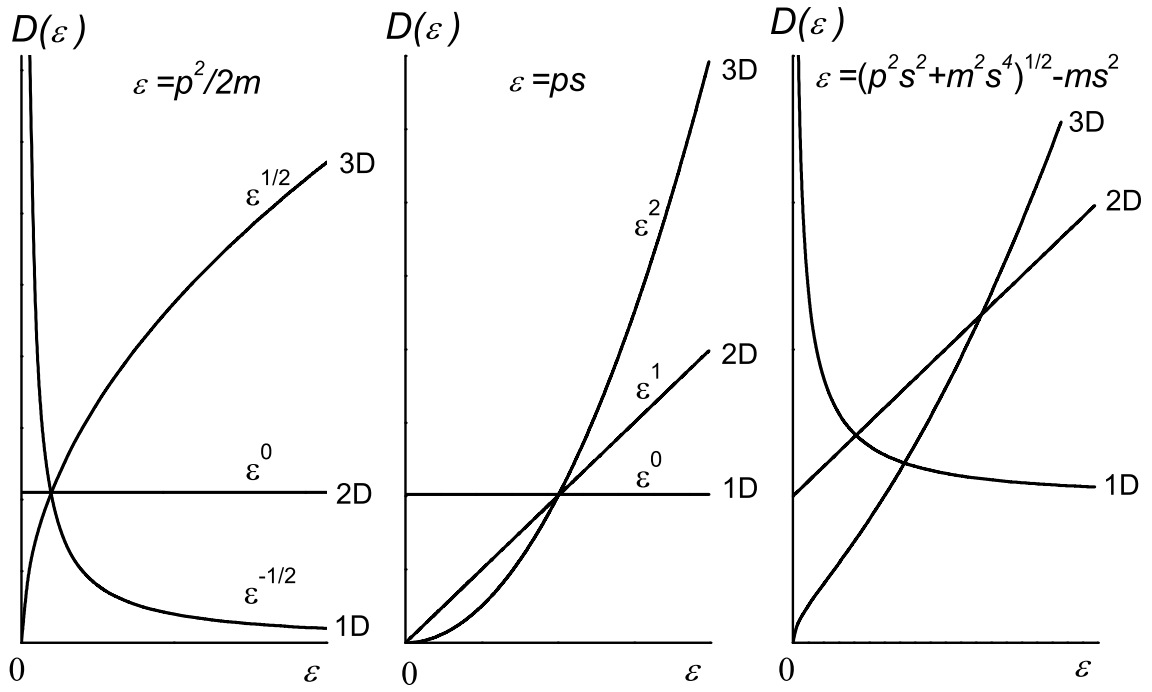

Рис. 2.2: Характер зависимости плотности состояний от энергии для электронного газа различной размерности при квадратичном, линейном и кейновском (в его двухзонной аппроксимации) законах дисперсии . Заметьте, что пересечение кривых носит фиктивный характер – плотности состояний для различных размерностей пространства измеряются в разных физических единицах, т. е. по величине не могут сравниваться друг с другом. Рисунки только иллюстрируют различный характер энергетических зависимостей плотностей состояний.

плотность состояний $D_d \propto S_d/v_g \propto p^{d-2}$, т. е. не зависит от p только при $d = 2$. Однако это не совсем так. Независимость двумерной плотности состояний от энергии имеет место *только при квадратичном законе дисперсии* и является, по-существу, следствием того, что, как объем двумерного пространства квазиимпульсов $\tau_{\mathbf{p}}(\mathbf{p})$, так и энергия $\varepsilon(\mathbf{p})$, являются квадратичными функциями импульса. Например, в более общем случае кейновского закона дисперсии

$$\varepsilon(p) = \sqrt{p^2 s^2 + (ms^2)^2} - ms^2. \tag{2.29}$$

формула (2.27) приводит к зависящей от энергии плотности состояний

$$D_{2D}(\varepsilon) = \frac{ms^2 + \varepsilon}{\pi \hbar^2 s^2}. \tag{2.30}$$

При $\varepsilon << ms^2$ (2.30) переходит в (2.27), тогда как в пределе высоких энергий (или в пределе $ms^2 \to 0$, что соответствует линейному закону дисперсии) D_{2D} линейно возрастает с энергией.

Нетрудно видеть, что по той же причине (одинаковой, линейной зависимости $\tau_{\mathbf{p}}$ и ε от импульса) плотность состояний не зависит от энергии и в одномерной системе с линейным законом дисперсии. Действительно, для $d = 1$ из (2.23) следует

$$D_{1D}(\varepsilon) = \frac{1}{\pi\hbar}\frac{2}{d\varepsilon/dp} \tag{2.31}$$

и для дисперсионного соотношения $\varepsilon = ps$ имеем

$$D_{1D}(\varepsilon) = \frac{2}{\pi s\hbar} = const \tag{2.32}$$

(рис. 2.2). Отметим, что одномерные системы проявляют особое поведение и в случае квадратичного закона дисперсии, а именно – одномерная плотность состояний

$$D_{1D}(\varepsilon) = \frac{2m}{\pi\hbar p(\varepsilon)} = \frac{\sqrt{2m}}{\pi\hbar\sqrt{\varepsilon}} \tag{2.33}$$

расходится при $\varepsilon \to 0$. Однако, поскольку такая корневая расходимость интегрируема, она не приводит к каким либо особенностям при вычислении концентрации (2.18). Для кейновского дисперсионного соотношения, подставляя в (2.31) $dp/d\varepsilon$ из (2.29), для одномерной плотности состояний получаем

$$D_{1D}(\varepsilon) = \frac{2}{\pi\hbar s}\frac{\varepsilon + ms^2}{\sqrt{\varepsilon(\varepsilon + 2ms^2)}}. \tag{2.34}$$

Если масса покоя не равна нулю, то плотность состояний D_{1D} имеет ту же корневую особенность при $\varepsilon = 0$, что и для параболического закона дисперсии (2.33), поскольку для $m \neq 0$ кейновский спектр при малых энергиях квадратичен.

2.2.1 Обобщенное выражение для плотности состояний при степенном законе дисперсии

Из представленных в табл. 2.1 выражений для плотности состояний легко увидеть, что зависимости $D_d(\varepsilon)$ могли бы быть представлены общим

для всех размерностей пространства выражениями (но различными для разных законов дисперсии)

$$D_d(\varepsilon) \propto \left(\frac{\sqrt{2m}}{\hbar}\right)^d \varepsilon^{\frac{d-2}{2}} \quad \textit{для} \quad \varepsilon = p^2/2m, \tag{2.35a}$$

$$D_d(\varepsilon) \propto \left(\frac{1}{s\hbar}\right)^d \varepsilon^{d-1} \quad \textit{для} \quad \varepsilon = ps, \tag{2.35b}$$

если бы не отличия в численных множителях, содержащих степени 2 и π. Последнее связано с тем, что использованные выше стандартные формулы для площади поверхности d – мерной сферы, выражающие ее через число π, различны для пространств четной и нечетной размерности (то же имеет место и для объема шара)

$$S_d = \frac{d2^d\pi^{\frac{d-1}{2}}\left(\frac{d-1}{2}\right)!}{d!}R^{d-1} \quad (d = 1, 3, 5...), \tag{2.36a}$$

$$S_d = \frac{d\pi^{\frac{d-1}{2}}}{\left(\frac{d}{2}\right)!}R^{d-1} \quad (d = 2, 4, 6...). \tag{2.36b}$$

Это неудобство может быть обойдено, если воспользоваться гамма функцией $\Gamma(x)$, определяемой для положительных значений аргумента интегралом Эйлера второго рода

$$\Gamma(x) = \int\limits_0^\infty t^{x-1}e^{-t}dt. \tag{2.37}$$

Для наших целей понадобятся только значения $\Gamma(x)$ при целых и полуцелых аргументах, когда $\Gamma(x)$ определяется очень простыми формулами. Для $x = 1$ функция (2.37) есть интеграл от экспоненты и ответ тривиален $\Gamma(1) = 1$. Для $x = \frac{1}{2}$ интеграл (2.37) элементарно сводится к хорошо известному интегралу ошибок $\Gamma(\frac{1}{2}) = 2\int\limits_0^\infty e^{-t^2}dt = \sqrt{\pi}$. Все другие значения гамма функции $\Gamma(n)$ и $\Gamma(n+\frac{1}{2})$ при целых n могут быть найдены из

$\Gamma(1)$ и $\Gamma(\frac{1}{2})$ с помощью рекуррентного соотношения [9],

$$\Gamma(x+1) = x\Gamma(x) \tag{2.38}$$

которое приводит к простым выражениям

$$\Gamma(n) = 1 \cdot 2 \cdot 3 \cdot \cdot (n-1) = (n-1)!$$

и

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n}\sqrt{\pi}.$$

Для представляющих в дальнейшем интерес аргументов значения гамма функции приведены в таблице

x	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4
$\Gamma(x)$	$\sqrt{\pi}$	1	$\frac{\sqrt{\pi}}{2}$	1	$\frac{3\sqrt{\pi}}{4}$	2	$\frac{15\sqrt{\pi}}{8}$	6

С помощью гамма функции площадь поверхности и объем d – мерной сферы радиуса R записываются в единой для всех размерностей форме [10]

$$S_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}R^{d-1}, \quad V_d = \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}R^d. \tag{2.39}$$

Используя первое из этих соотношений для пространства квазиимпульсов, получаем из (2.23) (при $g = 2$) формулу

$$D_d(\varepsilon) = \frac{1}{2^{d-2}\pi^{\frac{d}{2}}\hbar^d\Gamma(\frac{d}{2})}\frac{[p(\varepsilon)]^{d-1}}{d\varepsilon/dp}. \tag{2.40}$$

[9] Интегрированием по частям имеем

$$\Gamma(x+1) = \int\limits_0^\infty t^x e^{-t}dt = \left[-t^x e^{-t}\right]_0^\infty + x\int\limits_0^\infty t^{x-1}e^{-t}dt = x\Gamma(x).$$

[10]

$$S_d = R^{d-1}\int_0^\pi \sin^{d-2}(\theta_{d-2})d\theta_{d-2}\ldots\int_0^\pi \sin(\theta_1)d\theta_1\int_0^{2\pi} d\varphi = R^{d-1}\oint d\Omega_d = R^{d-1}\Omega_d,$$

где $d\Omega_d$ элемент d – мерного телесного угла, Ω_d – полный телесный угол.

описывающую зависимость плотности состояний от размерности пространства d, параметров (изотропного) закона дисперсии и энергии. Для $\varepsilon = p^2/2m$ отсюда получаем

$$D_d(\varepsilon) = \frac{1}{2^{d-1}\Gamma\left(\frac{d}{2}\right)} \left(\frac{2m}{\pi\hbar^2}\right)^{\frac{d}{2}} \varepsilon^{\frac{d-2}{2}}. \tag{2.41}$$

Для линейного закона дисперсии $\varepsilon = ps$

$$D_d(\varepsilon) = \frac{1}{2^{d-2}\Gamma\left(\frac{d}{2}\right)} \left(\frac{1}{\sqrt{\pi}s\hbar}\right)^{d} \varepsilon^{d-1}. \tag{2.42}$$

Однако и последние две формулы могут быть объединены в одну. Для степенной зависимости энергии от импульса

$$\varepsilon(p) = c_r p^r \tag{2.43}$$

($c_r = s$ для линейного и и $c_r = 1/2m$ для квадратичного закона дисперсии) последний сомножитель в (2.40) выражается через энергию в виде

$$p^{d-1}/(d\varepsilon/dp) = r^{-1} c_r^{-d/r} \varepsilon^{d/r-1}$$

и мы приходим к *обобщенной* (на случай произвольной размерности пространства и показателя степени в дисперсионном соотношении) формуле для плотности состояний

$$D_d(\varepsilon) = \frac{1}{r2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d} \left(\frac{1}{c_r^{1/r}\hbar}\right)^{d} \varepsilon^{\frac{d-r}{r}} \propto \left(\frac{1}{c_r^{1/r}\hbar}\right)^{d} \varepsilon^{\frac{d-r}{r}}. \tag{2.44}$$

Из этого выражения видно, что, если размерность пространства совпадает с показателем степени в законе дисперсии ($d = r$), то плотность состояний от энергии не зависит, как это было уже показано для одномерных и двумерных систем. Для свободного электрона закон дисперсии с $r = 3$ физически бессмысленнен (в принципе, это справедливо и для для зонных электронов в полупроводниках [11]) и трехмерная плотность состояний всегда зависит от энергии. Отметим, что в случае произвольной зависимости $\varepsilon(p)$, последняя, как правило, может быть разложена в

[11] Это не исключает поправок к квадратичному спектру, содержащих члены с $r > 2$.

сходящийся или асимптотический ряд по степеням импульса и формула (2.44) обеспечивает, в принципе, описание плотности состояний для любой системы с изотропным спектром.

2.3 Эффективные массы зонных носителей

Законы дисперсии, как мы видим, полностью определяют как динамику ($\mathbf{v}(\mathbf{p}) = \nabla\varepsilon(\mathbf{p})$), так и статистику зонных носителей заряда. Простейший, квадратичный, закон дисперсии, характеризуемый единственным параметром – эффективной массой, играет особую роль в физике полупроводников. Дело не только, и даже не столько, в том , что он хорошо описывает зонный спектр большинства полупроводников вблизи краев их энергетических зон, состояния которых дают основной вклад в электронные свойства трехмерных кристаллов. Будучи простейшим соотношением, связывающим энергию с импульсом [12], которое вытекает из однородности пространства и времени, т. е. из наиболее общих симметрийных соображений, квадратичный закон дисперсии описывает спектр

[12]На первый взгляд, простейшим является линейный закон дисперсии $\varepsilon = ps$. В действительности это, конечно, не так, поскольку водящая сюда величина p есть модуль импульса, т. е квадратичная форма его компонент. Линейный закон дисперсии в его явной покомпонентной записи

$$\varepsilon = s\sqrt{\sum_{i=1}^{d} p_i^2},$$

разумеется, не только сложнее квадратичного, но и не обладает одной из наиболее важных особенностей последнего – аддитивностью энергии по отношению к ее составляющим, относящимся к различным (ортогональным) степеням свободы (отметим, что это определяет и особое место декартовой системы координат в физике)

$$\varepsilon(\mathbf{p}) = \sum_{i=1}^{d} \frac{p_i^2}{2m_i} = \sum_{i=1}^{d} \varepsilon_i(p_i).$$

При линейном законе дисперсии (как и при любом другом неквадратичном дисперсионном соотношении) кинетическая энергия не равна сумме кинетических энергий, соответствующих движению вдоль ортогональных направлений в пространстве

$$\varepsilon = s\sqrt{\sum_{i=1}^{d} p_i^2} \neq \sum_{i=1}^{d} \varepsilon_i = s\sum_{i=1}^{d} p_i.$$

Движения в различных направлениях, в том числе, что наиболее важно, во внешних полях, оказываются взаимозависимыми. Линейный закон дисперсии может однако рассматриваться, как *простейший для одномерного движения*, поскольку для однокомпонентного импульса $\varepsilon(p) = s\sqrt{p^2} = sp$ (импульс вырождается в псевдоскаляр), а проблемы перемешивания движений для одномерной системы попросту не возникает.

свободного электрона, т. е. лежит в основе как классической, так и квантовой механики (в их нерелятивистском варианте), в том числе, разумеется, и физики конденсированного состояния, рассматривающей электроны проводимости, как свободные электроны в периодическом потенциале решетки. В связи с этим возникает естественное желание описать и само поведение зонных электронов, в том числе и во внешних полях, по аналогии с хорошо понятым движением свободных электронов, вплоть до использования близких по смыслу понятий, таких как масса зонных носителей, их скорость, импульс и т. п. Уже само по себе введение понятия о зонных законах дисперсии, равно как и развитие метода эффективной массы, есть, в определенном смысле, та же дань механике свободного электрона. Другая, экспериментальная, сторона такой тенденции связана с тем обстоятельством, что целый ряд эффектов в полупроводниках и металлах удобно и полезно интерпретировать с использованием таких ясных по смыслу понятий, как циклотронная или оптическая эффективная масса, эффективная масса плотности состояний и т.д. Такие явления, как циклотронный или магнитофононный резонанс, целый ряд магнитооптических эффектов являются по-существу прямыми методами определения этих параметров.

Конечно, проведение аналогии со свободным электроном весьма ограничено. Уже из того факта, что энергия зонных электронов есть периодическая функция квазиимпульса, следует, что их закон дисперсии не может быть сведен к квадратичному по всей зоне и, следовательно, не может быть описан единственным параметром. Число и значения фигурирующих в законах дисперсии зонных электронов материальных параметров определяются симметрией и величиной кристаллического потенциала. Так, в изотропной модели Кейна зонный спектр характеризуется эффективной массой m, кейновской скоростью s и спин-орбитальным расщеплением валентной зоны Δ_{so}, а ее двухзонный вариант описывается только первыми двумя из этих параметров (в последнем случае может быть проведена аналогия с релятивистским электроном, также характеризуемым двумя параметрами - массой покоя и скоростью света).

Если аналогичное свободному электрону однопараметрическое опи-

сание зонных электронов и возможно, то только в ограниченной области энергий, обычно вблизи экстремумов энергетических зон. В кристаллах с анизотропными зонами, даже при квадратичной зависимости энергии от компонент импульса, спектр не может быть охарактеризован единственной скалярной массой. Спектр и движение электронов в этом случае должен описываться тензорной величиной (тензором обратной эффективной массы) с несколькими независимыми компонентами. Очевидно, что в случае произвольного (поэтому многопараметрического) закона дисперсии величины, выступающие в качестве массы в том или ином явлении, должны быть, по меньшей мере, некоторой комбинацией описывающих дисперсионное соотношение параметров, а при неквадратичной дисперсии будут зависеть и от компонент импульса (или его модуля для изотропного спектра). Эти комбинации и зависимости различны, вообще говоря, для различных эффектов, используемых для измерения такого рода эффективных масс.

Так как законы дисперсии полностью описывает электронную систему в кристалле и ее поведение при внешних воздействиях, то введение новых характеристик (таких как различного рода эффективные массы) выражающихся, в конечном счете, через те же параметры законов дисперсии, выглядит, по-крайней мере, излишним (зачастую, это так и есть). Однако, даже если отвлечься от полезности аналогии со свободным электроном, что само по себе имеет несомненную эвристическую ценность, некоторые "специфические" эффективные массы оказываются в ряде случаев весьма полезными и широко используются при описании электронных свойств полупроводников и металлов, как и структур на их основе. Рассмотрение этих парметров имеет немаловажное значение и в методологическом отношении. Выяснение того, какая величина играет роль, аналогичную массе в случае свободного электрона, в том или ином явлении, как она зависит от вида спектра и зонных параметров или размерности пространства, несомненно, способствует лучшему пониманию физики явления, механизмов лежащих в его основе.

2.3.1 Эффективная масса плотности состояний

Одним из наиболее часто употребляемых понятий такого рода, является эффективная масса, вводимая для описания плотности состояний. Последняя является, пожалуй, наиболее важным понятием физики конденсированных сред, поскольку определяет такие важнейшие их характеристики, как концентрация носителей, их распределение по энергетическим состояниям, плотность состояний на уровне Ферми, играющую решающую роль в металлах, вырожденных полупроводниках и, особенно, в квантовых гетероструктурах. Эти же характеристики, в свою очередь, необходимы для описания любых явлений в кристалле, будь то кинетические, магнитные, оптические или другие свойства. Удобство введения и использования эффективной массы для описания плотности состояний обусловлено в первую очередь тем, что плотность состояний является скалярной величиной и характеризующая ее масса даже в случае анизотропного закона дисперсии выражается через простые скалярные комбинации компонент тензора обратной эффективной массы (см. ниже раздел 2.3.2).

Согласно выражению (2.22) плотность состояний в общем виде определяется интегралом по изоэнергетической поверхности от величины, обратной скорости электрона $v(\mathbf{p}) = |\nabla_{\mathbf{p}}\varepsilon|$

$$D_d(\varepsilon) = \frac{2}{(2\pi\hbar)^d} \int\limits_S \frac{dS_d}{|\nabla_{\mathbf{p}}\varepsilon|}.$$

Ничто однако не мешает представить эту *скалярную* величину в той же форме, что и выражение для плотности состояний в случае квадратичного закона дисперсии (2.41)

$$D_d(\varepsilon) = \frac{2}{2^d\Gamma\left(\frac{d}{2}\right)} \left(\frac{2m_D}{\pi\hbar^2}\right)^{\frac{d}{2}} \varepsilon^{\frac{d-2}{2}},$$

в котором однако величина m_D уже не совпадает с массой в законах дисперсии, а должна быть определена из условия, чтобы формула (2.41) приводила к тому же результату, что и (2.40). Приравнивая правые части (2.22) и (2.41) получаем обобщенное выражение для величины m_D, называемой *эффективной массой плотности состояний*

$$m_D(\varepsilon) = \frac{1}{2\pi}\left[\Gamma\left(\frac{d}{2}\right)\int\limits_S \frac{dS_d}{|\nabla_{\mathbf{p}}\varepsilon|}\right]^{2/d} \varepsilon^{\frac{2-d}{d}} \tag{2.45}$$

Из этого выражения видно, что и в этом отношении двумерные системы занимают особое положение. При $d = 2$ выражение (2.45) максимально упрощается и приобретает вид, с точностью до постоянного (размерного) коэффициента, эквивалентный выражению для плотности состояний. Таким образом плотность состояний может быть записана в форме

$$D_{2D}(\varepsilon) = \frac{m_D(\varepsilon)}{\pi\hbar^2}, \tag{2.46}$$

аналогичной выражению (2.28) для D_n при квадратичном спектре , но с эффективной массой $m \to m_D$. Подчеркнем, что это соотношение справедливо для двумерных систем *с любом законом дисперсии.* Для любой другой размерности пространства функциональные зависимости $D_n(\varepsilon)$ и $m_D(\varepsilon)$ отличаются и столь простое сопоставление плотности состояний и эффективной массы не может быть проведено.

2.3.2 Анизотропный квадратичный закон дисперсии

Покажем полезность введения эффективной массы плотности состояний на примере *параболической анизотропной* энергетической *зоны*, когда изоэнергетические поверхности представляют собой d – мерные эллипсоиды (конечно, в одномерной системе такой "эллипсоид" – всегда "сфера", сводящаяся к двум точкам) и закон дисперсии имеет вид

$$\varepsilon(\mathbf{p}) = \sum_{i=1}^{d} \frac{p_i^2}{2m_i}, \tag{2.47}$$

где m_i – эффективные массы вдоль главных осей эллипсоида. Для введения m_D в этом случае удобно исходить не из общего определения (2.45), а воспользоваться простым преобразованием координат в пространстве импульсов (см. рис. 2.3.2)

$$p_i = \sqrt{m_i}p_i'.$$

Закон дисперсии при этом преобразуется в *изотропный квадратичный*

$$\varepsilon(\mathbf{p}') = \sum_{i=1}^{d} p_i'^2.$$

При указанном преобразовании, пространство импульсов деформируется таким образом, что плотность состояний в фазовом пространстве $D_p = (2\pi\hbar)^{-d}$ изменяется согласно соотношению

$$D_p = \left(\prod_{i=1}^{d} m_i\right)^{1/2} D_p'.$$

В результате задача вычисления плотности состояний полностью сводится к уже решенной нами задаче для изотропного квадратичного закона дисперсии с точностью до замены в (2.41) изотропной массы на *эффективную массу плотности состояний* $m_D = \left(N_m^2 \prod_{i=1}^{d} m_i\right)^{1/d}$, где N_m – число эллипсоидов в первой зоне Бриллюэна.

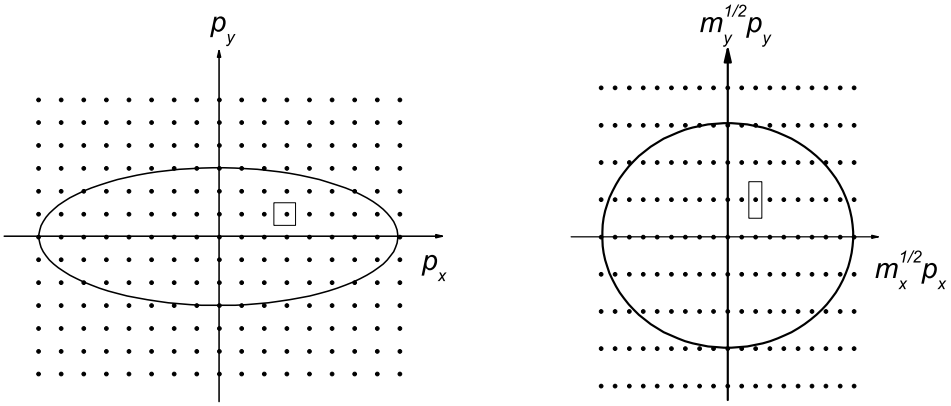

Рис. 2.3: К вычислению плотности состояний для электронного газа с анизотропным квадратичным законом дисперсии.

2.3.3 Изотропный закон дисперсии

В случае изотропного, но произвольного во всех других отношениях закона дисперсии плотность состояний в d – мерном пространстве определяется формулой (2.40). Приравнивая это выражение для $D_n(\varepsilon)$ выражению (2.41) для квадратичного спектра получаем обобщенную формулу

для *эффективной массой плотности состояний*

$$m_D(\varepsilon) = \frac{1}{2}\left[\frac{2p^{d-1}}{d\varepsilon/dp}\right]^{2/d} \varepsilon^{\frac{2-d}{d}} \tag{2.48}$$

в энергетической зоне с изотропным спектром.

В общем случае m_D зависит как от параметров закона дисперсии, так и от энергии и размерности пространства. Вытекающие из (2.48) общие выражения для m_D и выражения для частного случая линейного закона дисперсии при разных размерностях пространства приведены в табл. 2.2.

Таблица 2.2: Эффективная масса плотности состояний m_D и динамическая масса m_c для одномерного двумерного и трехмерного электронного газа. В последних двух столбцах приведены значения m_D и m_c для линейного закона дисперсии.

d	$m_D(\varepsilon)$	$m_D(\varepsilon = \varepsilon\lvert p\rvert)$	$m_D(\varepsilon = ps)$	$m_c(\varepsilon = ps)$
$1D$	$2\varepsilon\left(\frac{d\varepsilon}{dp}\right)^{-2}$	$2\varepsilon\left(\frac{d\varepsilon}{dp}\right)^{-2}$	$\frac{2}{s^2}\varepsilon$	$\frac{1}{s^2}\varepsilon$
$2D$	$\frac{1}{2\pi}\int\limits_S \frac{dS}{\lvert\nabla_p\varepsilon\rvert}$	$\frac{pdp}{d\varepsilon}$	$\frac{1}{s^2}\varepsilon$	$\frac{1}{s^2}\varepsilon$
$3D$	$\frac{1}{2}\left(\frac{1}{2\pi\sqrt{\varepsilon}}\int\limits_S \frac{dS}{\lvert\nabla_p\varepsilon\rvert}\right)^{2/3}$	$\frac{1}{2}\left(\frac{2p^2}{\sqrt{\varepsilon}}\frac{dp}{d\varepsilon}\right)^{2/3}$	$\frac{1}{\sqrt[3]{2}s^2}\varepsilon$	$\frac{1}{s^2}\varepsilon$

Введением эффективной массы m_D, продиктованным стремлением оперировать привычными соотношениями для свободного электрона, более сложная функциональная зависимость плотности состояний от энергии, по сравнению с таковой для простейшего (квадратичного) дисперсионного соотношения, переносится, попросту, на эффективную массу $m_D(\varepsilon)$. В случае сложных законов дисперсии такого рода улучшения похожи, конечно, на "тришкин кафтан". В принципе, любой произвольный

закон дисперсии и сам по себе может быть сведен к квадратичному введением соответствующей "эффективной массы для закона дисперсии", зависящей от параметров исходного закона дисперсии и от энергии (а в случае анизотропных спектров и от компонент импульса). Поскольку такого рода эффективные массы лишаются основного достоинства, которым обладает масса свободной частицы, а именно постоянства, то при сложной зависимости от импульса (энергии) таких "специальных" эффективных масс более разумно оперировать непосредственно законами дисперсии. Повторим, что это в значительной мере не так в случае двумерных систем, для которых как общее выражение для эффективной массы $m_D(\varepsilon)$ (формула (2.45) при $d = 2$)

$$m_D(\varepsilon) = \frac{1}{2\pi} \int\limits_S \frac{dS}{|\nabla_p \varepsilon|}, \tag{2.49}$$

так и его изотропное приближение (формула (2.48) при $d = 2$)

$$m_D(\varepsilon) = \frac{p}{d\varepsilon/dp}, \tag{2.50}$$

по существу эквивалентны выражениям (2.22) и (2.27) для самой плотности состояний.

2.3.4 Динамическая эффективная масса

Эффективная масса является не только материальным параметром, характеризующим плотность состояний, но и важнейшей характеристикой динамических свойств частицы. Для свободного электрона масса определяет соотношение между импульсом и скоростью $\mathbf{v}(\mathbf{p}) = m^{-1}\mathbf{p}$, а, следовательно, и нерелятивистское уравнение движения $d\mathbf{p}/dt = md\mathbf{v}/dt = \mathbf{F}$. В случае произвольного закона дисперсии скорость частицы определяется общим соотношением $\mathbf{v}(\mathbf{p}) = \nabla_{\mathbf{p}}(\varepsilon)$. Если мы хотим описать связь между скоростью и импульсом в виде, аналогичном таковой для свободного электрона

$$\mathbf{v}(\mathbf{p}) = \nabla_{\mathbf{p}}(\varepsilon) = \mathbf{m}_v^{-1}\mathbf{p}, \tag{2.51}$$

т. е. используя понятие массы ($\mathbf{m}_v$), то последняя в общем случае будет функцией импульса. При анизотропном зонном спектре величина $\mathbf{m}_v^{-1}$ в

(2.51), которую естественно назвать *обратной динамической эффективной массой*, как нетрудно видеть, есть симметричный d – мерный тензор второго ранга

$$\mathbf{m}_v^{-1} = \nabla_{\mathbf{p}} \nabla_{\mathbf{p}}(\varepsilon)$$

с компонентами [13]

$$\left(\frac{1}{m_v}\right)_{ij} = \frac{\partial^2 \varepsilon}{\partial p_i \partial p_j}. \tag{2.52}$$

При сферической форме поверхностей постоянной энергии такой тензор вырождается в скаляр и мы можем, используя определяющее величину m_v соотношение (2.51) в его скалярной форме (для модулей векторов $\nabla_{\mathbf{p}}(\varepsilon)$ и $\mathbf{p}$, поскольку они коллинеарны), ввести скалярную *динамическую эффективную массу*

$$m_v = \frac{p}{d\varepsilon/dp}. \tag{2.53}$$

В случае кейновского дисперсионного соотношения формула (2.53) дает

$$m_v(\varepsilon) = m + \frac{\varepsilon}{s^2}, \tag{2.54}$$

откуда в частном случае $m = 0$ следует выражение для динамической эффективной массы при линейном законе дисперсии

$$m_v(\varepsilon) = \frac{\varepsilon}{s^2}. \tag{2.55}$$

Сравним эту величину с эффективной массой плотности состояний для изотропного спектра. Рассмотрим вначале случай двух измерений. Ранее мы уже нашли двумерную плотность состояний для кейновского закона дисперсии (формула (2.28))

$$D_{2D}(\varepsilon) = \frac{ms^2 + \varepsilon}{\pi \hbar^2 s^2}.$$

Сопоставляя это выражение с формулой (2.30) для двумерной плотности состояний в параболической зоне $D_{2D}(\varepsilon) = m/\pi\hbar^2$, находим эффективную массу для двумерной плотности состояний

$$m_D(\varepsilon) = m + \frac{\varepsilon}{s^2} \tag{2.56}$$

[13] Обратные им величины $\left(\frac{1}{m_v}\right)^{-1}_{ij}$ не являются тензором, поскольку величины, обратные компонентам тензора, вообще говоря, тензора не образуют .

(очевидно, что это частный случай нашего общего утверждения (2.46)). Как видим, в случае кейновского закона дисперсии и двумерного электронного газа, определения для эффективной массы плотности состояний и динамической эффективной массы (последняя в изотропном случае от размерности пространства не зависит) эквивалентны. Эффективные массы (2.56) и (2.54) совпадают с фигурирующей в кейновском законе дисперсии массой m ("массой покоя") только при $\varepsilon = 0$, возрастая линейно с энергией со "скоростью", определяемой вторым параметром (s) кейновского дисперсионного соотношения. Как и должно быть, в области малых энергий, или в "нерелятивистском" пределе $s \to \infty$, массы m_D и m_v близки к "истинной" эффективной массе m закона дисперсии, тогда как в "ультрарелятивистском" пределе $\varepsilon >> ms^2$ переходят в соответствующие выражения для линейного закона дисперсии (см. табл. 2.2).

Вывод об эквивалентности динамической массы и массы для плотности состояний в случае двух измерений носит, между тем, более общий характер, а, именно, он остается верным для любого закона дисперсии, лишь бы он был изотропен. Это сразу следует из сопоставления выражений (2.50) для m_D и (2.53) для m_v. В дальнейшем (Гл. 6) мы дадим определение еще для одной "специальной" массы, пожалуй наиболее широко используемой в физике полупроводников и металлов, – циклотронной эффективной массы m_c, характеризующей движение носителей заряда в магнитном поле, как при его классическом рассмотрении (m_c определят частоту циклотронного вращения носителей заряда), так и при квантовом (m_c определяет квантование спектра в магнитном поле). Как мы увидим, зависимость циклотронной массы от параметров изотропного закона дисперсии определяется тем же соотношением (2.53), так что для двумерного изотропного электронного газа все три упомянутые выше определения эффективных масс оказываются эквивалентными (но не совпадающими с массой в законах дисперсии!).

Из общей формулы (2.48) нетрудно видеть, что при любой другой размерности пространства $d \neq 2$ эффективная масса для плотности состояний не может быть представлена в виде (2.53) и, следовательно, m_D

не совпадает с m_v и m_c. Ранее мы отмечали, что двумерная система занимает особое место в том отношении, что только для двух измерений плотность состояний для стандартного (квадратичного) закона дисперсии не зависит энергии. Теперь мы видим, что двумерные системы специфичны в более широком отношении, а именно, при произвольном изотропном законе дисперсии в двумерной, и только в двумерной системе, эффективные массы для плотности состояний равны циклотронным и динамическим эффективным массам. Более того, как будет далее показано (см. Гл. 6), тождественность между m_D и m_c для двумерных систем имеет место при любом (не обязательно изотропном) законе дисперсии.

Излишне проверять, что для изотропного параболического спектра формулы (2.50) и (2.53), равно как и общие формулы (2.45), (2.48) и (2.52), приводят для произвольной размерности пространства к значениям m_D и m_v, совпадающим с эффективной массой закона дисперсии m. Ничего иного и не должно получиться, поскольку эти соотношения выведены как раз из требования, чтобы выражения для плотности состояний и скорости были аналогичны таковым для свободного электрона (с точностью до замены $m_0 \to m$). Единственно, что дает такая проверка, это уверенность в том, что при выводе этих формул мы не сделали ошибок в выкладках!

2.4 Концентрация электронов в d − мерном электронном газе

При отсчете энергий от дна зоны $E_c = 0$ (в общем случае $\varepsilon(\mathbf{p}-\mathbf{p}_0)-E_c = 0$, где $\mathbf{p}_0$ – соответствующий нулевой кинетической энергии квазиимпульс; для энергетических зон, минимумы которых расположены в центре зоны Бриллюэна $\mathbf{p}_0 = 0$) выражение (2.18) для концентрации электронов после подстановки функции распределения (2.16) принимает вид

$$n = \int\limits_0^\infty D_d(\varepsilon)\frac{d\varepsilon}{\exp\left(\frac{\varepsilon-\mu}{kT}\right)+1} \tag{2.57}$$

(поскольку химпотенциал μ есть энергетическое расстояние между уровнем Ферми E_F и дном зоны проводимости $\mu = E_F - E_c$, то при отсчете энергий от дна зоны проводимости обозначения μ и E_F эквивалентны). Интегрирование в (2.57) ведется от дна зоны (при нашем отсчете энергий $E_c = 0$) до ее потолка, однако, ввиду больших значений ширины разрешенных зон и экспоненциального спадания функции распределения с ростом энергии, замена верхнего предела в интеграле (2.57) на ∞ не может сколь либо заметно сказаться на точности. Хотя вычисление не берущегося в квадратурах при произвольных температурах даже для простейших законов дисперсии интеграла в (2.57) не представляет особого труда для компьютерного расчета, мы вначале рассмотрим два предельных случая, допускающих простые аналитические аппроксимации . Это представляет интерес не только с точки зрения наглядности аналитического представления зависимости концентрации носителей от размерности пространства, зонных параметров материала и температуры, но и в виду физической важности этих двух пределов: случаев классического и вырожденного газа зонных носителей.

2.4.1 Вырождение электронного газа

Термин “вырожденный газ” возник в теории идеального газа частиц при описании его свойств, когда эти свойства отличаются от свойств классического идеального газа вследствие взаимного квантово-механического влияния частиц, обусловленного их неразличимостью в квантовой механике. Ясно, что такое отличие существенно только при высокой плотности газа, поскольку при достаточно большом расстоянии между частицами газа r_s (разреженный классический газ) их взаимное влияние мало и учитывать принцип тождественности частиц нет необходимости. Характерным расстоянием для проявления квантовых эффектов является длина волны де Бройля $\lambda\!\!\!^{-}$, т. е. вырождение наступает при условии, когда межчастичное расстояние становится близко к $\lambda\!\!\!^{-}$. Так как $r_s =$, а $\lambda\!\!\!^{-} = 2\pi\hbar/p = 2\pi\hbar/mv$ (здесь n_d- концентрация d –мерного газа, v - скорость частиц), то условием наступления вырождения является нера-

венство

$$\frac{2\pi\hbar n^{1/d}}{m_n v} > 1.$$

Скорость частиц в общем случае меняется (растет) с температурой и это условие определяет критическую температуру T_0, при которой наступает вырождение, и которая тем выше, чем меньше масса частиц газа и больше его плотность. Величину T_0 можно оценить из условия, что тепловая энергия T_0 при таком переходе порядка "квантовой энергии" $E \sim \hbar^2/2mr_s^2$, соответствующей локализации частицы в области порядка $\lambda\!\!\!^{-} \simeq r_s$, что, с точностью до коэффициента порядка единицы, приводит к значению

$$kT_0 \approx \hbar^2 n^{2/d}/2m. \tag{2.58}$$

Таким образом критерием сильного вырождения газа частиц является соотношение

$$kT << \hbar^2 n^{2/d}/2m. \tag{2.59}$$

Хотя, за исключением небольшой разницы в численных коэффициентах, выражение для критической температуры не зависит от спиновых свойств частиц газа, поведение *ферми-* и *бозе-газа* при вырождении кардинально различно (классические же их свойства одинаковы), поскольку характер влияния квантовой тождественности частиц различен для частиц с целым (бозоны) и полуцелым (фермионы) спином.

При вырождении бозе-газа с отличной от нуля массой (газ атомов, молекул), часть частиц газа ($N_{E=0}/N = 1 - (T/T_0)^{3/2}$) переходит в основное состояние, которому в этом случае соответствует нулевая энергия. Принцип Паули в случае бозонов не препятствует накоплению в одном состоянии любого числа частиц и в пределе $T = 0$ все частицы газа $N = N_{E=0}$ конденсируются в нижайшем энергетическом состоянии (основное состояние). Поскольку в основном состоянии $\varepsilon(p) = 0$ частицы имеют нулевой импульс, они при конденсации не вносят вклада в давление газа и в результате последнее не зависит от объема, как это имеет место в классической картине конденсации насыщенного пара, что и послужило причиной для обозначения этого явления как конденсации Бозе-Эйнштейна. Отметим, что, хотя газ фотонов всегда вырожден (вви-

ду нулевой массы фотонов $T_0 = \infty$ и при любой температуре $T < T_0$), тем не мене конденсации Бозе-Эйнштейна в нем произойти не может, поскольку фотоны не могут иметь нулевого импульса (при $T = 0$ фотонов не существует).

В случае представляющего для нас основной интерес ферми-газа, ввиду действия принципа Паули, запрещающего заполнение одного состояния несколькими частицами, даже при полном вырождении ($T = 0$) частицы газа не могут сконденсироваться в основном состоянии и заполняют состояния до некоторой максимальной энергии E_F, вышележащие же энергетические уровни пусты. Именно для этой пограничной энергии в металлах [14] и был первоначально введен термин - энергия Ферми $E_F - E_c = \mu(T = 0)$, который в дальнейшем стал использоваться в более широком смысле, как обозначение для μ при произвольной температуре. Наиболее важным следствием, вытекающим из принципа Паули, является тот факт, что в сильно вырожденной системе фермионов электроны, лежащие под уровнем Ферми, не могут участвовать в ее отклике на внешние воздействия, связанном с изменением энергии (если такое изменение не превышает энергетического расстояния до уровня Ферми), поскольку все соседние по энергии состояния заняты. Это относится и к тепловому возбуждению частиц Ферми–газа – все состояния, лежащие под уровнем Ферми оказываются "замороженными". Этим объясняется и отсутствие вклада в электронные свойства кристаллов от полностью заполненных низколежащих энергетических зон, соответствующих внутренним электронным орбиталям образующих кристаллическую решетку атомов. Непрерывно изменять свое энергетическое состояние могут только частицы в узкой полосе энергий вблизи уровня Ферми. В этом отношении статистика Ферми–Дирака кардинально отличается как от статистики Бозе–Эйнштейна, так и от статистики Больцмана, справедливой для газа малой плотности, когда вероятность заполнения состояний настолько мала, так что принцип Паули практически не накладывает

[14] Для электронного газа в металлах, благодаря его высокой концентрации, а в полуметаллах и малости электронной массы, температура вырождения (2.58) достигает значений 10^4 K. Последняя намного превосходит температуру плавления, так что практически при всех условиях электроны в металлах могут рассматриваться как сильно вырожденный газ.

ограничений на распределение электронов по энергиям.

Благодаря наличию резкой границы между заполненными и пустыми состояниями, концентрация носителей в d –мерном электронном газе в случае сильного вырождения может быть легко найдена из простых соображений, не используя (2.57) и вообще не прибегая к понятию плотности состояний $D_d(\varepsilon)$. Действительно, поскольку все состояния ниже уровня Ферми заполнены электронами, то концентрация электронов попросту равна числу этих состояний в расчете на единицу объема координатного пространства в объеме d – мерного импульсного пространства, ограниченного изоэнергетической поверхностью $\varepsilon(\mathbf{p}) = \mu = E_F$. В случае изотропного закона дисперсии это есть объем d – мерного шара радиуса p_F, соответствующего предельной энергии $\varepsilon(p_F) = \mu$. Поскольку объем такого шара

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)} p_F^d,$$

то d -мерная концентрация равна

$$n_d = 2\frac{V_d}{(2\pi\hbar)^d} = \frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\left(\frac{p_F}{\hbar}\right)^d = \frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d} k_F^d, \tag{2.60}$$

Зависимость n_d от величины химпотенциала дается той же формулой (2.60), если p_F выразить через μ, используя закон дисперсии $\mu = \varepsilon(p_F)$. Соответствующие выражения для зависимостей n_d от фермиевского волнового вектора $n_d(k_F)$ и энергии Ферми для простейших законов дисперсии $\varepsilon = p^2/2m$ и $\varepsilon = ps$ приведены в табл. 2.3.

Конечно, те же самые зависимости $n_d(\mu)$ могут быть получены, если исходить из выражения (2.57). Для распределения Ферми вероятность заполнения

$$f_0(\varepsilon) \to 0 \quad при \quad \varepsilon - \mu >> k_B T \tag{2.61a}$$

$$f_0(\varepsilon) \to 1 \quad при \quad \varepsilon - \mu << -k_B T. \tag{2.61b}$$

Поскольку в пределе $T \to 0$ это дает $f_0(\varepsilon) = 0$ для $\varepsilon > \mu$, мы можем заменить верхний предел в интеграле (2.57) на μ [15]. С другой стороны,

[15]Очевидно, что при близких к нулю температурах формула (2.57) даст ненулевую концентрацию

Таблица 2.3: Концентрация электронов для сильно вырожденного электронного газа как функция фермиевского волнового вектора и энергии Ферми для изотропных квадратичного и линейного законов дисперсии в случае трех, двух и одного измерений ($g = 2$)

d	$n(k_F)$	$n(\mu, T=0)$	
		$(\varepsilon = \frac{p^2}{2m})$	$(\varepsilon = ps)$
1	$\frac{2}{\pi}k_F$	$\frac{2(2m)^{\frac{1}{2}}}{\pi\hbar}\mu^{1/2}$	$\frac{2}{\pi s\hbar}\mu$
2	$\frac{1}{2\pi}k_F^2$	$\frac{(2m)}{2\pi\hbar^2}\mu$	$\frac{1}{2\pi s^2\hbar^2}\mu^2$
3	$\frac{1}{3\pi^2}k_F^3$	$\frac{(2m)^{3/2}}{3\pi^2\hbar^3}\mu^{3/2}$	$\frac{1}{3\pi^2 s^3\hbar^3}\mu^3$

при $\varepsilon \leq \mu$ функция $f_0(\varepsilon) = 1$ и выражение для концентрации в случае сильного вырождения принимает вид

$$n_d = \int\limits_0^{\mu} D_d(\varepsilon)d\varepsilon. \tag{2.62}$$

Легко убедиться, что это приводит к тем же формулам для $n_d(\mu)$, что приведены в табл. 2.3.

Поскольку интервал энергий, в котором функция распределения изменяется от 1 до 0, составляет несколько k_BT (величина $(\varepsilon - \mu)/k_BT$ в выражении для вероятности заполнения (2.16) входит в показатель экспоненты, так что сильных неравенств в условиях (2.61) не требуется), замена $f_0(\varepsilon - \mu)$ на ступенчатую функцию Хевисайда $\Theta(\mu - \varepsilon)$

$$\Theta(x) = \begin{cases} 0, & \textit{если} \quad x < 0 \\ 1, & \textit{если} \quad x > 0 \end{cases}, \tag{2.63}$$

только в том случае, если уровень Ферми лежит выше дна зоны проводимости, т. е. при положительных μ. В противном случае ($\mu < 0$), поскольку интегрирование ведется по состояниям зоны $\varepsilon > 0$, при $T \to 0$ член в знаменателе функции распределения $\exp[(\varepsilon - \mu)/k_BT] \to \infty$ при всех ε и вероятность заполнения $f_0(\varepsilon - \mu)$ всюду в зоне равна нулю.

использованная выше для вырожденного заполнения, оправдана при $\mu >> k_BT$. Таким образом, критическая температура, соответствующая переходу от классической статистики Больцмана к квантовой статистке $k_BT_0 \sim \mu$. Для квадратичного закона дисперсии $\mu = \hbar^2 k_F^2/2m$, а согласно (2.60) $k_F \sim n_d^{1/d}$, то $k_BT_0 \sim \hbar^2 n_d^{2/d}/2m$, что совпадает с оценкой (2.58).

2.4.2 Обобщенная формула для d -мерной концентрации электронов со степенным законом дисперсии

При произвольной температуре мы должны обратиться к общей формуле (2.57). Во многих важных случаях зависимость плотности состояний от энергии для зонных электронов имеет степенной вид (2.43) или может быть представлена в виде ряда (сходящегося или асимптотического) по степеням энергии. Для плотности состояний такого вида, как сразу видно из (2.57), вычисление концентрации, как функции химпотенциала и температуры, сводится к вычислению интегралов

$$\begin{aligned}\int\limits_0^\infty \frac{\varepsilon^j d\varepsilon}{\exp\left(\frac{\varepsilon-\mu}{k_BT}\right)+1} &= (k_BT)^{j+1}\int\limits_0^\infty \frac{x^j dx}{\exp\left(x-\frac{\mu}{kT}\right)+1} \\ &= (k_BT)^{j+1}\Gamma(j+1)\mathcal{F}_j\left(\frac{\mu}{kT}\right), \end{aligned} \tag{2.64}$$

где введена функция

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right) = \frac{1}{\Gamma(j+1)}\int\limits_0^\infty \frac{x^j dx}{\exp\left(x-\frac{\mu}{k_BT}\right)+1}, \tag{2.65}$$

известная под названием "интеграл Ферми–Дирака индекса j"[16]. В дальнейшем нам понадобится простое, но полезное дифференциальное соот-

[16]Интегралы Ферми–Дирака встречаются во многих других разделах физики твердого тела, в том числе в таком важном ее разделе, как теория кинетических явлений. Часто вместо интегралов (2.65) рассматривают интегралы $F_j = \Gamma(j+1)\mathcal{F}_j$. Преимуществом использования функции (2.65) является то, что она существуют также и для отрицательных целочисленных индексов, а при достаточно больших отрицательных значениях аргумента (классический предел) не зависит от индекса (в этом случае, как мы увидим ниже, $\mathcal{F}_j(x) = \exp x$).

ношение для интегралов Ферми

$$\frac{d}{d\eta}\mathcal{F}_j(\eta) = \mathcal{F}_{j-1}(\eta), \tag{2.66}$$

с очевидностью следующее из определения (2.65) (после дифференцирования нужно воспользоваться соотношением (2.38) $\Gamma(j) = \Gamma(j+1)/j$).

Подстановка выражения (2.44) для плотности состояний в (2.57) с использованием определения (2.64) приводит с к обобщенной формуле для концентрации в d –мерной системе со степенным законом дисперсии

$$n_d = \frac{\Gamma(d/r)}{r2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\frac{1}{\hbar^d}\left(\frac{k_BT}{c_r}\right)^{\frac{d}{r}}\mathcal{F}_{\frac{d}{r}-1}\left(\frac{\mu}{k_BT}\right). \tag{2.67}$$

Введением параметра

$$N_{cd} = \frac{\Gamma(d/r)}{r2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\frac{1}{\hbar^d}\left(\frac{k_BT}{c_r}\right)^{\frac{d}{r}}, \tag{2.68}$$

который принято называть эффективной плотностью состояний на дне зоны (N_{cd} имеет размерность концентрации, а не плотности состояний! – см. Примечание 20), выражение для концентрации принимает особенно простой вид

$$n_d = N_{cd}\mathcal{F}_{\frac{d}{r}-1}\left(\frac{\mu}{k_BT}\right). \tag{2.69}$$

Отсюда сразу следуют выражения для концентрации в случае квадратичного ($r = 2, c_r = 1/2m$)

$$n_d = N_{cd}\mathcal{F}_{\frac{d}{2}-1}\left(\frac{\mu}{k_BT}\right), \tag{2.70a}$$

$$N_{cd} = \frac{1}{2^{d-1}}\left(\frac{2mk_BT}{\pi\hbar^2}\right)^{\frac{d}{2}}, \tag{2.70b}$$

и линейного ($r = 1, c_r = s$)

$$n_d = N_{cd}\mathcal{F}_{d-1}\left(\frac{\mu}{k_BT}\right), \tag{2.71a}$$

$$N_{cd} = \frac{1}{2^{d-2}}\left(\frac{k_BT}{\sqrt{\pi}s\hbar}\right)^d\frac{\Gamma(d)}{\Gamma\left(\frac{d}{2}\right)}. \tag{2.71b}$$

законов дисперсии.

В случае, когда плотность состояний не зависит от энергии ($D(\varepsilon) \propto \varepsilon^0$ при квадратичном спектре в двумерной системе и линейном в одномерной) в выражениях (2.70а) при $d = 2$ и (2.71а) при $d = 1$ фигурирует интеграл Ферми–Дирака нулевого индекса, единственный из этого семейства, выражающийся в элементарных функциях. Поскольку для $j = 0$ гамма–функция $\Gamma(j+1) = 1$, интеграл (2.65)

$$\mathcal{F}_0\left(\frac{\mu}{k_BT}\right) = \int\limits_0^\infty \frac{dx}{\exp\left(x - \frac{\mu}{k_BT}\right) + 1}$$

заменой переменных $y = \exp\left(x - \mu/k_BT\right)$ приводится к виду

$$\mathcal{F}_0\left(\frac{\mu}{k_BT}\right) = \int\limits_{\exp\frac{-\mu}{k_BT}}^\infty \frac{dy}{y(1+y)} = \int\limits_{\exp\frac{-\mu}{k_BT}}^\infty \frac{dy}{y} - \int\limits_{\exp\frac{-\mu}{k_BT}}^\infty \frac{dy}{1+y} = \ln\left(1 + \exp\frac{\mu}{k_BT}\right). \tag{2.72}$$

Подставляя это выражение в (2.70) при $d = 2$, получаем для двумерной (поверхностной) концентрации электронов системы с квадратичным спектром аналитическое представление

$$n_s \equiv n_{2D} = \frac{mk_BT}{2\pi\hbar^2}\ln\left(1 + \exp\frac{\mu}{k_BT}\right), \tag{2.73}$$

справедливое при любых температурах и химпотенциалах. В случае сильного вырождения $\mu >> -k_BT$ (2.73) приводит к уже полученному нами ранее выражению (см. табл. 2.3)

$$n_s = \frac{m}{2\pi\hbar^2}\mu. \tag{2.74}$$

Соответственно, для невырожденного электронного газа $\mu << -k_BT$, разлагая $\ln(1+x)$ в (2.73) в ряд, в первом приближении получаем

$$n_s = \frac{mkT}{2\pi\hbar^2}\exp\frac{\mu}{kT}. \tag{2.75}$$

Совершенно аналогично из (2.71) находится и выражение для одномерной концентрации при линейном дисперсионном соотношении

$$n_{1D} = \frac{2kT}{\pi s\hbar}\ln\left(1 + \exp\frac{\mu}{k_BT}\right) \tag{2.76}$$

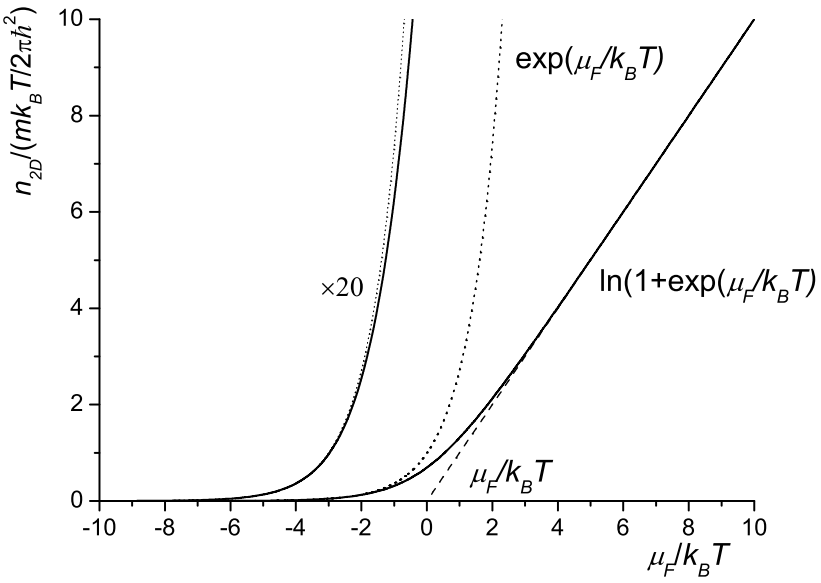

Рис. 2.4: Зависимость концентрации электронов от приведенного химпотенциала μ/k_BT, рассчитанная по формуле (2.73) и по аппроксимирующим ее формулам в пределах сильного вырождения (2.74) и классической статистики (2.75) .

с аналогичными (2.74) и (2.75) аппроксимациями для классического и вырожденного заполнения. Как видно из рис. 2.4, уже при $|\mu| = 2k_BT$ простейшие формулы (2.74) и (2.75) являются очень хорошими аппроксимациями для общей формулы (2.73). Поскольку типичная экспериментальная ошибка определения концентрации, как правило, не превышает нескольких процентов, то использование выражений для вырожденной и классической статистик совершенно оправдано уже при $|\mu| > (3 \div 4)k_BT$. Конечно, в рассматриваемой ситуации ($d = r$) вычисление концентраций по точным формулам (2.73), (2.76), ввиду простоты последних, не представляет каких-либо неудобств. Однако во всех других случаях интегралы Ферми $\mathcal{F}_{j\neq 0}$ не выражаются в элементарных функциях и отыскание достаточно точных и простых аппроксимаций, соответствующих d – мерному вырожденному или классическому электронному газу, актуально не только с методической, но и с практической точки зрения.

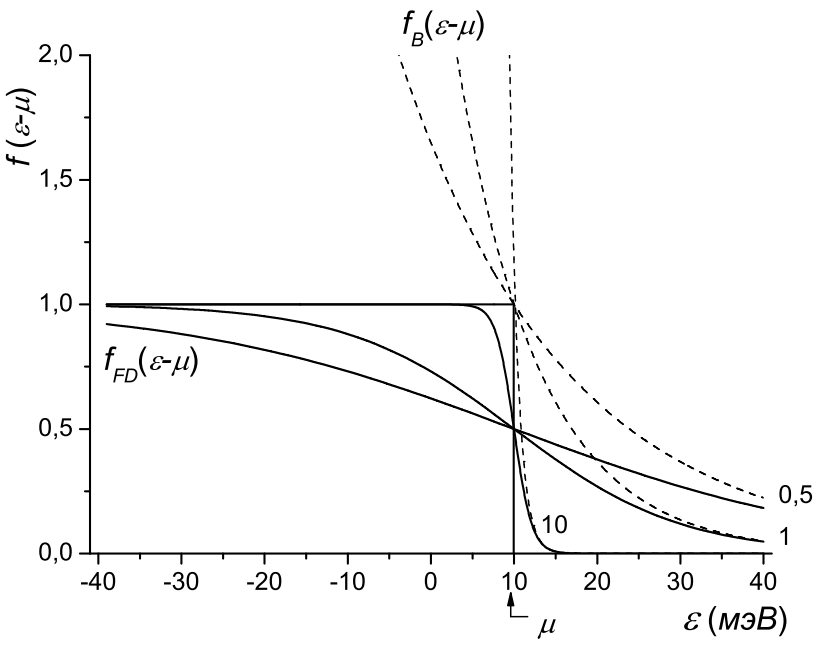

Рис. 2.5: Функции распределения Ферми–Дирака $f_{FD}(\varepsilon-\mu)$ и Больцмана $f_B(\varepsilon-\mu)$ для вырожденного электронного газа ($\mu > 0$) при различных степенях вырождения μ/k_BT (цифры у кривых). Видно, что при $\mu > 0$ распределение Больцмана приводит к недопустимо завышенному заполнению зонных состояний в области энергий $0 \div (\mu + 2k_BT)$ (Сравните со случаем $\mu < 0$ на рис. 2.6).

2.4.3 Аппроксимации для сильно вырожденного и классического электронного газа

Выше мы уже рассматривали случай предельно сильного вырождения $T \to 0$, основываясь на простой модели подсчета числа состояний в d – мерной сфере пространства импульсов (формула 2.60). Представляет интерес проследить, как этот результат следует из общего выражения 2.67.

Для этого можно воспользоваться тем, что входящая в интеграл Ферми (2.65) функция распределения $f_0(x) = \left[\exp(x - \frac{\mu}{k_BT}) + 1\right]^{-1}$ при больших положительных μ/k_BT в окрестности $x = \mu/k_BT$ имеет вид крутой ступеньки (рис. 2.5) и ее производная, следовательно, отлична от нуля только в узкой области аргументов x, т. е. близка к дельта-функции [17]. Чтобы использовать это свойство $\partial f_0/\partial x$ для вырожденного распределе-

[17]Действительно, в пределе $T \to 0$ производная $-\partial f_0/\partial x$ отлична от нуля только вблизи точки

ния, нужно выразить (2.65) через $\partial f_0/\partial x$, для чего нужно проинтегрировать (2.65) по частям

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right)=\frac{1}{\Gamma(j+1)}\left[\frac{1}{j+1}x^{j+1}\ f_0(x)|_0^\infty-\int\limits_0^\infty\frac{1}{j+1}x^{j+1}\frac{\partial f_0}{\partial x}dx\right] \quad (2.77)$$

Поскольку первое слагаемое в (2.77) обращается в ноль при подстановке обоих пределов, а $\partial f_0/\partial x\cong-\delta(x-\mu/k_BT)$, то (2.65) приводится к виду

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right)=\frac{1}{(j+1)\Gamma(j+1)}\int\limits_0^\infty x^{j+1}\delta(x-\mu/k_BT)dx=\frac{1}{\Gamma(j+2)}\left(\frac{\mu}{k_BT}\right)^{j+1}, \quad (2.78)$$

где в последнем равенстве использовано рекуррентное соотношение $x\Gamma(x)=\Gamma(1+x)$.

Подставляя эту аппроксимацию при $j=\frac{d}{r}-1$ в (2.69) получаем для концентрации

$$n_d\left(\mu\right)=N_{Cd}\frac{1}{\Gamma(\frac{d}{r}+1)}\left(\frac{\mu}{k_BT}\right)^{\frac{d}{r}}. \quad (2.79)$$

Поскольку $N_{Cd}\propto(k_BT)^{d/r}$, то концентрация носителей при сильном вырождении, как и должно быть, от температуры не зависит. Подстановка N_{Cd} из (2.68) с учетом $\Gamma\left(\frac{d}{r}+1\right)=\frac{d}{r}\Gamma\left(\frac{d}{r}\right)$ дает окончательно для концентрации

$$n_d\left(\mu\right)=\frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\frac{1}{\hbar^d}\left(\frac{\mu}{c_r}\right)^{\frac{d}{r}}. \quad (2.80)$$

Из сравнения этого соотношения с формулой (2.44) видно, что концентрация при сильном вырождении может быть выражена через плотность состояний на уровне Ферми

$$n_d\left(\mu\right)=\frac{r}{d}D_d(\mu)\mu. \quad (2.81)$$

$x=\mu/k_BT$. С другой стороны, выполняется и условие нормировки для δ – функции

$$\int\limits_0^\infty\left(-\frac{\partial f_0}{\partial x}\right)dx=-\int\limits_0^\infty df_0=f_0(0)-f_0(\infty)=1,$$

так как $f_0(0)=1$ и $f_0(\infty)=0$.

Легко убедиться, что для степенной зависимости энергии от импульса выражения (2.80) и (2.60) по-сути аналогичны. Однако в отличие от подхода, использованного при выводе (2.60), теперь можно получить поправки к формулам для концентраций, учитывающие "тепловое размытие" фермиевской ступеньки - ее конечную ширину для отличной от нуля температуры. В дальнейшем нам понадобятся такие поправки для интегралов более общего вида

$$\int\limits_0^\infty g(\varepsilon) f_0(\varepsilon-\mu) d\varepsilon \approx \int\limits_0^\mu g(\varepsilon) d\varepsilon + \frac{\pi^2 (k_B T)^2}{6} \frac{\partial^2 g(\varepsilon)}{\partial \varepsilon^2}\bigg|_{\varepsilon=\mu} \tag{2.82}$$

и

$$\int\limits_0^\infty g(\varepsilon) \frac{\partial f_0(\varepsilon)}{\partial \varepsilon} d\varepsilon \approx -g(\mu) - \frac{\pi^2 (k_B T)^2}{6} \frac{\partial g(\varepsilon)}{\partial \varepsilon}\bigg|_{\varepsilon=\mu} \tag{2.83}$$

с независящей от температуры и достаточно медленно меняющей по сравнению с

$$\partial f_0/\partial \varepsilon = -(4k_B T)^{-1} \operatorname{ch}^{-2}[(\varepsilon-\mu)/2k_B T] \tag{2.84}$$

функцией $g(\varepsilon)$. Рассмотрим вначале интеграл (2.83), в котором $g(\varepsilon)$ можно разложить в ряд по степеням $(\varepsilon-\mu)$ в окрестности μ, где подынтегральное выражение заметно отлично от нуля[18]

$$g(\varepsilon) = g(\mu) + (\varepsilon-\mu) g'(\mu) + (1/2)(\varepsilon-\mu)^2 g''(\mu) + ...$$

Поскольку функция (2.84) вдали от $\varepsilon=\mu$ близка к нулю, нижний предел в интеграле (2.83) может быть заменен на $-\infty$. В виду четности $\partial f_0/\partial\varepsilon$ нечетные члены разложения дадут ноль при интегрировании. Оставшиеся интегралы равны

$$\int\limits_{-\infty}^\infty \frac{\partial f_0}{\partial \varepsilon} d\varepsilon = (f_0(\infty) - f_0(-\infty)) = -1,$$

[18]При таком разложении мы имеем дело только с состояниями в окрестности уровня Ферми $\varepsilon=\mu$ шириной, соответствующей тепловому размытию фермиевской ступеньки, исключая тем самым замороженные принципом Паули состояния в "глубине" ферми–жидкости. По существу, такая процедура соответствует описанию на языке квазичастичных возбуждений ферми–жидкости с энергиями квазичастиц $\xi = |\varepsilon-\mu| \sim k_B T$ (см. разд. (2.4.5)).

$$\int\limits_{-\infty}^{\infty} \frac{\partial f_0}{\partial \varepsilon}(\varepsilon-\mu)^2 d\varepsilon = -\frac{1}{4k_BT}\int\limits_{-\infty}^{\infty} \frac{x^2dx}{\mathrm{ch}^2(x/2k_BT)} = -\frac{\pi^2(k_BT)^2}{3}$$

и мы приходим к разложению (2.83). Разложение (2.82) получается интегрированием по частям с последующим использованием (2.83).

Применяя формулу (2.82) к интегралу Ферми (2.65) (т. е. для функции $g(\varepsilon) = (\varepsilon/k_BT)^j/\Gamma(j+1)$) получаем аппроксимацию

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right) = \frac{1}{\Gamma(j+2)}\left(\frac{\mu}{k_BT}\right)^{j+1}\left[1+\frac{\pi^2 j(j+1)}{6}\left(\frac{k_BT}{\mu}\right)^2\right], \quad (2.85)$$

позволяющую с точностью до членов второго порядка по k_BT/μ описать температурную зависимость концентрации (при фиксированной величине химпотенциала) или химпотенциала (при фиксированной концентрации). Подставляя это выражение в (2.69), получаем

$$n_d = N_{cd}\frac{1}{\Gamma(\frac{d}{r}+1)}\left(\frac{\mu}{k_BT}\right)^{d/r}\left[1+\frac{\pi^2 d(d-r)}{6r^2}\left(\frac{k_BT}{\mu}\right)^2\right], \quad (2.86)$$

или, вводя концентрацию при нулевой температуре $n_d(\mu_0)$, где μ_0 – химпотенциал при $T=0$,

$$n_d(\mu) = n_d(\mu_0)\left(\frac{\mu}{\mu_0}\right)^{d/r}\left[1+\frac{\pi^2 d(d-r)}{6r^2}\left(\frac{k_BT}{\mu}\right)^2\right]. \quad (2.87)$$

При фиксированном значении химпотенциала $\mu = \mu_0$ это соотношение описывает температурную зависимость концентрации

$$n_d(T) = n_d(\mu_0)\left[1+\frac{\pi^2 d(d-r)}{6r^2}\left(\frac{k_BT}{\mu_0}\right)^2\right]. \quad (2.88)$$

Если же концентрация с температурой не меняется, как это имеет место, например, в объемных металлах и вырожденных сильнолегированных полупроводниках, т. е. $n_d(\mu) = n_d(\mu_0)$, то соотношение (2.87) определяет температурный сдвиг химпотенциала. Поскольку поправка к химпотенциалу предполагается малой, в поправочном члене в квадратных скобках

(2.87) можно положить $\mu = \mu_0$ и разрешая это уравнение относительно μ получим

$$\mu = \mu_0 \left[1 + \frac{\pi^2 d(d-r)}{6r^2} \left(\frac{k_B T}{\mu_0} \right)^2 \right]^{-r/d}, \tag{2.89}$$

или, разлагая далее квадратную скобку в ряд и ограничиваясь первыми двумя членами

$$\mu = \mu_0 \left[1 - \frac{\pi^2 (d-r)}{6r} \left(\frac{k_B T}{\mu_0} \right)^2 \right]. \tag{2.90}$$

Как видно из этих формул, знак температурной поправки к концентрации или химпотенциалу для вырожденного электронного газа зависит от соотношения между размерностью пространства d и показателем степени в дисперсионном соотношении r. Так, для квадратичного закона дисперсии величина химпотенциала при постоянной концентрации в одномерном электронном газе увеличивается с температурой, в отличие от (известного) результата для трехмерной системы, где $\partial\mu/\partial T|_{n_d=const} < 0$. Ясно, что знак эффекта определяется энергетической зависимостью плотности состояний, и может быть предсказан из простых соображений. Размытие фермиевской ступеньки при ненулевой температуре приводит к тому, что часть состояний в энергетическом интервале порядка нескольких $k_B T$ ниже уровня химпотенциала μ_0 опустошается, а часть пустых при $T = 0$ состояний выше μ_0 в таком же интервале частично заполняется электронами. Если плотность состояний растет с энергией, то, ввиду симметричности функции $\partial f_0/\partial\varepsilon$ относительно точки μ_0, число добавленных в систему электронов при неизменном значении химпотенциала будет больше, чем покинувших состояния с $\varepsilon < \mu_0$, что соответствует положительному температурному коэффициенту для концентрации (при $\mu = const$) и отрицательному для химпотенциала (при $n_d = const$). Последнее следует из того, что, поскольку концентрация выросла, чтобы сохранить ее такой же, как и при $T = 0$, нужно уменьшить величину химпотенциала, поскольку для всех имеющих физический смысл размерностей пространства $d > 0$ и дисперсионных показателей $r > 0$ концентрации, конечно, возрастают с μ_0

[19]. Исходя из тех же аргументов, следует, что в случае уменьшающейся с энергией плотности состояний, как это имеет место для одномерной системы со стандартным спектром $r = 2$, знаки температурных коэффициентов $\partial n/\partial T|_{\mu=const}$ и $\partial\mu/\partial T|_{n=const}$ будут противоположны случаю $\partial D_n/\partial\varepsilon > 0$.

Если плотность состояний от энергии не зависит, то согласно нашей качественной картине температурные поправки должны быть равны нулю. Это находится в полном согласии с формулами (2.86–2.90). Таким образом, системы с $d = r$, в которых $\partial D_n/\partial\varepsilon = 0$, занимают особое положение и в этом отношении. Напомним однако, что поправки в (2.86–2.90) вычислены с точностью до членов второго порядка малости по k_BT/μ , а поскольку при $d = r$ они равны нулю, то нужно рассмотреть члены более высокого порядка (четного) по малому параметру k_BT/μ . Как показано выше, при $d = r$ выражения для концентраций допускают простые аналитические представления (2.73) и (2.76), справедливые во всем температурном диапазоне, и для вычисления поправок можно исходить из этих точных формул. Для наших целей их удобно переписать в виде

$$\begin{aligned} n(\mu) &= n(\mu_0)\frac{k_BT}{\mu_0}\ln\left[\exp\frac{\mu}{k_BT}\left(1+\exp-\frac{\mu}{k_BT}\right)\right] \\ &= n(\mu_0)\frac{k_BT}{\mu_0}\left[\frac{\mu}{k_BT}+\ln\left(1+\exp-\frac{\mu}{k_BT}\right)\right], \end{aligned} \tag{2.92}$$

выделяя доминирующий в вырожденном электронном газе член $\exp(\mu/k_BT)$ и малую в этом пределе поправку к нему $\exp(-\mu/k_BT)$. Разлагая логарифм в ряд, получаем

$$n(\mu) = n(\mu_0)\frac{\mu}{\mu_0}\left[1+\frac{k_BT}{\mu}\exp-\frac{\mu}{k_BT}\right], \tag{2.93}$$

откуда следует, что при $d = r$ температурные поправки к n или μ для сильно вырожденного газа экспоненциально малы, а относительные изменения концентраций $(n(\mu) - n(\mu_0))/n(\mu_0)$ не зависят от размерности

[19]Согласно (2.69) $n_d = N_{cd}\mathcal{F}_{\frac{d}{r}-1}\left(\frac{\mu}{k_BT}\right)$, откуда

$$dn_d/d\mu = \frac{N_{cd}}{k_BT}\mathcal{F}_{\frac{d}{r}-2}\left(\frac{\mu}{k_BT}\right), \tag{2.91}$$

где использовано соотношение (2.66). Так как интеграл Ферми – всегда положительная функция, то $dn/d\mu > 0$.

пространства. Знаки поправок $\partial n/\partial T|_{\mu=const} > 0$ и $\partial \mu/\partial T|_{n=const} < 0$, т. е. те же что и в случае $d > r$.

Подчеркнем еще раз, что речь идет о температурных *поправках* к концентрации или химпотенциалу к их значениям при $T \to 0$ только для *сильновырожденного* газа, в котором в пределе $\mu/k_BT \to \infty$ эти величины от температуры не зависят. При этом для трехмерных металлов или сильнолегированных полупроводники с вырожденным при $T = 0$ газом зонных носителей (с проводимостью металлического типа – см. ниже) физический смысл имеет поправка $\partial \mu/\partial T|_{n=const}$, поскольку концентрации носителей в этих материалах определяются только числом узлов решетки или примесей (полностью ионизованных при $T = 0$) в единице объема и поэтому с температурой изменяться не могут (в пренебрежении слабым тепловым расширением кристаллов). В случае низкоразмерных электронных (дырочных) систем в принципе могут быть реализованы оба режима, как $n = const$, так и $\mu = const$ (возможно и смешанное поведение, сопровождающееся одновременным изменением концентрации и химпотенциала).

Проведенный выше анализ показывают, что в случае степенной зависимости энергии от квазиимпульса $\varepsilon \propto p^r$ концентрации зонных носителей в случае сильного вырождения $\mu = E_F - E_c > (2 \div 3)k_BT$ меняются с энергии Ферми по степенному закону $n_d \propto \mu^{\frac{d}{r}}$, причем показатель степени зависит как от закона дисперсии, так и от размерности пространства. Иначе обстоит дело в случае, когда уровень Ферми лежит значительно ниже дна зоны проводимости $\mu = E_F - E_c < -(2 \div 3)k_BT$. Поскольку кинетическая энергия электрона ε, а следовательно и $x = \varepsilon/k_BT$, величина положительная, то при выполнении этого условия единицей в знаменателе (2.65) можно пренебречь. Поскольку этот член ограничивает сверху вероятность заполнения состояний величиной равной единице, т. е. учитывает принцип Паули, такое пренебрежение эквивалентно переходу к статистике Больцмана. Конечно, принцип Паули сохраняет свою силу для фермионного газа любой плотности и при любой температуре, однако, поскольку при $\mu << -k_BT$ вероятность обнаружить электрон в любом энергетическом состоянии зоны проводимости ($\varepsilon > 0$) мала, то,

тем более, малы шансы на то, что два электрона могут занять один и тот же уровень (т. е. у них окажутся одинаковыми импульс и проекция спина). Таким образом, принцип Паули изменит функцию распределения в области положительных (зонных) энергий незначительно (см. рис. 2.6).

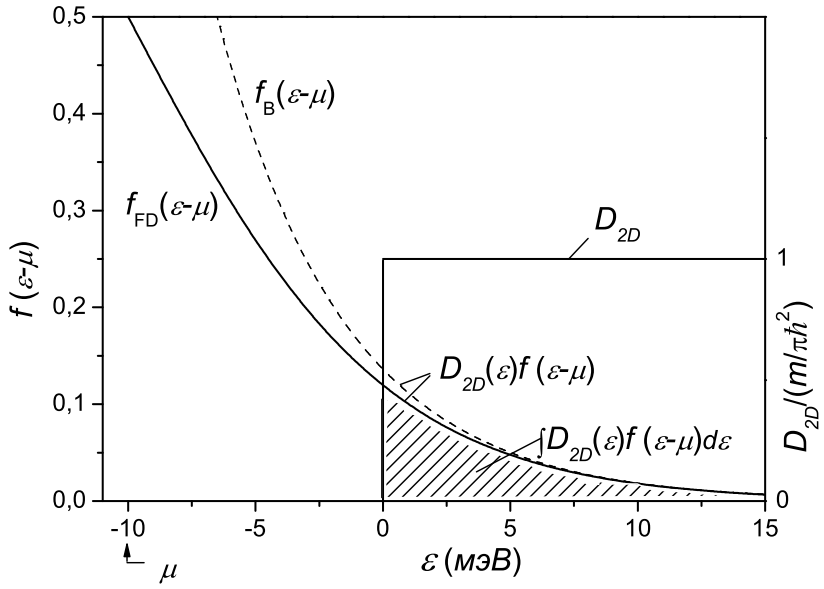

Рис. 2.6: Двумерная плотность состояний $D_{2D}(\varepsilon)$ и функции распределения Ферми–Дирака $f_{FD}(\varepsilon-\mu)$ и Больцмана $f_B(\varepsilon-\mu)$ для невырожденного электронного газа ($\mu = -2k_BT$) в зависимости от энергии. Видно, что фермиевское распределение в актуальной области энергий $\varepsilon > 0$ хорошо аппроксимируется функцией распределения Больцмана. (Сравните со случаем $\mu > 0$ на рис. 2.5).

Для невырожденной статистики выражение (2.65) приобретает, таким образом, вид

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right) = \frac{1}{\Gamma(j+1)}\int\limits_0^\infty x^j e^{\left(\frac{\mu}{k_BT}-x\right)}dx = e^{\frac{\mu}{k_BT}}\frac{1}{\Gamma(j+1)}\int\limits_0^\infty x^j e^{-x}dx,$$

а поскольку фигурирующий здесь интеграл есть просто интегральное представление гамма–функции (2.37) для индекса $j+1$, то интеграл Ферми сводится к одной и той же для всех индексов j экспоненциальной функции

$$\mathcal{F}_j\left(\frac{\mu}{k_BT}\right) = \exp\left(\frac{\mu}{k_BT}\right). \tag{2.94}$$

Концентрации носителей (2.67), следовательно, также экспоненциально зависят от химпотенциала

$$n_d\left(\frac{\mu}{k_BT}\right) = N_{cd}\exp\left(\frac{\mu}{k_BT}\right), \tag{2.95}$$

независимо от размерности пространства или закона дисперсии (если последний допускает разложение в степенной ряд) [20].

2.4.4 Статистика зонных и примесных состояний в трехмерном электронном газе

Поскольку для физики полупроводниковых наноструктур с квантовым пространственным ограничением основной интерес представляют электронные системы в двумерных слоях и одномерных квантовых нитях, можно было бы, на первый взгляд, не делать подробного анализа статистики в трехмерной системе, ограничившись только проведенным выше общим сопоставлением статистики носителе для различных размерностей. Однако низкоразмерные системы реализуются в трехмерных кристаллических структурах с ограничивающими движение носителей заряда потенциальными ямами, формы которых определяется трехмерным распределением зарядов примесей и зонных носителей, т. е. типом и уровнем легирования образующих гетероструктуру слоев. В связи с этим статистика трехмерного электронного газа и статистика заполнения состояний на легирующих примесях оказывается востребованной при описании всех реальных систем в гетероструктурах на основе полупроводников и металлов.

Концентрация электронов в трехмерном кристалле, как это следует из обобщенной формулы (2.67), равна

[20]Величины N_{cd}, конечно, зависят как от зонных параметров, так и от размерности системы (см. (2.70b) и (2.71b)). Формула (2.95) разъясняет смысл названия "эффективная плотность состояний в зоне" для величины N_{cd}. Экспонента в этом выражении дает вероятность заполнения квантового состояния с энергией E_c ($\mu = E_F - E_c$) для статистики Больцмана (при выбранном нами отсчете энергии $E_c = 0$!). Согласно выражению (2.95), концентрация электронов получается такой же, как если бы, вместо непрерывного распределения состояний по зоне, имели дело с совокупностью N_{cd} состояний на единицу объема с одинаковой для всех состояний энергией E_c.

$$n = N_c \mathcal{F}_{\frac{1}{2}}\left(\frac{\mu}{k_B T}\right), \qquad N_c = \frac{2}{\hbar^3}\left(\frac{m_c k_B T}{2\pi}\right)^{\frac{3}{2}} \tag{2.96}$$

для квадратичного закона дисперсии и

$$n = N_c \mathcal{F}_2\left(\frac{\mu}{k_B T}\right), \qquad N_c = \frac{2(k_B T)^3}{\pi^2 \hbar^3 s^3} \tag{2.97}$$

для линейного. Поскольку далее в этом разделе будут рассматриваться как электроны зоны проводимости, так и дырки валентной зоны мы, во избежание путаницы, для эффективных масс носителей в этих зонах будем использовать обозначения m_c и m_v (обозначение m_c эквивалентно m в предыдущих формулах). Формулы (2.96) или (2.97), по существу, определяют значение химпотенциала при данной температуре, если известна концентрация электронов (последняя может быть измерена экспериментально, например из эффекта Холла). Концентрация же электронов n в зоне проводимости определяется процессами их термического возбуждения из состояний валентной зоны, в результате чего в последней возрастает концентрация дырок p, а в случае легированного донорами материала, забросом с донорных уровней, что приводит к их ионизации, т. е. возрастанию концентрации заряженных доноров N_D^+. Если в материале имеются акцепторные уровни, то они также могут ионизоваться в результате захвата электронов из валентной зоны или с донорных уровней, что приводит к изменению концентрации отрицательно заряженных акцепторов N_A^-. В условиях термодинамического равновесия система должна описываться одним и тем же для всех составляющих ее частиц электрохимическим потенциалом, что однозначно определяет его величину, а, следовательно, и концентрации зонных носителей заряда n, p, и заряженных примесей N_D^+, N_A^-.

При известном законе дисперсии, а, следовательно, и плотности состояний валентной зоны, формулы для концентрации дырок могут быть непосредственно получены из соответствующих формул для электронов, для чего необходимо только учесть, что кинетическая энергия дырок $\varepsilon_v(\mathbf{p})$ отсчитывается вниз от потолка валентной зоны, отделенной от дна зоны проводимости энергетическим интервалом $\varepsilon_g = E_c - E_v$ (терми-

ческая щель). Если величину химпотенциала для дырок выразить через электронный химпотенциал, то для концентрации дырок остаются справедливыми все рассмотренные выше выражения для электронов с точностью до замены $\mu \to -\varepsilon_g - \mu$ [21] и параметров законов дисперсии электронов на параметры валентной зоны. Так, для степенных законов дисперсии формулы (2.96) и (2.97) переходят в аналогичные выражения для дырок

$$p = N_v \mathcal{F}_{\frac{1}{2}} \left(\frac{-\varepsilon_g - \mu}{k_B T} \right), \qquad N_v = \frac{2}{\hbar^3} \left(\frac{m_v k_B T}{2\pi} \right)^{\frac{3}{2}} \tag{2.98}$$

и

$$p = N_v \mathcal{F}_2 \left(\frac{-\varepsilon_g - \mu}{k_B T} \right), \qquad N_v = \frac{2(k_B T)^3}{\pi^2 \hbar^3 s_v^3}. \tag{2.99}$$

Для дискретных примесных уровней с энергией E_i и объемной концентрацией N_i плотность состояний (если уровни не размываются в примесную зону) равна бесконечности. Однако, поскольку интеграл по всем возможным состояниям для одного уровня должен быть равным единице, это интегрируемая бесконечность, представляемая δ–функцией (при описании состояний в запрещенной зоне в этом разделе мы используем для энергии обозначение E, чтобы отличать ее от кинетической энергии носителей в зонах ε)

$$D_i(E) = N_i \delta(E - E_i). \tag{2.100}$$

Концентрация электронов на примесных центрах, таким образом, равна

$$n_i = \int D_i(E) f_i(E_i - E_F) dE = N_i \int \delta(E - E_i) f_i(E_i - E_F) dE. \tag{2.101}$$

[21] Определенная выше термическая ширина запрещенной зоны $\varepsilon_g = E_c - E_v$ (напомним, что она может отличаться от оптической щели) есть величина положительная. Можно ввести ту же величину, придерживаясь стандартного для нашего рассмотрения отсчета энергий от дна зона проводимости. Тогда $\varepsilon_g = E_v - E_c$ будет величиной отрицательной. При таком определении во всех формулах для дырок нужно просто изменить знак при ε_g. "Двухвариантность" определения ε_g не должна приводить к затруднениям, поскольку выбор знака легко проверяется из условия, чтобы при $\mu = -|\varepsilon_g|$ химпотенциал для дырок $E_v - E_F$ (аргумент интеграла Ферми для дырок) был равен нулю $(\pm)\varepsilon_g - |\varepsilon_g|$=0. Концентрация дырок, вычисляемая при замене $F(\varepsilon_g - |\varepsilon_g|)$ на $\exp[(\varepsilon_g - |\varepsilon_g|)/k_B T]$, при этом равна эффективной плотности состояний для валентной зоны N_v.

Для функции распределения, описывающей вероятность заполнения примесных центров электронами $f_i(E_i - E_F)$ в выражении (2.101) использовано обозначение с индексом i, поскольку $f_i(E_i - E_F)$ отличается от функции распределения Ферми–Дирака f_0. В отличие от зонных носителей, когда в одном орбитальном состояний могут находиться два носителя с различными проекциями спина, примесные уровни, как правило, не могут захватить двух электронов из-за их сильного кулоновского отталкивания. Присоединение второго электрона, вне зависимости от его спинового состояния, приводит к сильному росту энергии системы и, если даже дополнительный электрон может образовать связанное состояние (многозарядные центры), энергия этого состояния будет значительно отличаться от энергии уровня с одним электроном, т. е. мы имеем дело с совершенно другим примесным уровнем. Тем не менее спин существенно сказывается на статистике распределения электронов по примесным состояниям. Например, в случае одновалентных доноров, захваченный на них электрон (нейтральный донор) может находится в двух различных состояниях, отличающихся направлением спина. Как следствие, нейтральное состояние донора имеет вдвое больший статистический вес по сравнению с ионизованным состоянием, соответствующим положительно заряженному донору. Аналогично, для акцепторных уровней, захватывающих дополнительный электрон (отрицательно заряженный акцептор) для укомплектования парной связи, двояким образом может быть описано отсутствие этого электрона (нейтральный акцептор). Функция распределения, учитывающая спиновое вырождение примесных центров, должна иметь вид

$$f_i = \frac{1}{1 + g_i \exp \frac{E_i - E_F}{k_B T}},$$

где g_i – кратность спинового вырождения ($g_i = 1/2$ для доноров и $g_i = 2$ – для акцепторов). Это выражение может быть записано в другом виде

$$f_i = \frac{1}{1 + \exp \frac{E_i + k_B T \ln g_i - E_F}{k_B T}} = \frac{1}{1 + \exp \frac{E_i' - E_F}{k_B T}},$$

откуда видно, что с точки зрения статистики учет вырождения формально эквивалентен "перенормировке" энергии примесного центра $E_i \to$

$E_i + k_B T \ln g_i$. При не слишком высоких температурах "сдвиг" энергетического уровня $k_B T \ln g_i$, как правило, мал по сравнению с энергией ионизации примеси и может не учитываться. Подставляя выражение для f_i в (2.101), и учитывая нормировку δ–функции, получаем для концентрации электронов на примесном уровне

$$n_i = N_i \frac{1}{\exp\left(\frac{E_i - E_F}{k_B T}\right) + 1}.$$

В дальнейшем нас, как правило, будет интересовать концентрация заряженных доноров $N_D^+ = N_D - N_D^0 = N_D - n_D$ и акцепторов $N_A^- = N_A - N_A^0 = n_A$. Будучи выраженными через электрохимический потенциал для электронов μ, эти величины, как легко видеть, определяются формулами

$$N_D^+ = N_D \frac{1}{\exp\left(\frac{\mu - \varepsilon_D}{k_B T}\right) + 1}, \tag{2.102}$$

$$N_A^- = N_A \frac{1}{\exp\left(\frac{\varepsilon_A - \mu}{k_B T}\right) + 1}, \tag{2.103}$$

где $\varepsilon_D = E_D - E_c$ и $\varepsilon_A = E_A - E_c$ – энергии донорных и акцепторных уровней, отсчитанные от дна зоны проводимости (энергии ионизации доноров и акцепторов в этих обозначениях равны $E_c - E_D = -\varepsilon_D$ и $E_A - E_v = \varepsilon_g + \varepsilon_A$ соответственно).

Одной из важных величин, используемых при описании гетероструктур, равно как и гомопереходов, образованных полупроводниками с различной степенью легирования, является плотность объемного (пространственного) заряда

$$\varrho(\mu) = e[p(\mu) + N_D^+(\mu) - n(\mu) - N_A^-(\mu)], \tag{2.104}$$

общее выражение для которой легко выписать, подставляя сюда приведенные выше обобщенные формулы (или подходящие аппроксимации) для зонных концентраций в трехмерной системе, с учетом законов дисперсии, описывающих соответствующую зону, и выражения (2.102), (2.103) для объемной концентрации заряженных примесей. Отметим, несколько

забегая вперед (см. следующую главу), что в неоднородных структурах, как и в присутствии внешних или встроенных электрических полей, электронный химпотенциал является, в общем случае, функцией координат, так что выражение (2.104) определяет локальную плотность объемного заряда $\varrho(z) = \varrho(\mu(z))$.

В случае однородного легирования и в отсутствии внешних или встроенных электрических полей [22] полупроводник, или некоторая его область, нейтрален и плотность объемного заряда равна нулю [23]. Выражение (2.104) в этом случае (т. е при равной нулю левой части; уравнение $\varrho(\mu) = 0$ называют *уравнением электронейтральности*) при известных зонных параметрах и параметрах легирования, однозначно определяет величину химпотенциала и его температурную зависимость. Как это будет рассмотрено в следующей главе, внешние поля и поля, обусловленные внутренним объемными зарядами, в полупроводниках эффективно экранируются свободными носителями заряда. В гетероструктурах размеры по крайней мере некоторых составляющих ее полупроводниковых слоев значительно превышают характерные длины экранирования, так что в глубине этих слоев полупроводник может считаться нейтральным. Это позволяет определить положение уровня Ферми в таких областях из уравнения электронейтральности (чаще актуально решение обратной задачи - выбрать зонные параметры полупроводника и режимы его легирования, обеспечивающие нужное положение уровня Ферми).

Как правило зонные носители в таких слоях подчиняются невырожденной статистике. Выражения для концентраций зонных носителей в этом случае существенно упрощаются, что позволяет установить некоторые полезные соотношения и ввести новые параметры, удобные для описания распределения электронной плотности в гетероструктурах. Для невырожденной статистики интегралы Ферми переходят в экспоненты,

[22] Вообще говоря, это связанные вещи, поскольку электрическое поле нарушает однородность, а неоднородность в легировании приводит к появлению локальных электрических полей - см. следующую главу.

[23] Это справедливо и в случае, если в полупроводнике отсутствуют примеси, а температура недостаточна для ионизации валентной зоны (полупроводник является, по-существу, диэлектриком).

что дает для зон с параболическим спектром

$$n = N_c \exp\left(\frac{\mu}{k_B T}\right) = \frac{2}{\hbar^3}\left(\frac{m_c k_B T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\frac{\mu}{k_B T}\right) \tag{2.105}$$

$$p = N_v \exp\left(\frac{-\varepsilon_g - \mu}{k_B T}\right) = \frac{2}{\hbar^3}\left(\frac{m_v k_B T}{2\pi}\right)^{\frac{3}{2}} \exp\left(\frac{-\varepsilon_g - \mu}{k_B T}\right) \tag{2.106}$$

Из этих выражений видно, что произведение концентраций и дырок

$$np = N_c N_v \exp\left(-\frac{\varepsilon_g}{k_B T}\right) = \frac{4}{\hbar^6}\left(\frac{k_B T}{2\pi}\right)^3 (m_c m_v)^{\frac{3}{2}} \exp\left(-\frac{\varepsilon_g}{k_B T}\right) = n_i^2 \tag{2.107}$$

не содержит химпотенциала, т. е. не зависит от типа и уровня легирования полупроводника, пока носители обеих зон подчиняются статистике Больцмана. Для этого необходимо, чтобы уровень Ферми находился в запрещенной зоне и отстоял достаточно далеко от краев обеих зон, т. е.

$$-\varepsilon_g + (2 \div 3)kT < \mu < -(2 \div 3)kT > .$$

Введением безразмерного параметра, характеризующего тип и уровень легирования материала

$$\lambda^2 = \lambda^2(\mu) = \frac{p}{n} = \frac{N_v}{N_c} \exp\left(\frac{-\varepsilon_g - 2\mu}{k_B T}\right) = \left(\frac{m_v}{m_c}\right)^{3/2} \exp\left(\frac{-\varepsilon_g - 2\mu}{k_B T}\right) \tag{2.108}$$

выражения для объемных зонных концентрации носителей

$$p = n_i \lambda, \quad n = n_i \lambda^{-1}. \tag{2.109}$$

и плотности объемного заряда $\varrho = |e|\, \varrho_n(\mu)$ (мы ввели концентрационный аналог плотности объемного заряда $\varrho_n(\mu)$)

$$\begin{aligned} \varrho_n(\mu) &= n_i \left[\lambda(\mu) - \lambda^{-1}(\mu) + \frac{N_D^+}{n_i}\left(\exp\frac{\mu - \varepsilon_D}{k_B T} + 1\right)\right] \\ &- n_i \left[\frac{N_A^-}{n_i}\left(\exp\frac{\varepsilon_A - \mu}{k_B T} + 1\right)\right] \end{aligned} \tag{2.110}$$

приобретают особенно простой вид. Если температура настолько высока, что все примеси ионизованы, уравнение (2.110) упрощается еще больше

$$\varrho_n(\mu) = n_i \left[\lambda(\mu) - \lambda^{-1}(\mu) + \frac{N_D}{n_i} - \frac{N_A}{n_i} \right]. \tag{2.111}$$

При $\varrho_n = 0$ уравнения (2.110), (2.111 определяют положение уровня Ферми в нейтральной области полупроводника. Легко видеть, что в случае классической статистики уравнение нейтральности сводится к алгебраическому уравнению относительно $\exp(\mu/k_BT)$ (в общем случае четвертого порядка). Если полупроводник не легирован, то из уравнения нейтральности ((2.110) при $\varrho_n = 0$) следует, что $\lambda = 1$, а концентрации зонных носителей равны друг другу

$$p = n = n_i,$$

как и должно быть для нелегированного, собственного полупроводника. Именно это определяет общепринятое обозначение для величины n_i и ее наименование - *концентрация собственных (intrinsic) носителей.*

На первый взгляд, в случае невырожденной статистики столь же простые соотношения получаются, если обе или одна из зон описываются линейным законом дисперсии. Так для полупроводника с линейным законом дисперсии в зоне проводимости и квадратичным в валентной зоне (электроны и тяжелые дырки в кейновском полупроводнике) соотношения (2.97) и (2.98) для невырожденной статистики ($\mathcal{F}(x) \to \exp(x)$) дают также независящее от легирования произведение

$$np = \frac{4}{\pi^2\hbar^6} \left(\frac{m_v(k_BT)^3}{2\pi s^2} \right)^{\frac{3}{2}} \exp\left(-\frac{\varepsilon_g}{k_BT} \right). \tag{2.112}$$

В действительности это соотношение имеет мало смысла для реальных полупроводников. Так, в кейновском полупроводнике линейный закон дисперсии для электронов имеет место при близкой к нулю их "массе покоя" $m_c \to 0$. Однако при этом близка к нулю и ширина запрещенной зоны $\varepsilon_g \simeq m_c s^2 \to 0$, т. е. зона проводимости смыкаются с валентной. Очевидно, что уровень Ферми в такой системе может проходить либо вблизи точки касания зон (слабое легирование), либо попадает в одну

из зон (в зависимости от типа легирования). В результате, требование о справедливости статистики Больцмана одновременно в обеих зонах, на котором основано соотношение (2.112), ни при каких условиях не может быть выполнено.

2.4.5 Зонная теория и многоэлектронные эффекты

В заключение нашего рассмотрения статистики d – мерного электронного газа, зададимся вопросом: а оправдан ли вообще использованный подход к расчету плотности состояний и концентраций в реальных системах, а если оправдан, то каковы границы его применимости? Оказывается, имеются достаточно серьезные основания для постановки такого вопроса. Дело в том, что, говоря о статистике носителей мы тем самым уже предполагаем, что имеем дело с многочастичной задачей. Между тем, в нашем подходе по-существу использована модель независимых частиц, предполагающая, что движение любой из них не зависит от координат и импульсов всех других частиц системы (это не совсем так в том отношении, что учтен принцип Паули, обязанный в некотором смысле также своеобразному взаимодействию, обусловленному квантово-механической неразличимостью частиц и по разному проявляющемуся для частиц с целым и полуцелым спином). Фактически рассматривалась "статистика" одного электрона в пустой зоне.

Согласно предыдущему рассмотрению статистика частиц в пространстве любой размерности полностью определена, если задан закон дисперсии в энергетической зоне и известен спин частиц, определяющий тип статистики. Однако заряд, являющийся важнейшей, наряду с эффективной массой и спином, характеристикой частицы, нигде явно не учитывался. В неявном виде заряд носителя присутствует, конечно, в дисперсионном соотношении, поскольку именно он играет решающую роль при (электромагнитном) взаимодействии электрона с периодическим потенциалом решетки, которое, в одночастичной картине, целиком определяет закон дисперсии. Однако, помимо взаимодействия с составляющими решетку атомами, *заряженная частица* при своем движении, даже в отсутствии внешних электромагнитных полей, испытывает ку-

лоновское воздействие (в обычных условиях отталкивание) со стороны других частиц. Межэлектронное кулоновское взаимодействие приводит к тому, что движение электронов в кристалле является скоррелированным. В системе фермионов имеется еще один корреляционный эффект – обусловленная принципом Паули обменная корреляция. Она также проявляется в том, что электроны стараются держаться подальше друг от друга, однако в отличие от чисто кулоновского отталкивания обменной корреляции подвержены только электроны в одинаковом спиновом состоянии. Другое отличие состоит в том, что если кулоновское взаимодействие относительно медленно уменьшается с расстоянием между частицами (в отсутствии экранировки $\propto r_s^{-1}$- см. ниже), то обменно-корреляционные эффекты являются короткодействующими.

Ясно, что роль кулоновских и обменных эффектов возрастает с уменьшением среднего межэлектронного расстояния r_s и должна быть особенно существенна в газе носителей высокой плотности (как мы видели, такой газ является вырожденным). В первую очередь это относится к металлам, где число занятых электронами разрешенных состояний в зоне составляет значительную долю от их полного числа. Поскольку же число состояний на единицу объема в каждой энергетической зоне решетки равно числу состояний в зоне Бриллюэна, а последнее совпадает с числом узлов решетки в единице объема, то межэлектронное расстояние в металлах очень близко к периоду решеточного потенциала $r_s \sim a_L$. Справедливость одночастичной, игнорирующей межэлектронные корреляции, зонной теории в таких условиях не имеет, вообще говоря, под собой оснований, по крайне мере требует обоснования.

Уже в силу самой многочастичности задачи очевидно, что последовательный учет межэлектронного взаимодействия относится к одной из наиболее трудных проблем физики конденсированного состояния. К счастью, имеются серьезные аргументы в пользу того, что межэлектронное взаимодействие не меняет классификации уровней – состояние электрона можно по прежнему характеризовать заданием его квазиимпульса. Строго говоря, роль частиц электронного газа принимают на себя коллективные элементарные возбуждения ферми–жидкости, каждое из которых

обладает определенным квазиимпульсом. Квазичастицы такого, близкого по своим свойствам к идеальному, газа также подчиняются статистике Ферми–Дирака, а их число совпадает с числом частиц в жидкости. Таким образом, учет межэлектронного взаимодействия не затрагивает, по существу, главных результатов зонной теории, одночастичной в своей основе. Можно по-прежнему пользоваться всеми понятиями и подходами, разработанными для системы невзаимодействующих электронов в периодическом потенциале решетки, в том числе важнейшей из таких характеристик - законом дисперсии $\varepsilon(\mathbf{p})$, который учитывает не только решеточный потенциал, но и взаимодействие электронов друг с другом. Поскольку дисперсионное соотношение $\varepsilon(\mathbf{p})$ в рамках всего нашего рассмотрения считается заданным (параметры законов дисперсии в любом случае определяются из сопоставления с экспериментом, который, естественно, "зондирует" все, в том числе многочастичные, эффекты), то его перенормировка за счет межэлектронных взаимодействий, никак не отразится на получаемых нами результатах, основанных на методе эффективной массы.

Обратимся теперь к полупроводникам. Даже при максимальных используемых уровнях легирования $N_{D,A} \sim 10^{19}$ см$^{-3}$ и полной ионизации примесей среднее расстояние между носителями в зоне $r_s = 1/N_{D,A} \sim 10^{-6}$ см на несколько порядков превышает постоянную решетки. Вклад межэлектронных эффектов в этих условиях можно, по-видимому, считать малым по сравнению с влиянием периодического потенциала решетки, так что, на первый взгляд, межэлектронное взаимодействие не должно играть столь критической роли, как в металлах. Однако это далеко не всегда так.

Выше мы значительное, если не основное, внимание уделили расчету электронных концентраций в условиях вырождения электронного газа. Однако при внимательном рассмотрении, легко увидеть, что случай сильно вырожденной статистики в однородном объемном (трехмерном) полупроводнике при использованной выше стандартной модели δ – образной плотности примесных состояний практически никогда не реализуется. Это очевидно для собственного полупроводника, а в случае низ-

ких температур, меньших температуры истощения примесей, и для материала, легированного примесями, поскольку носителей в зоне в этом режиме мало, а концентрация нейтральных примесей наоборот велика, для чего уровень Ферми доложен находится в запрещенной зоне. То же самое остается справедливым и при более высоких температурах. Например, в электронном полупроводнике (легированном донорами; очевидно, что то же справедливо и для дырочного полупроводника) для реализации сильно вырожденной статистики уровень Ферми должен находиться по крайней мере на несколько k_BT выше дна зоны проводимости. Однако при этом он неизбежно будет находиться и выше донорных уровней (на еще большее расстояние!), а последние при таком удалении от уровня Ферми также должны быть почти полностью занятыми электронами, т. е. в своем большинстве не могут быть ионизованными, что не согласуется с предположением о высокой концентрации носителей в зоне, поскольку последние появляются там как раз за счет ионизации доноров. Это качественное рассмотрение (можете проверить, что это находится в полном согласии с анализом уравнения нейтральности) показывает, что сильное вырождение газа носителей $\mu >> k_BT$ в полупроводниках с дискретными уровнями примесей ни при каких условиях не может быть достигнуто (при некоторых весьма специальных условиях положительные значения химпотенциала возможны, однако всегда $\mu < k_BT$).

В связи с этим, может создаться впечатление, что проведенный выше анализ статистики электронов в полупроводниках при сильном вырождении трехмерного электронного газа, имеет разве что только академический интерес. Однако это не так. Эксперимент показывает, что при достаточно сильном легировании электронный газ в полупроводниках проявляет типичное для вырожденных электронов поведение (аналогичное электронам в металле) – в сильнолегированных полупроводниках зонные носители существуют вплоть до самых низких температур $T \to 0$, их концентрация равна концентрации доноров и практически не зависит от температуры, что, как мы видели, характерно для сильно вырожденного электронного газа. Все это свидетельствует, что при сильном легировании примеси полностью ионизованы уже при абсолютном

нуле температуры. Термической активации для их ионизации не требуется. Таким образом, предположение о дискретных примесных уровнях в запрещенной зоне, несомненно требующих термической активации для заброса носителей в зону, должно быть пересмотрено. А именно на этом предположении и был основан наш вывод о невозможности сильного вырождения газа носителей в трехмерных полупроводниках.

Тот факт, что для перехода в металлическое состояние требуется сильное легирование, позволяет предположить, что в основе это явления лежат многочастичные эффекты. Как и в случае металлов, теоретическое рассмотрение такой многоэлектронной системы представляет принципиальные трудности, не преодоленные до конца до настоящего времени. Однако основные идеи, на которых основано объяснение перехода электронной (дырочной) системы в металлическое состояние при сильном легировании полупроводников, просты. Здесь наиболее существенны два эффекта.

Во первых с ростом концентрации примесных центров растет вероятность межпримесного туннелирования, в результате чего, благодаря конечному времени жизни электрона на примесной орбите уровни, согласно принципу неопределенности для энергии-времени $\Delta\varepsilon\Delta t \sim \hbar$, уширяются, а с другой стороны электроны уже не могут рассматриваться принадлежащими какой нибудь одной примеси, а коллективизированы, подобно электронам основной решетки зонной теории. В результате система первоначально независимых дискретных примесных уровней, преобразуется в размытую примесную зону, состояний в которой благодаря спиновому вырождению заполнены не полностью (в простейшем случае наполовину), и потому вырождение электронного газа примесной зоны, подобно электронам в металле, сохраняется при самых низких температурах. С ростом концентрации примесей ширина полосы примесных состояний увеличивается и при некотором уровне легирования примесная зона должна слиться с основной энергетической зоной полупроводника. Естественно, что термической активации носителей для заполнения зоны разрешенных состояний исходного (нелегированного) полупроводника теперь не требуется.

Имеется еще один многоэлектронный эффект, который при сильном легировании также приводит к безактивационному заполнению зонных состояний в примесном полупроводнике – экранирование примесного потенциала электронным газом (рассмотрению этого явления посвящена следующая глава). Благодаря экранированию, кулоновский потенциал примеси, определяющий энергию связи (= энергии активации) электрона на нем, на больших расстояниях эффективно экранируется зарядами свободных зонных электронов. Если характерная длина экранирования, уменьшающаяся по мере роста концентрации носителей (которая в свою очередь определяется легированием, так что проблема носит самосогласованный характер), будет меньше исходного (без учета экранирования) радиуса орбиты связанного на примеси электрона (эффективного боровского радиуса), примесь не сможет "удержать" электрон с той же энергией. Благодаря ослаблению притягивающего потенциала примеси при экранировании, энергия связи электрона уменьшается, и, в принципе, при достаточно сильном легировании может вообще обратиться в ноль. Таким образом, экранирование примесного потенциала при сильном легировании, приводит к тому же результату, что и модель образования примесной зоны, а именно к переходу полупроводника, при слабом легировании и низких температурах являющегося диэлектриком, в состояние металлического типа. Совершенно очевидно, что оба указанных эффекта играют существенную роль, причем, будучи взаимозависимыми, они не могут рассматриваться изолированно. Хотя точная картина явления, повторимся, не до конца пóнята, это, опять таки, не меняет наших результатов, коль скоро определены законы дисперсии электронов и дырок. Несомненно, что сильное легирование видоизменяет зонные дисперсионные соотношения, однако это относится, в основном, только к состояниям вблизи краев энергетических зон [24].

[24] в сильнолегированных полупроводниках понятие запрещенной зоны, а вместе с ним и краев разрешенных зон теряет строгий смысл из-за появления обусловленных размытием уровней "хвостов" плотности состояний в запрещенной зоне. Однако, виду экспоненциального спадания плотности этих состояний вглубь энергетической щели, края разрешенных зон остаются выделенными резким увеличением вблизи них плотности состояний. По этой же причине число состояний в области "хвоста" , отнесенное к единице объема, невелико по сравнению с полной концентрацией электронов (примесей), так что искажения законов дисперсии в актуальном для вырожденного газа носителей диапазоне энергий вблизи уровня Ферми, не могут быть существенными (здесь однако возможен

Необходимо подчеркнуть, что из сказанного*совсем не следует, что сильное легирование является необходимым условием для сильного вырождения* электронного газа! Приведенные выше рассуждения относятся исключительно к случаю нейтрального трехмерного полупроводника, когда заполнение зонных состояний возможно только за счет "внутреннего резерва" – ионизации примесей, расположенных в той же локальной области полупроводника. Однако носители могут быть и "позаимствованы" в соседних областях полупроводника или гетероструктуры, используя внешние или встроенные электрические поля. При этом, конечно, локальная нейтральность нарушается – появляется объемный заряд $\varrho(z) = \varrho(\mu(z))$, и, следовательно, изменяется и величина химпотенциала μ, определяемая соотношением (2.104) (или его эквивалентами (2.110), (2.111)). Изменением объемного заряда в нужной области полупроводниковой структуры можно в достаточно широких пределах управлять концентрацией носителей в той или иной зоне (в зависимости от знака ϱ) этой области, а изменяя температуру можно менять степень вырождения (рассматриваемая область может быть вообще нелегированной!). Именно таким образом, в конечном счете, формируются в полупроводниковых гетероструктурах слои с низкоразмерным электронным газом. Создаваемый прикладываемыми извне и (или) встроенными за счет объемных зарядов электрическими полями рельеф потенциала играет здесь двоякую роль - с одной стороны приводит к размерному квантованию спектра, к его двумеризации, с другой, обеспечивает заполнение носителями подзон размерного квантования.

Возможность независимого от легирования и контролируемого управления концентрацией носителей в низкоразмерном электронном газе является одним из важных преимуществ таких систем перед трехмерным электронным газом объемного полупроводника. В трехмерном кристалле изменить равновесную концентрацию носителей можно либо изменяя температуру, либо меняя уровень легирования. Что касается максимальной концентрации носителей, достижимой при температурах меньших

многоэлектронный эффект другого рода - образование "фермиевской" щели в плотности состояний).

перехода к собственному полупроводнику (область истощения примесей), а тем более концентрации носителей в сильнолегированном полупроводнике с вырожденным электронным газом, то они могут быть изменены единственным путем - изменением концентрации легирующих примесей. Это означает, что исследуя трехмерные электронные системы с различной концентрацией носителей мы по необходимости исследуем различные образцы, которые вовсе не обязательно одинаковы во всех других отношениях, кроме концентрации. Выше мы уже отмечали, что одной из важнейших, а с другой стороны одной из наиболее трудных проблем физики, является учет межэлектронных корреляций (своим утверждением, что нам не о чем особенно беспокоиться, коль известен закон дисперсии, мы вовсе не решили проблемы, а попросту ушли от ее решения; вопрос о том, каким образом многочастичные эффекты сказываются на зонном спектре $\varepsilon(\mathbf{p})$ остался в принципе не решенным). Ясно, что решение этой задачи требует исследования систем с различной плотностью электронного газа. При изучении трехмерных систем неконтролируемые вариации других параметров образцов, помимо концентрации носителей, могут в значительной степени исказить достаточно тонкие эффекты, связанные с межэлектронным взаимодействием.

Двумерные и одномерные системы, благодаря возможности менять в широких переделах концентрацию носителей в одной и той же структуре, в этом отношении являются идеальным объектом для исследования многочастичных эффектов. Тем более, что роль последних возрастает с понижением размерности системы. С уменьшением числа степеней свободы электроны при своем движении имеют меньше "маневра" избежать своих соседей, влияния их кулоновского потенциала . В одномерной системе положение становится вообще критическим. Траектории всех электронов в одномерной "нити" совпадают и их взаимные столкновения становятся неизбежными. Применимость результатов, основанных на модели независимых электронов, к реальным многочастичным системам в этом случае вызывает самые серьезные сомнения. Эта проблема не может быть решена в рамках упомянутой выше модели возбуждений ферми–жидкости, позволяющей свести описание многочастичной систе-

мы к опять таки одночастичной, по своей сути, картине газа невзаимодействующих квазичастиц, как это возможно в случае трехмерного и двумерного электронного газа. Даже при $T = 0$ и в отсутствии внешних полей электроны под уровнем Ферми имеют разные скорости $v = d\varepsilon/dp$ (для квадратичного закона дисперсии $v = \sqrt{2\varepsilon/m}$; обратите внимание - *исключением является линейный закон дисперсии!*, при котором скорости всех частиц одинаковы). Поскольку в одномерной системе все они вынуждены двигаются вдоль одной траектории, то быстрые электроны неизбежно будут догонять медленных и должны тормозится благодаря кулоновскому отталкиванию (медленные ускоряться). Однако в одночастичной картине ни ускоряться ни тормозиться они не могут в силу того, что "заморожены" принципом Паули! Мы , таким образом, в принципе не можем рассматривать электроны в одномерной системе как газ независимых частиц. Такая предельно скоррелированная система взаимодействующих электронов должна рассматриваться скорее как конденсат наподобие, если не твердого тела (тогда невозможна проводимость), то текучей несжимаемой жидкости (*жидкость Латтинжера*). Если мы и можем описать электронные свойства такой системы на языке одночастичных возбуждений, т. е. некоторых независимых квазичастиц, также несущих электрический заряд, то их скорости должны быть одинаковыми, иначе пред нами встанут те же проблемы что и в системе исходных частиц. Закон дисперсии неизбежно должен перенормироваться в линейный! Означает ли это, что мы должны пересмотреть полученные ранее в одночастичном приближении результаты для одномерного газа? Оказывается, нет. Одночастичное описание остается справедливым для рассматриваемого нами круга явлений. Что касается перенормировки закона дисперсии, то в одномерной системе, как мы увидим далее, его вид вообще не играет роли, дисперсионное соотношение выпадает из конечных формул, описывающих реакцию системы на внешнее поле или явления переноса.

Возвращаясь к вопросу о преимуществах структур с низкоразмерным газом с точки зрения экспериментального исследования многоэлектронных эффектов, подчеркнем, что в таких системах имеется возможность

независимо управлять как концентрацией, так и степенью вырождения носителей μ/k_BT, меняя температуру. Методы управления величиной ϱ, а следовательно μ и концентрацией носителей в гетероструктурах достаточно разнообразны и зависят от конкретного вида гетероструктуры и предъявляемых к ней требований. В МДП-структурах обогащенная электронами (дырками) приповерхностная область полупроводника формируется притягивающим к поверхности электроны (дырки) электрическим полем. В полупроводниковых композиционных гетероструктурах источниками электронов служат, как правило, примеси, расположенные, однако, в других слоях многослойной структуры. Концентрацией в этих структурах можно также дополнительно управлять и электрическим полем, аналогично тому, как это делается в МДП–структурах. Возможность пространственного разделения проводящего электронного (дырочного) канала от ионизованных примесей (*селективное легирование*), обеспечивающих заполнение носителями состояний в канале, относится к одному из важнейших преимуществ электронных систем в гетероструктурах. При таком разделении фактически устраняется рассеяние носителей канала на кулоновских потенциалах примеси, являющееся одним из наиболее важных факторов (при низких температурах основным), ограничивающим подвижность. Количественное описание пространственного распределения объемного заряда ϱ, и связанного с ним распределения химпотенциала и концентрации носителей в некоторых гетероструктурах будет дано в последующих главах. В частности, в следующей главе будет обрисован общий подход к такому описанию и проанализировано пространственное распределение в частном, но важном случае малых ϱ.

2.5 Задачи

1. Докажите справедливость представленных в табл. 2.1 выражений.

2. Найдите плотность состояний для кейновского закона дисперсии $\varepsilon = \sqrt{p^2s^2 + m^2s^4} - ms^2$ для трехмерного электронного газа.

Ответ:

$$D_{3D} = \frac{1}{\pi^2\hbar^3 s^3}(\varepsilon - ms^2)\sqrt{\varepsilon(\varepsilon + 2ms^2)}$$

3. Покажите, что в предельных случаях малых и больших энергий результат предыдущего задания и формулы (2.30) для двумерного и (2.34) для одномерного электронного газа согласуются с табл. 2.1. Найдите скорость электрона в этих предельных случаях.

4. Исходя из формулы (2.48) найдите выражения для эффективных масс плотности состояний для одномерного и трехмерного электронного газа с кейновским законом дисперсии $\varepsilon = \sqrt{p^2s^2 + m^2s^4} - ms^2$.

   Ответ:

$$m_{1D}(\varepsilon) = \frac{2(\varepsilon + ms^2)^2}{s^2(\varepsilon + 2ms^2)}$$

$$m_{3D}(\varepsilon) = \frac{1}{2s^2}\left[4(\varepsilon + ms^2)^2(\varepsilon + 2ms^2)\right]^{1/3}$$

5. Покажите, что в предельных случаях малых и больших энергий формулы предыдущего задании согласуются с формулами табл. 2.2.

6. Рассмотрите энергетическую зону в одномерном кристалле с (гипотетическим) законом дисперсии

$$\varepsilon(p) = \frac{E_0}{2}(1 - \cos\frac{ap}{\hbar}).$$

   Какой физический смысл параметров a и E_0? Как зависят фазовая и групповая скорость электрона в зоне от квазиимпульса?

   Ответ:

   a -постоянная решетки, E_0 -ширина зоны.

$$v_f(p) = \frac{E_0}{p}\sin^2\frac{pa}{2\hbar},$$

$$v_g(p) = \frac{aE_0}{2\hbar}\sin\frac{pa}{\hbar}$$

7. Дайте выражения для динамической эффективной массы и эффективной массы для закона дисперсии в одномерном кристалле с законом дисперсии
$$\varepsilon(p) = \frac{E_0}{2}(1 - \cos\frac{ap}{\hbar}).$$

Ответ:

$$m_\varepsilon(p) = \frac{2}{E_0 a^2 \sin\frac{pa}{\hbar}},$$
$$m_v(p) = \frac{1}{p} E_0 \sin^2\frac{pa}{2\hbar}$$

8. Найдите концентрацию электронов для кейновского закона дисперсии $\varepsilon = \sqrt{p^2 s^2 + m^2 s^4} - ms^2$ при сильном вырождении электронного газа ($T = 0$) в случае одного и двух измерений. Проверьте, что результаты согласуются с данными табл. 2.3.

Ответ:

$$n_{1D}(\mu) = \frac{2}{\pi s\hbar}\sqrt{\mu(\mu + 2ms^2)} = \frac{2\sqrt{2m\mu}}{\pi\hbar}\sqrt{1 + \frac{\mu}{2ms^2}}$$

$$n_{2D}(\mu) = \frac{m\mu}{\pi\hbar^2} + \frac{\mu^2}{2\pi\hbar^2 s^2} = \frac{\mu}{\pi\hbar^2}\left(m + \frac{\mu}{2s^2}\right)$$

9. Найдите двумерную концентрацию электронов в кейновском полупроводнике при произвольной температуре. Проверьте, что результаты согласуются с данными табл. 2.3.

Ответ:

$$n_{2D}(\mu) = \frac{mkT}{\pi\hbar^2}\left[\ln\left(1 + \exp\frac{\mu}{kT}\right) + kT\Gamma(2)\mathcal{F}_1(\mu)\right]$$

10. Выразите плотность состояний на уровне Ферми через концентрацию электронов для вырожденного газа с квадратичным и линейным законом дисперсии в случае одного, двух и трех измерений.

Ответ: Обобщенное выражение для дисперсионного соотношения $\varepsilon = c_r p^r$:

$$D_d(\mu) = \frac{d}{r}\left[\frac{1}{d2^{d-2}\Gamma(d/2)}\right]^{r/d} \frac{1}{(\sqrt{\pi}\hbar)^r}\frac{1}{c_r} n_d^{\frac{d-r}{d}}.$$

Отсюда для квадратичного ($c_r = 1/2m, r = 2$) и линейного ($c_r = s, r = 1$) зонных спектров:

d	$(\varepsilon = p^2/2m)$	$(\varepsilon = ps)$
$1D$	$\frac{4}{\pi}\frac{m}{\pi\hbar^2}n^{-1}$	$\frac{2}{\pi}\frac{1}{s\hbar}n^0$
$2D$	$\frac{m}{\pi\hbar^2}n^0$	$\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{s\hbar}n^{1/2}$
$3D$	$\left(\frac{3}{\pi}\right)^{1/3}\frac{m}{\pi\hbar^2}n^{1/3}$	$\left(\frac{3}{\pi}\right)^{2/3}\frac{1}{s\hbar}n^{2/3}$

2.6 Вопросы для самопроверки

1. Имеются ли ограничения на величину скорости и импульса частицы:

   (a) в бесконечном и свободном от внешних полей пространстве?

   (b) в свободном от внешних полей, но ограниченном пространстве?

   Одинаков ли ответ для случаев классического и квантового рассмотрения?

2. То же, что и в предыдущем вопросе, но для электрона проводимости в кристалле.

3. Каковы ограничения на величину квазиимпульса электрона проводимости в одномерном кристалле с постоянной решетки a? Имеются ли такие ограничения для импульса свободного электрона?

4. Чем отличаются подходы к подсчету числа квантовых состояний в пространстве импульсов основанные на квантовании в ящике (нулевые граничные условия) и Борна-Кармана (циклические граничные условия)? Какой из них более физичен и почему?

5. Зависит ли двумерная плотность состояний в пространстве квазиимпульсов от закона дисперсии? То же для трехмерной.

6. Можно ли ввести физически содержательное понятие плотности состояний (а следовательно и рассчитать концентрацию частиц) в рамках классической механики? Что меняет переход к квантово-механическому рассмотрению? Входит ли постоянная Планка в выражение для плотности состояний?

7. Каков характер зависимости от постоянной Планка плотности состояний для полупроводниковых наноструктур с двумерным и одномерным электронным газом? Зависит ли этот характер от закона дисперсии в квантовой плоскости или квантовой нити?

8. Возрастает или уменьшается плотность состояний $D(\varepsilon)$ с ростом скорости частицы с изотропным законом дисперсии при заданной энергии. Зависит ли характер этой зависимости от размерности пространства?

9. При каких условиях плотность состояний в d - мерном пространстве не зависит от энергии?

10. Для каких законов дисперсии плотность состояний в энергетическом пространстве постоянна в случае

    (a) одномерного электронного газа?

    (b) двумерного электронного газа?

11. Каким должен был бы быть закон дисперсии трехмерного газа, чтобы его плотность состояний не зависела от энергии? Возможно ли это для свободного электрона?

12. В каком случае плотность состояний стремится к бесконечности при стремящейся к нулю энергии частицы? Какой характер этой расходимости и приводит ли она к расходимости в величине концентрации?

13. В пространстве скольких измерений плотность состояний фотонов не зависит от энергии?

14. Каков характер зависимости плотности состояний нейтрино от размерности пространства?

15. Как ведет себя плотность состояний в двумерном кейновском газе электронов с ростом энергии в области малых и предельно больших энергий?

16. Как изменяются плотности состояний с ростом энергии в одномерном газе электронов и акустических фононов?

17. Что такое эффективная масса плотности состояний и когда оправдано введение этого понятия? Одинаково ли выражение для этой величины для электронов проводимости в трехмерном полупроводнике, в структурах с двумерном электронным газом газом и в квантовых нитях?

18. Что такое динамическая эффективная масса носителя заряда? Отличается ли она от эффективной массы плотности состояний? Зависят ли эти величины от энергии?

19. В каких случаях эффективная масса плотности состояний и динамическая эффективная масса совпадают с эффективной массой, фигурирующей в законе дисперсии. Зависят ли в этих случаях масса от энергии?

20. В системе какой размерности определения для эффективной массы плотности состояний и динамической эффективной масс эквивалентны? Какое ограничение накладывается при этом на закон дисперсии?

21. Чем отличаются невырожденный газ зонных носителей от вырожденного?

22. Можно ли объяснить сильное вырождение электронного газа в полупроводниках без учета многоэлектронных эффектов?

23. Почему при слабом легировании трехмерных полупроводников нельзя достичь сильного вырождения в газе носителей заряда? Может ли быть это достигнуто в полупроводниковых квантовых ямах?

24. Можно ли определить концентрацию электронов при сильном вырождении, не зная закона дисперсии, но имея возможность измерить величину фермиевского квазиимпульса? Можно ли определить энергию Ферми при тех же условиях?

25. Какова зависимость d -мерной концентрации от фермиевского волнового вектора для одномерного, двумерного и трехмерного электронных газов?

26. Какой квантово-механический принцип лежит в основе разделения газа носителей заряда на вырожденный и невырожденный?

27. Какие параметры определяют степень вырождения газа? Для какого газа раньше наступает вырождение с понижением температуры - для газа протонов или электронов при их одинаковой плотности?

28. Чем отличается поведение фермионного и бозонного газов по мере роста степени их вырождения (при понижении температуры или увеличения плотности)?

29. Какова температура вырождения газа фотонов (бозоны) и газа нейтрино (фермионы)?

30. Возможна ли Бозе-конденсация в газе фотонов? Могут ли существовать фотоны при $T = 0$?

31. Зависит ли интеграл Ферми в выражениях для концентрации носителей в случае невырожденного заполнения от

(a) размерности пространства?

(b) закона дисперсии?

32. Чему равна температурная поправка первого порядка по параметру k_BT/μ к энергии Ферми μ при нулевой температуре для сильно вырожденного d -мерного газа электронов, если концентрация с температурой не изменяется?

33. Как зависит температурная поправка к энергии Ферми от температуры (при тех же условиях, что в предыдущем вопросе) в нижайшем ненулевом порядке по k_BT/μ? Зависят ли поправки от размерности электронного газа? В каком случае эта поправка равна нулю?

34. Дайте качественное объяснение тому, что знак температурной поправки к концентрации d -мерного сильно вырожденного электронного газа при постоянной энергии Ферми однозначно определяется тем, растет или уменьшается плотность состояний с энергией. Каков этот знак? Зависит ли он от размерности пространства и закона дисперсии? Приведите примеры систем, для которых концентрация (при постоянной энергии Ферми) растет, не меняется и уменьшается с ростом температуры.

35. Нетрудно показать (см. задание 9 к Гл. 2), что для изотропного электронного газа со степенным законом дисперсии $\varepsilon \propto p^r$ (какие значения показателя степени физически допустимы?) средняя энергия электрона равна $\overline{\varepsilon} = \frac{d}{r}k_BT$, для классического d -мерного газа и $\overline{\varepsilon} = \frac{d}{d+r}\mu$ для вырожденного. Прокомментируйте эти выражения.

36. По какой причине не зависящая от легирования (для невырожденного полупроводника!) "собственная концентрация"$n_i = \sqrt{np}$ не имеет смысла для полупроводников с линейной дисперсией?

37. Выпишите выражение для плотности состояний $D(\varepsilon)$, обобщенное

на случай произвольной размерности пространства и произвольного закона дисперсии.

38. Чему равна плотность состояний для d -мерного электронного газа с произвольным, но изотропным законом дисперсии.

Глава 3

Теплоемкость и экранирование в d –мерном электронном газе

Важнейшим свойством системы носителей заряда является ее ответ на внешние воздействия. Именно исследование такой реакции и дает информацию о свойствах системы и правильности наших представлений о законах, управляющих ее поведением. Практически все представляющие интерес воздействия на систему электронов в кристалле, сводятся в конечном счете, к тепловому и электромагнитному. Последнее относится не только к эффектам в электрическом и магнитном полях, или в поле электромагнитной волны (оптические эффекты), но и к эффектам в поле звуковой волны, при деформации кристаллов, и т. д. Некоторые из этих эффектов, а именно теплоемкость электронного газа и экранирование им статических внешних полей могут быть рассмотрены уже на этом этапе, опираясь практически только на результаты предыдущей главы.

3.1 Теплоемкость d –мерного газа

Далее всюду рассматривается удельная (на единицу объема) теплоемкость при постоянном объеме системы V и постоянной концентрации газа n_d. Для вычисления теплоемкости $c_d = \partial\mathcal{E}/\partial T|_V$ необходимо знать температурную зависимость энергии электронного газа. Начнем с рассмотрения системы со степенным законом дисперсии. Вычисление энер-

гии (или любых ее степеней) системы частиц в этом случае

$$\mathcal{E} = \int\limits_0^\infty \varepsilon D_d(\varepsilon) f_0(\varepsilon - \mu) d\varepsilon, \tag{3.1}$$

как видно из сопоставления (3.1) с (2.57) и (2.64), совершенно аналогично расчету концентрации, за исключением того, что в выражении для $\mathcal{E}$ индекс интеграла Ферми возрастает на единицу. Если использовать введенную ранее плотность состояний на дне зоны N_{cd}, выражение для $\mathcal{E}$ записывается в виде (см. задание 9 к Гл. 2)

$$\mathcal{E} = N_{cd} k_B T \frac{\Gamma(d/r+1)}{\Gamma(d/r)} \mathcal{F}_{\frac{d}{r}+1}. \tag{3.2}$$

3.1.1 Классический газ

Для классического газа, согласно (2.94), $F_j(\mu/k_BT) = \exp(\mu/k_BT)$, и (3.2) сводится к простой линейной зависимости от температуры и концентрации

$$\mathcal{E} = N_{cd} k_B T \frac{\Gamma(d/r+1)}{\Gamma(d/r)} \exp(\mu_0/k_BT) = N_{cd} k_B T \frac{d}{r} \exp(\mu_0/k_BT) = \frac{d}{r} n_d k_B T, \tag{3.3}$$

где в последнем равенстве использовано выражение (2.95) для предполагающейся независящей от температуры концентрации невырожденного газа $n_d\left(\mu_0/k_BT\right) = N_{cd} \exp\left(\mu_0/k_BT\right)$. В отличие от рассматриваемого ниже случая сильного вырождения, величину химпотенциала в выражении (3.3) можно считать не зависящей от температуры и равной ее значению при $T = 0$. Действительно, в формуле (2.95) степенной температурной зависимостью $N_{cd} \propto T^{d/r}$ (см. (2.68)) можно пренебречь по сравнению с экспоненциальной, откуда при $n_d(\mu) = n_d(\mu_0)$ следует $\mu \cong \mu_0$. Таким образом, для теплоемкости получаем формулу

$$c_d = \frac{d}{r} n_d k_B, \tag{3.4}$$

являющуюся очевидным обобщением на случай произвольных d и r хорошо известного выражения для теплоемкости классического трехмерного идеального газа $c_{3D} = 3n_d k_B/2$.

3.1.2 Вырожденный газ

Для вырожденного газа в пределе $T \to 0$ энергия, как и концентрация, в первом приближении от температуры не зависит и для вычисления зависимости $\mathcal{E}(T)$ нужно воспользоваться разложением (2.78) для функции распределения

$$\mathcal{F}_{\frac{d}{r}+1}\left(\frac{\mu}{k_BT}\right) = \frac{1}{\Gamma(d/r+2)}\left(\frac{\mu}{k_BT}\right)^{\frac{d}{r}+1}\left[1+\frac{\pi^2 d(d+r)}{6r^2}\left(\frac{k_BT}{\mu}\right)^2\right]. \tag{3.5}$$

Поскольку мы вычисляем теплоемкость при постоянной концентрации, то химпотенциал в (3.5) также зависит от температуры. Эта зависимость была уже нами найдена и во втором порядке по μ/k_BT описывается выражением (2.89)

$$\mu = \mu_0\left[1+\frac{\pi^2 d(d-r)}{6r^2}\left(\frac{k_BT}{\mu_0}\right)^2\right]^{-r/d}.$$

Фигурирующий в (3.5) перед квадратной скобкой множитель $(\mu/k_BT)^{\frac{d}{r}+1}$ принимает вид

$$\begin{aligned}\left(\frac{\mu}{k_BT}\right)^{\frac{d}{r}+1} &= \left(\frac{\mu_0}{k_BT}\right)^{\frac{d}{r}+1}\left[1+\frac{\pi^2 d(d-r)}{6r^2}\left(\frac{k_BT}{\mu_0}\right)^2\right]^{-\frac{d+r}{r}} \\ &\approx \left(\frac{\mu_0}{k_BT}\right)^{\frac{d}{r}+1}\left[1-\frac{\pi^2 (d+r)(d-r)}{6r^2}\left(\frac{k_BT}{\mu_0}\right)^2\right],\end{aligned}$$

где второе равенство получено разложением $[\ \ ]^{-\frac{d+r}{r}}$ по биному Ньютона. Подставляя это выражение в (3.5) (в поправочном члене в квадратных скобках можно положить $\mu = \mu_0$) получим , пренебрегая членом четвертого порядка по μ/k_BT

$$\mathcal{F}_{\frac{d}{r}+1}\left(\frac{\mu}{k_BT}\right) = \frac{1}{\Gamma(d/r+2)}\left(\frac{\mu_0}{k_BT}\right)^{\frac{d}{r}+1}\left[1+\frac{\pi^2 (d+r)}{6r}\left(\frac{k_BT}{\mu_0}\right)^2\right]. \tag{3.6}$$

После подстановки (3.6) в (3.2) приходим к выражению для энергии

$$\mathcal{E} = \left[N_{cd} k_B T \frac{1}{\Gamma(\frac{d}{r}+1)} \left(\frac{\mu_0}{k_B T} \right)^{\frac{d}{r}+1} \right] \frac{\Gamma^2(\frac{d}{r}+1)}{\Gamma(\frac{d}{r})\Gamma(\frac{d}{r}+2)} \mu_0 \left[1 + \frac{\pi^2(d+r)}{6r} \left(\frac{k_B T}{\mu_0} \right)^2 \right],$$

где члены сгруппированы таким образом, чтобы выделить выражение для концентрации при $T = 0$ (первая квадратная скобка – сравните с (2.79); напомним, что n_d полагается не зависящей от температуры). Сводя с помощью рекуррентного соотношения (2.38) гамма-функции к функции $\Gamma(d/r)$, находим искомую зависимость энергии от температуры

$$\mathcal{E} = \frac{d}{d+r} n_d \mu_0 \left[1 + \frac{\pi^2(d+r)}{6r} \left(\frac{k_B T}{\mu_0} \right)^2 \right]. \tag{3.7}$$

В рамках используемого предположения о постоянстве n_d единственным температурно-зависимым членом в этом выражении является второе слагаемое в квадратных скобках, так что для удельной теплоемкости (в расчете на единицу объема) получаем

$$c_d = \partial \mathcal{E} / \partial T|_V = \frac{\pi^2}{3} \frac{d}{r} \frac{n_d}{\mu_0} k_B^2 T. \tag{3.8}$$

Для всех размерностей пространства теплоемкость при постоянной концентрации при абсолютном нуле температуры c_d равна нулю, в соответствии с теоремой Нернста, и линейно возрастает с температурой. При $d = 3$ и $r = 2$ (3.8) приводит к стандартной формуле трехмерной физики металлов и сильнолегированных полупроводников с параболическими зонами

$$c_{3D} = \frac{\pi^2}{2} \frac{n}{\mu_0} k_B^2 T = \left(\frac{\pi}{3} \right)^{2/3} \frac{m}{\hbar^2} n^{1/3} k_B^2 T = \frac{(2m)^{3/2}}{6\hbar^3} \mu_0^{1/2} k_B^2 T. \tag{3.9}$$

Нетрудно убедиться, что, если бы мы пренебрегли температурной зависимостью μ, то получили бы такое же выражение, но с другим "размерно-дисперсионным" множителем: d^2/r^2 вместо d/r. Ясно, что результаты не будут отличаться, если $d = r$, что, как мы знаем, соответствует независящей от энергии плотности состояний. Это является, разумеется, простым следствием установленной нами ранее независимости химпотенциала от температуры при $d = r$ (см. (2.89) и (2.90)).

Как мы знаем, в случае сильного вырождения, в отклике электронной системы на внешнее, в том числе тепловое, воздействие, в силу принципа Паули, могут участвовать только электроны в узкой полосе состояний вблизи уровня Ферми, поскольку электроны лежащие под уровнем Ферми не могут перейти в уже занятые вышележащие состояния. Между тем, в выражение (3.8) для теплоемкости входит полная концентрация электронов, как будто в отклике системы на нагревание участвуют электроны во всем энергетическом интервале от $\varepsilon = 0$ до $\varepsilon = \mu_b$. В действительности это не так. Согласно (2.81) для сильно вырожденного газа отношение n_d/μ для степенного закона дисперсии пропорционально плотности состояний

$$\frac{n_d}{\mu} = \frac{r}{d} D_d(\mu).$$

Используя это соотношение выражение (3.8) может быть записано в виде, вообще не содержащем концентрации носителей

$$c_d = \frac{\pi^2}{3} D_d(\mu) k_B^2 T. \tag{3.10}$$

Единственным параметром, помимо температуры, определяющим величину теплоемкости вырожденного газа является плотность состояний на уровне Ферми.

Это выражение получено в предположении степенного закона дисперсии, однако сейчас мы покажем, что оно справедливо при любом дисперсионном соотношении (лишь бы плотность состояний $D_n(\varepsilon)$ не изменялась слишком быстро в окрестности $\varepsilon = \mu$). Поскольку полная энергия газа $\mathcal{E} = \int\limits_0^\infty \varepsilon D_n(\varepsilon) f_0(\varepsilon) d\varepsilon$ зависит от температуры только через функцию распределения, то теплоемкость

$$c_d = \frac{\partial \mathcal{E}}{\partial T} = \int\limits_0^\infty \varepsilon D_d(\varepsilon) \frac{\partial f_0}{\partial T} d\varepsilon = -\int\limits_0^\infty \varepsilon D_d(\varepsilon) \frac{\partial f_0}{\partial \varepsilon} \left(\frac{\varepsilon - \mu}{k_B T} + \frac{\partial \mu}{k_B \partial T} \right) d\varepsilon,$$

где использовано выражение для $\partial f_0/\partial T$

$$\frac{\partial f_0}{\partial T} = -\frac{\partial f_0}{\partial \varepsilon} \left(\frac{\varepsilon - \mu}{k_B T} + \frac{\partial \mu}{k_B \partial T} \right) \tag{3.11}$$

Применяя к полученному выражению для теплоемкости разложение (2.83), после простого дифференцирования получаем

$$c_d = \mu D_n(\mu)\frac{\partial\mu}{k_B\partial T} + \frac{\pi^2 k_B T}{3}\frac{\partial}{\partial\mu}(\mu D_n(\mu)). \tag{3.12}$$

Поскольку речь идет о теплоемкости при постоянной концентрации, производная $\partial\mu/\partial T$ должна быть найдена из условия

$$\frac{\partial n_d}{\partial T} = \int\limits_0^\infty D_n(\varepsilon)\frac{\partial f_0(\varepsilon)}{\partial T}d\varepsilon = 0.$$

Выражение для этой производной получается, как видно из сопоставления выражений для $\partial n_d/\partial T$ и $\partial\mathcal{E}/\partial T$, простым переписыванием (3.12) с заменой $\mu D_n(\mu)$ на $D_n(\mu)$

$$\frac{\partial n_d}{\partial T} = D_n(\mu)\frac{\partial\mu}{k_B\partial T} + \frac{\pi^2 k_B T}{3}\frac{\partial}{\partial\mu}(D_n(\mu)).$$

Находя отсюда $\partial\mu/\partial T$ и подставляя в (3.12) приходим к формуле (3.10), доказывая тем самым справедливость последней при любом дисперсионном соотношении. Таким образом, теплоемкость любой вырожденной системы частиц линейно возрастает с температурой с коэффициентом пропорциональности, определяемым только плотностью состояний на уровне Ферми.

Теплоемкость характеризуется двумя параметрами - плотностью состояний на уровне Ферми и приращением энергии частиц при нагревании. Однако это приращение $k_B T$ (k_B в расчете на единичный температурный интервал) является одинаковой для всех частиц термодинамической характеристикой и от концентрации не зависит. Таким образом, теплоемкость аддитивна по отношению к вкладу в нее отдельных частиц газа и ее концентрационная зависимость также определяется только плотностью состояний на уровне Ферми. Выражение (3.8) имеет чрезвычайно простой физический смысл. Возбуждение электронов, лежащих глубоко под уровнем Ферми, в том числе тепловое, заморожено принципом Паули и вклад в c_d дают только частицы в полосе энергий порядка $k_B T$ вблизи $\varepsilon = \mu$. В этом малом интервале энергий ($k_B T << \mu$) плотность состояний D_n можно считать постоянной, поэтому число активных

частиц - величина порядка $D_n k_B T$. Каждая из них вносит в теплоемкость вклад порядка k_B, откуда получаем для теплоемкости $D_n k_B^2 T$, что с точностью до численного множителя совпадает с (3.8).

Благодаря аддитивному вкладу отдельных частиц в теплоемкость, в ней особенно отчетливо проявляется паулевское замораживание электронов под уровнем Ферми. Если бы в разогреве электронной системы участвовали все электроны, c_d должна была бы расти линейно с n_d, как это имеет место для теплоемкости невырожденного газа любой размерности (несколько забегая вперед, отметим для рассматриваемой в разд. (3.2.3) длины экранирования такая пропорциональность, вообще говоря, отсутствует). Если же относительная доля активных (не замороженных) электронов с ростом полной концентрации уменьшается, как это неизбежно имеет место при вырождении, зависимости $c_d(n_d)$ должны быть сублинейными. То, что это действительно так, легко видеть в случае степенного закона дисперсии, при котором согласно (2.44) плотность состояний меняется с энергией по закону $D_d(\varepsilon) \propto \varepsilon^{(d-r)/r}$. По тому же закону меняется и плотность состояний на уровне Ферми при изменении химпотенциала μ

$$D_d(\mu) \propto \mu^{(d-r)/r}.$$

Поскольку же, согласно (2.80), μ и n_d связаны соотношением

$$\mu = \left(d 2^{d-2} \Gamma(d/2) \sqrt{\pi}^d\right)^{r/d} \hbar^r c_r n_d^{r/d} \propto n_d^{r/d}, \tag{3.13}$$

то зависимость $D_d(\mu)$ от концентрацией описывается выражением

$$D_d(\mu) \propto n_d^{(d-r)/d}. \tag{3.14}$$

и, следовательно, теплоемкость меняется с концентрацией по закону

$$c_d \propto D_d(\mu) \propto n_d^{(d-r)/d}. \tag{3.15}$$

При любых имеющих физический смысл значениях r и d показатель степени в этом выражении меньше единицы. Если выполняется соотношение $r = d$, то теплоемкость от концентрации вообще не зависит. Это означает, без особого преувеличения, что (если система остается вырожденной при любой концентрации ($T \to 0$)) в системе со 100 электронами

достигается тот же прирост энергии при нагревании, что и при 10^{10} электронов в том же (d – мерном) объеме (цифры потому и не важны, что концентрация в (3.15) не входит!) . Этот парадоксальный с классической точки зрения результат (в классическом газе $c_d \propto n_d$) является как раз совершенно естественным с точки зрения принципа Паули. Поскольку в реальных системах $r \leq 2$, то такое ярко выраженное проявление свойств вырожденной статистики имеет место только в системах пониженной размерности.

Более того, для одномерного газа при $r > 1$ из выражения (3.15) следует что, чем больше концентрация электронов, тем теплоемкость меньше! В частности, для квадратичного закона дисперсии ($r = 2$) из соотношений (2.44) и (2.80) находим, что теплоемкость

$$c_{1D} = \frac{4}{3}\frac{m}{\hbar^2}\frac{1}{n_{1D}}k_B^2 T, \tag{3.16}$$

обратно пропорциональна концентрации. Очевидная причина такого поведения – уменьшение плотности состояний $D(\varepsilon)$ с ростом энергии в одномерном газе (рис. 2.2), а следовательно и уменьшение плотности состояний на уровне Ферми $D(\mu)$ по мере заполнения состояний одномерной зоны при увеличении концентрации. Для одномерного газа с линейным законом дисперсии и плотностью состояний (2.32) теплоемкость c_d от концентрации не зависит и определяется только параметром закона дисперсии s и мировыми константами [1]

$$c_{1D} = \frac{2\pi}{3}\frac{1}{s\hbar}k_B^2 T. \tag{3.17}$$

Подобное универсальное поведение теплоемкости имеет место и для двумерного газа с квадратичным законом дисперсии. Это непосредственно следует из (2.28)

$$c_{2D} = \frac{\pi}{3}\frac{m}{\hbar^2}k_B^2 T. \tag{3.18}$$

С другой стороны, *линейный характер зависимости* теплоемкости *от температуры* является прямым свидетельством (особенно в низкоразмерных системах, где реализуются условия для независящей от энер-

[1] Разумеется, те же результаты могут быть получены и из выражения (3.8), если выразить в нем химпотенциал через концентрацию.

гии плотностьи состояний) того, что теплоемкость определяется исключительно концентрацией способных к возбуждению электронов с энергиями вблизи уровня Ферми $n_{dF} \sim D_n(\mu)k_BT$, поскольку n_{dF} пропорциональна ширине фермиевской ступеньки k_BT. В этом смысле можно провести прямую аналогию между теплоемкостью вырожденного газа (3.10) $c_d = (\pi^2/3)n_{dF}k_B$ и ее классическим значением (3.4) $c_d = (d/r)n_dk_B$, если под концентрацией в первом случае понимать величину n_{dF}.

3.1.3 Электронная и решеточная теплоемкость.

При исследовании термических свойств кристаллов, связанных только с его тепловым возбуждением, таких как теплоемкость или теплопроводность (в отличие от эффектов, обусловленных электрическим зарядом носителей) экспериментально всегда измеряется суммарный эффект от электронной и решеточной подсистем кристалла. В связи с этим возникает вопрос об их относительном вкладе. На первый взгляд, по крайней мере в полупроводниках, где концентрации электронов даже при наивысших уровнях легирования на несколько порядков меньше числа атомов (узлов) основной решетки в единице объема, электронная теплоемкость должна быть незначительной на фоне решеточной. В общем случае это однако не так. Вывод о доминирующем вкладе последней, как мы увидим, справедлив только в рамках классического описания решеточной теплоемкости, которое не применимо при низких температурах.

Для описания нормальных колебаний решетки удобно воспользоваться квазичастичной картиной, в рамках которой элементарным энергетическим возбуждениям решетки сопоставляется бозе–газ фононов, т. е. возбуждений с энергиями $\hbar\omega_L(n+ 1/2)$, превышающими энергию нулевых колебаний решетки - осцилляторов с $n = 0$. Последние образуют неизменный "фон" , который не проявляется в каких-либо эффектах, наподобие состояний фермионов под уровнем Ферми. Состояние фонона, подобно зонным электронам, полностью характеризуется квазиимпульсом p_{ph} и энергией, связанных друг с другом законом дисперсии $\varepsilon(\mathbf{p}_{ph})$. Процедура вычисления теплоемкости решетки, таким образом, принципиально не отличается от использованной выше для расчета электронной

и сводится к нахождению температурной зависимости полной внутренней энергии решетки (за вычетом не зависящей от температуры энергии нулевых колебаний), т. е. к вычислению фононного аналога выражения (3.1)

$$\mathcal{E} = \int \varepsilon D_d(\varepsilon) f_{BE}(\varepsilon - \mu) d\varepsilon. \tag{3.19}$$

В связи с этим обобщенным выражением следует сделать замечание, касающиеся размерности решетки. Вообще говоря, последние достижения в области полупроводниковой нанотехнологии позволяют создавать структуры с пространственным квантованием не только электронного, ни и фононного спектра, в результате чего обе подсистемы можно рассматривать низкоразмерными. Однако такие структуры следует отнести, скорее, к экзотическим. В подавляющем большинстве реальных структур с низкоразмерным электронным газом решеточная система является, по существу, трехмерной, так как потенциальный рельеф, обеспечивающий пространственное ограничение зонных носителей, как и наличие гетерограниц, обладающих высоким совершенством с точки зрения согласования решеток различных гетерослоев, слабо сказывается на решеточном спектре. Однако, поскольку это не никак не усложняет выкладок, мы, как и в случае электронного газа, будем рассматривать решетку произвольной размерности, тем более, что сопоставление результатов для решеток разных размерностей, как и в случае электронов, представляет отдельный физический интерес.

Вернемся к выражению (3.19) для энергии фононной системы. Помимо того очевидного отличия этого выражения от (3.1), что распределение по энергетическим состояниям фононов, являющихся безспиновыми квазичастицами, должно описываться функцией распределения Бозе–Эйнштейна

$$f_{BE} = \frac{1}{\exp \frac{\varepsilon - \mu}{k_B T} - 1},$$

имеется несколько других важных особенностей, отличающих систему фононов (являющихся "квази" -частицами) от газа обычных частиц:

1. Для обычных частиц (атомов, молекул, свободных или зонных элек-

тронов и др.) объем V, число частиц N (для единичного объема концентрация n) и температура T являются независимыми термодинамическими параметрами, описывающими систему. Так, для зонных носителей в полупроводнике при одной и той же температуре концентрация при различных уровнях легирования различна, и ее значение определяет величину химпотенциала. Фононы же, в том числе при взаимодействии друг с другом или с другими частицами, могут возникать и исчезать. Для такой системы с переменным числом частиц переменные V, N (n) и T не являются независимыми. Число фононов в термодинамическом равновесии определяется из минимума свободной энергии $\Phi(T, V, N)$. При заданных V и T это означает $(\partial\Phi/\partial N)|_{T,V} = 0$, а поскольку изменение Φ при внесении или удалении из системы одной частицы есть по определению ее химпотенциал, то в выражении для фононной (как и фотонной) функции распределения следует положить $\mu = 0$ [2]. Это кардинальным образом изменяет поведение функции распределения Бозе-Эйнштейна. Так при ненулевом химпотенциале $\mu << -k_BT$ функция $f_{BE}(\varepsilon)$ была бы мала во всем интервале представляющих интерес положительных энергий , тогда как при $\mu = 0$ она неограниченно возрастает с уменьшением энергии $f_{BE}(\varepsilon \to 0) \to \infty$. С другой стороны, при положительных $\mu >> k_BT$ функция распределения для малых энергий вообще теряет физический смысл $f_{BE}(\varepsilon < \mu) < 0$.

2. В области небольших квазиимпульсов фононы являются либо безмассовыми (акустические фононы), т. е. описываются линейным законом дисперсии $\varepsilon = ps_{ph}$ с независящей от величины квазиимпульса скоростью s_{ph} [3](скорость звука), либо фононы с независя-

[2]Разумеется при некоторых условиях химпотенциал может быть близким к нулю и для системы обычных частиц. Но уже при малом изменении параметров системы равенство $\mu = 0$ нарушается, тогда как для фоноов в равновесии $\mu \equiv 0$.

[3]Для решеток размерностью $d \geq 2$ фононный спектр состоит из нескольких ветвей, соответствующих продольным и поперечным фононам, скорости которых $s_{ph(l)}$, $s_{ph(t)}$ отличаются (спектр анизотропен). При вычислении плотности состояний можно, однако, ввести "эффективную скорость" для плотности состояний фононов, подобно тому, как была введена эффективная масса для плотности состояний электронов в анизотропной зоне в разд. 2.3.2 (см. задание 1 настоящего раздела). Именно это эффективная величина имеется в виду здесь и далее.

щей от квазиимпульса энергией $\varepsilon(p) = \varepsilon_{opt} p^0$, т. е. с нулевой скоростью $\mathbf{s}_{ph(opt)} = \nabla_{\mathbf{p}}\varepsilon$ (с бесконечной массой). Хотя при больших квазиимпульсах законы дисперсии акустических фононов становятся сублинейными, а энергия оптических уменьшается с ростом p, в дальнейшем будем считать, в соответствии с *моделями Дебая* и *Эйнштейна*, что указанные дисперсионные соотношения справедливыми для всей области изменения p. Тогда плотность состояний для акустических фононов определяется выражением (2.42). Однако при этом надо учесть, что в приведенных выше формулах предполагалось двукратное вырождение уровней, связанное с двумя спиновыми поляризациями электронов. У фононов спиновое вырождение отсутствует, однако возможны различные пространственные поляризации, соответствующие продольному и $d-1$ поперечным фононам. Таким образом выражение (2.42) должно быть умножено на фактор $d/2$, что дает для плотности состояний в акустической ветви фононного спектра

$$D_d(\varepsilon) = K_d \varepsilon^{d-1}, \quad \text{где} \quad K_d = \frac{2d}{\Gamma\left(\frac{d}{2}\right)} \left(\frac{1}{2\sqrt{\pi} s_{ph}\hbar}\right)^d. \tag{3.20}$$

Для оптических фононов, возбуждаемых в решетках, содержащих более одного атома на элементарную ячейку, плотность состояний, согласно обобщенной формуле (2.44) $D_d(\varepsilon) \propto \varepsilon^{(d-r)/r}$ (для оптических фонов $r = 0$) при любой размерности пространства равна бесконечности. Это следует и из общего выражения (2.22)

$$D_d(\varepsilon) \propto \int\limits_S \frac{dS_d}{v_g}.$$

- плотность состояний обязательно имеет особенность, если (групповая) скорость частиц $\nabla_{\mathbf{p}}\varepsilon$ обращается в ноль, т. е. если дисперсионная зависимость $\varepsilon(\mathbf{p})$ является "плоской" (при локально плоской зависимости $\varepsilon(\mathbf{p})$ такие особенности называют *сингулярностями ван Хова*). При принятом предположении, что для оптических фонов зависимость $\varepsilon(\mathbf{p})$ является "плоской" по всей зоне (*модель*

Эйнштейна), ситуация по существу аналогична случаю рассмотренных в разд. 2.4.4 дискретных примесных уровней в полупроводниках. Плотность состояний в этом случае носит δ – образный характер

$$D_d(\varepsilon) = n_L m_b d\delta(\varepsilon - \varepsilon_{opt}), \tag{3.21}$$

где n_L – число элементарных ячеек решетки в единице объема и m_b – число атомов в элементарной ячейке в d –мерного кристалла.

3. В отличие от зонных носителей, концентрации которых как правило меньше числа состояний в зоне Бриллюэна (вследствие чего верхний предел в интегралах Ферми мог быть положен равным бесконечности), фононы, число которых не является заданной величиной, могут при высоких температурах занимать состояния вплоть до потолка энергетической зоны. Верхняя же граница разрешенных значений квазиимпульса и энергии фонона (и следовательно верхний предел в интеграле (3.19)) определяется тем, что в решетке не могут возбуждаться колебания с длиной волны меньшей, чем постоянная решетки a_L, откуда следует

$$E_{max} = s_{ph} p_{max} \leq 2\pi s_{ph}\hbar/a_L \sim 2\pi s_{ph}\hbar n_L^{1/d}, \tag{3.22}$$

где $n_L = (1/a_L)^d$. Величину предельной энергии E_{max} можно рассчитать из других соображений, а именно исходя из того, что полное число возможных колебательных мод решетки должно быть равно числу степеней свободы, т. е. произведению $n_L d$ числа узлов на размерность пространства

$$\int\limits_0^{E_{max}} D_d(\varepsilon)d\varepsilon = K_d \int\limits_0^{E_{max}} \varepsilon^{d-1} d\varepsilon = n_L d, \tag{3.23}$$

откуда

$$E_{max} = \left(\frac{n_L d^2}{K_d}\right)^{1/d}. \tag{3.24}$$

Подстановка в это выражение K_d из (3.20) показывает, что по порядку величины результат (3.24) близок к оценке (3.22). Точному

значению E_{max} не следует придавать особого смысла, поскольку, как отмечено выше, линейное дисперсионное соотношение, а следовательно и выражение (3.20) для плотности состояний заведомо не справедливы при энергиях, близких к E_{max}.

С учетом этих замечаний выражение для внутренней энергии газа акустических фононов принимает вид

$$\mathcal{E} = K_d \int\limits_0^{E_{max}} \frac{\varepsilon^d d\varepsilon}{\exp\frac{\varepsilon}{k_B T} - 1} = K_d (k_B T)^{d+1} \int\limits_0^t \frac{x^j dx}{e^x - 1} = K_d (k_B T)^{d+1} \mathfrak{J}_d \left(\frac{\Theta_D}{T}\right), \tag{3.25}$$

где проведена замена переменных в определенном интеграле $x = \varepsilon / k_B T$ и введены функция (*интеграл Дебая*)

$$\mathfrak{J}_j(t) = \int\limits_0^t \frac{x^j dx}{e^x - 1} \tag{3.26}$$

и температурный эквивалент предельной энергии (*температура Дебая* Θ_D)

$$k_B \Theta_D = E_{max} = \left(\frac{d^2 \cdot n_L}{K_d}\right)^{1/d} = 2\sqrt{\pi} \left(\frac{d \cdot \Gamma(d/2)}{2}\right)^{1/d} \frac{s_{ph}\hbar}{a_L} \approx 2\sqrt{\pi} \frac{s_{ph}\hbar}{a_L}. \tag{3.27}$$

Температура Дебая фактически не зависит от размерности пространства (если исходить из соотношения (3.22), то d в выражение для $\Theta_D = 2\pi s_{ph}\hbar / a_L$ в явном виде не входит) и определяется двумя материальными параметрами – скоростью звука и постоянной решетки. Типичные значения Θ_D лежат в интервале $(100 \div 400)$ K.

Совершенно аналогично вычисляется концентрации акустических фононов

$$n_{d(ph)} = K_d (k_B T)^d \mathfrak{J}_{d-1} \left(\frac{\Theta_D}{T}\right). \tag{3.28}$$

Сопоставление этого выражения с формулой для концентрации зонных носителей (2.69) показывает, что функция $\mathfrak{J}_j(\Theta_D/T)$ играет ту же роль для фононов, что интеграл Ферми в статистике фермионов (с аргументом $k_B\Theta_D/k_BT$ вместо μ/k_BT), а множитель $K_d(k_BT)^{d+1}$ эквивалентен эффективной плотности состояний N_c.

Интеграл (3.26) $\mathfrak{J}_d(\Theta_D/T)$ легко вычисляется численно, однако характер его температурной зависимости (зависимости от параметра $t = \Theta_D/T$) может быть понят из рассмотрения предельных случаев высокой $T >> \Theta_D$ и низкой $T << \Theta_D$ температуры, допускающих аналитические представления [4]. В первом случае ($t << 1$) в интервале актуальных значений $x << 1$ экспонента может быть разложена в ряд и в первом, линейном по t приближении получаем

$$\mathfrak{J}_j(t) = \frac{t^j}{j}. \tag{3.29}$$

При больших t (низких температурах) подынтегральная функция экспоненциально убывает с ростом x и верхний предел интеграла можно устремить к бесконечности. Тогда функция $\mathfrak{J}_j(t)$ сводится к сходящемуся несобственному интегралу, т. е. к зависящей только от индекса константе [5]

$$\mathfrak{J}_j(\infty) = \int\limits_0^\infty \frac{x^j dx}{e^x - 1} = \zeta(j+1)\Gamma(j+1),$$

где дзэта–функция Римана $\zeta(j)$ при $j \geq 2$ близка к единице. Использование этих аппроксимаций приводит к выражениям для полной энергии фононов

$$\mathcal{E} = K_d (k_B T)^{d+1} \frac{1}{d} \left(\frac{\Theta_D}{T}\right)^d = d n_L k_B T \quad (T >> \Theta_D) \tag{3.30}$$

[4] Акустические фононы, как и фотоны, будучи безмассовыми, вырождены при любой температуре. Критерий высокой и низкой температуры в этом случае определяется не соотношением между тепловой энергией и энергией Ферми $k_B T/\mu$, как это имеет место для фермионов (для фононов, как мы видели, $\mu = 0$), а, согласно (3.26), соотношением между $k_B T$ и предельно допустимой энергией фонов $k_B \Theta_D$.

[5] Разложением $f_{BE}(x)$ в ряд, комбинированием членов и используя определение гамма–функции (2.37) получаем ($j > -1$)

$$\begin{aligned}\int\limits_0^\infty \frac{x^j dx}{e^x - 1} &= \int\limits_0^\infty x^j \left(\sum_1^\infty e^{-nx}\right) dx = \sum_1^\infty \int\limits_0^\infty x^j e^{-nx} dx \\ &= \sum_1^\infty \frac{1}{n^{j+1}} \int\limits_0^\infty t^j e^{-t} dt = \Gamma(j+1) \sum_1^\infty \frac{1}{n^{j+1}} = \Gamma(j+1)\zeta(j+1),\end{aligned}$$

где $\zeta(j) = \sum_{n=1}^\infty \frac{1}{n^j}$ для первых целочисленных аргументов равна $\zeta(2) = \pi^2/6$, $\zeta(3) \approx \pi^3/25,75$, $\zeta(4) \approx \pi^4/90$. При $j = 1$ этот ряд логарифмически расходится $\zeta(1) = 1 + 1/2 + 1/3 + \ldots + 1/n = C + \ln n$, где $C \approx 0.5772$ – число Бернулли.

(последнее равенство получено подстановкой Θ_D из (3.27)) в высокотемпературном пределе и

$$\mathcal{E} = K_d(k_BT)^{d+1}\zeta(d+1)\Gamma(d+1) = d^2\zeta(d+1)\Gamma(d+1)\left(\frac{T}{\Theta_D}\right)^d n_L k_B T) \tag{3.31}$$

для температур много меньших дебаевских (в последнем равенстве параметр K_d выражен через температуру Дебая Θ_D из того же соотношения (3.27)). Совершенно аналогично получаем, что зависимость числа фонов в единице d – мерного объема от температуры и размерности пространства дается формулами

$$n_{d(ph)} = K_d(k_BT)^d\frac{1}{d-1}\left(\frac{\Theta_d}{T}\right)^{d-1}\frac{d^2}{d-1}n_L\frac{T}{\Theta_D} \qquad (T >> \Theta_D) \tag{3.32}$$

$$n_{d(ph)} = K_d(k_BT)^d\zeta(d)\Gamma(d) = d^2\zeta(d)\Gamma(d)n_L\left(\frac{T}{\Theta_D}\right)^d \quad (T << \Theta_D) \tag{3.33}$$

Обращает на себя внимание, что для одномерной системы концентрация фононов в обоих пределах равна бесконечности ($\zeta(1) = \infty$). Этот результат не связан с используемыми при выводе (3.32) и (3.33) приближениями и носит общий характер, являясь следствием линейности фононного дисперсионного соотношения и статистики Бозе-Эйнштейна. Действительно, поскольку при линейном законе дисперсии плотность состояний в одномерной системе не зависит от энергии $D_{1D}(\varepsilon) = const$, концентрация фононов определяется интегралом от функции распределения Бозе-Эйнштейна

$$n_{1D(ph)} = D_{1D}\int\limits_0^{E_{max}}\frac{d\varepsilon}{\exp\frac{\varepsilon}{k_BT}-1} = D_{1D}\ln\frac{1}{1-e^{-\varepsilon/k_BT}}\Bigg|_0^{E_{max}}, \tag{3.34}$$

который логарифмически расходится на нижнем пределе. В обычной, трехмерной физике такой проблемы не возникает. Для систем с большим числом измерений $d > 1$, характерная для бозонной статистики инфракрасная расходимость в среднем числе частиц в одном квантовом

состоянии при нулевой энергии $f_{BE}(\varepsilon \to 0) \to \infty$ (хорошо известный пример - фотоны) при подсчете полного числа фононов в единице объема устраняется нулевой плотностью состояний $D_{d>1}(\varepsilon \to 0) \to 0$. Легко видеть, что для бозонов интегралы $\int\limits_0^E D_d(\varepsilon) f(\varepsilon)$ расходятся при любой сублинейной зависимости $D_d(\varepsilon)$, а поскольку при степенным законе дисперсии $D_d(\varepsilon) \propto \varepsilon^{\frac{d}{r}-1}$ (см. (2.44)), то условием их сходимости является неравенство $d \geq 2r$. Расходимость интегралов этого вида сохраняется и для бозонов с ненулевым химпотенциалом, если $\mu > 0$ (интеграл расходится при $\varepsilon = \mu > 0$).

Следует отметить, что такая особенность одномерного газа фононов получена в рамках подхода, игнорирующего межчастичные взаимодействия в многочастичной системе, в том числе и на этапе вычисления плотности состояний - рассчитывается, по существу, плотность "пустых ", незаполненных состояний. В одномерных системах, как это уже обсуждалось в разделе 2.4.5 для случая фермионного газа одномерных носителей заряда, такой подход вызывает принципиальные возражения, поскольку движение взаимодействующих частиц в системе с одной степенью свободы является предельно коррелированным. Это в равной мере относится как к движению взаимодействующих по закону Кулона заряженных электронов, так и атомов решетки, без взаимодействия между которыми (и между фононами и электронами) нет и самого кристалла.

Как уже отмечалось, число частиц не является независимым параметром в системах безмассовых бозонов (являющихся системами с переменным числом частиц) и расходимость в $n_{1D(ph)}$ не отражается на физических свойствах одномерной решеточной системы. Как видно из (3.30) и (3.31) важнейшая характеристика решетки - полная энергия фононной системы является хорошо определенной величиной при любой размерности пространства, что позволяет решить рассматриваемый в этом разделе вопрос о теплоемкости решетки. В пренебрежении термическим расширением величина n_L в выражениях (3.30) и (3.31) от температуры не зависит и теплоемкость d – мерной решетки дается их элементарным

дифференцированием

$$c_{d(ph)} = dn_L k_B \qquad (T >> \Theta_D), \tag{3.35}$$

$$c_{d(ph)} = d^2\zeta(d+1)\Gamma(d+1)\left(\frac{T}{\Theta_D}\right)^d n_L k_B \qquad (T << \Theta_D), \tag{3.36}$$

что приводит, в частности, к известным выражениям для теплоемкости трехмерной решетки

$$c_{3D(ph)} = 3n_L k_B \qquad (T >> \Theta_D), \tag{3.37}$$

$$c_{3D(ph)} = \frac{12\pi^4}{5}\left(\frac{T}{\Theta_D}\right)^3 n_L k_B \qquad (T << \Theta_D). \tag{3.38}$$

Высокотемпературная теплоемкость решетки, как и классическая теплоемкость носителей (3.4) $c_d = \frac{d}{r} n_d k_B$, равна теплоемкости идеального газа, т. е. произведению числа степеней свободы $d \cdot n_L$ системы на постоянную Больцмана. Поскольку в полупроводниках концентрация носителей на много порядков меньше числа узлов решетки в единице объема, вклад электронно-дырочной подсистемы в теплоемкость кристалла при высоких температурах действительно пренебрежимо мал, в соответствии с приведенными в начале этого раздела качественными соображениями. Хотя в металлах концентрации электронов n_d близки к n_L, однако электронный газ в них вырожден и классическая формула (3.4) даже при самых высоких температурах неприменима. Используя выражение для теплоемкости вырожденного электронного газа (3.8) для отношения $c_d/c_{d(ph)}$ при высокой температуре получим

$$\frac{c_d}{c_{d(ph)}} = \frac{\pi^2}{3r}\frac{n_d}{n_L}\frac{k_B T}{\mu}. \tag{3.39}$$

Поскольку в металлах отношение $\mu/k_B T > 10^2$, то $c_d << c_{d(ph)}$ даже при $n_d \sim n_L$ и вклад электронов в высокотемпературную теплоемкость, как и в полупроводниках, может не учитываться.

Совершенно иная ситуация может реализоваться при низких температурах. Согласно (3.8) теплоемкость вырожденного электронного газа при любой его размерности линейна по температуре, тогда как решеточная (3.36) меняется с температурой как T^d. Поскольку в системах

размерности $d > 1$ теплоемкость решетки убывает с понижением температуры быстрее электронной, то ниже некоторой температуры в теплоемкости будет доминировать ее электронная составляющая. Как отмечалось, в типичных реальных структурах с низкоразмерным электронным газом фононная система является по существу трехмерной с теплоемкостью (3.38), поэтому неравенство $c_d >> c_{d(ph)}$ может быть, в принципе, достигнуто и для одномерной электронной системы.

Вычисление концентрации и энергии газа оптических фононов с плотностью состояний (3.21), благодаря δ – функции, не представляет никакого труда

$$n_{d(ph)} = d \cdot n_L m_b \int \frac{\delta(\varepsilon - \varepsilon_{opt}) d\varepsilon}{\exp \frac{\varepsilon}{k_B T} - 1} = \frac{d \cdot n_L m_b}{\exp \frac{\varepsilon_{opt}}{k_B T} - 1}, \tag{3.40}$$

$$\mathcal{E} = d \cdot n_L m_b \int \frac{\varepsilon \delta(\varepsilon - \varepsilon_{opt}) d\varepsilon}{\exp \frac{\varepsilon}{k_B T} - 1} = \frac{\varepsilon_{opt} d \cdot n_L m_b}{\exp \frac{\varepsilon_{opt}}{k_B T} - 1}. \tag{3.41}$$

Из последней формулы получаем выражение для решеточной теплоемкости, связанной с оптической ветвью фононного спектра

$$c_{d(Oph)} = \frac{\partial \mathcal{E}}{\partial T} = d \cdot n_L m_b k_B \frac{(\Theta_E/T)^2 \exp(\Theta_E/T)}{(\exp(\Theta_E/T) - 1)^2}, \tag{3.42}$$

где введена *температура Эйнштейна*, определяемая соотношением

$$k_B \Theta_E = \varepsilon_{opt}.$$

Типичные значения Θ_E несколько выше Θ_D и приходятся на интервал температур $\sim (10^2 \div 10^3)$ K, так что уже в области температур $T < 30$ K вклад экспоненциально убывающей с уменьшением T теплоемкости оптических фононов

$$c_{d(Oph)} \approx d \cdot n_L m_b k_B (\Theta_E/T)^2 \exp(-\Theta_E/T) \qquad (T << \Theta_E)$$

в полную теплоемкость решетки совершенно незначителен.

Как следствие, теплоемкость кристалла при низких температурах определяется только зонными носителями (напомним, что предполагается независимость их концентрации от температуры вплоть до $T \to 0$)

и акустическими фононами. В виду аддитивности теплоемкости по отношению к вкладам от разных подсистем, температурная зависимость полной теплоемкости описывается вытекающей из (3.10) и (3.36) формулой

$$\begin{aligned} c = c_{d(e)} + c_{d_L(ph)} &= \frac{\pi^2}{3} D_d(\mu) k_B^2 T + d_L^2 \zeta(d_L+1)\Gamma(d_L+1) \left(\frac{T}{\Theta_D}\right)^{d_L} n_L k_B \\ &= AT + BT^{d_L}, \end{aligned} \tag{3.43}$$

где учтено, что размерности d электронной и d_L решеточной подсистем могут отличаться. Из этого выражения видно, что зависимость $c(T)$, отложенная в координатах $c/T = f(T^{d_L-1})$, представляет собой прямую с отсечкой $c_{d(e)}/T$, определяющей электронную теплоемкость и углом наклона, тангенс которого $c_{d_L(ph)}/T^d$ дает величину решеточной теплоемкости. Таким образом изучение температурной зависимости теплоемкости позволяет разделить вклад в нее решетки и зонных носителей и, как следствие, определить такие параметры, как скорость звука и, что наиболее важно с точки зрения исследования электронных свойств, плотность состояний зонных носителей на уровне Ферми - важнейшую характеристику вырожденного электронного газа, определяющую его свойства.

3.2 Экранирование статических полей

Все полученные до сих пор результаты справедливы для ансамбля любых частиц, будь то свободные электроны, зонные носители заряда в кристалле или газ атомов или молекул, лишь бы частицы могли быть описаны законом дисперсии (и подчинялись статистике Ферми или Бозе–Эйнштейна). Теперь примем во внимание, что электроны и дырки в проводниках являются носителями электрического заряда и, следовательно, могут изменять свое состояние не только за счет теплового возбуждения, но и при воздействии электромагнитного, в простейшем случае, электростатического поля. Однако прежде чем изучать ответ системы зарядов на электрическое поле, нам нужно рассмотреть, как следует видоизменить выражения для концентрации носителей в присутствии поля.

Таблица 3.1: Удельные электронные (для квадратичного и линейного законов дисперсии) и решеточные (для акустической ветви) теплоемкости в случае одного, двух и трех измерений. В скобках указаны законы дисперсии.

	$\frac{c_{d(e)}}{mk_B^2T/\hbar^2}$ $(\varepsilon = p^2/2m)$	$\frac{c_{d(e)}}{k_B^2T/s\hbar}$ $(\varepsilon = ps)$	$\frac{c_{d(ph)}}{n_Lk_B}$ $(\varepsilon = ps_{ph})$	$\frac{\Theta_D}{s_{ph}\hbar/k_Ba_L}$ $(\varepsilon = ps_{ph})$
$1D$	$\frac{4}{3}n^{-1}$	$\frac{2\pi}{3}n^0$	$\frac{\pi^2}{6}\left(\frac{T}{\Theta_D}\right)$	π
$2D$	$\frac{\pi}{3}n^0$	$\frac{\pi\sqrt{2\pi}}{3}n^{1/2}$	$\frac{8\pi^3}{25.75}\left(\frac{T}{\Theta_D}\right)^2$	$2\sqrt{\pi}$
$3D$	$\left(\frac{\pi}{3}\right)^{2/3}n^{1/3}$	$\pi\left(\frac{\pi}{3}\right)^{1/3}n^{2/3}$	$\frac{12\pi^4}{5}\left(\frac{T}{\Theta_D}\right)^3$	$\sqrt[3]{6\pi^2}$

3.2.1 Статистика электронов в присутствии электростатического потенциала

Если на систему зарядов действует электрическое поле, это означает, что имеется градиент электростатического потенциала $V(\mathbf{r})$ ($\mathbf{E} = -\nabla_r V(\mathbf{r})$) и полная энергия носителей $E = E(\mathbf{p}, \mathbf{r}) = \varepsilon(\mathbf{p}) + \varphi(\mathbf{r})$ теперь зависит не только от импульса, но и от координаты ввиду координатной зависимости потенциальной энергии электрона $\varphi(\mathbf{r}) = -eV(\mathbf{r})$ [6]. Поскольку в рамках метода эффективной массы внешнее электрическое поле предполагается малым по сравнению с внутрикристаллическими полями, оно в окрестности каждой локальной точки не изменяет зонной структуры полупроводника (ширины энергетических щелей и зонных законов дисперсии). Таким образом, все энергетические уровни в присутствии потенциала сдвигаются на одну и ту же величину $\varphi(\mathbf{r})$, включая дно зоны E_c. В общем выражении (2.57) для концентрации (ниже мы вводим в явном виде уровень E_c, поскольку он теперь может зависеть от координаты)

$$n = \int\limits_0^\infty D_d(\varepsilon) f_0(\varepsilon - \mu) d\varepsilon = \int\limits_{E_c}^\infty D_d(E - E_c) f_0 \left[(E - E_c) - (E_F - E_c)\right] dE.$$

такой "синхронный" сдвиг уровней

$$\begin{aligned} E &\rightarrow E + \varphi(r) \\ E_c &\rightarrow E_c + \varphi(r) \end{aligned} \tag{3.44}$$

не изменяет, конечно, кинетической энергии $\varepsilon = E - E_c$ в выражениях для плотности состояний и функции распределения, как и пределов интегрирования, ведущегося по прежнему по всем состояниям энергетической зоны, дно которой теперь имеет энергию $E_c + \varphi(\mathbf{r})$).

Единственное изменение в выражении для концентрации (а, следовательно, и для всех вытекающих из него выражений раздела 2.4)

$$n = \int\limits_{E_c+\varphi}^\infty D_d(\varepsilon) f_0[\varepsilon - (E_F - (E_c + \varphi))] d\varepsilon.$$

[6] Если потенциал не зависит от координаты $V(\mathbf{r}) = const$, он всегда может быть выбран равным нулю, поскольку как в классической, так и в квантовой физике V определен с точностью до произвольной постоянной, меняющей только систему отсчета энергии.

состоит в том, что вместо химпотенциала $\mu = E_F - E_c$ стоит величина $E_F - (E_c + \varphi(\mathbf{r}))$. Однако последняя есть не что иное, как энергетическое расстояние между дном зоны $E_c(\mathbf{r})$ и уровнем Ферми E_F, т. е. локальное значение зависящего теперь от координаты химпотенциала $\mu(\mathbf{r}) = E_F - E_c(\mathbf{r})$. Таким образом все полученные выше выражения для концентраций остаются в силе для ее локальных значений $n_d(\mu(\mathbf{r}), T)$. Термодинамический параметр E_F, характеризующий многочастичную систему носителей в состоянии термодинамического равновесия, имеет теперь смысл не химического, а электрохимического потенциала. Поскольку $E_F - E_c - \varphi(\mathbf{r}) = \mu(\mathbf{r})$, а электрохимический потенциал при равновесии одинаков для всей системы, то сумма химического и электрического потенциалов [7]

$$\varphi(\mathbf{r}) + \mu(\mathbf{r}) = E_F - E_{c0} \tag{3.45}$$

(потенциал φ определен с точностью до произвольной постоянной E_{c0}, указывающей начало отсчета энергии) не зависит от координаты. Возрастание одной из этих величин с координатой сопровождается таким же уменьшением другой и, таким образом, задачи определения пространственного распределения $\varphi(\mathbf{r})$ и $\mu(\mathbf{r})$ эквивалентны.

3.2.2 Уравнение Пуассона

Рассмотрим такое распределение в частом, но важном случае кристалла, в электрическом поле. Поместим трехмерный кристалл между обкладками плоского конденсатора, одна из которых образует омический контакт с образцом, а вторая отделена от поверхности кристалла диэлектрическим зазором (рис. 3.1). Фактически мы рассматриваем задачу электростатики, базовым уравнением которой является закон Гаусса (в его дифференциальной форме)

$$\mathrm{div}(\epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})) = 4\pi\varrho(\mathbf{r}), \tag{3.46}$$

где ϵ –диэлектрическая проницаемость, ϱ –*трехмерная* плотность объемного заряда. Если кристалл всюду нейтрален и в нем нет подвиж-

[7] Как это почти общепринято, здесь и далее для краткости термин "потенциал"используется как для собственно потенциала V, так и для потенциальной энергии электрона $\varphi = -eV$.

ных носителей заряда (диэлектрик, невырожденный полупроводник при $T \to 0$), то внешнее поле не может нарушить локальной нейтральности материала и привести к образованию пространственного заряда. Следовательно, согласно (3.46), поле внутри полупроводника также однородно, а его величина определяется непрерывностью $\epsilon\mathbf{E}$ на границе полупроводника с изолятором. Если же в кристалле имеются зонные (свободные) носители заряда, их перемещение в поле приведет к образованию объемного заряда у поверхности полупроводника, который будет экранировать внешнее поле, последнее будет затухать по мере удаления от поверхности. Величина $\varrho(\mathbf{r})$ определяется локальными концентрациями свободных носителей и заряженных примесей. Координатная же зависимость последних описывается координатной зависимостью химпотенциала $\mu(\mathbf{r})$, который в условиях термодинамического равновесия однозначно связан с электростатическим потенциалом соотношением (3.45). Поскольку пространственное распределение электрического поля в (3.46) также описывается электростатическим потенциалом $\mathbf{E} = -\nabla V = \nabla\varphi/e$, то мы имеем дело с задачей самосогласованного определения $\varphi(\varrho(\varphi))$ и $\varrho(\varphi(\varrho))$. Таким образом, нам нужно преобразовать, уравнение Гаусса в уравнение для потенциала φ, т. е. перейти, используя $\mathbf{E} = \nabla\varphi/e$, к *уравнению Пуассона*

$$\mathrm{div}(\epsilon(\mathbf{r})\nabla\varphi(\mathbf{r})) = 4\pi\varrho(\mathbf{r}). \tag{3.47}$$

Полагая диэлектрическую проницаемость полупроводника независящей от координаты (в гетероструктурах ϵ, конечно, может быть различна для разных ее областей) приходим к уравнению

$$\sum_i \frac{d^2\varphi}{dx_i^2} = \triangle\varphi = \frac{4\pi e\varrho(\varphi(\mathbf{r}), \mathbf{r})}{\epsilon}. \tag{3.48}$$

Следует обратить внимание, что в общем случае объемная плотность заряда зависит от координаты не только через $\varphi(\mathbf{r})$, но и может содержать явную координатную зависимость, если легирование образца неоднородно ($\varphi(\mathbf{r})$ определяет только вероятность ионизации примеси, полная же концентрация заряженных центров, конечно, пропорциональна полной концентрации примесей $N_D(\mathbf{r})$, $N_A(\mathbf{r})$). Однако в дальнейшем мы будем,

как правило, предполагать, что концентрации легирующих примесей в пределах каждой из компонент гетероструктуры постоянны. В этом приближении для рассматриваемого случая направленного по оси z однородного (в плоскости x, y) внешнего электрического поля, уравнение Пуассона сводится к одномерному уравнению [8]

$$\frac{d^2\varphi}{dz^2} = \frac{4\pi e^2 \varrho_n(\varphi)}{\epsilon} \tag{3.49}$$

с независящей в явном виде от координаты правой частью, в которой вместо объемной плотности заряда использован ее концентрационный эквивалент $\varrho_n = \varrho/|e|$. Поскольку, согласно (3.45) $\nabla\varphi = -\nabla\mu$, то уравнение (3.49) может быть записано в полностью эквивалентной форме

$$\frac{d^2\mu}{dz^2} = -\frac{4\pi e^2 \varrho_n(\mu)}{\epsilon}, \tag{3.50}$$

где

$$\varrho_n(\mu) = p(\mu) + N_D^+(\mu) - n(\mu) - N_A^-(\mu), \tag{3.51}$$

Последняя модификация уравнения Пуассона обладает тем удобством, что локальные концентрации зонных носителей и заряженных примесей в (3.51) непосредственно определяется приведенными в предыдущем разделе формулами, в которых химический потенциал, будучи разностью энергий $\mu(z) = E_F - E_c(z)$, не зависит, в отличие от $\varphi(z)$, от начала отсчета энергии.

К одномерному уравнению Пуассона сводятся многие задачи физики полупроводниковых структур, с которыми мы в дальнейшем встретимся неоднократно. Ниже мы рассмотрим решение уравнения Пуассона общего вида (для сокращения записи введем обозначение $f(\mu)$ для правой части (3.50))

$$\frac{d^2\mu}{dz^2} = f(\mu) \tag{3.52}$$

с произвольной функцией $f(\mu)$, предполагая однако, что $f(\mu)$ не содержит явной зависимости от координаты. Умножим левую и правую части

[8] В действительности мы, конечно, имеем дело с 1+2 – мерной задачей: одномерной вдоль оси z и двумерной в плоскости x, y. Однако, если система однородна в этой плоскости для любого сечения $z = const$, то переменные в уравнении Пуассона разделяются, а свойства системы от координат x, y не зависят. Наличие боковых границ в реальных образцах (конечных размеров), нарушает такую однородность, однако при достаточном удалении от границ их влияние будет незначительным.

(3.52) на $d\mu/dz$ и используем тождество

$$\frac{d\mu}{dz}\frac{d^2\mu}{dz^2} = \frac{d\mu}{dz}\frac{d}{dz}\left(\frac{d\mu}{dz}\right),$$

что, после умножения обеих сторон уравнения на dz, приводит к уравнению

$$\frac{d\mu}{dz}d\left(\frac{d\mu}{dz}\right) = f(\mu)d\mu.$$

Граничные условия, заданные для некоторой токи $z = z_0$ (правильнее говорить о сечении $z = z_0$, однако с учетом примечания 8 это не может вызвать недоразумений), автоматически учитываются, если при интегрировании этого уравнения взять определенный интеграл в пределах от $\mu(z_0)$ до μ

$$\int\limits_{\frac{d\mu}{dz}\big|_{z_0}}^{\frac{d\mu}{dz}} \frac{d\mu}{dz}d\left(\frac{d\mu}{dz}\right) = \int\limits_{\mu(z_0)}^{\mu} f(\mu)d\mu,$$

откуда находим первый интеграл уравнения Пуассона

$$\frac{d\mu}{dz} = \pm\sqrt{\left(\frac{d\mu}{dz}\bigg|_{z=z_0}\right)^2 + 2\int\limits_{\mu(z_0)}^{\mu} f(\mu)d\mu} = f_1(\mu). \tag{3.53}$$

Правая часть этого дифференциального уравнения первого порядка также не содержит явной зависимости от аргумента. Его интегрирование с использованием второго граничного условия для μ в точке $z = z_1$ приводит к формуле

$$\int\limits_{z_1}^{z} dz = z - z_1 = \int\limits_{\mu(z_1)}^{\mu} \frac{d\mu}{f_1(\mu)}, \tag{3.54}$$

описывающей пространственное распределение потенциала (в виде обратной функции $z = z(\mu)$). В общем, вычисление как первого, так и второго интегралов уравнения Пуассона требует численного интегрирования, однако во многих важных случаях удается получить хорошие аналитические аппроксимации.

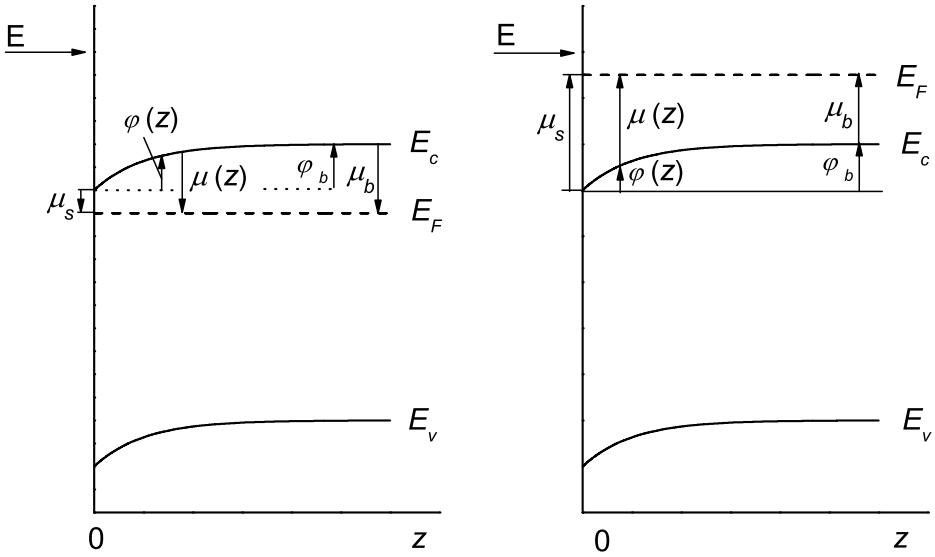

Рис. 3.1: Зонные диаграммы невырожденного и вырожденного полупроводника n – типа в электрическом поле.

3.2.3 Режим линейного экранирования

Вернемся к нашей задаче экранирования приложенного к поверхности полупроводника внешнего электрического поля, предполагая, что полупроводник однородно легирован донорными примесями, а зарядом дырок можно пренебречь (материал n –типа, рис. 3.1). Мы не будем накладывать каких-либо ограничений на степень вырождения электронного газа и вид закона дисперсии, однако в этом разделе ограничимся слабыми электрическими полями, такими, что химпотенциал слабо изменяется в приповерхностной области по сравнению с его значением в объеме полупроводника $\mu_b = \mu(z \to \infty)$

$$\mu(z) - \mu_b << \mu_b.$$

Поскольку в глубине полупроводника, где поле полностью экранировано и полупроводник нейтрален, концентрация электронов равна концентрации ионизированных доноров $n(\mu_b) = N_D^+$ то в пренебрежении перезарядкой примесей в приповерхностной области

$$\varrho_n(\mu) = N_d^+ - n(\mu) = n(\mu_b) - n(\mu).$$

Разлагая $n(\mu)$ в ряд Тейлора в окрестности μ_b

$$n(\mu) = n(\mu_b) + \left.\frac{dn}{d\mu}\right|_{\mu_b} (\mu - \mu_b) + \ldots$$

и ограничиваясь, виду малости $\mu - \mu_b$, линейным членом разложения [9] приходим к уравнению

$$\frac{d^2\mu}{dz^2} = \frac{4\pi e^2 (dn/d\mu)|_{\mu_b}}{\epsilon}(\mu - \mu_b). \tag{3.55}$$

Из уравнения (3.55) видно, что коэффициент перед скобкой в его правой части имеет размерность квадрата обратной длины. Введением характерной длины

$$L_{scr} = \sqrt{\frac{\epsilon}{4\pi e^2 (dn/d\mu)|_{\mu_b}}} \tag{3.56}$$

(физический смысл этого параметра выяснится ниже) уравнение, описывающее пространственное распределение потенциала, принимает простой вид

$$\frac{d^2\mu}{dz^2} = \frac{(\mu - \mu_b)}{L_{scr}^2}, \tag{3.57}$$

с граничными условиями

$$\mu = \mu_b; \quad \frac{d\mu}{dz} = 0 \quad при \quad z = \infty. \tag{3.58}$$

Вместо электрического поля на поверхности полупроводника

$$E(z=0) = \left.\frac{d\varphi}{edz}\right|_{z=0} = -\left.\frac{d\mu}{edz}\right|_{z=0}$$

удобно задать величину поверхностного химпотенциала (связь между ними будет установлена ниже)

$$\mu(0) = \mu_s. \tag{3.59}$$

Для нашего линеаризованного уравнения (3.57) вычисление первого и второго интегралов Пуассона элементарно. Подставляя $f(\mu) = (\mu -$

[9] По этой причине рассматриваемое приближение слабого поля называют режимом линейного экранирования, или режимом плоских зон, поскольку изгибы зон в этом случае малы, т. е. зоны близки к плоским.

$\mu_b)/L_{scr}^2$ в (3.53) и используя граничные условия (3.58) для $z_0 = \infty$, получаем

$$\frac{d\mu}{dz} = -\frac{1}{L_{scr}}\sqrt{(\mu - \mu_b)^2} = -\frac{\mu - \mu_b}{L_{scr}}, \tag{3.60}$$

где выбор знака перед корнем обусловлен тем, что при $\mu > \mu_b$ химпотенциал уменьшается с ростом z, а при $\mu < \mu_b$ возрастает. Уравнение (3.54), после подстановки этого выражения для производной, с использованием граничного условия (3.59) на поверхности ($z_1 = 0$) дает простую экспоненциальную зависимость химпотенциала от расстояния до поверхности полупроводника

$$\frac{\mu - \mu_b}{\mu_s - \mu_b} = \exp\left(-\frac{z}{L_{scr}}\right). \tag{3.61}$$

Если отсчитывать электростатический потенциал φ от дна зоны проводимости на поверхности полупроводника $\varphi(0) = 0$, то его пространственное распределение описывается формулой

$$\varphi = \mu_s - \mu = \varphi_b\left(1 - \exp -\frac{z}{L_{scr}}\right), \tag{3.62}$$

где $\varphi_b = \mu_s - \mu_b$ – электростатический потенциал в объеме полупроводника (см. рис. 3.1). Из уравнения (3.60) следует, что электрическое поле

$$E = -\frac{1}{e}\frac{d\mu}{dz} = \frac{\mu_s - \mu_b}{eL_{scr}}\exp\left(-\frac{z}{L_{scr}}\right) \tag{3.63}$$

также спадает по экспоненциальному закону, а введенная нами выше *величина* L_{scr} (3.56) характеризует глубину проникновения электрического поля в проводник и *носит название длины экранирования.* При $z = 0$ выражение (3.63) устанавливает связь между приложенным к поверхности полупроводника электрическим полем и поверхностным химпотенциалом

$$E(0) = \frac{\mu_s - \mu_b}{eL_{scr}}. \tag{3.64}$$

Фигурирующая в (3.61) – (3.64) разность $\mu_s - \mu_b$ называется *приповерхностным изгибом зон.* В случае положительного изгиба зон (зоны в приповерхностной области изогнуты вниз по отношению к объему, образуя вместе с потенциальной стенкой на границе потенциальную яму

для электронов) электрическое поле, согласно (3.64), $E(0) > 0$, т. е. направлено в положительном направлении оси z, как это и должно быть, поскольку именно при таком направлении поля имеющие отрицательный заряд $-e$ электроны притягиваются к поверхности.

Рассмотрим зависимость длины экранирования (3.56)

$$L_{scr} = \sqrt{\frac{\epsilon}{4\pi e^2 (dn/d\mu)|_{\mu_b}}}$$

от зонных параметров материала и температуры, содержащихся в производной $dn/d\mu|_{\mu_b}$. Хотя последняя легко находится в обобщенном виде из формулы (2.69) с использованием соотношения (2.66)

$$dn_d/d\mu = \frac{N_{cd}}{k_B T} \mathcal{F}_{\frac{d}{r}-2}\left(\frac{\mu}{k_B T}\right), \tag{3.65}$$

мы ограничимся анализом наиболее простых, но интересных с физической точки зрения пределов классической и вырожденной статистики. Хотя в настоящем разделе мы рассматриваем трехмерную задачу, однако, имея в виду дальнейшие обобщения, будем использовать обобщенные формулы для концентрации и плотности состояний.

Приближение Дебая–Хюккеля

В случае невырожденного электронного газа зависимость d – мерной концентрации от химпотенциала, согласно (2.95), имеет вид

$$n_d = C \exp \frac{\mu}{k_B T}$$

с независящим от μ коэффициентом C, откуда

$$\frac{dn_d}{d\mu} = \frac{1}{k_B T} C e^{\frac{\mu}{k_B T}} = \frac{n_d}{k_B T}. \tag{3.66}$$

Следует подчеркнуть, что это соотношение между $dn_d/d\mu$ и n_d не зависит ни от размерности пространства, ни от закона дисперсии, ни от величины химпотенциала (само значение $dn_d/d\mu$ от этих величин, конечно, зависит, поскольку $n_d = n_d(\mu_b, d, r)$ – см.(2.95) и (2.68)). Поставляя это значение при $\mu = \mu_b$ и $d = 3$ ($n_d = n_{3D} \equiv n$) в (3.56), получаем для длины

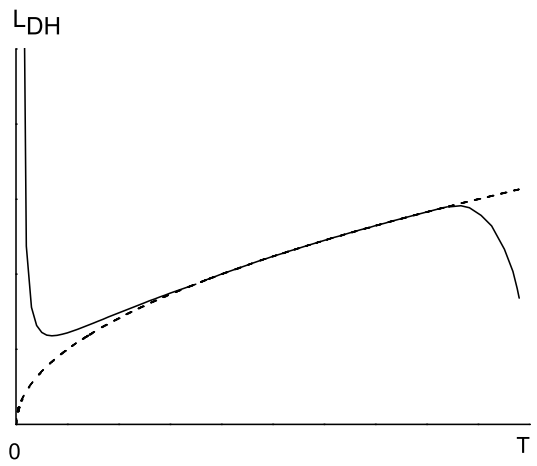

Рис. 3.2: Температурная зависимость длины экранирования Дебая–Хюккеля. Пунктиром показана корневая зависимость L_{DH} в формуле (3.67).

экранирования в невырожденном трехмерном электронном газе формулу *Дебая–Хюккеля*

$$L_{DH} = \sqrt{\frac{\epsilon k_B T}{4\pi e^2 n_b}}, \tag{3.67}$$

где $n_b = n(\mu_b)$.

Возрастание длины экранирования Дебая – Хюккеля с температурой согласуется с простыми физическими соображениями. Для разреженного (классического) газа именно тепловое движение электронов с характерной скоростью $v_T \propto \sqrt{T}$ препятствует им собраться в тонкий слой у поверхности полупроводника. Однако корневая температурная зависимость справедлива, если только концентрация электронов в (3.67) не изменяется с температурой. В рассматриваемом случае легированного полупроводника n – типа, концентрация зонных носителей определяется их температурным забросом в зону с донорных примесей и остается постоянной только при достаточно высокой температуре, соответствующей полной ионизации примесей (выше так называемой температуры истощения примесей T_D, которая может быть найдена из уравнения электронейтральности). При понижении температуры $T < T_D$ концентрация быстро уменьшается (по закону $n_b \propto \exp(-\varepsilon/2k_B T)$, где ε_D – энергия ионизации доноров), длина экранирования возрастает и в пределе

$T \to 0$ $L_{DH} \to \infty$ (полупроводник становится изолятором). Зависимость $L_{DH} \propto \sqrt{T}$ нарушается и со стороны высоких температур, а именно при переходе к режиму собственной проводимости, когда становится существенным возрастание концентрации электронов проводимости за счет их термической активации из валентной зоны. Это приводит к уменьшению L_{DH} с ростом температуры (рис. 3.2). Однако в этом температурном интервале формула (3.67) уже не справедлива, поскольку объемный заряд возникает теперь в основном за счет пространственного разделения внешним полем электронов и дырок. В уравнении Пуассона (3.50) необходимо учесть вклад дырок валентной зоны в объемную плотность заряда (3.51). Для металлов и сильнолегированных полупроводников с металлическим типом проводимости при $T = 0$ концентрация от температуры не зависит, однако и выражение (3.67), полученное для классического распределения, в этих случаях неприменимо ввиду вырождения электронного газа и должно быть пересмотрено.

Приближение Томаса–Ферми

Для сильно вырожденного газа согласно (2.80)

$$n_d = C_1 \mu^{\frac{d}{r}},$$

откуда находим (коэффициент C_1 от химпотенциала не зависит)

$$\frac{dn_d}{d\mu} = \frac{d}{r} \cdot C_1 \mu^{\frac{d}{r}-1} = \frac{d}{r} \cdot \frac{C_1 \mu^{\frac{d}{r}}}{\mu} = \frac{d}{r} \cdot \frac{n_d}{\mu}. \tag{3.68}$$

Подстановка этого выражения при $d = 3$ и $\mu = \mu_b$ в (3.56) приводит к формуле для длины экранирования в вырожденном электронном газе (длина *Томаса-Ферми*)

$$L_{TF} = \sqrt{\frac{r\epsilon\mu_b}{12\pi e^2 n_b}}, \tag{3.69}$$

откуда

$$L_{TF} = \sqrt{\epsilon\mu_b / 6\pi e^2 n_b}$$

для квадратичного ($r = 2$) и

$$L_{TF} = \sqrt{\epsilon\mu_b / 12\pi e^2 n_b}$$

для линейного ($r = 1$) законов дисперсии.

Как это уже неоднократно подчеркивалось, в случае сильного вырождения, в отклике электронной системы на воздействие внешних полей, в силу принципа Паули, могут участвовать только электроны в узкой полосе состояний вблизи уровня Ферми, поскольку электроны лежащие под уровнем Ферми не могут перейти в уже занятые вышележащие состояния и, следовательно, не могут ускоряться электрическим полем, а, следовательно, и участвовать в экранировании. В выражение же для длины экранирования (3.69) входит полная концентрация электронов, как будто в экранировании поля участвуют все электроны. Для прояснения этого вопроса вернемся к общей формуле (3.56), которая содержит два материальных параметра: ϵ и $dn_d/d\mu|_{\mu_b}$. Диэлектрическая проницаемость ϵ характеризует экранирование поля за счет поляризации решетки. Для вычисления второго, представляющего наибольший интерес, обратимся к общей формуле для концентрации (2.57). Производная концентрации (2.57) по химпотенциалу равна

$$\begin{aligned}\frac{dn_d}{d\mu} &= \frac{\partial}{\partial\mu}\int\limits_0^\infty D_d(\varepsilon)f(\varepsilon-\mu)d\varepsilon = \int\limits_0^\infty D_d(\varepsilon)\frac{\partial f(\varepsilon-\mu)}{\partial\mu}d\varepsilon \\ &= -\int\limits_0^\infty D_d(\varepsilon)\frac{\partial f(\varepsilon-\mu)}{\partial\varepsilon}d\varepsilon,\end{aligned}$$

где учтено, что плотность состояний от μ не зависит, а функция распределения зависит только от разности $\varepsilon - \mu$. Поскольку при сильном вырождении производная функции распределения по энергии близка к δ–функции с обратным знаком, то

$$\frac{dn_d}{d\mu} = -\int\limits_0^\infty D_d(\varepsilon)\frac{\partial f(\varepsilon-\mu)}{\partial\varepsilon}d\varepsilon \approx \int\limits_0^\infty D_d(\varepsilon)\delta(\varepsilon-\mu)d\varepsilon = D_d(\mu), \qquad (3.70)$$

т. е. производная $dn_d/d\mu$ есть плотность состояний на уровне Ферми. Подчеркнем, что этот результат не зависит ни от размерности пространства, ни от закона дисперсии (в частном случае степенного закона дисперсии соотношение (3.70) непосредственно следует из (3.68) и (2.81)).

Таким образом

$$L_{TF} = \sqrt{\frac{\epsilon}{4\pi e^2 D_{3D}(\mu_b)}} \tag{3.71}$$

т. е. экранирование при сильном вырождении, как и рассмотренная выше теплоемкость, действительно обусловлено только электронами с энергиями вблизи уровня Ферми, в полном соответствии с принципом Паули.

"Зависимость" же длины экранирования от концентрации обусловлена просто тем фактом, что согласно, (3.14)

$$D_d(\mu) \propto n_d^{(d-r)/d},$$

в трехмерном электронном газе с изменением концентрации меняется и плотность состояний на уровне Ферми. В результате, длина экранирования (3.71) уменьшается (эффективность экранирования растет) с ростом концентрации

$$L_{TF} \propto n_b^{(r-3)/6}$$

(для квадратичного и линейного законов дисперсии соответственно по закону $L_{TF} \propto 1/n_b^{1/6}$ и $L_{TF} \propto 1/n_b^{1/3}$). Зависимость длины экранирования в электронном газе от концентрации носит опосредованный характер. Она обусловлена не зависимостью эффективности экранирования от общего числа свободных зарядов, а тем, что в *трехмерном* электронном газе увеличение концентрации сопровождается возрастанием плотности состояний на уровне Ферми. Это утверждение носит, конечно, общий для вырожденных систем характер. Оно справедливо для любых явлений и описывающих их параметров [10] (одно из них - теплоемкость мы уже рассмотрели ранее) в системах любой размерности и с любым законом дисперсии. Отклик вырожденной системы всегда определяется не полной концентрацией частиц, основная часть которых заморожена

[10]Имеются в виду явления при близких к термодинамическому равновесию условиях, имеющие место, в принципе, при сколь угодно малой интенсивности обуславливающих их внешних воздействий, т. е. характеризуемые непрерывным изменением энергии. Воздействие может носить и дискретный характер, например при поглощении частицей в энергетическом состоянии E, лежащим глубоко под уровнем Ферми, кванта энергии $E_F - E >> k_B T$, в результате которого частица окажется в состоянии выше уровня Ферми, а в континууме состояний появится дырка. Такие процессы приводят к отклонению системы от состояния термодинамического равновесия и ее описание единым электрохимическим потенциалом, т. е. равновесной функцией распределения, становится невозможным.

принципом Паули ($\mu >> k_BT$!), а частицами в энергетическом интервале порядка нескольких k_BT вблизи энергии Ферми. Концентрация таких "активных" частиц $\sim D_n(\mu)k_BT$ (это и есть квазичастицы упомянутой в разделе 2.4.5 модели элементарных возбуждений Ферми–жидкости) не зависит напрямую от полного их числа. Однако эти величины могут быть связаны, если плотность состояний $D_n(\mu)$ или другие существенные для данного явления характеристики частиц с энергиями $\varepsilon \sim \mu$ (скорость, интенсивность рассеяния на дефектах, соответствующая эффективная масса и т. п.) изменяются с концентрацией. Такая не обусловленная причинно связь (скорее корреляция) параметров системы с концентрацией может вводить в заблуждение, особенно, если в силу тех или иных причин она линейна, как это, например, имеет место для проводимости (см. раздел 8.3).

В разд. 3.1.2 мы уже отмечали, что отсутствие прямой связи между эффектами в вырожденном электронном газе с полной концентрацией электронов предельно ярко проявляется в системах пониженной размерности. В случае длины экранирования (3.71) единственным материальным параметром, характеризующим эффективность экранирования при низких температурах, является плотность состояний на уровне Ферми. При степенном законе дисперсии из соотношения (3.14) видно, что плотность состояний $D_n(\mu)$ и концентрация n_d никак не коррелированы друг с другом, если $d = r$, т. е. в случае двумерных систем с квадратичным спектром и одномерных с линейным законом дисперсии. Это, конечно, является прямым следствием установленной нами ранее независимости плотности состояний от энергии при $d = r$ (см. (2.28) и (2.33)) и предположения о сильном вырождении электронного газа. Очевидно, что в трехмерном газе при физически осмысленных законах дисперсии $r \leq 2$ такая ситуация не может быть реализована. Важнейшая характеристика вырожденной системы – плотность состояний на уровне Ферми всегда возрастает с трехмерной концентрацией.

К сожалению, сопоставление длин экранирования в газе заряженных частиц различной размерности не может быть проведено непосредственно, основываясь только на анализе плотности состояний, как это было

сделано в случае теплоемкости. Дело в том, что выражение (3.56) не может носить универсального характера (справедливого для всех размерностей пространства) уже из соображений размерности, поскольку длина L_{scr} выражается через удельную характеристику (единственную!) - $dn/d\mu$ и имеет размерность длины только в трехмерном случае. Для каждой размерности задача об экранировании требует отдельного рассмотрения.

3.2.4 Экранирование в двумерном электронном газе

Для рассмотрения экранирования в системах меньшей размерности, мы по-прежнему должны исходить из уравнения Пуассона, которое по своей природе, будучи следствием закона Кулона, является трехмерным [11]. Обратимся к задаче о "поверхностном" экранировании (также в линейном приближении) в двумерном электронном газе. Рассмотрим перпендику-

[11] В рамках настоящего курса мы не рассматриваем вопросов, относящихся к двумерной и одномерной электродинамике, носящих в значительной степени абстрактный характер. Укажем только на модификацию закона Кулона. Отправной точкой для "объяснения" закона Кулона является требование независимости связанного с элементарным зарядом e потока потенциального векторного поля, пронизывающего замкнутую поверхность вокруг заряда от ее формы или размеров (силовые линии не могут быть замкнутыми). Поскольку из соображений сферической симметрии величина поля должна зависеть только от расстояния до заряда (изотропность пространства), в качестве такой поверхности удобно выбрать сферу с центром на заряде. Отсюда для трехмерной системы следует, что $E = c_3 e/r^2$ (c_3 – зависящий от системы единиц коэффициент пропорциональности), поскольку только в этом случае поток

$$\Phi_{3D} = \oint\limits_S \mathbf{E} \cdot d\mathbf{s} = ce \oint\limits_S \frac{1}{r^2} ds = c_3 \frac{e}{r^2} \oint\limits_S ds = c_3 \frac{e}{r^2} 4\pi r^2 = c_3 4\pi e$$

не зависит от r (трехмерная терема Гаусса). В случае двумерного векторного поля (силовые линии не выходят из плоскости) совершенно аналогично следует, что указанным выше требованиям удовлетворяет только поле $E = c_1 e/r$

$$\Phi_{2D} = \oint\limits_L \mathbf{E} \cdot d\mathbf{l} = c_2 e \oint\limits_L \frac{1}{r} dl = c_2 \frac{e}{r} \oint\limits_L dl = c_1 2\pi e,$$

а потенциал двумерного кулоновского поля зависит от расстояния по логарифмическому закону $\varphi(r_1) - \varphi(r_2) = c_1 e \ln(r_1/r_2)$. Очевидно, что одномерное векторное поле вообще не зависит от расстояния $E = c_1 e r^0$, поскольку силовая линия полностью локализована в "одномерной нити"

$$\Phi_{1D} = 2c_1 e.$$

Используя (2.39), теорему Гаусса можно записать в обобщенном виде

$$\Phi_d = \frac{c_d e}{r^{d-1}} S_d = c_d \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} e.$$

лярный оси y двумерный слой, помещенный в полубесконечную среду с той же диэлектрической проницаемостью ϵ (обобщение на случай отличающихся ϵ по сути тривиально), полагая, что внешнее электрическое поле по-прежнему приложено в направлении оси z и однородно в плоскости x, y. *Трехмерная* плотность индуцированного внешним полем заряда ϱ_n в таком двумерном слое, фигурирующая в уравнении Пуассона

$$\frac{d^2\mu}{dz^2} = \frac{4\pi\varrho(\mu)}{\epsilon},$$

может быть выражена через *двумерную* концентрацию в виде

$$\varrho_n(\mu) = e(n_s(\mu) - n_s(\mu_b))\delta(y),$$

где μ - величина химпотенциала в точке z, усредненная по распределению электронов вдоль перпендикулярного двумерному слою направления y (мы ограничиваемся рассмотрением простейшего случая идеальной двумерной системы, описываемой дельта-функцией $\psi^*(y) \cdot \psi(y) = \delta(y)$). Поскольку изменения потенциала считаются малыми, выражение для ϱ может быть линеаризовано, подобно тому как это было сделано для трехмерно случая

$$\varrho_n(\mu) = e^2 \frac{dn_s}{d\mu}|_{\mu_b} \delta(y)(\mu(z) - \mu_b).$$

В результате уравнение Пуассона может быть представлено в виде

$$\frac{d^2\mu}{dz^2} = \frac{4\pi e^2}{\epsilon} \left.\frac{dn_s}{d\mu}\right|_{\mu_b} \delta(y)\mu = \frac{2\delta(y)}{L_{scr}}(\mu - \mu_b), \tag{3.72}$$

где

$$L_{scr} = \frac{\epsilon}{2\pi e^2 (dn_s/d\mu)|_{\mu_b}}. \tag{3.73}$$

Сопоставление этого уравнения с соответствующим уравнением для трехмерного электронного газа (3.55) показывает, что величина L_{scr} в (3.72) играет в двумерном случае ту же роль, что и длина экранирования (3.56) в трехмерном [12].

[12]Очевидно, что $\delta(y)$ имеет размерность обратной длины, поскольку $\int\limits_{-\infty}^{\infty} \delta(y)dy = 1$. То же самое, конечно, справедливо и для модуля волновой функции $\psi(y)$ произвольного вида в силу ее нормировки.

Как мы видели, в случае сильного вырождения $(dn_d/d\mu)|_{\mu_b}$ есть плотность состояний на уровне Ферми, вне зависимости от размерности пространства или закона дисперсии. Поскольку при квадратичном законе дисперсии плотность состояний в двумерном электронном газе (2.28) при всех энергиях одна и та же , то длина экранирования при сильном вырождении не зависит от энергии Ферми, а следовательно и от концентрации электронов

$$L_{scr} \equiv L_{TF} = \frac{\epsilon}{2\pi e^2 D(\mu)} = \frac{\epsilon\hbar^2}{2me^2} = const \tag{3.74}$$

Таким образом, вырожденный двумерный газ как высокой, так и низкой концентрации (при квадратичном законе дисперсии!) одинаково эффективно экранирует статическое электрическое поле. Это неожиданное с классической точки зрения поведение, как и рассмотренное ранее подобное поведение теплоемкости вырожденного газа (не обязательного заряженного), является еще одной яркой иллюстрацией проявления принципа Паули в низкоразмерных системах - реакция системы электронов на внешнее поле, как и на термическое возбуждение не зависит от концентрации носителей заряда. В трехмерной системе ситуация, конечно, та же самая - длина экранирования определяется только плотностью состояний на уровне Ферми, последняя в этом случае, как мы видели, растет с энергией Ферми и, как следствие, с увеличением трехмерной концентрации носителей. В результате мы приходим к обычному поведению - чем больше свободных носителей заряда, тем более эффективно экранируются внешние поля (тем меньше длина экранирования).

В связи с результатом (3.74) следует сделать важное замечание. Ясно, что электронный газ не может образовывать пространственных структур масштаба, меньшего, чем соответствующая ему минимальная длина волны де Бройля λ_B. Поэтому эффекты экранирования важны, только если длина экранирования велика по сравнению с λ_B. В вырожденном электронном газе минимальная длина волны соответствует электронам на уровне Ферми и указанное условие имеет вид

$$L_{TF}k_F >> 1. \tag{3.75}$$

Для двумерного электронного газа $k_F = \sqrt{2\pi n_{2D}}$ и формула (3.74), согласно (3.75), в действительности имеет смысл только при достаточно высокой концентрации, удовлетворяющей, с точностью до коэффициента порядка единицы, условию

$$n_{2D} >> \left(\frac{me^2}{\epsilon\hbar^2}\right)^2. \tag{3.76}$$

Условие (3.75) должно выполняться при любой размерности пространства и соотношение, аналогичное (3.76), имеет место и при экранировании в трехмерной системе. Для квадратичного закона дисперсии длина Томаса–Ферми (3.71), как функция концентрации имеет вид

$$L_{TF} = \frac{1}{2}\left(\frac{\pi}{3}\right)^{\frac{1}{6}}\sqrt{\frac{\epsilon\hbar^2}{me^2}}\frac{1}{n^{1/6}}, \tag{3.77}$$

где учено, что $D_{3D}(\mu) = 2m^{3/2}\mu^{1/2}/3\pi^2\hbar^3$ (см. Табл. 2.1), а химпотенциал связан с концентрацией соотношением $\mu = (3\pi^2\hbar^3 n)^{2/3}/2m$ (см. Табл. 2.3). Поскольку $k_F = (3\pi^2 n)^{1/3}$ (Табл. 2.3), то условие (3.75) сводится к соотношению

$$n >> \left[\frac{me^2}{\epsilon\hbar^2}\right]^3, \tag{3.78}$$

определяющему область концентраций, для которой существенны эффекты экранирования в трехмерной системе.

Как и следует ожидать, формула (3.73) совместно с (2.74) приводят к уже полученному нами выражению для независящей от концентрации длины экранирования (3.74). С другой стороны, используя (2.75), получаем, что в невырожденном двумерном электронном газе (при квадратичном же законе дисперсии) длина экранирования обратно пропорциональна поверхностной плотности электронов

$$L_{scr} \equiv L_{DH} = \frac{kT}{\epsilon 2\pi e^2 n_s}, \tag{3.79}$$

т. е. в этом случае мы имеем привычное поведение – эффективность экранирования растет при увеличении концентрации свободных носителей заряда.

В реальных двумерных системах с вырожденным электронным газом низкой плотности при рассмотрении экранирования следует учитывать наличие хвостов плотности состояний (плотность состояний на этих участках зависит от энергии) и отличие температуры от абсолютного нуля (см. ниже).

3.2.5 Экранирование кулоновского потенциала

Введенные выше при описании поверхностного экранирования характеристические длины L_{scr} в равной степени характеризуют и экранирование кулоновского поля точечного заряда, потенциал которого в диэлектрической среде с проницаемостью ϵ равен

$$V(r) = \frac{Ze}{\epsilon r}.$$

В присутствии свободных носителей заряда потенциал, благодаря экранированию, спадает быстрее. Для трехмерной системы уравнение Пуассона, правая часть которого теперь наряду с индуцированным зарядом, содержит точечный внешний заряд, благодаря сферической симметрии кулоновского поля сводится к одномерному уравнению, решением которого является экранированный кулоновский потенциал (*потенциал Юкавы*)

$$V(r) = -\frac{Ze}{r\epsilon} \exp -\frac{r}{L_{scr}}. \tag{3.80}$$

с длиной экранирования, определяемой тем же соотношением (3.56).

Для двумерного электронного слоя в диэлектрической среде вместо сферически симметричной задачи мы имеем дело с 1+2 мерной задачей, обладающей центральной симметрией только в плоскостях, параллельных двумерному слою (предполагается, что экранируемый заряд Ze расположен в точке с цилиндрическими координатами $r = 0, z = z_0$). Решение уравнения Пуассона для такой системы более низкой симметрии значительно сложнее. В виду осевой симметрии задачи результирующий экранированный потенциал выражается через функций Бесселя нулевого порядка. Мы ограничимся только приведением конечного результата

для идеальной двумерной системы, когда все электроны лежат на плоскости $z = 0$ и для достаточно больших расстояний от заряда $r >> L_{scr}$

$$V(r) = -\frac{Ze}{r\epsilon}\left(\frac{L_{scr}}{r}\right)^2\left(1 + \frac{z_0}{L_{scr}}\right), \tag{3.81}$$

где L_{scr} - двумерная длина экранирования (3.73). Степенное спадание экранированного кулоновского потенциала с расстоянием (3.81) значительно слабее экспоненциального для трехмерной системы (3.80) – двумерный электронный газ слабо экранирует кулоновский потенциал по сравнению с трехмерным. В этом состоит одно из качественных отличий экранирования в двумерной и трехмерной электронных системах.

При рассмотрении экранирования кулоновского потенциала, удобно выразить длину (радиус) линейного экранирования через характеристическую длину, описывающую протяженность волновой функции электрона, связанного на таком потенциале, а именно через эффективный боровский радиус

$$a_B = \frac{\epsilon\hbar^2}{me^2}. \tag{3.82}$$

Если также ввести среднее расстояние между электронами $r_s = n_d^{-1/d}$, то формулы для радиуса экранирования принимают вид

$$L_{TF} = \frac{1}{2}\left(\frac{\pi}{3}\right)^{1/6}\sqrt{a_B r_s} \tag{3.83}$$

для трехмерного и

$$L_{TF} = \frac{a_B}{2} \tag{3.84}$$

для двумерного электронного газа. Условия квазиклассичности (3.76) и (3.78), будучи выраженными в этих параметрах могут быть записаны в общей для обеих размерностей форме (с точностью до коэффициентов порядка единицы)

$$r_s << a_B. \tag{3.85}$$

Действительно, используя выражение (3.82) для a_B, неравенство (3.85) может быть преобразовано к в виду

$$\frac{e^2}{\epsilon r_s} << \frac{\hbar^2}{2mr_s^2} = \frac{\hbar^2}{2mn_d^{2/d}}.$$

Поскольку по порядку величины $n_d \simeq k_f^{-d}$, то это соотношение сводится к

$$\frac{e^2}{\epsilon r_s} << \frac{\hbar^2 k_F^2}{2m}, \tag{3.86}$$

что эквивалентно требованию, чтобы кинетическая энергия электронов W_k была велика по сравнению с потенциальной энергией V_{ee} их кулоновского взаимодействия

$$\frac{W_k}{V_{ee}} = \frac{\mu}{(e^2/\epsilon r_s)} >> 1. \tag{3.87}$$

В этом выражении мы учли, что для вырожденного газа средняя кинетическая энергия электронов близка к их энергии Ферми μ. Легко видеть, что это условие одновременно обеспечивает и линейность экранирования кулоновского потенциала [13]. Для линейности экранирования требуется, чтобы потенциальная энергия электронов в кулоновском поле на расстояниях порядка длины экранирования была мала по сравнению с энергией Ферми. Однако, поскольку, согласно (3.83) – (3.85),

$$L_{TF} >> r_s \tag{3.88}$$

(в области экранирования должно быть достаточно много носителей, в противном случае задача вообще теряет смысл), то неравенство (3.86) будет тем более выполнено, если в левой его стороне заменить r_s на L_{TF}

$$\frac{e^2}{\epsilon L_{TF}} << \mu. \tag{3.89}$$

Отметим, что в идеальной одномерной системе (силовые линии электрического поля локализованы в квантовой нити) достаточно одного электрона, чтобы полностью экранировать потенциал кулоновского заряда.

Фигурирующее в (3.87) соотношение между кинетической энергией и энергией межэлектронного кулоновского взаимодействия относится к числу наиболее важных физических параметров (эти величины определяют гамильтониан в уравнении Шредингера, а, следовательно, и свойства системы), в том числе в физике конденсированного состояния и

[13] Обратное утверждение, вообще говоря, не справедливо – экранирование может быть линейным, но не квазиклассичным! В частности, как мы увидим позднее, при поверхностном экранировании в вырожденном полупроводнике длина линейного экранирования не превышает минимальной длины волны де Бройля электронов.

физике квантовых низкоразмерных систем в особенности. В случае изотропного степенного закона дисперсии отношение (3.87) может быть легко записано в обобщенном виде, если использовать соотношение (2.80) между μ $(= W_k)$ и n_d

$$\frac{W_k}{V_{ee}} = \left[d\pi^{\frac{d}{2}} 2^{d-2} \Gamma(\frac{d}{2}) \right]^{\frac{r}{d}} \hbar^r c_r \frac{\epsilon}{e^2} n_d^{\frac{r-1}{d}}, \tag{3.90}$$

где также учтено, что $r_s = (1/n_d)^{1/d}$. В случае свободного электронного газа и эквивалентного ему газа зонных носителей с параболическим спектром это отношение равно

$$\frac{W_k}{V_{ee}} = c_d \frac{\hbar^2 \epsilon}{me^2} n_d^{\frac{1}{d}} = c_d a_B n_d^{\frac{1}{d}}, \tag{3.91}$$

где зависящий от размерности пространства численный множитель c_d равен: $c_{1D} = \pi^2/8 \sim 1$, $c_{2D} = \pi \sim 3$, и $c_{3D} = (3\pi^2)^{2/3}/2 \sim 5$. Хотя с ростом концентрации межэлектронное расстояние уменьшается, а следовательно растет кулоновское взаимодействие $V_{ee} \propto r_s^{-1} = (n_d)^{1/d}$, увеличивается также и кинетическая энергия электронов, причем для любой размерности пространства, последняя увеличивается быстрее $W_k \propto k_F^2 \propto (n_d)^{2/d}$. В результате, с ростом плотности электронного газа влияние многоэлектронных эффектов уменьшается. Отметим, что благодаря малости эффективной массы и больших диэлектрических проницаемостей, условия для рассмотрения электронного газа, как газа невзаимодействующих электронов, в полупроводниках выполняются при значительно меньших концентрациях, чем в металлах, в которых эффективный боровский радиус близок к его величине для свободного электрона.

Совершенно иная ситуация имеет место в случае линейного дисперсионного соотношения. Как видно из общего выражения (3.90), концентрационная зависимость отношения $W_k/V_{ee} \propto n_d^{(r-1)/d}$ при $r = 1$ исчезает для систем любой размерности, что является следствием того, что при линейной дисперсии кинетическая энергия $W_k \propto k_F \propto (n_d)^{1/d}$ увеличивается с той же скоростью, что и потенциальная. Как следствие, соотношение между W_k и V_{ee} для "ультрарелятивистского" электронного газа является константой, определяемой только размерностью пространства,

материальными параметрами и мировыми постоянными

$$\frac{W_k}{V_{ee}} = c_d \left[\frac{s\hbar\epsilon}{e^2}\right], \tag{3.92}$$

причем зависимость от размерности системы d (через коэффициенты c_d: $c_{D1} = \pi/2 \sim 1,6$, $c_{D2} = \sqrt{2\pi} \sim 2,5$, $c_{3D} = \sqrt[3]{3\pi^2} \sim 3.1$) не очень существенна. Легко видеть, что величина $\alpha = e^2/\epsilon s\hbar$, обратная множителю в квадратных скобках в (3.92), аналогична постоянной тонкой структуры $\alpha_0 = e^2/c\hbar$, с точностью до замены заряда на его экранированное в диэлектрической среде с проницаемостью ϵ значение $e/\sqrt{\epsilon}$, и скорости света c на кейновскую скорость s. Выражая эту величину, которую можно назвать эффективной постоянной тонкой структуры (подобно тому как мы ввели эффективный боровский радиус), через α_0, запишем формулу (3.92) в виде

$$\frac{W_k}{V_{ee}} = c_d \frac{s\epsilon}{c} \left[\frac{e^2}{c\hbar}\right]^{-1} = c_d \frac{s\epsilon}{c} \alpha^{-1}. \tag{3.93}$$

Как уже отмечалось, скорость $s \cong 10^8$ см/сек и практически одинакова для всех полупроводников с законом дисперсии кейновского типа, т. е. составляет $\sim 1/300$ скорости света, а типичные значения диэлектрической проницаемости $\epsilon \sim 10$. Поскольку $\alpha^{-1} \approx 137$, а величина размерного коэффициента $c_d \sim 2$, то кинетическая энергия электронного газа с линейной дисперсией примерно на порядок величины превышает энергию кулоновского отталкивания, при всех концентрациях и для любой размерности пространства.

В заключение раздела подчеркнем наиболее общие и существенные черты линейного экранирования.

- Длина (радиус) линейного экранирования не зависит от величины экранируемого потенциала и определяется лишь концентрацией (в вырожденном газе плотностью состояний на уровне Ферми) электронов и параметрами законов дисперсии.

- Экранирование зарядов системы происходит независимо. Так суммарный экранированный потенциал, создаваемый всеми кулоновскими центрами, равен сумме их экранированных потенциалов, т. е. каждый центр экранируется сам по себе. Этот вывод непосредственно следует из линейности уравнения Пуассона, постулируемой в рамках используемого в этом разделе приближения.

Практически любой прибор полупроводниковой электроники, а тем более приборы микро– и наноэлектроники, представляет собой неоднородную структуру, различные области которой отличаются типом или степенью легирования, а в гетероструктурах также и видом используемого полупроводника или металла (если учесть контакты, как правило, металлические, то все полупроводниковые приборы являются гетероструктурами). Эффекты экранирования играют определяющую роль в распределении концентрации и потенциала неоднородно легированных полупроводников (неоднородности в пространственном распределении примесей неизбежны, вопрос только в степени такой неоднородности). Даже если в пространственном распределении заряженных примесей имеется резкая ступенька, отделяющая две области с разным легированием, благодаря экранированию примесей электростатический потенциал, а вместе с ним и химпотенциал, а следовательно и концентрации носителей, будут плавными функциями координат, с пространственным масштабом порядка длины экранирования. Как правило (при не слишком высоких плотностях токов), такие неоднородные системы могут рассматриваться находящимися в условиях, близких к термодинамическому равновесию, т. е. с пространственно почти постоянным значением электрохимического потенциала. Как будет показано в следующем разделе, такое равновесие всегда носит динамический характер, а именно диффузионные потоки носителей, обусловленные неоднородностью концентрации носителей (градиентами химпотенциала) уравновешиваются потоками, связанными с их дрейфом в электрических полях, обусловленных градиентами электростатического потенциала.

В дальнейшем, при изучении потенциальных профилей в гетероструктурах с квантовыми ямами, мы, в основном, будем иметь дело с нелиней-

ным экранированием. Однако и этих случаях также удается, как правило, ввести характерные длины, описывающие пространственное распределение потенциала. Значения таких параметров, в общем случае, являются, конечно, функциями величины внешних или встроенных электрических полей.

3.3 Обобщенное соотношение Эйнштейна

Обратимся еще раз к зонной диаграмме полупроводника в электрическое поле на рис. 3.1. Поскольку в приповерхностной области пространственного заряда имеется электрическое поле и имеются подвижные (зонные) носители заряда, должен возникнуть электрический ток плотностью $\mathbf{j}_E(z) = ne\mu_n\mathbf{E}(z)$, где $\mu_n(= e\tau_p/m)$ - подвижность носителей. С другой стороны, поскольку полупроводник отделен от металлического электрода изолятором (т. е. на границе полупроводника имеется потенциальный барьер, толщина и высота которого настолько велики, что носители не могут туннелировать в состояния металлического электрода), ток в любом сечении полупроводника $z = const$ в стационарных условиях равен нулю. Это кажущееся противоречие легко разрешается, если учесть, что наряду с электрическим полем $\mathbf{E} = -\nabla\varphi/e$, т. е. градиентом электрического потенциала, в приграничной области полупроводника имеется и градиент химпотенциала, а , следовательно, и концентрации. Как следствие, возникает диффузионный поток электронов, пропорциональный градиенту концентрации. В условиях термодинамического равновесия (когда все переходные процессы, связанные с включением электрического поля закончились) $j = 0$ и диффузионный ток $\mathbf{j}_D = eD_n\nabla n$ (D_n – коэффициент диффузии) должен в точности компенсировать полевой (омический) ток (ниже мы полагаем, что поле направлено вдоль оси z $\mathbf{E} = (0, 0, E)$ и система изотропна)

$$j = ne\mu_n E + eD_n\nabla n = ne\mu_n E + eD_n\frac{dn}{dz} = ne\mu_n E + eD_n\frac{dn}{d\mu}\frac{d\mu}{dz} = 0 \quad (3.94)$$

Поскольку электрохимический потенциал $E_F = \mu(z) + \varphi(z)$ при $j = 0$ постоянен по образцу, то

$$\frac{d\mu(z)}{dz} = -\frac{d\varphi(z)}{dz} = -e\frac{dV(z)}{dz} = -eE$$

и последнее равенство в (3.94) устанавливает связь между коэффициентом диффузии и подвижностью

$$D_n = \frac{1}{e}\left(\frac{dn}{nd\mu}\right)^{-1}\mu_n. \tag{3.95}$$

Используя выражение для трехмерной концентрации при $\varepsilon = p^2/2m$ получаем отсюда для невырожденного трехмерного газа *соотношение*

$$D_n = \frac{k_B T}{e}\mu_n, \tag{3.96}$$

установленное впервые *Эйнштейном* и носящее его имя. В случае вырожденной статистики подобным образом получаем

$$D_n = \frac{2}{3}\frac{\mu}{e}\mu_n.$$

Нетрудно видеть, что рассуждения, использованные при выводе (3.95), остаются справедливыми при любой размерности пространства. Хотя размерности удельных величин j и n зависят от d, однако концентрация в (3.95) входит в комбинации $n/(dn/d\mu)$ с размерностью энергии (для величины $n/(dn/d\mu)$ часто используют термин "*энергия диффузии*"). В случае невырожденного газа, используя (3.66), имеем для любой размерности пространства и произвольного закона дисперсии

$$\frac{1}{n_d}\frac{dn_d}{d\mu} = \frac{1}{k_B T}.$$

Таким образом, соотношение Эйнштейна (3.95) для невырожденной статистики носит универсальный характер - для всех размерностей пространства и законов дисперсии коэффициент диффузии и подвижность связаны формулой (3.96).

Для сильно вырожденного газа, подставляя (3.68) в (3.95), получаем столь же простое соотношение

$$D_n = \frac{r}{d}\cdot\frac{\mu}{e}\mu_n, \tag{3.97}$$

зависящее однако, в отличие от случая классической статистики, от размерности пространства и закона дисперсии. Согласно (3.96) и (3.97), отношение D_n/μ_n линейно растет со средней кинетической энергией носителей в расчете на одну степень свободы ($\propto k_BT$ для невырожденной и $\propto \mu/d$ для вырожденной статистики).

Соотношение Эйнштейна позволяет просто получить выражение для проводимости, если она обусловлена диффузией электронов (далее мы увидим, что именно такова природа проводимости в квантующем магнитном поле). Выражая в стандартной формуле для проводимости $\sigma = en_d\mu_n$ подвижность через коэффициент диффузии, получаем

$$\sigma = e^2 D_n \frac{dn_d}{d\mu}$$

Для вырожденной статистики величина $dn_d/d\mu$ (см. (3.70)) есть плотность состояний на уровне Ферми, откуда

$$\sigma = e^2 D_n D_d(\mu), \tag{3.98}$$

т. е., как и экранирование, диффузионная проводимость вырожденного электронного газа (3.98) определяется только состояниями вблизи уровня Ферми. В случае $d = r$ (двумерный газ с квадратичным и одномерный с линейным законом дисперсии) плотность состояний на уровне Ферми не меняется с концентрацией и, следовательно, формула для диффузионной проводимости (3.98) не содержит в явном виде концентрации носителей и является еще одним ярким проявлением принципа Паули в системах пониженной размерности.

3.4 Задачи

1. Найдите среднюю энергию электронов

$$\overline{\varepsilon} = \frac{\mathcal{E}_d}{n_d} = \frac{\int \varepsilon D_n(\varepsilon) f_0(\varepsilon - \mu) d\varepsilon}{\int D_n(\varepsilon) f_0(\varepsilon - \mu) d\varepsilon}$$

в классическом и сильно вырожденном d – мерном газе с изотропным степенным законом дисперсии.

Ответ:

$$\overline{\varepsilon} = \frac{d}{r} k_B T,$$

для классического газа и

$$\overline{\varepsilon} = \frac{d}{d+r} \mu$$

для вырожденного газа. Из этих выражений видно, что если для классического газа средняя энергия линейно растет с увеличением числа степеней свободы, то для вырожденного стремится к насыщению при значении $\overline{\varepsilon} \approx \mu$.

Указание: Из выражения для энергии $\mathcal{E}$ видно, что ее вычисление полностью аналогично вычислению концентрации $n_d = N_c \mathcal{F}_{\frac{d}{r-1}}(\mu)$ (формулы (2.65)–(2.69))

$$\mathcal{E} = k_B T N_c \frac{\Gamma(d/r+1)}{\Gamma(d/r)} \mathcal{F}_{\frac{d}{r}}(\mu) = k_B T N_c \frac{d}{r} \mathcal{F}_{\frac{d}{r}}(\mu).$$

Использование аппроксимаций интегралов Ферми $\mathcal{F}_j = \exp(\mu/k_B T)$ для невырожденного и $\mathcal{F}_j(\mu/k_B T) = \frac{1}{\Gamma(j+2)}(\mu/k_B T)^{j+1}$ для сильновырожденного газа приводит к приведенным выше простым выражениям для $\overline{\varepsilon}$.

2. Что можно сказать, основываясь на результатах предыдущего задания, о теплоемкости при постоянном объеме классического и вырожденного газа при $T = 0$, если полагать плотность газа и соотношение между его концентрацией и химпотенциалом не зависящими при низких температурах от T?

3. Покажите, что "эффективная скорость"для плотности состояний для d – мерных акустических фононов равна

$$s_{ph} = s_{ph(l)} s_{ph(t)} \left(\frac{d}{s_{ph(t)}^d + (d-1) s_{ph(l)}^d} \right)^{1/d},$$

где $s_{ph(l)}$ и $s_{ph(t)}$ – скорости продольных и поперечных фононов.

4. Найдите среднюю энергию d – мерных акустических фононов в низко- и высокотемпературном пределах.

   Ответ:

$$\begin{aligned} \bar{\varepsilon}_{ph} &= \frac{d-1}{d} k_B \Theta_D \qquad (T >> \Theta_D) \\ \bar{\varepsilon}_{ph} &= d \frac{\zeta(d+1)}{\zeta(d)} k_B T \qquad (T << \Theta_D). \end{aligned}$$

   Для одномерной системы $\zeta(1) = \infty$ и в обоих пределах средняя энергия фононов равна нулю, что обусловлено бесконечной концентрацией одномерных фононов.

5. Выразите теплоемкости двумерного и трехмерного вырожденного газа с линейным законом дисперсии через энергию Ферми и концентрацию.

   Ответ:

$$\begin{aligned} c_d &= \frac{\pi k_B^2 T}{3 s^2 \hbar^2} \mu = \frac{\sqrt{2}\pi^{3/2}}{3} \frac{k_B^2 T}{s\hbar} \sqrt{n} \qquad (2D), \\ c_d &= \frac{k_B^2 T}{3 s^3 \hbar^3} \mu^2 = \left(\frac{\pi^4}{3}\right)^{1/3} \frac{k_B^2 T}{s\hbar} \sqrt[3]{n^2} \qquad (3D). \end{aligned}$$

6. Выведите соотношение Эйнштейна для двумерного электронного газа при произвольной температуре (в предположении квадратичного закона дисперсии).

7. Найдите длину экранирования в двумерном электронном газе при произвольной температуре для квадратичного закона дисперсии

   Ответ:

$$L_{scr} = \frac{\epsilon \hbar^2}{2me^2} \left(1 + \exp - \frac{\mu_b}{k_B T}\right).$$

3.5 Вопросы для самопроверки

1. Как модифицируется выражение для теплоемкости классического трехмерного идеального газа $c_{3D} = 3n_d k_B/2$ при его обобщении на

случай электронной системы произвольной размерности и произвольного степенного закона дисперсии (показатель степени r)?

2. Как модифицируется выражение для теплоемкости вырожденного трехмерного электронного газа $c_d = \frac{3}{2}\frac{\pi^2}{3}\frac{n_d}{\mu_0}k_B^2 T$. при его обобщении на случай электронной системы произвольной размерности и произвольного степенного закона дисперсии (показатель степени r)?

3. Согласно выражению для теплоемкости сильно вырожденного d -мерного электронного газа $c_d = \partial\mathcal{E}/\partial T|_V = \frac{\pi^2}{3}\frac{d}{r}\frac{n_d}{\mu_0}k_B^2 T$, последняя пропорциональна полной концентрации электронов в системе и, конечно, линейно возрастает с концентрацией. С другой стороны, согласно принципу Паули, электроны, лежащие по уровнем Ферми, не могут участвовать в отклике системы на внешнее воздействие, в том числе и на термическое возбуждение. Как разрешается это (кажущиеся) противоречие?

4. Чему равна теплоемкость вырожденного электронного газа в квантовой нити при абсолютном нуле температуры?

5. Теплоемкость вырожденного электронного газа однозначно определяется плотностью состояний на уровне Ферми. Исходя из этого, укажите при каких условиях теплоемкость не зависит от концентрации электронов в

   (a) в квантовой нити?

   (b) в квантовой плоскости?

6. Может ли теплоемкость электронного газа не зависеть от концентрации в реальных трехмерных полупроводниках?

7. В электронной системе какой размерности, и при каких условиях теплоемкость может уменьшаться с ростом концентрации электронов!?

8. На основании каких аргументов из линейного характера зависимости удельной теплоемкости от температуры в вырожденном низ-

коразмерном ($d < 3$) электронном газе с независящей от концентрации носителей теплоемкостью может сделан вывод о том, что в согласии с принципом Паули, нагревание меняет энергетическое состояние только электронов вблизи уровня Ферми?

9. Какие наиболее важные с точки зрения расчета теплоемкости особенности отличают фононную подсистему в полупроводнике от электронной?

10. В системе какой размерности и почему может возникать расходимость при подсчете концентрации невзаимодействующих акустических фонов, приводящая к бесконечно высокой их концентрации? Почему этого затруднения не возникает при вычислении концентрации электронов в той же системе?

11. Что позволяет выделить пренебрежимо малый по сравнению с фононами вклад вырожденных электронов в теплоемкость при низких температурах? Какая важная информация может быть извлечена из измерений низкотемпературной теплоемкости?

12. Как видоизменяются выражения для концентрации зонных носителей в присутствии электростатического потенциала?

13. Выполняется ли локально уравнение электронейтральности при неоднородном легировании?

14. Как зависит приповерхностный изгиб зон от координаты в режиме линейного экранирования? Какие параметры описывает эту зависимость в случае вырожденного и невырожденного газа?

15. Выпишите общее выражение для длины экранирования в трехмерном электронном газе, связывающее ее с производной концентрации по химпотенциалу. При каких условиях эта производная эквивалентна плотности состояний на уровне Ферми?

16. Выпишите выражение для длины экранирования Дебая-Хюккеля и укажите при каких условиях она справедлива? Как изменяется

длина Дебая-Хюккеля с ростом температуры. Чему она равна при $T = 0$ и при очень высоких температурах?

17. Выпишите выражение для длины экранирования в трехмерном вырожденном полупроводнике. Как изменяется длина экранирования Томаса-Ферми с температурой.

18. Согласно принципу Паули при вырождении электронного газа носители с энергиями меньшими энергии Ферми не могут участвовать в экранировании электрического поля (как внешнего так и встроенного). С чем в таком случае связана зависимость длины экранирования Томаса-Ферми от полной концентрации электронов в трехмерном электронном газе? При каком (гипотетическом) законе дисперсии длина Томаса-Ферми в трехмерном газа не зависела бы от температуры?

19. В каком случае длина экранирования Томаса-Ферми в двумерном газе не зависит от его концентрации. Как может быть объяснен этот неожиданный результат, указывающий на независимость эффективности экранирования от числа экранирующих свободных зарядов? Справедливо ли это для двумерной длины экранирования Дебая-Хюккеля?

20. Каковы пределы применимости классического рассмотрения экранирования в модели Томаса-Ферми? Какому условию общего вида должна удовлетворять классическая длина Томаса-Ферми? Свяжите это условие с концентрацией носителей и укажите со стороны больших или малых концентраций ограничена область применимости модели Томаса-Ферми?

21. По каким законам спадает потенциал заряженной примеси при учете экранирования в трехмерном и двумерном электронном газе? Дайте качественное объяснение более слабой экранировки кулоновского потенциала в двумерном электронном газе.

22. Приведите условия справедливости квазиклассического рассмотрения экранирования в d -мерном электронном газе в виде

    (a) соотношения между межэлектронным расстоянием и эффективным боровским радиусом;

    (b) соотношения между энергией межэлектронного взаимодействия и кинетической энергией электрона.

23. Как изменяется отношение энергии Ферми к энергии межэлектронного кулоновского взаимодействия при возрастании концентрации и размерности пространства для электронного газа с

    (a) квадратичным законом дисперсии?

    (b) линейным законом дисперсии? электрона.

24. Когда экранирование кулоновского потенциала можно описывать в линейном приближении? Можно ли доверять результату, полученному в приближении линейного экранирования, если энергия отталкивания двух электронов, расстояние между которыми равно полученной при таком расчете длине Томаса-Ферми, превышает энергию Ферми?

25. Если к МДП-структуре приложено напряжение, то в приповерхностной области полупроводника имеется электрическое поле. Если в этой области есть также и свободные носители заряда, то должен возникнуть электрический ток. Между тем, из-за присутствия диэлектрика (изолятора!) постоянного (равновесного) тока в такой структуре протекать не может (все переходные процессы полагаются закончившимися, т.е. рассматривается стационарное состояние). Как разрешается это кажущееся несоответствие?

26. Выпишите (или выведите) соотношение Эйнштейна для коэффициента диффузии и подвижности носителей (зависят ли размерности этих величин от размерности системы носителей?) в его наиболее

общем (справедливом для любого закона дисперсии и любой размерности пространства) виде. Каков физический смысл фигурирующей в этом выражении энергии диффузии?

27. Как изменится отношение коэффициента диффузии к подвижности при уменьшении размерности пространства при прочих равных параметрах в случаях

    (a) невырожденной статистики?

    (b) вырожденной статистики?

28. Зависит ли отношение коэффициента диффузии к подвижности от закона дисперсии для

    (a) вырожденного газа носителей?

    (b) невырожденного газа носителей?

29. Укажите при каких условиях проводимость, обусловленная диффузией носителей, не зависит от их концентрации?

Глава 4

Классическая теория полупроводниковых гетероструктур

4.1 Одиночный гетеропереход

Гетеропереходы представляют собой контакт между двумя веществами, обладающими различными свойствами и являются основой практически всех приборов электроники, в том числе микро – , нано – и оптоэлектроники. Наиболее важные из них:

1. гетеропереходы между различными полупроводниками;

2. гетеропереходы между металлами и полупроводниками, такие как барьеры Шоттки или омические контакты металл-полупроводник;

3. гетеропереходы в структурах металл-диэлектрик-полупроводник (МДП структуры), являющихся основным элементом в современных компьютерных технологиях.

Нас в первую очередь будет интересовать энергетическая структура зон в гетеропереходах с участием полупроводников, являющаяся основой при описании эффектов размерного квантования и двумеризации электронного газа. Структура энергетических зон и ее описание в гетероструктурах несколько сложнее, чем в случае традиционных полупроводниковых

гомо p-n переходов в первую очередь за счет разрывов энергии, соответствующей краям зоны проводимости и валентной зоны. Эти разрывы обусловлены различием в величинах работы выхода и ширины запрещенной зоны контактирующих полупроводников.

Смысл энергетической диаграммы гетероперехода и используемые при ее описании подходы легче всего понять при рассмотрении конкретного гетероперехода (из довольно широкого ряда различных типов полупроводниковых гетеропереходов). В качестве такого примера рассмотрим контакт полупроводника p–типа (левая сторона гетероструктуры) и n-типа (правая сторона) предполагая, что объемные свойства полупроводников сохраняются неизменными вплоть до границы раздела, где имеет место резкий переход от одного материала к другому. Строго говоря, такое предположение справедливо, только если постоянные решетки (и тем более тип кристаллической решетки) контактирующих материалов одинаковы. В противном случае возникающие механические напряжения приводят к изменению параметров зонной структуры вблизи гетероперехода, либо при достаточно большой величине напряжений к образованию дислокационной сетки на границе, также нарушающей однородность свойств. Положим для определенности, что ширина запрещенной зоны материала n типа E_{gn} больше, чем дырочного полупроводника E_{gn}. При дальнейшем рассмотрении будем также предполагать:

1. пренебрежимо малую плотность пограничных состояний в запрещенной зоне;

2. величину разрывов зон не зависящей от легирования;

3. величину скачков краев зон единственной характеристикой данной пары материалов, т. е. не зависящей от связей на границе раздела, деталей расположения атомов внутри границы и кристаллографической ориентации.

В рамках такого рассмотрения зонная диаграмма гетероперехода полностью определяется, исходя из зонной структуры массивных материалов, составляющих гетеропару. До установления контакта двух различных

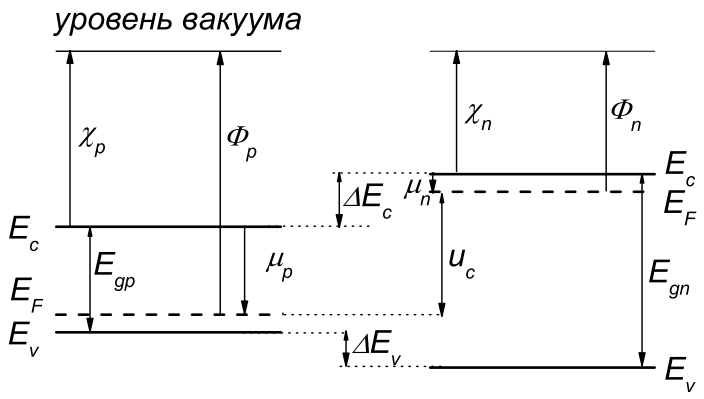

Рис. 4.1: Зонные диаграммы полупроводниковой $p-n$ гетеропары до установления контакта.

полубесконечных полупроводников расположение зон имеет вид, показанный на рис. 4.1 в предположении, что уровень Ферми на свободной поверхности совпадает с его положением в объеме (случай плоских зон). Каждый полупроводник характеризуется тремя энергетическими уровнями: краем зоны проводимости E_c, потолком валентной зоны E_v (или, что эквивалентно, шириной запрещенной зоны $E_g = E_c - E_v$) и уровнем Ферми E_F. Для определения относительно положения энергетических уровней в разных материалах, составляющих гетероструктуру (положения дна какой-либо зоны или уровней Ферми – остальные уровни при известных зонной структуре и уровне легирования определяется однозначно) мы можем исходить из работы, совершаемой при переносе электрона с одного уровня на другой, либо разности работ совершаемых при его переносе на какой-либо третий уровень. В случае полупроводников и металлов в качестве такой характеристики чаще всего используется работа выхода Ф– работа, совершаемая при переносе электрона с уровня Ферми через поверхность, ограничивающую объем, на уровень энергии свободного пространства достаточно далеко за пределами материала (уровень вакуума). Работа выхода, однако, зависит от положения уровня Ферми и, следовательно, меняется при легировании (для невырожденного полупроводника эта величина в некотором отношении фиктивна, поскольку уровень Ферми в этом случае расположен в запрещенной зоне, где электронные состояния в общем случае отсутствуют). Более естественно

использовать электронное сродство χ, определяемое как работа по переносу электрона со дна зоны проводимости на уровень вакуума. Эта величина представляет собой характеристику материала, не зависящую от уровня и типа легирования или ширины запрещенной зоны. Положим, что для рассматриваемой структуры $\chi_p > \chi_n$, т.е . до приведения в контакт зона проводимости полупроводника p-типа лежит ниже дна зоны проводимости материала n –типа.

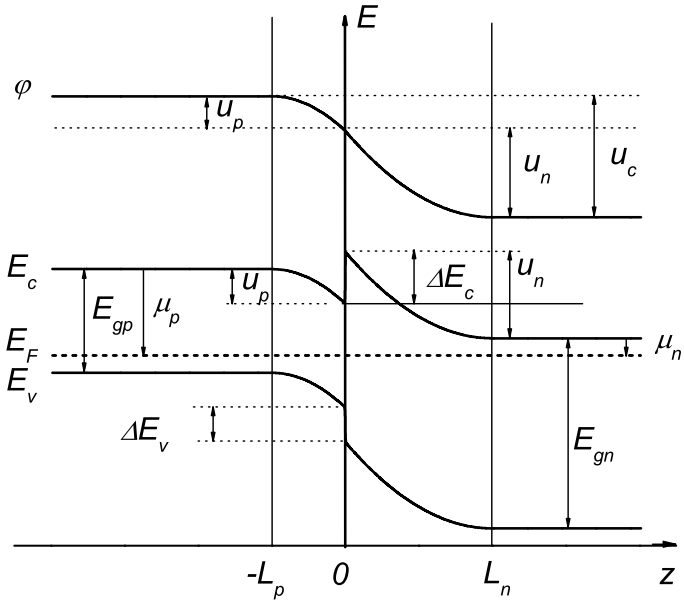

Рис. 4.2: Зонная диаграмма полупроводникового $p-n$ гетероперехода в модели Шокли – Андерсона.

При формировании гетероперехода при контакте полупроводников часть свободных (подвижных) носителей из-за градиента их концентраций на гетерогранице диффундируют в соседний материал. При этом оставшийся нескомпенсированным заряд неподвижных примесей (отрицательно заряженных акцепторов на p-стороне гетероперехода и положительно заряженных доноров на n-стороне) образует двойной слой объемного заряда. Перетекание носителей будет происходить до тех пор, пока возникающее при таком переносе контактное электрическое поле двойного слоя не уравновесит оба потока, т.е. пока уровни Ферми в двух материалах не выровняются, что соответствует установлению термодинамического равновесия в системе (рис. 4.2). Возникающий при выравнивании уровней Ферми изгиб зон приводит к появлению потенциаль-

ных барьеров, которые и препятствуют дальнейшему уходу свободных носителей из контактирующих материалов. Величина изгиба зон равна контактной разности потенциалов - разности работ выхода

$$u_c = \Phi_p - \Phi_n. \tag{4.1}$$

Из-за разности ширин запрещенной зоны, как видно из рис. 4.2, края зон (в отличие от гомо $p-n$ перехода) испытывают разрывы, величины которых легко определяется из простых геометрических соображений

$$\Delta E_c = \chi_p - \chi_n \tag{4.2}$$

$$\Delta E_v = E_{gp} - E_{gn} - (\chi_p - \chi_n) = \Delta E_g - \Delta E_c \tag{4.3}$$

Ясно, что в зависимости от конкретных зонных параметров контактирующих полупроводников величина скачка зон может отличаться не только по величине, но и по знаку (для большинства полупроводниковых пар, образующих достаточно качественные гетеропереходы, разрыв валентной зоны, как правило, превышает скачек в зоне проводимости). В рассматриваемом нами модельном гетеропереходе взяты такие параметры гетеропары, что знаки разрыва зон противоположны для валентной зоны и зоны проводимости. Из рис. 4.2 видно, что это приводит к качественным отличиям в зонной диаграмме для электронов и дырок. Если в валентной зоне наличие разрыва зон приводит просто к увеличению величины потенциального барьера для дырок по сравнению с гомопереходом при тех же параметрах легирования ($u_c + \Delta E_v$ вместо u_c в гомо $p-n$ переходе), то в зоне проводимости наряду с уменьшением потенциального барьера для электронов в глубине полупроводника ($u_c \to u_c - \Delta E_c$) вблизи гетерограницы возникает “пичек” в потенциале. Такой профиль потенциала представляет особый интерес, поскольку при этом на p - стороне перехода образуется потенциальная яма, и следовательно можно ожидать пространственного квантования электронного спектра при достаточно малых ее размерах.

4.1.1 Зонная диаграмма одиночного гетероперехода (модель Шокли–Андерсона)

При заданных зонных параметрах контактирующих полупроводников и уровнях легирования, распределение потенциала в условиях термодинамического равновесия дается уравнением Пуассона

$$\frac{d^2\varphi}{dz^2} = \frac{4\pi e^2}{\epsilon}\varrho_n$$

с величиной объемного заряда $\varrho_n/|e|$, определяемой выражением (3.51). Граничные условия в случае неограниченного образца имеют вид

$$\varphi(-\infty) = \varphi_p, \quad \varphi(\infty) = \varphi_n, \tag{4.4}$$

$$\frac{d\varphi}{dz}(-\infty) = 0, \quad \frac{d\varphi}{dz}(\infty) = 0. \tag{4.5}$$

Равенство нулю $d\varphi/dz$ отражает отсутствие электрического поля $E = d\varphi/edz$ в глубине полупроводника ввиду его экранирования свободными носителями заряда в проводящей системе. Для системы, состоящей из двух разнородных материалов, эти условия должны быть дополнены условиями на границе их контакта друг с другом: непрерывностью потенциала и непрерывностью электрической индукции (при отличающихся диэлектрических проницаемостях электрическое поле $d\varphi/edz$ испытывает разрыв на гетерогранице)

$$\varphi(-0) = \varphi(+0) = \varphi(0) = \varphi_0, \tag{4.6}$$

$$\epsilon_p \left.\frac{d\varphi}{dz}\right|_{z=-0} = \epsilon_n \left.\frac{d\varphi}{dz}\right|_{z=+0}. \tag{4.7}$$

В дальнейшем мы будем предполагать однородное легирование как p-, так n-области гетероструктуры

$$N_A(z) = const = N_A,$$

$$N_D(z) = const = N_D.$$

В этом случае правая часть уравнения Пуассона не содержит явно аргумента и является только функцией химпотенциала. В условиях термодинамического равновесия последний, в свою очередь, однозначно определяется электрическим потенциалом, поскольку в отсутствие тока $\mu_F(z)+\varphi(z) = const$. Из энергетической диаграммы зон на рис. 4.2 легко видеть, что для левой области гетероструктуры $\mu(z) = \mu_{pb} - (\varphi_p - \varphi(z))$, где $\mu_{pb} = \mu(-\infty)$ химпотенциал в объеме -области. Аналогично для n области $\mu(z) = \mu_{bn} + (\varphi(z) - \varphi_n)$, $\mu_{nb} = \mu(\infty)$. Индексы p и n здесь и далее соответствуют принадлежности соответствующей величины p и n областям гетероструктуры, а индекс b значениям соответствующих величин в квазинейтральной области (на внешних границах структуры, при $z = \mp\infty$ для неограниченного образца). В отличие от электрического потенциала химпотенциал испытывает разрыв на самом гетеропереходе равный скачку зоны проводимости ΔE_c.

Наряду с однородностью легирования в дальнейшем мы будем также предполагать, что как в квазинейтральных областях, так и во всей ОПЗ примеси полностью ионизованы

$$N_A^-(z) = const = N_A$$

$$N_D^+(z) = const = N_D.$$

Это означает, что уровень Ферми на обеих сторонах структуры лежит, по крайней мере, на несколько kT ниже донорных уровней, но выше акцепторных ($E_A(z) + 3kT < \mu_F < E_D - 3kT$). Поскольку уровни примесных центров лежат в запрещенной зоне, это предположение автоматически обеспечивает справедливость использования статистики Больцмана при вычислении локальных концентраций зонных носителей.

Из диаграммы на рис. 4.2 видно, что при приближении к гетеропереходу дно валентной зоны в $p-$ области и дно зоны проводимости в $n-$ области удаляются от уровня Ферми, т.е. концентрации основных носителей в этих областях уменьшаются. Для статистики Больцмана такое уменьшение носит экспоненциальный характер

$$p_p(z) = e^{-\frac{E_{gp}+\mu}{kT}} = e^{-\frac{E_{gp}+\mu_{pb}}{kT}} e^{-\frac{\mu-\mu_{pb}}{kT}} = p_{pb} e^{\frac{\varphi-\varphi_p}{kT}}$$

$$n_n(z) = e^{\frac{\mu}{kT}} = e^{\frac{\mu_{pb}}{kT}} e^{\frac{\mu-\mu_{nb}}{kT}} = n_{nb} e^{\frac{\varphi_n-\varphi}{kT}},$$

где p_{pb} и n_{nb} – концентрации основных носителей в квазинейтральных областях. Поскольку все примеси ионизованы, то из условий электронейтральности в объеме следует, что $p_{pb} = N_A$, $n_{nb} = N_D$ и эти выражения могут быть записаны в виде

$$p_p(\varphi) = N_A \exp -\frac{\varphi_p - \varphi}{kT}, \tag{4.8}$$

$$n_n(\varphi) = N_D \exp -\frac{\varphi - \varphi_n}{kT}. \tag{4.9}$$

Таким образом, в той части приконтактной области, где выполнены условия

$$\varphi_p - \varphi, \quad \varphi - \varphi_n > 3kT_b, \tag{4.10}$$

концентрации примесей более чем на порядок превышают концентрацию основных носителей и вкладом последних в объемный заряд можно пренебречь.

С другой стороны, концентрации неосновных носителей n_p и p_n при приближении к гетеропереходу возрастают

$$n_p(z) = n_{pb} \exp \frac{\varphi_p - \varphi}{kT} = p_{pb}\lambda_p^{-2} \exp \frac{\varphi_p - \varphi}{kT} = N_A \lambda_p^{-2} \exp \frac{\varphi_p - \varphi}{kT}, \tag{4.11}$$

$$p_n(z) = p_{nb} \exp \frac{\varphi - \varphi_n}{kT} = n_{nb}\lambda_n^{2} \exp \frac{\varphi - \varphi_n}{kT} = N_D \lambda_n^{2} \exp \frac{\varphi - \varphi_n}{kT}, \tag{4.12}$$

где использовано соотношение (2.108), связывающее концентрации основных и неосновных носителей в нейтральной области полупроводника. Поскольку своего максимального значения концентрации неосновных носителей достигают на самой гетерогранице $z = 0$, то нетрудно убедиться, что в структурах, где выполнены условия

$$\begin{aligned} \varphi_p - \varphi_0 &< kT(2\ln\lambda_p + 3), \\ \varphi_0 - \varphi_n &< -kT(2\ln\lambda_n - 3) \end{aligned} \tag{4.13}$$

(достаточно сильно легированные материалы - $\lambda_p >> 1$, $\lambda_n << 1$, и невысокие парциальные контактные разности потенциалов u_p, и u_n) вклад неосновных носителей в объемный заряд также несущественен.

Таким образом, при выполнении условий (4.10), (4.13) в рамках рассматриваемой модели распределение потенциала в ОПЗ описывается уравнениями

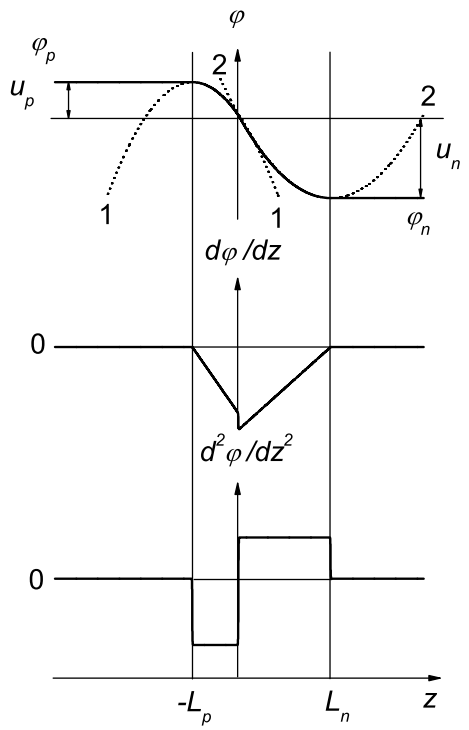

Рис. 4.3: Пространственное распределение потенциала φ, его первой и второй производной в $p-n$ гетеропереходе в модели Шокли - Андерсона. Кривые 1 и 2 соответствуют зависимостям (4.30) $\varphi = \varphi_p - u_p(1 + z/L_p)^2$ и (4.31) $\varphi = \varphi_n + u_n(1 - z/L_n)^2$.

$$\frac{d^2\varphi}{dz^2} = -\frac{4\pi e^2 N_A}{\epsilon_p} \quad (z < 0), \tag{4.14}$$

$$\frac{d^2\varphi}{dz^2} = \frac{4\pi e^2 N_D}{\epsilon_p} \quad (z > 0) \tag{4.15}$$

для p и n областей структуры соответственно. Используя формулу для первого интеграла уравнения Пуассона (3.53) (для рассматриваемого случая функция $f(\varphi)$ есть константа, а $z_0 = \mp\infty$) получаем для p -области

$$\frac{d\varphi}{dz} = -\sqrt{\frac{8\pi e^2 N_A}{\epsilon_p}(\varphi_p - \varphi)}. \tag{4.16}$$

Аналогично для n-области структуры

$$\frac{d\varphi}{dz} = -\sqrt{\frac{8\pi e^2 N_D}{\epsilon_n}(\varphi_n - \varphi)}. \tag{4.17}$$

Поскольку $\varphi(z)$ является всюду убывающей функцией z, то $d\varphi/dz < 0$ – рис. 4.3, что и определяет выбор знака перед корнями. Из условия непрерывности электрической индукции при $z = 0$ и учитывая, что

$$\begin{aligned} \varphi_p - \varphi_0 &= u_p, \\ \varphi_0 - \varphi_n &= u_n, \end{aligned} \tag{4.18}$$

приходим к следующему соотношению для величин u_p и u_n

$$\frac{u_p}{u_n} = \frac{\epsilon_n N_D}{\epsilon_p N_A}, \tag{4.19}$$

показывающему, что на материал с меньшим уровнем легирования приходится большая часть полной контактной разности потенциалов. Совместно с равенством

$$u_c = u_p + u_n \tag{4.20}$$

соотношение (4.19) определяет величины u_p и u_n

$$u_p = u_c\left(1 + \frac{\epsilon_p N_A}{\epsilon_n N_D}\right)^{-1}, \quad u_n = u_c\left(1 + \frac{\epsilon_n N_D}{\epsilon_p N_A}\right)^{-1} \tag{4.21}$$

Таким образом, при известном значении полной контактной разности потенциалов энергетические параметры u_p и u_n полностью определяются параметрами легирования и величинами диэлектрических проницаемостей.

Используя далее (3.54) с граничными условиями при $z_1 = \mp\infty$ ($\varphi(-\infty) = \varphi_p$, $\varphi(+\infty) = \varphi_n$) приходим к следующим уравнениям

$$\int\limits_{\varphi_p}^{\varphi} \frac{d\varphi}{\sqrt{\varphi_p - \varphi}} = -\sqrt{\frac{8\pi e^2 N_A}{\epsilon_p}} \int\limits_{-\infty}^{z} dz, \tag{4.22}$$

$$\int\limits_{\varphi_n}^{\varphi} \frac{d\varphi}{\sqrt{\varphi - \varphi_n}} = -\sqrt{\frac{8\pi e^2 N_D}{\epsilon_n}} \int\limits_{\infty}^{z} dz, \tag{4.23}$$

описывающим пространственное распределение потенциала соответственно в левой и правой сторонах гетероперехода.

Нетрудно, однако, видеть, что интегралы слева конечны, тогда как справа расходится для бесконечного образца или принимают произвольное значение, определяемое выбором размеров контактирующих полупроводников. Между тем очевидно, что ввиду экранирования контактного поля ход потенциала не должен зависеть от размеров образца при достаточно большой его толщине. Причина отмеченного противоречия очевидна и заключается в том, что уравнения (4.14) и (4.15) и вытекающее из них выражения для электрического поля (4.16) и (4.17) справедливы только в области, где можно пренебречь вкладом заряда свободных носителей. Последнее предположение, очевидно, не выполняется в квазинейтральных областях в глубине полупроводника ($z \to \mp\infty$), где $p_{pb} \approx N_A$, $n_{nb} \approx N_D$ и объемный заряд близок к нулю. Чтобы обойти эту трудность заменим граничные условия (4.4), (4.5) на

$$\varphi(-L_p) = \varphi_p, \quad \varphi(L_n) = \varphi_n, \tag{4.24}$$

$$\left.\frac{d\varphi}{dz}\right|_{-L_p} = 0, \quad \left.\frac{d\varphi}{dz}\right|_{L_n} = 0, \tag{4.25}$$

где L_p и L_n соответствуют максимальным расстояниям от гетероперехода, где еще выполняются неравенства $p_p << N_A$ и $n_n << N_D$, т.е. L_p и L_n определяются приближенными соотношениями $\varphi_p - \varphi(L_p) \approx 3kT$ и $\varphi(L) - \varphi \approx 3kT$. Ограничиваясь рассмотрением этих областей, т.е. заменяя нижние пределы интегрирования по z в (4.22) и (4.23) на $-L_p$ и L_n приходим к следующим выражениям для потенциала в p и n областях слоя объемного заряда гетероперехода

$$\varphi = \varphi_p - \frac{2\pi e^2 N_A L_p^2}{\epsilon_p}\left(1 + \frac{z}{L_p}\right)^2 \quad (z \le 0) \tag{4.26}$$

$$\varphi = \varphi_n + \frac{2\pi e^2 N_D L_n^2}{\epsilon_n}\left(1 - \frac{z}{L_n}\right)^2 \quad (z \ge 0) \tag{4.27}$$

Воспользуемся теперь непрерывностью потенциала на гетерогранице

$$\varphi(-0) = \varphi(+0) = \varphi_0$$

(в общем случае $\epsilon_p \neq \epsilon_n$ зависимость $\varphi(z)$, будучи непрерывной, имеет при $z = 0$ излом – см. рис. 4.3). Записывая выражения (4.26) и (4.27) для $z = 0$ и учитывая (4.18) находим выражения для изгибов зон приходящихся на p – и n – области гетероперехода

$$u_p = \frac{2\pi e^2 N_A L_p^2}{\epsilon_p} \tag{4.28}$$

$$u_n = \frac{2\pi e^2 N_D L_n^2}{\epsilon_n} \tag{4.29}$$

С учетом этого, зависимость $\varphi(z)$ может быть записана в простом виде

$$\frac{\varphi_p - \varphi}{u_p} = \left(1 + \frac{z}{L_p}\right)^2 \quad (z \leq 0) \tag{4.30}$$

$$\frac{\varphi - \varphi_n}{u_n} = \left(1 - \frac{z}{L_n}\right)^2 \quad (z \geq 0), \tag{4.31}$$

содержащем два (для каждой стороны гетероперехода) характерных параметра: толщину области пространственного заряда L_p (L_n) и падение контактной разности потенциала u_p (u_n), приходящейся на p (n) область. Для определения характерных длин L_p и L_n воспользуемся соотношениями (4.28), (4.29), из которых следует

$$\frac{u_p}{u_n} = \frac{\epsilon_n N_A L_p^2}{\epsilon_p N_D L_n^2}, \tag{4.32}$$

что совместно с (4.19) дает

$$\frac{L_p}{L_n} = \frac{N_D}{N_A}. \tag{4.33}$$

Из этого выражения следует, что контактное поле проникает на бóльшую глубину в ту область гетероструктуры, где концентрация легирующей примеси меньше. Нетрудно видеть, что соотношение (4.33) является следствием электронейтральности гетероперехода в целом: заряд акцепторов слева $-|e|\, L_p N_A$ и доноров справа $+|e|\, L_n N_D$ от гетероперехода полностью компенсируют друг друга. Подставляя в выражение для полной контактной разности потенциалов (4.20) формулы (4.28), (4.29) для u_p и u_n , выражая из (4.33) L_n через L_p и разрешая уравнение относительно L_p получаем:

$$L_p = \sqrt{\frac{\epsilon_p \epsilon_n u_c}{2\pi e^2 \left(\epsilon_p N_A + \epsilon_n N_D\right)} \frac{N_D}{N_A}}. \tag{4.34}$$

До сих пор мы рассматривали гетеропереход в отсутствии приложенного к системе внешнего напряжения. При этом параметры, описывающие потенциал в ОПЗ, как видно из выражений (4.34), (4.33) и (4.29) полностью определяются величиной контактной разности потенциалов u_c, т.е. разностью электрических потенциалов в p – и n – областях $u_c = \varphi(L_p) - \varphi(L_n)$. При приложении к структуре внешней разности потенциалов V энергетическое расстояние между уровнями Ферми в объеме (совпадающих при $V = 0$) изменится на величину eV. Поскольку область объемного заряда обеднена свободными носителями, она (при не слишком большой толщине образца) дает основной вклад в сопротивление структуры и при не очень больших смещениях (меньших высоты барьера u_c/e для прямых смещений и меньших пробойного напряжения для обратных) основная часть внешнего напряжения падает на гетеропереходе. В результате, разность электрических потенциалов (величина полного изгиба зон) станет равной $\varphi(L_p) - \varphi(L_n) = u_c - eV$. Пока приложенное напряжение при положительных значениях V(плюс на p области) остается меньше контактной разности потенциалов объемный заряд в области гетероперехода по-прежнему создается в основном заряженными примесями (предполагается также, что приложенное к структуре напряжение не приводит к перезарядке примесных центров) и использованное нами уравнение Пуассона, как и все вытекающие из него соотношения, остается справедливым с единственной заменой $u_c \to u_c - eV$ (конечно, сами толщины обедненных слоев при этом изменяются).

Таким образом, в общем случае выражение для L_p имеет вид:

$$L_p = \sqrt{\frac{\epsilon_p \epsilon_n (u_c - eV)}{2\pi e^2 \left(\epsilon_p N_A + \epsilon_n N_D\right)} \frac{N_D}{N_A}}, \tag{4.35}$$

откуда приходим к выражениям для L_n :

$$L_n = \sqrt{\frac{\epsilon_p \epsilon_n (u_c - eV)}{2\pi e^2 \left(\epsilon_p N_A + \epsilon_n N_D\right)} \frac{N_A}{N_D}} \tag{4.36}$$

и полной толщины области пространственного заряда $W = L_p + L_n$

$$W = \sqrt{\frac{\epsilon_p \epsilon_n (u_c - eV)}{2\pi e^2 (\epsilon_p N_A + \epsilon_n N_D)} \frac{(N_A + N_D)^2}{N_A N_D}}. \tag{4.37}$$

Из последней формулы видно, что при сильном отличии уровней легирования n и p областей толщина ОПЗ определяется низколегированной областью. Так при $N_A >> N_D$ и $\epsilon_p \approx \epsilon_n \approx \epsilon$

$$W = \sqrt{\frac{\epsilon (u_c - eV)}{2\pi e^2 N_D}}. \tag{4.38}$$

Это, в общем-то, очевидно, поскольку для выполнения условия электронейтральности гетероперехода ($L_p N_A = L_n N_D$) области с более высокой концентрацией примесей соответствует меньшая толщина.

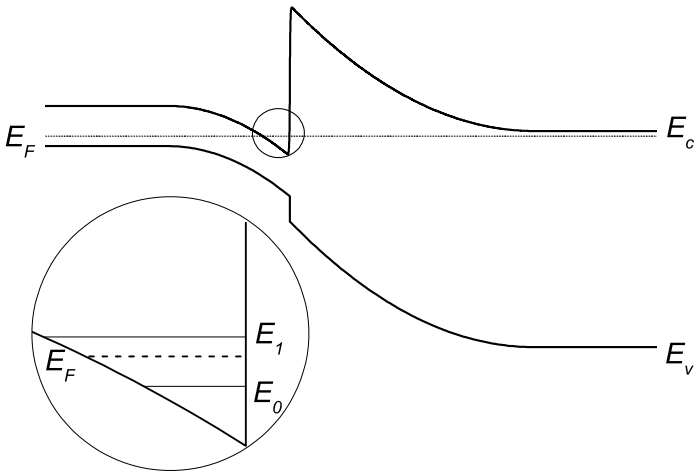

Рис. 4.4: Образование квантовой ямы для электрона в одиночном $p - n$ гетеропереходе. Показаны два первых уровня размерного квантования, нижайший по энергии из которых лежит ниже уровня Ферми, что соответствует заполнению основной двумерной подзоны.

Полученные выражения для потенциала и определяющих его характерных параметров не содержат каких-либо зонных параметров полупроводника (эффективной массы - в общем случае параметров законов дисперсии). Это и неудивительно, поскольку мы полностью пренебрегли вкладом зонных носителей заряда. Само же наличие свободных (подвижных) носителей заряда конечно необходимо, поскольку слой пространственного заряда может возникнуть только за счет ухода из этой

области подвижных электронов и дырок. За исключением диэлектрической проницаемости, эти выражения не содержат и характеристик гетероперехода (ширины запрещенных зон, электронного сродства и т. п), что также неудивительно, поскольку распределение потенциала, описываемого уравнением Пуассона, в принятом приближении, игнорирующем заряд зонных носителей в ОПЗ, определяется только пространственным распределением примесей (в рассмотренной модели однородным – но сказанное остается справедливым и в общем случае зависящих от координат концентраций примесей) и диэлектрическими проницаемостями. Таким образом, зависимость $\varphi(z)$ в рассматриваемом приближении одинакова для гомо – и гетеропереходов. Однако в гетеропереходе потенциальная энергия электронов и дырок в общем случае имеет разрывы из-за разрывов зон. В результате носители вблизи гетероперехода могут оказаться в потенциальной яме (как это имеет место для рассмотренной нами структуры - рис. 4.4) и их движение в направлении нормальном границе раздела оказывается ограниченным с одной стороны барьером в $\varphi(z)$ а с другой барьером, обусловленным скачком дна зоны проводимости (или валентной зоны или той и другой). В случае, если характерный размер такой потенциальной ямы сравним с характерной длиной волны де Бройля носителей (фермиевской длиной волны для вырожденного газа носителей или длиной волны, соответствующей тепловой энергии kT для невырожденных носителей) такая яма является квантовой и движение носителей в ней квантуется.

На представленных на рис. 4.1 и 4.2 зонных диаграммах знаки и величина смещений зон выбраны таким образом, что на границе раздела более узкая запрещенная зона лежит внутри более широкой. Такие гетеропереходы с "охватывающей" конфигурацией называют гетеропереходами I рода. В зависимости от зонных параметров и электронного сродства возможны, однако, и гетеропереходы с иным взаимным расположением зон, показанные на рис. 4.5. В частности, край валентной зоны широкозонного полупроводника может оказаться выше края валентной зоны полупроводника с более узкой запрещенной зоной. Соответствующие этой конфигурации гетеропереходы называют гетеропереходами II

рода.

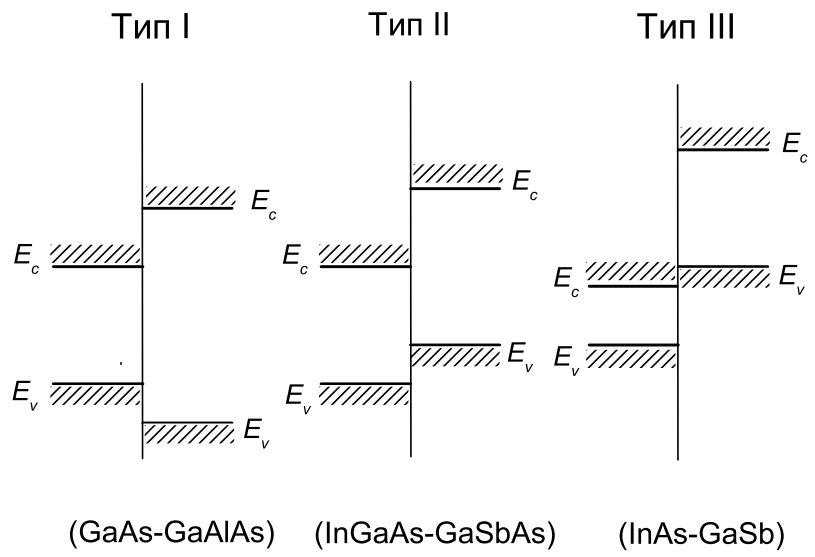

Рис. 4.5: Схематическое изображение расположения зон в гетеропереходах I и II рода. Указаны типичные пары полупроводников, реализующих эти расположения. На показанных здесь малых расстояниях от границы раздела изгибом зон можно пренебречь.

4.1.2 Методы определения параметров гетероструктур

Величина барьера на гетерогранице (величина разрыва зоны проводимости), как сказано выше, определяется разностью электронного сродства полупроводников $\Delta E_c = \chi_n - \chi_p$. К сожалению, точность экспериментального определения электронного сродства и работ выхода не превышает нескольких долей электрон-вольта, так что разность $\chi_n - \chi_p$ известна с точностью не лучше 0,5-1 эВ, тогда как в большинстве случаев рельеф потенциала в квантовых ямах требуется знать с точностью до нескольких мэВ.

Однако при известных значениях контактной разности потенциалов и химпотенциалов μ_{pb} и μ_{nb} в p и n областях структуры (т.е. при известном уровне их легирования) величина разрывов зон на гетерогранице может быть определена из соотношения (это легко видеть из рис. 4.2)

$$u_c - \Delta E_c = \mu_{nb} - \mu_{pb},$$

что дает

$$\Delta E_c = u_c - \mu_{nb} + \mu_{pb}$$

для разрыва зоны проводимости и

$$\Delta E_v = \Delta E_g - u_c + \mu_{nb} - \mu_{pb}$$

для скачка валентной зоны. Ошибка определения контактной разности потенциалов из значений работ выхода входящих в гетеропару материалов $u_c = \Phi_n - \Phi_p$ того же порядка что и для электронного сродства. Однако входящая в эти выражения величина u_c может быть непосредственно определена из измерений электроемкости гетероструктуры, как функции приложенного к переходу внешнего напряжения. Действительно, по определению, дифференциальная удельная емкость гетероперехода (емкость на единицу площади) равна

$$C_s = \frac{dQ_s}{dV} = eN_A \frac{dL_p}{dV} \left(= -eN_D \frac{dL_n}{dV} \right), \tag{4.39}$$

где $Q_s = -eN_A L_p = eN_D L_n$ - поверхностная плотность объемного заряда в ОПЗ. Дифференцируя (4.35) по V и выражая dL_p/dV через полную толщину ОПЗ (4.37) приходим к формуле, аналогичной выражению для емкости плоского конденсатора

$$C = \frac{1}{4\pi W} \frac{\epsilon_p \epsilon_n (N_A + N_D)}{\epsilon_p N_A + \epsilon_n N_D}. \tag{4.40}$$

Для гомоперехода или при близких значениях проницаемостей $\epsilon_p \approx \epsilon_n = \epsilon$ (довольно типичный случай для наиболее важных гетеропереходов) это выражение совпадает с емкостью плоского конденсатора с расстоянием между пластинами W

$$C \approx \frac{\epsilon}{4\pi W}. \tag{4.41}$$

Поскольку, согласно (4.37), $W \propto \sqrt{u_c - eV}$, то величина обратная квадрату емкости является линейной функцией напряжения

$$\frac{1}{C^2} = \frac{8\pi}{e^2} \frac{\epsilon_p N_A + \epsilon_n N_D}{\epsilon_p \epsilon_n N_A N_D} (u_c - eV)$$

с отсечкой на оси напряжений V, равной u_c/e и наклоном

$$\frac{d(1/C^2)}{dV} = -\frac{8\pi}{e}\frac{\epsilon_p N_A + \epsilon_n N_D}{\epsilon_p \epsilon_n N_A N_D},$$

определяемым легированием структуры, что используется для определения концентрации примесей. Нетрудно видеть, что в случае сильной асимметрии легирования наклон зависимости $C^{-2}(V)$ определяется только концентрацией примеси в слаболегированной области перехода. Так, при $N_A >> N_D$ предыдущая формула дает

$$N_D = -\frac{8\pi}{\epsilon_n e^2}\left(\frac{d(1/C^2)}{deV}\right)^{-1}$$

Отметим, что согласно выражению (4.40) измерения емкости непосредственно дают значения характерного размера W области пространственного заряда.

4.2 МДП-структуры

При рассмотрении гетероперехода формирование потенциала в приконтактной области связано с диффузионным переносом свободных носителей, имеющих разные химпотенциалы в двух контактирующих материалах. Такое разделение происходит до тех пор, пока образующееся за счет некомпенсированного заряда примесей внутренне встроенное электрическое поле двойного слоя не уравновесит диффузионные потоки подвижных зарядов. Теперь мы рассмотрим другую ситуацию, а именно, потенциал, формирующийся в полупроводнике при помещении его во внешнее электрическое поле, как это имеет место в гетероструктурах металл-диэлектрик полупроводник. В зависимости от направления приложенного к поверхности полупроводника поля, имеющиеся свободные носители (электроны и дырки в собственном полупроводнике, или только электроны (дырки) в полупроводнике n- (p-) типа), будут притягиваться или отталкиваться от поверхности. Изменение концентрации носителей у поверхности по сравнению с ее значением в объеме приводит к изменению химпотенциала и, как следствие, к появлению диффузионного

потока, обратного омическому току, обусловленному электрическим полем.

4.2.1 Пространственное распределение потенциала

В отличие от рассмотренного в предыдущем разделе гетероперехода вкладом заряда свободных носителей в объемный заряд приповерхностной области МДП – структур не только нельзя пренебречь, но в ряде случаев он дает доминирующий вклад. В этом случае уравнение Пуассона удобнее записать не для электрического потенциала, а для химпотенциала, непосредственно определяющего локальные концентрации зонных носителей, т. е. в форме (3.50) с объемной плотностью заряда (мы ограничимся случаем степенных законных законов дисперсии $\varepsilon = c_r p^r$, параметры которых c_r, p^r в разных зонах могут отличаться)

$$\begin{aligned} \rho_n(\mu) &= N_v \left[\mathcal{F}_{\frac{3}{r_v}-1}\left(\frac{-E_g-\mu}{k_BT}\right) - \mathcal{F}_{\frac{3}{r_v}-1}\left(\frac{-E_g-\mu_b}{k_BT}\right)\right] \\ &- N_c \left[\mathcal{F}_{\frac{3}{r_c}-1}\left(\frac{\mu}{k_BT}\right) - \mathcal{F}_{\frac{3}{r_c}-1}\left(\frac{\mu_b}{k_BT}\right)\right], \end{aligned} \tag{4.42}$$

где согласно(2.68) при $d = 3$

$$N_{c,v} = \frac{\Gamma(3/r)}{r\pi^2\hbar^3}\left(\frac{k_BT}{c_r}\right)^{3/r}, \tag{4.43}$$

а индексы c и v соответствуют состояниям в зоне проводимости и в валентной зоне [1].

Как и в случае полупроводникового гетероперехода мы рассматриваем полупроводник в электрическом поле в условиях термодинамического равновесия, т. е. в предположении, что все переходные процессы, связанные с перезарядкой МДП - конденсатора при изменении внешнего напряжения, закончены (из-за наличия изолирующего слоя диэлектрика постоянный ток в МДП-структуре отсутствует). Как и в разд. 4.1, мы ограничимся рассмотрением однородно легированного полупроводника

[1] Мы пренебрегаем легкими дыркам, в виду их малого вклада в концентрацию носителей валентной зоны.

с полностью ионизованными примесями, когда правая часть уравнения Пуассона не зависит явно от координаты и первый интеграл Пуассона определяется формулой (3.53). С использованием граничного условия в глубине полупроводника ($z_0 = \infty$) $\mu(z_0) = \mu_b,\ d\mu/dz|_{z_0} = 0$ эта формула приводит к выражению

$$\frac{d\mu}{dz} = \sqrt{\frac{8\pi e^2 k_B T}{\epsilon}} \digamma(\mu, \mu_b), \tag{4.44}$$

с функцией $\digamma = \mp(\digamma_c - \digamma_p)^{1/2}$, где

$$\begin{aligned} \digamma_c &= N_c \left[\mathcal{F}_{\frac{3}{r_c}} \left(\frac{\mu}{k_B T} \right) - \mathcal{F}_{\frac{3}{r_c}} \left(\frac{\mu_b}{k_B T} \right) - \frac{\mu - \mu_b}{k_B T} \mathcal{F}_{\frac{3}{r_c}-1} \left(\frac{\mu_b}{k_B T} \right) \right] \\ \digamma_p &= N_v \left[\mathcal{F}_{\frac{3}{r_v}} \left(\frac{-E_g - \mu}{k_B T} \right) - \mathcal{F}_{\frac{3}{r_v}} \left(\frac{-E_g - \mu_b}{k_B T} \right) - \frac{\mu - \mu_b}{k_B T} \mathcal{F}_{\frac{3}{r_v}-1} \left(\frac{-E_g - \mu_b}{k_B T} \right) \right] \end{aligned} \tag{4.45}$$

Функция $\digamma$ должна быть взята со знаком $-$ для положительных изгибов зон $\mu - \mu_b > 0$ (электрическое поле в ОПЗ направлено вглубь полупроводника от его поверхности, создавая потенциальную яму для электронов и барьер для дырок) и со знаком $+$ для $\mu - \mu_b < 0$ (очевидно, что $\digamma(\mu = \mu_b) = 0$). При интегрировании (3.53) использовано вытекающее из (2.66) соотношение для интегралов Ферми

$$\int \mathcal{F}_j(\frac{\mu}{k_B T}) = k_B T \mathcal{F}_{j+1}(\frac{\mu}{k_B T}). \tag{4.46}$$

4.2.2 Поверхностная плотность заряда и дифференциальная емкость ОПЗ

Одной их важных характеристик приповерхностной области полупроводника является поверхностная плотность индуцированного электрическим полем заряда в ОПЗ (заряд в расчете на единицу площади структуры)

$$Q_{sc} = e \int_0^\infty (p(z) + N_D^+ - n(z) - N_A^-) dz = e \int_{\mu_s}^{\mu_b} (p(\mu) + N_D^+ - n(\mu) - N_A^-) d\mu, \tag{4.47}$$

Поскольку локальные концентрации зонных носителей $p(\mu(z)),\ n(\mu(z))$ определяются величиной химпотенциала, то, на первый взгляд, для расчета Q_s необходимо знать ход потенциала в ОПЗ , т. е. требуется решение уравнения Пуассона $\mu = \mu(z)$. Это действительно так, если речь идет о вычислении поверхностных плотностей электронов или дырок в отдельности. Однако, как мы сейчас увидим, для определения поверхностной плотности *полного* индуцированного заряда, как и для расчета некоторых других важных физических характеристик, достаточно знать первый интеграл уравнения Пуассона, т. е. функцию $\digamma(\mu, \mu_b)$. Для определения плотности заряда в ОПЗ полупроводника, заключенного в слое между параллельными поверхности полупроводника плоскостями $z - z_1$ и $z - z_2$, выделим в этом слое цилиндр с площадью основания A, ограниченный перпендикулярными поверхности боковыми гранями и проинтегрируем по объему цилиндра уравнение Гаусса (3.46)

$$\int\limits_V \mathrm{div}\mathbf{E}(\mathbf{r})dV = \frac{4\pi}{\epsilon}\int\limits_V \varrho(\mathbf{r})dV, \tag{4.48}$$

Заменяя первый интеграл по объему от дивергенции поля интегралом по ограничивающей этот объем поверхности от поля $\mathbf{E}$ и учитывая, что интеграл по параллельным полю боковым граням равен нулю, получаем для рассматриваемого случая однородного (в параллельной поверхности плоскости) поля напряженностью $\mathcal{E}$

$$A(\mathcal{E}(z_2) - \mathcal{E}(z_1)) = \frac{4\pi}{\epsilon}Q(z_1, z_2), \tag{4.49}$$

где $Q(z_1, z_2) = \int\limits_V \varrho(\mathbf{r})dV$ - заряд в выделенном объеме. Поскольку $e\mathcal{E}(z) = -d\mu/dz$, то, используя (4.44), получаем

$$Q_s(z_1, z_2) = \frac{Q(z_1, z_2)}{A} = \sqrt{\frac{\epsilon k_B T}{2\pi}}\left[\digamma(\mu(z_1)) - \digamma(\mu(z_2)\right]. \tag{4.50}$$

Для поверхностной плотности полного индуцированного в полупроводнике заряда (заключенного в слое от $z = 0$ до $z = \infty$) отсюда имеем

$$Q_{sc} = \sqrt{\frac{\epsilon k_B T}{2\pi}}\digamma(\mu_s) \tag{4.51}$$

в согласии с известной формулой электростатики для электрического поля заряженной плоскости

$$\mathcal{E}_s = \frac{4\pi}{\epsilon} Q_{sc}). \tag{4.52}$$

Используя это выражение, легко определить одну из наиболее важных физических характеристик ОПЗ, измеряемую экспериментально, - равновесную (низкочастотную) дифференциальную емкость

$$C_{sc} = \frac{dQ_{sc}}{dV_s} = -e\frac{dQ_{sc}}{d\mu_s} = -e\sqrt{\frac{\epsilon k_B T}{2\pi}}\frac{\partial F(\mu_s)}{\partial \mu_s}. \tag{4.53}$$

Легко видеть, что при любом знаке приповерхностного изгиба зон $\mu_s - \mu_b$ производная $\partial F/\partial \mu_s$ отрицательна (электрическое поле вследствие экранировки убывает вглубь полупроводника), так что емкость всегда является величиной положительной.

4.2.3 Режимы обогащения, обеднения и инверсии

При анализе параметров ОПЗ полупроводника в приповерхностных слоях МДП-структур принято различать несколько режимов, соответствующих различным интервалам изгибов зон (ниже для определенности рассматривается материал $n-$типа):

а) Режим плоских зон (или линейного экранирования), соответствующий случаю изгибов зон, при которых концентрация носителей в ОПЗ мало отличается от объемной, что, как нетрудно видеть, удовлетворяется при $|\mu - \mu_b| << kT$ в случае статистики Больцмана и $|\mu - \mu_b| << \mu_b$ при вырождении газа носителей. Приповерхностная область полупроводника в этих условиях была уже рассмотрена в разделе (3.2.3).

б) Режим обогащения, соответствующий изгибам зон, при которых всюду в приповерхностной области выполнено условие $n(\mu(z)) > n_b$.

в) Режим обеднения, соответствующий изгибам зон, при которых концентрации как основных, так и неосновных носителей меньше концентрации ионизованных примесей $(n(z), p(z) < N_D - N_A$ для рассматриваемого материала n-типа). Объемный заряд ОПЗ полупроводника в этом режиме, как и в рассмотренном в предыдущем разделе гетеропереходе,

определяется, в основном, зарядом примесей и соответствующий слой называется слоем обеднения или истощения.

г) Режим инверсии, соответствующий изгибам зон, при которых концентрации неосновных носителей вблизи поверхности превышают концентрацию ионизованных примесей ($p(z) > N_D - N_A$ для материала n-типа). Объемный заряд в инверсионном слое полупроводника определяется зарядом неосновных носителей и примесей. Инверсионный слой всегда отделен от квазинейтральной области объема полупроводника слоем обеднения, толщина которого, как правило (за исключением случая узкощелевых полупроводников), значительно больше толщины слоя инверсии.

В отличие от случая гетероперехода величина изгиба зон, определяемая приложенным к МДП напряжением, ограничена только диэлектрической прочностью диэлектрика и глубина приповерхностной потенциальной ямы может оказаться столь большой, что уровень Ферми в значительной части ОПЗ (а в случае вырожденного полупроводника во всей ОПЗ) может оказаться значительно выше дна потенциальной ямы для электронов (или ниже потолка ямы для дырок при изгибе зон вверх) и электронный (дырочный) газ в этой области будет вырожден. Таким образом, первый интеграл уравнения Пуассона (а следовательно и само распределение поверхностного потенциала $\varphi(z) = E_F - \mu(z)$) в общем случае может быть определен только численными методами. Ниже мы ограничимся двумя предельными ситуациями: (1) невырожденного заполнения зонных состояний (достаточно слабые изгибы зон) и (2) вырожденной статистики носителей в приповерхностной потенциальной яме (сильный изгиб зон).

4.2.4 Классический предел (статистика Больцмана)

Невырожденное заполнение зонных состояний соответствует условию, когда уровень Ферми как в объеме полупроводника, так и всюду в приповерхностной ОПЗ удален от разрешенных зон по крайней мере на несколько k_BT. Хотя выражение для $d\mu/dz$ для невырожденной статистики может быть без труда найдено из общих выражений (4.44) и (4.45)

заменой $\mathcal{F}_j(\mu) = \exp(\mu/k_BT)$, полезно стартовать с исходного уравнения Пуассона, чтобы ввести некоторые широко используемые параметры, характеризующие систему с невырожденным газом носителей в ОПЗ. Для локальной концентрации электронов $n(\mu(z))$ в этом случае имеем

$$n = N_c \exp\frac{\mu}{kT} = N_c \exp\frac{\mu-\mu_b}{kT}\exp\frac{\mu_b}{kT} = n_b \exp y = n_i\lambda^{-1}\exp y,$$

где $\mu_b = \mu(\infty)$ и $n_b = n(\infty) = N_D$ - соответственно химпотенциал и концентрация электронов в объеме полупроводника и введен безразмерный изгиб зон

$$y(z) = \frac{\mu(z)-\mu_b}{kT} \tag{4.54}$$

(для n_b мы воспользовались соотношением (2.109)). Аналогично, для концентрации дырок в тех же обозначениях имеем

$$p = p_i\lambda \exp -y.$$

Из условия нейтральности полупроводника вдали от поверхности $\rho_n(\infty) = 0$ и постоянства концентрации ионизованных примесей следует, что

$$N_D - N_A = p_b - n_b = n_i(\lambda - \lambda^{-1}).$$

Уравнение Пуассона в новых обозначениях принимает вид

$$\frac{d^2y}{dz^2} = -\frac{4\pi e^2 n_i}{\epsilon kT}\left[\lambda(e^{-y}-1) - \lambda^{-1}(e^y-1)\right]. \tag{4.55}$$

Отсюда, используя граничные условия $y(\infty) = 0$, $\left.\frac{dy}{dz}\right|_{z=\infty} = 0$, для первого интеграла уравнения Пуассона имеем

$$\frac{dy}{dz} = \sqrt{\frac{8\pi e^2 n_i}{\epsilon kT}}F(y,\lambda), \tag{4.56}$$

где введена безразмерная функция

$$F(y,\lambda) = \mp\left(\lambda^{-1}(e^y-1) + \lambda(e^{-y}-1) + y(\lambda-\lambda^{-1})\right)^{\frac{1}{2}}. \tag{4.57}$$

Как и в (4.45) знак + в этом выражении соответствует изгибу зон вниз (положительный изгиб зон $y > 0$) и знак − для отрицательного изгиба зон. Из структуры выражения (4.56) следует, что коэффициент при

безразмерной функции $F(y,\lambda)$ имеет размерность обратной длины. Введением характерной длины

$$L_i = \sqrt{\frac{\epsilon kT}{4\pi e^2 n_i}} \tag{4.58}$$

уравнение (4.56) принимает вид

$$\frac{dy}{dz} = \frac{\sqrt{2}}{L_i} F(y,\lambda). \tag{4.59}$$

Интегрируя это уравнение в пределах от y_s до y по y и от 0 до z по z с использованием граничного условия на поверхности полупроводника $y(0) = y_s = (\mu_s - \mu_b)/kT)$ окончательно приходим к следующему выражению, описывающему распределение приповерхностного потенциала в полупроводнике (в виде обратной функции $z(y)$)

$$z(y) = \frac{L_i}{\sqrt{2}} \int\limits_{y_s}^{y} \frac{dy}{F(y,\lambda)}.$$

Выражение для электрического поля на поверхности полупроводника в рассматриваемом приближении невырожденной статистики, как функция безразмерного поверхностного потенциала y_s

$$\mathcal{E}_s = \mathcal{E}(z=0) = -\left.\frac{\partial V}{\partial z}\right|_{V=V_s} = e\left.\frac{\partial \mu_s}{\partial z}\right|_{\mu=\mu_s} = \frac{e}{k_B T}\left.\frac{\partial y}{\partial z}\right|_{y=y_s}$$

непосредственно дается формулой (4.59)

$$\mathcal{E}_s = \frac{e}{k_B T}\frac{\sqrt{2}}{L_i} F(y_s)$$

Отсюда для поверхностной плотности заряда в ОПЗ (4.52) получаем

$$Q_{sc} = Q_{i0}\sqrt{2} F(y_s), \tag{4.60}$$

где введена независящая от изгиба зон масштабная поверхностная плотность заряда, определяемая параметрами полупроводника и температурой

$$Q_{i0} = \frac{e}{k_B T}\frac{1}{L_i} = e n_i L_i.$$

Дифференцируя (4.60) по поверхностному изгибу зон находим дифференциальную емкость ОПЗ

$$\begin{aligned} C_{sc}(y_s) &= -\frac{e}{k_BT}\frac{\partial Q_{sc}}{dy_s} = -C_{i0}\sqrt{2}\frac{\partial F(y_s)}{\partial y_s} \\ &= -C_{i0}\frac{\lambda^{-1}\exp y_s - \lambda\exp -y_s + (\lambda - \lambda^{-1})}{F(y_s)}, \end{aligned} \tag{4.61}$$

где масштабная емкость

$$C_{i0} = \frac{\epsilon}{4\pi L_i}.$$

В случае полупроводника n – типа $\lambda << 1$ и выражение (4.57) для функции $F(y)$ может быть приближенно записано в виде

$$F(y,\lambda) \approx \lambda^{-\frac{1}{2}}F_n, \tag{4.62}$$

где

$$F_n(y,\lambda) = \mp\left(e^y + \lambda^2 e^{-y} - (y+1)\right)^{\frac{1}{2}}. \tag{4.63}$$

Уравнение (4.59) запишется в виде

$$\frac{dy}{dz} = \frac{\sqrt{2}}{L_n}F_n(y,\lambda), \tag{4.64}$$

где введена длина экранирования в полупроводнике n – типа

$$L_n = \sqrt{\lambda}L_i = \sqrt{\frac{\epsilon kT}{4\pi e^2 n_b}} \tag{4.65}$$

(здесь использовано равенство $n_b = \lambda^{-1}n_i$). Нетрудно видеть, что это выражение совпадает с введенной в разд. 3.2.3 длиной Дебая–Хюккеля (3.67), характеризующей линейное экранирование в невырожденном электронном газе. Линейное экранирование соответствует слабому изменению концентрации носителей во внешнем электрическом поле $n(z) - n_b << n_b$ и, как легко видеть, описывается уравнением (4.64) в предположении $|y(z)| << 1$ (режим близких к плоским зон). При этом условии выражение (4.63) для функции F_n с точностью до членов второго порядка (в первом порядке по y $F_n \equiv 0$) принимает вид

$$F_n(y,\lambda) = \mp\left(e^y - y - 1)\right)^{\frac{1}{2}} \approx \mp\left(1 + y + \frac{y^2}{2} - y - 1)\right)^{\frac{1}{2}} = \mp\frac{y}{\sqrt{2}}$$

и уравнение (4.64) для первой производной сводится к

$$\frac{dy}{dz} = \mp \frac{y}{L_n},$$

где для положительных изгибов зон (слабое обогащение приповерхностной области электронами $0 < y << 1$) в правой стороне следует положить знак - (y уменьшается вглубь полупроводника). Интегрирование этого уравнения

$$\int\limits_{y_s}^{y} \frac{dy}{y} = -\frac{1}{L_n} \int\limits_{0}^{z} dz = -\frac{z}{L_n}$$

приводит к экспоненциально спадающему решению

$$y(z) = y_s \exp -\frac{z}{L_n}, \tag{4.66}$$

что, учитывая (4.54), согласуется с решением (3.61) при $L_{csr} = L_{DH} = L_n$. При интегрировании уравнения (4.64) для отрицательных изгибов зон (со знаком + в правой части - слабое обеднение приповерхностной области электронами $-1 << y < 0$) необходимо сделать замену переменной $y \to -y$ (чтобы первообразная $\ln y$ имела смысл при $y < 0$), что приводит к тому же решению (4.54).

Из выражения (4.57) видно, что функция $F(y, \lambda)$ инвариантна про отношению к замене $y \to -y$, $\lambda \to \lambda^{-1}$. Таким образом, область пространственного заряда дырочного полупроводника ($\lambda >> 1$) описывается теми же соотношениями, что и ОПЗ материала n – типа, но с обратным по знаку изгибом зон. Так, уравнение (4.59) для дырочного полупроводника запишется в виде

$$\frac{dy}{dz} = \frac{\sqrt{2}}{L_p} F_p(y, \lambda), \tag{4.67}$$

с длиной экранирования в полупроводнике p – типа

$$L_p = \sqrt{\lambda^{-1}} L_i = \sqrt{\frac{\epsilon k T}{4\pi e^2 p_b}} \tag{4.68}$$

и функцией

$$F_p(y, \lambda) = \mp \left(e^{-y} + \lambda^{-2} e^{y} - (1 - y)\right)^{\frac{1}{2}}. \tag{4.69}$$

4.2.5 ОПЗ с вырожденным электронным газом.

С точки зрения исследования низкоразмерных систем на основе приповерхностных слоев в МДП - структурах наибольший интерес представляет режим, соответствующий вырождению электронного газа в ОПЗ полупроводника. Мы ограничимся рассмотрением случая обогащающих изгибов зон в полупроводнике n - типа (см. зонную диаграмму на правой панели рис. 3.1). Для простоты выкладок предполагается, что вырожденная статистика для электронов справедлива как в объеме, так и во всей приповерхностной ОПЗ полупроводника, а зарядом дырок можно пренебречь. Объемный заряд в этом случае определяется концентрацией электронов и доноров, которые предполагаются полностью ионизованными

$$\varrho_n(\mu) = N_D^+ - n(\mu) = n(\mu_b) - n(\mu), \tag{4.70}$$

где использовано условие электронейтральности в объеме полупроводника $n(+\infty) = n(\mu_b) = n_b = N_D^+ = N_D$. Как и ранее мы рассмотрим случай полупроводника с квадратичным и линейным законом дисперсии. Зависимость электронной концентрации от химпотенциала в этих случаях определятся соотношениями (табл. 2.3)

$$n(\mu) = \frac{(2m)^{\frac{3}{2}}}{3\pi^2\hbar^3}\mu^{\frac{3}{2}} \qquad (\varepsilon = \frac{p^2}{2m}), \tag{4.71}$$

$$n(\mu) = \frac{1}{3\pi^2\hbar^3 s^3}\mu^3 \qquad (\varepsilon = ps), \tag{4.72}$$

которые удобно для дальнейшего записать в одинаковых переменных, выразив эффективную массу в (4.71) через ширину запрещенной зоны и кейновскую скорость

$$2ms^2 = E_g \tag{4.73}$$

(это соотношение между m и E_g с близкими значениями s справедливо для широкого класса представляющих интерес полупроводников - см. табл. 4.1).

Используя граничные условия $\mu(\infty) = \mu_b$ и $\partial\mu/\partial z|_\infty = 0$ (электрическое поле в объеме полупроводника полностью экранировано свободными носителями заряда) для первого интеграла уравнения Пуассона

Таблица 4.1: Ширина запрещенной зоны E_g, эффективная масса на дне зоны проводимости m_n (в единицах массы свободного электрона m_0), χ- диэлектрическая постоянная, кейновская скорость $s = E_g/2m_n$, модифицированная постоянная тонкой структуры $\alpha = e^2/\chi s\hbar$ и параметр $\beta = \sqrt{8\pi\alpha/3}$ для некоторых соединений A_3B_5 и A_2B_6.

	E_g, мэВ	m_n/m_0	χ	$s \cdot 10^{-8}$, см/с	α	β
GaAs	+1520	0,067	12,5	1,410	0,123	1,02
InAs	+418	0,024	14,9	1,238	0,119	1,00
InSb	+237	0,0138	17,9	1,229	0,099	0,91
HgCdTe(x=0,22)	+104	0,0087	17,9	1,025	0,120	1,00
HgCdTe(x=0,22)	+104	0,0087	17,9	1,025	0,120	1,00
HgTe	-303	0,031	20,3	0,927	0,116	0,99

(3.53) получаем

$$\frac{d\mu}{dz} = -\sqrt{\frac{8\pi e^2}{\epsilon}\frac{(2m)^{3/2}}{3\pi^2\hbar^3}}\left[\int\limits_{\mu_b}^{\mu}\mu^{\frac{3}{2}}d\mu\right]^{\frac{1}{2}} = -\sqrt{\frac{8\pi e^2}{3\epsilon s\hbar}\frac{E_g^{3/2}}{\pi^2\hbar^2 s^2}\frac{2}{5}}\left(\mu^{\frac{5}{2}} - \mu_b^{\frac{5}{2}}\right)^{\frac{1}{2}} \tag{4.74}$$

для полупроводника с квадратичным электронным спектром и

$$\frac{d\mu}{dz} = -\sqrt{\frac{8\pi e^2}{\epsilon}\frac{1}{3\pi^2\hbar^3 s^3}}\left[\int\limits_{\mu_b}^{\mu}\mu^3 d\mu\right]^{\frac{1}{2}} = -\sqrt{\frac{8\pi e^2}{3\epsilon s\hbar}}\frac{1}{s\hbar\pi}\frac{1}{2}\left(\mu^4 - \mu_b^4\right)^{\frac{1}{2}} \tag{4.75}$$

для материала со спектром $\varepsilon = ps$. Для простоты выкладок далее предполагается слабое легирование подложки, так что можно положить $n_b \approx$

0, $\mu_b \approx 0$.[2] Нетрудно видеть, что входящая в (4.74), (4.75) величина

$$\alpha' = \frac{e^2}{\epsilon s \hbar} \tag{4.76}$$

является аналогом постоянной тонкой структуры $\alpha = e^2/c\hbar$ с точностью до замены скорости света на кейновскую скорость и заряда e на его перенормированную в среде с диэлектрической проницаемостью ϵ величину $e/\sqrt{\epsilon}$. [3] Как видно из таблицы 4.1, численные значения α' практически одинаковы для представляющих наибольший интерес полупроводниковых соединений A_3B_5 и A_2B_6 (и на порядок величины выше α). Введением безразмерного параметра β

$$\beta = \sqrt{\frac{8\pi\alpha'}{3}} = \sqrt{\frac{8\pi e^2}{3\epsilon s \hbar}} \tag{4.77}$$

(с хорошей точностью можно полагать $\beta = 1$ - см. табл. 4.1) выражения (4.74) и (4.75) при $\mu_b \to 0$ принимают вид

$$\frac{d\mu}{dz} = -\frac{\beta}{\pi\hbar s}\sqrt{\frac{2}{5}} E_g^{\frac{3}{4}} \mu^{\frac{5}{4}} \qquad (\varepsilon = \frac{p^2}{2m}), \tag{4.78}$$

$$\frac{d\mu}{dz} = -\frac{\beta}{2\pi s\hbar}\mu^2 \qquad (\varepsilon = ps), \tag{4.79}$$

Интегрирование первого из этих уравнений с использованием граничного условия на поверхности полупроводника $\mu(z=0) = \mu_s$

$$\int_{\mu_s}^{\mu} \frac{d\mu}{\mu^{5/4}} = -\frac{\beta}{\pi\hbar s}\sqrt{\frac{2}{5}} E_g^{\frac{3}{4}} \int_0^z dz$$

приводит к соотношению

$$\frac{4}{\mu^{1/4}} - \frac{4}{\mu_s^{1/4}} = \frac{\beta}{\pi\hbar s}\sqrt{\frac{2}{5}} E_g^{\frac{3}{4}} z.$$

[2] В случае $\mu_b \neq 0$ (напомним, что электронный газ в объеме предполагается вырожденным, так что $\mu_b \geq 0$) полученные ниже результаты будут справедливы при достаточно большом приповерхностном изгибе зон $\mu_s - \mu_b >> \mu_b$ в той области пространственного заряда, где $\mu >> \mu_b$.

[3] Поскольку кейновское дисперсионное соотношение формально аналогично закону дисперсии релятивистского электрона, то появление в теории полупроводников параметров релятивистского типа не является удивительным.

Умножая обе части этого равенства на $\mu_s^{1/4}$ и вводя характерную длину

$$z_s = \frac{2\pi s\hbar}{\beta\mu_s}\sqrt{10}\left(\frac{\mu_s}{E_g}\right)^{\frac{3}{4}} \tag{4.80}$$

получаем

$$\left(\frac{\mu_s}{\mu}\right)^{\frac{1}{4}} - 1 = \frac{z}{z_s} \tag{4.81}$$

Таким образом распределение химпотенциала в обогащенном слое полупроводника с квадратичным законом дисперсии описывается зависимостью

$$\mu = \frac{\mu_s}{[1 + z/z_s]^4}. \tag{4.82}$$

с характерным размером приповерхностной потенциальной ямы z_s (4.80). На расстоянии от поверхности $z = z_s$ химпотенциал уменьшается от его максимального значения $\mu = \mu_s$ более чем на порядок, а объемная концентрация электронов

$$n \sim \mu^{3/2} \sim \frac{\mu_s^{3/2}}{(1 + z/z_s)^6}$$

почти на два порядка, т. е. почти все индуцированные в полупроводнике внешним электрическим полем электроны локализованы в приповерхностном слое шириной z_s. Если отсчитывать энергии от дна зоны на поверхности, то потенциальная энергия электрона в приповерхностной ОПЗ обогащенного слоя равна

$$U(z) = \mu_s - \mu = \mu_s\left[1 - \frac{1}{(1 + z/z_s)^4}\right]. \tag{4.83}$$

При небольших расстояниях от поверхности $z \ll z_s$ потенциал (4.83) близок к линейному

$$U(z) \approx \frac{4\mu_s}{z_s}z \quad (z \ll z_s), \tag{4.84}$$

а при $z \gg z_s$ описывается степенной функцией

$$U(z) \approx \mu_s\left[1 - \frac{z_s^4}{z^4}\right] \qquad (z \gg z_s). \tag{4.85}$$

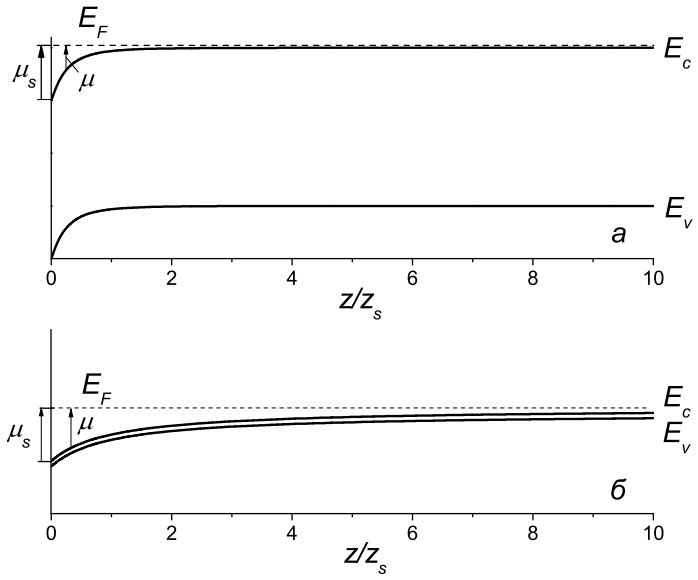

Рис. 4.6: Зонные диаграммы приповерхностной области полупроводника с большой (верхняя диаграмма) и предельно малой (нижняя диаграмма) шириной запрещенной зоны в режиме сильных обогащающих изгибов зон.

Аналогичным образом из уравнения (4.79) при том же граничном условии $\mu(z=0)=\mu_s$ находим пространственное распределение химпотенциала в полупроводнике с линейным дисперсионным соотношением

$$\mu = \frac{\mu_s}{1+z/z_s}, \tag{4.86}$$

где характерная длина

$$z_s = \frac{2\pi s\hbar}{\beta\mu_s} \tag{4.87}$$

обратно пропорциональна величине приповерхностного изгиба зон. Отметим, что в случае полупроводника со стандартной дисперсией эффективная ширина приповерхностной потенциальной ямы (4.80) уменьшается с ростом μ_s существенно слабее, а именно $z_s \propto \mu_s^{-1/4}$. Поскольку квадратичный закон дисперсии справедлив при достаточно малых энергиях $\varepsilon << E_g$, то выражение (4.82) (как и (4.80)) применимо только в области изгибов зон, удовлетворяющих условию $\mu_s << E_g$. Из сравнения выражений (4.80) и (4.87) с учетом того что $\beta \simeq 1$ следует, что при одинаковых приповерхностных изгибах зон μ_s приповерхностная ОПЗ в

широкозонных полупроводниках, как правило, значительно тоньше, чем в материалах с близким к линейному законом дисперсии.

4.2.6 Квантовый характер приповерхностных потенциальных ям в МДП – структурах

В структурах с пространственным ограничением потенциальная яма, в пределах которой локализовано движение носителей заряда, является квантовой, если ее размеры сопоставимы или меньше длины волны де Бройля, соответствующей характерной энергии носителей в яме. В композиционных гетероструктурах размер потенциальной ямы является независимым параметром (задается толщиной узкозонного слоя структуры в процессе ее выращивания) и практически не зависит от внешних условий или от концентрации локализованных в ней носителей. Эффективность квантования спектра в этом случае (квантовый характер ямы) определяется параметрами электронного газа - эффективной массой и характерной энергией носителей, которая в актуальном случае вырожденного газа определяется, в свою очередь, их концентрацией. В приповерхностных потенциальных ямах МДП - структур ситуация существенно сложнее. В этом случае от внешних условий (приложенного к полупроводнику внешнего электрического поля или температуры) зависит как концентрация газа носителей, а следовательно и их характерная длина волны, так и сама ширина потенциальной ямы, определяемая из самосогласованного решения уравнения Пуассона. Оказывается,однако, что в случае вырожденной статистики потенциальная яма в обогащенных слоях полупроводников всегда является квантовой, т. е. ее размеры близки или даже меньше длины волны де Бройля локализованных в ней электронов. Действительно, в случае потенциальной ямы с приповерхностным изгибом зон μ_s в полупроводнике с квадратичной дисперсией наименьшая длина волны электрона равна

$$\lambda_{dB} = \frac{2\pi\hbar}{p_F} = \frac{2\pi\hbar}{\sqrt{2m\mu_s}} = \frac{2\pi s\hbar}{\sqrt{E_g\mu_s}}, \tag{4.88}$$

где использована замена $m = E_g/s^2$. Выражение для характерной толщины обогащенного слоя (4.80) может быть переписано в виде

$$z_s = \frac{2\pi s\hbar}{\sqrt{\mu_s E_g}}\sqrt{10}\frac{1}{\beta}\left(\frac{\mu_s}{E_g}\right)^{\frac{1}{4}},$$

откуда, используя (4.88), следует

$$z_s = \sqrt{10}\frac{1}{\beta}\left(\frac{\mu_s}{E_g}\right)^{\frac{1}{4}}\lambda_{dB}\left(\mu_s\right). \tag{4.89}$$

Выше отмечалось, что это выражение справедливо при $\mu_s << E_g$, а так как $\beta \simeq 1$, то даже для наиболее коротковолновых электронов (с энергией $\varepsilon \sim \mu_s$)

$$z_s \lesssim \lambda_{dB} \tag{4.90}$$

т. е. приповерхностная потенциальная яма является квантовой. Для больших отношений μ_s/E_g (большие изгибы зон в узкощелевых полупроводниках) мы должны использовать формулу (4.87). Поскольку при близком к линейному дисперсионном соотношении наименьшая длина волны электрона в приповерхностной потенциальной яме равна

$$\lambda_{dB} = \frac{2\pi s\hbar}{\mu_s}, \tag{4.91}$$

то из (4.87) следует

$$z_s = \frac{2\pi s\hbar}{\beta\mu_s} = \frac{1}{\beta}\lambda_{dB}. \tag{4.92}$$

Таким образом условие (4.90) оказывается выполненным при любом соотношении между μ_s и E_g, т. е. приповерхностную яму с вырожденным электронным газом всегда следует рассматривать, как квантовую.

Приведенные оценки справедливы при достаточно сильных приповерхностных изгибах зон, удовлетворяющих условию $\mu_s >> \mu_b$ (см. примечание 2). В обратном пределе $\mu_s << \mu_b$ мы имеем дело с рассмотренным ранее режимом линейного экранирования, где в качестве характерного размера приповерхностной потенциальной ямы выступает длина экранирования (3.71). Подставляя в (3.71) объемную плотность состояний на уровне Ферми для квадратичного закона дисперсии (табл. 2.1)

$$D_{3D}(\mu_b) = \frac{(2m)^{3/2}}{2\pi^2\hbar^3}\mu_b^{1/2} = \frac{E_g^{3/2}}{2\pi^2\hbar^3 s^3}\mu_b^{1/2}$$

(в последнем равенстве использовано соотношение (4.73)) и используя введенный выше параметр β (4.77) приходим к выражению

$$L_{scr} = \frac{2\pi s\hbar}{\mu_b}\frac{1}{\beta\sqrt{3}}\left(\frac{\mu_b}{E_g}\right)^{\frac{3}{4}}, \tag{4.93}$$

отличающемуся только численным множителем от выражения (4.80) для параметра z_s. Аналогично, с случае линейного дисперсионного соотношения с плотностью состояний

$$D_{3D}(\mu_b) = \frac{1}{\pi^2 s^3 \hbar^3}\mu_b^2$$

(табл. 2.1) получаем

$$L_{scr} = \frac{2\pi s\hbar}{\mu_b}\frac{1}{\beta\sqrt{6}}. \tag{4.94}$$

Учитывая, что в вырожденном полупроводнике характерная длина волны де Бройля в режиме линейного экранирования определяется соотношениями (4.88) и (4.91) с точностью до замены $\mu_s \to \mu_b$, длины экранирования (4.93) и (4.94) могут быть выражены через фермиевские длины волн де Бройля в виде

$$\begin{aligned} L_{scr} &= \frac{1}{\beta\sqrt{3}}\left(\frac{\mu_s}{E_g}\right)^{\frac{3}{4}}\lambda_{dB} \qquad (\varepsilon = p^2/2m), \\ L_{scr} &= \frac{1}{\beta\sqrt{6}}\lambda_{dB} \qquad (\varepsilon = ps), \end{aligned}$$

Отсюда следует, что и в случае малых изгибов зон ширина приповерхностной потенциальной ямы при любом законе дисперсии также оказывается порядка или меньше фермиевской длины де Бройля, указывая на квантовый характер движения локализованных в приповерхностном слое электронов.

4.3 Задачи

1. Найдите (в предположении невырожденной статистики и полной ионизации примесей) пространственное распределение безразмерного потенциала $y(z)$, объемной концентрации электронов $n_{3D}(z)$ и

емкость ОПЗ C_{sc} в приповерхностной области полупроводника n – типа при сильном обогащении.

Ответ:

Из условия сильного обогащения $n(z) = n_b \exp y >> n_b$ следует $\exp y >> 1$, $y(z) >> 1$ и для функции F_n (4.63) получаем

$$F_n \approx -\exp\frac{y}{2}.$$

В результате уравнение (4.64) с граничным условием $y(0) = y_s$ дает

$$y(z) = y_s - \ln\left(1 + \frac{z}{\sqrt{2}L_{ns}}\right)^2,$$

где выражение для характерной толщины обогащенного слоя

$$L_{ns} = \sqrt{\frac{\epsilon k_B T}{4\pi e^2 n_s}} = L_n \exp -\frac{y_s}{2}$$

аналогично выражению для длины линейного экранирования (4.65) с точностью до замены концентрации электронов в объеме n_b на их объемную концентрацию вблизи поверхности $n_s = n(0) = n_b \exp y_s$. Аналогично, используя характерную длину L_{ns} могут быть получены простые выражения для объемной концентрации электронов

$$n(z) = n_s\left(\frac{1}{1 + z/\sqrt{2}L_{ns}}\right)^2$$

и дифференциальной емкости

$$C_{sc} = \frac{\epsilon}{4\pi\sqrt{2}L_{ns}} = \frac{\epsilon}{4\pi\sqrt{2}L_n}\exp\frac{y_s}{2}. \tag{4.95}$$

2. Найдите ход потенциала $y(z)$, величину объемного заряда доноров на единицу площади структуры Q_{sc} и емкость приповерхностной области C_{sc} для полупроводника n – типа в режиме обеднения ($n(z) << n_b$, $p(z) << n_b$).

   Ответ:

   Из условия $n(z) = n_b \exp y << n_b$ имеем $\exp y << 1$ ($y < -3$), а из условия $p(z) << n_b$ следует $p(z) = p_b \exp -y(z) = n_b\lambda^2 \exp -y(z) <<$

n_b, т. е. $\lambda^2 \exp -y << 1$ ($y > 2\ln\lambda + 3$). Из (4.63) видно, что в области обедняющих изгибов зон ($-3 > y > 2\ln\lambda + 3$) для функции F_n справедлива аппроксимация

$$F_n(y) \approx \sqrt{-y}.$$

Уравнение (4.64) с граничным условием $y(0) = y_s$ приводит к параболическому потенциалу

$$y(z) = y_s \left(1 - \frac{z}{W}\right)^2 \tag{4.96}$$

с характерной толщиной слоя обеднения

$$W = \sqrt{2} L_n \sqrt{-y_s} >> L_n \tag{4.97}$$

Поскольку $n_b = N_D$, то

$$W = \sqrt{\frac{\epsilon k_B T(-y_s)}{2\pi e^2 N_D}},$$

что совпадает с выражением (4.38) для толщины слоя обеднения в асимметричном ($N_A >> N_D$) $p - n$ – переходе при изгибе зон $u_c - eV = -k_B T y_s = \mu_b - \mu_s$.

Будучи выраженными через W, заряд и емкость соответственно равны

$$Q_{sc} = e n_b W,$$

$$C_{sc} = \frac{\epsilon}{4\pi W} = \frac{\epsilon}{4\pi\sqrt{2} L_n} \frac{1}{\sqrt{-y_s}}. \tag{4.98}$$

3. Найдите (в предположении невырожденной статистики и полной ионизации примесей) пространственное распределение безразмерного потенциала $y(z)$, объемной концентрации дырок $n_{3D}(z)$ и емкость C_{sc} инверсионного слоя в полупроводнике n – типа в режиме сильной инверсии.

   Ответ:

При сильной инверсии доминирующий вклад в объемный заряд в инверсионном слое дают неосновные носители, т. е. дырки в материале n – типа. Ввиду этого $p(z) = p_b \exp -y(z) = n_b \lambda^2 \exp -y >> n_b$, т. е. $\lambda^2 \exp -y >> 1$. А поскольку $\lambda << 1$ и $y < 0$, то $\lambda^2 \exp -y >> |y| >> 1 >> \exp y$ и в выражении (4.63) для функции F_n существенно только второе слагаемое

$$F_n(y, \lambda) = \lambda \exp -\frac{y}{2}.$$

Уравнение (4.64) с такой функцией и граничным условием $y(0) = y_s$ приводит, аналогично случаю сильного обогащения, к логарифмическому ходу поверхностного потенциала

$$y(z) = y_s + \ln \left(1 + \frac{z}{\sqrt{2} L_{ps}}\right)^2 \tag{4.99}$$

с характерной длиной

$$L_{ps} = \sqrt{\frac{\epsilon k_B T}{4\pi e^2 p_s}} = L_n \lambda^{-1} \exp \frac{y_s}{2}, \tag{4.100}$$

определяемой объемной концентрацией дырок на поверхности

$$p_s = p(0) = p_b \exp(-y_s) = n_b \lambda^2 \exp(-y_s).$$

Пространственное распределение дырок в инверсионном слое с точностью до замены $L_{ns} \to L_{ps}$ аналогично таковому для электронов при сильном обогащении

$$p(z) = n_s \left(\frac{1}{1 + z/\sqrt{2} L_{ps}}\right)^2.$$

Дифференциальная емкость может быть записана в виде, аналогичном емкости плоского конденсатора с расстоянием между обкладками равном $\sqrt{2} L_{ps}$

$$C_{sc} = \frac{\epsilon}{4\pi \sqrt{2} L_{ps}} = \frac{\epsilon}{4\pi \sqrt{2} L_n} \lambda \exp -\frac{y_s}{2}. \tag{4.101}$$

4. Сопоставьте толщину инверсионного слоя в полупроводнике n – типа с толщиной отделяющего его от объема слоя обеднения, предполагая невырожденную статистику и полную ионизацию доноров.

Ответ:

Инверсионному слою соответствует интервал расстояний от поверхности от $z = 0$ до $z = z_i$, в пределах которого выполнено условие $\lambda^2 \exp -y >> 1$, т. е. соответствующий интервалу изгибов зон

$$2 \ln \lambda = y_i > y(z) > y(0) = y_s. \tag{4.102}$$

Ход потенциала в этой области описывается приближенно уравнением (4.99), подставляя в которое $y(z_i) = y_i = \ln \lambda^2$ и разрешая уравнение относительно $z(= z_i)$ находим толщину инверсионного слоя

$$z(y_i) = z_i = \sqrt{2} L_{ps} \left(\lambda \exp -\frac{y_s}{2} - 1 \right) \approx \sqrt{2} L_{ps} \lambda \exp -\frac{y_s}{2} = \sqrt{2} L_n, \tag{4.103}$$

где учтено, что при сильной инверсии $y_s < 2 \ln \lambda^2$, а в последнем равенстве использовано соотношение (4.100).

Слою обеднения соответствует интервал расстояний от поверхности, для которого

$$2 \ln \lambda = y_i = y(z_i) < y(z) < 1 \tag{4.104}$$

и безразмерный изгиб зон удовлетворяет уравнению (4.96), в котором, если отсчитывать расстояние от точки $z = z_i$, соответствующей границе между инверсионным и обедненным слоем, следует положить $y_s = y_i$. Толщина следующего за инверсионным каналом слоя обеднения определяется соотношением (4.97) при $y_s = y_i$, что дает

$$W_i = \sqrt{2} L_n \sqrt{-y_i} = \sqrt{2} L_n \sqrt{\ln \lambda^{-2}}. \tag{4.105}$$

Отсюда и из (4.103) для искомого отношения толщин обедненного и инверсионного слоев получаем

$$\frac{W_i}{z_i} = \sqrt{\ln \lambda^{-2}}. \tag{4.106}$$

что, используя определение (2.108)

$$\lambda^2 = \lambda^2(\mu_b) = \frac{p_b}{n_b} = \left(\frac{m_v}{m_c}\right)^{3/2} \exp\left(\frac{-\varepsilon_g - 2\mu_b}{k_B T}\right),$$

может быть записано в виде

$$\frac{W_i}{z_i} = \sqrt{\frac{E_g + 2\mu_b}{k_B T} + \frac{3}{2}\ln\frac{m_c}{m_v}}. \tag{4.107}$$

В случае широкозонных полупроводников толщина слоя обеднения даже при комнатных температурах, как правило, во много раз превышает толщину слоя инверсии.

5. Проанализируйте зависимость дифференциальной емкости ОПЗ полупроводника n – типа от безразмерного приповерхностного изгиба зон y_s.

   Ответ:

   Введением характерной для полупроводника n – типа масштабной емкости

$$C_{n0} = \frac{\epsilon}{4\pi\sqrt{2}L_n} \tag{4.108}$$

   выражения (4.95), (4.98) и (4.101) для дифференциальной емкости ОПЗ в режимах сильного обогащения, обеднения и сильной инверсии могут быть записаны соответственно в виде

$$C_{sc} = C_{n0} \times \begin{cases} \exp(y_s/2) & (y_s > 3) \\ 1/\sqrt{-y_s} & (2\ln\lambda + 3 < y_s < -3) \\ \lambda\exp(-y_s/2) & (y_s < 2\ln\lambda - 3) \end{cases} \tag{4.109}$$

   С ростом абсолютной величины $|y_s|$ емкость экспоненциально возрастает в области обогащающих и инверсионных изгибов зон и уменьшается при изгибах зон, соответствующих обеднению. Таким образом, на вольт-фарадной характеристике имеется минимум, приходящийся на значение y_s, соответствующее переходу от

режима обеднения к режиму инверсии $y_s \approx y_i = -2\ln\lambda$. Значение емкости в минимуме приблизительно равно

$$C_{sc}^{\min} = \frac{C_{n0}}{\sqrt{-y_i}} = \frac{e}{4}\sqrt{\frac{\epsilon n_i}{\pi k_B T}}\frac{1}{\sqrt{-\lambda\ln\lambda}},$$

что, при известных температуре и зонных параметрах полупроводника (известной собственной концентрации n_i), позволяет определить параметр легирования λ, т. е. концентрацию легирующих примесей.

6. Прокомментируйте представленные на рис. 4.7 расчетные зависимости индуцированного в полупроводнике заряда и дифференциальной емкости от поверхностного изгиба зон для двух полупроводников n–типа с одинаковыми зонными параметрами. Найдите изгибы зон y_s, соответствующие переходам к режиму обогащения и инверсии. Сравните эти величины для разных образцов и с теоретическими значениями.

7. Прокомментируйте представленные на рис. 4.8 зависимости $Q_{sc}(y_s)$ и $C_{sc}(y_s)$ для собственного, электронного и дырочного полупроводников с одинаковыми зонными параметрами. Найдите изгибы зон y_s, соответствующие режиму обеднения. Сравните эти величины для разных образцов и с теоретическими значениями. Укажите, каким преобразованием связаны значения $Q_{sc}(y_s)$ и $C_{sc}(y_s)$ для двух полупроводников с $\lambda_1 = \lambda_2^{-1}$.

4.4 Вопросы для самопроверки

1. Перечислите основные типы наноструктур, используемых для создания электронных систем пониженной размерности и укажите их основные особенности.

2. Нарисуйте равновесную энергетическую диаграмму n-p гетероперехода, в котором ширина запрещенной зоны дырочного полупроводника больше и весь разрыв зон приходится на валентную зону.

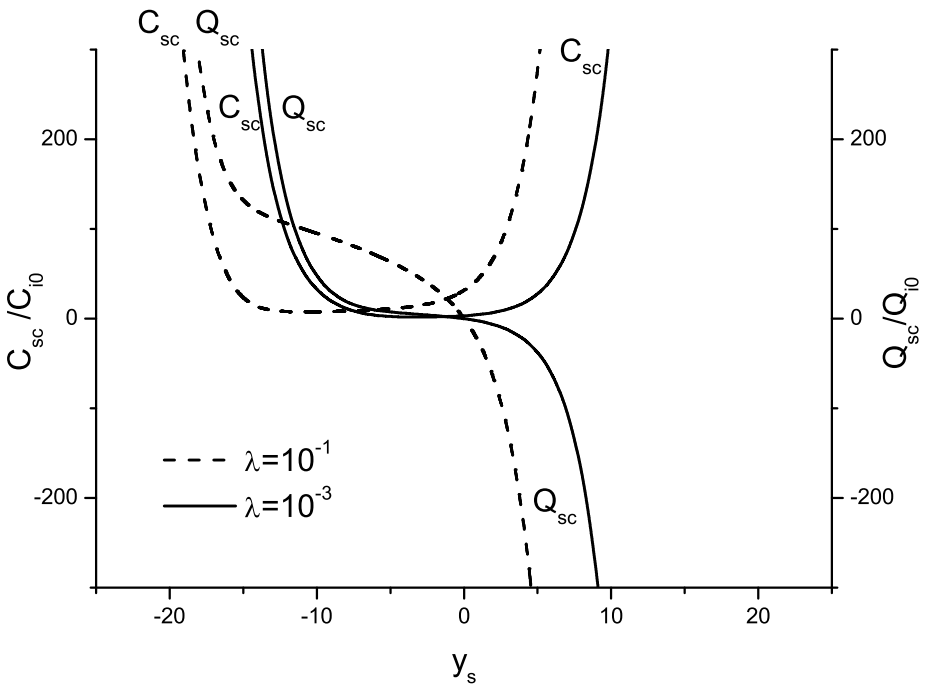

Рис. 4.7: Зависимости нормированных поверхностной плотности заряда и дифференциальной емкости от поверхностного изгиба зон в полупроводнике n–типа при различных уровнях легирования, рассчитанные в предположении невырожденного заполнения зонных состояний.

Для какого типа носителей и в какой области образуется в таком гетеропереходе квантовая яма? Возникнет ли такая яма, если энергетическая щель больше в материале n -типа. Что будет, если разрыв зон испытывает только зона проводимости?

3. Отличается ли ход потенциала в области пространственного заряда в модели Андерсона для гетероперехода от случая гомоперехода при том же уровне легирования материалов n - и p -типа?

4. Какими параметрами образующих гетеропару полупроводников определяются

   (a) контактная разность потенциалов?

   (b) величина и знак разрывов зон на гетерогранице?

5. Каковы наиболее существенные приближения, используемые в модели Андерсона для распределения потенциала в гетеропереходе?

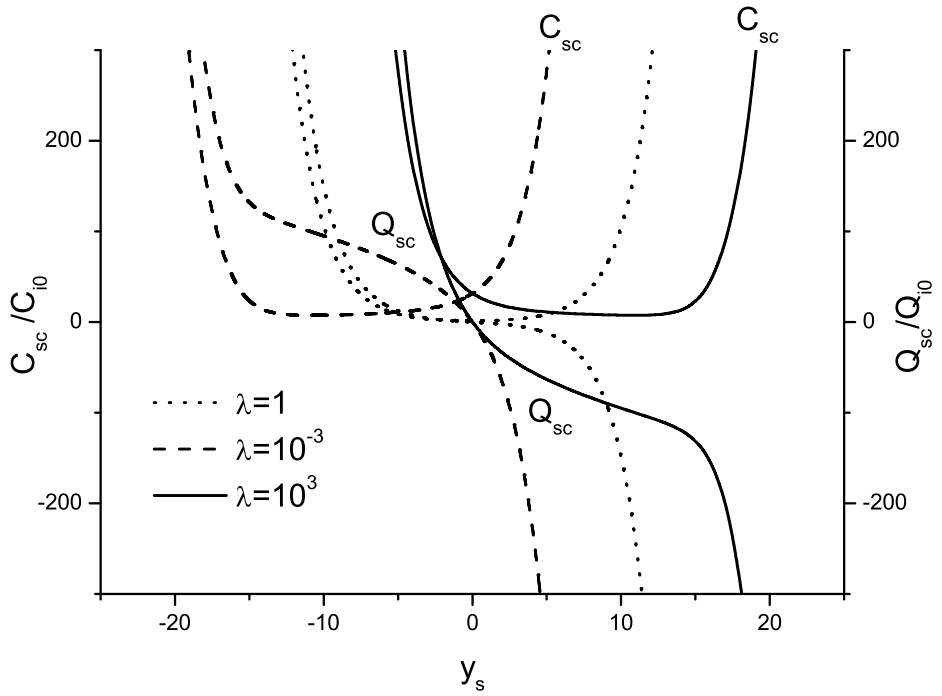

Рис. 4.8: Зависимости нормированных поверхностной плотности заряда и дифференциальной емкости от поверхностного изгиба зон в собственном, электронном и дырочном полупроводниках.

В каких областях приконтактного слоя эти условия нарушены наверняка, а в каких могут быть нарушены, но не обязательно (укажите когда не нарушены)?

6. Каким соотношением связаны изгибы зон приходящиеся на n - и p -области гетероперехода с концентрациями легирующей примеси и диэлектрическими проницаемостями?

7. Какими параметрами определяется ход потенциала в области гетероперехода и каков его характер зависимости от расстояния до гетероперехода?

8. Найдите из условия электронейтральности слоя обеднения в целом (при известных концентрациях легирующих примесей и в предположении их полной ионизации) отношение толщин переходных областей в двух сторонах перехода. В какой области толщина слоя обеднения больше в случае сильной асимметрии легирования?

9. Как зависит электрическое поле в слое обеднения от координаты? Изобразите схематически эту зависимость для гетеро- и гомоперехода.

10. Увеличивается или уменьшается толщина слоя обеднения при приложении разности потенциалов к переходу в запорном направлении? Укажите к какой области перехода для реализации этого режима должен быть приложен положительный полюс внешнего источника напряжения.

11. Меняются ли толщины переходных областей в двух сторонах полупроводникового гетероперехода и их отношение при приложении к гетеропереходу внешнего напряжения?

12. Как из измерений дифференциальной электроемкости гетероперехода можно определить контактную разность потенциалов? Какие параметры полупроводников гетеропары необходимо дополнительно знать, чтобы из этих измерений определить величину разрывов зон на гетрогранице? Из каких экспериментов можно определить эти параметры?

13. По какой причине расчет классического потенциала в приповерхностной области полупроводника МДП-структур, в отличие от гетероперехода (в модели полного истощения), требует самосогласованного определения потенциала и заряда? Какое дифференциальное уравнение дает решение этой классически (в действительности квазиклассической - почему?) самосогласованной задачи? Каковы граничные условия, накладываемые на потенциал?

14. Изобразите схематически зонные диаграммы для приповерхностного слоя полупроводника n -типа, соответствующие режимам линейного экранирования, обогащения, обеднения и инверсии. Рассмотрите случаи вырожденного и невырожденного полупроводника.

15. Как известно, в трехмерном полупроводнике при не слишком высоком уровне легирования электронный газ не может быть сильно

вырожденным. Можно ли достичь сильного вырождения электронного (или дырочного) газа в приповерхностной области нелегированного полупроводника в МДП-структуре?

16. Как (качественно) ведет себя дифференциальная емкость области пространственного заряда полупроводника n -типа и МДП - структуры в целом при изменении приповерхностного изгиба зон от больших отрицательных до больших положительных значений?

17. По какому закону изменяется приповерхностный потенциал с координатой в режиме плоских зон? На какую глубину проникает внешнее электрическое поле в полупроводник в этом режиме. Как называется эта величина и чему она равна в вырожденном и невырожденном полупроводнике?

18. Каким законом может быть аппроксимирован потенциал при обогащении приповерхностной области полупроводника электронами на небольших, по сравнению с глубиной проникновения электрического поля в полупроводник, расстояниях от поверхности?

19. По какой простой формуле электростатики можно вычислить величину заряда в приповерхностной области полупроводника МДП-структуры, если известно электрическое поле на поверхности полупроводника? Как из измерений емкости можно определить индуцированный в приповерхностной области полупроводника заряд (а следовательно и число носителей заряда) при известной разности потенциалов, приложенной к структуре?

20. Приповерхностная область полупроводника (электронный канал полевого транзистора) может быть обогащена электронами как в режиме обогащения (в полупроводнике n -типа), так и при сильной инверсии (в полупроводнике p -типа). В каком из этих двух случаев на электроны, при их одинаковой поверхностной концентрации, действует большее электрическое поле? Или оно одинаково?

21. Химпотенциал в вырожденном полупроводнике при сильном при-

поверхностном изгибе зон μ_s может меняться с расстоянием от поверхности z как

$$\mu = \frac{\mu_s}{[1 + z/z_s]^4},$$

или

$$\mu = \frac{\mu_s}{1 + z/z_s}$$

Каким случаям соответствуют эти распределения потенциала? Одинаковы ли фигурирующие здесь характеристические длины?

22. Как соотносятся эффективная ширина квазиклассически самосогласованной приповерхностной потенциальной ямы и минимальная длина волны де Бройля электронов для вырожденного полупроводника

    (a) с линейной дисперсией?

    (b) с квадратичной дисперсией?

    Зависит ли это соотношение от глубины приповерхностной ямы?

23. Что можно сказать о результате расчета квазиклассически самосогласованного приповерхностного потенциала, если полученный при расчете размер ямы больше фермиевской длины волны электрона?

Глава 5

Размерное квантование в структурах с пространственным ограничением

Поскольку типичные размеры приповерхностной потенциальной ямы оказываются порядка, а зачастую и меньше, характерной длины де Бройля, т.е. такая яма является квантовой, энергетический спектр электронов в яме должен быть рассчитан исходя из квантового описания. Электроны вблизи гетерограниц (границы раздела с диэлектриком в МДП-структурах) находятся в поле довольно сложного потенциала. С одной стороны на них действует поле периодического потенциала решетки внутри квантовой ямы, на которое накладывается относительно медленно меняющееся в пространстве электрическое поле, создаваемое примесями и приложенным к структуре напряжением при использовании полевых электродов (затворов). На границах раздела формируется, как правило, высокий потенциальный барьер, что приводит к отталкиванию электронов. Свойства переходного слоя на гетерограницах могут существенно отличаться от свойств образующих гетеропереход материалов. Полное описание системы электронов в структурах с пространственным ограничением лежит, вообще говоря, вне рамок существующих теорий. Для определения электронного спектра в таких системах необходим, таким образом, ряд упрощающих предположений, не искажающих физической

сущности рассматриваемой системы, с последующей оценкой их возможного влияния на полученные результаты. Наиболее существенные из таких упрощений, которые будут предполагаться в дальнейшем, следующие:

1. Прежде всего, в дальнейшем мы полагаем, что характерный размер ямы L_z существенно превышает постоянную решетки $a << L_z \leq \lambda\!\!\!^{-}$ и потенциал мало изменяется на ее масштабе. Таким образом, может быть использован метод эффективной массы, в рамках которого некоторым образом учитывается микроскопическая структура полупроводника, а именно, роль периодического решеточного потенциала сводится к перенормировке параметров законов дисперсии (в простейшем случае квадратичного закона дисперсии к перенормировке массы электрона (дырки)), которые в дальнейшем полагаются известными. Волновая функция электрона вблизи гетерограницы может затухать на расстояниях порядка нескольких постоянных решетки и применимость метода эффективной массы в этой области, вообще говоря, сомнительна. В связи с этим однако следует заметить, что недостаток надежной информации о физических параметрах приграничных слоев в значительно большей мере ограничивает точность расчетов, нежели использование метода эффективной массы.

2. При описании спектра мы не будем принимать во внимание многочастичные эффекты, в том числе, связанные с наличием спиновой степени свободы – обменно-корреляционное взаимодействие, т.е. ограничимся простейшим приближением Хартри.

3. Границы гетеропереходов полагаются резкими и любые поверхностные эффекты, в том числе наличие поверхностных состояний и их возможная перезарядка в дальнейшем при расчете спектра игнорируются.

4. Диэлектрические проницаемости материалов, образующих гетеропереходы, предполагаются близкими, так что влиянием сил изоб-

ражения можно пренебречь (в случае кремниевых МДП-структур, являющихся одним из наиболее распространенных объектов в физике низко-размерных систем, учет потенциала изображения приводит к заметной модификации энергетического спектра).

5.1 Общая формулировка проблемы. Основные уравнения

В случае однозонного приближения проблема расчета спектра, таким образом, сводится к решению уравнения Шредингера с эффективной массой m (для простоты мы ограничимся материалами с изотропным спектром и для изотропной эффективной массы электрона $m*$ в дальнейшем используем обозначение m, ось z выбрана в направлении ограничивающего потенциала)

$$\left[\frac{1}{2m}(\hbar^2\hat{k}_z^2 + \hbar^2\hat{k}_x^2 + \hbar^2\hat{k}_y^2) + U(z)\right]\phi(z,x,y) = \varepsilon\phi(z,x,y) \tag{5.1}$$

Полагая, что потенциал имеет вид

$$U(z,x,y) = U(z) + U(x,y),$$

(1+2 -мерная задача) переменные в уравнении Шредингера могут быть разделены (такое разделение, вообще говоря, несправедливо, если на границах гетероструктуры нарушено условие $\varphi(L_i) = 0$, которое в дальнейшем предполагается выполненным). Далее мы будем полагаем, что движение носителей в двумерной (x,y) плоскости является свободным, т.е. $U(x,y) = const = 0$.

В этом случае волновая функция может быть представлена в виде

$$\phi(z,x,y) = \varphi(z)\exp i(k_x x + k_y y),$$

где $(k_x^2+k_y^2)/2m$. Нетрудно видеть, что действие операторов $\hat{k}_x^2 = (-i\partial/\partial x)^2$ и $\hat{k}_y^2 = (-i\partial/\partial y)^2$ на волновую функцию приводит к появлению слагаемого $(k_x^2 + k_y^2)/2m$ в левой части уравнения (5.1), соответствующего кинетической энергии свободного движения электрона в двумерной плоскости. Таким образом, энергия, соответствующая движению электрона в

двумерной плоскости, и энергия, отвечающая квантованному движению в направлении ограничивающего потенциала (в направлении конфайнмента), аддитивны (следует подчеркнуть, что вывод об аддитивности $\varepsilon_\perp$и $\varepsilon_{||}$ не носит универсального характера и справедлив только в случае полупроводников с параболическим спектром - см. ниже). Представив полную энергию в виде

$$\varepsilon = \varepsilon_\perp + \varepsilon_{||} = \varepsilon_n + \hbar^2 k_s^2/2m, \tag{5.2}$$

где

$$\hbar k_s = \hbar\sqrt{k_x^2 + k_y^2} \tag{5.3}$$

– двумерный квазиимпульс, приходим к следующему уравнению для собственных значений $\varepsilon_{\perp i}$ и волновой функции, описывающей распределение электронной плотности в направлении конфайнмента

$$\left[\frac{\hat{k}_z^2}{2m} + U(z)\right]\varphi_i(z) = \varepsilon_{\perp i}\varphi_i(z) \tag{5.4}$$

Каждому такому квантовому уровню в потенциальной яме (уровню размерного квантования) принадлежит континуум состояний, отличающихся значениями двумерного квазиимпульса

$$\varepsilon_i = \varepsilon_{\perp i} + \varepsilon_{||i} = \varepsilon_{\perp i} + \hbar^2 k_{si}^2/2m \tag{5.5}$$

(двумерная подзона).

Потенциал $U(z)$, как и при классическом рассмотрении, должен быть определен из уравнения Пуассона. Ясно, однако, что использованный ранее при классическом описании потенциала метод локальной плотности (вклад зонных носителей заряда в объемный заряд определялся исходя из объемных законов дисперсии) не дает адекватного описания распределения заряда, а, следовательно, и распределения потенциала в квантовой яме. Наиболее ярко такая разница проявляется вблизи гетерограниц с высокими потенциальными барьерами. Так, при классическом рассмотрении приповерхностной потенциальной ямы в МДП-структурах, локальная плотность электронов $n(z)$ максимальна вблизи границы с диэлектриком, где максимально значение химпотенциала μ_s, тогда как

квантово-механическое описание предполагает $n(0) \approx 0$ в виду сильного затухания волновой функции вблизи потенциального барьера на границе. (При $U_b = \infty$, что зачастую оправдано для приповерхностных квантовых ям в МДП-структурах и, как правило, предполагается при теоретическом их описании, $n(0) = 0$.)

В случае квантовой ямы для электронов их концентрация $n(z)$, фигурирующая в уравнении Пуассона, определяется волновой функцией

$$n(z) = \sum_i n_{si} \varphi_i^*(z) \varphi_i(z), \tag{5.6}$$

где n_{si} - число носителей в i-й двумерной подзоне на единицу площади (двумерная концентрация), определяемое в условиях термодинамического равновесия из условия общности уровня Ферми для всех заполненных подзон (энергии Ферми в различных двумерных подзонах $\mu_i = E_F - E_i$ при этом, конечно, отличаются, поскольку энергии, соответствующие дну двумерных подзон $E_i = E_i(k_{si} = 0)$ -уровни размерного квантования- различны). Уравнения Шредингера и Пуассона должны еще быть дополнены условием нормировки волновых функций

$$\int_{-\infty}^{\infty} \varphi_i^*(z) \varphi_i(z) dz = 1$$

Так как уравнение Пуассона для одноэлектронного потенциала теперь само содержит волновые функции, то решение уравнения Шредингера является нелинейной задачей на собственные значения $\varepsilon_{\perp i}$. Самосогласованное решение уравнений Шредингера и Пуассона возможно только численными методами, основанными на использовании итерационной процедуры – решается уравнение Шредингера с предполагаемым затравочным потенциалом и определяются волновые функции, с помощью которых можно получить следующую итерацию для потенциала и т.д. до получения желаемой точности.[1]

[1]Следует отметить, что полученный в результате n-й итерации потенциал, вообще говоря, нельзя использовать для $(n+1)$-й итерации уравнения Шредингера, поскольку такая процедура может оказаться расходящейся. Простейший метод заключается в использовании линейной комбинации n-го затравочного потенциала и потенциала полученного в результате этой итерации.

Основной недостаток численных методов – отсутствие достаточно ясной физической картины и как следствие крайне малая эвристическая и методологическая ценность такого подхода. С методической же точки зрения важна не столько точность полученного результата (ни одна задача в физике не решается точно), сколько возможность получения достаточно простой физической картины рассматриваемого явления, соответствующей наиболее важным, определяющим, его особенностям. Умение пренебрегать какими-либо воздействиями, не искажая основной сути рассматриваемой проблемы, умение оценивать возникающие при этом погрешности имеет не меньшее методологическое значение. При дальнейшем рассмотрении мы кардинально упростим проблему квантования спектра в гетероструктурах предполагая, во первых, ограничивающий потенциал заданным (устраняя тем самым задачу самосогласования) и, во вторых, при решении самого уравнения Шредингера ограничимся квазиклассическим приближением, в рамках которого нахождение спектра сводится к простому интегрированию. Получаемые при таком подходе решения могут быть зачастую представлены в аналитическом виде, обеспечивая ясное понимание качественных особенностей спектра и характера его зависимости от основных, определяющих, параметров гетеросистемы.

Что касается первого упрощения, т.е. игнорирования вклада квантовых поправок в распределение плотности электронного заряда при вычислении потенциала, то оно представляется достаточно оправданным для композиционных гетероструктур, в которых ограничивающий потенциал определяется в первую очередь задаваемой в процессе выращивания гетероструктуры толщиной слоя с двумерным электронным газом и высотой барьеров на его границе с более широкозонным материалом (величиной разрыва зон на гетерограницах). При не слишком больших концентрациях электронов это приближение справедливо и в случае квантовых ям на основе одиночных гетеропереходов и дельта-легированных слоев, где рельеф потенциала определяется в основном зарядом легирующей примеси с высокой концентрацией, а вклад электронов в квантовой яме, а тем самым и квантового характера распределения электрон-

ной плотности, в формирование результирующего потенциала невелик. В случае приповерхностных квантовых ям в МДП-структурах вкладом заряда электронов в формирование приповерхностного потенциала можно пренебречь только в случае электронных инверсионных слоев, при условии, что двумерная концентрация электронов в инверсионном слое мала по сравнению с поверхностной плотностью заряда в слое обеднения. Однако в общем случае (особенно в обогащенных электронных слоях и в инверсионных каналах на слаболегированных подложках) вклад заряда электронов приповерхностной квантовой ямы в определяемый уравнением Пуассона результирующий приповерхностный потенциал может оказаться доминирующим. Тем не менее и здесь можно избежать необходимости самосогласованного решения уравнений Пуассона и Шредингера, используя самосогласованный классический поверхностный потенциал, полученный нами в Разд.4.2. По существу такое приближение аналогично теории возмущений, когда спектр системы в первом приближении определяется волновыми функциями нулевого приближения (в рассматриваемой задаче распределением электронной плотности не квантованной системы - см. (5.6)). В большинстве представляющих интерес случаев использование классически самосогласованного потенциала при квантовании спектра в квантовых гетероямах приводит к результатам, весьма близким к точным решениям, использующим сложную процедуру самосогласования.

Как правило, не приводит к существенным погрешностям при расчете спектра в квантовых ямах и использование квазиклассического метода решения уравнения Шредингера. Квазиклассическое приближение дает разумные результаты даже в тех случаях, когда условие его применимости (малость дебройлевской длины волны по сравнению с размерами ямы, или, что тоже самое большие номера квантовых чисел i), казалось бы, не выполняется. Однако, на самом деле, параметром разложения в квазиклассическом подходе является не величина $1/i$, а отношение $1/\pi^2 i^2$. В результате стационарное решение уравнения Шредингера достаточно хорошо описывается квазиклассическим приближением даже для основного (погрешность не превышает 0.1) и первого возбужден-

ного (вносимая ошибка менее 0,025) состояний. Для ряда же наиболее важных физических задач квазиклассический спектр совпадает со спектром, даваемым точным решением (движение в кулоновском потенциале, в магнитном поле), либо весьма близок нему (финитное движение в однородном электрическом поле).

5.2 Прямоугольная потенциальная яма в композиционной гетероструктуре

В рамках метода эффективной массы расчет электронного спектра в полупроводниковых наноструктурах выглядят часто как практические занятия по квантовой механике. Мы начнем с простейшего случая трехслойной композиционной гетероструктуры, дно зоны проводимости узкого внутреннего слоя которой лежит ниже дна зоны проводимости окружающих его барьерных слоев. Потенциал такой одиночной электронной квантовой ямы может рассматриваться как прямоугольный, решение уравнения Шредингера для которого имеется в любом учебнике по квантовой механике. За направление квантования здесь и далее выбирается ось z, движение по осям x, y предполагается свободным - $U(x, y) = const$.

5.2.1 Приближение бесконечно высоких барьеров

Особенно просто может быть получено решение для ямы с бесконечно высоким потенциальными барьерами. Электрон при такой потенциальной стенке не может проникнуть в барьер, его волновая функция на границе с барьером равна нулю и, следовательно, для стационарных состояний на ширине ямы L_z должно укладываться полуцелое число длин волн де Бройля

$$\lambda\!\!\!^{-}\frac{i}{2} = \frac{\pi\hbar}{p_z} i = L_z, \tag{5.7}$$

что и определяет разрешенные (квантованные) значения перпендикулярной границе раздела компоненты квазиимпульса p_z

$$L_z p_z(\varepsilon_i) = \pi\hbar i. \tag{5.8}$$

Для определения энергетического спектра электронов необходимо задать их закон дисперсии $\varepsilon = \varepsilon(p(p_z))$. Рассмотрим простейшие квадратичный и линейный законы дисперсии, являющиеся предельными низко- и высокоэнергетическими случаями кейновского закона дисперсии (или релятивистского закона дисперсии для свободного электрона)

Квадратичный закон дисперсии

Для системы электронов с изотропным (обобщение на случай анизотропии спектра в этом случае тривиально) квадратичным законом дисперсии, движущихся в одномерной потенциальной яме

$$p_z = \sqrt{2m\varepsilon - (p_x^2 + p_y^2)} = \sqrt{2m\left(\varepsilon - \frac{p_s^2}{2m}\right)}, \tag{5.9}$$

что после подстановки в условие квантования (5.8) дает для спектра (энергии отсчитываются от дна потенциальной ямы)

$$\varepsilon_i(p_s) = \frac{\hbar^2}{2m}\frac{\pi}{L_z^2}i^2 + \frac{p_s^2}{2m} = \varepsilon_i(0) + \frac{p_s^2}{2m} = \varepsilon_\perp + \varepsilon_{||}, \tag{5.10}$$

где $p_s = \sqrt{p_x^2 + p_y^2}$ - двумерный квазиимпульс. Полная энергия электронов в яме в этом случае может быть разделена на две независимые части: энергию, связанную с движением в направлении, перпендикулярном границам раздела гетероструктуры, т.е. в направлении ограничивающего потенциала

$$\varepsilon_i(p_s = 0) = \frac{\hbar^2}{2m}\frac{\pi}{L_z^2}i^2 \tag{5.11}$$

(уровни размерного квантования в одномерной яме) и кинетическую энергию свободного движения в двумерном слое. Таким образом, спектр электрона в квантовой яме, характеризуемый в ее отсутствии квантовыми числами p_x, p_y, p_z, разбивается на несколько ветвей, называемых подзонами размерного квантования или просто подзонами и характеризуемых номером ветви (квантовое число i) и двумерным квазиимпульсом (квантовые числа p_x, p_y).

Линейный закон дисперсии

Отмеченная в предыдущем разделе взаимная независимость движений в направлении ограничивающего потенциала и перпендикулярно ему имеет место только в случае параболического дисперсионного соотношения, являясь в конечном счете следствием аддитивности энергий свободного движения по различным степеням свободы $\varepsilon = \varepsilon_x + \varepsilon_y + \varepsilon_z = p_x^2/2m + p_y^2/2m + p_z^2/2m$. Простой и яркой иллюстрацией взаимной зависимости этих движений является система с линейным дисперсионным соотношением, для которой даже в отсутствии внешнего потенциала энергии движения по различным степеням свободы не аддитивны

$$\varepsilon = ps = \sqrt{p_x^2 s^2 + p_y^2 s^2 + p_z^2 s^2} \neq p_x s + p_y s + p_z s. \tag{5.12}$$

Рассмотрим такую систему в бесконечно глубокой прямоугольной потенциальной яме. Внутри ямы величина z-й компоненты квазиимпульса выражается через полную энергию через соотношение

$$p_z = \frac{1}{s}\sqrt{\varepsilon^2 - p_s^2 s^2}. \tag{5.13}$$

Подстановка (5.13) в (5.8) дает для спектра

$$\varepsilon_i(p_s) = \sqrt{\frac{\pi^2 s^2 \hbar^2}{L_z^2} i^2 + p_s^2 s^2}. \tag{5.14}$$

Как и в случае квадратичного закона дисперсии спектр распадается на подзоны размерного квантования, характеризуемые квантовыми числами i, p_x, p_y, однако теперь он не может быть представлен как сумма энергии размерного квантования $\varepsilon_\perp = \varepsilon_i(p_s = 0)$ и кинетической энергии свободного движения внутри двумерной подзоны $\varepsilon_{||} = p_s s$. Действительно, из (5.14) видно, что

$$\varepsilon_i(p_s) = \sqrt{\varepsilon_\perp^2 + \varepsilon_{||}^2} \neq \varepsilon_\perp + \varepsilon_{||}, \tag{5.15}$$

т.е. определение спектра не может быть сведено к нахождению уровней размерного квантования $\varepsilon_i(0)$ с последующим добавлением энергии свободного движения в двумерной плоскости (в рассматриваемом частном

случае линейной дисперсии аддитивны квадраты энергий $\varepsilon^2 = \varepsilon_\perp^2 + \varepsilon_\parallel^2$). Повторим, что взаимозависимость движений в направлении квантования и перпендикулярно ему (невозможность свести 1+2 (пространственно) – мерную задачу к одномерной) присуща любой системе с неквадратичным спектром и при любой форме ограничивающего потенциала.

Уровни размерного квантования при $p_s = 0$ (энергетическое положение дна двумерных подзон) даются при линейной дисперсии, согласно (5.14), формулой

$$\varepsilon_i(0) = \frac{\pi s \hbar}{a} i. \tag{5.16}$$

Сопоставление этой формулы с (5.11) является хорошей иллюстрацией того, что спектр электрона в потенциальной яме определяется не только потенциалом, но и кардинальным образом зависит от описывающего систему дисперсионного соотношения. В отличие от случая электронов с квадратичной дисперсией, для которого энергии уровней размерного квантования растут по квадрату с номером подзоны $\varepsilon_i(0) \propto i^2$, для линейной дисперсии $\varepsilon_i(0) \propto i$, т.е. спектр эквидистантен.

В реальных полупроводниковых наноструктурах с квантовыми ямами высота барьера ΔU конечна и полученные выше результаты (5.10) и (5.14) имеют смысл только для подзон, дно которых $\varepsilon_i(0)$ лежат ниже вершины барьера, т.е. для достаточно малых номеров подзон

$$i < \frac{L_z}{\hbar}\sqrt{\frac{2m\Delta U}{\pi}} \tag{5.17}$$

при параболическом

$$i < \frac{L_z}{\hbar}\frac{\Delta U}{s\pi} \tag{5.18}$$

и линейном законах дисперсии. Для подзон, уровни размерного квантования которых лежат близко к верхнему краю потенциального барьера точность использованного приближения бесконечно глубокой ямы может оказаться недостаточной. В этом случае требуется решение уравнения Шредингера для прямоугольной ямы конечной высоты, решение которого требует использования численных методов (может быть найдено в стандартном учебнике по квантовой механике). В применении к гетероструктурам при решении этой задачи необходимо учесть, что

законы дисперсии полупроводников, образующих центральный слой с двумерным газом и барьерные слои отличаются. В частности, в случае стандартных полупроводников с квадратичным спектром, отличаются эффективные массы электрона. Поскольку материалы барьерных слоев, как правило, имеет бóльшую запрещенную зону, эффективная масса электрона в них бóльше. Поскольку при конечной высоте барьера волновая функция электрона частично проникает в барьер (часть времени электрон проводит в барьерных слоях), это должно приводить к эффективному увеличению массы электрона и, как следствие, к опусканию уровней размерного квантовая в яме, рассчитанных без учета этого эффекта (см. формулу (5.11), согласно которой $\varepsilon_i(0) \propto m^{-1}$). Наиболее существенным этот эффект будет, конечно, для верхних подзон, волновые функции которых более глубоко проникают под барьер.

5.3 Потенциальная яма произвольной формы.

В прямоугольной потенциальной яме потенциальная энергия электрона постоянна как внутри ямы, так и в области барьеров, отличаясь в этих областях на величину разрывов зоны проводимости. Как следует изменить приблизительное (для ямы с конечной высотой барьера) условие квантования (5.8) в случае произвольной формы ограничивающего потенциала $U(z)$? Ответ на этот вопрос можно получить, если обратить внимание, что левая часть равенства (5.8) есть площадь траектории электрона в фазовом пространстве, при его движении от одного края ямы до другого

$$L_z p_z = \int\limits_0^{L_z} p_z dz. \tag{5.19}$$

Если в прямоугольной яме ввиду постоянства потенциальной энергии компонента квазиимпульса p_z внутри ямы постоянна (траектория в фазовом пространстве - прямоугольник), при произвольном пространственной зависимости потенциала p_z также будет функцией координаты (фазовая траектория произвольна). Исходя из этих соображений прихо-

дим к обобщенному на случай произвольной потенциальной ямы правилу квазиклассического квантования

$$\int\limits_{z_1}^{z_2} p_z(\varepsilon, U(z), p_s)dz = \pi\hbar(i + \gamma_1 + \gamma_2), \tag{5.20}$$

известному как правило квантования Бора-Зоммерфельда. Координаты z_1, z_2 соответствуют точкам поворота при классическом движении в яме и определяются из условия $p_z(z_1) = 0$; $p_z(z_2) = 0$. Фазовые факторы γ_1 и γ_2 приближенно учитывают поведение волновой функции на границах классически доступной области и равны $\gamma = 1/2$ для бесконечно высокой потенциальной стенки (волновая функция в точке поворота равна нулю) и $\gamma = 1/4$ для плавного потенциала. Как и выше энергетический спектр определяется спектром разрешенных значений p_z и законом дисперсии, устанавливающим связь между кинетической энергией (равной разности полной и потенциальной энергий) и квазиимпульсом. Для параболического закона дисперсии $\varepsilon = p^2/2m$ такая связь дается формулой

$$p_z(z) = \sqrt{2m\left(\varepsilon - U(z) - \frac{p_s^2}{2m}\right)}, \tag{5.21}$$

из которой видно, что аддитивность энергий движения вдоль и поперек двумерного слоя сохраняется при любой форме потенциала. Для кейновского закона дисперсии $\varepsilon = \sqrt{p^2s^2 + m^2s^4} - ms^2$ поперечная компонента квазиимпульса связана с энергией соотношением

$$p_z = \frac{1}{s}\sqrt{(\varepsilon - U(z))\,(\varepsilon - U(z) + 2ms^2) - p_s^2s^2}, \tag{5.22}$$

откуда при $m \to 0$ (квази-ультрарелятивистский предел) следует и соответствующее выражение для случая линейного дисперсионного соотношения

$$p_z = \frac{1}{s}\sqrt{(\varepsilon - U(z))^2 - p_s^2s^2}. \tag{5.23}$$

(выражение (5.21) также может быть получено из (5.22) стандартным переходом к нерелятивистскому пределу $s \to \infty$). Таким образом, вычисление электронного спектра в квазиклассическом приближении при произвольном одномерном потенциале в полупроводнике с кейновском зако-

ном дисперсии (переход к простейшим случаям квадратичной и линейной дисперсий легко осуществляется указанным выше способом $s \to \infty$ или $m \to 0$) сводится с простому интегрированию

$$\int\limits_{z_1}^{z_2} \sqrt{(\varepsilon - U(z))\,(\varepsilon - U(z) + 2ms^2) - p_s^2 s^2} dz = \pi s \hbar (i + \gamma_1 + \gamma_2), \quad (5.24)$$

(за исключением случая прямоугольной ямы, координаты точек поворота, определяемые корнями подынтегрального выражения, также зависят от энергии), что значительно проще задачи на нахождение собственных значений и собственных функций уравнения Шредингера.

Из выражения (5.24) следует, что как и в случае прямоугольной ямы, задача вычисления электронного спектра в произвольной яме на основе материала с непараболическим спектром не может быть сведена к вычислению уровней размерного квантования с последующим добавлением кинетической энергии свободного движения в двумерной плоскости. Движения вдоль и поперек ограничивающего потенциала взаимозависимы. Законы дисперсии в подзонах размерного квантования $\varepsilon(p_s)$ в результате не совпадает с законом дисперсии исходного материала, как это имеет место в случае квадратичного закона дисперсии. Присутствие одномерного ограничивающего потенциала приводит к перенормировке параметров закона дисперсии в двух других направлениях.

В заключение этого раздела приведем удобную для вычисления спектра в квазиклассически самосогласованных потенциальных ямах модификацию правила квантования Бора-Зоммерфельда. Как мы видели в разд. 3.2.2, в случае независящей от координаты концентрации легирующей примеси в отдельных областях гетероструктур, уравнение Пуассона (5.20), определяющее такой самосогласованный потенциал, не содержит координаты в явном виде и нахождение его первого интеграла сводится к вычислению определенного интеграла. В результате мы получаем простое выражение для производной потенциала (электрического поля) как функции потенциала (формула 3.53). В то же время определение координатной зависимости потенциала, требует дополнительного интегрирования, дающего к тому же не зависимость $U(z)$, а $z(U)$. Нетрудно однако

видеть, что для квазиклассического квантования достаточно знать зависимость $\frac{dU}{dz}(U)$. Действительно, заменой переменных в интеграле (переходом к интегрированию по потенциалу), правило Бора-Зоммерфельда модифицируется к виду

$$\int_{U(z=0)}^{U(k_z=0)} k_z\left(\varepsilon, k, U\right)\left(\frac{dU}{dz}\right)^{-1} dU = \pi\left(i + \frac{3}{4}\right). \tag{5.25}$$

Использование этой модификации существенно упрощает вычисление спектра подзон размерного квантования в классически самосогласованных квантовых ямах, а в случае численных расчетов кардинально уменьшает время вычислений.

5.4 Одиночный гетеропереход

Рассмотрим энергетический спектр электронов в квантовой яме одиночного $p-n$ гетероперехода, образованной разрывом зоны проводимости Δ_c на гетерогранице и электростатическим потенциалом в глубине $p-$ области структуры (рис. 4.4). Классически рассчитанное распределение потенциала в такой системе описывается формулой (4.30). При отсчете координаты от гетерограницы вглубь $p-$ области, а энергии от дна потенциальной ямы выражение для потенциальной энергии электрона преобразуется к виду

$$U(z) = u_p - (\varphi_p - \varphi(z)) = u_p \frac{z}{L_p}\left(2 - \frac{z}{L_p}\right) \quad (0 \le z \le L_p), \tag{5.26}$$

$$U(z) = u_p \quad (z > L_p) \tag{5.27}$$

При достаточно сильном легировании $p-$ области (это одновременно обеспечивает и малую ширину потенциальной ямы, делая ее квантовой) распределение потенциала определяется в основном зарядом примесей, а вкладом электронов в яме в уравнение Пуассона можно пренебречь. Таким образом вид потенциальной ямы не будет зависеть от закона дисперсии зоны проводимости. Последний, однако, как это уже неоднократно

подчеркивалось, существенен при расчете электронного спектра в квантовой яме. Рассмотрим в качестве примера, как и для прямоугольной ямы, случаями квадратичного и линейного законов дисперсии.

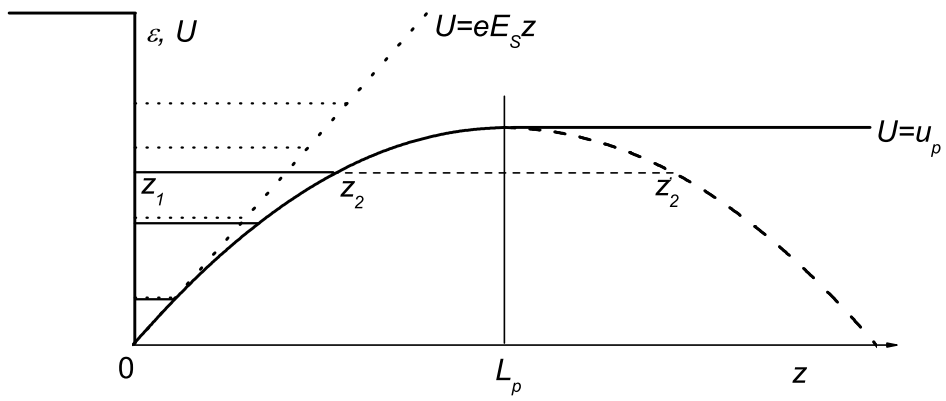

Рис. 5.1: К квантованию в одиночном $p-n$ гетеропереходе. Показаны классический потенциал (сплошная линия), потенциал (5.26) (пунктир) и линейный потенциал в приближении треугольной ямы (5.31) (точечная линия).

Параболический закон дисперсии

Поскольку на гетерогранице ($z_1 = 0$) волновая функция близка к нулю, а во второй точке поворота потенциал плавный, то фазовый фактор для такой ямы $\gamma = 1/2 + 1/4 = 3/4$. В результате формула Бора-Зоммерфельда для потенциальной ямы в одиночном гетеропереходе на основе полупроводников с квадратичным спектром принимает вид

$$\int\limits_0^{z_2} \sqrt{2m\left(\varepsilon_\perp - u_p \frac{z}{L_p}\left(2 - \frac{z}{L_p}\right)\right)} dz = \pi\hbar(i + 3/4) \qquad (5.28)$$

(поскольку для параболической зоны $\varepsilon = \varepsilon_\perp + p^2/2m$, достаточно определить уровни размерного квантования в одномерной яме $\varepsilon_\perp = \varepsilon_i(0)$).

Линейный закон дисперсии

Аналогично для линейного закона дисперсии получаем

$$\int\limits_0^{z_2} \sqrt{\left[\varepsilon - u_p \frac{z}{L_p}\left(2 - \frac{z}{L_p}\right)\right]^2 - p^2 s^2} dz = \pi s\hbar(i + 3/4) \qquad (5.29)$$

Хотя интегралы в (5.28) и (5.29) легко вычисляются[2] и выражаются в элементарных функциях, мы не будем выписывать получающиеся уравнения для спектра ввиду их громоздкости, ограничившись в следующем разделе более простой аппроксимацией для потенциала. Уравнение для линейного закона дисперсии (5.29) существенно упрощается для случая дна двумерных подзон $\varepsilon = \varepsilon(p_s = 0)$

$$\int\limits_0^{z_2} \left[\varepsilon - u_p \frac{z}{L_p}\left(2 - \frac{z}{L_p}\right)\right] dz = \pi s\hbar(i + 3/4), \tag{5.30}$$

однако и здесь получается уравнение третьей степени относительно величины $\sqrt{1 - \varepsilon/u_p}$, так что аналитическое выражение для уровней размерного квантования также довольно громоздко и его получение может быть предложено в качестве задачи (см. задание 2.).

5.4.1 Модель треугольной потенциальной ямы

Нетрудно видеть, что потенциал (5.26) вблизи границы раздела ($z << L_p$) близок к линейному $U(z) \approx 2u_p z/L_p$. Поскольку на самой гетерогранице имеется близкий к вертикальному барьер, то потенциальная яма гетероперехода вблизи ее дна может быть аппроксимирована треугольной

$$U(z) = e\surd_s z \quad (z > 0), \tag{5.31}$$

$$U(z) \approx \infty \quad (z \leq 0), \tag{5.32}$$

где $e\mathrm{E}_s = 2u_p/eL_p$ - величина электрического поля в $p-$ области на гетерогранице. Приближение треугольной потенциальной ямы, игнорирующее экранирование электрического поля зарядами в глубине полупроводника ($e\mathrm{E}(z) = const = e\mathrm{E}_s$), достаточно часто используется при квантовании спектра в полупроводниковых наноструктурах, поскольку, с одной стороны, дает простое аналитическое выражение для спектра, а, с другой, приводит к неплохим оценкам уровней размерного квантования для нижних подзон.

[2]При вычислении этих интегралов необходимо помнить, что решения, соответствующие фиктивным точкам поворота $z_i > L_p$ (рис. 5.1) должны быть, согласно (5.26) и (5.27), отброшены.

Для параболического закона дисперсии уравнение (5.28) в этой аппроксимации для потенциала переходит в

$$\int\limits_0^{z_2} \sqrt{2m\left(\varepsilon_\perp - e\mathrm{E}_s z\right)}dz = \pi\hbar(i+3/4). \tag{5.33}$$

Из условия равенства нулю компоненты p_z (подкоренного выражения в интеграле) для координаты второй точки поворота (первая точка поворота $z=0$ совпадает с гетрограницей) получаем

$$z_2 = \varepsilon_\perp / e\mathrm{E}_s, \tag{5.34}$$

что после подстановки в (5.33) приводит к уравнению

$$\sqrt{2me\mathrm{E}} \int\limits_0^{z_2} \sqrt{\left(\frac{\varepsilon_\perp}{e\mathrm{E}_s} - z\right)}dz = \pi\hbar(i+3/4), \tag{5.35}$$

решение которого

$$\varepsilon_{\perp i} = \left[\frac{\hbar}{\sqrt{2m}} e\mathrm{E}_s \frac{3}{2}\pi\left(i+\frac{3}{4}\right)\right]^{\frac{2}{3}} \tag{5.36}$$

определяет уровни размерного квантования и, тем самым, электронный спектр в треугольной яме

$$\varepsilon = \varepsilon_{\perp i} + \frac{p_s^2}{2m}. \tag{5.37}$$

В отличие от бесконечно глубокой прямоугольной ямы постоянной ширины, в которой расстояние между уровнями размерного квантования увеличиваются с ростом энергии (согласно (5.11) $\varepsilon_{\perp i} \propto i^2$), для треугольной ямы, ширина которой увеличивается с энергией, уровни сгущаются по мере роста индекса подзоны (зависимость $\varepsilon_{\perp i}(i)$ сублинейна).

Сравним квазиклассический спектр уровней размерного квантования (5.36) с точными решениями уравнения Шредингера для линейного потенциала (движение в однородном электрическом поле)

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial z^2} - e\mathrm{E}_s z\right]\psi(z) = \varepsilon\psi(z) \tag{5.38}$$

удовлетворяющего граничным условиям $\psi(0) = 0$, $\psi(\infty) = 0$. Это уравнение тривиальной заменой переменных легко сводится к уравнению Эйри

$$\frac{\partial^2 y}{\partial x^2} = xy. \tag{5.39}$$

Точные значения энергии, определяемые нулями функции Эйри $Ai(z) = 0$, даются для первых трех подзон тем же выражением (5.36), в котором $i + 3/4$ следует заменить на 0,7587, 1,7540, 2,7575. Отсюда видно, что даже для основной подзоны отличие квазиклассического результата от точного решения немногим выше 1%.

Совершенно аналогично получаем уравнение для квазиклассического спектра в случае линейного закона дисперсии

$$\int\limits_0^{z_2} \sqrt{(\varepsilon - e\mathrm{E}_s z)^2 - p_s^2 s^2} dz = \pi s\hbar(i + 3/4). \tag{5.40}$$

Для дна двумерных подзон $\varepsilon(p_s = 0)$ (точка поворота по-прежнему определяется выражением (5.34) $z_2 = \varepsilon/e\mathrm{E}$) решение этого уравнения тривиально

$$\varepsilon_{\perp i} = \left[2\pi s\hbar e\mathrm{E}_s \left(i + \frac{3}{4}\right)\right]^{\frac{1}{2}}. \tag{5.41}$$

Как видим, при линейной дисперсии уровни размерного квантования в том же приближении треугольной потенциальной ямы слабее зависят от номера подзоны и электрического поля, чем уровни в полупроводнике с квадратичным спектром. Отметим, что, используя соотношение $m = E_g/2s^2$, уровни (5.36) для параболического закона дисперсии могут быть выражены через те же переменные, что и в (5.41)

$$\varepsilon_\perp = \left[\frac{3}{4\sqrt{E_g}} 2\pi s\hbar e\mathrm{E}_s \left(i + \frac{3}{4}\right)\right]^{\frac{2}{3}}. \tag{5.42}$$

Напомним, что электрическое поле в приведенных выражениях есть поле на границе гетероперехода $\mathrm{E}_s = 2u_p/eL_p$, т.е. максимальное поле в гетеропереходе. Учет экранирования уширяет потенциальную яму, в виду чего реальные уровни размерного квантования лежат ниже тех,

что дает аппроксимация треугольного потенциала. Очевидно, что отличия наиболее существенны для высоколежащих подзон. Поскольку линейный потенциал в рассматриваемой аппроксимации неограниченно растет с координатой z, а реальный потенциал ограничен величиной $U(z > L_p) = u_p$ (см. (5.27)), то решения, соответствующие подзонам, дно которых $\varepsilon_i(p_s = 0)$ лежит выше u_p, вообще не имеют физического смысла.

5.5 Структуры с δ-легированными слоями

Как мы видели, ход потенциала в n-области гетероперехода определяется выражением

$$\frac{\varphi_n - \varphi}{u_n} = \left(1 - \frac{z}{L_n}\right)^2.$$

Если энергии и z-координату отсчитывать от дна потенциальной ямы, то выражение для потенциальной энергии электрона $U(z) = \varphi(z) - \varphi_n$ в структуре с δ–легированным слоем (близкий к моноатомному слой n-типа в матрице p-типа) принимает вид

$$U(z) = u_n \left(\frac{z}{L_n}\right)^2 = \frac{2\pi e^2 N_D}{\varepsilon} z^2. \tag{5.43}$$

Таким образом, мы имеем дело с потенциалом гармонического осциллятора, гамильтониан которого

$$H = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2} z^2 \tag{5.44}$$

дает хорошо известный эквидистантный спектр

$$\varepsilon_\perp = \hbar\omega \left(n + \frac{1}{2}\right). \tag{5.45}$$

Приравнивая

$$\frac{2\pi e^2 N_D}{\varepsilon} z^2 = \frac{m\omega^2}{2} z^2 \tag{5.46}$$

получаем для частоты осциллятора в нашей системе

$$\omega = \omega_p = \sqrt{\frac{4\pi e^2 N_D}{\varepsilon m}}, \tag{5.47}$$

т.е. выражение, совпадающее с формулой для плазменной частоты.

Нетрудно убедиться, что квазиклассическое квантование в этом случае приводит к точному выражению для спектра. В рассматриваемом случае квадратичной дисперсии условие Бора-Зоммерфельда для параболического потенциала записывается в виде

$$\int\limits_{z1}^{z_2} \sqrt{2m\left[\varepsilon_\perp - \frac{m\omega^2}{2}z^2\right]}dz = m\omega_p \int\limits_{z1}^{z_2} \sqrt{\left[\frac{2\varepsilon_\perp}{m\omega^2} - z^2\right]}dz = \pi\hbar(i + 1/4 + 1/4). \tag{5.48}$$

Обозначая $2\varepsilon_\perp/m\omega^2 = a^2$, находим для точек поворота $z_{1,2} = \pm a$ и приходим к уравнению

$$\int\limits_{-a}^{+a} \sqrt{[a^2 - z^2]}dz = \frac{\pi\hbar}{m\omega_p}(i + 1/2) \tag{5.49}$$

интегрирование которого ($\int = \frac{1}{2}\left[\sqrt{[a^2 - z^2]} + a^2 \arccos\frac{z}{a}\right]$) дает в точности спектр (5.45). Таки образом, электронный спектр в полупроводниковых слоях с δ –легированием имеет в случае квадратичного закона дисперсии вид

$$\varepsilon_\perp = \hbar\omega_p\left(i + \frac{1}{2}\right) + \frac{p_s^2}{2m} = \hbar\sqrt{\frac{4\pi e^2 N_D}{\varepsilon m}}\left(i + \frac{1}{2}\right) + \frac{p_s^2}{2m}. \tag{5.50}$$

Соответствующие квазиклассические выражения для кейновского и линейного законов дисперсии, вывод которых предлагается в качестве упражнения, также совпадают с точными решениями уравнения Шредингера для параболического потенциала.

5.6 Приповерхностная квантовая яма в МДП-структурах

Из результатов раздела 4.2.6 следует, что при низких температурах ширина приповерхностной потенциальной ямы в МДП-структурах при любом законе дисперсии порядка или меньше фермиевской длины де Бройля, указывая на квантовый характер движения локализованных в приповерхностном слое электронов. Как и выше, мы рассмотрим квантование

в квазиклассическом приближении. Из выражений (4.82) и (4.86) легко видеть, что при отсчете энергии от дна потенциальной ямы (дна зоны проводимости при $z = 0$) потенциальная энергия электрона описывается выражением

$$U(z) = \mu_s - \mu(z) = \mu_s \left(1 - \frac{1}{(1 + z/z_s)^4}\right) \tag{5.51}$$

в случае материалов с параболическим законом дисперсии и

$$U(z) = \mu_s - \mu(z) = \mu_s \left(\frac{z}{z + z_s}\right) \tag{5.52}$$

для линейного закона дисперсии, где характерные длины определяются выражениями (4.80) и (4.87).

Полагая, что волновая функция на границе с диэлектриком равна нулю для фазового фактора имеем $\gamma = 1/2 + 1/4$ и правило квантования Бора-Зоммерфельда записывается в виде

$$\int\limits_0^{z_2} \sqrt{E_g \left[\varepsilon_\perp - \mu_s \left(1 - \frac{1}{(1 + z/z_s)^4}\right)\right]} dz = \pi s \hbar (i + 3/4) \tag{5.53}$$

для полупроводника с квадратичной дисперсией и

$$\int\limits_0^{z_2} \sqrt{\left[\varepsilon - \mu_s \left(\frac{1}{1 + z/z_s}\right)\right]^2 - p_s^2 s^2} = \pi s \hbar (i + \frac{3}{4}) \tag{5.54}$$

для материалов с законом дисперсии $\varepsilon = ps$, где точки поворота z_2 определяются из условия равенства нулю подкоренных выражений ($p_z = 0$).

5.6.1 Модель треугольной потенциальной ямы

Как и в случае квантовой ямы в одиночном гетеропереходе для невысоких индексов подзон (уровней, лежащих вблизи дна потенциальной ямы) поверхностный потенциал может быть аппроксимирован моделью треугольной потенциальной ямы. Величина электрического поля на поверхности полупроводника с квадратичным законом дисперсии при данном поверхностном изгибе зон μ_s, как видно из выражения (5.51) для

$U(z)$, равна

$$\mathrm{E}_s = \frac{1}{e}\,\frac{dU(z)}{dz}\bigg|_{z=0} = \frac{4\mu_s}{ez_s}. \tag{5.55}$$

Уровни размерного квантования в рассматриваемом приближении даются простой подстановкой этого выражения в полученную нами выше формулу (5.42), что совместно с выражением (4.80) для характеристического параметра z_s дает

$$\varepsilon_\perp = \frac{\mu_s}{4}\left(\frac{9\beta}{4\sqrt{5}}\right)^{\frac{2}{3}}\left(1+\frac{3}{4}i\right)^{\frac{2}{3}}\left(\frac{E_g}{\mu_s}\right)^{\frac{1}{6}}. \tag{5.56}$$

Поскольку, как указывалось, параметр $\beta \cong 1$, то величина численного коэффициента в первых скобках также близка к единице и двумерные подзоны могут быть с хорошей точностью (в рассматриваемой аппроксимации для потенциала) описаны выражением

$$\varepsilon_i(ps) \approx \frac{\mu_s}{4}\left(1+\frac{3}{4}i\right)^{\frac{2}{3}}\left(\frac{E_g}{\mu_s}\right)^{\frac{1}{6}} + \frac{p_s^2}{2m}. \tag{5.57}$$

Аналогичного, для линейного закона дисперсии поверхностное электрическое поле, согласно (5.52) и (4.87), равно

$$\mathrm{E}_s = \frac{1}{e}\,\frac{dU(z)}{dz}\bigg|_{z=0} = \frac{\mu_s}{ez_s} = \frac{\beta\mu_s^2}{2\pi s\hbar}, \tag{5.58}$$

что совместно с (5.41) дает энергетическое положение дна двумерных подзон ($p_s = 0$)

$$\varepsilon_{\perp i} = \frac{\mu_s}{2}\left[\beta\left(4i+3\right)\right]. \tag{5.59}$$

Поскольку глубина реальной приповерхностной потенциальной ямы равна μ_s, то физический смысл в рассматриваемом приближении треугольного потенциала имеют только решения, удовлетворяющие условию $\varepsilon_\perp < \mu_s$. В достаточно широкозонных полупроводниках, с близким к квадратичному законом дисперсии, последний множитель в (5.57) при типичных изгибах зон слабо зависит от μ_s) и близок к 1.5. В результате число уровней размерного квантования, как нетрудно убедиться, не превышает 5-6.

В случае же узкозонных полупроводников, для которых закон дисперсии уже при небольших энергиях близок линейному, в треугольном потенциале имеется только один уровень размерного квантования $\varepsilon_{\perp 0} \approx \frac{\sqrt{3}}{2}\mu_s$, удовлетворяющий условию $\varepsilon_\perp < \mu_s$. Показательно, что этот вывод справедлив для любой глубины квантовой ямы - выражение (**??**) для $\varepsilon_{\perp i}$ масштабно инвариантно относительно приповерхностного изгиба зон ($\varepsilon_{\perp 0}/\mu_s = \frac{\sqrt{3}}{2} = const$). В действительности, уравнение для точного потенциала (5.54) имеет бесконечное число решений (см. задание 8 к настоящей главе), причем и в этом случае энергии уровней размерного квантования масштабно инвариантны относительно приповерхностного изгиба зон. Таким образом, в отличие от случая широкозонных полупроводников, модель треугольной потенциальной ямы совершенно не адекватна реальной ситуации в приповерхностной потенциальной яме узкозонного полупроводника.

5.7 Задачи

1. Найдите квазиклассический спектр уровней размерного квантования в симметричной квантовой яме композиционной гетероструктуры на основе полупроводника с кейновским законом дисперсии. Ширина ямы a, разрыв зоны проводимости на гетерогранице Δ.

   Ответ:
   $$\varepsilon_i(p_s) = \sqrt{\frac{\pi^2 s^2 \hbar^2}{a^2} i^2 + p_s^2 s^2 + m^2 s^4} - ms^2. \tag{5.60}$$

   Поскольку в квазиклассическом приближении квантовая яма полагается бесконечно глубокой, следует оставить только решения для
   $$i < \frac{a}{\pi s \hbar}\sqrt{\Delta(\Delta + 2ms^2)}. \tag{5.61}$$

2. Докажите, что законы дисперсии в подзонах размерного квантования в полупроводнике с линейным объемным законом дисперсии отличаются от линейных, но близки к ним при больших подзонных квазиимпульсах.

3. Найдите квазиклассические уровни размерного квантования в гетероструктуре с симметричной треугольной ямой ($U(z) = |eEz|$) на основе полупроводника с квадратичной дисперсией.

   Ответ:

$$\varepsilon_{\perp i} = \left[\frac{\hbar}{\sqrt{2m}} e\mathrm{E}_s \frac{3}{4}\pi \left(i + \frac{1}{2}\right)\right]^{\frac{2}{3}} \tag{5.62}$$

4. В гетероструктуре с прямоугольной квантовой ямой, образованной слоем GaAs, между слоями диэлектрика с большой запрещенной зоной, GaAs заменен на InAs. Можно ли изменить параметры ямы таким образом, чтобы уровни размерного квантования остались без изменения, в предположении, что проникновением волновых функций в барьерные слои и непараболичностью законов дисперсии можно пренебречь? Как изменится положение уровней, если учесть: а) непараболичнось и б) туннельное проникновение в барьерные слои?

5. Найдите энергетическое положение дна двумерных подзон $\varepsilon = \varepsilon(p_s = 0)$ в одиночном гетеропереходе с потенциалом (5.26) на основе полупроводника с линейным законом дисперсии.

6. Найти уровни размерного квантования в δ–легированном слое с параболическим потенциалом в предположении линейного закона дисперсии для материала ямы.

7. Найдите энергетическое положение дна двумерных подзон в квантовой яме с δ–легированным слоем для случая кейновского дисперсионного соотношения. Убедитесь, что в предельных случаях малой и высокой энергии полученное решение приводит к правильным результатам для квадратичной и линейной дисперсий.

8. Найдите квазиклассические энергии связи в квазиклассически самосогласованной приповерхностной потенциальной яме глубиной μ_s в нелегированном полупроводнике ($\mu_b = 0$) с линейным законом дисперсии (T=0).

Ответ: При $p_s = 0$ интеграл в (5.54) элементарно вычисляется и для нормированной энергии связи

$$\eta_i = \frac{\mu_s - \varepsilon_{\perp i}}{\mu_s}$$

получаем трансцендентное уравнение

$$-\eta_i - \ln \eta_i = \frac{\beta}{2}\left(i + \frac{3}{4}\right) + 1. \tag{5.63}$$

Это уравнение, как мы убедимся ниже, может быть удовлетворено при сколь угодно больших i (в яме любой глубины имеется бесконечное число уровней размерного квантования). Поскольку $\eta_i < 1$ и уменьшается с ростом i, для высокоэнергетических подзон можно пренебречь членом η_i, что приводит к простой аппроксимации для нормированной на μ_s энергии связи

$$\eta_i = \exp\left[-\frac{\beta}{2}\left(i + \frac{3}{4}\right) - 1\right] \approx 0.253 \exp -\frac{i}{2}, \tag{5.64}$$

где в последнем равенстве учтено, что $\beta \cong 1$. Поскольку в уравнение (5.63) входит только отношение $(\mu_s - \varepsilon_{\perp i})/\mu_s$, спектр уровней размерного квантования масштабно инвариантен относительно глубины ямы.

5.8 Вопросы для самопроверки

1. В какой области гетероструктуры широкощелевой полупроводник p-типа - узкозщелевой полупроводник p-типа с нулевым разрывом зон валентной зоны образуется потенциальная яма и для каких носителей? При каких условиях яма будет заполнена носителями и с каких примесей они переходят в яму?

2. Каковы необходимые и достаточные условия для проявления квантовых эффектов при пространственном ограничении газа зонных носителей в полупроводниках и металлах?

3. По какой причине условия для размерного квантования легко реализуются в полупроводниковых наноструктурах и трудно достижимы в металлах? Сделайте оценки.

4. Какими квантовыми числами характеризуется энергетический спектр в отсутствие внешних полей

   (a) в объемном кристалле?

   (b) в наноструктурах с двумерным электронным газом?

   (c) в квантовой нити?

   (d) в квантовой точке?

5. Перечислите известные Вам (из наиболее часто используемых) приближения при расчете спектра носителей в полупроводниковых наноструктурах.

6. Каким образом можно избежать сложной процедуры самосогласованного решения уравнений Шредингера и Пуассона и когда это оправдано? Для каких полупроводниковых наноструктур наиболее оправдан подход, основанный на квазиклассическом расчете профиля потенциальной ямы с последующим решением уравнения Шредингера для потенциала в этой яме?

7. Где более оправдано пренебрежение квантованием спектра электронов при расчете профиля квантовой ямы - в композиционных гетероструктурах или МДП-структурах?

8. Укажите пределы применимости квазиклассического подхода при расчете энергетического спектра в полупроводниковых наноструктурах и какова его наибольшая погрешность? Назовите примеры ограничивающих потенциалов, для которых квазиклассический подход дает решения, совпадающие или очень близкие к точному решению уравнения Шредингера.

9. Выпишите квазиклассические уравнения, определяющие спектр двумерных подзон в треугольном потенциале для случаев квадратичного и линейного законов дисперсии.

10. Чем отличается пространственное распределение электронной плотности в приповерхностной квантовой яме МДП-структуры, рассчитанное при квазиклассическом и квантовом самосогласовании?

11. При каких условиях 2+1-мерная задача определения спектра двумерных подзон в структурах с одномерным пространственным ограничением может быть сведена к одномерной задаче размерного квантования? Необходимо ли выполнение этих условий при определении энергетического положения дна двумерных подзон?

12. Приведите примеры полупроводниковых наноструктур с двумерным электронным газом, в которых энергетическое расстояние между подзонами размерного квантования

    (a) растет с номером подзоны;

    (b) уменьшается с номером подзоны;

    (c) уровни размерного квантования эквидистантны.

    Зависит ли ответ на этот вопрос от законов дисперсии в исходных трехмерных кристаллах?

13. Как отличаются энергии связи в квантовой яме гетероструктур на основе полупроводников, отличающихся эффективной массой при условии, что ограничивающий потенциал одинаков? Может ли быть выполнено последнее условие в приповерхностных ямах на основе МДП-структур?

14. Каков вид ограничивающего потенциала в одиночном гетеропереходе и как ведут себя энергетические расстояния между подзонами размерного квантования с ростом индекса подзоны - уровни сгущаются или разрежаются? Сравните со случаем композиционной гетероструктуры с прямоугольным потенциалом.

15. Каково относительное энергетическое положение уровней размерного квантования с одинаковыми квантовыми числами в параболическом ограничивающем потенциале в полупроводнике со стан-

дартной зоной проводимости и в кейновском полупроводнике, если дно зон и эффективные массы на дне зон совпадают?

16. Изобразите схематически (в координатах энергия-квазиимпульс) энергетический спектр электронов в квантовой нити с прямоугольным ограничивающим потенциалом высотой V, создающим три связанных состояния. В каких состояниях могут находиться электроны с одной и той же полной энергией? Рассмотрите энергии выше и ниже верхнего края потенциальной ямы. Возможны ли переходы между этими состояниями с одной энергией? Если да, то какие процессы ответственны за такие переходы?

17. Движение электронов двумерного газа со стандартным спектром и эффективной массой m ограничивают в направлении оси x областью размерном a. Как модифицируется электронный спектр, если волновая функция вне этой области равна нулю?

18. Двумерный электронный газ с квадратичным законом дисперсии помещают в прямоугольный потенциальный ящик с размерами a и b с бесконечно высокими потенциальными стенками. Как модифицируется энергетический спектр?

19. Что можно сказать о вырождении энергетических уровней двумерного газа, помещенного в квадратный потенциальный ящик с бесконечно высокими стенками (закон дисперсии полагаем квадратичным)? Попробуйте найти состояние, вырожденное с уровнем $\varepsilon_{4,7}$.

Глава 6

Полуклассическая и квантовая механика электрона проводимости в электрическом и магнитном поле

Электрон проводимости является квантовым объектом, и его законы движения могут быть рассмотрены только на основании квантовомеханического подхода. Только квантовая механика объясняет “свободу “ электрона в кристалле - способность перемещаться на макроскопические расстояния в кристаллической решетке и позволяет ввести такие понятия как квазиимпульс, закон дисперсии и т.д. Однако , если такие понятия сформулированы, можно говорить об электроне в решетке как о свободной частице с законом дисперсии $\varepsilon(\mathbf{p})$ (($\varepsilon(\mathbf{p})$ следует рассматривать как кинетическую энергию частицы, в дальнейшем зависимость $\varepsilon(\mathbf{p})$ полагаем известной). Если размеры траектории значительно превышают постоянную решетки, т.е. внешние поля малы по сравнению с внутри- и межатомными, применим метод эффективной массы. Для материалов с квадратичным изотропным законом дисперсии уравнение Шредингера для огибающей волновой функции в этом случае записывается в виде

$$[\frac{\hat{\mathbf{p}}^2}{2m} + U(r)]\phi(\mathbf{r}) = \varepsilon\phi(\mathbf{r}), \tag{6.1}$$

совпадающим по форме с уравнением Шредингера для свободного электрона, но с перенормированной эффективной массой. В случае материалов с произвольным законом дисперсии эффективное уравнение Шредингера эквивалентно (6.1) с точностью до замены кинетического члена $\hat{\mathbf{p}}^2/2m$ на $\varepsilon(\hat{\mathbf{p}})$

$$[\varepsilon(\hat{\mathbf{p}}) + U(\mathbf{r})]\phi(\mathbf{r}) = \varepsilon\phi(\mathbf{r}). \tag{6.2}$$

6.1 Классическая динамика частицы с произвольным законом дисперсии

Существует широкий класс явлений, когда поведение отдельного электрона хорошо описывается классическими уравнениями движения. В то же время далеко не очевидно, что поведение ансамбля электронов, подчиняющихся статистике Ферми–Дирака, также можно описать классически. Если принцип Паули так сильно влияет на статистику электронов, почему бы ему не отразиться столь же кардинально на их динамике? В действительности, для невзаимодействующих электронов это не так - динамическое поведение системы электронов полностью определяется динамикой составляющих его носителей, что это является прямым следствием теоремы Лиувилля.

Полуклассическая теория позволяет предсказать как в отсутствии столкновений меняются со временем координата и квазиимпульс (волновой вектор) под воздействием внешних полей. [1] Причем такое предсказание можно сделать исходя лишь из закона дисперсии, не используя никакой другой информации о периодическом потенциале решетки. При известном законе дисперсии состояние электрона в приближении метода эффективной массы описывается его координатой, квазиимпульсом и номером энергетической зоны. Последний в рамках этой модели есть интеграл движения (межзонные переходы запрещены). Изменение со временем координаты и квазиимпульса электрона проводимости опре-

[1] когда мы говорим, что электрон имеет определенную координату и импульс мы имеем в виду волновой пакет, ширина которого много меньше размеров зоны Бриллюэна.

деляются уравнениями движения

$$\frac{d\mathbf{p}}{dt} = e\left\{\mathbf{E} + \frac{1}{c}\left[\mathbf{v}\times\mathbf{H}\right]\right\} \tag{6.3a}$$

$$\frac{d\mathbf{r}}{dt} = \mathbf{v} = \nabla_p\varepsilon. \tag{6.3b}$$

Несмотря на формальное сходство этих уравнений с уравнениями для свободного электрона, описываемое ими движение электрона в кристалле отличается от движения свободной частицы. Во-первых, $\mathbf{p}$ здесь не импульс, а квазиимпульс, определенный с точностью до вектора обратной решетки, т.е. изменяющийся в пределах первой зоны Бриллюэна. Во-вторых, вследствие периодической зависимости энергии от квазиимпульса, скорость также есть периодическая функция $\mathbf{p}$. Это придает много своеобразия движению электрона кристалла во внешних полях. В частности ниже мы увидим, что электрическое поле приводит к финитному характеру движения электрона проводимости в пределах разрешенной энергетической зоны, и, следовательно, оказывает квантующее действие на состояния электрона в кристалле, тогда как движение свободного электрона в электрическом поле не ограничено и спектр является непрерывным.

6.2 Движение в постоянном и однородном электрическом поле

В случае постоянного и однородного электрического поля уравнение движения принимает вид

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E}, \tag{6.4}$$

решение которого

$$\mathbf{p}(t) = \mathbf{p}(0) + e\mathbf{E}t \tag{6.5}$$

описывает движение в p пространстве с постоянной и одинаковой для всех носителей скоростью $e\mathbf{E}$. По виду формула (6.5) указывает на неограниченное возрастание со временем $\mathbf{p}$. Однако в периодическом потенциале состояния, отличающиеся на вектор обратной решетки, физически полностью эквивалентны, так что утверждение о "неограниченном

возрастании" $\mathbf{p}$ в данном случае лишено смысла. Достигнув границы зоны Бриллюэна электрон попадает и на противоположную ее границу, откуда вновь начинается линейное во времени возрастание квазиимпульса. Если электрическое поле параллельно одному из векторов обратной решетки, квазиимпульс будет меняться периодически, как это показано на рис. 6.1.

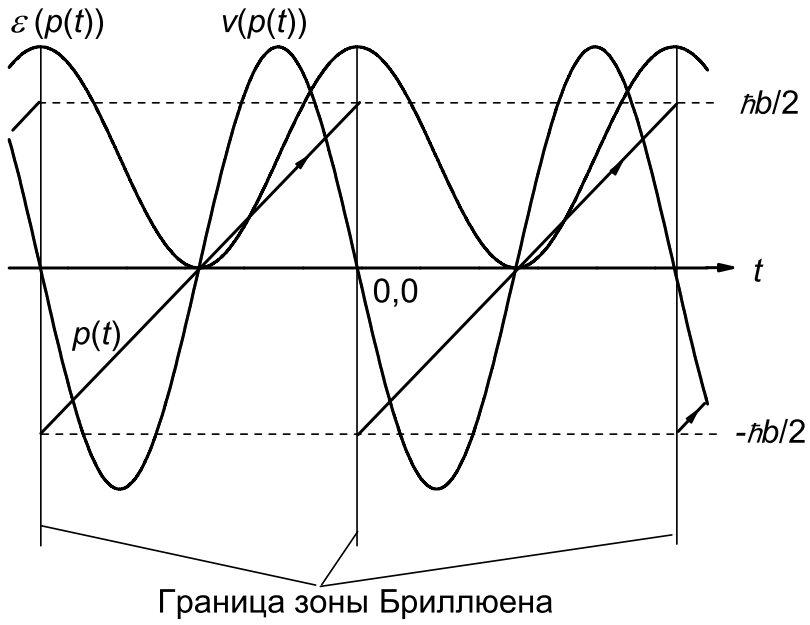

Рис. 6.1: Эволюция во времени квазиимпульса p, энергии ε и скорости $v = |\nabla_p \varepsilon|$ электрона проводимости в постоянном и однородном электрическом поле, иллюстрирующая периодический характер движения в координатном пространстве.

Соответственно, периодически меняется и кинетическая энергия электрона $\varepsilon(p)$. Поскольку в нижней и верхней областях энергетической зоны $\nabla_{\mathbf{p}}\varepsilon$ отличается по знаку, скорости электронов направлены в противоположные стороны. Зависимости кинетической энергии и скорости от времени также проиллюстрированы на рис. 6.1 (ввиду линейного характера зависимости $\mathbf{p}(t)$ (6.5), эти рисунки, с точностью до постоянного множителя, в равной мере иллюстрируют зависимости ε и v от $\mathbf{p}$). Хотя скорость линейно зависит от $\mathbf{p}$ вблизи дна зоны, при приближении к границе зоны Бриллюэна она достигает максимума, а затем уменьшается, обращаясь в ноль на самой границе. Уменьшение скорости с ростом квазиимпульса вблизи потолка энергетической зоны означает, что

ускорение электрона противоположно приложенной внешней силе! Такое необычное поведение обязано исключительно воздействию периодического решеточного потенциала, который хотя и не входит в явном виде в полуклассические уравнения движения (6.3), тем не менее учтен через закон дисперсии $\varepsilon(\mathbf{p})$.

Для рассматриваемых нестационарных состояний из закона сохранения энергии

$$\varepsilon(\mathbf{p}) - e\mathbf{E}\cdot\mathbf{r} = const$$

следует, что, поскольку для электрона в периодическом потенциале решетки $\varepsilon(\mathbf{p})$ есть периодическая функция, ограниченная в пределах энергетической зоны, а поле постоянно во времени и пространстве, электрон будет совершать вдоль поля периодическое движение по траектории (блоховские колебания)

$$e\mathbf{E}\cdot(\mathbf{r}(t) - \mathbf{r}(0)) = \varepsilon(\mathbf{p}(0) + e\mathbf{E}t) - \varepsilon(\mathbf{p}(0)). \tag{6.6}$$

Таким образом, периодичность $\varepsilon(\mathbf{p})$ приводит к тому, что движение электрона в периодическом потенциале (идеальной!) решетки при наложении постоянного электрического поля финитно и периодично в координатном пространстве. Это совершенно не похоже на поведение свободных электронов в электрическом поле, для которых скорость пропорциональна импульсу и растет со временем (в нерелятивистском пределе неограниченно и линейно), так что движение электрона в координатном пространстве инфинитно. [2]

Период блоховских колебаний, т.е. время, за которое электрон проходит всю зону Бриллюэна и возвращается в исходное состояние, определяется уравнением (6.5) при $p(t) - p(0) = \Delta p_E$, где Δp_E – период функции $\varepsilon(\mathbf{p})$ в направлении электрического поля, что дает $T = \Delta p_E/eE$. Поскольку $\Delta p_E/eE = 2\pi\hbar/a_E$, где a_E - период решетки в направлении электрического поля, то частота колебаний равна $\omega_E = 2\pi eE/\Delta p_E \approx eEa/\hbar$,

[2]При стремлении решеточного потенциала U_r к нулю финитное движение электрона в кристалле должно переходить в инфинитное движение. Чтобы понять каким образом это происходит, достаточно вспомнить, что при $U_r \to 0$ зазоры между энергетическими зонами исчезают и вероятности межзонных переходов в электрическом поле возрастают до единицы.

а их амплитуда $\Delta r_{\max}$, определяемая уравнением (6.6), дается соотношением $\Delta r_{\max} = \Delta\varepsilon_c/eE$, где $\Delta\varepsilon_c$ - ширина энергетической зоны. [3]

Чтобы электрон "почувствовал" размеры зоны (граничные условия на границах зоны Бриллюэна) он должен пройти всю эту зону, и, следовательно, всю энергетическую зону $\varepsilon(p)$, без разрушения траектории, т.е. не испытав столкновений с дефектами решетки. Следовательно, амплитуда блоховских осцилляций должна быть меньше длины свободного пробега (а частота осцилляций больше частоты столкновений с дефектами решетки), что накладывает весьма жесткое условие обнаружения осцилляционного движения электронов в координатном пространстве и квантования его спектра в зоне. [4]. При всех разумных значениях поля амплитуда колебаний из-за больших значений ε_c (следствие малости постоянной решетки) чрезвычайно велика. При ширине зоны $\Delta\varepsilon \sim 1$ эВ в типичных для металлов полях $E \sim 10^{-6} \div 10^{-3}$ В/см размер блоховской траектории $\Delta r \sim 10^3 - 10^6$ см, т.е. не только на порядки превышает длину свободного пробега, но и типичные размеры кристаллов. Значительно большие поля, до $E \sim 10^6$ В/см, можно создать в достаточно чистых полупроводниках, но и в этом случае период блоховских осцилляций по крайней мере на порядок выше типичных времен релаксации импульса (на малых же участках траектории между столкновениями электрон, конечно, движется поступательно). Проявления блоховских колебаний можно ожидать в полупроводниковых сверхрешетках, в которых период искусственного периодического рельефа потенциала может быть сделан значительно бóльшим, так что и ширина сверхрешеточных мини-зон мала по сравнению с шириной разрешенных зон материнского кристалла.

[3]Локализация (периодического) движения электронов в области пространства $\Delta r_{\max}$, размеры которой определяются электрическим полем, приводит к квантованию спектра – в энергетической зоне возникают эквидистантные уровни – уровни Ванье, отстоящие друг от друга на величину $\Delta\varepsilon = \hbar\omega_E \approx eEa$.

[4]Интерес к реализации такого режима движения электрона связан с заманчивой идеей генерации переменного электрического тока в статическим электрическом поле (создания блоховского генератора)

6.3 Классическое движение электрона проводимости в постоянном однородном магнитном поле

Рассмотрим теперь движение электрона в постоянном и однородном магнитном поле. Из уравнения Лоренца, связывающего скорость электрона в пространстве квазиимпульсов $d\mathbf{p}/dt$ со скоростью в координатном пространстве $\mathbf{v} = d\mathbf{r}/dt$ в присутствии магнитного поля,

$$\frac{d\mathbf{p}}{dt} = \frac{e}{c}\left[\mathbf{v} \times \mathbf{H}\right] \tag{6.7}$$

следует, что проекция импульса на направление магнитного поля является интегралом движения, в чем легко убедиться, умножая это уравнение скалярно на $\mathbf{H}$, что дает

$$\frac{d(\mathbf{p} \cdot \mathbf{H})}{dt} = \frac{e}{c}\left[\mathbf{v} \times \mathbf{H}\right] \cdot \mathbf{H}.$$

Поскольку смешанное произведение $[\mathbf{v} \times \mathbf{H}] \cdot \mathbf{H}$ равно нулю, то $d(\mathbf{p} \cdot \mathbf{H})/dt = 0$. Таким образом, скалярное произведение $\mathbf{p} \cdot \mathbf{H}$, т.е. проекция $\mathbf{p}$ на $\mathbf{H}$, сохраняется во времени. В дальнейшем мы будем полагать ось $\mathbf{z}$ направленной вдоль магнитного поля, тогда этот результат можно записать в виде $p_z = const$. С другой стороны, энергия зависит только от квазиимпульса и может, следовательно, меняться только за счет изменения $\mathbf{p}$

$$\frac{d\varepsilon}{dt} = \nabla_p \varepsilon \frac{d\mathbf{p}}{dt} = \frac{e}{c}\mathbf{v} \cdot \left[\mathbf{v} \times \mathbf{H}\right] = 0$$

(здесь мы использовали выражение для средней квантовомеханической скорости $\mathbf{v} = \nabla_p \varepsilon$, а при переходе ко второму равенству уравнение Лоренца (6.7). Этот результат показывает, что и энергия электрона при движении в магнитном поле также сохраняется $\varepsilon(\mathbf{p}) = const$ в полном соответствии с тем, что сила Лоренца перпендикулярна скорости и работы не совершает. Таким образом, траектория электрона в пространстве квазиимпульсов в магнитном поле определяется пересечением изоэнергетической поверхности $\varepsilon(\mathbf{p}) = const$ и плоскости $p_z = const$ (рис. 6.2). В полупроводниках изоэнергетические поверхности для всех представляющих интерес энергий являются замкнутыми и, таким образом, движение

электрона в пространстве квазиимпульсов в магнитном поле финитно и периодично. Поскольку в направлении параллельном магнитному полю электрон в **p**- пространстве (в пространстве квазиволновых векторов $\mathbf{k} = \mathbf{p}/\hbar$) покоится ($dp_z/dt = 0$), его движение в этом пространстве носит чисто двумерный характер (в двумерной системе среднее квантовомеханическое значение p_z вообще равно нулю).

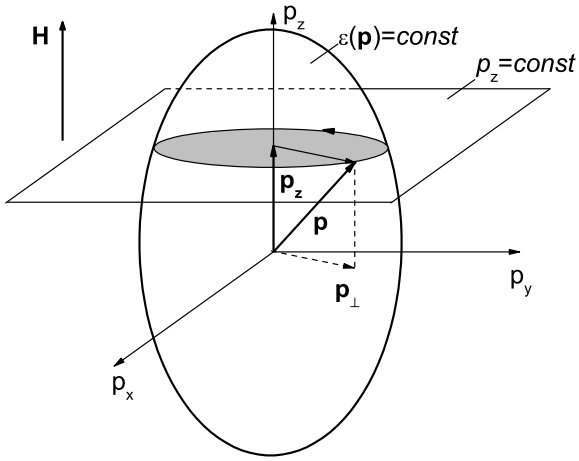

Рис. 6.2: Орбита электрона проводимости в пространстве квазиимпульсов в магнитном поле.

Следует обратить внимание еще на одну существенную разницу в поведении электронов в электрическом и магнитном полях, связанную со статистикой электронного газа и проявляющуюся в системах с вырожденным электронным газом. В силу принципа Паули, изменить энергию, а, следовательно, и модули импульса и скорости, могут только электроны, для которых существуют незанятые вышележащие энергетические уровни. Следовательно, в электрическом поле, которое всегда увеличивает кинетическую энергию заряженной частицы, электроны в состояниях, лежащих глубоко под уровнем Ферми в движении участвовать не могут – для этих состояний уравнение (6.4) несовместимо с принципом Паули. Поскольку магнитное поле не меняет энергии (если поле постоянно и однородно на масштабах порядка радиуса циклотронной орбиты), то для магнитного поля указанное ограничение отсутствует и в циклотронном движении участвуют все электроны.

6.3.1 Циклотронная частота и циклотронная масса

Вычислим частоту периодического движения электрона по траектории в магнитном поле, для чего спроектируем уравнение Лоренца на плоскость (x, y)

$$\frac{dp_s}{dt} = \frac{eH}{c} v_s.$$

Здесь dp_s – элемент дуги траектории в пространстве квазиимпульсов, а $v_s = \sqrt{v_x^2 + v_y^2}$ - скорость в плоскости спроектированной траектории обычного пространства. Интегрирование этого уравнения по плоской траектории C, задаваемой уравнениями $\varepsilon(\mathbf{p}) = const$, $p_z = const$ дает для времени движения

$$t = \int\limits_0^t dt = \frac{c}{eH} \int\limits_C \frac{dp_s}{v_s}.$$

Если траектория замкнутая, то, беря этот интеграл по всему контуру, получаем период обращения

$$T_c = \frac{c}{eH} \oint\limits_C \frac{dp_s}{v_s}.$$

Угловая частота вращения в магнитном поле $\omega_c = 2\pi/T_c$ (циклотронная частота), таким образом, дается формулой

$$\omega_c = \frac{eH}{c} \left(\frac{1}{2\pi} \oint\limits_C \frac{dp_s}{v_s} \right)^{-1}. \tag{6.8}$$

Сравнение этого выражения с циклотронной частотой свободного электрона

$$\omega_c = \frac{eH}{m_0 c}$$

показывает, что величину

$$m_c(E) = \frac{1}{2\pi} \oint\limits_C \frac{dp_s}{v_s} \tag{6.9}$$

(в общем случае она зависит от энергии) естественно назвать циклотронной эффективной массой, которая является функцией только сохраняющихся в магнитном поле величин p_z и ε, что делает ее использование

значительно более удобным, чем использование тензора обратной эффективной массы, который в общем случае зависит от всех компонент вектора $\mathbf{p}$. Выражение для m_c можно представить в другом виде, воспользовавшись тем, что вектор скорости $\mathbf{v}_s$ перпендикулярен плоской кривой, задаваемой уравнениями $p_z = const$ и $\varepsilon(\mathbf{p}) = const$. Вычислим площадь, ограниченную этой замкнутой не пересекающейся кривой

$$\sigma_p(\varepsilon, p_z) = \iint dp_x dp_y = \int d\varepsilon \oint\limits_C \frac{dp_s}{v_s},$$

Отсюда следует, что фигурирующий в выражении для m_c интеграл по замкнутой плоской траектории равен производной площади, ометаемой траекторией, по энергии

$$\oint\limits_C \frac{dp_s}{v_s} = \frac{\partial \sigma_p}{\partial \varepsilon}$$

Отсюда, используя (6.9), видно, что эффективная циклотронная масса может быть определена также через дифференциал $\partial\sigma_p/\partial\varepsilon$

$$m_c = \frac{1}{2\pi}\frac{\partial \sigma_p}{\partial \varepsilon}. \tag{6.10}$$

Поскольку циклотронная частота (и отвечающая ей циклотронная масса) зависит от ε (а в трехмерном случае и от p_z), то, в отличие от газа свободных электронов, электроны в кристалле в общем случае не вращаются с общей для всех электронов частотой. Величина ω_c (ее значение может быть достаточно просто определено экспериментально, например, из циклотронного резонанса) однозначно определяется формой изоэнергетической поверхности и ориентацией плоскости $p_z = const$ (ориентацией магнитного поля), в связи с чем исследование электронного газа в магнитном поле является одним из наиболее мощных средств для восстановления формы изоэнергетических поверхностей и определения законов дисперсии в полупроводниках и металлах. Представляет интерес сопоставить формулу для циклотронной эффективной массы (6.9) с полученным ранее выражением для эффективной массы плотности состояний (2.45) (см. также Табл. 2.2). Нетрудно видеть, что в двумерном, и только в двумерном, случае эти массы тождественны. Действительно,

интегрирование в соотношении

$$m_D = \frac{1}{2\pi} \int\limits_S \frac{dS}{|\nabla_p \varepsilon|}$$

ведется по изоэнергетической поверхности S, а в выражении для m_c

$$m_c = \frac{1}{2\pi} \oint\limits_C \frac{dp_s}{v_s}$$

по траектории электрона C, определяемой пересечением этой поверхности с плоскостью $p_z = const$. Очевидно, что в двумерном случае траектория в магнитном поле и изоэнергетическая поверхность есть одно и то же. С другой стороны, в двумерной системе перпендикулярная магнитному полю скорость v_s совпадает с полной групповой скоростью $v = |\nabla_p \varepsilon|$ и, следовательно, определения для m_D и m_c в двух измерениях эквивалентны.

В разд .2.3.1 мы видели, что особенностью двумерной системы с квадратичной дисперсией является независимость плотности состояний (и, следовательно, эффективной массы плотности состояний) от энергии. В дальнейшем, в разд. 2.3.4 было доказано более общее утверждение об эквивалентности в двумерной системе эффективной динамической массы (легко видеть, что она совпадает с циклотронной массой) и эффективной массы плотности состояний для произвольного, но изотропного спектра. Теперь мы видим, что это утверждение носит общий характер – в двумерной (и только в двумерной) системе $m_D(E) \equiv m_c(E)$ при произвольном законе дисперсии.

6.3.2 Связь двумерного движения в координатном и импульсном пространствах

Мы до сих пор говорили о движении электронов в пространстве импульсов. Уравнение Лоренца фактически устанавливает связь скорости электрона в пространстве квазиимпульсов $d\mathbf{p}/dt$ со скоростью в координатном пространстве $\mathbf{v} = d\mathbf{r}/dt$ и, как нетрудно видеть, дает связь траектории электронов в p пространстве с проекцией траектории в координатном пространстве на плоскость, перпендикулярную магнитному

полю. В случае двумерной системы движение в плоскости перпендикулярной магнитному полю есть единственно возможное движение. Поскольку $\mathbf{v} = d\mathbf{r}/dt$, уравнение (6.7) может быть записано в виде

$$\frac{d\mathbf{k}}{dt} = \frac{e}{c\hbar}\left[\frac{d\mathbf{r}}{dt} \times \mathbf{H}\right] = \frac{eH}{c\hbar}\left[\frac{d\mathbf{r}}{dt} \times \mathbf{1}_H\right], \tag{6.11}$$

где $\mathbf{1}_H$- единичный орт в направлении магнитного поля (здесь удобнее описание в пространстве волновых векторов $\mathbf{k} = \mathbf{p}/\hbar$; хотя теперь $\mathbf{r}$ и $\mathbf{k}$ двумерные векторы, мы для удобства по-прежнему используем трехмерную векторную алгебру). Очевидно, что множитель при векторном произведении во втором равенстве имеет размерность квадрата обратной длины. Эту характерную длину

$$\lambda_H = \sqrt{\frac{c\hbar}{|e|H}}, \tag{6.12}$$

зависящую только от магнитного поля, но не зависящую от параметров материала, называют магнитной длиной. Из уравнения (6.11) сразу следует, что двумерная траектория в обычном пространстве (или ее проекция на плоскость x, y для трехмерной системы) может быть получена из траектории в пространстве волновых векторов путем поворота на угол $\pi/2$ (поскольку в каждый момент времени скорости в обычном и координатном пространствах взаимно перпендикулярны друг другу и магнитному полю) и изменения масштаба - умножением на квадрат магнитной длины (рис. 6.3) [5]. Отсюда, в частности, сразу следует, что направление, период и частота циклотронного движения в координатном пространстве те же самые, что и в k пространстве.

В дальнейшем нам понадобится соотношение между площадями, ограниченными такими подобными друг другу траекториями в обычном про-

[5]Связь движений в координатном и импульсном пространстве вдоль магнитного поля в трехмерной системе в общем случае не носит столь простого характера. Для движения вдоль магнитного поля в координатном пространстве уравнения движения имеют вид

$$z = \int_0^z v_z(t)dz, \quad v_z = \frac{\partial\varepsilon}{\hbar\partial k_z},$$

где в случае произвольного закона дисперсии величина v_z не обязательно постоянна, даже если $k_z = const$. Поэтому, покоящаяся в импульсном пространстве в направлении магнитного поля частица может двигаться в координатном пространстве вдоль поля в общем случае неравномерно.

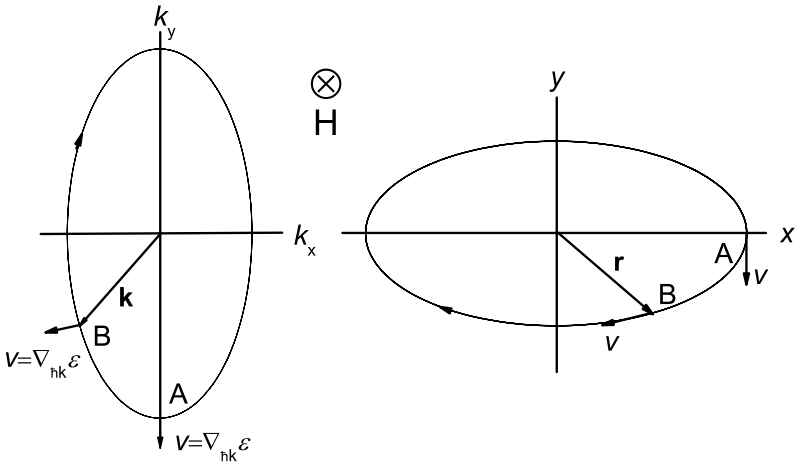

Рис. 6.3: Двумерные траектории электрона проводимости в магнитном поле в координатном пространстве и пространстве квазиволновых векторов.

странстве и пространстве волновых векторов. Геометрически легко ви-

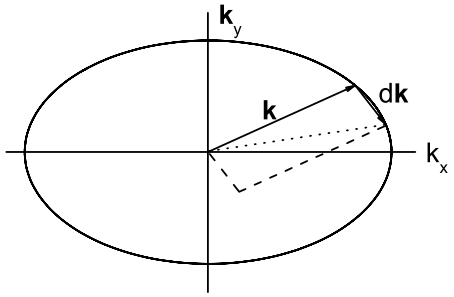

Рис. 6.4: .

деть (рис. 6.4), что площадь в k пространстве, охватываемая циклотронной орбитой равна

$$\sigma_k = \frac{1}{2} \oint \mathbf{k} \times d\mathbf{k}. \tag{6.13}$$

Умножая (6.11) на скаляр dt получаем, что дифференциалы полярных векторов $d\mathbf{r}$ и $d\mathbf{k}$ связаны соотношением

$$d\mathbf{k} = \lambda_H^{-2} \left[d\mathbf{r} \times \mathbf{1}_H \right].$$

Интегрирование уравнения Лоренца во времени дает также соотношение, связывающее радиус векторы в координатном и импульсном пространствах

$$\mathbf{k} = \frac{e}{c\hbar} \left[\mathbf{r} \times \mathbf{H} \right] = \lambda_H^{-2} \left[\mathbf{r} \times \mathbf{1}_H \right] \tag{6.14}$$

(в покомпонентном виде $k_x = \lambda_H^{-2} y, k_y = -\lambda_H^{-2} x$). Используя эти соотношения, получаем из (6.13)

$$\sigma_k = \frac{1}{2} \oint \mathbf{k} \times d\mathbf{k} = \frac{\lambda_H^{-4}}{2} \oint \mathbf{r} \times d\mathbf{r}.$$

Поскольку интеграл в правой части есть удвоенная площадь, обметаемая траекторией в координатном пространстве, то отсюда следует искомое соотношение

$$\frac{\sigma_r}{\sigma_k} = \lambda_H^4 = \left(\frac{c\hbar}{eH}\right)^2. \tag{6.15}$$

6.4 Квантовая механика электрона в магнитном поле

Поскольку для двумерного электронного газа движение в магнитном поле носит финитный характер (в случае, когда время жизни на циклотронной орбите значительно больше времени между столкновениями), энергетический спектр электронов квантуется. Для определения спектра необходимо решить уравнение Шредингера, для чего необходимо знать гамильтониан заряженной частицы в электромагнитном поле. Последний должен быть получен из гамильтонова описания классического движения, а именно заменой импульса частицы в соответствующей классической функции Гамильтона оператором импульса. В отличие от электрического поля, влияние которого описывается введением в уравнение Шредингера зависящего от координаты электростатического потенциала $\varphi(r)$ (в случае однородного поля $\varphi(r) = e\mathbf{E} \cdot \mathbf{r}$), обусловленная магнитным полем сила Лоренца $\mathbf{F}_L = e\left[\mathbf{v} \times \mathbf{H}\right]/c$ определяется скоростью $\mathbf{v} = \nabla_p \varepsilon(p)$ частицы, а, следовательно, ее импульсом, но не зависит для однородного и постоянного магнитного поля от ее координаты. Таким образом, воздействие магнитного поля должно описываться членами гамильтониана, содержащими импульс частицы, т.е. его кинетической частью $\varepsilon(p)$. С другой стороны, может показаться неожиданным, что в гамильтониан входит векторный потенциал, поскольку статическое магнитное поле не производит работы и не может изменить

энергию частицы. Это “противоречие” объясняется тем, что в магнитном поле входящий в гамильтониан импульс есть полный (обобщенный) импульс $\mathbf{P} = \mathbf{p}_{kin} + \mathbf{p}_{fld} = \mathbf{p} + e\mathbf{A}/c$, являющийся суммой двух членов: обычного импульса $\mathbf{p}_{kin}$ (количества движения $m\mathbf{v}$ для квадратичного закона дисперсии) и потенциального импульса $\mathbf{p}_{fld}$, импульса поля. Сопутствующий частице импульс в электромагнитном поле есть интеграл по объему от вектора Пойтинга

$$\mathbf{p}_{fld} = \frac{1}{4\pi c}\int dV[\mathbf{E}\times\mathbf{H}].$$

В нерелятивистском пределе можно считать, что магнитное поле $\mathbf{H} = rot\mathbf{A}$ обусловлено только внешними источниками, а электрическое поле $\mathbf{E} = -\nabla\varphi$ создается только зарядом самой частицы. Если заряд расположен в точке $\mathbf{r}_i$, то потенциал дается уравнением Пуассона $\nabla^2\varphi = -4\pi e\delta(\mathbf{r}-\mathbf{r}_i)$ и, следовательно

$$p_{fld} = -\frac{1}{4\pi c}\int dV[\nabla\varphi\times\mathbf{rot}\mathbf{A}].$$

Векторное произведение под интегралом преобразуется к виду

$$\nabla\varphi\times\mathbf{rot}\mathbf{A} = \mathbf{A}\times\mathbf{rot}\,(\nabla\varphi) - \mathbf{A}div\,(\nabla\varphi) - (\nabla\varphi)div\mathbf{A} = \mathbf{A}\nabla^2\varphi,$$

где в последнем равенстве учтено, что $\mathbf{rot}\,(\nabla\varphi)$=0 и выбрана поперечная калибровка $div\mathbf{A} = 0$. В результате получаем формулу

$$\mathbf{p}_{fld} = -\frac{1}{4\pi c}\int dV\mathbf{A}\nabla^2\varphi = -\frac{1}{c}\int dV\mathbf{A}e\delta(\mathbf{r}-\mathbf{r}_i) = e\mathbf{A}/c,$$

раскрываюую физический смысл вклада поля в обобщенный (полный) импульс. Из сказанного ясно, что входящий в гамильтонову функцию обычный импульс $\mathbf{p}\equiv\mathbf{p}_{kin}$, будучи выраженным через полный, принимает вид $\mathbf{P}-e\mathbf{A}/c$ (часто записывают как $\mathbf{p}-e\mathbf{A}/c$, молчаливо понимая под символом $\mathbf{p}$ обобщенный импульс; для электрона встречается запись $\mathbf{p}+e\mathbf{A}/c$, когда под зарядом понимают его абсолютную величину $|e|$).

Выражение $\mathbf{P} - e\mathbf{A}/c$ инвариантно относительно любого калибровочного преобразования, поскольку при изменении $\mathbf{A}$ на $\delta\mathbf{A}$ в полевом

импульсе обобщенный импульс $\mathbf{P}$ переходит в $\mathbf{P} + e\delta\mathbf{A}/c$. Для частицы с квадратичным законом дисперсии видно, что гамильтонова функция имеет вид $\mathcal{H} = (\mathbf{P} - e\mathbf{A}/c)^2/2m$, т.е. кинетическая энергия электрона $\mathbf{p}_{kin}^2/2m$, как и должно быть при классическом рассмотрении, не изменяется в магнитном поле. Переход к квантово-механическому описанию производится заменой обобщенного импульса $\mathbf{P}$ на оператор $\hat{\mathbf{P}} = \hbar\hat{\mathbf{K}} = -i\hbar\nabla$, т.е. эффективный гамильтониан в магнитном поле получается заменой $\mathbf{p}$ в законе дисперсии $\varepsilon(p)$ оператором $-i\hbar\nabla - e\mathbf{A}/c = \hbar(\hat{\mathbf{k}} - e\mathbf{A}/c\hbar)$.

Как и в случае однородного электрического поля, гамильтониан взаимодействия с магнитным полем неограничен (содержит члены, пропорциональные r^2- см. ниже), поэтому даже слабое поле резко перестраивает спектр. Для магнитного поля, однако, накладываются значительно более слабые ограничения на возможность наблюдения такой перестройки, чем в случае электрического. Принципиальное отличие между воздействием электрического и магнитного поля заключается в том, что первое увеличивает кинетическую энергию электрона, тогда как второе работы над электроном не совершает и не изменяет его энергию. Поэтому невозможно создать очень сильное электрическое поле вследствие разогрева или разрушения (электрического пробоя) образца, тогда как для магнитного поля таких ограничений в полупроводниках практически не существует. [6]

6.4.1 Квантование в магнитном поле (квадратичный закон дисперсии

Рассмотрение квантовых явлений в магнитном поле удобно начать для материалов, зонный спектр которых описывается квадратичным законом дисперсии. Задача в этом случае полностью аналогична рассмотрению квантового движения свободного электрона в магнитном поле и допус-

[6] В некоторых металлах в сильном магнитном поле может наблюдаться аналог зинеровского электрического пробоя - магнитный пробой, связанный с межзонным туннелированием электрона при достаточно малой щели. Ясно, что условия, при которых возможен электрический или магнитный пробой, несовместимы с полуклассическим описанием, поскольку последнее запрещает межзонное туннелирование, являющееся чисто квантовым явлением, и требует, чтобы энергия электрона оставалась в пределах той зоны, в которой он находился первоначально.

кает точное и в то же время достаточно простое математическое описание. Поскольку, как показано выше при полуклассическом рассмотрении, движение электрона в магнитном поле носит по-существу двумерный характер, мы в последующем рассматриваем двумерный электронный газ. Обобщение на случай трех измерений, как мы увидим, тривиально. Подстановка $\hat{\mathbf{p}} \to \hat{\mathbf{p}} - e\mathbf{A}/c$ в уравнение (6.1) приводит к уравнению Шредингера

$$\frac{\hbar^2}{2m}[(\hat{k}_x - \frac{e}{c\hbar}A_x)^2 + (\hat{k}_y - \frac{e}{c\hbar}A_y)^2]\phi(x,y) = \varepsilon\phi(x,y), \tag{6.16}$$

где мы перешли от квазиимпульса к волновому вектору $\mathbf{k} = \mathbf{p}/\hbar$. Выбирая векторный потенциал в виде $\mathbf{A} = (0, Hx)$ (калибровка Ландау), что соответствует перпендикулярному двумерной плоскости магнитному полю

$$H_z = \frac{\partial A_y}{dx} - \frac{\partial A_x}{dy} = H$$

($H_x = H_y = 0$ – напомним, что в двух измерениях ротор является псевдоскаляром), уравнение приобретает вид

$$\frac{\hbar^2}{2m}[\hat{k}_x^2 + (\hat{k}_y - \frac{eH}{c\hbar}x)^2]\phi(x,y) = \varepsilon\phi(x,y). \tag{6.17}$$

Использованная калибровка Ландау (в этой калибровке компоненты обобщенного волнового вектора $\mathbf{K} = \mathbf{P}/\hbar$ равны: $K_x = k_x, K_y = k_y - eHx/c\hbar$) обеспечивает наиболее простой вид уравнения Шредингера с потенциалом, зависящим только от одной пространственной координаты x, так что переменные разделяются, а волновая функция может быть записана в виде

$$\phi(x,y) = \varphi(x)\varphi(y) = \varphi(x)e^{ik_y y}. \tag{6.18}$$

Раскрывая в (6.17) операторное выражение $(\hat{k}_y - eHx/c\hbar)^2$, действуя операторами $\hat{k}_x = -i\nabla_x$ и $\hat{k}_y = -i\nabla_y$ на волновую функцию $\phi(x,y)$ и сокращая обе стороны уравнения на $\exp ik_y y$ получаем уравнение для волновой функции $\varphi(x)$, описывающей движение в направлении, перпендикулярном векторному потенциалу $\mathbf{A}$

$$\frac{\hbar^2}{2m}[\hat{k}_x^2 + (k_y - \frac{eH}{c\hbar}x)^2]\varphi(x) = \varepsilon\varphi(x), \tag{6.19}$$

где k_y - собственное значение оператора $\hat{k}_y$. Введением магнитной длины λ_H гамильтониан в этом уравнении приводится к виду

$$\hat{\mathcal{H}} = \frac{\hbar^2 \hat{k}_x^2}{2m} + \frac{\hbar^2 \lambda_H^{-4}}{2m}(k_y \lambda_H^2 - x)^2.$$

Замечая, что $\hbar^2 \lambda_H^{-4}/m = m\omega_c^2$, где ω_c – циклотронная частота, и, вводя вместо квантового числа k_y параметр размерности длины $x_0 = \lambda_H^2 k_y$, уравнение для $\varphi(x)$ окончательно сводится к рассмотренному уже нами ранее уравнению Шредингера для одномерного гармонического осциллятора с гамильтонианом

$$\hat{\mathcal{H}} = \frac{\hbar^2 \hat{k}_x^2}{2m} + \frac{m\omega_c^2 (x - x_0)^2}{2},$$

из вида которого ясно, что квантовое число $x_0 = \lambda_H^2 k_y$ играет роль x -координаты центра осциллятора. Таким образом, спектр двумерного электронного газа, характеризуемый в отсутствии внешних полей квантовыми числами k_x и k_y, вырождается в магнитном поле в дискретный спектр

$$\varepsilon_n = \hbar\omega_c (n + \frac{1}{2}) \tag{6.20}$$

(двумерный магнитный “атом” с эквидистантным спектром). Состояния электрона в магнитном поле теперь описываются квантовыми числами - n (номер уровня Ландау) и k_y (или положением центра осциллятора x_0), причем энергия электрона в магнитном поле зависит только от n. В силу произвольности k_y, центр осциллятора при одной и той же энергии может находиться где угодно в образце, т. е. имеет место вырождение по координате центра осциллятора (к вопросу о кратности вырождения мы вернемся позднее). Этого и следовало ожидать, поскольку однородное магнитное поле не зависит от координат и все точки пространства физически эквивалентны.

Нетрудно видеть, что в использованной калибровке $A = (0, Hx)$ [7] координатная зависимость волновой функции (6.18) не согласуется с классической круговой траекторией электрона [8]– в направлении векторного

[7] из симметрии задачи очевидно, что эквивалентный результат (а для спектра, конечно, тот же самый) получится и в калибровке $\mathbf{A} = (-Hx, 0)$

[8] Как мы увидим в Гл. 7, калибровка Ландау дает результат, согласующийся с классической картиной движения в случае скрещенных магнитного и электрического полей.

потенциала $\varphi(y) = \exp ik_y y$ подобна волновой функции для свободного электрона, тогда как в направлении перпендикулярном $\mathbf{A}$ соответствует состояниям, локализованным в параболической потенциальной яме $m\omega_c^2(x-x_0)^2/2$ с центром при x_0. Отсюда однако не следует, что электрон двигается в направлении оси y со скоростью $\hbar k_y/m$, как это имеет место для свободного электрона с волновой функцией того же вида. Правильное определение средней (усредненной по интервалу времени, значительно превышающему циклотронный период) квантовомеханической скорости есть $\mathbf{v} = \nabla_{\mathbf{p}}\varepsilon$. Поскольку же энергия в магнитном поле $\varepsilon_n = \hbar\omega_c(n+1/2)$ не зависит от квазиимпульса, обе компоненты скорости $v_y = \partial\varepsilon/\partial p_y$, $v_x = \partial\varepsilon/\partial p_x$ равны нулю, в полном согласии с классической картиной циклотронного движения с неподвижным центром. В случае свободного электрона энергия $\varepsilon = \hbar^2\mathbf{k}^2/2m$ зависит от импульса, что дает для скорости $v_y = \partial\varepsilon/\partial\hbar k_y$ стандартное выражение $v_y = \hbar k_y/m$ (для рассматриваемого случая квадратичной дисперсии). Более близкое к классическому движению решение может быть получено в калибровке $A = (-Hy/2, Hx/2)$, соответствующей тому же магнитному полю с единственной ненулевой компонентой вдоль оси z. Как и классическое движение заряженной частицы, решения уравнения Шредингера с таким гамильтонианом соответствует двум гармоническим осцилляторам с частотой $\omega_c = eH/mc$ и энергией, определяемой той же формулой, что и при калибровке Ландау.

Наиболее близкое к классической картине решение можно получить, если выбрать аксиально-симметричную калибровку $\mathbf{A} = \frac{1}{2}[\mathbf{H}\times\mathbf{r}]$ (в полярных координатах $A_\varphi = \frac{1}{2}H\rho$, $A_\rho = 0$) и искать решение уравнения Шредингера в полярных координатах в виде

$$\phi(\rho,\varphi) = \rho^{\frac{1}{2}}F(\rho)e^{il\varphi},$$

что приводит к радиальному уравнению

$$\frac{\hbar^2}{2m}\frac{d^2F}{d\rho^2} + (\varepsilon - V)F = 0.$$

Эффективный потенциал в этом уравнении

$$V = \frac{\hbar^2}{2m}\left(\frac{\rho^2}{4\lambda_H^4} + \frac{l}{\lambda_H^2} + \frac{l^2 - 1/4}{\rho^2}\right),$$

как нетрудно убедиться, приравняв нулю производную $\partial V/\partial\rho$, имеет минимум при

$$\rho_0 = \lambda_H(4l^2-1)^{\frac{1}{4}},$$

а по мере удаления от минимума возрастает квадратично $m\omega_c^2(\rho-\rho_0)^2/2$. Таким образом электрон связан в яме очень близкой к гармонической с характеристической частотой ω_c и его спектр описывается теми же дискретными квантовыми уровнями $\hbar\omega_c(l+1/2)$. Прямой подстановкой можно убедиться, что точное решение радиального уравнения имеет вид

$$F(\rho) = \rho^{(l+\frac{1}{2})}\exp\left(-\frac{\rho^2}{4\lambda_H^2}\right),$$

откуда видно, что максимум волновой функции приходится на $\rho_c = \lambda_H\sqrt{2l+1}$, что соответствует классическому радиусу циклотронной орбиты r_c электрона с энергией $\hbar\omega_c(l+1/2)$, а радиальная полуширина функции F^2 равна λ_H. Это меньше радиуса в $\sqrt{2l+1}$ раз и для больших номеров l волновая функция ограничена узкой областью около классической орбиты.

Полученные результаты легко обобщаются на случай трех измерений. Добавление в гамильтониан члена $\hbar^2\hat{k}_z^2/2m$, описывающего третью степень свободы вдоль магнитного поля, приводит при той же калибровке $A = (0, Hx, 0)$, ввиду отсутствия зависимости потенциала от координаты z, к волновой функции $\phi(x,y,z) = \varphi(x)\exp ik_y y\exp ik_z z$, а решение уравнения Шредингера отличается от соответствующего двумерного решения добавлением кинетической энергии в направлении магнитного поля

$$\varepsilon_n = \hbar\omega_c(n+\frac{1}{2}) + \frac{\hbar^2k_z^2}{2m}. \tag{6.21}$$

Центр циклотронной орбиты остается неподвижным в направлении, перпендикулярном магнитному полю ($v_x = v_y = 0$), тогда как вдоль магнитного поля частица двигается со скоростью $v_z = \partial\varepsilon/\partial p_z = \hbar k_z/m$.

6.4.2 Квантование в магнитном поле (произвольный закон дисперсии)

В случае неквадратичного закона дисперсии задача квантования спектра в магнитном поле является далеко нетривиальной. Однако если движение электрона в магнитном поле может рассматриваться как квазиклассическое (а это справедливо в весьма широком круге представляющих интерес ситуаций), решение может быть получено достаточно просто. В общем случае произвольного закона дисперсии эффективный двумерный гамильтониан в магнитном поле имеет вид

$$\hat{\mathcal{H}} = \varepsilon(\hat{k}_x - \frac{eA_x}{c\hbar}, \hat{k}_y - \frac{eA_y}{c\hbar})$$

с достаточно общей функциональной зависимостью $\hat{\mathcal{H}}(k, r)$. Поскольку двумерное движение в магнитном поле периодично, правило квантования Бора Зоммерфельда запишется в виде

$$\oint\limits_C \mathbf{P} \cdot d\mathbf{r} = \oint\limits_C \left(\mathbf{p} + \frac{e}{c}\mathbf{A}\right) \cdot d\mathbf{r} = 2\pi\hbar(n + \frac{1}{2}),$$

где $\mathbf{P}$ - обобщенный импульс и интегрирование ведется по циклотронной траектории $\varepsilon(\mathbf{p}) = const$ (в этом разделе во избежание путаницы мы обозначаем фигурирующий в адиабатическом инварианте обобщенный импульс посредством $\mathbf{P}$). Или, переходя к волновым векторам

$$\oint\limits_C \mathbf{k} \cdot d\mathbf{r} + \frac{e}{c\hbar} \oint\limits_C \mathbf{A} \cdot d\mathbf{r} = 2\pi(n + \frac{1}{2}). \tag{6.22}$$

При вычислении первого интеграла используем установленную ранее связь двумерных траекторий в магнитном поле в координатном пространстве и пространстве волновых векторов (6.15)

$$\oint\limits_C \mathbf{k} \cdot d\mathbf{r} = \lambda_H^{-2} \oint\limits_C \left[\mathbf{r} \times \frac{\mathbf{H}}{H}\right] \cdot d\mathbf{r}.$$

Перестановкой членов в смешанном произведении фигурирующий здесь интеграл приводится к виду

$$\oint\limits_C \left[\mathbf{r} \times \frac{\mathbf{H}}{H}\right] \cdot d\mathbf{r} = -\frac{\mathbf{H}}{H} \oint\limits_C \mathbf{r} \times d\mathbf{r}, a‘$$

что с точностью до знака совпадает с удвоенной площадью, охватываемой циклотронной орбитой. Таким образом, для первого члена в (6.22) имеем:

$$\oint\limits_C \mathbf{k} \cdot d\mathbf{r} = -2\lambda_H^{-2}\sigma_r.$$

Используя теорему Стокса и учитывая, что $\mathbf{H} = \mathbf{rotA}$, второй интеграл в (6.22) вычисляется также элементарно

$$\frac{e}{c\hbar}\oint\limits_C \mathbf{A} \cdot d\mathbf{r} = \frac{e}{c\hbar}\oint\limits_\sigma \mathbf{rotA} \cdot d\sigma_r = \frac{e}{c\hbar}\oint\limits_\sigma \mathbf{H} \cdot d\sigma_r = \frac{eH}{c\hbar}\oint\limits_\sigma d\sigma_r = \lambda_H^2\sigma_r.$$

В результате, квазиклассическое квантование в магнитном поле (6.22) сводится к квантованию площади, охватываемой циклотронной орбитой в координатном пространстве

$$\sigma_r = 2\pi\lambda_H^2(n + \frac{1}{2}) = \frac{2\pi c\hbar}{eH}(n + \frac{1}{2}). \tag{6.23}$$

Это соотношение может также интерпретироваться как квантование магнитного потока $\Phi = H\sigma_r$ через площадь орбиты

$$\Phi_n = \frac{ch}{e}(n + \frac{1}{2}) = \Phi_0(n + \frac{1}{2}) \tag{6.24}$$

Величина

$$\Phi_0 = 2\pi c\hbar/e = ch/e \tag{6.25}$$

представляет собой квант магнитного потока. Используя установленную ранее связь (6.15) для площади орбиты в $k-$пространстве получаем

$$\sigma_k = \lambda_H^{-4}\sigma_r = 2\pi\lambda_H^{-2}(n + \frac{1}{2}) = \frac{eH}{ch}(n + \frac{1}{2}). \tag{6.26}$$

Поскольку циклотронная орбита в $k-$пространстве, а, следовательно, и охватываемая ею площадь, определяется изоэнергетической поверхностью $\varepsilon(\mathbf{p}) = const$ (в трехмерном случае плоской кривой, определяемой уравнениями $\varepsilon(\mathbf{p}) = const$, $p_z = const$)), это соотношение устанавливает зависимость энергии от магнитного поля и номера уровня Ландау. Разность энергий двух соседних уровней Ландау равна

$$\varepsilon_{n+1} - \varepsilon_n = \lambda_H^{-2}\left(\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon)}\right)^{-1} = \hbar\omega_c(\varepsilon), \tag{6.27}$$

в полном согласии с принципом соответствия (равенство частот циклотронного движения и энергетических переходов). Формально, квазиклассическое квантование справедливо только для больших номеров квантовых уровней. Однако, поскольку для малых n (малых энергий) закон дисперсии в полупроводниках, как правило, близок к квадратичному, а квазиклассическое приближение в этом случае совпадает с точным решением даже при $n = 0$, условие квантования (6.26) можно считать применимым практически без ограничений (разумеется, в рамках применимости самого эффективного гамильтониана).

Отметим, что использованный выше подход при выводе условия квантования (6.26) носит общий характер, не требующий какого-либо предположения о калибровке для векторного потенциала. Конечно, тот же результат может быть получен и для конкретного выбора $\mathbf{A}(\mathbf{r})$. Так, в калибровке Ландау $\mathbf{A} = (0, Hx)$ эффективный двумерный гамильтониан в магнитном поле принимает вид $\hat{\mathrm{H}} = \varepsilon(\hat{k}_x, \hat{k}_y - eHx/c\hbar)$. Поскольку гамильтониан не зависит от y, мы имеем единственное условие квантования $\oint\limits_C K_x \cdot dx = 2\pi(n + 1/2)$. При использованной калибровке $K_x = k_x + eA_x/c\hbar = k_x$, а согласно (6.14) $dx = \lambda_H^2 dk_y$. Учитывая также, что

$$\oint\limits_C k_x \cdot dk_y = \sigma_k,$$

приходим к тому же правилу квантования (6.26).

Рассмотрим в качестве простых примеров использование правила квантования (6.26) для изотропных квадратичного и линейного законов дисперсии. Площадь орбиты при изотропной зависимости $\varepsilon(|\mathbf{k}|)$ равна $\sigma_k = \pi k^2$. Для квадратичного закона дисперсии имеем $\sigma_k = 2\pi m\varepsilon/\hbar$, что после подстановки в условие квантования приводит к уже полученному нами выше эквидистантному спектру:

$$\varepsilon_n = \hbar\omega_c(n + \frac{1}{2}).$$

Для линейного закона дисперсии $\varepsilon = ps$ ($\varepsilon = s\sqrt{p_x^2 + p_y^2}$) условие

квантования запишется в виде

$$\frac{\pi\varepsilon^2}{\hbar^2 s^2} = 2\pi\lambda_H^{-2}(n+\frac{1}{2}),$$

что дает для спектра в магнитном поле

$$\varepsilon_n = s\hbar\sqrt{2\frac{eH}{c\hbar}(n+\frac{1}{2})}. \tag{6.28}$$

Отметим, что и в этом случае квазиклассическое приближение приводит к результату, совпадающему с точным решением уравнения Клейна–Гордона–Фока (или уравнения Дирака при игнорировании спина). Расстояние между магнитными уровнями при линейном законе дисперсии уменьшается с ростом магнитного поля и номера уровня Ландау. Если спектр в магнитном поле представить в стандартном виде (как при квадратичном законе дисперсии) с некоторой эффективной циклотронной частотой и массой

$$\varepsilon_n = \hbar\omega_c(\varepsilon)(n+\frac{1}{2}) = \hbar\frac{eH}{m_c(\varepsilon)c}(n+\frac{1}{2}),$$

то последняя будет возрастающей функцией магнитного поля и номера уровня Ландау

$$m_c(H,n) = \frac{1}{s}\sqrt{\frac{e\hbar H}{2c\hbar}(n+\frac{1}{2})}$$

(легко проверить, что к тому же результату приводит и формула (6.9)). Нетрудно видеть, что это выражение согласуется со стандартным для ультрарелятивистского электрона соотношением

$$m_c(H,n) = m_c(\varepsilon) = \frac{\varepsilon_n}{2s^2}.$$

6.5 Кратность вырождения уровней Ландау

Как следует из двумерного спектра в магнитном поле, энергия не зависит от квантового числа k_y - уровни Ландау вырождены по положению центра осциллятора $x_0 = \lambda_H^2 k_y$. Для неограниченной двумерной системы разрешены произвольные значения x_0 (k_y) и кратность вырождения равна бесконечности. Для образца конечного размера, как и в случае

невозмущенной магнитным полем двумерной системы, число разрешенных значений k_y будет пробегать ограниченный дискретный ряд значений. Определим кратность вырождения магнитных уровней на единицу площади, полагая вначале закон дисперсии квадратичным, т.е. для спектра (6.20). Поскольку, согласно теореме Лиувилля, объем фазового пространства, а вместе с ним и число квантовых состояний, для консервативной системы не изменяется, число состояний, приходящихся на n-й уровень Ландау с энергией $\varepsilon_n = \hbar\omega_c(n + 1/2)$, может быть найдено из условия, что состояния из интервала энергий $n\hbar\omega_c \div (n+1)\hbar\omega_c$ в отсутствии магнитного поля группируются в магнитном поле в один уровень $\varepsilon_n = \hbar\omega_c(n + 1/2)$ (рис. 6.5). Это дает для кратности вырождения n_H уровня с энергией ε_n (на единицу площади)

$$n_H = \int\limits_{n\hbar\omega_c}^{(n+1)\hbar\omega_c} D(\varepsilon)d\varepsilon = \frac{m}{2\pi\hbar^2} \int\limits_{n\hbar\omega_c}^{(n+1)\hbar\omega_c} d\varepsilon = \frac{m}{2\pi\hbar}\omega_c,$$

где мы использовали выражение для плотности состояний в случае квадратичного закона дисперсии $D(\varepsilon) = m/2\pi\hbar^2$ (в этом разделе мы считаем, что спиновое и долинное вырождения сняты). Поскольку при квадратичном законе дисперсии $\omega_c = eH/mc$, то для кратности вырождения получаем

$$n_H = \frac{eH}{ch}. \tag{6.29}$$

Как видим, n_H определяется исключительно величиной магнитного поля и мировыми константами и не зависит ни от параметров материала (эффективной массы), ни от номера уровня Ландау. В приведенном выводе формулы (6.29) такая независимость обусловлена, по существу, эквидистантностью уровней Ландау и независимостью плотности состояний $D(\varepsilon)$ двумерной системы от энергии в нулевом магнитном поле (эквидистантностью уровней квазидискретного спектра при $H = 0$) для квадратичного закона дисперсии. Возникает вопрос - является ли такая независимость только следствием квадратичности закона дисперсии (как это имеет место для независимости двумерной плотности состояний от энергии) или носит более общий характер?

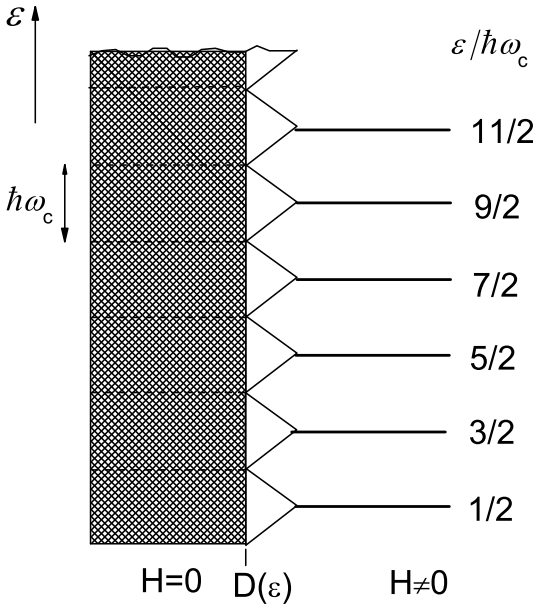

Рис. 6.5: .

Нетрудно видеть, что (6.29) остается в силе и в случае произвольной зависимости $\varepsilon(k)$, если движение в магнитном поле квазиклассично, что соответствует большим номерам уровней Ландау. В рамках картины о "схлапывании" состояний из интервала энергий $\varepsilon_{n+1} \div \varepsilon_n$ в один уровень мы по-прежнему имеем

$$n_H = \int\limits_{\varepsilon_n}^{\varepsilon_{n+1}} D(\varepsilon)d\varepsilon = \frac{1}{2\pi\hbar^2} \int\limits_{\varepsilon_n}^{\varepsilon_{n+1}} m_D(\varepsilon)d\varepsilon,$$

где, однако, теперь эффективная масса плотности состояний m_D зависят от энергии. При больших n энергетический интервал $\varepsilon_{n+1} - \varepsilon_n$ мал по сравнению с ε_n, уровни "квазиэквидистантны", и масса плотности состояний m_D может быть вынесена за знак интеграла, что дает

$$n_H \approx \frac{m_D(\varepsilon_n)}{2\pi\hbar^2}\hbar\omega_c(\varepsilon_n) = \frac{m_D(\varepsilon_n)}{2\pi\hbar^2}\frac{\hbar eH}{m_c(\varepsilon_n)c} = \frac{eH}{ch},$$

где учтено, что $\varepsilon_{n+1} - \varepsilon_n = \hbar\omega_c = \hbar eH/cm_c(\varepsilon)$, а в последнем равенстве мы использовали установленную нами ранее эквивалентность m_D и m_c для двумерных систем при произвольном законе дисперсии. Этот квазиклассический результат может быть получен и иначе. Площадь пространства волновых векторов, приходящаяся на одно состояние, согласно

(6.26) равна $\Delta\sigma_k = \sigma_{k,n+1} - \sigma_{k,n} = 2\pi\lambda_H^{-2}$ и, следовательно, число состояний на единицу площади равно $1/\Delta\sigma_k = 2\pi\lambda_H^{-2} = eH/ch$, что совпадает с (6.29).

Тот факт, что кратность вырождения как в области малых, так и больших номеров уровней Ландау не зависит от параметров материала (от закона дисперсии), позволяет предположить, что формула (6.29) носит универсальный характер и, следовательно, может, быть получена из общих соображений, не использующих при выводе материальных параметров образца (параметров закона дисперсии). Действительно, подобно тому, как это было сделано при подсчете числа состояний в кристалле без магнитного поля (разд. 2.1), число квантовых состояний может быть получено наложением циклических граничных условий на волновую функцию. В случае калибровки $A = (0, Hx)$ условие периодичности $\varphi(0) = \varphi(L)$ может быть наложено только на волновую функцию $\varphi(y) = \exp ik_y y$, что дает для разрешенных значений волнового вектора

$$k_{yn} = 2\pi n/L_y. \tag{6.30}$$

Граничные же условия по x для образца с достаточно большими размерами в этом направлении $L_x >> \lambda_H$, вообще несущественны, поскольку пространственная протяженность волновой функции $\varphi(x)$ порядка магнитной длины, а $\lambda_H << L_x$ (так, для случая параболического спектра электрон локализован в магнитном потенциале $m\omega_c^2(x - x_0)^2/2)$). Однако собственное значение импульса k_{yn} не может принимать сколь угодно больших значений, поскольку его величина связана с положением центра осциллятора $x_0 = \lambda_H^2 k_y$, а последний не может выходить за границы образца. Поскольку $x_0 \leq L_x$, то максимально возможная величина k_{yn} равна $\lambda_H^{-2} L_x$, что совместно с (6.30) дает для верхнего значения квантового числа n, а, следовательно, для числа квантовых состояний $n_{\max} = \lambda_H^{-2} L_x L_y/2\pi$ или, в расчете на единицу двумерного объема, $n_{\max}/L_x L_y = eH/ch$, т.е. приходим к выражению (6.29) [9].

[9]Укажем еще на один элегантный способ вывода соотношения (6.29). Каждому электронному состоянию в магнитном поле (циклотронной орбите) соответствует квант потока (6.25) $\Phi_0 = ch/e$. Поскольку поток Φ в расчете на единицу площади равен H, то число состояний на единицу площади $n_H = \Phi/\Phi_0 = H/\Phi_0 = eH/ch$.

6.6 Особенности циклотронного движения в 2D системах

В случае идеальной двумерной системы (квантовая плоскость) можно задать только две компоненты векторного потенциала (A_x, A_y), в связи с чем может быть определена только одна компонента магнитного поля

$$H = \mathbf{rot}\mathbf{A} = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}, \tag{6.31}$$

(соответствующая z-ой компоненте в трехмерном случае; как всякое векторное произведение, ротор в двумерном случае вырождается в псевдоскаляр). Таким образом, в идеальной двумерной системе мы всегда имеем дело с магнитным полем, перпендикулярным двумерной плоскости. Физическая причина этого достаточно очевидна. Действительно, только нормальная к 2D - слою компонента магнитного поля создает силу Лоренца, лежащую в 2D плоскости, и только она может обеспечить циклотронное движение электронов. Компонента, параллельная двумерному электронному газу, приводит к силе Лоренца, направленной перпендикулярно плоскости движения и не может привести к перемещению электрона, поскольку этому препятствует высокий потенциальный барьер, двумеризующий электронный газа. В реальных двумерных системах, как толщина двумерного слоя, так и высота барьеров имеют конечные значения, однако, до тех пор, пока радиус циклотронной орбиты значительно превышает толщину двумерного слоя, а высота барьера больше циклотронной энергии $\hbar\omega_c$, ситуация, по существу, остается той же самой – орбитальное движение в магнитном поле определяется только нормальной компонентой магнитного поля $H_\perp = H\cos\theta$. Величину последней в двумерной системе можно изменять как путем изменения величины магнитного поля H, так и использованием наклонных магнитных полей, когда изменяют ориентацию магнитного поля по отношению к нормали к двумерному слою (угол θ между ними). Эта особенность является наиболее мощным средством для установления двумерного характера движения носителей заряда в структурах с пространственным ограничением.

Следует обратить внимание, что спектр в магнитном поле для двумерной системы принципиально отличается от спектра трехмерного электронного газа (см. рис. 6.6). В идеальной двумерной системе, а все полученные выше результаты предполагают стационарный характер состояний, спектр дискретен, между уровнями Ландау ε_n электронных состояний нет, и плотность состояний, как функция энергии, есть δ- функции, центрированные при $\varepsilon = \varepsilon_n$

$$D(\varepsilon) = n_H \sum \delta(\varepsilon - \varepsilon_n).$$

В трехмерной системе между энергетическими уровнями ε_n, соответ-

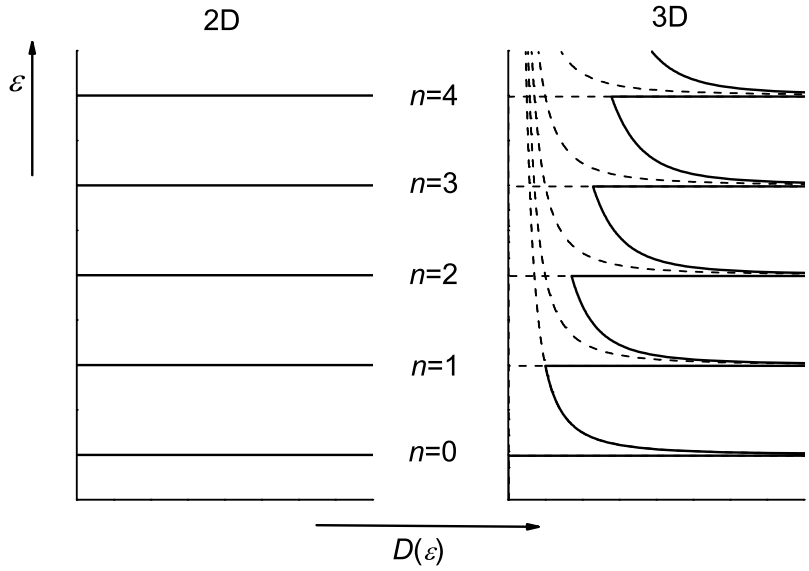

Рис. 6.6: Плотность состояний в магнитном поле для двумерного и трехмерного электронного газа.

ствующими дну одномерных подзон Ландау $\varepsilon_n(k_z = 0)$, даже в отсутствии рассеяния имеются состояния с квазинепрерывным спектром, соответствующие различным значениям сохраняющегося во времени квазиимпульса в направлении магнитного поля $\hbar k_z$. С использованием полученных здесь результатов и результатов разд. (2.2) для плотности состояний в одномерной системе, легко получить выражение для плотности состояний трехмерного электронного газа, спектр которого есть совокупность одномерных подзон, соответствующих каждому двумерному уровню Ландау. В частном случае квадратичного закона дисперсии, суммируя по одномерным подзонам Ландау со спектром $\varepsilon_n(k_z) =$

$\hbar\omega_c(n+1/2)+\hbar^2k_z^2/2m$ и плотностью состояний (2.33), сразу получаем

$$D_H(\varepsilon)=\frac{\sqrt{2m}n_H}{\pi\hbar}\sum_{n=0}^{n_{\max}}[\varepsilon-\varepsilon_n(0)]^{-\frac{1}{2}}=\frac{\sqrt{2m}eH}{2\pi^2\hbar^2c}\sum_{n=0}^{n_{\max}}[\varepsilon-\varepsilon_n(0)]^{-\frac{1}{2}},\quad(6.32)$$

где суммирование ведется по всем подзонам Ландау, дающим вклад в плотность состояний при заданной энергии ε, т.е. дно которых $\varepsilon_n(0)=\varepsilon_n=\hbar\omega_c(n+1/2)$ лежит ниже ε (подкоренное выражение под знаком суммы неотрицательно). Необходимо подчеркнуть, что и в трехмерной системе значения энергий ε_n физически выделены, поскольку плотность состояний (2.33) в каждой одномерной подзоне Ландау $\varepsilon_n(k_z)$ при $k_z=0$, имеет особенность вида $1/k_z$.

В применении к реальным системам понятие о строго дискретных уровнях энергии в магнитном поле имеет смысл, если можно пренебречь процессами рассеяния электронов на циклотронной орбите. Последние приводят к нестационарности состояний, соответствующих уровням Ландау. На каждом уровне электрон находится лишь конечное время τ_p - характерное среднее время между столкновениями с дефектами. Согласно принципу неопределенности между энергией и временем это означает, что при учете рассеяния уровни Ландау ε_n определены с точностью не превышающей $\Delta\varepsilon=\Gamma=h/\tau_p$. Полученные выше формулы имеют смысл, если эта ширина мала по сравнению с расстоянием между уровнями $\Gamma<<\varepsilon_{n+1}-\varepsilon_n=\hbar\omega_c$, что требует все более высоких магнитных полей по мере роста концентрации рассеивающих центров. Условие слабого размытия уровней совпадает с классическим определением сильных магнитных полей $\omega_c\tau_p>>1$, как полей, при которых еще имеет смысл говорить о циклотронной орбите (см Гл. 8).

В связи со столкновительным уширением уровней Ландау в двумерном электронном газе следует сделать еще одно замечание. Как в классической картине, так и квантовой трактовке движения электронов проводимости в магнитном поле, циклотронная орбита в двумерной системе остается локализованной в пространстве (это не так в трехмерном случае, где имеется свободное движение центра орбиты вдоль магнитного поля). Отсюда, на первый взгляд, следует, что, если циклотронная орбита уже сформировалась (на пути электрона по орбите дефекты от-

сутствуют), то и в дальнейшем она не может быть разрушена, поскольку ее пространственное положение не изменяется, а рассеивающие центры неподвижны (исключением является рассеяние на фононах, несущественное в области низких температур). В реальной системе, однако, всегда имеются электрические поля, связанные с зарядами примесей, либо с флуктуациями заряда, формирующего ограничивающий потенциал квантовой ямы. Присутствие таких полей приводят к дрейфу циклотронной орбиты (см. Гл. 7), и, как следствие, к рассеянию на дефектах, имеющихся на пути дрейфа.

Вычисление электронного спектра в магнитном поле с учетом уширения уровней Ландау является далеко не тривиальной задачей, при решении которой зачастую возникают трудности принципиального характера, связанные, например, с сингулярным видом плотности состояний. Анализ этих вопросов выходит за рамки нашего рассмотрения, однако, поскольку окончательные ответы достаточно просты и физически ясны, мы приведем наиболее часто используемые выражения для плотности состояний, полученные в рамках некоторых приближений. Для рассеяния на заряженных центрах рассмотрение, основанное на низшем борновском приближении при расчете рассеяния на каждом отдельном центре и самосогласованной процедуре при расчете эффектов уширения уровней, приводит в достаточно сильных полях, когда можно пренебречь взаимодействием между разными уровнями Ландау, к полуэллиптической форме плотности состояний

$$D(\varepsilon) = n_H \sum_n \frac{1}{\pi\Gamma_n} \left[1 - \left(\frac{\varepsilon - \varepsilon_n}{\Gamma_n}\right)^2\right]^{1/2} \tag{6.33}$$

с шириной Γ_n, в общем случае зависящей как от магнитного поля, так и номера уровня Ландау. Другой подход, основанный на интегралах по траекториям, приводит к гауссовой форме магнитных уровней

$$D(\varepsilon) = n_H \sum_n \sqrt{\frac{2}{\pi}} \frac{1}{\Gamma_n} \exp\left[-2\left(\frac{\varepsilon - \varepsilon_n}{\Gamma_n}\right)^2\right], \tag{6.34}$$

не имеющей нефизических особенностей, как это имеет место для в плотности состояний (6.33). Уширение уровней Ландау сильно зависит от ра-

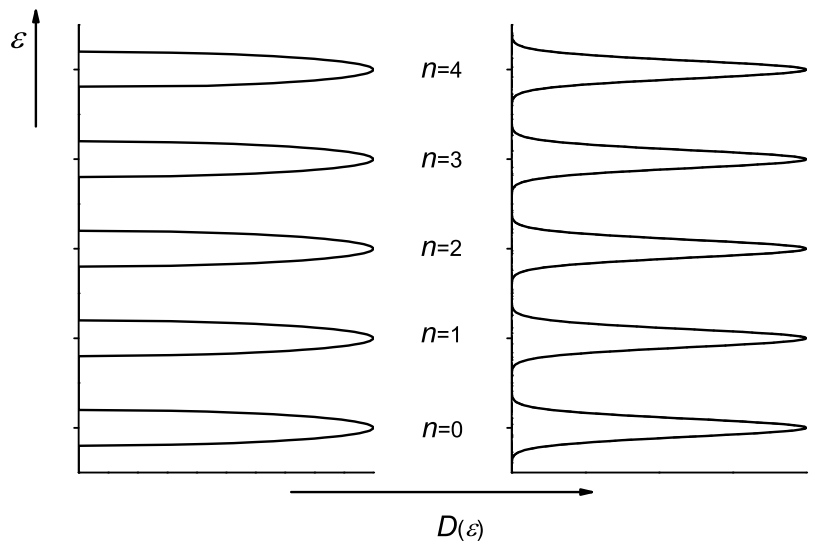

Рис. 6.7: Плотность состояний двумерного электронного газа в магнитном поле (6.33) и (6.34) при учете рассеяниия на короткодействующих центрах.

диуса действия d_V рассеивающего потенциала. Для короткодействующих рассеивателей, радиус действия которых меньше длины волны электрона в магнитном поле $d_V < \lambda_H/\sqrt{2n+1}$, уширение

$$\Gamma_n = \Gamma = \sqrt{\frac{2}{\pi}\hbar\omega_c\frac{\hbar}{\tau_p}} \tag{6.35}$$

(τ_p - время релаксации импульса при $H = 0$) растет с магнитным полем, но не зависит от номера уровня Ландау. Профили плотности состоянии, описываемые формулами (6.33) и (6.34) для короткодействующих рассеивателей, сравниваются на рис. 6.7. Для предельно дальнодействующих рассеивателей уширение отсутствует. Очевидно, что для больших номеров уровней Ландау, для которых длина волны мала (хотя циклотронный радиус велик $r_c = \lambda_H\sqrt{2n+1}$!), роль радиуса действия более существенна. Учет рассеяния устраняет расходимость в плотности состояний двумерного газа в магнитном поле, однако, при выполнении условия $\Gamma << \hbar\omega_c$, плотность состояний в интервале энергий между магнитными уровнями мала по сравнению с ее величиной на уровне Ландау даже для гауссова уширения.

6.7 Движение в наклонном магнитном поле

Рассмотрим систему электронов в двумерном слое конечной толщины, пространственное распределение электронной плотности в которой описывается волновой функцией $\varphi(z)$ и которая находится в наклонном магнитном поле $(0, H_y, H_z)$, лежащем в плоскости y, z. При выборе калибровки в виде $\mathbf{A} = (H_y z, H_z x, 0)$ система описывается эффективным гамильтонианом [10]

$$\hat{\mathcal{H}} = \frac{\hbar^2}{2m}[(\hat{k}_x - \frac{eH_y}{c\hbar}z)^2 + (\hat{k}_y - \frac{eH_z}{c\hbar}x)^2]. \tag{6.36}$$

Поскольку, обычно радиус циклотронной орбиты даже при наибольших магнитных полях намного превышает типичную толщину 2D – слоя, компоненту магнитного поля, направленную вдоль двумерного слоя, можно рассматривать как возмущение. Ввиду этого можно ограничиться описанием движения электронов в 2D подзоне первым по H_y порядком теории возмущений, что приводит к гамильтониану

$$\hat{\mathcal{H}} = \frac{\hbar^2}{2m}[(\hat{k}_x - \lambda_{Hy}^{-2} < z >)^2 + (\hat{k}_y - \lambda_{Hz}^2 x)^2 + \lambda_{Hy}^{-4}\delta_z)^2], \tag{6.37}$$

где $\lambda_{Hy}^2 = c\hbar/eH_y$, $\lambda_{Hz}^2 = c\hbar/eH_z$, $\delta_z = [< z^2 > - < z >^2]^{1/2}$ - эффективная толщина двумерного слоя ($< z >= \int \varphi^*(z) z \varphi(z) dz$ и $< z >^2 = \int \varphi^*(z) z^2 \varphi(z) dz$ - средние значения координаты и ее квадрата в 2D подзоне, определяемые формирующим двумерный слой ограничивающим потенциалом).

В случае строго параллельного двумерному слою поля ($H_z = 0$) гамильтониан (6.37) вообще не содержит координат ($\lambda_{Hz} = 0$) и сразу определяет законы дисперсии (заменой операторов на их собственные значения). Воздействие поля H_y сводится, таким образом, к модификации исходной (в нулевом магнитном поле) дисперсионной параболы

$$\varepsilon(k_x, k_y) \to \varepsilon(k_x - \lambda_{Hy}^{-2} < z >, k_y) + \frac{\hbar^2 \lambda_{Hy}^{-4} \delta_z^2}{2m},$$

[10] Здесь мы ограничиваемся простейшим случаем квадратичного закона дисперсии. При произвольном виде $\varepsilon(\mathbf{p})$, в том числе для кейновского закона дисперсии, рассмотрение двумерной системы в наклонном магнитном поле наталкивается на серьезные, зачастую принципиального характера, трудности.

а именно к простому смещению ее минимума в пространстве волновых векторов в направлении k_x на величину $\lambda_{Hy}^{-2} < z >$ и сдвигу по энергии на $\hbar^2\lambda_{Hy}^{-4}\delta_z^2/2m$ (диамагнитный сдвиг - рис. 6.8). Происхождение такой модификации очевидно. В магнитном поле, параллельном оси y, на электроны двумерного слоя, движущиеся в противоположных направлениях оси x, действуют силы Лоренца, направленные в противоположных направлениях оси размерного квантования. Как следствие, это ведет к различной деформации волновых функций для разных k_x и нарушению симметрии $\varepsilon(k_x) = \varepsilon(-k_x)$.

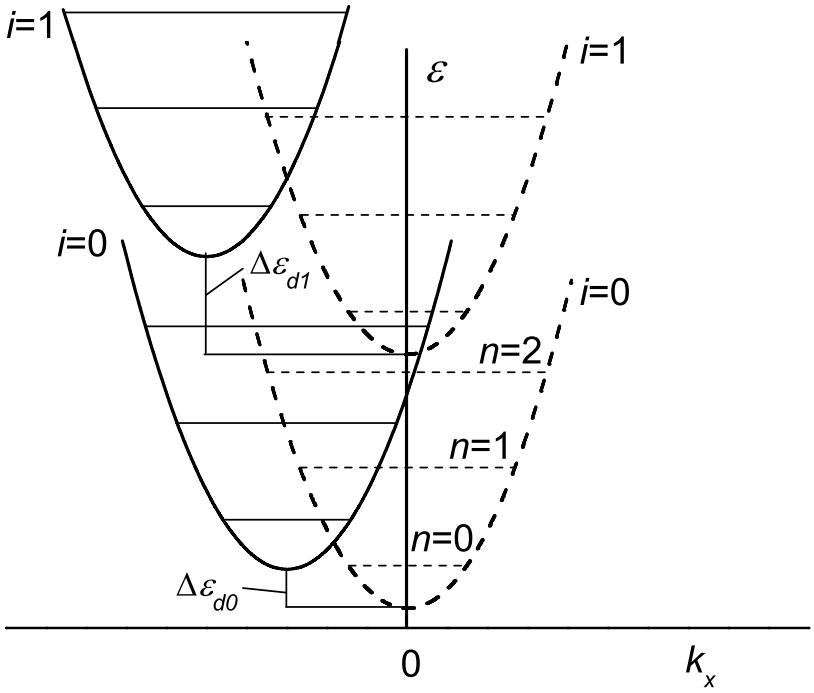

Рис. 6.8: Структура двумерных подзон и уровней Ландау в перпендикулярном (пунктир) и наклонном (сплошные линии) магнитном поле, лежащем в плоскости yz. Перпендикулярная двумерной плоскости компонента магнитного поля в обоих случаях одинакова. Параболы описывают спектр при равной нулю перпендикулярной компоненте магнитного поля. Параллельное двумерному слою магнитное поле приводит к диамагнитному сдвигу уровней размерного квантования. Предполагается, что пространственная протяженность волновой функции в направлении ограничивающего потенциала в основной подзоне меньше.

В случае, если наряду с H_y имеется также и перпендикулярная поверхности компонента магнитного поля H_z, гамильтониан содержит од-

ну координату x, допуская разделение переменных, и легко сводится к рассмотренному ранее гамильтониану для гармонического осциллятора, определяя уровни Ландау

$$\varepsilon_n = \hbar\omega_{cz}(n + \frac{1}{2}) + \frac{e^2 H_y^2}{2mc^2}\delta_z^2, \tag{6.38}$$

где циклотронная частота ω_{cz} определяется только перпендикулярной двумерному слою компонентой магнитного поля $\omega_{cz} = eH\cos\theta/mc = \omega_c\cos\theta$ ($H = \sqrt{H_y^2 + H_z^2}$, θ- угол между магнитным полем $\mathbf{H}$ и нормалью к двумерному слою). Таким образом, дискретный спектр в наклонном магнитном поле, с точностью до общего для всех уровней Ландау диамагнитного сдвига

$$\Delta\varepsilon_d = \frac{e^2(H\sin\theta)^2}{2mc^2}\delta_z^2, \tag{6.39}$$

полностью аналогичен структуре уровней Ландау в перпендикулярном поле величиной $H/\cos\theta$.

Квадратичная по параллельному магнитному полю величина диамагнитного сдвига растет (также квадратично) с увеличением протяженности волновой функции в направлении двумеризующего потенциала в соответствии с бóльшим нарушением критерия малости размера двумерного слоя по сравнению с магнитной длиной. Ясно, что в структурах, в которых ширина квантовой ямы растет с энергией (квантовые ямы в МДП – структурах, в одиночной гетерояме, дельта-легированных слоях и др.) диамагнитный сдвиг для высокоэнергетических 2D подзон, имеющих бóльшую протяженность волновых функций, будет заметно больше. Как следствие, межподзонные энергетические расстояния будут расти с параллельным магнитным полем, что можно наблюдать, исследуя, например, межподзонное поглощение света. Экспериментальные исследования подобных эффектов в параллельных магнитных полях, с другой стороны, дают информацию о параметрах пространственного распределения электронной плотности в направлении ограничивающего потенциала (в направлении конфайнмента). Отметим также, что из-за диамагнитного сдвига, дно двумерных подзон (особенно возбужденных) может с

ростом параллельного двумерному слою магнитного поля подняться выше уровня Ферми, что приведет к опустошению соответствующих двумерных состояний, и, как следствие, может существенно сказаться на свойствах двумерной системы.

6.8 Спиновое расщепление

Ранее мы нигде не учитывали в явном виде спин электрона (неявно полуцелое значение спина предполагалось при использовании статистики Ферми – Дирака). В отсутствии магнитного поля и если игнорировать взаимодействие спина с магнитным полем, порождаемым электроном при его движении (спин-орбитальное взаимодействие), спиновая степень свободы не сказывается на орбитальном движении. В магнитном поле связанный со спином магнитный момент электронов проводимости взаимодействует с полем и, в соответствии с двумя возможными значениями проекции спина, спиновое вырождение снимается и каждый уровень Ландау (каждая магнитная подзона в случае трех измерений) расщепляется на два спиновых подуровня (подзоны), смещенные относительно первоначально положения ε_n в простейшем случае квадратичного закона дисперсии на $\sigma_s g \mu_B H/2$, где $\sigma_s = \pm 1$, величина g есть g- фактор для данной энергетической зоны, μ_B - магнетон Бора (зеемановское расщепление).

В рамках метода эффективной массы, на котором целиком основано наше рассмотрение, влияние периодического потенциала решетки U_{lat} сведено к модификации закона дисперсии, тогда как уравнения движения и эффективный гамильтониан не зависят от U_{lat} и содержат только члены, связанные с внешними приложенными полями. Однако, наряду с перенормировкой массы электрона $m_0 \to m_{eff}$ решеточный потенциал при наличии спин-орбитальной связи может изменить и величину g- фактора, описывающего в эффективном гамильтониане взаимодействие магнитного момента электрона с внешним магнитным полем. Спин-орбитальное взаимодействие, будучи чисто релятивистским эффектом, определяется скоростью электрона, последняя же имеет наи-

большие значения вблизи остовов атомов формирующих решетку и растет с атомным номером. Если в кремнии величина g- фактора практически равна его значению для свободного электрона $g_0 = 2$, то в полупроводниках с кейновским спектром, особенно в узкощелевых и бесщелевых полупроводниках, эффективные g- факторы перенормируются столь же кардинально, как и эффективные массы ($g/g_0 \approx -m_0/m$) и могут достигать значений, в десятки и сотни раз превышающих g - фактор свободного электрона. Это, в общем-то, и неудивительно, поскольку сама малость ширин запрещенной зоны и эффективных массы в этих материалах обусловлена релятивистскими эффектами, в том числе (в бесщелевых полупроводниках почти исключительно) сильным спин-орбитальным взаимодействием.

Поскольку зеемановское расщепление $g\mu_B H$ не зависит от калибровки векторного потенциала, а в однородном магнитном поле гамильтониан вообще не содержит координаты электрона, его учет, по-существу, не изменяет уравнение Шредингера в магнитном поле - не сказывается на орбитальном движении электрона и на спектре, связанном с циклотронным движением. Формально член $g\mu_B H/2$ может быть отнесен к энергии в исходном уравнении и таким образом аддитивен к спектру, обусловленному орбитальным магнитным квантованием. В случае квадратичного закона дисперсии это сразу видно из исходного уравнения Шредингера (учет зеемановского расщепления аналогичен учету энергии движения в третьем измерении для трехмерной системы).

$$\frac{\hbar^2}{2m}\left[[(\hat{k}_x - \frac{e}{c\hbar}A_x)^2 + (\hat{k}_y - \frac{e}{c\hbar}A_y)^2] \pm \frac{1}{2}g\mu_B H\right]\phi(x,y) = \varepsilon\phi(x,y) \quad (6.40)$$

Отсюда для энергии в магнитном поле имеем

$$\varepsilon_n = \hbar\omega_c(n + \frac{1}{2}) \pm \frac{1}{2}g\mu_B H. \quad (6.41)$$

Наиболее сильно учет спина проявляется в специфике поведения двумерного электронного газа в наклонном магнитном поле. Дело в том, что если орбитальное квантование (структура уровней Ландау) определяется только перпендикулярной двумерному слою компонентой магнитного поля $H\cos\theta$, то зеемановское расщепление, обусловленное внутренней,

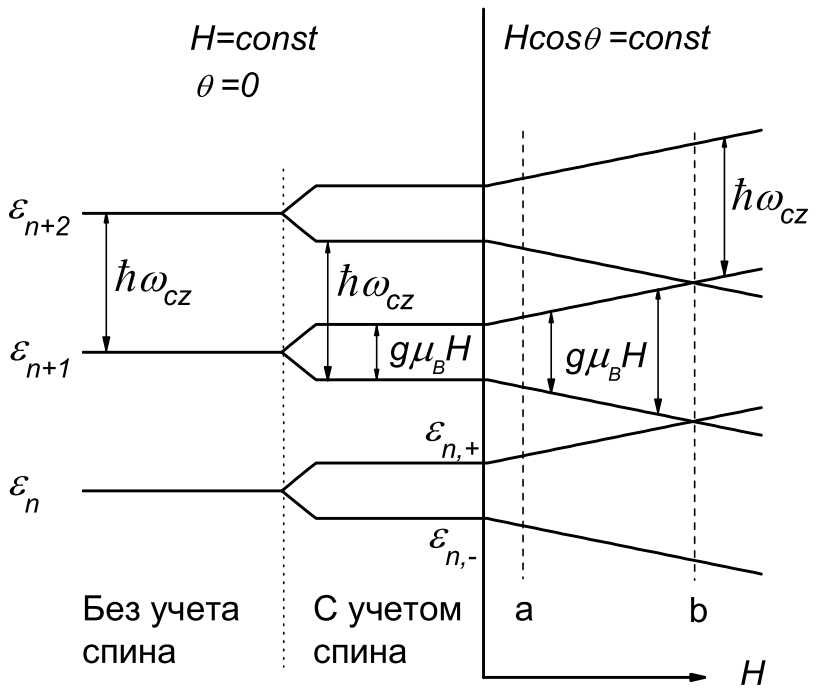

Рис. 6.9: Структура уровней Ландау двумерного электронного газа с учетом спинового расщепления в перпендикулярном ($\theta = 0$) и в наклонном магнитном поле (как функция H при постоянной компоненте $H_\perp = H\cos\theta$.

спиновой степенью свободы, не зависит от ориентации магнитного поля и определяется полным магнитным полем $\Delta\varepsilon_s = \pm g\mu_B H/2$. В простейшем случае квадратичного закона дисперсии спектр структура уровней магнитного квантования определяется, таким образом, формулой

$$\varepsilon_{n,\pm} = \frac{\hbar eH\cos\theta}{mc}(n+\frac{1}{2}) \pm \frac{1}{2}g\mu_B H + \frac{(eH\sin\theta)^2}{2mc^2}\delta_z^2. \qquad (6.42)$$

В двумерном электронном газе мы имеем возможность независимо управлять величиной энергетического зазора между уровнями Ландау и зазора между двумя спиновыми состояниями одного и того же уровня Ландау (в трехмерной системе отношение этих величин является константой), что позволяет кардинально изменять структуру спектра в магнитном поле, вплоть до изменения нормального порядка следования уровней Ландау $\varepsilon_{n+1} > \varepsilon_n$. Изменяя магнитное поле, но оставляя постоянной его нормальную двумерному слою компоненту $H_z = H\cos\theta = const$ путем изменения угла наклона, можно менять величину зеемановского расщепления $g\mu_B H$ при постоянном циклотронном расщеплении $\hbar\omega_{cz} =$

$\hbar eH\cos\theta/mc$ (рис. 6.9). И, наоборот, увеличивая угол наклона магнитного поля при постоянной его величине, можно уменьшать расстояние между уровнями Ландау (вплоть до нуля) при неизменном спиновом расщеплении [11]. В частности, таким путем можно достичь того, чтобы спиновое расщепление $\varepsilon_{n,+} - \varepsilon_{n,-}-$ сравнялось с энергетическим интервалом между двумя разными спиновыми состояниями соседних уровней Ландау $\varepsilon_{n+1,-} - \varepsilon_{n,+}-$ (уровни в сечении "a" на рис. 6.9). Такая эквидистантная структура спектра отчетливо проявляется в магнетоосцилляционных эффектах и исследования в наклонных магнитных полях позволяют по величине угла θ, соответствующего выполнению условия для такой эквидистантности

$$\frac{\hbar eH\cos\theta}{mc} = 2g\mu_B H \tag{6.43}$$

определить g -фактор в двумерном электронном газе. Эквидистантность имеет место также и при выполнении условия

$$\frac{\hbar eH\cos\theta}{mc} = g\mu_B H, \tag{6.44}$$

т.е. при углах наклона, когда зеемановское расщепление становится равным циклотронному (сечение "b" на рис. 6.9). Легко видеть, что структура уровней Ландау в этом случае (двукратное вырождение соседних уровней Ландау с противоположной проекцией спина, за исключением нижайшего по энергии состояния $\varepsilon_{0,-}$) аналогична таковой для свободного электрона.

6.9 Задачи

1. Как меняется скорость и координата электрона со временем в одномерном кристалле с законом дисперсии

$$\varepsilon(p) = \frac{E_0}{2}(1 - \cos\frac{ap}{\hbar}),$$

[11] Диамагнитный сдвиг, будучи независим от номера уровня Ландау, не изменяет структуры (относительного положения) уровней в магнитном поле.

помещенном в однородное электрическое поле E. Зависят ли частота и амплитуда блоховских осцилляций от ширины энергетической зоны?

Ответ:

$$v_g(t) = \frac{aE_0}{2\hbar}\sin\left(\frac{e\mathrm{E}a}{\hbar}t\right),$$

$$x(t) = \frac{E_0}{2e\mathrm{E}}\left[1-\cos\left(\frac{e\mathrm{E}a}{\hbar}t\right)\right] = \frac{E_0}{e\mathrm{E}}\sin^2\left(\frac{e\mathrm{E}a}{2\hbar}t\right),$$

Частота блоховских осцилляций определяется только энергией набираемой электроном в электрическом поле на длине постоянной решетки и не зависит от ширины энергетической зоны. Амплитуда же осцилляций пропорциональна последней.

2. Выведите соотношение между площадями орбит в магнитном поле в координатном пространстве и пространстве квазиволновых векторов

$$\frac{\sigma_r}{\sigma_k} = \left(\frac{c\hbar}{eH}\right)^2 = \lambda_H^4$$

для электрона с квадратичным законом дисперсии, используя выражение для классического циклотронного радиуса.

3. Покажите, что квазиклассическое выражение для спектра в магнитном поле электрона с линейным законом дисперсии

$$\varepsilon_n = s\hbar\sqrt{2\frac{eH}{c\hbar}(n+\frac{1}{2})}.$$

совпадает с точным решением (в пренебрежении спином) .

Указание: Используйте уравнение Клейна-Гордона-Фока в магнитном поле для частицы с нулевой массой

$$s^2(\hat{k}_x - \frac{e}{c\hbar}A_x)^2 + s^2(\hat{k}_y - \frac{e}{c\hbar}A_y)^2\phi = \varepsilon^2\phi$$

и сведите его подходящим выбором калибровки к уравнению для линейного осциллятора.

4. Найдите точный (исходя из уравнения Клейна-Гордона-Фока) и квазиклассический спектр в магнитном поле для частицы с кейновским законом дисперсии.

   Ответ:
$$\varepsilon_n = \sqrt{2s^2\hbar\frac{eH}{c}\left(n+\frac{1}{2}\right)+m^2s^4} - ms^2$$

6.10 Вопросы для самопроверки

1. Что образуют компоненты электромагнитного поля - вектор или тензор? Что такое векторный потенциал? Как он связан с напряженностью магнитного поля?

2. Каким образом учитывается воздействие электрического и магнитного полей в классической механике свободного электрона? Выпишите уравнение движения. Отличается ли соответствующее уравнение для электронов проводимости?

3. Запишите уравнение Шредингера для свободного электрона и электрона в кристалле в однородном электрическом и магнитном полях в предположении квадратичного закона дисперсии. Как модифицируется это уравнение в случае произвольного закона дисперсии? Что такое эффективный гамильтониан?

4. Назовите интегралы движения для электрона в магнитном поле. Как они определяют траекторию электрона?

5. Дайте качественное объяснение, почему при гамильтоновом описании движения заряженной частицы в однородных электрическом и магнитном полях электрическое поле входит в потенциальную энергию гамильтониана, а магнитное в оператор кинетической энергии?

6. Свободный электрон в однородном электрическом поле двигается в пространстве импульсов равномерно и неограничено (напишите выражение для скорости в этом пространстве) и ускоренно и также

неограничено в координатном пространстве. Так ли это в случае электрона проводимости в идеальном и неограниченном кристалле, помещенном в однородное электрическое поле?

7. Что такое блоховские осцилляции? Чем определяются их частота и амплитуда? Что препятствует практической реализации блоховского генератора в трехмерных полупроводниках и в металлах? По какой причине более перспективны для реализации такого генератора сверхрешетки на основе полупроводниковых наноструктур?

8. Как называется частота вращения электрона проводимости в магнитном поле? Выпишите выражение для нее в случае квадратичного и линейного законов дисперсии.

9. Дайте определение циклотронной эффективной массы в двумерном электронном газе. Выразите ее через интеграл от обратной скорости электрона по траектории электрона в пространстве квазиволновых векторов. Как соотносится эта величина с эффективной массой двумерной плотности состояний? Различно ли это соотношение для частиц с квадратичным и линейным законом дисперсии?

10. Как соотносятся магнитные длины легкой и тяжелой дырок? В какие известные Вам выражения входит этот параметр? Какие из этих выражений относятся к классическому и квантовому описанию электрона в магнитном поле?

11. Как можно из траектории электрона в магнитном поле в пространстве квазиимпульса получить траекторию в координатном пространстве? Зависят ли эти траектории от закона дисперсии? Как соотносятся площади траекторий в этих пространствах? Зависит ли это соотношение от законов дисперсии?

12. Может ли магнитный поток через площадь, ометаемую траекторией электрона при его движении в магнитном поле, принимать произвольные значения?

13. Площадь орбиты электрона в магнитном поле в пространстве волновых векторов принимает квантованные значения

$$\sigma_k = \frac{eH}{ch}(n + \frac{1}{2}).$$

Как, используя это соотношение, можно определить энергию уровней Ландау в двумерном газе, если известен закон дисперсии? Продемонстрируйте это на примере линейного изотропного закона дисперсии $\varepsilon = s\hbar k$.

14. Могут ли наблюдаться эффекты в магнитном поле в идеальной квантовой нити? Если да, то какие?

15. В какой области магнитных полей могут наблюдаться эффекты, связанные с квантованием Ландау электронов в квантовой нити диаметром a? Рассмотрите случаи квадратичного и линейного законов дисперсии.

16. Чем отличаются спектры в магнитном поле в идеальных трехмерном, двумерном и одномерном электронном газах? Напишите выражения для этих спектров в случае двумерного и трехмерного газов частиц с квадратичной дисперсией.

17. Волновая функция электрона в двумерной плоскости в магнитном поле в калибровке Ландау с точностью до нормировочного множителя равна $\psi(x, y) = \varphi(x)\exp(ip_y y/\hbar)$, т.е. ее зависимость от координаты y совпадает с волновой функцией свободного электрона, двигающегося по оси y. В последнем случае скорость, как известно, равна $v_y = p_y/m$. Из какой формулы следует последнее выражение? Что дает эта формула для квантово-механической скорости электрона в магнитном поле?

18. Чему равна плотность состояний идеального двумерного газа электронов проводимости в магнитном поле на уровне Ландау и между уровнями Ландау? Есть ли отличия в этом отношении между двумерным и трехмерным газом?

19. По какой причине спектр в магнитном поле является сильно вырожденным? Какой термин используется для числа состояний на одном уровне Ландау, отнесенного к единице двумерного объема? Зависит ли эта величина от номера уровня Ландау? Что можно сказать об отношении числа состояний на нулевых уровнях Ландау тяжелых и легких дырок в одном и том же магнитном поле?

20. Какие подходы к вычислению кратности вырождения уровней в магнитном поле Вам известны? Выпишите выражение для этой величины. Если затрудняетесь, то попытайтесь это сделать из соображений размерности, принимая во внимание, что вырождение должно зависеть от магнитного поля, заряда (магнитное поле действует только на заряженные частицы) и, будучи квантовой характеристикой, должно содержать постоянную Планка.

21. Каким соотношением связаны кратность вырождения уровней Ландау и магнитная длина?

22. Какие процессы устраняют дельтаобразный характер плотности состояний в двумерном газе электронов проводимости, помещенном в магнитное поле?

23. Как изменяется энергетическое положение уровней Ландау при изменении ориентации магнитного поля в изотропных двумерном и трехмерном электронных газах при нулевом g-факторе?

24. Как изменяется положение энергетических уровней в магнитном поле при изменении его ориентации в изотропных двумерном и трехмерном электронных газах с учетом зеемановского расщепления?

25. По какой причине параллельное двумерному слою магнитное поле не приводит к образованию уровней Ландау? Какие условия должны быть выполнены в реальных структурах с двумерным электронным газом, чтобы проявились эффекты, связанные с влиянием

параллельного двумерному слою магнитного поля на орбитальное движение носителей заряда? Что такое диамагнитный сдвиг?

Глава 7

Двумерный электронный газ в скрещенных электрическом и магнитном полях

7.1 Граничные состояния

Полученные в предыдущей главе результаты по динамике и спектру электронов в магнитном поле относятся по-существу к однородной и потому бесконечной системе, поскольку как полуклассическое, так и квантовое рассмотрение предполагало отсутствие каких либо других, помимо магнитного, полей, в том числе полей, связанных со скачком потенциала вблизи границ образца. Легко видеть, что наличие границы кардинально меняет характер классического движения электронов, расстояние центра циклотронной орбиты которых от границы меньше циклотронного радиуса $r_H = v/\omega_c = \sqrt{2\varepsilon/m}/\omega_c$ и которые, следовательно, испытывают влияние границы. Поскольку r_H при всех достижимых магнитных полях значительно превышает постоянную решетки (в противном случае мы выходим и за рамки метода эффективной массы) граница может рассматриваться как высокая прямоугольная потенциальная стенка. Если, помимо того, такая стенка является гладкой на расстояниях масштаба r_H и столкновения носят упругий характер, то взаимодействие с границей образца может рассматриваться как зеркальное отражение класси-

ческой частицы, при котором параллельная поверхности образца компонента импульса сохраняется, а нормальная к границе компонента меняет знак. Результатом такого взаимодействия является перемещение (дрейф) заряженной частицы вдоль границы образца по "скачущим магнитным траекториям", форма которых зависит от энергии (скорости) электрона и расстояния центра его орбиты до границы (рис. 7.1). По мере прибли-

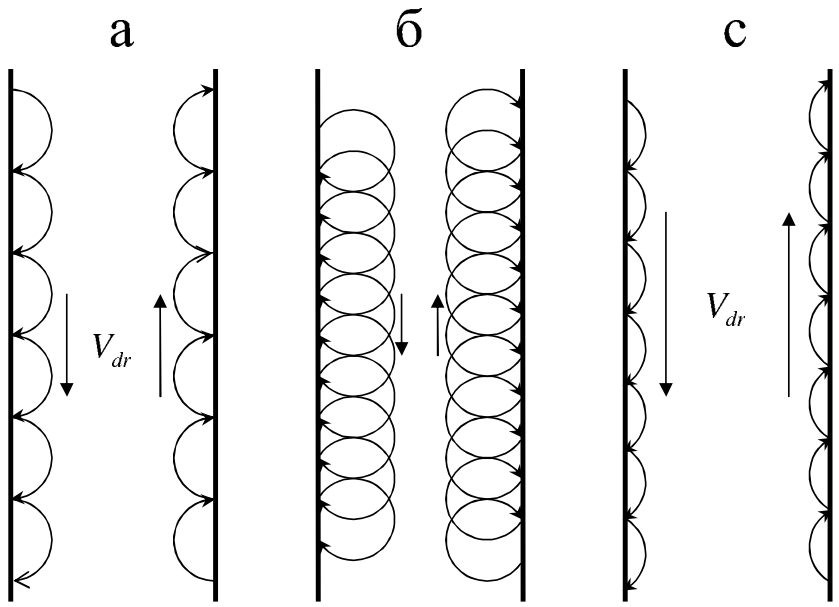

Рис. 7.1: Скачущие пограничные траектории в магнитном поле при различных расстояниях центра циклотронной орбиты от границы.

жения центра орбиты к границе среднее (за "усеченный"циклотронный период или за большой промежуток времени) значение параллельной границе компоненты скорости $\bar{v}_y = v_d$ возрастает и достигает максимального значения $\omega_c r_H$, когда центр находится вне образца на расстоянии близком к r_H (предельно малые углы падения, почти скользящие траектории - рис. 7.1c). В перпендикулярном границе направлении средняя скорость $\bar{v}_x$ остается равной нулю, тогда как компонента v_x мгновенной скорости в точке отражения в указанном интервале положений центра орбиты возрастает от нуля до максимального значения $\omega_c r_H$, когда центр орбиты совпадает с границей, и затем снова уменьшается до нуля.

При квантовом рассмотрении полученные выше решения для спектра в магнитном поле с дискретными уровнями энергии ε_n являются решениями, соответствующими нулевым граничным условиям для волновой

функции φ_n на бесконечности и поэтому, строго говоря, справедливы только для бесконечного образца. Однако, в виду быстрого (экспоненциального) затухания волновой функции, нулевые условия с хорошей точностью удовлетворяются уже при расстояниях, не намного превышающих радиус циклоторонной орбиты (при $x > x_0$ и при $r > r_0$ для линейной и центрально – симметричной калибровки). В результате для состояний ε_{n,x_0}, центры орбит которых x_0 расположены от границы на расстояние, превышающее радиус циклотронной орбиты

$$r_{Hn} = \sqrt{2\varepsilon_n/m}/\omega_c = \sqrt{2\hbar\omega_c(n+1/2)/m}/\omega_c = \lambda_H\sqrt{2n+1},$$

наличие границ не скажется существенно на спектре. Если же расстояние центра осциллятора до границы меньше r_{nH} (классическая орбита не помещается полностью внутри образца), то из-за высокого приграничного потенциального барьера на границе образца нулевое граничное условие должно выполняться на расстоянии, меньшем радиуса циклотронной орбиты, что не может быть удовлетворено про том же значении энергии, что и для случая объемных орбит. Таким образом, положение энергетического уровня изменяется. Поскольку возмущение спектра границей зависит от положения центра осциллятора $x_0 = \lambda_H^2 k_y$, характерное для объемного спектра вырождение по k_y снимается (граница нарушает однородность пространства, ответственного за такое вырождение в объеме, хотя магнитное поле по-прежнему однородно) и энергия будет функцией k_y. Поскольку k_y фактически непрерывная переменная, обусловленное границей изменение энергии будет также непрерывным, т.е. относящиеся к каждому уровню Ландау пограничные состояния образуют фактически неквантованный континуум уровней.

Легко видеть, что к тем же выводам приводит и использование квазиклассической процедуры квантования. В условии квантования магнитного потока (6.24) теперь фигурируют “урезанные” границей траектории, а, следовательно, ввиду подобия траекторий в прямом и обратном пространстве, урезанные траектории входят и в условие квантования для орбит в k пространстве, площади которых теперь зависят от положения центров осцилляторов относительно границы, и, следовательно, энергия зависит не только от номера уровня Ландау, но и от $k_y = \lambda_H^{-2} x_0$.

Поскольку движение в перпендикулярном границе направлении остается финитным и при зеркальном отражении характеризуется одним и тем же значением x_0, то есть одним и тем же периодом, то спектр электрона, как и в случае объемных орбит, квантуется. (При случайном, диффузном рассеянии от границы x_0 является случайной величиной и говорить о квантовании спектра не приходится, поскольку ширина уровней будет того же порядка что и расстояние между ними). Для количественного анализа мы ограничимся случаем квадратичного закона дисперсии, а квантование проведем в использующем предельно простую математику квазиклассическом приближении. Выпишем еще раз уравнение, определяющее спектр электрона в магнитном поле (как и выше, используем линейную калибровку $A(0, xH)$)

$$\int_{x_1}^{x_2} k_x \cdot dx = \pi(n+\gamma) = \pi(n+\gamma_1+\gamma_2),$$

выделив в фазовом факторе γ части, проистекающие от левой (γ_1) и правой (γ_2) точек поворота и определяемые из сшивки волновых функций внутри и вне магнитной потенциальной ямы $m\omega_c^2(x-x_0)^2/2$. Входящая в интеграл величина k_x должна быть выражена через энергию и потенциал из функции Гамильтона

$$\mathcal{H} = \varepsilon = \frac{\hbar^2 k_x^2}{2m} + \frac{m\omega_c^2(x-x_0)^2}{2}, \quad x_0 = \lambda_H^2 k_y.$$

Далее удобно за начало отсчета принять координату центра осциллятора. Тогда $k_x = (m\omega_c/\hbar)\sqrt{a^2-x^2}$, где $a^2 = 2\varepsilon/m\omega_c^2$ и уравнение для определения уровней энергии после простейшего интегрирования сводится к виду

$$\varepsilon = \hbar\omega_c\pi(n+\gamma_1+\gamma_2)\left[\frac{x_2}{a}\sqrt{1-\frac{x_2^2}{a^2}} - \frac{x_1}{a}\sqrt{1-\frac{x_1^2}{a^2}} + \arcsin\frac{x_2}{a} - \arcsin\frac{x_1}{a}\right]^{-1} \tag{7.1}$$

Для состояний в объеме (удаленных от границ на расстояние большее радиуса циклотронной орбиты $a = r_H$ - см. рис. 7.2 а) точки поворота определяются условием $k_x = 0$, что дает $x_1 = -a$, $x_2 = a$. Поскольку

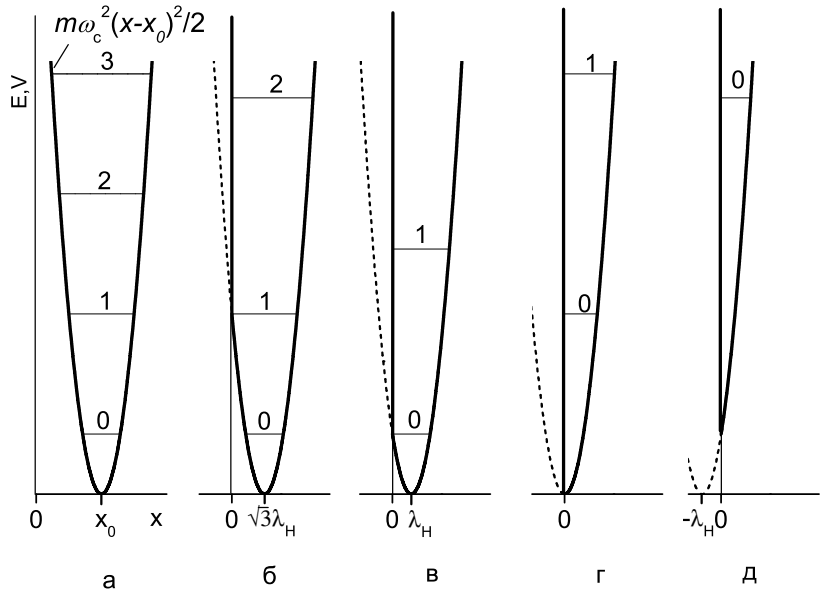

Рис. 7.2: Эффективный потенциал и уровни энергии в магнитном поле при различных расстояниях x_0 (в единицах магнитной длины λ_H) центра осциллятора от границы образца.

потенциал в этом случае определяется только магнитным потенциалом $m\omega_c^2 x^2/2$, а последний в малой окрестности обеих точек поворота линеен, то фазовые факторы равны $\gamma_1 = \gamma_2 = 1/4$. В результате получаем известный нам результат $\varepsilon = \hbar\omega_c(n + 1/2)$, не зависящий от положения центра осциллятора относительно границ образца (математически это есть следствие того, что координаты точек поворота в этом случае "выпадают" из выражения в квадратных скобках уравнения (7.1)).

Если же координата какой либо границы (или обеих границ) попадает внутрь интервала $-a < x < a$, то движение частицы определяется не только магнитным потенциалом, но и потенциалом границы образца, который можно, не умаляя общности, рассматривать как непроницаемую для электрона потенциальную стенку. Точки поворота $x_{1,2} = x(k_x = 0)$ в этом случае определяются положением потенциальной стенки относительно центра осциллятора, а величина фазового фактора для соответствующей точки поворота равна $1/2$ (в соответствии с обращением волновой функция на непроницаемой стенке в ноль). Далее мы ограничимся только состояниями на левой границе, полагая, что для всех магнитных орбит расстояние до второй стенки, т.е. ширина двумерной систе-

мы, намного превышает циклотронный радиус. В этом случае $\gamma_1 = 1/2$, $\gamma_2 = 1/4$, $-a < x_1 < a$ и $x_2 = a$, что приводит к трансцендентному уравнению (поскольку a также зависит от энергии) для энергии

$$\varepsilon_n(x_1) = \hbar\omega_c\pi(n + \frac{3}{4})\left[\frac{\pi}{2} - \frac{x_1}{a}\sqrt{1 - \frac{x_1^2}{a^2}} - \arcsin\frac{x_1}{a}\right]^{-1} \tag{7.2}$$

Для каждого значения x_1 (каждой "скачущей" орбиты) мы имеем дискретный спектр, описываемый квантовым числом n (поскольку $a = a(\varepsilon)$, то зависимость ε от n не имеет простой формы $\varepsilon \propto (n + 3/4)!$), однако для разных x_1 значения энергий отличаются, а именно ε_n монотонно растут по мере приближения центра орбиты к границе. Такое возрастание энергии ясно и из простого физического соображения – ширина потенциальной ямы по мере приближения орбиты к границе, что соответствует приближению потенциальной стенки к правой точке поворота, уменьшается – рис. 7.2 а-д (а при изменении x_1 в интервале от $x_1 = 0$ до $x_1 = a$ поднимается и дно потенциальной ямы (рис. 7.2 д)) и уровни "выталкиваются" вверх. Простое аналитическое выражение для спектра получается для частного случая орбит с центром осциллятора на границе (рис. 7.2 г). В этом случае $x_1 = 0$ и выражение (7.1) сразу приводит к эквидистантному спектру $\varepsilon_n(x_1 = 0) = 2\hbar\omega_c(n + 3/4)$. Расстояние между энергетическими уровнями для этих состояний вдвое больше, чем это имеет место для уровней Ландау в объеме.

Поскольку положение центра осциллятора может принимать произвольные значения из интервала $-a < x_1 < a$, спектр граничных магнитных состояний является фактически непрерывным. Вырождение по положению центров осциллятора, а, следовательно, и по $k_y = \lambda_H^{-2}x_1$ (здесь фигурирует x_1, поскольку за начало отсчета нами принята координата центра осциллятора $x_0 = 0$!) для пограничных состояний снимается в полном соответствии с выводами проведенного в начале настоящего раздела качественного рассмотрения. Поскольку энергия поверхностных состояний зависит от $k_y = \lambda_H^{-2}x_1$, то групповая скорость электрона в параллельном границе направлении

$$\bar{v}_y = \nabla_{\hbar k_y}\varepsilon_n(k_y) = \lambda_H^2\partial\varepsilon_n(x_1)/\hbar\partial x_1$$

уже не равна нулю, как это имело место для объемных магнитных орбит. Это полностью согласуется с классической картиной скачущих траекторий вблизи границы. Направленному потоку электронов вдоль границы соответствует электрический ток $J_y \propto ev_y$, существующий и без приложенного к двумерной системе электрического поля. Ясно, что на противоположной границе образца течет ток обратного направления, так что полный ток через любое сечение образца равен нулю. Поскольку с таким циркулирующим по периметру образца током связан магнитный поток, то поверхностные магнитные орбиты определяют диамагнетизм электронов проводимости.

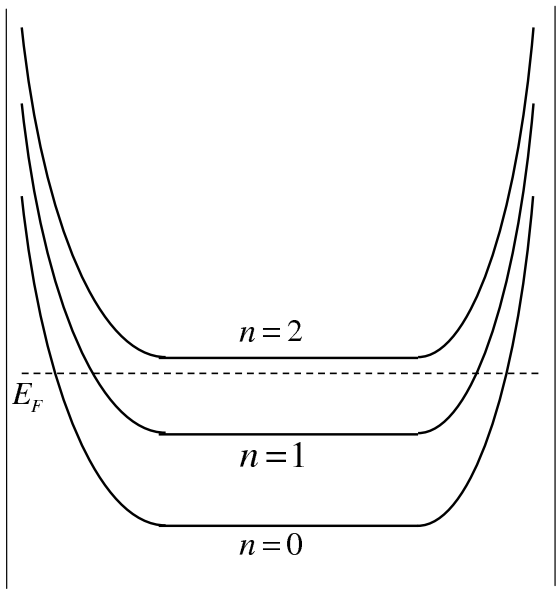

Рис. 7.3: Ход уровней Ландау в двумерном электронном газе внутри и вблизи границ образца.

Как уже указывалось, понятие о квантованных поверхностных состояниях имеет смысл только для случая зеркального отражения от поверхности. Характер рассеяния определяется соотношением между характерными размерами искажений поверхности Δ и дебройлевской длиной волны $\lambda\!\!\!^{-}$ падающих электронов - при больших $\lambda\!\!\!^{-}$ электроны не “чувствуют” мелких деталей $\Delta << \lambda\!\!\!^{-}$ рельефа рассеивающей поверхности. Поскольку длина волны электрона в перпендикулярном границе направлении $\lambda\!\!\!^{-}_y = h/mv_y$ растет с уменьшением компоненты скорости v_y, то наиболее

благоприятные условия для зеркального отражения имеют место для почти скользящих орбит с $(a - x_1)/a << 1$ $(x_1 > 0)$.

Квантование для таких состояний удобнее провести в рамках также квазиклассического, но несколько отличающегося от использованного выше, подхода. При малых углах падения θ параллельная поверхности компонента скорости достигает своего экстремального значения, близкого к скорости движения по циклотронной орбите $\omega_c r_H$ (см. выше), и слабо зависит от θ. На движущийся вдоль поверхности с почти постоянной скоростью электрон (поскольку $v_y(\theta) \approx const$)) действует почти постоянная, прижимающая его к поверхности (к связанному с границей потенциальному барьеру), сила Лоренца

$$F_L = \frac{e}{c} v_y H.$$

Соответствующий этой силе потенциал, удовлетворяющий граничному условию $V(0) = 0$, равен $V(x) = -eHv_y x/$, $(F_L = -dV/dx)$, что эквивалентно притягивающему к поверхности однородному электрическому полу напряженностью $E_s = Hv_y/c$ (сравните с результатами разд. 7.2). Поскольку со стороны границы имеется непроницаемый барьер, задача полностью эквивалентна уже решенной нами в разд. 5.6.1 задаче квантования спектра в треугольной потенциальной яме, энергетические уровни в которой описываются выражением (5.56). Подстановка $E_s = Hv_y/c$ в (5.56) сразу дает искомый энергетический спектр

$$\varepsilon_{n,v_y} = \left[\frac{3}{2\sqrt{m}}\pi\hbar\frac{eHv_y}{c}(n + \frac{3}{4})\right]^{2/3} = \left[\frac{3}{2\sqrt{m}}\pi\hbar\omega_c p_y(n + \frac{3}{4})\right]^{2/3} \qquad (7.3)$$

Отметим, что связанные с квантованием скользящих граничных траекторий осцилляции плотности состояний с изменением магнитного поля при фиксированной энергии (в реальных магнитоосцилляционных эффектах, связанных с этими особенностями, фиксирована энергия Ферми) периодичны по магнитному полю - номер осцилляции (номер уровня Ландау, совпадающего с уровнем Ферми) определяется вытекающим из выражения (7.3) соотношением

$$n = \frac{2}{3}\varepsilon_F^{3/2}\frac{c\sqrt{2m}}{\pi\hbar e v_y}H - \frac{3}{4} \qquad (7.4)$$

Следует подчеркнуть, что подобная зависимость для стандартного (внутри образца) двумерного спектра в магнитном поле $\varepsilon = \hbar\omega_c(n+1/2)$ имеет вид

$$n = \frac{\varepsilon_F mc}{e\hbar}\frac{1}{H} - \frac{1}{2}, \tag{7.5}$$

т.е. осцилляции в этом случае (осцилляции Шубникова де Газа) периодичны не по H, а по величине обратной магнитному полю $1/H$.

Существование граничных магнитных состояний не является спецификой только двумерных систем, поскольку динамика электронов в магнитном поле для трехмерной системы, как неоднократно уже подчеркивалось, аналогична случаю двумерного движения (за исключением наличия дополнительной степени свободы в направлении магнитного поля в координатном пространстве). В заключение отметим, что вклад граничных состояний в общую плотность состояний двумерной системы незначителен (за исключением образцов крайне малых размеров), поскольку число таких состояний пропорционально периметру образца (в случае трехмерной системы площади), тогда как полное число состояний его площади (объему в трехмерной системе). Тем не менее, в целом ряде эффектов именно приграничные области могут давать значительный вклад в измеряемую величину (например, при исследовании высокочастотных эффектов, ввиду затухания электромагнитного излучения в проводящей среде на сравнительно небольшой глубине от поверхности), а для некоторых явлений такие состояния играют определяющую роль (диамагнетизм электронов проводимости, квантование двумерной холловской проводимости в одной из моделей этого явления (см. разд. 9.4)).

7.2 Динамика электрона проводимости в скрещенных полях

В большинстве эффектов, лежащих в основе экспериментальных методов исследования электронных систем, наряду с магнитным полем к системе приложено также электрическое поле. Поскольку в случае двумерных систем доминирующее влияние на электроны проводимости оказывает перпендикулярная двумерному слою компонента магнитного поля,

а тянущее электрическое поле приложено в двумерной плоскости, мы по существу почти всегда имеем дело со случаем скрещенных (перпендикулярных) магнитного и электрического полей. Следует сразу отметить, что поскольку электрический потенциал для однородного электрического поля линеен по координате $V(r) = -eEr$, а эффективный магнитный потенциал $m\omega_c^2(r-r_0)^2/2$ квадратичен, магнитное поле приводит к большей перестройке спектра и электрическое поле зачастую можно рассматривать как малое возмущение. Если магнитное поле кардинально модифицирует спектр, делая его полностью дискретным, то электрическое поле не меняет характера энергетического спектра (связанное с отражением от границ зоны Бриллюэна квантование Ванье, как показано ранее, не может проявиться практически ни при каких условиях).

Для рассматриваемой ориентации полей нетрудно видеть, что задача может быть фактически сведена к случаю чисто магнитного поля. Действительно, мы знаем, что псевдоскаляр $(\mathbf{E} \cdot \mathbf{H})$ и скаляр $H^2 - E^2$ являются инвариантами электромагнитного поля при преобразованиях Лоренца. Из первого инварианта следует, что скрещенные поля $\mathbf{E} \perp \mathbf{H}$ при любом преобразовании остаются скрещенными, а соответствующим преобразованием Лоренца можно добиться, чтобы электрическое поле во втором инварианте обратилось в ноль $H^2 - E^2 = H'^2$. Для этого нужно перейти в систему отсчета, движущуюся со скоростью $\mathbf{v} = c[\mathbf{E} \times \mathbf{H}]/H^2$, записать в ней уже известные нам уравнения движения в магнитном поле, и затем преобразовать их при обратном переходе в исходную лабораторную систему. Мы, однако, обойдемся без этой достаточно длинной процедуры - для нас это важно только как подсказка, что такое сведение задачи скрещенных полей к задаче одного магнитного поля возможно.

7.2.1 Классическая динамика

Покажем это вначале для классического движения [1] электрона в скрещенных полях. При этом мы для общности будем исходить из трехмерного уравнения движения

$$\frac{d\mathbf{p}}{dt} = e\left\{\mathbf{E} + \frac{1}{c}\left[\mathbf{v}\times\mathbf{H}\right]\right\} \tag{7.6}$$

с целью продемонстрировать, что и в этом случае наиболее существенные особенности движения носят двумерный характер. Нетрудно видеть, что в случае $\mathbf{E}\perp\mathbf{H}$ справедливо тождество

$$\mathbf{E} = -\frac{[\mathbf{E}\times\mathbf{H}]\times\mathbf{H}}{H^2}.$$

Подставляя это выражение для электрического поля в уравнение движения и вводя имеющую размерность скорости величину

$$\mathbf{v}_{HE} = c\frac{[\mathbf{E}\times\mathbf{H}]}{H^2}, \tag{7.7}$$

видим, что уравнение движения в скрещенных полях действительно преобразуется к виду

$$\frac{d\mathbf{p}}{dt} = \frac{e}{c}\left[(\mathbf{v}-\mathbf{v}_{HE})\times\mathbf{H}\right], \tag{7.8}$$

аналогичному уравнению для движения только в магнитном поле, если под скоростью понимать величину $\mathbf{v}-\mathbf{v}_{HE}$. Таким образом, в скрещенных полях наряду с циклотронным (вращательным) движением электрон приобретает не изменяющуюся со временем дрейфовую скорость в направлении, перпендикулярном электрическому и магнитному полю и равную по величине $v_{HE} = cE/H$. Классические двумерные траектории

[1] Еще раз напомним, что в случае электронов в кристаллической решетке под движением электрона мы всегда понимаем полуклассическое движение, поскольку только квантовая механика оправдывает возможность описания электрона проводимости во внешних полях, как свободного (не чувствующего в явном виде периодического потенциала) электрона с перенормированным законом дисперсии (метод эффективной массы). В контексте настоящего раздела термин "классическое движение" используется в том смысле, что определяемые внешними полями (в том числе полями рассеивающих центров) характерные размеры траектории электрона намного превышают его дебройлевскую длину волны, т.е. движение квазиклассично и может быть описано классически. С точки зрения математического описания полуклассическое движение электрона проводимости в этом случае полностью аналогично классической динамике свободного электрона, что и оправдывает использование понятия "классическое движение"

электронов проводимости в скрещенных полях (рис. 7.4 см. *задание* 1 настоящей главы) представляют собой в общем случае трахоиды (укороченные и удлиненные циклоиды), вид которых зависит от вектора начальной (в момент включения магнитного поля или после столкновения с дефектом) скорости электрона [2].

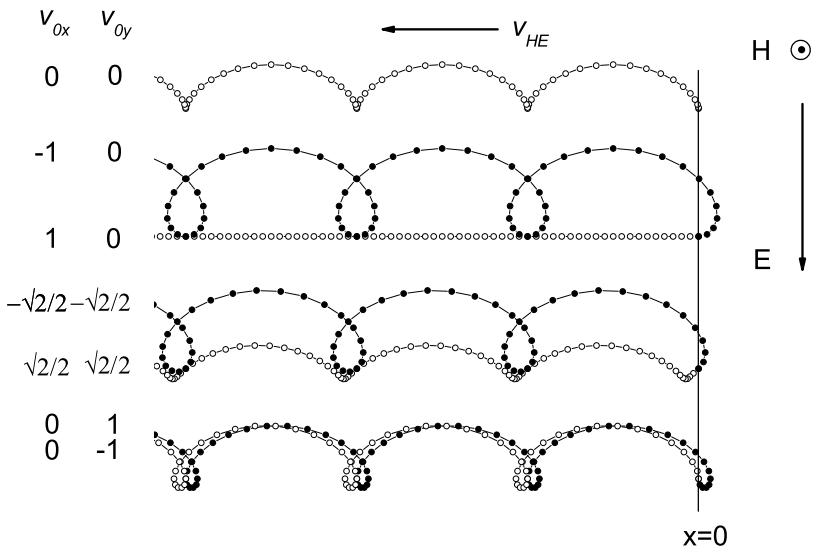

Рис. 7.4: Классические траектории электрона в двумерной плоскости в скрещенных электрическом и магнитном полях при различных начальных скоростях ($x = 0$ при $t = 0$, компоненты начальной скорости приведены в единицах v_{HE}). Соседним точкам соответствуют одинаковые интервалы времени. Циклоида (верхняя кривая) соответствует первоначально покоящемуся электрону. Скорость электрона, имевшего в начальный момент времени скорость равную по величине и направлению дрейфовой, не изменяется во времени (третья сверху траектория).

Естественно, говорить о дрейфе электрона можно только на масштабах, значительно превышающих радиус циклотронной орбиты r_H (или за промежутки времени, значительно превышающие период циклотронный период). С ростом магнитного поля r_H уменьшается и указанный критерий может выполняться на достаточно малых пространственных масштабах (допускаемых, конечно, пределами применимости самого метода эффективной массы). При этом, однако, уменьшается и величина

[2]В трехмерном случае электрон имеет также постоянную компоненту скорости в направлении магнитного поля, поэтому дрейфует под углом к магнитному полю в плоскости, перпендикулярной **E**.

дрейфовой скорости и в пределе $H \to \infty, v_{HE} \to 0$, те электрон останавливается (и становится близок к точечному объекту поскольку $\lambda_H \to 0$). Физически это понятно, поскольку в случае предельно высоких магнитных полей влиянием электрического поля, сообщающего электрону дрейфовое движение (вдоль $\mathbf{E}$ в нулевом магнитном поле и перпендикулярно $\mathbf{E}$, если $H \neq 0$), можно пренебречь.

Более критичным является предел слабых магнитных полей, поскольку скорость дрейфа даваемая выражением (7.7) при $H \to 0$ стремится к бесконечности, что, конечно лишено смысла – при заданном магнитном поле электрическое поле не может превышать H, поскольку в противном случае $v_{HE} > c$! Более того, так как наши результаты основаны на нерелятивистском приближении (в кристаллических электронных системах это совершенно оправдано, поскольку неизбежные процессы рассеяния ограничивают предельные значения скоростей значениями $v_{HE} << c$), то диапазон допустимых полей ограничен, по-существу, требованием $E << H$. Ранее уже подчеркивалось, что, как правило, электрическое поле может рассматриваться слабым по сравнению с магнитным. Теперь мы видим, что предполагать малость E по сравнению с H не только можно, но и должно с точки зрения применимости полученных результатов.

Выше рассматривалось бесстолкновительное движение носителей в идеальной решетке, однако ясно, что в промежутках между столкновениями характер движения в скрещенных полях остается по-существу таким же, если длина свободного пробега значительно больше циклотронного радиуса. Поскольку $\mathbf{E} = -\nabla U(x)$, электрон дрейфует вдоль эквипотенциали $U(x) = const$, которая в однородном поле есть прямая, перпендикулярная $\mathbf{E}$ (наряду с дрейфом имеется, конечно, циклотронное вращение заряда по орбите с радиусом r_H, величина которого уменьшается с ростом магнитного поля). Однако, даже если эквипотенциаль искривлена, но кривизна мала по сравнению с r_H, по-прежнему можно говорить о локальном движении в скрещенных полях, характеризуемым дрейфом в направлении перпендикулярном локальному электрическому полю. Показательно, что дрейф в скрещенных полях носит универсаль-

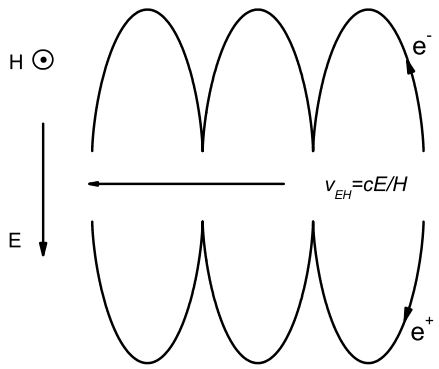

Рис. 7.5: Траектории покоившихся в начальный момент времени электрона и дырки в скрещенных полях.

ный характер - скорость дрейфа (7.7) не зависит ни от массы, ни от величины или знака заряда частицы, хотя конечно, сам дрейф возможен только при наличии у частицы заряда. Таким образом, как электроны, так и дырки дрейфуют с одинаковой скоростью и в одном направлении (рис. 7.5). Именно благодаря этому по знаку эффекта Холла, являющегося проявлением дрейфа в скрещенных полях, можно определить знак носителей заряда, поскольку переносимый носителями электрический ток имеет противоположное направление для движущихся в одном направлении электронов и дырок.

Плотность тока, переносимого носителями заряда при их дрейфе в скрещенных полях, по определению равна $j = dQ/Sdt = nev_{HE} = necE/H$ (здесь S - сечение образца, n - двумерная концентрация носителей). Поскольку для рассматриваемой ориентации полей дрейфовый ток в двумерной системе перпендикулярен электрическому полю, проводимость, являющаяся коэффициентом пропорциональности между током и полем, является тензором. Полагая электрическое поле направленным вдоль оси x $\mathbf{E} = (E_x, 0)$, для плотности тока имеем $\mathbf{j} = (0, j_y)$, т.е. недиагональная (холловская) компонента тензора проводимости равна $\sigma_{yx} = nec/H$, тогда как диагональная компонента σ_{xx} равна нулю, поскольку переноса заряда в направлении электрического поля в среднем за циклотронный период нет. Хотя процессы рассеяния приводят к тому, что появляется перенос заряда как поперек, так и вдоль электрического поля, в некоторых условиях, как мы увидим позднее при рассмотрении

эффекта Холла в двумерной системе, тензор проводимости в магнитном поле действительно имеет вид, характерный для переноса в идеальной (в отсутствии рассеяния) системе, в образцах макроскопического размера (намного превышающего длину свободного пробега электрона в отсутствии магнитного поля).

Обратимся теперь к вопросу о классическом движении в скрещенных полях электрона с произвольным законом дисперсии $\varepsilon(\mathbf{p})$. В случае магнитного поля, вследствие законов сохранения энергии и продольной компоненты импульса, траектория описывается пересечением изоэнергетической поверхности и плоскости $p_z = const$. Из уравнения движения (7.8), умножением обеих его сторон скалярно на $\mathbf{H}$, сразу следует, что вне зависимости от вида закона дисперсии компонента p_z остается интегралом движения и в случае скрещенных полей. Что касается энергии, то при включении электрического поля появляется дополнительный кинетический член, связанный с дрейфовой скоростью. Существенно, однако, что поскольку дрейфовая скорость перпендикулярна электрическому полю, последнее не производит работы по перемещению заряда в направлении $\mathbf{v}_{HE}$ (именно по этой причине дрейфовая скорость от времени не зависит). То же самое относится и к среднему движению электрона в направлении электрического поля. Поскольку (циклотронное) движение в этом направлении периодично, работа поля в среднем период равна нулю (перемещение заряда в направлении $\mathbf{E}$ в среднем равно нулю, как и в случае движения только в магнитном поле) [3]. Можно сказать, что магнитное поле не только само не производит работы над зарядом, но и не дает совершать ее и электрическому полю, если последнее перпендикулярно $\mathbf{H}$.

Однако мгновенное значение энергии (на отдельных участках траектории) $\varepsilon(\mathbf{p})$ теперь не является инвариантом движения. Действительно,

$$\frac{d\varepsilon}{dt} = \nabla_p \varepsilon \frac{d\mathbf{p}}{dt} = \frac{e}{c}\mathbf{v}\left[(\mathbf{v} \times \mathbf{H}\right] - \frac{e}{c}\mathbf{v}\left[\mathbf{v}_{HE} \times \mathbf{H}\right],$$

где первый член, как и в чисто магнитном поле, по прежнему равен нулю,

[3] Следует подчеркнуть однако, что классическое рассмотрение не является последовательным, поскольку излучение электрона при его ускоренном (циклотронном) движении не позволяет, как и в случае атома, объяснить существование стационарных орбит

а второй после подстановки выражения для $\mathbf{v}_{HE}$ приводит к ненулевому значению $d\varepsilon(p)/dt$

$$\frac{d\varepsilon}{dt} = \mathbf{v}\frac{d\mathbf{p}}{dt} = e\mathbf{v}\mathbf{E}. \tag{7.9}$$

Легко, однако, видеть, что в скрещенных полях, в силу независимости $\mathbf{v}_{HE}$ от времени, инвариантом движения является разность $\varepsilon(p) - \mathbf{p}\mathbf{v}_{HE}$. Действительно,

$$\frac{d}{dt}(\varepsilon(p) - \mathbf{p}\mathbf{v}_{HE}) = \nabla_p\varepsilon\frac{d\mathbf{p}}{dt} - \mathbf{v}_{HE}\frac{d\mathbf{p}}{dt} = (\mathbf{v} - \mathbf{v}_{HE})\frac{d\mathbf{p}}{dt},$$

что после подстановки $d\mathbf{p}/dt$ из уравнения движения (7.8) дает

$$\frac{d}{dt}(\varepsilon(p) - \mathbf{p}\mathbf{v}_{HE}) = \frac{e}{c}(\mathbf{v} - \mathbf{v}_{HE})\left[(\mathbf{v} - \mathbf{v}_{HE}) \times \mathbf{H}\right] = 0.$$

Таким образом, и в случае произвольного закона дисперсии $\varepsilon(\mathbf{p})$ движение частицы в скрещенных полях также может трактоваться, как движение только в магнитном поле частицы с законом дисперсии

$$\varepsilon^*(p) = \varepsilon(p) - \mathbf{p}\mathbf{v}_{HE}. \tag{7.10}$$

Траектория частицы в импульсном пространстве определяется теперь плоской кривой, задаваемой уравнениями

$$\varepsilon^*(p) = \varepsilon(p) - \mathbf{p}\mathbf{v}_{HE} = const, \quad p_z = const. \tag{7.11}$$

Связь траекторий в пространстве импульсов и в координатном пространстве дается следующим из (7.9) уравнением

$$\mathbf{v}\left(\frac{d\mathbf{p}}{dt} - e\mathbf{E}\right) = \frac{d\mathbf{r}}{dt}\frac{d}{dt}(\mathbf{p} - e\mathbf{E}t) = 0,$$

из которого видно, что скорость в координатном пространстве перпендикулярна скорости в импульсном пространстве, если под импульсом понимать величину $\mathbf{p}^* = \mathbf{p} - e\mathbf{E}t$. Если траектории в импульсном пространстве замкнуты, то совершенно ясно, что полученные ранее формулы для вычисления таких параметров движения в магнитном поле, как периоды обращения электрона, циклотронные массы и др. просто переносятся на случай скрещенных полей путем замены $\varepsilon(\mathbf{p}) \to \varepsilon^*(\mathbf{p})$ и $\mathbf{p} \to \mathbf{p}^*$

7.2.2 Квантование спектра в скрещенных электрическом и магнитном полях

В соответствии с установленной выше аналогией движения в скрещенных полях с движением только в магнитном поле с модифицированным законом дисперсии, задача квантования спектра может быть сведена к уже решенной нами задаче определения спектра в магнитном поле. В присутствии электрического поля в гамильтониане системы появляется новый член – потенциальная энергия электрона $V(r)$ в электрическом поле, равная в случае направленного по оси x однородного поля $V(x) = -eEx$. При выборе калибровки для векторного потенциала $\mathbf{A}(0, xH)$ выбор направления электрического поля$\mathbf{E}$ перпендикулярно $\mathbf{A}$ приводит к уравнению Шредингера, зависящему только от одной координаты . Волновая функция, таким образом, может быть представлена в точно том же виде, что и в случае магнитного поля $\phi(x, y) = \varphi(x) \exp ik_y y$.

Ниже мы ограничимся случаем частицы с квадратичным законом дисперсии. Действуя совершенно аналогично тому, как это делалось при квантовании спектра в магнитном поле (разд. 6.4) приходим к гамильтониану

$$\hat{\mathcal{H}} = \frac{\hbar^2 \hat{k}_x^2}{2m} + \frac{m\omega_c^2 (x - x_0)^2}{2} - eEx, \tag{7.12}$$

отличающемуся от гамильтониана при $E = 0$ только потенциальной энергией электрона в электрическом поле (для использованного выбора $\mathbf{A}(0, xH)$ и $V(x) = -eEx$ это, собственно, физически очевидный результат, который можно выписать сразу). Очевидно, что соответствующее гамильтониану (7.12) уравнение может быть преобразовано к стандартному для гармонического осциллятора виду

$$\left[\frac{\hbar^2 \hat{k}_x^2}{2m} + \frac{m\omega_c^2 (x - x_0')^2}{2} - C\right] \varphi(x) = \varepsilon \varphi(x)$$

с независящими от координаты величинами x_0' и C. Таким образом, мы возвращаемся к уравнению Шредингера для заряда в магнитном поле

$$\left[\frac{\hbar^2 \hat{k}_x^2}{2m} + \frac{m\omega_c^2 (x - x_0')^2}{2}\right] \varphi(x) = \varepsilon' \varphi(x), \tag{7.13}$$

($\varepsilon' = \varepsilon + C$) с известными нам собственными значениями $\varepsilon'_n = \hbar\omega_c(n + 1/2)$. Величины x'_0 и C могут быть найдены методом неопределенных коэффициентов, что приводит к модифицированной координате центра осциллятора

$$x'_0 = \lambda_H^2 \left(k_y + k_{HE}\right), \tag{7.14}$$

где

$$k_{HE} = \frac{mcE}{\hbar H} = \frac{mv_{HE}}{\hbar}$$

и выражению для введенного выше параметра C

$$C = c\hbar k_y \frac{E}{H} + \frac{mc^2}{2}\frac{E^2}{H^2}.$$

Таким образом, двумерный спектр в скрещенных электрическом и магнитном полях имеет вид

$$\varepsilon_{nk_y} = \hbar\omega_c(n + \frac{1}{2}) - c\hbar k_y \frac{E}{H} - \frac{mc^2}{2}\frac{E^2}{H^2}. \tag{7.15}$$

Физический смысл дополнительных, по сравнению со случаем чисто магнитного поля, второго и третьего слагаемых в (7.15) становится особенно ясным, если их переписать в виде

$$\Delta\varepsilon_{x_0} = c\hbar k_y \frac{E}{H} = eEx_0, \tag{7.16}$$

$$\Delta\varepsilon_{HE} = \frac{mc^2}{2}\frac{E^2}{H^2} = \frac{mv_{HE}^2}{2}. \tag{7.17}$$

Первый из них есть потенциальная энергия заряда с координатой x_0 в электрическом поле, второй - кинетическая энергия, отвечающая дрейфу электрона со скоростью v_{HE} в перпендикулярном электрическому и магнитному полю направлении.

Как это и следует из общих соображений, изложенных в начале настоящего раздела, вырождение спектра по квантовому числу $k_y = \lambda_H^{-2} x_0$ и сопряженному с ним положению центра осциллятора x_0 снимается, поскольку электрическое поле нарушает изотропность пространства (в двумерной плоскости появляется выделенное направление) и осцилляторы с разными координатами их центров теперь не эквивалентны. Поскольку энергия в скрещенных полях зависит от k_y, y-я компонента средней

групповой скорости не равна нулю, как это имело место в случае одного магнитного поля, а равна

$$\bar{v}_y = \nabla_{\hbar k_y}\varepsilon = cE/H$$

в полном соответствии с классическим результатом для скорости дрейфа в этом направлении (перемещение в параллельном электрическому полю направлении в среднем за период по-прежнему отсутствует $\bar{v}_x = \nabla_{\hbar k_x}\varepsilon = 0$). Мы видим, что и в случае скрещенных полей скорость по оси y не равна $\hbar k_y/m$, как могло бы показаться, исходя из подобия волновой функции $\exp(ik_y y)$ и волновой функции, описывающей действительно свободное по оси y движение электрона проводимости. Хотя в отличие от случая $E = 0$ скорость в этом направлении $\bar{v}_y = \nabla_{\hbar k_y}\varepsilon$ теперь не равна нулю, ее величина от квазиимпульса $\hbar k_y$ не зависит. Как мы видим, в случае скрещенных полей калибровка Ландау хорошо согласуется с классической картиной движения электрона.

Воздействие электрического поля на двумерную систему электронов проводимости в магнитном поле приводит, таким образом, к одинаковому для всех значений квантового числа n сдвигу $\Delta\varepsilon_{HE}$ уровней Ландау $\varepsilon_n = \hbar\omega_c(n + 1/2)$ и, также одинаковому для всех уровней, "размазыванию" вокруг ε_n в полосу с квазинепрерывным спектром, расстояние между квазидискретными уровнями которой и полная ширина определяются разрешенными значениями $k_y = \lambda_H^{-2}x_0$. Отношение энергетического сдвига $\Delta\varepsilon_{HE}$ и ширины полосы к расстоянию между уровнями Ландау $\hbar\omega_c$, как видно из полученных выше выражений, быстро убывает с ростом магнитного поля ($\Delta\varepsilon_{HE}/\hbar\omega_c \propto E^2/H^3$, $\Delta\varepsilon_{x_{0\,\max}}/\hbar\omega_c \propto E/H^2$) и для актуальных значений полей величина сдвига и размытие малы по сравнению с циклотронной энергией. Это полностью подтверждает основанный ранее на качественных соображениях вывод о том, что в большинстве эффектов в магнитных полях электрическое поле можно рассматривать как слабое возмущение, т.е. макроскопические свойства в скрещенных полях определяются в первую очередь динамикой электронов проводимости и их спектром в магнитном поле.

Поскольку, согласно теореме Лиувилля, полное число состояний консервативной системы не зависит от внешних воздействий, их полное чис-

ло в полосе состояний, возникшей вокруг уровней Ландау, благодаря снятию вырождения спектра по положению осцилляторов, в точности равно кратности вырождения исходных (в отсутствии электрического поля) уровней Ландау. Это непосредственно следует и из того, что циклические граничные условия, налагаемые на волновую функцию в направлении дрейфа $\exp(ik_y y)$, те же самые, что и в случае $E = 0$, что, очевидно, приводит к тому же числу состояний $n_H = eH/ch$.

В случае частиц с произвольным законом дисперсии, ввиду аналогии движения носителей заряда с законом дисперсии $\varepsilon(\mathbf{p})$ в скрещенных полях и движением носителей с модифицированным законом дисперсии $\varepsilon^*(\mathbf{p}) = \varepsilon(\mathbf{p}) - \mathbf{p}\mathbf{v}_{HE}$ только в магнитном поле, квантование может быть проведено аналогично тому, как это было сделано в случае $E = 0$. В правиле квазиклассического квантования

$$\sigma_k^* = 2\pi\lambda_H^{-2}(n + \frac{1}{2})$$

σ_k^* — теперь площадь, охватываемая двумерной траекторией $\varepsilon^*(\mathbf{k}) = \varepsilon(\mathbf{k}) - \hbar\mathbf{k}\mathbf{v}_{HE} = const$ (предполагается, что траектория в скрещенных полях замкнута) [4]. Расстояние между квазиэквидистантными уровнями определяется формулой, подобной выражению (6.27)

$$\Delta\varepsilon = \lambda_H^{-2}\left(\frac{\partial\sigma^*(\varepsilon)}{\partial\varepsilon^*)}\right)^{-1} = \hbar\omega_c(\varepsilon).$$

Поскольку в модифицированный закон дисперсии входит скорость дрейфа, частота $\omega_c(\varepsilon)$ а, следовательно, и расстояние между уровнями в случае скрещенных полей зависят от электрического поля (аналог эффекта Штарка; следует, однако, подчеркнуть, что для квадратичного закона дисперсии этот эффект пропадает - см. *задание* 2.).

Представляет интерес сравнить движение электрона в скрещенных полях с рассмотренным выше движением в магнитном поле вблизи границ образца. Последнее формально также может рассматриваться как движение в магнитном поле в присутствии электростатического потенциала $U(\mathbf{r})$ (в виде бесконечно высокой потенциальной стенки - ($U(\mathbf{r}) = 0$

[4] площадь сечения поверхности $\varepsilon^*(\mathbf{k}) = const$ плоскостью $k_z = const$ в случае трех измерений.

внутри и $V(\mathbf{r}) = \infty$ вне образца), градиент которого перпендикулярен $\mathbf{H}$. Характер движения в обоих случаях аналогичен – финитное движение в направлении $\nabla V(\mathbf{r})$ и дрейф в перпендикулярном $\nabla V(\mathbf{r})$ направлении. В обоих случаях спектр квантуется, а вырождение по положению центра осциллятора и сопряженному с ним волновому вектору p_y снимается. Однако аналогия здесь не полная. В случае скрещенных полей направление дрейфа не зависит от знака заряда носителей (зато связанный с дрейфом ток отличается по знаку), направления же скачущих приграничных траекторий для электронов и дырок, определяемые направлением вращения заряда в магнитном поле, противоположны (соответственно связанные с таким переносом токи неразличимы по направлению).

7.3 Калибровочная инвариантность и эффект Бома–Ааронова

В заключительном разделе этой главы рассмотрим один важный аспект вопроса, относящегося к сопоставлению (*полу*)классического и квантового описания движения электронов проводимости в электромагнитном поле. При полуклассическом описании (как и при классическом описании свободного электрона) система полностью описывается уравнениями движения (7.6), где в случае электромагнитного поля сила, действующая на заряженную частицу, дается формулой Лоренца, в которой фигурируют магнитное и электрическое поля. При квантово-механическом подходе рассмотрение основано на гамильтоновом описании, при котором воздействие электромагнитного поля описывается не напряженностями полей, а векторным и скалярным потенциалами $\mathbf{A}(\mathbf{r})$, $\varphi(\mathbf{r})$, определяющими гамильтониан системы

$$\frac{\hbar^2}{2m}[(\hat{\mathbf{k}} - \frac{e}{c\hbar}\mathbf{A}(\mathbf{r}))^2 + \varphi(\mathbf{r})]\Psi(\mathbf{r}) = \varepsilon\Psi(\mathbf{r}). \tag{7.18}$$

Полевые $\mathbf{H}$, $\mathbf{E}$ и потенциальные $\mathbf{A}$, $\varphi(\mathbf{r})$ характеристики электромагнитного поля связаны соотношениями

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla\varphi(\mathbf{r}), \quad \mathbf{H} = \mathbf{rot}\mathbf{A}. \tag{7.19}$$

Существенно, однако, что такая связь не является однозначной! При заданных значениях электрического и магнитного поля, векторный и скалярный потенциалы определены только с точностью до калибровочного (градиентного) преобразования

$$\mathbf{A}(\mathbf{r}) \;\rightarrow\; \mathbf{A}(\mathbf{r}) - \nabla\chi(\mathbf{r},t), \tag{7.20a}$$

$$\varphi(\mathbf{r}) \;\rightarrow\; \varphi(\mathbf{r}) + \frac{\partial\chi(\mathbf{r},t)}{\partial t}, \tag{7.20b}$$

поскольку при произвольной (см. однако, конец этого раздела) скалярной функции $\chi(\mathbf{r},t)$ преобразование (7.20) оставляет инвариантными определяемые уравнениями (7.19) полевые величины $\mathbf{E}$ и $\mathbf{H}$. В частности, почти очевидно (пример см. ниже), что всегда можно выбрать такую калибровку, при которой магнитное и (или) электрическое поле во всем пространстве (или в некоторой его области) равны нулю, тогда как векторный (скалярный) потенциал отличен от нуля. Для этого достаточно, чтобы $\chi(\mathbf{r},t)$ имела ненулевой градиент, что, в виду произвольности калибровочной функции $\chi(\mathbf{r},t)$, всегда возможно.

В классической электродинамике поведение системы полностью определяются силой Лоренца и в тех областях, где силовые поля $\mathbf{E}(\mathbf{r})$ и $\mathbf{H}(\mathbf{r})$ равны нулю, физическое воздействие электромагнитного поля отсутствует, независимо от того, равны там потенциалы $\mathbf{A}(\mathbf{r})$, $\varphi(\mathbf{r})$ нулю или нет. Так ли это в квантовой электродинамике, где воздействие электромагнитного поля определяется как раз потенциалами $\mathbf{A}(\mathbf{r})$, $\varphi(\mathbf{r})$, явным образом входящими в гамильтониан? Как правило, это, действительно, так. Как и векторный потенциал, волновая функция также определена не однозначно, а именно с точностью до произвольного фазового множителя, не влияющего на физически наблюдаемые эффекты, определяемые модулем волновой функции. Легко убедиться, что при калибровочном преобразовании $\mathbf{A}(\mathbf{r})$ уравнение Шредингера не изменяет своего вида, если одновременно с заменой $\mathbf{A}(\mathbf{r}) \rightarrow \mathbf{A}(\mathbf{r}) - \nabla\chi(\mathbf{r})$ произвести подстановку для волновой функции

$$\Psi(\mathbf{r}) \rightarrow \Psi(\mathbf{r}) \exp\left(i\frac{e}{c\hbar}\chi(\mathbf{r},t\right) = \Psi(\mathbf{r}) \exp\left(i\alpha\right)$$

Таким образом, хотя фазы волновых функций α в магнитном поле

определены лишь с точностью до произвольной функции координат, это не сказывается на решениях уравнения Шредингера. Указанная неоднозначность волновых функций в магнитном поле не отражается на каких-либо имеющих физический смысл величинах (квадрат волновой функции, величины матричных элементов и пр.), в определение которых не входит явным образом векторный потенциал. Однако даже в последнем случае не обязательно, что это может отразиться на наблюдаемых физических величинах. Например, векторный потенциал явно входит в выражение для плотности потока в магнитном поле

$$j = \frac{i\hbar}{2m}(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi) - \Psi\frac{e}{mc}\mathbf{A}(\mathbf{r})\Psi^*.$$

Однако легко непосредственным вычислением убедиться, что и это выражение вполне однозначно при одновременном калибровочном преобразовании векторного потенциала и волновой функции.

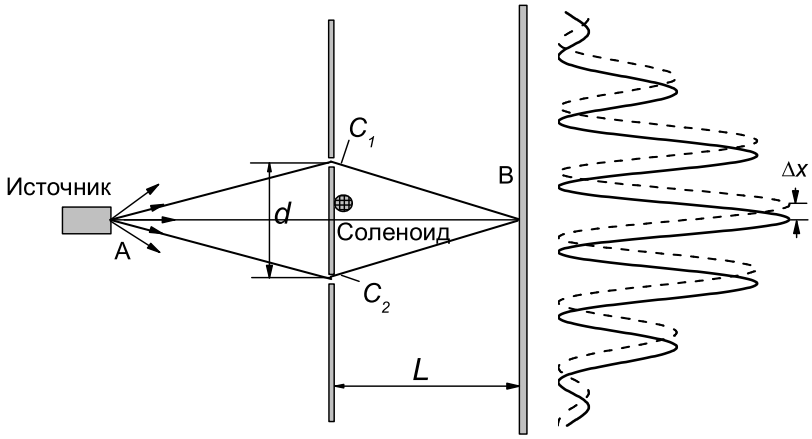

Рис. 7.6: Схема опыта по интерференции электронных волн. Сдвиг интерференционной картины при помещении соленоида внутри контура, ограниченного двумя возможными траекториями электрона, обусловлен исключительно изменением фазы волновой функции $\Delta\delta = (e/c\hbar)\int_c \mathbf{A}d\mathbf{r}$ при движении электрона по траектории в векторном поле $\mathbf{A}$. Вне соленоида магнитное поле $\mathbf{H}$ равно нулю и на орбитальное движение соленоид не влияет - все траектории в поле $\mathbf{A}$ остаются теми же, что и в отсутствии соленоида. Сравните со схемой опыта на рис. (7.8), где траектории проходят через области, где как $\mathbf{A} \neq 0$, так и $\mathbf{H} \neq 0$.

И тем не менее, ответ на поставленный выше вопрос не всегда положительный. При квантовом рассмотрении могут иметь место вполне

наблюдаемые физические эффекты в областях, где $\mathbf{E}$ и $\mathbf{H}$ равны нулю, но потенциалы $\mathbf{A}(\mathbf{r})$ или (и) $\varphi(\mathbf{r})$ отличны от нуля. Очевидно, что зависимость фазы волновой функции от векторного потенциала может отразиться на интерференции электронных волн. Остановимся на одном весьма красивом и интересном эффекте, связанном с калибровочной инвариантностью электродинамики – эффекте Бома–Ааронова. Рассматриваемое явление относится к хорошо известному квантово - механическому опыту по интерференции электронов, проходящих через две щели (рис. 7.6). В силу волновой природы электрона, пока мы не зафиксировали, через какую щель электрон попадает на экран, на последнем возникает характерная интерференционная картина. Положение максимумов и минимумов на экране определяется разницей "оптических" путей a электронных волн попадающих в данную точку экрана. Если длина волны электрона $\lambda\!\!\!^{-}$, то разность фаз для волн, приходящих от двух щелей, равна

$$\delta = 2\pi\frac{a}{\lambda\!\!\!^{-}}. \tag{7.21}$$

Если плоскость, в которой лежат щели, параллельна экрану с расстоянием между ними L, то разница путей a для точки с координатой x, отсчитываемой от точки на экране, проекция которой на линию соединяющую щели расположена посередине между ними, дается при $L >> x$ выражением $a = (xd/L)$. Отсюда получаем формулу, описывающую интерференционную картину

$$x = \frac{L\lambda\!\!\!^{-}}{2\pi d}\delta. \tag{7.22}$$

В точках, где $\delta = 2\pi n$ имеем максимумы (конструктивная интерференция), а в точках, где $\delta = \pi(2n+1)$ минимумы (деструктивная интерференция) электронной плотности.

Теперь посмотрим, как повлияет на интерференционную картину наличие в области траектории электрона ненулевого векторного потенциала, такого, однако, что напряженность магнитного поля в этой области равна нулю. Такая ситуация может быть реализована с помощью соленоида, в котором, как известно, линии магнитной индукции сосредоточены внутри, но отсутствуют вне соленоида. Нетрудно найти вид векторного потенциала, соответствующего этой ситуации. Рассмотрим длинный

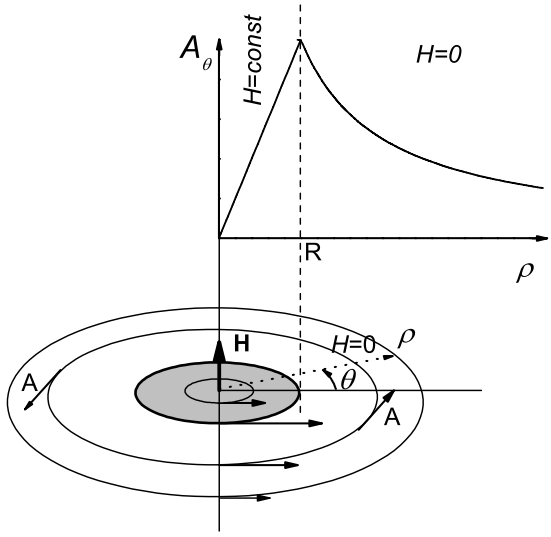

Рис. 7.7: Векторный потенциал и магнитное поле внутри и вне бесконечного соленоида с током радиусом R. Радиальная компонента векторного потенциала в выбранной калибровке равна нулю. Вне соленоида магнитного поля нет, но векторный потенциал не равен нулю.

и для простоты круглый в сечении соленоид радиуса R. Из дальнейшего будет видно, что результаты не зависят от размеров или точного положения соленоида, лишь бы он не пересекал траектории интерферирующего электрона (напомним, что в квантовой механике речь всегда идет об интерференции электрона самого с собой - разные электроны не интерферируют), поскольку последнее нарушает наше исходное предположение о нулевой магнитной индукции в области движения электрона (сила Лоренца на электрон не действует). В виду аксиальной симметрии векторный потенциал разумно выбрать независящим от полярного угла, т.е с единственной ненулевой компонентой перпендикулярной радиус вектору. В полярной системе координат (ρ, ϑ) для соленоида с полем H это соответствует калибровке (рис. 7.7))

$$A_\vartheta = \frac{H\rho}{2}, \quad A_\rho = 0 \quad \textit{внутри соленоида} \tag{7.23}$$

$$A_\vartheta = \frac{HR^2}{2\rho}, \quad A_\rho = 0 \quad \textit{вне соленоида}^5. \tag{7.24}$$

Посмотрим теперь, как действует такое поле на волновую функцию электрона. В отсутствии внешних полей мы имеем дело со свободным

[5]В выбранных координатах выражение для напряженности магнитного поля $\mathbf{H} = \mathbf{rot}\mathbf{A}$, будучи

движением электрона и волновая функция есть плоская волна

$$\Psi = |\Psi|\, e^{i\mathbf{k}\cdot\mathbf{r}} \equiv |\Psi|\, e^{i\alpha}$$

При наличии векторного потенциала волновой вектор $\mathbf{k}$ заменяется обобщенным $\mathbf{k} \to \mathbf{k} - e\mathbf{A}/c\hbar$, что приводит, после подстановки в выражение для Ψ, к изменению фазы волновой функции

$$\alpha \to \alpha - \frac{e}{c\hbar}\mathbf{A}\cdot\mathbf{r}.$$

Приращение фазы, набираемое электроном при движении вдоль траектории C, тогда равно

$$\Delta\alpha = -\frac{e}{c\hbar}\int\limits_C \mathbf{A}\cdot d\mathbf{r}.$$

Таким образом, для траекторий электрона C_1, C_2 в интерференционном эксперименте мы имеем

$$\Delta\alpha_1 = -\frac{e}{c\hbar}\int\limits_{C_1} \mathbf{A}\cdot d\mathbf{r}, \quad \Delta\alpha_2 = -\frac{e}{c\hbar}\int\limits_{C_2} \mathbf{A}\cdot d\mathbf{r}.$$

Общее изменение разности фаз может быть представлено в виде контурного интеграла по замкнутой траектории $C = C_1 + C_2$ [6]. Из написанных выше выражений сразу следует

$$\Delta\delta = \Delta\alpha_1 - \Delta\alpha_1 = \frac{e}{c\hbar}\oint\limits_C \mathbf{A}\cdot d\mathbf{r} = \frac{e}{c\hbar}\int\limits_\sigma \mathbf{rot}\mathbf{A}\cdot d\sigma, \tag{7.25}$$

где в последнем равенстве мы использовали теорему Грина. Поскольку $\mathbf{rot}\mathbf{A} = H$, то последний интеграл есть магнитный поток Φ через стягивающую контур площадь, и мы приходим к выражению

$$\Delta\delta = \frac{e}{c\hbar}\Phi = 2\pi\frac{\Phi}{\Phi_0}, \tag{7.26}$$

записано в покомпонентном виде, дает

$$H = \frac{1}{r}\left[\frac{\partial A_\vartheta}{\partial\rho} + A_\vartheta - \frac{\partial A_\rho}{\partial\vartheta}\right] \quad ,$$

откуда для калибровки (7.23–7.24) получаем $H = H$ *внутри соленоида* и $H = 0$ *вне соленоида*, что и требовалось.

[6] При вычислении контурного интеграла по C нужно помнить, что направление обхода для путей C_1 и C_2 должно быть одинаковым, например, против часовой стрелки, когда охватываемая контуром площадь σ при обходе остается слева - такое направление принимается в дифференциальной геометрии за положительное.

где Φ_0 – квант магнитного потока. В результате сдвиг интерференционной картины равен

$$\Delta x = \frac{L\lambda\!\!\!^{-}}{2\pi d}\Delta\delta = \frac{L\lambda\!\!\!^{-}}{d}\frac{\Phi}{\Phi_0}. \tag{7.27}$$

Теперь рассмотрим два случая.

1). Соленоид расположен вне контура. В этом случае магнитное поле и соответственно магнитный поток через контур равны нулю и ненулевое значение векторного потенциала в этой области не приводит к физически наблюдаемому эффекту, в полном соответствии с классической электродинамикой. По существу это есть геометрический результат, теорема дифференциальной геометрии о независимости циркуляции векторного поля $\mathbf{A}$ по замкнутому контуру от размеров или формы контура. Поскольку в рассматриваемой ситуации векторное поле не имеет особенностей в области, охватываемой контуром (оно регулярно и однозначно, поскольку вне соленоида $A_\vartheta = HR^2/2\rho = \Phi/2\pi\rho$ и $\rho \neq 0$ - см. (7.24)), контур может быть стянут в точку, для которой циркуляция, конечно, равна нулю [7].

2). Ситуация кардинальным образом отличается в случае, когда тонкий соленоид пронизывает площадь охватываемую траекторией. Магнитное поле на траектории и в этом случае равно нулю. Однако, поскольку полный магнитный поток Φ через охватываемую контуром площадь теперь не ноль, а равен πHR^2, возникает разность фаз $\Delta\delta = 2\pi\Phi/\Phi_0$ и, следовательно, имеет место наблюдаемый макроскопический эффект – сдвиг интерференционной картины. Совершенно очевидно, что в обеих случаях результат не зависит ни от положения ни от размеров соленоида (согласно (7.24) изменение R эквивалентно изменению калибровки), оставаясь справедливым и при $R \to 0$, если при этом $\Phi = const$, ни от формы или размеров траектории [8]. Хотя векторный потенциал $A_\vartheta \propto 1/\rho$ вне соленоида уменьшается с ρ, длина контура $C = 2\pi\rho$ растет, так что интеграл $\int\limits_{C_1} \mathbf{A}\cdot d\mathbf{r}$ при $\rho \to \infty$ не изменяется ($= \Phi$). Принципиальное

[7]Отсюда и следует известный результат - циркуляция по замкнутому контуру безвихревого векторного поля $\mathbf{A}$ ($\mathbf{rot}\mathbf{A} = 0$ всюду внутри контура) равна нулю.

[8]Поскольку соленоид может быть сколь угодно тонким и может быть помещен в непосредственной близости у стенки между щелями, дифракция электронов на щелях не приведет к заметной вероятности того, что траектория электрона пересечет соленоид.

отличие от случая 1) заключается в том, что хотя циркуляция векторного поля **A** по замкнутому контуру по прежнему не зависит от размеров или формы контура, контур теперь не может быть стянут в точку без пересечения области, занимаемой соленоидом, где нарушается регулярность векторного потенциала. В результате, определяющая сдвиг фаз циркуляция векторного поля, хотя и не зависит от размеров и формы траектории, однако она не может быть обращена в ноль[9].

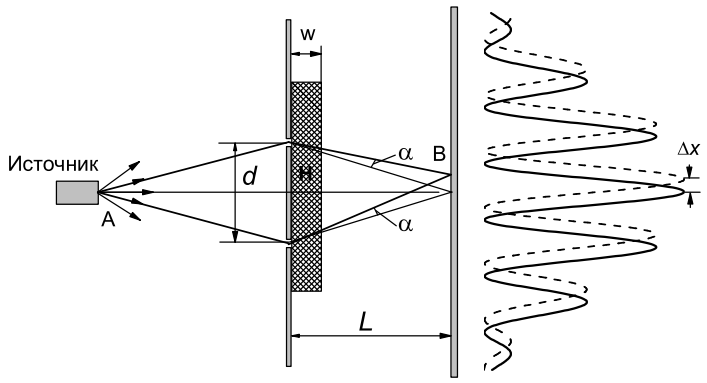

Рис. 7.8: То же что на рис. (7.6), но вместо соленоида за щелями имеется однородное магнитное поле сосредоточенное в узкой полосе шириной $W << L$. Сдвиг интерференционной картины в таком интерференционном эксперименте обусловлен отклонением траекторий (на угол α) при прохождении электроном области с магнитным полем. Квантовомеханический результат, основанный на расчете сдвига фаз через векторный потенциал, для этого случая совпадает с классическим, в рамках которого изменение траектории обусловлено силой Лоренца, действующей на электрон в области плоски. Сравните со схемой опыта на рис. (7.6).

Нетрудно видеть, что квантовомеханический результат (7.27) остается в силе, если траектория электрона пересекает область, где магнитное поле не равно нулю. Рассмотрим случай, когда позади щелей в полосе шириной w имеется однородное магнитное поле напряженности H. Чтобы подсчитать сдвиг по фазе, мы должны взять интеграл ((7.26)) от **A** по контуру, охватываемому траекторией, который по прежнему равен потоку магнитного поля пронизывающему контур, что и приводит к формуле (7.27). Полагая для простоты $w << L$, так что поток равен Hwd, для

[9]Нетрудно видеть, что, вообще говоря, случай 1) непосредственно следует из 2), однако для получения более отчетливой физической (и топологической) картины полезно рассмотреть их по отдельности.

разности фаз получаем

$$\Delta\delta = \frac{e}{c\hbar} Hwd.$$

Сдвиг же интерференционной картины равен

$$\Delta x = L\lambda\!\!\!^{-}\frac{e}{c\hbar} Hw.$$

Ввиду малости w, такой сдвиг равноценен тому, как если бы все траектории отклонились на небольшой угол α, равный

$$\alpha = \frac{\Delta x}{L} = \frac{\lambda\!\!\!^{-}}{c\hbar} eHw.$$

Однако, в отличие от случая тонкого соленоида расположенного вне траектории электрона, в рассматриваемой модификации опыта эффект должен иметь место и при чисто классическом рассмотрении, поскольку теперь электрон проходит область ненулевого магнитного поля H, где он испытывает воздействие силы Лоренца $\mathbf{F}_L = e[\mathbf{v} \times \mathbf{H}]/c$. При классическом рассмотрении эта сила, действующая в течение времени $t = w/v$ в узкой полоске магнитного поля, отклонит (в действительности отклонит!) все траектории на некоторый небольшой (т.к. $w << L$) угол β. Величина этого углового отклонения определяется отношением поперечного импульса электрона, приобретенного при прохождении поля, $\Delta p_x = F_L t = ewH/c$ к полному импульсу

$$\beta = \frac{\Delta p_x}{p} = \frac{ewH}{cp}.$$

Поскольку же частице с импульсом p ставится в соответствие волна де Бройля с $\lambda\!\!\!^{-} = \hbar/p$, то выражения для α и β идентичны, т.е. квантово-механическое и полуклассическое рассмотрения приводят к одной и той же величине эффекта.

Следует подчеркнуть, что хотя наличие векторного потенциала в области, где отсутствует электромагнитное поле, играет решающую роль в эффекте Бома–Ааронова, в явном виде векторный потенциал в окончательный результат – сдвиг интерференционной картины - по прежнему не входит (входит через инвариантную величину $\mathbf{rotA}$ в (7.25)). Таким

образом, калибровочная инвариантность наблюдаемых физических явлений, как и должно быть с точки зрения однозначности наблюдаемых эффектов, не нарушается и в квантовой области.

Тем не менее, мы видим, что векторный потенциал имеет более глубокий физический смысл, чем это дает классическая электродинамика. Эффект Бома – Аронова свидетельствует, что векторный потенциал воздействует на электрон даже тогда, когда последний движется только в областях где $H = 0$. В то же время, поскольку физические эффекты зависят только от полевой формы $\mathbf{rot}\mathbf{A}$, приходится сделать вывод, что на электрон воздействуют поля, которые отличны от нуля только в областях, недоступных для него, то есть что воздействие носит нелокальный характер. С формальной точки зрения это есть следствие нелокального характера контурного интеграла $\oint\limits_C \mathbf{A} \cdot d\mathbf{r}$, величина, которого определяется областью пространства, лежащей вне контура. На траектории электрона мы имеем $H = 0$ и $E = 0$, так что плотность энергии электромагнитного поля равна нулю, т.е. траектория находится в физическом вакууме. Но поскольку $\mathbf{A} \neq 0$, вакуум обладает некоторой “структурой”.

На том основании, что в области траектории поле безвихревое ($\mathbf{rot}\mathbf{A} = 0$), а безвихревое поле является полем потенциальным, мы можем написать $\mathbf{A} = \nabla\chi$, где χ можно найти, заметив, что согласно формуле (7.25)

$$A_\vartheta = \frac{1}{r}\frac{\partial\chi}{\partial\vartheta} = \frac{HR^2}{2\rho}. \tag{7.28}$$

Второе из равенств (7.28) приводит к решению

$$\chi = \frac{HR^2}{2}\vartheta, \tag{7.29}$$

где опущена несущественная для дальнейшего произвольная константа интегрирования. Калибровочная функция χ, как видим, не зависит от ρ и полностью определяется направлением радиус-вектора (полярным углом ϑ). На первый взгляд, из выражения (7.29) следует, что поскольку функция $\chi(\vartheta)$ не имеет особенностей (всюду регулярна), то $\mathbf{rot}(\nabla\chi) \equiv 0$ и магнитный поток, а вместе с этим и сам эффект Бома - Ааронова, должен отсутствовать. Однако, будучи регулярной, функция $\chi(\vartheta)$ не является однозначной, при замене $\vartheta \to \vartheta + 2\pi$ она возрастает на величину

$\pi H R^2 = \Phi$. В самом деле,

$$\int \mathbf{A} \cdot d\mathbf{r} = \int \nabla\chi \cdot d\mathbf{r} = \frac{HR^2}{2} \int\limits_{\vartheta}^{\vartheta+2\pi} d\vartheta = [\chi(\vartheta + 2\pi) - \chi(\vartheta)] = \Phi. \quad (7.30)$$

Регулярные неоднозначные функции могут существовать лишь в неодносвязных пространствах, т.е. в пространствах, в которых не все кривые могут быть непрерывно сжаты в точку. В нашем случае мы имеем дело с пространством вакуума вне соленоида, которое действительно не является односвязным – это есть двумерное пространство с дыркой (проколом) в месте, где соленоид пронизывает плоскость. Траектории, охватывающие такой прокол, не могут быть стянуты в точку. Очевидно также, что и произвольная траектория, обходящая соленоид n раз, не может быть деформирована в кривую, обходящую соленоид $m \neq n$ раз (нельзя размотать клубок со сращенными концами, пока не удален соленоид). Именно такого класса траектории (независимо от их пространственного размера и формы) и соответствуют эффекту Бома – Ааронова, который, таким образом, является по существу топологическим эффектом, обусловленным неодносвязностью калибровочной группы электродинамики.

В его классической постановке, эффект Бома – Ааронова может регистрироваться как (периодическое) изменение интенсивности потока электронов в какой либо точке экрана при изменении магнитного потока соленоида, помещенного между щелями. Период осцилляций Бома – Ааронова соответствует изменению разности фаз $\Delta\alpha$ на 2π (эта разность фаз при движении по двум траекториям C_1 и C_2, охватывающих соленоид с разных сторон, в точности равна изменению фазы волновой функции электрона по замкнутой траектории $\Delta\delta$). Согласно формуле (7.26)

$$\Delta\delta = \frac{e}{c\hbar}\Phi = 2\pi\frac{\Phi}{\Phi_0},$$

следует, что осцилляции Бома–Ааронова периодичны по магнитному потоку с периодом, равным кванту потока. Таким образом, потоки Φ и $\Phi + n\Phi_0$ дают идентичную интерференционную картину и, следователь-

но, физически неразличимы в полном согласии со сделанным выше выводом о многозначности калибровочной функции.

Поскольку электрическое поле также инвариантно относительно калибровочного преобразования скалярного потенциала (7.20b), имеется и электростатический вариант эффекта Бома – Ааронова. Воздействие электрического поля $\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla\varphi(\mathbf{r})$ в этом случае может быть устранено путем заключения электронных траекторий в фарадеевские клетки. Формальная сторона этой модификации эффекта фактически аналогична рассмотренному выше случаю векторного потенциала. Нетрудно видеть, что преобразование (7.20b) с одновременной заменой волновой функции

$$\Psi(\mathbf{r}) \to \psi(\mathbf{r}) = \Psi(\mathbf{r}) + \partial\chi(\mathbf{r}, t)/\partial t$$

так же не изменяет решений уравнения Шредингера. Единственное отличие в том, что фазовый сдвиг в этом случае определяется интегралом по времени от скалярного потенциала

$$\Delta\delta = \frac{e}{c\hbar}\int \varphi \cdot dt. \tag{7.31}$$

По отдельности взятые, выражения (7.25) и (7.31) справедливы только для статических полей, однако объединив их мы получим правильный результат для произвольного электромагнитного поля. По сути этот закон заменяет в квантовой механике классическую формулу Лоренца $\mathbf{F} = e(\mathbf{E} + [\mathbf{v} \times \mathbf{H}])/c$.

7.4 Задачи

1. Исходя из уравнения Лоренца покажите, что траектория электрона проводимости с квадратичным изотропным законом дисперсии в скрещенных полях описывается уравнениями

$$x = x_0 + v_{EH}t + \frac{v_{0y}}{\omega_c}(\cos\omega_c t - 1) + \frac{v_{0x} - v_{EH}}{\omega_c}\sin\omega_c t \tag{7.32}$$

$$y = y_0 + \frac{v_{EH} - v_{0x}}{\omega_c}(\cos\omega_c t - 1) + \frac{v_{0y}}{\omega_c}\sin\omega_c t, \tag{7.33}$$

если в начальный момент электрон находился в точке (x_0, y_0) и имел скорость (v_{0x}, v_{0y}).

2. Покажите, что для квадратичного закона дисперсии циклотронная частота и циклотронная масса не зависят от электрического поля.

3. Покажите, что квазиклассическое квантование в скрещенных полях приводит к результату, совпадающему с (7.15).

4. Найдите квазиклассический спектр в скрещенных полях для частиц с линейным законом дисперсии.

7.5 Вопросы для самопроверки

1. Что такое скачущие краевые траектории в магнитном поле? Изобразите такую траекторию для электрона и дырки на какой-нибудь границе образца при выбранном направлении магнитного поля. Может ли находиться центр такой дрейфующей циклотронной орбиты за пределами образца?

2. Изобразите схематически положение уровней Ландау двумерного электронного газа в образце конечного размера. Дайте качественное физическое объяснение. Какое основное качественное отличие спектра в объеме и у границы системы?

3. При каком положении центра циклотронной орбиты относительно границы средняя скорость приграничного дрейфа электрона максимальна? Чему она равна?

4. Изобразите схематически эффективный магнитный потенциал с центром орбиты в объеме и вблизи границы двумерного слоя. Объясните качественно, используя квазиклассический подход, как смещаются уровни Ландау при приближении центра осциллятора к границе? Какова наибольшая энергия уровня Ландау при известной высоте потенциального барьера на границе?

5. От чего зависит вклад пограничных состояний в общую плотность состояний? Укажите некоторые физические явления, где такие состояния играют определяющую роль?

6. Выпишите квазиклассическое уравнение эффективной массы для двумерного электрона в скрещенных магнитном и электрическом полях без учета и с учетом рассеяния на дефектах.

7. Какие соображения можно привести для обоснования того, что уравнение для движения электрона в скрещенных полях можно свести к уравнению движения только в магнитном поле?

8. Является ли энергия интегралом движения в случае трехмерного электронного газа в скрещенных магнитном и электрическом полях? Существует ли аналог этого интеграла движения?

9. Изобразите схематически траекторию электрона и дырки в идеальном двумерном газе в скрещенных полях. Что можно сказать о токе, связанном с перемещением заряда в такой ориентации полей? Чему равна проводимость?

10. Отличаются ли скорости и направления дрейфа в скрещенных полях

    (a) для электрона и легкой дырки в случае кейновского закона дисперсии?

    (b) для легкой и тяжелой дырки?

11. Чем отличается спектр электрона в скрещенных полях от его спектра в магнитном поле? Как можно из спектра определить среднюю квантовомеханическую скорость электрона? Чему она равна?

12. Отличается ли число состояний, приходящихся на один уровень Ландау, в скрещенных полях и в том же магнитном поле, но в отсутствии электрического? Отличается ли степень вырождения энергетических уровней в этих двух случаях?

13. Как связаны векторный потенциал и напряженность магнитного поля? Что такое калибровочная инвариантность электродинамики? Изменяет ли калибровочное преобразование волновые функции системы? Сказывается ли неоднозначность векторного потенциала

    (a) на уравнениях движения классической физики?

    (b) на волновых функциях?

    (c) на наблюдаемых квантовомеханических эффектах?

14. В чем заключается эффект Бома–Ааронова? Какое явление лежит в основе классического опыта по его наблюдению? Какая величина в конечном счете измеряется? Зависит ли ее значение от калибровки? Входит ли в ее выражение квант магнитного потока? Какая периодичность характерна для эффекта Бома–Ааронова?

Глава 8

Явления переноса

Кинетические процессы или явления переноса, осуществляемые носителями заряда в полупроводниках и металлах, как и в структурах на их основе, являются, с одной стороны, наиболее мощным средством изучения их физических свойств, а, с другой стороны, лежат в основе большинства практических применений таких структур, в том числе в микро- и наноэлектронике. В отличие от рассмотренных в Гл.3 статистически равновесных свойств, кинетические процессы зависят от взаимодействий внутри системы, что делает теорию явлений переноса значительно более сложной и менее достоверной, чем теория равновесных систем. Строгое квантово-механическое рассмотрение явлений переноса в твердом теле является чрезвычайно сложной задачей, встречающей на своем пути трудности, носящие зачастую принципиальный характер. Интуитивно понятно, что, поскольку ток (или поток энергии) есть скорость переноса заряда (энергии) через некоторое сечение образца, то мы, в той или иной степени, вынуждены рассматривать траекторию носителей, что, согласно соотношению неопределенностей, несовместимо, вообще говоря, со строгим квантово-механическим описанием. К счастью, почти неизменно оказывается, что полуклассическая теория представляет собой прекрасную и надежную аппроксимацию, обладая при этом огромным достоинством - позволяет давать прозрачные физические интерпретации явлениям. Что касается явлений чисто квантового происхождения, то они практически всегда могут быть рассмотрены как поправки к результатам полуклассического рассмотрения. Более того, сами такие яв-

ления (например эффекты, обусловленные интерференцией блоховских электронных волн – типичное квантовое явление) также во многих случаях допускают квазиклассическое описание.

8.1 Кондактанс и удельная проводимость в магнитном поле

Прежде всего, коснемся вопроса о соотношении между интегральной, полной, проводимостью (величина, измеряемая экспериментально, которую принято называть кондактансом, чего мы и будем придерживаться в дальнейшем) и удельной проводимостью (величина, рассчитываемая в рамках той или иной теоретической модели, которую мы в дальнейшем будем называть проводимостью) для систем различной пространственной размерности. В отличие от проводимости, кондактанс, определяемый как отношение тока к падению напряжения на образце, представляет собой свойство конкретного образца (конечных размеров) в целом, а не внутреннюю характеристику материала. Рассмотрим образец в виде d –мерного параллелепипеда со сторонами L_i $(i = x, y, z)$. В общем случае проводимость является тензором. Для нас наиболее важно, что даже в изотропном кристалле проводимость является тензором, если рассматриваемая система помещена в магнитное поле, поскольку перпендикулярная скорости электрона сила Лоренца приводит к дрейфу носителей заряда в направлении, перпендикулярном электрическому полю. При пропускании тока J_x через образец в направлении L_x (в такой геометрии образца при разомкнутых потенциальных контактах на противоположных боковых гранях $J_y = 0$) возникает разность потенциалов U_x в направлении L_x и U_y в направлении L_y (холловская разность потенциалов; магнитное поле предполагается направленным перпендикулярно плоскости L_x, L_y). Из определений кондактанса $G_{ij} = J_j/U_j$, удельной проводимости $\sigma_{ij} = j_i/E_j$ (в этом разделе мы предполагаем, что локальные значения σ_{ij} по всему образцу одинаковы), плотности тока $j_i = J_i/S$ (S - площадь сечения плоскостью, перпендикулярной L_x, L_y) и соотношения, связывающего разность потенциалов и электрическое поле $U_i = E_i L_i$

элементарно следует, что связь между кондактансом и проводимостью в трехмерной системе определяется соотношениями

$$G_{xx}^{(3D)} = \frac{J_x}{U_x} = \frac{j_x L_y L_z}{E_x L_x} = \sigma_{xx}^{(3D)} \frac{S}{L_x}, \tag{8.1}$$

$$G_{xy}^{(3D)} = \frac{J_x}{U_y} = \frac{j_x L_y L_z}{E_y L_y} = \sigma_{xy}^{(3D)} L_z. \tag{8.2}$$

Для двумерной системы имеем

$$G_{xx}^{(2D)} = \frac{J_x}{U_x} = \frac{j_x L_y}{E_x L_x} = \sigma_{xx}^{(2D)} \frac{L_y}{L_x}, \tag{8.3}$$

$$G_{xy}^{(2D)} = \frac{J_x}{U_y} = \frac{j_x L_y}{E_y L_y} = \sigma_{xy}^{(2D)}. \tag{8.4}$$

Для одномерного электронного газа $S = 1$ (площадь сечения в плоскости, перпендикулярной оси L_x определятся соотношением $S = dV/dL_x$, где V- объем системы; для одномерной нити $V = L_x$), так что плотность тока совпадает с током, и мы имеем

$$G_{xx}^{(1D)} = G_x = \frac{J_x}{U_x} = \frac{j_x \cdot 1}{E_x L_x} = \sigma_{xx}^{(1D)} \frac{1}{L_x} \tag{8.5}$$

В случае одного измерения вектор вырождается в псевдоскаляр и тензор не может быть определен. Проводимость в этом случае имеет единственную компоненту. Очевидно, что идеальный одномерный электронный газ не может испытывать воздействия магнитного поля, поскольку имеется единственная компонента квазиимпульса и скорости, т. е. $\mathbf{p}$ и $\mathbf{v}$ всегда коллинеарны и, следовательно, $d\mathbf{p}/dt = 0$ (импульс может сменить свое направление только на противоположное). В этом случае само понятие траектории, не совпадающей с L_x, в том числе циклотронной орбиты, не имеет смысла, поскольку говорить о траектории можно только в том случае, если ее размеры больше размера частицы. Размер квантовой частицы в поперечном квантовой нити направлении есть длина волны де Бройля в этом направлении, которая в квантовой одномерной системе порядка ее поперечного размера.

Уже на этом этапе мы видим, что двумерные системы занимают особое положение.

- Во-первых, размерности кондактанса и удельной проводимости для двумерного электронного газа совпадают.

- Во-вторых, диагональная (диссипативная) компонента кондактанса инвариантна относительно масштабного преобразования – диссипативные кондактансы образцов различного размера, но имеющих подобную геометрию, одинаковы (в связи с этим в качестве единицы удельного сопротивления в двумерных системах естественно рассматривать величину Ом/квадрат).

- И, наконец, недиагональная компонента кондактанса $G_{xy}^{(2D)}$ вообще не содержит размеров образца и эквивалентна холловской удельной проводимости σ_{xy}. Это обстоятельство имеет крайне важное значение с точки зрения экспериментального определения компонент тензора удельной проводимости, а, следовательно, и с точки зрения сопоставления эксперимента с теоретическими моделями, поскольку рассчитывается, как правило, именно удельная проводимость. Экспериментально σ_{xy} определяется из G_{xy} и размеров образца. Что касается компонент кондактанса G_{xy}, то они могут быть измерены с высокой точностью, определяемой точностью измерения тока и холловской разности потенциалов U_y. Основной вклад в погрешность определения σ_{xy} вносят факторы, связанные с размерами и геометрией образца (включая геометрию расположения и размеры потенциальных контактов, между которыми измеряются U_x и U_y), которые ограничивают точность измерения σ_{xy} величинами, не превышающими, как правило, нескольких процентов. Поскольку в двумерном случае соотношение между $\sigma_{xy}^{(2D)}$ и $G_{xy}^{(2D)}$ не содержит каких-либо размеров, измерения холловского кондактанса двумерного электронного газа дают непосредственно значение $\sigma_{xy}^{(2D)}$.

Соотношение $\sigma_{xy}^{(2D)} = G_{xy}^{(2D)}$, может быть, в частности, интерпретировано в том смысле, что холловский кондактанс двумерного электронного газа не изменится, если в образце произвольным образом “выколоть” отверстия произвольной формы. Интуитивно такое поведение может быть

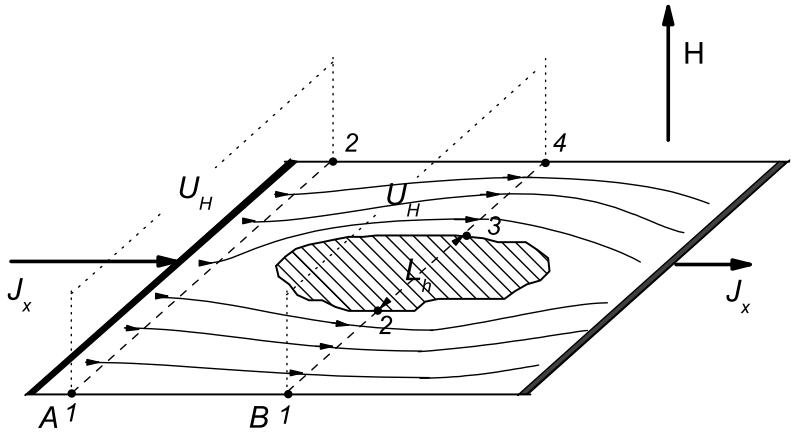

Рис. 8.1: Холловская разность потенциалов в двумерном образце с выколотой внутри него областью (заштриховано).

легко понято в рамках модели Друде-Лоренца. Рассмотрим двумерный образец шириной L_y с полостью внутри (Рис. 8.1). В перпендикулярном току J_x сечении, проходящем через полость (сечение B с участком шириной L_h внутри полости), "площадь" сечения, приходящаяся на токонесущие участки, в $(L_y - L_h)/L_y$ раз меньше, чем в области без полости (сечение A шириной L_y). Поскольку полный ток в любом сечении одинаков, то плотность тока $j_x = en_s v_x$ в этих участках, а, следовательно, и скорость электронов, а вместе с ней и сила Лоренца $F_y = ev_x H/c$ во столько же раз больше. Если направленный поток носителей в направлении L_y отсутствует ($j_y = 0$), электрическое поле E_y во всех сечениях в точности компенсирует силу Лоренца

$$eE_y = -ev_x H/c \tag{8.6}$$

(собственно говоря, это соотношение следует из известного уже нам выражения для скорости дрейфа электронов в скрещенных полях E_y, H), т.е. величина холловского поля в сечении с полостью в $L_y/(L_y - L_h)$ раз выше. Холловская же разность потенциалов $U_H = U_y = E_y(L_y - L_h)$ в любом сечении образца а, следовательно, и измеряемая величина кондактанса $G_{xy} = J_x/U_y$ остается той же самой. Ясно, что сказанное остается в силе при любом числе и расположении "выколок" произвольной формы и размера. Обратим внимание, что в приведенных рассуждениях неявно

предполагается, что потенциалы противоположных точек (точки 2 и 3 на (рис. 8.1) в сечении полости одинаковы (см. также разд. 9.2).

Для дальнейшего удобно рассматривать образец в виде d – мерного гиперкуба со стороной L (это не изменяет общности результатов). Для рассмотренной выше однородной d - мерной системы имеем

$$G_{xx}^{(d)} = \sigma_{xx}^{(d)} L^{d-2} \tag{8.7}$$

$$G_{xy}^{(d)} = \sigma_{xy}^{(d)} L^{d-2} \tag{8.8}$$

(в последнем соотношении предполагается $d > 1$). В реальной проводящей системе удельная проводимость неизбежно является флуктуирующей величиной. Такие флуктуации могут быть обусловлены как флуктуациями концентрации носителей, так и флуктуациями подвижности (транспортного времени релаксации импульса). Рассмотрим образец с неоднородным распределением концентрации носителей тока, которое возникает либо за счет неоднородности распределения легирующей примеси, либо за счет обусловленных флуктуациями толщины подзатворного диэлектрика флуктуаций потенциала в 2D и 1D системах с контролируемыми полевым электродом проводящими каналами. За счет флуктуаций электрического потенциала дно зоны проводимости будет флуктуировать относительно положения уровня Ферми (в условиях близких к термодинамическому равновесию уровень Ферми стабилизирован), т.е. величина химпотенциала μ является случайной функцией координат и всегда возможна настолько сильная флуктуация, что уровень Ферми опустится ниже дна зоны проводимости. Для простоты рассмотрим образец при нулевой температуре, когда электронный (дырочный) газ полностью вырожден. В этом случае такой флуктуации соответствует нулевая концентрация носителей и, следовательно, локальная проводимость будет равна нулю. Система является проводящей, если возможен перенос носителя заряда между двумя контактами к образцу (пусть для определенности контакты расположены на гранях, перпендикулярных L_x, соответствующие области контактов принято называть “берегами”). Для достаточно малых размеров образца есть всегда вероятность, что независимо от размерности, мы попадем на такую непроводящую область.

Зададимся вопросом о поведении кондактанса при увеличении размера куба. В случае одного измерения, даже если мы стартуем с ситуации, когда проводимость малого образца ненулевая, при увеличении размеров нити по закону больших чисел при достаточно большом объеме (длине нити L) обязательно встретится флуктуация потенциала, соответствующая появлению изолирующей локальной области. Для одномерного канала это означает невозможность переноса тока между берегами, расположенными на концах нити. Таким образом, одномерная система при достаточно большом ее размере всегда (при сколь угодно малом среднем масштабе флуктуаций потенциала, который с ростом L уменьшается) является изолятором. В трехмерном кристалле с металлическим типом проводимости, даже если мы стартуем с области (малого размера), являющейся в силу флуктуации изолятором, при увеличении размеров куба возникнут обходящие такие области пути, по которым будет возможно протекание носителей с одного берега системы на другой. Трехмерная система при достаточно большом ее размере является проводником при сколь угодно сильном масштабе флуктуаций. Отметим, что оба эти вывода не противоречат приведенным выше соотношениям для однородных систем: при $L \to \infty$ проводимость $G_x^{(1D)} \to 0$ для одномерной нити и $G_x^{(3D)} \to \infty$ для трехмерного образца. Однако в случае неоднородных систем такой переход имеет место, хотя и при сколь-угодно большой, но конечной величине L (характерная длина зависит от параметров флуктуирующего потенциала). Двумерная система занимает промежуточное положение между случаем одномерной и трехмерной системы. Однако, при нулевой температуре, как показывает строгое рассмотрение, двумерный электронный газ, как и одномерный, при достаточно большом размере образца также должен быть изолятором.

8.2 Тензоры двумерной удельной проводимости и сопротивления в магнитном поле

Как это следует из уравнения движения заряда в электрическом поле, электроны проводимости двигаются ускоренно (с одинаковым для всех электронов ускорением $e\mathbf{E}/m$), т.е. в отсутствии рассеяния их скорости а, следовательно, и связанный с переносом заряда ток

$$\mathbf{j}_d = en_d\mathbf{v} = \frac{e^2 n_d}{m} t\mathbf{E}$$

неограниченно возрастают со временем [1]. Проводимость в случае такого нестационарного процесса является функцией времени

$$\sigma_d = \frac{e^2 n_d}{m} t$$

и не может служить однозначным параметром, характеризующим электрические свойства материала. Закон Ома в идеальной проводящей системе не выполняется. Если наряду с электрическим полем, система помещена в магнитное поле (в двумерном случае, как мы видели, актуальна взаимно перпендикулярная ориентация $\mathbf{E}$ и $\mathbf{H}$), то, согласно результатам разд. 7.2, в отсутствии рассеяния носители заряда дрейфуют в направлении, перпендикулярном электрическому и магнитному полю, с постоянной во времени и одинаковой для всех носителей скоростью $\mathbf{v}_H = c[\mathbf{E} \times \mathbf{H}]/H^2$. Поскольку ток в направлении электрического поля вообще отсутствует, то проводимость в магнитном поле, связанная с переносом вдоль электрического поля, равна нулю. Дрейфовый ток $j_y = nev_{EH}$ связан с величиной электрического поля соотношением

$$j_y = necE_x/H = \sigma_{yx}E_x \tag{8.9}$$

[1] При больших временах это соотношение, конечно, не справедливо, поскольку $v < c$, и должно быть заменено его релятивистскими аналогом

$$\mathbf{j}_d = \frac{e^2 n_d c}{\sqrt{1 + (mc^2/ce\mathbf{E}t)^2}}.$$

где введена холловская проводимость $\sigma_{yx} = nec/H$. Таким образом, в случае скрещенных полей возникает холловский ток, пропорциональный напряженности электрического поля, т. е. закон Ома (по отношению к соотношению между j_y и E_x) выполняется и в отсутствии рассеяния.

Для того чтобы при действии внешних полей, ускоряющих носители заряда, потоки заряда (или энергии) были стационарными необходимо, чтобы возрастание импульса было ограничено, т. е. носители должны отдавать приобретенный от внешнего поля импульс. В реальном кристалле всегда имеются нарушения периодичности решеточного потенциала (как за счет статических локальных потенциалов, создаваемых примесями, вакансиями, дислокациями и т.д, так и за счет зависящих от времени акустических и оптических колебаний решетки), так что электрон проводимости может рассматриваться как свободный блоховский электрон только на участках траектории между локальными нарушениями периодичности, что и обеспечивает стационарный характер (при усреднении по достаточно большому промежутку времени) процессов переноса. В виду случайного характера, воздействие таких нарушений не может быть описано на основании динамических закономерностей и требует статистического подхода для их учета. Радиус действия обусловленных дефектами потенциалов порядка нескольких постоянной решетки и при типичных скоростях носителей $\bar{v}$ (средняя тепловая скорость для невырожденного электронного газа или скорость Ферми в случае вырождения) времена взаимодействия носителей с потенциалом дефектов чрезвычайно малы по сравнению с воздействием внешних полей, и взаимодействие носит характер столкновений, приводящих к случайному и резкому изменению скоростей и импульсов носителей.

В простейшей модели неидеальность кристалла может быть описана путем введения феноменологического параметра – среднего времени движения электрона между столкновениями τ (иначе - времени релаксации импульса, поскольку после выключения вешнего воздействия система электронов релаксирует к равновесному состоянию за время порядка τ - см. следующий раздел этой главы). Если размеры электронной системы намного превышают путь, проходимый электроном между столкновени-

ями (длину свободного пробега $\ell = \tau\bar{v}$), движение носит диффузионный характер. Если размеры образца в различных направлениях существенно различаются, это условие может выполняться не для всех направлений - в одних направлениях движение может быть диффузионным, в других баллистическим. При учете рассеяния воздействие электрического поля ограничено временами порядка τ и дополнительная (дрейфовая) скорость электронов, приобретаемая в поле между столкновениями, имеет конечную, в среднем равную $v_{dr} = eE\tau/m$, величину. Если время релаксации не зависит от импульса, v_{dr} для всех электронов одинакова и плотность тока в электрическом поле

$$j_d = en_d v_{dr} = ne^2\tau/mE = \sigma E,$$

то есть подчиняется закону Ома с проводимостью Друде-Лоренца

$$\sigma_d = n_d e^2\tau/m = n_d e\mu_n,$$

где подвижность μ_n является коэффициентом пропорциональности между дрейфовой скоростью и электрическим полем $v_{dr} = \mu_n E$.

В рамках феноменологического рассмотрения рассеяние носителей сказывается и на их движении в присутствии магнитного поля. Центры дрейфующих в скрещенных полях циклотронных орбит будут при столкновениях с дефектами испытывать скачки, так что в системе появится диффузия “центров осцилляторов”, в том числе и в направлении электрического поля, так что при рассеянии появится и перенос, направленный по (для дырок) или против (для электронов) электрического поля. Как следствие, изменятся обе компоненты тензора электропроводности. Тензор электропроводности определяется соотношениями

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

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

описывающими линейный отклик системы носителей заряда в кристалле на внешнее электрическое поле (в этой записи мы использовали теорему Онзагера, согласно которой кинетические коэффициенты удовлетворяют соотношениям симметрии $\sigma_{xx} = \sigma_{yy}$, $\sigma_{xy} = -\sigma_{yx}$).

В рамках подхода Друде-Лоренца расчет удельной электропроводности основан на уравнении движения, рассеяние в котором может быть учтено введением пропорциональной импульсу “силы трения “ $-\mathbf{p}/\tau$

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + \frac{e}{c}\left[\mathbf{vH}\right] - \frac{\mathbf{p}}{\tau} \tag{8.10}$$

Следует подчеркнуть, что при рассмотрении электропроводности, т.е. параметра, описывающего реакцию системы на однородное внешнее электрическое поле, неявно предполагается, что система бесконечна в координатном пространстве, поскольку наличие любой границы, ограничивающей движение носителей, означает присутствие потенциального барьера, и, как следствие, нарушение однородности электрического поля. Фактически в этом случае мы бы имели дело с рассмотренной в разд. 3.2 задачей экранирования электрического поля.

В обычной прямоугольной геометрии образца влияние его границ не может быть устранено. Интуитивно ясно, что такого влияния можно избежать путем соединения противоположных границ образца (подобно тому, как это было нами использовано при определении числа разрешенных состояний путем задания циклических граничных условий для волновой функции), т. е. в образцах сферической, цилиндрической или дисковой геометрии. Абстрактное рассмотрение таких геометрий опыта конечно возможно, но экспериментальная реализация для первых двух случаев весьма затруднительна не только ввиду сложности изготовления образцов, но и с точки зрения реализации скрещенной ориентации электрического и магнитного полей на цилиндрической или сферической поверхности. На практике может быть достаточно просто реализована кольцевая геометрия образца (диск Корбино), используемая для определения диагональной компоненты тензора проводимости в магнитном поле.

Для системы конечных размеров удобнее рассматривать тензор сопротивления, описывающий линейный отклик системы на пропускание через нее тока

$$E_x = \rho_{xx} j_x + \rho_{xy} j_y$$

$$E_y = -\rho_{yx} j_x + \rho_{xx} j_y$$

Обычно именно компоненты тензора удельного сопротивления и определяются экспериментально, поскольку при измерениях может быть легко реализовано условие $j_y = 0$, тогда как реализация эксперимента с заданным (одинаковым по всему образцу) направлением и величиной электрического поля, как мы выше видели, далеко не тривиальна. При $j_y = 0$ первое уравнение дает

$$E_x = \rho_{xx} j_x = \rho_{xx} n e v_{xE},$$

где мы выразили плотность тока в рамках кинетического метода Лоренца-Друде через концентрацию носителей и дрейфовую скорость электронов, приобретаемую в направлении электрического поля за время τ между столкновениями

$$v_x = (eE_x/m)\tau = \mu_n E_x.$$

В результате для диагональной компоненты тензора удельного сопротивления получаем известное выражение

$$\rho_{xx} = \frac{1}{ne\mu_n}$$

Из второго уравнения имеем

$$E_y = \rho_{yx} j_x = \rho_{yx} n e v_x$$

Поскольку $j_y = 0$, мы можем воспользоваться соотношением (8.6), что приводит к следующему выражению для недиагональной компоненты тензора сопротивления

$$\rho_{yx} = -\rho_{xy} = \frac{H}{nec}.$$

Холловская компонента удельного сопротивления в магнитном поле, в отличие от его диагональной компоненты, не зависит от рассеяния и равна обратному значению вычисленной нами ранее для идеального двумерного газа холловской проводимости (8.9). Нетрудно видеть, что это является следствием независимости скорости дрейфа носителей в скрещенных полях от их массы и величины заряда. Как видно, две компоненты тензора удельного сопротивления связаны соотношением

$$\frac{\rho_{xy}}{\rho_{xx}} = \frac{H\mu_n}{c} = \omega_c \tau \tag{8.11}$$

и тензор может быть записан в виде матрицы

$$\rho = \frac{1}{en\mu_n}\begin{pmatrix} 1 & \omega_c\tau \\ -\omega_c\tau & 1 \end{pmatrix} = \frac{1}{\sigma_0}\begin{pmatrix} 1 & \omega_c\tau \\ -\omega_c\tau & 1 \end{pmatrix}, \tag{8.12}$$

где $\sigma_0 = ne^2m/\tau$ – друдевская проводимость. Компоненты тензора проводимости могут быть получены обращением тензора ρ_{ij}

$$\sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2}, \tag{8.13}$$

$$\sigma_{xy} = \frac{-\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2}, \tag{8.14}$$

что приводит к следующему матричному представлению

$$\sigma = \frac{en\mu_n}{1+\omega_c^2\tau^2}\begin{pmatrix} 1 & -\omega_c\tau \\ \omega_c\tau & 1 \end{pmatrix} = \frac{\sigma_0}{1+\omega_c^2\tau^2}\begin{pmatrix} 1 & -\omega_c\tau \\ \omega_c\tau & 1 \end{pmatrix}. \tag{8.15}$$

Как можно видеть из этого выражения, характер зависимости компонент тензора проводимости от магнитного поля и времени релаксации, существенно различен для случаев, когда $\omega_c\tau << 1$ (область классически слабых магнитных полей) и $\omega_c\tau >> 1$ (классически сильные магнитные поля) (см. рис. 8.2). Фигурирующий здесь характеристический параметр $\omega_c\tau$ зависит как от времени релаксации, так и от параметров закона дисперсии (см. формулы (6.8) – (6.10)) и , следовательно, понятие сильного или слабого магнитного поля является относительным – величина критического поля, соответствующего условию $\omega_c\tau = 1$ может в разных материалах отличаться на несколько порядков. При одинаковой степени беспорядка условие сильного поля в материалах с малой циклотронной массой реализуется в меньших магнитных полях. Если выразить циклотронную частоту через классический циклотронный радиус и скорость $\omega_c = r_c v$ и учесть, что длина свободного пробега $\ell = v\tau$, то легко видеть, что параметр $\omega_c\tau$ в (8.11) может быть записан в эквивалентной форме

$$\omega_c\tau = \frac{\ell}{r_c} = \frac{H\mu_n}{c} = \frac{\rho_{xy}}{\rho_{xx}}, \tag{8.16}$$

т. е. условие сильного классического поля удовлетворено, если на длине свободного пробега укладывается несколько радиусов циклотронной орбиты.

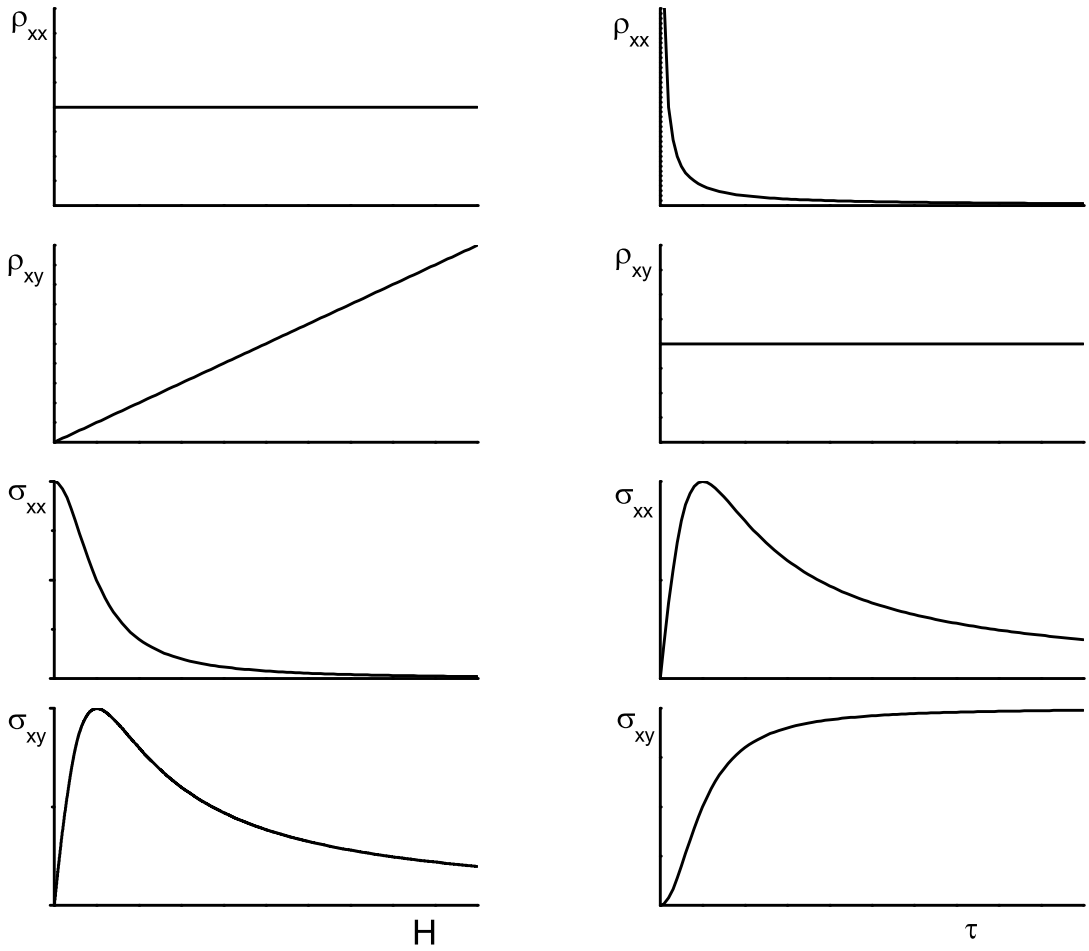

Рис. 8.2: Зависимость компонент тензоров сопротивления и проводимости от магнитного поля (при $\tau = const$) и от времени релаксации ($H = const$)

В классически слабых полях диагональная (диссипативная [2]) компонента проводимости σ_{xx} ведет себя стандартным образом, т. е. не зависит от магнитного поля (и близка к друдевской проводимости σ_0) и растет с уменьшением интенсивности рассеяния. В области же $\omega_c\tau >> 1$ диссипативная проводимость уменьшается как с магнитным полем, так и со временем релаксации, в соответствии с рассмотренным в разд. 7.2 характером движения заряда в скрещенных магнитных полях, когда в пределе $\tau \to \infty$ перенос в направлении электрического поля отсутствует, поскольку носители дрейфуют в направлении перпендикулярном $\mathbf{E}$ и $\mathbf{H}$. При этом, однако, в сильных полях $\omega_c\tau >> 1$ с ростом H уменьшается и недиагональная (холловская, бездиссипативная) компонента σ_{xy},

[2]Очевидно, что ненулевое значение компоненты σ_{xx} соответствует диссипации энергии в системе, поскольку джоулевы потери $\propto j_x E_x = \sigma_{xx} E_x^2$.

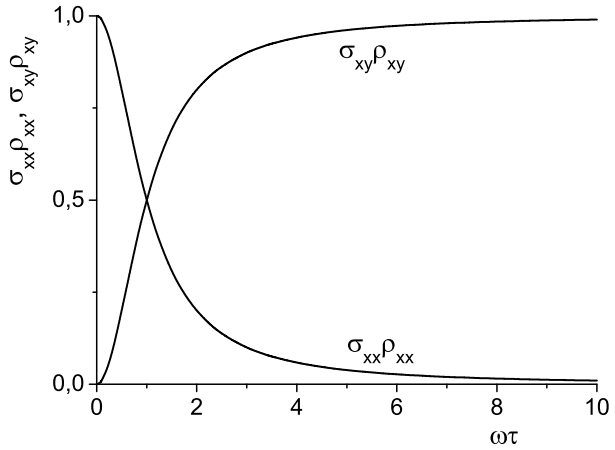

Рис. 8.3: Зависимость произведений $\rho_{xx}\sigma_{xx}$ и $\rho_{xy}\sigma_{xy}$ от $\omega_c\tau$

что связано с уменьшением скорости дрейфа ($v_{HE} = cE/H \to 0$ при $H \to \infty$).

Наиболее существенной для дальнейшего особенностью двумерного тензора электропроводности является тот факт, что в сильных магнитных полях при стремлении к нулю удельной проводимости σ_{xx} стремится к нулю и диссипативная компонента удельного сопротивления ρ_{xx}. Такое необычное поведение обусловлено тензорным характером σ и ρ в магнитном поле и следует непосредственно из общего соотношения (8.13), если $\rho_{xy} = H/nec \neq 0$ (то же справедливо и для компоненты σ_{xy}: $\sigma_{xy} \to 0$, если $\rho_{xy} \to 0$ при $\rho_{xx} = 1/ne\mu_n \neq 0$). Стандартное соотношение

$$\sigma = 1/\rho$$

выполняется, однако, для диссипативных компонент σ и ρ в области слабых полей и для холловских компонент в классически сильных полях, что проиллюстрировано на рис 8.3 ($\sigma_{ij} = 1/\rho_{ij}$ при $\sigma_{ij}\rho_{ij} = 1$).

Проведенное в этом разделе простое феноменологическое описание переноса не может считаться последовательным по-крайней мере в двух отношениях. В случае невырожденной статистики, в плотность тока системы вносят вклад носители, имеющие различные энергии и скорость. Времена свободного пробега этих носителей между столкновениями могут отличаться, тогда как фигурирующий в приведенных выше формулах феноменологический параметр τ полагается, по сути, одинаковым для всех носителей. Априори не очевидно, что процесс может быть адек-

ватно описан путем введения некоторого усредненного времени релаксации, а если это возможно, то остается неясной процедура усреднения τ по различным состояниям, и пределы ее применимости. На первый взгляд ситуация проще в случае вырожденной статистики, поскольку в этом случае в силу принципа Паули в переносе участвуют только носители вблизи уровня Ферми, времена релаксации для которых близки. Однако, с другой стороны, в выражения для σ_{ij} и ρ_{ij} входит полная концентрация носителей, что не согласуется с принципом Паули. Хотя, как будет показано ниже, более строгое рассмотрение, последовательно учитывающее вклад в ток носителей в различных энергетических состояниях, приводит к тем же выражениям для ρ_{ij} (8.12) и σ_{ij} (8.15), физическое смысл фигурирующих в этих формулах феноменологических параметров существенно различен.

8.3 Уравнение Больцмана

Последовательное полуклассическое описание явлений переноса основано на уравнении Больцмана, а, зачастую, очень хорошим приближением является его линеаризованная модификация – уравнение Блоха. Хотя уравнение Больцмана рассматривается подробно практически во всех учебниках по теории объемных (трехмерных) полупроводников или металлов, а его расширение на случай систем с пониженной размерностью не представляет особых затруднений, мы все же остановимся на основных используемых положениях и следующих из них соотношениях, предполагая, где это возможно, произвольную размерность пространства d.

Статистическое описание совокупности электронов проводимости основано на использовании функции распределения, описывающей распределение частиц по разрешенным квантовым состояниям системы. Выше мы уже пользовались подобной функцией при вычислении равновесной концентрации электронов. В случае термодинамического равновесия вероятность заполнения квантового состояния при заданной температуре зависит только от энергии состояния, отсчитываемой от выбранной подходящим образом нулевой энергии. Когда же на электрон действуют

внешние поля, изотропность системы нарушается, появляется выделенное направление и вероятность заполнения состояний теперь зависит не только от энергии, но и от отдельных компонент квазиимпульса (скорости, в случае изотропного закона дисперсии последняя коллинеарна с квазиимпульсом). Это непосредственно следует из уравнения движения: под действием внешних сил квазиимпульс $\mathbf{p}$, характеризующий состояние, в котором находится электрон в момент времени t, претерпевает в промежуток времени δt изменение $\delta\mathbf{p}$. Кроме того, поскольку равновесная вероятность зависит не только от энергии, но и от температуры и химпотенциала, можно ожидать, что и неравновесная функция распределения будет функцией координат, если имеются градиенты T или (и) μ.

Электронная подсистема кристалла должна быть описана, таким образом, функцией распределения $f = f(\mathbf{r}, \mathbf{p}, t)$, определяемой так, что число электронов в элементе объема $d\tau_{\mathbf{r}} d\tau_{\mathbf{p}}$ $2d$ - мерного фазового пространства в момент времени t равно

$$\frac{2}{(2\pi\hbar)^d} f(\mathbf{p}, \mathbf{r}, t) d\tau_{\mathbf{p}} d\tau_{\mathbf{r}}$$

В состоянии равновесия, т.е. при выключении внешних воздействий, функция $f(\mathbf{r}, \mathbf{p}, t)$ должна переходить в распределение Ферми-Дирака. В принципе электронные свойства полностью определены, коль скоро известна функция распределения. Действительно, вклад в ток электрона в состоянии $\mathbf{p}$ равен произведению его заряда на скорость $e\mathbf{v}$. Соответственно плотность тока в точке $\mathbf{r}$ получается просто суммированием вкладов отдельных электронов. Таким образом, вычисление плотности тока сводится к вычислению интеграла от скорости по всему d - мерному импульсному пространству с функцией распределения в качестве весовой функции

$$\mathbf{j}_d = e \int\limits_{\tau_{\mathbf{p}}} f(\mathbf{r}, \mathbf{p}, t) \mathbf{v}(\mathbf{p}) \frac{2}{(2\pi\hbar)^d} d\tau_{\mathbf{p}} \tag{8.17}$$

Совершенно аналогично описывается и плотность потока тепла

$$\mathbf{w}_d = \int\limits_{\tau_{\mathbf{p}}} f(\mathbf{r}, \mathbf{p}, t) \mathbf{v}(\mathbf{p}) (\varepsilon - \mu) \frac{2}{(2\pi\hbar)^d} d\tau_{\mathbf{p}} \tag{8.18}$$

Отметим, что в уравнение 8.18 входит не энергия ε, а разность $\varepsilon - \mu$, поскольку "количество теплоты" равно внутренней энергии за вычетом свободной энергии, т. е. в расчете на один электрон за вычетом химпотенциала. Таким образом, описание явлений переноса и вычисление соответствующих кинетических коэффициентов сводится к отысканию правильной функции распределения в присутствии внешних полей при заданных граничных условиях.

Базовым уравнением для $f(\mathbf{r}, \mathbf{p}, t)$ и является уравнение Больцмана, основанное, по-существу, на теореме Лиувилля. Рассмотрим элемент объема $d\tau_{\mathbf{r}} d\tau_{\mathbf{p}}$ фазового пространства с центром в точке $(\mathbf{r}, \mathbf{p})$. Электроны могут покинуть или попасть в этот элемент объема в результате:

1. движения в координатном пространстве $\mathbf{r}$, поскольку они обладают скоростью $\mathbf{v} = \nabla_{\mathbf{p}}\varepsilon$;

2. движения в импульсном пространстве, поскольку согласно уравнению движения $d\mathbf{p}/dt = \mathbf{F}$ они приобретают под действием внешнего силового поля ускорение;

3. рассеяния под действием внутренних силовых полей, обусловленных присутствием дефектов.

В следствие этих процессов электроны окажутся в момент времени $t + dt$ в таком же по величине объеме $d\tau_{\mathbf{r}} d\tau_{\mathbf{p}}$ (согласно теореме Лиувилля), но с центом в точке $(\mathbf{r} + (\partial\mathbf{r}/\partial t)dt, \mathbf{p} + (\partial\mathbf{p}/\partial t)dt)$. Таким образом, если в момент времени t электроны занимали состояния, определяемые функцией распределения $f(\mathbf{r}(t), \mathbf{p}(t), t)$, то в момент времени $t + dt$ распределение электронов по состояниям описывается другой функцией распределения $f(\mathbf{r} + \mathbf{v}dt, \mathbf{p} + (\partial\mathbf{p}/\partial t)dt, t + dt)$. В соответствии с теоремой Лиувилля о неизменности фазового объема при движении системы вдоль фазовых траекторий и учитывая универсальность элемента объема, приходящегося на одно квантовое состояние $\Delta p \Delta r = 2\pi\hbar$ или, другими словами, учитывая сохранение числа состояний, полная производная от функции распределения

$$\frac{df}{dt} = 0.$$

Дифференцируя $f(\mathbf{r}, \mathbf{p}, t)$ по времени как сложную функцию, т. е. учитывая, что $\mathbf{r}$ и $\mathbf{p}$ являются функциями времени приходим к дифференциальному уравнению

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} f \cdot \frac{d\mathbf{r}}{dt} + \nabla_{\mathbf{p}} f \cdot \frac{d\mathbf{p}}{dt}. \tag{8.19}$$

Поскольку согласно уравнению движения $d\mathbf{p}/dt = \mathbf{F}_t$, где $\mathbf{F}_t$ - полная сила, действующая на электроны, а $d\mathbf{r}/dt = \mathbf{v}$ уравнение (8.19) принимает вид

$$-\frac{\partial f}{\partial t} = \nabla_{\mathbf{r}} f \cdot \mathbf{v} + \nabla_{\mathbf{p}} f \cdot (\mathbf{F} + \mathbf{F}_{sc}),$$

где в $\mathbf{F_t} = \mathbf{F} + \mathbf{F_{sc}}$ мы выделили два класса сил. Внешние силы $\mathbf{F}$, а это в общем случае есть сила Лоренца

$$\mathbf{F} = \mathbf{eE} + \frac{e}{c}\left[\mathbf{v} \times \mathbf{H}\right],$$

и внутренние силы $\mathbf{F_{sc}}$, обусловленные локальными нарушениями периодичности решетки. Действия внутренних и внешних сил противоположны. Если первые приводят к возникновению направленного движения частиц в координатном и импульсном пространстве, как это было описано в предыдущих разделах, то вторые приводят к разрушению такой направленности.

Анализ кинетических явлений, в общем случае представляющий очень сложную задачу, значительно упрощается если для описания столкновений можно ввести время релаксации. Если внешние поля, действующие на находящуюся в стационарном неравновесном состоянии систему, выключить, то функция распределения будет меняться только за счет столкновений, которые приводят систему частиц в равновесное состояние, т.е. восстанавливают нарушенное внешними полями равновесное состояние, описываемое равновесной функцией распределения Ферми-Дирака. Неравновесная функция распределения всегда может быть представлена в виде

$$f = f_0 + f_1,$$

если на описывающую отклонение системы от равновесия функцию f_1 не накладывается каких-либо ограничений. Приближение времени релакса-

ции соответствует простейшему предположению, что скорость релаксации системы к равновесию пропорциональна отклонению от равновесия

$$\frac{\partial f}{\partial t} = \left.\frac{\partial f}{\partial t}\right|_{sc} = -F_{sc} \cdot \nabla_{\mathbf{p}} f = -\frac{f - f_0}{\tau(p, r)} = -\frac{f_1}{\tau}$$

с коэффициентом пропорциональности $1/\tau$, что приводит к временной зависимости функции распределения вида

$$f(r, p, t) - f_0(r, p) = (f(r, p, 0) - f_0(r, p))\, e^{-\frac{t}{\tau}}$$

или

$$f_1(t) = f_1(0) e^{-\frac{t}{\tau}}.$$

Время релаксации, величина и характер зависимости от импульса которого определяется конкретными механизмами рассеяния, показывает насколько быстро в системе восстанавливается нарушенное внешними полями термодинамическое равновесие. В предположении, что τ не меняется в присутствии внешних полей, т. е. является однозначной характеристикой процессов рассеяния как в процессе релаксации, так и в стационарном неравновесном состоянии, уравнение Больцмана запишется в виде

$$\nabla_{\mathbf{r}}(f_0 + f_1) \cdot \mathbf{v} + \nabla_{\mathbf{p}}(f_0 + f_1) \cdot \left(e\mathbf{E} + \frac{e}{c}\,[\mathbf{v} \times \mathbf{H}]\right) = -\frac{f_1}{\tau}, \qquad (8.20)$$

где пространственная производная $\nabla_{\mathbf{r}} f_0$ от равновесной функции распределения

$$f_0 = \frac{1}{\exp\frac{(\varepsilon - E_F)}{kT} + 1} = \frac{1}{\exp\frac{(\varepsilon(\mathbf{p}) - \varphi(\mathbf{r}) - \mu(\mathbf{r}))}{kT(\mathbf{r})} + 1}$$

(в последнем выражении энергии отсчитываются от дна зоны проводимости) определяется координатными зависимостями электрического и химического потенциалов и температуры, а производная по квазиимпульсу $\nabla_{\mathbf{p}} f_0$ законом дисперсии $\varepsilon(p)$. Подстановка этих производных в (8.20) приводит к уравнению

$$[-\mathbf{v} \cdot \nabla_{\mathbf{r}}(\varphi + \mu) + (\varepsilon - \varphi - \mu)\nabla_{\mathbf{r}} \ln T]\,\frac{\partial f_0}{\partial \varepsilon} + \left(e\mathbf{E} \cdot \mathbf{v} + \frac{e}{c}\,[\mathbf{v} \times \mathbf{H}] \cdot \mathbf{v}\right)\frac{\partial f_0}{\partial \varepsilon} +$$
$$+ \mathbf{v} \cdot \nabla_{\mathbf{r}} f_1 + \left(e\mathbf{E} + \frac{e}{c}\,[\mathbf{v} \times \mathbf{H}]\right) \cdot \nabla_{\mathbf{p}} f_1 + \frac{f_1}{\tau} = 0, \qquad (8.21)$$

в котором мы выписали равный нулю член с магнитным полем, стоящий множителем при $\partial f_0/\partial\varepsilon$, чтобы показать, что в пренебрежении слагаемым, содержащим $\nabla_{\mathbf{p}} f_1$ уравнение Больцмана не описывает эффектов в магнитном поле. При небольших отклонениях от равновесия можно полагать, что производные от f_1 являются величинами второго порядка малости по сравнению с производными от равновесной функции распределения. В результате в приходим к линеаризованному уравнению Блоха для неравновесной добавки f_1 к функции распределения

$$\frac{\partial f_0}{\partial\varepsilon}\left[-\mathbf{v}\cdot\nabla_{\mathbf{r}}(\varphi+\mu)+(\varepsilon-\varphi-\mu)\nabla_{\mathbf{r}}\ln T\right]+\frac{\partial f_0}{\partial\varepsilon}\cdot e\mathbf{E}\cdot\mathbf{v}+\frac{e}{c}\left[\mathbf{v}\times\mathbf{H}\right]\cdot\nabla_{\mathbf{p}} f_1+\frac{f_1}{\tau}=0. \tag{8.22}$$

В присутствии внешнего поля изотропность пространства нарушается и функция распределения, которая в равновесии зависела только от энергии (модуля квазиимпульса), зависит теперь от отдельных компонент квазиимпульса, т. е. становится анизотропной в пространстве квазиимпульсов.

Уравнение (8.21) (или (8.22)) совместно с выражениями для тока (8.17) и потока тепла (8.18) полностью описывают все эффекты, связанные с носителями заряда в кристалле при воздействии на него электрического и (или) магнитного поля (электропроводность, гальваномагнитные явления), градиента температур (теплопроводность) или их комбинации (термоэлектрические и термомагнитные явления). Благодаря большому числу таких комбинаций, разнообразию взаимных ориентаций полей и граничных условий, накладываемых на потоки зарядов и (или) тепла, различию доминирующих механизмов рассеяния, степени фермиевского вырождения и законов дисперсии общее число различных случаев, которые могут быть рассмотрены, превышает несколько сотен. Очевидно, что рассмотрение всех этих теоретически возможных случаев не только невозможно, но и нецелесообразно. Наиболее существенные особенности физических механизмов, лежащих в основе полуклассического описания кинетических процессов переноса, могут быть достаточно полно выяснены уже при рассмотрении одного из этого класса явлений, поскольку все они основаны по сути на одном и том же уравнении. В качестве типичного, но важного примера может быть рассмотрена стационарная

электропроводность в электрическом поле.

8.4 Электропроводность d – мерного электронного газа

Мы ограничимся простейшим случаем электропроводности однородной системы в однородном электрическом поле. Оставляя в уравнении Блоха (8.22) только слагаемое, содержащее электрическое поле, сразу приходим к решению для функции $f_1 = -e\tau(\partial f_0/\partial\varepsilon)\mathbf{E}\cdot\mathbf{v}$, откуда

$$f(\mathbf{p}) = f_0 - e\tau\frac{\partial f_0}{\partial\varepsilon}\mathbf{E}\cdot\mathbf{v} \tag{8.23}$$

Легко видеть, что такое же выражение для $f(\mathbf{p})$ может быть получено, если разложить в ряд Тейлора модифицированную равновесную функцию распределения

$$f_0(\varepsilon(\mathbf{p}-e\tau\mathbf{E})) = f_0(\varepsilon(\mathbf{p})) - e\tau\frac{\partial f_0}{\partial\varepsilon}\nabla_{\mathbf{p}}\varepsilon\cdot\mathbf{E} = f_0 - e\tau\frac{\partial f_0}{\partial\varepsilon}\mathbf{E}\cdot\mathbf{v}, \tag{8.24}$$

т. е. неравновесная стационарная функция распределения имеет тот же вид, что и равновесная, но сдвинутая в пространстве импульсов в направлении электрического поля на величину $-e\mathbf{E}\tau$. Учитывая, что при рассматриваемом малом отклонении от равновесия $\varepsilon(\mathbf{p}-e\tau\mathbf{E}) \approx \varepsilon(\mathbf{p}) - e\tau\nabla_{\mathbf{p}}\varepsilon\cdot\mathbf{E}$ и следовательно

$$f(\mathbf{p}) = f_0(\varepsilon(\mathbf{p}) - e\tau\mathbf{E}\cdot\mathbf{v}), \tag{8.25}$$

приходим к еще одной интерпретации результата (8.23) - распределение по состояниям в электрическом поле эквивалентно равновесному, однако с энергией, измененной на величину, соответствующую энергии, приобретенной электроном при его движении со скоростью $\mathbf{v}$ в поле за время между столкновениями τ (в виду слабости отклонения от равновесия скорость между столкновениями изменяется слабо, т. е. в первом приближении может считаться почти постоянной). Однако необходимо отдавать отчет, что несмотря на формальное сходство представлений для неравновесной функции распределения (8.24), (8.25) и, с другой стороны, (8.23)

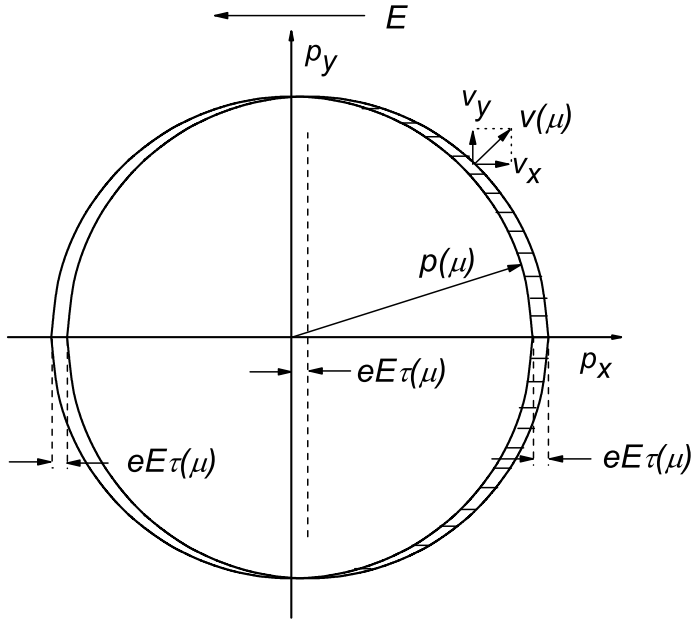

Рис. 8.4: Поверхность Ферми в электрическом поле

(и сходство конечных выражений для тока, в том числе и для вырожденного газа - см. ниже) их физическое содержание совершенно различно. Согласно формулам (8.24) и (8.25) внешнее поле изменяет импульс и энергию электронов *во всех состояниях*, что не согласуется с принципом Паули. Электроны в глубине "сферы" Ферми не могут ускоряться электрическим полем и по той же причине они не рассеиваются примесью. С этим полностью согласуется выражение для неравновесной функции распределения (8.23), согласно которому внешнее поле действует только на электроны, находящиеся в состояниях, для которых отлична от нуля производная $\partial f_0/\partial\varepsilon$. В случае сильного вырождения это соответствует состояниям в узкой области энергий $\sim k_BT << \mu$ вблизи поверхности Ферми. Появляется небольшая добавка к f_0 со той стороны поверхности Ферми, где $e\mathbf{E}\cdot\mathbf{v} > 0$, и равная ей по величине но отрицательная по знаку с противоположной стороны. Для невырожденного газа, в котором учет принципа Паули слабо влияет на заполнение состояний, все три выражения для функции распределения по сути эквивалентны. Представление (8.23) в точности соответствует лежащей в основе кинетического метода описания процессов переноса классической картине, когда *все носители*, ускоряясь внешним полем в промежутках между столкновени-

ями τ, приобретают дополнительную, дрейфовую скорость $\delta\mathbf{v} = e\mathbf{E}\tau/m$ и, следовательно, переносят ток плотностью $\mathbf{j}_d = n_d e\mathbf{v} = (n_d e^2\tau/m)\mathbf{E}$ (в случае, если τ, а следовательно и $\delta\mathbf{v}$ зависят от энергии, как это обычно и имеет место, в выражение для тока входит усредненное по всем токонесущим состояниям время релаксации $\bar{\tau}$ - см. ниже).

Подставляя $f = f_0 + f_1$ в выражение для плотности тока

$$\mathbf{j}_d = e\int\limits_{\tau_\mathbf{p}} f_0\mathbf{v}\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d} - e^2\int\limits_{\tau_\mathbf{p}} \tau\frac{\partial f_0}{\partial\varepsilon}\left(\mathbf{v}\cdot\mathbf{E}\right)\cdot\mathbf{v}\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d}$$

и учитывая, что первый интеграл, ввиду того, что подынтегральное выражение является нечетной функцией импульса, равен нулю (в равновесии число электронов имеющих скорость $\mathbf{v}$ в точности равно числу электронов со скоростью $-\mathbf{v}$) получаем

$$\mathbf{j}_d = -e^2\int\limits_{\tau_\mathbf{p}} \tau\frac{\partial f_0}{\partial\varepsilon}\left(\mathbf{v}\cdot\mathbf{E}\right)\cdot\mathbf{v}\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d}. \tag{8.26}$$

В тоже время необходимо отметить, что, хотя распределение электронов по энергиям во внешнем поле меняется (см. (8.25)), полная энергия электронов

$$\mathcal{E} = \int f_0\varepsilon(\mathbf{p})\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d} + \int f_1\varepsilon(p)\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d} = \int f_0\varepsilon(\mathbf{p})\frac{2d\tau_\mathbf{p}}{(2\pi\hbar)^d} = \mathcal{E}_0 \tag{8.27}$$

(на этот раз равен нулю интеграл, содержащий функцию f_1 в силу нечетности последней) в неравновесных условиях остается той же, что и в условиях термодинамического равновесия. Носители заряда в рассматриваемом приближении (τ описывает только релаксацию импульса носителей, а не энергии) получают от внешних полей и передают решетке (при столкновениях) только квазиимпульс, но не энергию, в полном согласии с тем, что джоулево тепло должно выражаться через квадрат электрического поля.

Как и при вычислении плотности состояний в разд. 2.3.2, удобно перейти в (8.26) от интегрирования по состояниям пространстве импульсов к интегралу по энергии и изоэнергетической поверхности (см. (2.24)). В

дальнейшем мы ограничимся наиболее простым в математическом отношении случаем изотропного закона дисперсии, поскольку обобщение на случай анизотропного спектра не приводит к каким либо существенно новым в физическом отношении особенностям, а в актуальном случае эллипсоидальных изоэнергетических поверхностей задача может быть сведена к изотропной соответствующим выбором масштаба вдоль осей эллипсоида, подобно тому, как это было сделано в разд. 2.3.2. В случае сферической поверхности постоянной энергии интеграл по $d\tau_{\mathbf{p}}$ преобразуется согласно соотношению (2.25), что приводит к выражению

$$\mathbf{j}_d = -e^2 \int\limits_0^\infty \oint\limits_\Omega \tau \frac{\partial f_0}{\partial \varepsilon} (\mathbf{v} \cdot \mathbf{E}) \cdot \mathbf{v} D_d(\varepsilon) d\varepsilon \frac{d\Omega_d}{\Omega_d}. \tag{8.28}$$

Вместо интегрирования по $d\Omega_d$, хотя и элементарного, удобно воспользоваться соображениями симметрии. В виду изотропности спектра (следовательно и времени релаксации) вектор плотности тока коллинеарен электрическому полю. Единственной векторной величиной в интеграле является произведение $\mathbf{v} \cdot \mathbf{v} \cdot \mathbf{E}$, единственная ненулевая компонента которого равна (полагаем, что $\mathbf{j_d}$ и $\mathbf{E}$ направлены по оси x)

$$\mathbf{v} \cdot \mathbf{v} \cdot \mathbf{E} = v_x^2 E.$$

Это дает $1/d$ долю вклада от квадрата скорости, $v^2 E = E \sum_{i=1}^d v_i^2$. Поэтому

$$j_d = -\frac{e^2 E}{d} \int\limits_0^\infty \tau \frac{\partial f_0}{\partial \varepsilon} v_x^2 D_d(\varepsilon) d\varepsilon \oint\limits_\Omega \frac{d\Omega_d}{\Omega_d} = -\frac{e^2 E}{d} \int\limits_0^\infty \tau \frac{\partial f_0}{\partial \varepsilon} v_x^2 D_d(\varepsilon) d\varepsilon = \sigma_d E, \tag{8.29}$$

откуда для проводимости получаем

$$\sigma_d = -\frac{e^2}{d} \int\limits_0^\infty \tau(\varepsilon) \frac{\partial f_0}{\partial \varepsilon} v_x^2(\varepsilon) D_d(\varepsilon) d\varepsilon \tag{8.30}$$

Поскольку $\partial f_0 / \partial \varepsilon < 0$, а заряд входит в квадрате, проводимость есть положительно определенная величина для любого типа носителей (при

одном и том же направлении поля потоки частиц $\delta\mathbf{v} \propto e\mathbf{E}$ с разным знаком заряда $\pm\, e$ направлены противоположно, но именно по этой причине связанные с ними токи $\mathbf{j} \propto e\delta\mathbf{v}$ имеют одно направление).

Для основных механизмов рассеяния зависимость времени релаксации от энергии носит степенной характер $\tau = c_\tau \varepsilon^{r_\tau}$ и для степенного закона дисперсии $\varepsilon = c_r p^r$ проводимость может быть выражена через интегралы Ферми. Подставляя в (8.30) выражения для τ, скорости

$$v(\varepsilon) = (\partial\varepsilon/\partial p) = r c_r^{1/r} \varepsilon^{(r-1)/r} \tag{8.31}$$

и плотности состояний (2.44) приходим к формуле

$$\sigma_d = -\frac{r}{d 2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d} c_r^{\frac{2-d}{r}} c_\tau \frac{e^2}{\hbar^d} \int\limits_0^\infty \varepsilon^{\frac{d+r-2}{r}+r_\tau} \frac{\partial f_0}{\partial\varepsilon} d\varepsilon. \tag{8.32}$$

Фигурирующий здесь интеграл интегрированием по частям сводится к интегралу Ферми (2.65)

$$\int\limits_0^\infty \varepsilon^j \frac{\partial f_0}{\partial\varepsilon} d\varepsilon = -j \int\limits_0^\infty \varepsilon^{j-1} f_0(\varepsilon - \mu/k_B T) d\varepsilon = -j\Gamma(j)(k_B T)^j \mathcal{F}_{j-1}(\mu/k_B T),$$

в результате чего выражение для проводимости принимает вид

$$\sigma_d = \frac{r}{d 2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d} c_r^{\frac{2-d}{r}} c_\tau \frac{e^2}{\hbar^d} \Gamma(j_d+1)(k_B T)^{j_d} \mathcal{F}_{j_d-1}(\mu/k_B T), \tag{8.33}$$

где

$$j_d = \frac{d+r-2}{r} + r_\tau \tag{8.34}$$

и использовано рекуррентное соотношение (2.38) $j\Gamma(j) = \Gamma(j+1)$. Если ввести концентрацию носителей (2.67), эта формула может быть максимально приближена к классическому выражению для проводимости Друде-Лоренца

$$\sigma_d = n_d e^2 \frac{\langle\tau\rangle}{m}, \tag{8.35}$$

где

$$\frac{\langle\tau\rangle}{m} = \frac{r^2 c_r^{2/r} c_\tau}{d\Gamma(d/r)} \Gamma\left(\frac{d+2r-2}{r} + r_\tau\right) (k_B T)^{\frac{r-2}{r}+r_\tau} \frac{\mathcal{F}_{\frac{d-2}{r}+r_\tau}(\mu/k_B T)}{\mathcal{F}_{\frac{d-r}{r}}(\mu/k_B T)}. \tag{8.36}$$

Выражение (8.36) конкретизирует точный смысл процедуры усреднения времени релаксации по состояниям носителей и ее зависимость от размерности пространства и параметров закона дисперсии.

Для невырожденного электронного газа интеграл Ферми от индекса не зависит и последний множитель в (8.36), содержащий отношение интегралов Ферми, равен единице. То же самое имеет место и для произвольной степени вырождения, если (независимо от размерности пространства) показатели r и r_τ связаны соотношением

$$r_\tau = (2 - r)/r, \tag{8.37}$$

при выполнении которого индексы фигурирующих в (8.36) интегралов Ферми совпадают. Легко видеть, что при выполнении этого условия проводимость при постоянной концентрации носителей от температуры не зависит, если не зависит от температуры τ. Рассмотрим простейший, но поучительный случай независящего от энергии времени релаксации $\tau = c_\tau = \tau_0$ $(r_\tau = 0)$

$$\sigma_d = n_d e^2 \frac{r^2 c_r^{2/r} \tau_0}{d\Gamma(d/r)} \Gamma\left(\frac{d + 2r - 2}{r}\right) (k_B T)^{\frac{r-2}{r}} \frac{\mathcal{F}_{\frac{d-2}{r}}(\mu/k_B T)}{\mathcal{F}_{\frac{d-r}{r}}(\mu/k_B T)}. \tag{8.38}$$

В случае стандартной зоны $r = 2$ $(c_r = 1/2m)$ выполняется условие (8.37) и мы приходим в точности к формуле Друде–Лоренца

$$\sigma_d = n_d \frac{e^2 \tau_0}{m}, \tag{8.39}$$

которая, таким образом, при $\tau = const$ справедлива при любой степени вырождения и для любой размерности пространства.

Рассмотрим теперь тот же случай $\tau = const = \tau_0$ для линейного дисперсионного соотношения $r = 1$. Индексы интегралов Ферми в (8.38) теперь различны и мы ограничимся невырожденным газом (когда $\mathcal{F}_j(\mu) = \exp(\mu/k_B T)$ не зависит от индекса), для которого (8.38) приводит к выражению

$$\sigma_d = n_d \frac{e^2 \tau_0}{k_B T/s^2}. \tag{8.40}$$

Отметим, что согласно (2.55) величина $k_B T/s^2$ играет роль динамической эффективной массы частицы с энергией $\varepsilon = k_B T$. В отличие от

(8.39), проводимость при линейном дисперсионном соотношении даже при независящих от температуры n_d и τ_0 зависит от температуры, а именно обратно пропорциональна T. Физическая причина столь разного поведения проводимости при квадратичном и линейном зонном спектре достаточно проста. Для невырожденного электронного газа $\partial f_0/\partial\varepsilon \approx -f_0/k_BT$ и согласно (8.30) проводимость пропорциональна

$$\sigma_d \propto \tau_0 \int\limits_0^\infty \frac{v^2}{k_BT} D_d(\varepsilon) f_0 d\varepsilon. \tag{8.41}$$

При линейном законе дисперсии скорость (8.31) от энергии вообще не зависит ($v \equiv s$), а значит не меняется и с изменением энергии теплового движения k_BT. Следовательно проводимость

$$\sigma_d \propto \frac{\tau_0 s^2}{k_BT} \int\limits_0^\infty D_d(\varepsilon) f_0 d\varepsilon = \frac{\tau_0 s^2}{k_BT} n_d \tag{8.42}$$

уменьшается с температурой за счет чисто "статистического" уменьшения отклонения функции распределения от ее равновесного значения (для невырожденного газа $f_1 \propto f_0/k_BT$). Для квадратичного же закона дисперсии скорость $v = \sqrt{2\varepsilon/m}$ носителей невырожденного газа порядка их средней тепловой скорости $v \sim \sqrt{k_BT/m}$. В результате, возрастание v^2 с температурой компенсирует уменьшение статистического множителя $1/k_BT$

$$\sigma_d \propto \frac{\tau_0}{k_BT} \int\limits_0^\infty \frac{k_BT}{m} D_d(\varepsilon) f_0 d\varepsilon = \frac{\tau_0}{m} n_d. \tag{8.43}$$

Для наиболее интересного с точки зрения выяснения особенностей низкоразмерных квантовых систем случая низких температур, соответствующих вырождению электронного газа, проводимость в первом приближении (δ – функционном) легко находится непосредственно из выражения (8.30)

$$\sigma_d = \frac{e^2}{d} \tau(\mu) v^2(\mu) D_d(\mu), \tag{8.44}$$

т. е., как и должно быть, определяется только параметрами носителей в ближайшей окрестности поверхности Ферми. Если ранее, при вычислении концентрации, теплоемкости и длины экранирования в качестве

такого параметра фигурировала только плотность состояний на уровне Ферми, то проводимость зависит также от фермиевской скорости носителей $v(\mu) = \partial\varepsilon/\partial p|_{\varepsilon=\mu}$ и фермиевского времени релаксации. Подчеркнем, что для вырожденного газа нет необходимости знать закон дисперсии носителей во всем энергетическом диапазоне от $\varepsilon = 0$ до $\varepsilon \sim \mu$, нужен лишь его вид в окрестности уровня Ферми. В остальном интервале энергий соотношение $\varepsilon(p)$ может быть любым, от него ничего не зависит [3]. Отметим, что это же относится и к концентрации носителей, величина которой при сильном вырождении полностью определяется величиной фермиевского квазиимпульса (см. (2.60))

$$n_d = \frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\left(\frac{p_F}{\hbar}\right)^d,$$

а для определения соответствующей энергии Ферми μ нужно только "локальное" дисперсионное соотношение $\mu = \varepsilon(p_F)$. Физическая причина этого проста. Исходным пунктом при вычислении плотности состояний, как мы видели (см. разд. 2.1), является квантование квазиимпульса, а отнюдь не энергии. Переход к квантованию в энергетическом пространстве (к плотности квантовых состояний $D_d(\varepsilon)$), для чего и необходимо знать зависимость $\varepsilon(p)$), связан только с тем, что заполнение состояний носителями определяется функцией распределения $f_0(\varepsilon)$, зависящей от энергии, а не от компонент квазиимпульса. Но в вырожденном газе $f_0(\varepsilon)$ вырождается в единичную ступенчатую функцию Хевисайда (2.63) $\Theta(\mu-\varepsilon)$, т. е по существу $f_0(\varepsilon)$ от энергии не зависит. Совершенно иначе обстоит дело при невырожденной статистике, когда вклад в ток дают все зонные носители - для вычисления проводимости (8.30) (для вычисления $\langle\tau\rangle/m$ (8.36) при степенном дисперсионном соотношении) необходимо знать закон дисперсии во всем диапазоне энергий, начиная со дна зоны.

Для закона дисперсии степенного вида, подставляя в (8.44) плотность состояний $D_d(\mu)$ и квадрат скорости Ферми (8.31) $v(\mu)$, получаем

[3]Это обстоятельство особенно важно для систем с низкоразмерным газом в кейновских полупроводниках с малой запрещенной зоной. Параметры законов дисперсии в низкоразмерных подзонах из-за взаимозависимости движений вдоль и поперек ограничивающих потенциалов, отличаются в этом случае от законов дисперсии исходного, "материнского" материала. Однако такие отличия существенны только при энергиях вблизи краев подзон размерного квантования.

$$\sigma_d = \frac{r c_r^{\frac{2-d}{r}}}{d 2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d} \frac{e^2\tau(\mu)}{\hbar^d} \mu^{\frac{d+r-2}{r}}. \tag{8.45}$$

Для двумерной системы это выражение принимает особенно простой вид

$$\sigma_{2D} = \frac{r}{2\pi}\frac{e^2\tau}{\hbar^2}\mu = r\frac{e^2}{h}\frac{\mu}{\hbar/\tau}, \tag{8.46}$$

где во втором равенстве мы ввели энергетическое размытие уровней $\hbar/\tau$, обусловленное конечным временем жизни τ носителей в стационарном состоянии. Поскольку в двумерной системе проводимость σ_{2D} совпадает по размерности с кондактансом (8.3), то комбинация мировых констант h/e^2 имеет, очевидно, размерность сопротивления (в системе СИ $h/e^2 = 25812,807$ Ом). Поскольку речь идет о вырожденном газе носителей, т. е. о проводимости металлического типа, то выражение (8.46) справедливо, только если можно говорить о достаточно резкой границе поверхности Ферми $\varepsilon(p_F) = \mu$. Последнее справедливо, только если неопределенность энергии $\hbar/\tau$ (размытие энергетических уровней) не превышает энергию Ферми μ, т. е. критерием металлического поведения является неравенство $\mu \gtrsim \hbar/\tau$. В связи с этим величину e^2/h естественно назвать *минимальной двумерной металлической проводимостью* [4]. Отметим, что эту универсальную константу можно выразить через скорость света c и постоянную тонкой структуры $\alpha = e^2/c\hbar$

$$\frac{e^2}{h} = \frac{c}{2\pi}\frac{e^2}{c\hbar} = \frac{\alpha}{2\pi}c \approx \frac{c}{861}, \tag{8.47}$$

т. е. проводимость двумерного газа в системе СГС имеет размерность скорости.

Для проводимости одномерного вырожденного газа формула (8.45) дает

$$\sigma_{1D} = \frac{2}{\pi}\frac{e^2\tau}{\hbar}[r c_r^{1/r}\mu^{(r-1)/r}]. \tag{8.48}$$

Если учесть, что множитель в квадратных скобках равен скорости носителей (8.31) на уровне Ферми $v(\mu)$, а произведение $v(\mu)\tau(\mu)$ есть фермиевская длина свободного пробега $\ell(\mu)$ ($\ell(\mu) = \tau(\mu)p(\mu)/m$ для $r = 2$

[4] Этот же термин (и та же величина) используется для обозначения минимальной проводимости в точке перехода металл-изолятор в модели сильной локализации в неупорядоченных системах.

и $\ell(\mu) = \tau(\mu)s$ для $r = 1$), то одномерная проводимость может быть представлена в виде

$$\sigma_{1D} = \frac{e^2}{h} 4\ell(\mu). \tag{8.49}$$

Легко видеть, что это соотношение носит более общий характер, а именно, оно справедливо для одномерного вырожденного электронного газа при любом (не обязательно степенном) законе дисперсии. Действительно, плотность состояний в одномерной системе (2.31) не зависит явно от p и определяется только скоростью носителей $D_{1D} = 4/hv$. В результате общее выражение (8.44) дает для проводимости ту же величину (8.49). В связи с этим выражением для одномерной проводимости представляет рассмотреть кондактанс одномерного электронного газа, который согласно (8.5) обратно пропорционален размеру системы L и, следовательно, подобно двумерному кондактансу, также может быть выражен через универсальную величину e^2/h

$$G_{1d} = \frac{4\ell}{L} \frac{e^2}{h} \tag{8.50}$$

с независящим от зонных параметров коэффициентом $4\ell/L$. В рамках рассматриваемой модели, основанной на уравнении Больцмана, предполагается, что размеры системы намного превышают длину свободного пробега и следовательно проводимость одномерного электронного газа при этих условиях не может превышать значения e^2/h. Можно однако задаться вопросом, какова будет проводимость в баллистическом режиме, когда длина квантовой нити меньше длины свободного пробега? Разумно полагать, что в этих условиях процессом, ограничивающим возрастание квазиимпульса в электрическом поле является рассеяние от границ нити и в качестве длины свободного пробега следует тогда рассматривать длину образца. В результате формула (8.50) приводит к кондактансу вырожденного одномерного газа носителей

$$G_{1d} = \frac{4e^2}{h}, \tag{8.51}$$

не зависящему от каких либо параметров системы носителей и определяемому только мировыми константами. Далее мы увидим, что этот,

основанный на качественных соображениях вывод полностью согласуется с результатами более последовательного рассмотрения одномерного кондактанса (см. разд. 9.3).

При экспериментальных исследованиях контролируемой величиной является, как правило не величина химпотенциала, фигурирующая в (8.45), а концентрация носителей. В структурах с низкоразмерным электронным газом , как это уже отмечалось ранее, n_d можно контролируемо изменять в довольно широких пределах в одном и том же образце и изучение концентрационных зависимостей кинетических коэффициентов является одним из наиболее мощных средств изучения таких систем. В случае степенного дисперсионного соотношения проводимость может быть выражена через концентрацию подстановкой выражения (3.13) для $\mu(n_d)$ в формулу (8.45)

$$\sigma_d = \left[\frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\right]^{\frac{2-r}{d}} rc_r \frac{e^2\tau}{\hbar^{2-r}} n_d^{\frac{d+r-2}{d}}. \tag{8.52}$$

Для квадратичного закона дисперсии это выражение принимает стандартный для друдевской проводимости вид

$$\sigma_d = \frac{e^2\tau(\mu)}{m} n_d. \tag{8.53}$$

Ранее мы видели (формула (8.39)), что подобный результат для стандартной зоны $r = 2$ справедлив и в случае произвольной степени вырождения, однако при условии, что время релаксации не зависит от энергии. Проводимость вырожденного газа определяется только временем релаксации на уровне Ферми, т. е. соотношения (8.52) и (8.53) не ограничены случаем $\tau(\varepsilon) = \tau_0 = const$. При известной зависимости $\mu(n_d)$ (формула 3.13) исследования концентрационной зависимости σ_d дают прямую информацию о характере зависимости $\tau(\varepsilon)$ и, как следствие, о доминирующих механизмах рассеяния.

Обращает на себя внимание, что в выражение (8.53) входит *полная концентрация* носителей, хотя физически ясно, что в вырожденной системе носители с энергией $\varepsilon \lesssim \mu - k_BT$ не могут дать вклада в ток (отметим, что металлическая проводимость сохраняется и при $T = 0$).

В полученные ранее формулы для теплоемкости (3.8) и длины экранирования (3.69) вырожденного газа также входила полная концентрация, но она входила в комбинации с энергией Ферми n_d/μ, а это отношение равно плотности состояний на уровне Ферми. Выражение же (8.53) содержит только концентрацию электронов ($\tau(\mu)$ является независимой характеристикой, которая не обязательно изменяется с концентрацией), причем проводимость для всех размерностей пространства линейна по n_d! Это создает впечатление, что и в вырожденной системе, подобно тому, как это имеет место для невырожденного газа носителей (см. выше), ток переносится всеми носителями заряда. Конечно, это является не более чем иллюзией, обусловленной тем обстоятельством, что при $r = 2$ плотность состояний на уровне Ферми в (8.44) растет с концентрацией как $D_d \propto n_d^{(d-2)/d}$, а квадрат фермиевской скорости как $v^2(\mu) \propto n_d^{2/d}$, что и приводит к линейному росту с n_d входящего в выражение для проводимости их произведения. В двумерном электронном газе, где плотность состояний при $r = 2$ не зависит от энергии, зависимость σ_{2D} от концентрации связана исключительно с концентрационной зависимостью скорости Ферми.

То, что электропроводность при сильном вырождении связана с участием в переносе заряда только электронов в состояниях вблизи уровня Ферми, наиболее явно проявляется в системах с линейным дисперсионным соотношением, для которых выражение (8.52)

$$\sigma_d = \left[\frac{1}{d2^{d-2}\Gamma(d/2)(\sqrt{\pi})^d}\right]^{\frac{1}{d}} s\frac{e^2\tau}{\hbar} n_d^{\frac{d-1}{d}}. \tag{8.54}$$

указывает на сублинейный характер зависимости $\sigma_d(n_d)$ (в предположении $\partial\tau/\partial n_d \leq 0$). Для одномерной системы концентрация носителей вообще выпадает из выражения для проводимости (8.54)

$$\sigma_{1D} = \frac{2}{\pi}\frac{e^2}{\hbar}\tau(\mu)s \tag{8.55}$$

(что, конечно, совпадает с выражением (8.49). В конечном счете это является следствием независимости одномерной плотности состояний и скорости от энергии при линейном законе дисперсии (см. ниже).

Как это часто бывает, простейшие в физическом отношении системы зачастую дают и наиболее яркую картину основных особенностей того или иного физического явления. Если отвлечься от взаимозависимости движений по различным степеням свободы (в простейшем еще в одном отношении случае одномерной системы этой проблемы попросту не возникает), то электронный газ с линейным дисперсионным соотношением следует в некотором отношении также отнести к наиболее простым физическим системам. Простота этой системы заключается в том, что все носители двигаются с одинаковой, не зависящей от импульса (энергии), скоростью. В связи с этим электрическое поле, как и вообще любое внешнее воздействие, в том числе тепловое, изменяя в силу уравнения движения (8.10) квантовое состояние составляющих систему частиц $\varepsilon(\mathbf{p})$, не может, тем не менее, изменить их скорости $v \equiv s = const$. Как результат, фигурирующая в обычной картине электропроводности дрейфовая скорость, т. е. изменение средней скорости в электрическом поле за время между столкновениями в системе с $r = 1$ тождественно равна нулю! Таким образом стандартное, друдевское выражение для плотности тока не только не отражает правильной картины физики токопереноса (связывая ток в вырожденном электронном газе с движением *всех* электронов, в противоречии с принципом Паули), но в случае линейного дисперсионного соотношения вообще приводит к нулевой плотности тока

$$j_d = en_d v_{dr} = \sigma E = 0, \tag{8.56}$$

а следовательно и к нулевой проводимости, вне зависимости от процессов рассеяния. Между тем здравый смысл подсказывает (и выражение (8.54) это подтверждает), что электропроводность, конечно, не исчезает и при линейном законе дисперсии. Рассматриваемая ситуация ($r = 1$) [5] является яркой иллюстрацией, что плотность тока в вырожденном газе должна выражаться не через полную концентрацию носителей и их среднюю дрейфовую скорость, а через концентрацию n_{dF} носителей, не заморо-

[5]Вспомните, что при последовательном рассмотрении линейный закон дисперсии является естественным для квазичастиц, осуществляющих перенос заряда в многоэлектронной одномерной системе (см. разд 2.4.5).

женных принципом Паули и скорость на уровне Ферми ($v = v(\mu) \neq v_{dr}$)

$$j_d = en_{dF}v(\mu) \tag{8.57}$$

(в отличие от (8.56) здесь не фигурирует дрейфовой скорости!).

Ранее, при рассмотрении теплоемкости вырожденного электронного газа в Гл. 3.1.2, в качестве n_{dF} фигурировала величина $D_d(\mu)k_BT$. Однако, в отличие от теплоемкости, которая по своему физическому смыслу требует термической активации системы, проводимость вырожденного электронного газа (металлическая проводимость) имеет место и при $T = 0$, когда определенная указанным выше образом концентрация "активных" носителей заряда ($n_{dF} = D_d(\mu)k_BT$) равна нулю. Для вычисления и установления физического смысла величины n_{dF} в случае электропроводности обратимся к выражению для плотности тока $j_d = \sigma_d E$ с проводимостью вырожденного газа, определяемой формулой (8.44)

$$j_d = \frac{e^2}{d}\frac{2S_d(\mu)}{(2\pi\hbar)^d}\tau(\mu)v(\mu)E = \frac{e}{d}\left[\frac{2V_d}{(2\pi\hbar)^d}\right]\frac{d}{p_(\mu)}v(\mu)\tau(\mu)eE, \tag{8.58}$$

где мы выразили плотность состояний на уровне Ферми через площадь поверхности Ферми $S_d(\mu)$ (см. формулу (2.23)), а в последнем равенстве площадь $S_d(\mu)$ выражена через ограниченный ею объем $V_d = S_d p(\mu)/d$. Поскольку выражение в квадратных скобках есть концентрация электронов при сильном вырождении (2.60), а произведение $\tau(\mu)eE$ равно полученному электроном от электрического поля (и переданному решетке в акте рассеяния) импульсу Δp за время между столкновениями, то

$$j_d = e\left[n_d\frac{\Delta p}{p(\mu)}\right]v(\mu). \tag{8.59}$$

Из сопоставления (8.57) и (8.59) видно, что концентрация "активных" носителей

$$n_{dF} = n_d\frac{\Delta p}{p(\mu)} \tag{8.60}$$

составляет $\Delta p/p(\mu)$ часть от их полной концентрации (в виду предполагаемой малости возмущения системы величину фермиевского импульса

$p(\mu)$ можно считать не зависящей от поля). Влияние электрического поля и рассеяния в этой картине сводится к линейному по полю и времени релаксации увеличению переданного носителям импульса, и, как следствие, такому же линейному росту концентрации участвующих в переносе носителей с энергиями вблизи уровня Ферми. В виду линейного характера зависимости $n_{dF}(E)$ выполняется закон Ома, однако понятие подвижности $\mu_n = v_{dr}/E = \sigma_n/en_d$ в общем случае, как мы видели, лишено физического смысла. При квадратичном дисперсионном соотношении $p = mv$ и

$$j_d = e\left[n_d\frac{\tau(\mu)eE}{v(\mu)m)}\right]v(\mu) = e\left[n_d\frac{v_{dr}}{v(\mu))}\right]v(\mu), \tag{8.61}$$

откуда следует, что $n_{dF} = n_d v_{dr}/v(\mu)$. Поскольку ширина фермиевской ступеньки при $T = 0$ равна нулю, равновесное распределение Ферми имеет совершенно четкую границу. Однако в присутствии электрического поля, в направлении $\mathbf{E}$ возникает "размытие" поверхности Ферми другого рода, а именно связанное с неопределенностью переданного импульса $\tau(\mu)eE$, обусловленной статистическим разбросом величин τ в отдельных актах рассеяния.

Аналогичным образом могут быть рассмотрены и другие кинетические эффекты, равно как и зависимость описывающих их кинетических коэффициентов от размерности пространства, степени вырождения или законов дисперсии. Так, в присутствии магнитного поля $\mathbf{H}\perp\mathbf{E}$ (но в отсутствии градиента температуры) уравнение (8.22) приводит к выражению для неравновесной добавки к функции распределения

$$f_1 = -\frac{\partial f_0}{\partial\varepsilon}\frac{\tau}{1+(\omega_c\tau)^2}\left[\mathbf{v}\cdot\mathbf{E} + \frac{\omega_c\tau}{H}\mathbf{v}[\mathbf{H}\times\mathbf{E}]\right]. \tag{8.62}$$

В случае изотропного закона дисперсии и при сильном вырождении подстановка этого выражения в формулу для тока (8.17) (из-за второго слагаемого в (8.62) векторы $\mathbf{j}$ и $\mathbf{E}$ теперь неколлинеарны) дает для компонент плотности тока в присутствии магнитного поля

$$j_x = \frac{\sigma_0}{1+(\omega_c\tau)^2}E_x - \frac{\sigma_0\omega_c\tau}{1+(\omega_c\tau)^2}E_y, \tag{8.63}$$

$$j_y = \frac{\sigma_0\omega_c\tau}{1+(\omega_c\tau)^2}E_x + \frac{\sigma_0}{1+(\omega_c\tau)^2}E_y, \tag{8.64}$$

в полном согласии с выражением для тензора электропроводности в магнитном поле (8.15), полученным в разд. 8.2 в рамках простого феноменологического рассмотрения.

В заключение этого раздела рассмотрим пределы применимости полуклассической теории кинетических явлений. Помимо ограничений, накладываемых использованием метода эффективной массы, применимость уравнения Больцмана ограничена рядом дополнительных приближений, использованных при его выводе.

1. Уравнения (8.21) и (8.22), как и сама неравновесная функция распределения, содержат как координату частицы, так и ее импульс, т. е. уравнение Больцмана основано на использовании понятия траектории. Для этого накладываемая соотношением Гейзенберга неопределенность координаты (длина волны де Бройля) должна быть намного меньше длины траектории, в качестве характерного размера которой в кинетических процессах выступает длина свободного пробега

$$\Delta r \sim \frac{2\pi\hbar}{p} = \lambda_{dB} << \ell = v\tau. \tag{8.65}$$

Поскольку произведение скорости и импульса есть по порядку величины энергия частицы ε, то это условие сводится к требованию, чтобы связанная с нестационарностью состояний неопределенность энергии была мала по сравнению с самой энергией

$$\frac{2\pi\hbar}{\tau} = \Delta\varepsilon(\tau) << \varepsilon = \begin{cases} k_B T & \text{– невырожденная статистика} \\ \mu & \text{– вырожденная статистика.} \end{cases}, \tag{8.66}$$

где в качестве характерной энергии фигурирует средняя тепловая энергия $k_B T$ для невырожденного газа и энергия Ферми для вырожденного. Эти условия носят весьма общий характер и должны быть выполнены при любых типах и величинах обуславливающих перенос внешних воздействий. Уравнение Больцмана таким образом неприменимо для описания процессов малой длительности, происходящих в малых объемах.

2. Поскольку линеаризованное уравнение (8.22) (уравнение Блоха) получено в предположении малого отклонения от равновесия $f_1 << f_0$, это накладывает ограничения на величину внешних полей. В случае электрического поля линеаризация уравнения (8.21) возможна, если выполнено условие

$$-\frac{\partial f_0}{\partial \varepsilon} eE\tau v << f_0. \tag{8.67}$$

Учитывая, что $eE\tau v = \Delta p v \sim \Delta\varepsilon$ это соотношение можно записать в виде

$$\Delta\varepsilon(E) << -\frac{\partial f_0}{f_0 \partial \varepsilon} = \frac{f_0}{k_B T} = \begin{cases} k_B T & \text{– невырожденная статистика} \\ \mu & \text{– вырожденная статистика.} \end{cases} \tag{8.68}$$

3. В выражение для неравновесной добавки f_1 помимо полей входит производная равновесной функции распределения $\partial f_0/\partial\varepsilon$. Для того чтобы эта величина имела физический смысл, необходимо, чтобы на ширине фермиевской ступеньки $\sim k_B T$ спектр мог рассматриваться квазинепрерывным, т. е. расстояние между соседними энергетическими состояниями было значительно меньше $k_B T$

$$\Delta\varepsilon_n << k_B T. \tag{8.69}$$

Нарушение требования квазинепрерывности спектра приводит к неопределенности и скорости электрона $\mathbf{v} = \nabla_{\mathbf{p}}(\varepsilon)$, лежащей в основе полуклассического описания переноса. Как мы видели в разд. 6.6, наиболее существенные изменения в спектре возникают при воздействии магнитного поля, особенно в двумерных системах, где, благодаря магнитному квантованию, спектр становится существенно дискретным. Хотя магнитное поле может быть классически сильным ($\omega_c\tau >> 1$), оно не должно быть квантующим, т. е. должно удовлетворять условию

$$\hbar\omega_c << k_B T. \tag{8.70}$$

Столкновительное уширение уровней Ландау может в значительной степени сгладить дискретный характер двумерного спектра в

магнитном поле, однако необходимо помнить, что наряду с условием (8.70) должно также выполняться требование (8.66), ограничивающее величину уширения сверху.

4. И, наконец, использованное рассмотрение основано на диффузионном приближении, что накладывает ограничения на размеры системы, которые должны значительно превышать длину свободного пробега. В противном случае теряет свой смысл и сама функция распределения $f(\mathbf{p}, \mathbf{r}, t)$, описывающая электронный газ, как стохастическую систему, распределение носителей в которой определяется исключительно процессами рассеяния (становится несправедливым и приближение времени релаксации). А при достаточно малых размерах, за счет размерного квантования будет модифицироваться и энергетический спектр носителей, определяющий, как мы видели, все существенные особенности процессов переноса заряда или энергии во внешних полях (см., например, условие (8.69)).

Нетрудно видеть, что все эти условия сводятся в наиболее существенных чертах к требованию, чтобы неопределенность энергии носителей и ее изменение при внешних воздействиях были малы по сравнению с типичным значением кинетической энергии носителей в термодинамическом равновесии.

8.5 Задачи

1. Получите выражение для тензора проводимости (8.15), исходя из уравнения Лоренца (8.10) для квадратичного закона дисперсии ($p = mv$).

   Указание: Используйте выражение для компонент плотности тока $j_i = env_i$, а для определения v_i распишите покомпонентно уравнение движения (8.10).

2. Найдите функциональную зависимость компонент тензора проводимости (8.15) от магнитного поля и времени релаксации на рис. 8.2 в пределе классически слабых и сильных полей.

3. Найдите аналитический вид представленных на рис. 8.3 зависимостей $\sigma_{xx} \cdot \rho_{xx}(\omega_c\tau)$, $\sigma_{xy} \cdot \rho_{xy}(\omega_c\tau)$.

4. Покажите, используя (8.44) и общее выражение для плотности состояний (2.23), что проводимость вырожденного d – мерного газа может быть представлена в виде

$$\sigma_d = \frac{e^2}{h}\frac{2}{d(2\pi)^{d-1}}\ell S_d(k_F),$$

где ℓ – фермиевская длина свободного пробега и $S_d(k_F)$ – площадь поверхности Ферми в пространстве волновых векторов. Убедитесь, что это выражение согласуется с формулами (8.45) – (8.49).

5. Рассмотрите в двумерной плоскости со случайным плавным потенциалом область, ограниченную некоторой замкнутой эквипотенциальной линией. Система помещена в магнитное поле и заполнен только нижайший уровень Ландау. Пользуясь квазиклассическим приближением найдите концентрацию электронов, заключенных внутри этой области. Магнитная длина λ_H предполагается известной, а ее отношение к длине эквипотенциали сколь угодно мало.

   Ответ: $n = 1/2\pi\lambda_H^2 = n_H$.

8.6 Вопросы для самопроверки

1. В электронном газе какой размерности совпадают по размерности

   (a) кондактанс и удельная проводимость?

   (b) ток и плотность тока?

2. Изменяются ли диссипативный и холловский кондактансы образца с двумерным электронным газом, если внутри образца сделать отверстие прямоугольной формы? Несколько таких отверстий? То же в случае круглого отверстия?

3. Как зависит критерий классически сильного и классически слабого поля от эффективной массы носителей заряда?

4. В отсутствии магнитного поля удельная проводимость равна величине, обратной удельному сопротивлению. Выполняется ли это соотношение для диссипативных (недиссипативных) компонент двумерных тензоров удельной проводимости и удельного сопротивления в магнитном поле?

5. К какой величине стремится диссипативная компонента двумерного тензора удельного сопротивления в области классически сильных магнитных полей, если диссипативная компонента удельной проводимости стремится к нулю?

6. Необходимо ли знать дисперсионное соотношение для вырожденного d-мерного электронного газа во всем диапазоне энергий от дна зоны до уровня Ферми при вычислении

   (a) проводимости?

   (b) концентрации носителей?

7. Выпишите вытекающую из уравнения Больцмана формулу для проводимости двумерного вырожденного газа, выражающую ее через квант проводимости, энергию Ферми и столкновительное уширение уровней. Каково минимальное значение проводимости двумерного вырожденного электронного газа?

8. Можно ли выразить проводимость электронного газа с линейным законом дисперсии, используя понятие скорости дрейфа (средней дополнительной скорости, набираемой электронами при их ускорении в электрическом поле за время между столкновениями)?

9. Выпишите формулы, выражающие удельную проводимость и кондактанс квантовой проволоки с вырожденным электронным газом через квант проводимости и длину свободного пробега. Что можно сказать о величине одномерного кондактанса в пределе больших и малых по сравнению с длиной свободного пробега образцов?

10. С какой скоростью связан перенос тока в вырожденном газе носителей - со скоростью дрейфа или фермиевской скоростью?

11. Согласно принципу Паули в переносе тока в вырожденном газе могут участвовать только носители вблизи уровня Ферми. Чем тогда обусловлена зависимость проводимости от полной концентрации вырожденного двумерного электронного газа

    (a) с квадратичным законом дисперсии?

    (b) с линейным законом дисперсии?

12. Для вырожденного электронного газа какой размерности и с каким законом дисперсии классическая проводимость не зависит от концентрации электронов (при независящей от концентрации длине свободного пробега)?

13. Каковы пределы применимости основанного на уравнении Больцмана подхода к описанию явлений переноса? Зависят ли они от размерности электронной системы?

Глава 9

Квантовый перенос в магнитном поле

Как это было отмечено в заключении предыдущей главы, важнейшим условием применимости классической теории переноса является предположение о слабом влиянии внешний полей на энергетический спектр носителей заряда. С другой стороны из результатов разд. 6.4 следует, что это приближение с очевидностью нарушено для двумерного электронного газа, спектр которого в магнитном поле претерпевает качественные изменения, из непрерывно становится близким к дискретному, так что фигурирующее в классической картине переноса понятие закона дисперсии теряет смысл. Понятие плотности состояний, определяющее процессы переноса при вырождении электронного газа, конечно остается, однако теперь это осциллирующая функция энергии (разд. 6.4). Результаты классической модели будут иметь смысл, только если такие осцилляции малы, что соответствует уширению уровней Ландау, намного превышающему циклотронную энергию, или температуры настолько велики, что на энергетический интервал, соответствующий ширине фермиевской ступеньки, приходится много уровней Ландау. В противном случае перестройка спектра в магнитном поле приводит к кардинальным изменениям в механизмах проводимости и такому необычному явлению, как квантование макроскопической величины - холловского кондактанса.

9.1 Квантовая диффузия в магнитном поле

В разделе 7.2.1 мы видели, что носители заряда в двумерной системе, помещенной в перпендикулярное магнитное поле, в отсутствии рассеяния дрейфуют перпендикулярно силовым линиям электрического поля и не создают направленного по полю электрического тока. Проводимость такой системы равна нулю[1]. Отметим, что это согласуется и с феноменологической картиной переноса - диссипативная проводимость в классически сильных магнитных полях $\omega_c\tau >> 1$ согласно (8.15) равна

$$\sigma_{xx} \approx n_s \frac{mc^2}{\tau H^2} \tag{9.1}$$

и при стремлении времени релаксации τ к бесконечности, если магнитное поле не равно нулю, стремится к нулю (см. рис. 8.2). Нетрудно видеть, что к тому же выводу приводит и квантовое рассмотрение. В двумерной системе спектр носителей вырождается в магнитном поле в дискретную систему уровней Ландау (двумерный "магнитный атом"). Если уровень Ферми расположен между уровнями Ландау, то электрическое поле в силу принципа Паули не может изменить энергетическое состояние носителей на лежащих ниже уровня Ферми полностью заполненных уровнях, если в системе нет процессов, способных возбудить носители на вышележащие пустые уровни. Последнее условие будет выполнено при достаточно низких температурах $k_BT << \hbar\omega_c$ (для простоты можно положить $T = 0$). То же самое остается в силе, если уровень Ферми совпадает с каким-либо уровнем Ландау, но в системе нет рассеяния. В этом случае, хотя на уровне Ландау наряду с занятыми есть и пустые состояния, электрическое поле по-прежнему не может сообщить носителям дополнительного импульса вдоль поля, поскольку в отсутствии рассеяния состояния являются полностью стационарными, т.е. энергетическая ширина уровня Ландау равна нулю.

Рассеяние вносит кардинальные изменения в эту картину движения носителей в магнитном поле. Только в присутствии рассеивающих центров в системе появляется перенос заряда в направлении электрического

[1]В этом разделе мы всюду рассматриваем перенос вдоль электрического поля, т. е. под проводимостью имеется в виду диагональная компонента тензора удельной проводимости.

поля. Столь необычное для классического переноса, при котором рассеяние уменьшает проводимость системы, поведение двумерного электронного газа в магнитном поле имеет достаточно простую физическую причину. Рассмотрим это на языке квазиклассических траекторий и квантованного магнитным полем энергетического спектра. Нетрудно видеть (рис. **??**), что упругие столкновения дрейфующих в скрещенных полях электронов с дефектами приводят к изменению их траекторий таким образом, что центры циклотронных орбит в каждом акте рассеяния скачком смещаются в направлении, преимущественно противоположном электрическому полю (для дырки по полю), что при многократных актах рассеяния приводит к направленному против поля диффузионному потоку электронов. Подчеркнем важность предположения о столкновительным механизме рассеяния, наподобие рассеяния в модели идеального газа. Носитель не должен чувствовать потенциала рассеивателя, пока он на него не "натолкнется". В качестве характерного параметра, аналогичного радиусу частиц газа, в нашей модели выступает радиус циклотронной орбиты r_H - рассеивающий потенциал выступает как рассеиватель, только если электрон пролетает на расстоянии от него, меньшем r_H. В случае плавного рассеивающего потенциала, слабо изменяющегося на расстояниях порядка магнитной длины, характер движения электрона не отличается от его дрейфа в скрещенных полях - электрон двигается вдоль эквипотенциальных линий результирующего поля, обусловленного совместным действием внешнего поля и медленно меняющегося в пространстве поля дефекта. Скачек ведущего центра циклотронной орбиты будет иметь место, только если на такой траектории встретится дефект с короткодействующим потенциалом.

Таким образом, в реальной, всегда в какой-то мере неупорядоченной, системе носители участвуют в двух движениях - дрейфуют перпендикулярно силовым линиям локального электрического поля со скоростью v_{HE} на участках его траектории между столкновениями и диффундируют за счет рассеяния в направлении электрического поля. Тот же вывод о появлении диссипативной проводимости при введении в систему беспорядка следует и из квантового рассмотрения. Поскольку в присутствии

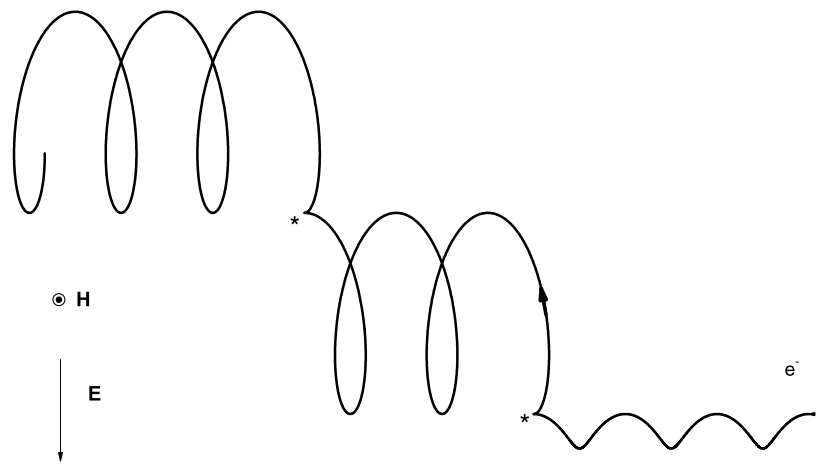

Рис. 9.1: Классические траектории в двумерной плоскости (x, y) в скрещенных электрическом и магнитном полях при рассеянии на короткодействующем потенциале. Центры орбит носителей с отрицательным зарядом при рассеянии смещаются в направлении, противоположном электрическому полю, что соответствует направленному по полю диффузионному току в согласии с законом Ома.

рассеивателей время пребывания носителя в стационарном состоянии τ ограничено, дискретный спектр уровней Ландау, согласно соотношению неопределенностей $\Delta\varepsilon\tau \sim \hbar$, модифицируется в размытые энергетические полосы шириной

$$\Gamma \sim h/\tau. \tag{9.2}$$

Если ширина столкновительно уширенного уровня Ландау намного меньше расстояния между ними и уровень Ферми расположен в области энергий между уровнями Ландау, где состояний мало, диссипативная проводимость при низкой температуре будет по-прежнему близка к нулю, как и в отсутствие дефектов. Однако, если уровень Ферми пересекает уровень Ландау, находясь в области энергий с большой плотностью состояний, для носителей имеются свободные состояния и они могут непрерывно изменять свое энергетическое состояние (кинетическую энергию) во внешнем электрическом поле, в отличие от случая идеальной системы с нулевой шириной уровня Ландау. Таким образом, при изменении относительного положения уровня Ферми и уровней Ландау (в двумерных структурах это может осуществятся как изменением магнитного поля, так и двумерной концентрации носителей) проводимость системы будет

изменяться таким образом, что в областях магнитных полей (или концентраций), соответствующих прохождению уровня Ферми между уровнями Ландау ее значения в достаточно сильных магнитных полях будут близки к нулю.

Для количественного описания проводимости для ситуации, когда уровень Ферми проходит через уровни Ландау, и появляется диссипативная проводимость, обусловленная диффузией центров осцилляторов, мы можем воспользоваться полученным в разд. 3.3 на основании соотношения Эйнштейна выражением для диффузионной проводимости (3.98) вырожденного газа носителей с энергией Ферми E_F

$$\sigma_{xx} \approx e^2 D_n D_{2D}(E_F), \tag{9.3}$$

в котором коэффициент диффузии может быть определен из выражения для диффузионной длины

$$L_d = \sqrt{D_n \tau}. \tag{9.4}$$

Как мы видели, центр орбиты электрона может переместиться в акте рассеяния на короткодействующем потенциале на расстояние, порядка радиуса циклотронной орбиты, т.е по порядку величины длина прыжка $L_d \approx r_H$. Размер циклотронной орбиты в квантующем магнитном поле находится из соотношения

$$\frac{mv^2}{2} = \frac{m\omega_c^2 r_H^2}{2} = \hbar\omega_c(n + \frac{1}{2})$$

(n- номер уровня Ландау), которое дает

$$r_H = \left[\frac{2c\hbar}{eH}(n + \frac{1}{2}\right]^{\frac{1}{2}} = \lambda_H(2n+1)^{\frac{1}{2}}. \tag{9.5}$$

Отметим, что поскольку радиус циклотронной орбиты растет с номерном магнитного уровня, то условие короткодействующего характера рассеивающего потенциала лучше удовлетворяется для высоких магнитных полей и больших номеров уровней Ландау. Подставляя в (9.4) вместо L_d это выражение получаем для коэффициента диффузии следующую оценку

$$D_n = \frac{\lambda_H^2}{\tau}(2n+1) = \frac{c\hbar}{\tau eH}(2n+1). \tag{9.6}$$

Плотность состояний вблизи центра уровня Ландау по порядку величины равна отношению числа состояний на уровне Ландау, т.е. кратности вырождения n_H, к ширине уровня (9.2)

$$D_{2D}(E_F) = n_H \frac{\tau}{h} = \frac{eH}{ch}\frac{\tau}{h}. \tag{9.7}$$

В результате формула (9.1) дает для пикового значения проводимости проводимости, соответствующего нахождению уровня Ферми в центре уровня Ландау

$$\sigma_{xx}^{max} \approx e^2 \frac{c\hbar}{\tau eH}(2n+1) \cdot \frac{eH}{ch}\frac{\tau}{h} = \frac{e^2}{\pi h}\left(n + \frac{1}{2}\right). \tag{9.8}$$

Это довольно неожиданный результат[2]. Мы использовали модель, важнейшими элементами которой было наличие магнитного поля и беспорядка в системе, тогда как формула (9.8) содержит только мировые постоянные. Из рассмотренной нами интуитивной картины видно, что независимость σ_{xx}^{max} от уширения уровней и магнитного поля обусловлена компенсацией двух эффектов. С ростом интенсивности рассеяния растет вероятность прыжка циклотронной орбиты. С другой стороны возрастает и уширение уровней, что в свою очередь уменьшает вероятность прыжка, поскольку уменьшается плотность состояний вблизи уровня Ферми, в которые должен перейти электрон при рассеянии. [3] Зависимость σ_{xx}^{max} от номера уровня Ландау обусловлена, очевидно, тем, что ширина уровней, в отличии от вероятности сбоя траектории (коэффициента диффузии) в рассматриваемой простейшей модели не зависит радиуса циклотронной орбиты. Вообще говоря, это может оказаться не совсем справедливым - фигурирующие в формулах (9.6) и 9.7) времена релаксации не могут априори считаться одинаковыми, хотя примерная

[2]Отметим, что даваемая формулой (9.8) величина всего вдвое меньше значения $\frac{e^2}{\pi\hbar}(n+1/2)$, полученного в рамках более последовательного рассмотрения для короткодействующего потенциала. Учитывая использованные в нашей модели приближения, следует признать, что согласие неплохое.

[3]Подобного рода компенсация, приводящая к независимости результата от определяющего параметра задачи, известна еще из школьного курса физики. Ускорение свободного падения не может быть получено без рассмотрения массы падающего тела, однако последняя выпадает из ответа, поскольку ускорение обратно пропорционально, а сила притяжения пропорциональна массе. (Отметим однако, что из факта эквивалентности инертной и тяжелой масс, Эйнштейн пришел к созданию общей теории относительности.)

пропорциональность между ними сохраняется. Нужно отметить, что помимо предположения о короткодействующем характере рассеяния, при выводе формулы неявно (9.8) предполагалось, что концентрация рассеивающих центров N_i должна быть настолько высокой, чтобы все электроны могли принимать участие в проводимости. Очевидно, что для этого расстояние между примесными центрами должно быть меньше радиуса циклотронной орбиты, откуда получаем

$$N_i >> \frac{eH}{2\pi c\hbar(2n+1} = \frac{n_H}{2n+1}.$$

Как и критерий малости радиуса действия рассеивающего потенциала, это условие может нарушиться при достаточно сильных магнитных полях, особенно для основного уровня Ландау - в этом случае концентрация рассеивателей должна превышать кратность вырождения уровней в магнитном поле n_H.

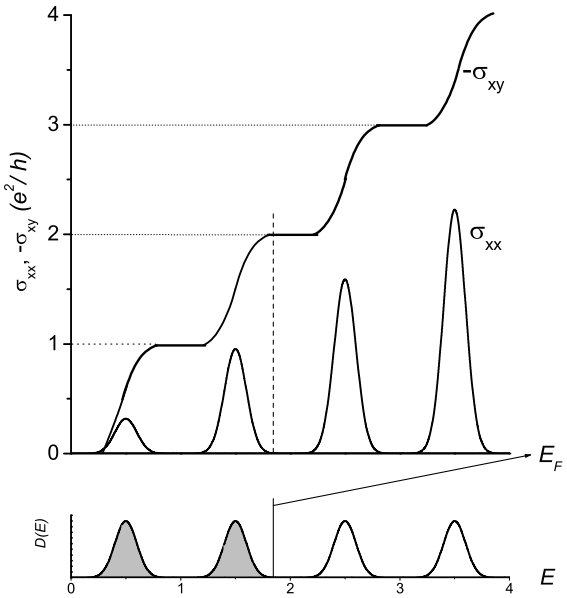

Рис. 9.2: Плотность состояний в магнитном поле в зависимости от энергии (нижний рисунок) и компоненты двумерного тензора проводимости в зависимости от энергии Ферми при изменении ее положения относительно уровней Ферми.

Зависимость квантовой проводимости от положения уровня Ферми относительно уровней Ландау проиллюстрирована на рис. 9.2, в нижней

части которого показана энергетическая зависимость двумерной плотности состояний в магнитном поле. Отметим, что качественно поведение проводимости при изменении энергии Ферми согласуется и с классическим выражением для σ, согласно которому проводимость пропорциональна плотности состояний на уровне Ферми. По мере заполнения уровней Ландау с ростом энергии Ферми, плотность состояний осциллирует, что приводит и к осцилляциям проводимости. В области не слишком сильных магнитных полей, такие осцилляции плотности состояний малы по сравнению с ее значением в отсутствие магнитного поля и близки к синусоидальным. Эта ситуация соответствует наблюдению осцилляций проводимости Щубникова - де Гааза. В случае же сильно выраженного квантования, соответствующего малой плотности состояний между уровнями Ландау, проводимость на достаточно протяженных участках ее зависимости от E_F близка к нулю, имея всплески лишь на участках при последовательном пересечении уровнем Ферми уровней Ландау, высота которых описывается выражением (9.8).

9.2 Квантовый эффект Холла

Обратимся теперь к зависимости от положения уровня Ферми холловской компоненты тензора проводимости $\sigma_{yx} = -\sigma_{xy}$. Особенно просто выражение для нее может быть получено для случая, когда уровень Ферми проходит между уровнями Ландау. Рассеяние на примесях в этом случае подавлено принципом Паули (электроны с уровней Ландау, расположенных ниже уровня Ферми, ввиду отсутствия состояний в щели между уровнями Ландау, и большой удаленности вышележащих уровней не могут при низких температурах их покинуть) и электроны в скрещенных полях дрейфуют вдоль эквипотенциальных линий со скоростью $v_{HE} = cE/H$. Связанная с таким дрейфом плотность тока $j_y = -en_s v$ равна (электрическое поле полагаем направленным по оси x)

$$j_y = -en_s c\frac{E_x}{H}. \tag{9.9}$$

Подчеркнем, что в отличие от диссипативного переноса, в холловском токе, всегда текущем вдоль эквипотенциалей, могут принимать участие

электроны, сколь угодно глубоко лежащие под уровнем Ферми. Из (9.9) для проводимости имеем

$$\sigma_{xy} = -n_s e \frac{c}{H}. \tag{9.10}$$

Нетрудно видеть, что это выражение для квантовой холловской проводимости полностью совпадает с классической формулой в пределе классически сильного магнитного поля $\omega_c \tau >> 1$ (см. (8.63)). Для случая, когда уровень Ферми находится между i-ым и $i+1$-ым уровнями Ландау, при низких температурах полностью заполнено i уровней, тогда как остальные пусты. Поскольку число состояний на одном уровне Ландау в расчете на единицу площади равно $n_H = eH/ch$ (см. разд. 6.5), то концентрация электронов в системе равна in_H и для холловской проводимости получаем

$$\sigma_{xy} = -i \frac{eH}{ch} n_s \frac{ec}{H} = \frac{e^2}{h} i. \tag{9.11}$$

Опять таки, как и в случае диссипативной компоненты проводимости, выражение для двумерной квантовой холловской проводимости определяется только мировыми физическими константами и не зависит от каких-либо параметров двумерной системы или величины внешних полей. Следует однако подчеркнуть, что в отличие от формулы (9.8), полученной с использованием ряда упрощающих, в точности никогда не реализуемых в реальных системах условий, выражение (9.11) совершенно точно (при $T \to 0$), если только уровень Ферми расположен в щели между уровнями Ландау.

9.2.1 Перколяционная модель

Зададимся теперь вопросом, можно ли наблюдать такие квантованные значения холловской проводимости (точность квантования определяется только ошибками в величинах мировых констант)? Поскольку в двумерном электроном газе удельная холловская проводимость эквивалентна интегральной (кондактансу, см. разд. 8.1), то точность измерения $\sigma_{xy} = G_{xy} = J_y/U_x$ определяется только погрешности измерения тока и напряжения. Благодаря такой эквивалентности устраняется основной

источник ошибок, связанный с размерами и геометрией образов (в случае трехмерных систем эти факторы ограничивают точность величинами порядка процента). Таким образом, остается только реализовать условия, чтобы уровень Ферми находился в щели между уровнями Ландау. Однако именно это единственное условие и не может быть выполнено со сколь-либо удовлетворительной точностью!

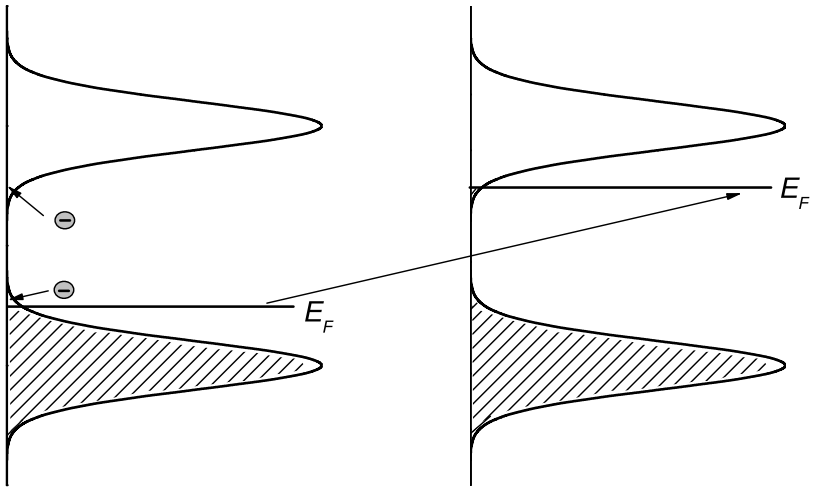

Рис. 9.3: При добавлении в двумерную систему, число электронов на единицу площади в которой на один меньше кратности вырождения уровней Ландау, всего двух электронов уровень Ферми при $T = 0$ скачком переходит на следующий по номеру уровень Ландау.

В отсутствии или в слабом магнитном поле мы можем изменять (а при известном законе дисперсии и определить) энергию Ферми изменяя концентрацию электронов (эта величина достаточно хорошо измеряется экспериментально), например изменяя напряжение на металлическом электроде полевого транзистора. В квантующем магнитном поле электроны двумерного газа распределяются по состояниям на уровнях Ландау. Нетрудно видеть, что в интересующем нас случае щели в спектре (в плотности состояний) уровень Ферми при низкой температуре фактически не может находиться в этой щели! Рассмотрим показанную на рис. 9.3 энергетическую структуру уровней в условиях, когда на верхнем из заполненных уровней Ландау осталось незаполненным одно состояние. Очевидно, что при $T = 0$ уровень Ферми находится у верхнего края этого уровня Ландау, поскольку выше его на уровне есть незаполненное

состояние. Добавим теперь в систему два электрона (увеличим концентрацию электронов). Теперь уровень Ферми будет проходить у нижнего края пустого до этого ближайшего уровня Ландау, поскольку нижайшее состояние на этом уровне заполнено . Таким образом, если в щели между уровнями Ландау нет состояний (а это непременное условие для квантования $\sigma_{x,y}$, то ничтожно малое изменение концентрации носителей (или магнитного поля, меняющего число состояний на уровне Ландау) приводит к скачку уровня Ферми с одного уровня Ландау на другой (сколь угодно мало отличаясь по концентрации, состояния на рис. 9.3 различаются по шкале энергий на достаточно большую величину ($\sim \hbar\omega_c$). Следовательно протяженным участкам зависимостей $\sigma(E_F)$ на рис. 9.3, соответствующим нулевой диссипативной проводимости и квантованным значениям $\sigma_{x,y}$ (9.11) в действительности соответствует лишь одна точка по концентрации или магнитному полю. Очевидно, что такая точка может быть найдена и при классическом переносе - мы всегда можем задать такую концентрацию двумерных электронов (или величину магнитного поля), при которой выполнено условие

$$\frac{n_s ec}{H} = i\frac{e^2}{h} \tag{9.12}$$

с целым значением i.

Самое удивительное заключается, однако, в том, что экспериментально квантованные значения σ_{xy} (9.11) наблюдаются не в одной точке, как это предсказывает теория, а именно на протяженных участках зависимостей $\sigma_{xy}(n_s)$ или $\sigma_{xy}()$, т. е. имеют тот же вид (в указанных координатах), что и на рис. 9.2. [4]. В полном соответствии с универсальностью выражения для (9.11) σ_{xy} наблюдаемые на таких участках (плато квантового эффекта Холла) значения холловской проводимости имеют совершенно одинаковые значения для разных образцов, вне зависимости от типа структур с двумерным газом или параметров используемых полупроводниковых материалов. Типичной пример такой экспериментальной зависимости приведен на рис. 9.4.

[4]Квантовый эффект Холла был открыт экспериментально в 1980г., а уже пять лет спустя за это открытие была присуждена Нобелевская премия (рекордно короткий для этой награды срок

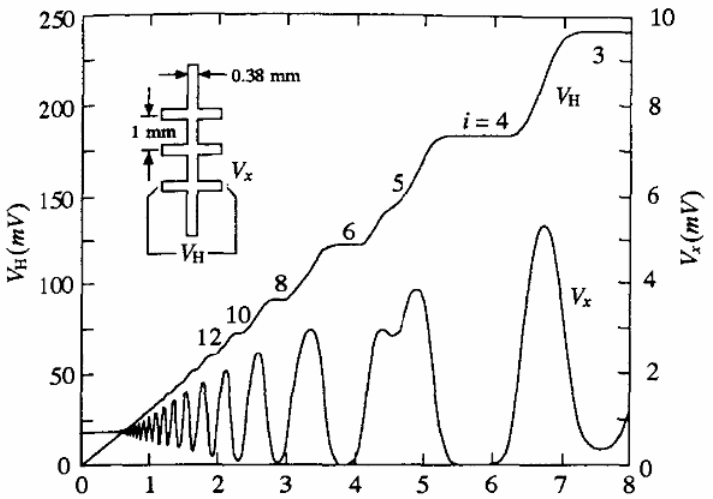

Рис. 9.4: Экспериментальные зависимости продольной и холловской разностей потенциалов от магнитного поля в двумерном электронном газе гетероструктуры AlGaAs-GaAs-AlGaAs.

Поскольку в случае двумерных систем взаимным положением уровней Ландау и энергии Ферми можно экспериментально управлять как изменением магнитного поля, так и концентрации носителей (при исследовании переноса в магнитном поле в трехмерных системах доступен только первый способ), то для дальнейшего анализа удобно выражение для холловской проводимости (9.10)

$$-\sigma_{xy} = \frac{n_s ec}{H} = \frac{e^2}{h} n_s \frac{ch}{eH} = \frac{e^2}{h} \frac{n_s}{n_H}$$

переписать в более универсальном виде

$$-\sigma_{xy} = \frac{e^2}{h}\nu, \tag{9.13}$$

выражая концентрацию носителей через кратность вырождения магнитных уровней $n_H = eH/ch$ путем введения фактора заполнения уровней Ландау $\nu = n_s/n_H$. Если условиться измерять проводимость в единицах кванта проводимости $g_0 = e^2/h$, то холловская проводимость

$$-\frac{\sigma_{xy}}{g_0} = \nu \tag{9.14}$$

просто равна фактору заполнения уровней Ландау, указывающему сколько уровней Ландау, расположенных ниже уровня Ферми, заполнено элек-

тронами (при неполном заполнении ν - дробное число) и величиной которого можно управлять изменяя концентрацию и магнитное поле

$$\nu = \frac{ch}{e}\frac{n_s}{H}.$$

В столь же универсальном виде можно записать и выражение (9.8) для пиковой величины продольной компоненты проводимости

$$\frac{\sigma_{xx}^{max}}{g_0} \approx \frac{\nu}{\pi}. \tag{9.15}$$

Напомним, что это выражение справедливо только для величины σ_{xx} в ее максимуме, когда уровень Ферми проходит посередине уровней Ландау, т.е. когда фактор заполнения принимает только полуцелые значения $\nu = (n + 1/2)$ (один из уровней заполнен наполовину).

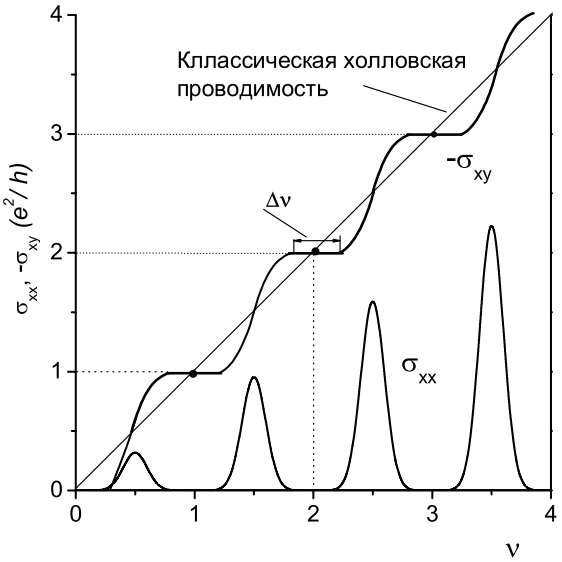

Рис. 9.5: Продольная и холловская компоненты двумерного тензора проводимости в зависимости от фактора заполнения уровней Ландау.

Будучи выраженной в таких координатах, экспериментально наблюдаемая зависимость представлена в главных чертах на рис. 9.5. Проведенный же выше теоретический анализ (как при классическом, так при квантовом подходе), показывает, что при отсутствии состояний между уровнями Ландау во всей области изменения фактора заполнения должен наблюдаться линейный рост σ_{xy} (диагональная прямая линия на рис. 9.5 с ростом фактора заполнения (горизонтальные плато не должны

наблюдаться), а области с нулевой продольной проводимостью должны отсутствовать (всплески должны смыкаться). Сходство экспериментальных кривых на рис. 9.5 с соответствующими зависимостями на рис. 9.2, полученными в рамках "наивной" теории квантового переноса, игнорирующей пининг уровня Ферми на уровнях Ландау, приводит к заключению, что в реальных структурах уровень Ферми может быть расположен в области энергий, где отсутствуют подвижные состояния, которыми в магнитном поле могут быть только состояния уровней Ландау. Об этом однозначно свидетельствуют экспериментально наблюдаемые широкие области факторов заполнений с близкой к нулю продольной проводимостью (экспериментальное отношение σ_{xx} в минимуме к его пиковому значению $\sim 10^{-7}$) с центрами вблизи целых значений ν, соответствующих расположению уровня Ферми посередине между уровнями Ландау.

Такое поведение может быть объяснено либо в предположении, что помимо состояний на уровнях Ландау в системе имеются не дающие вклада в ток (локализованные) состояния с энергиями в области щелей между уровнями Ландау, обеспечивающие локализацию уровня Ферми в щели, либо часть состояний на самих уровнях Ландау, а именно состояния вдали от их центров (где σ_{xx} близка к нулю), не дают вклада в ток, т.е. по какой-то причине являются локализованными. Первый механизм может быть связан с существованием способных обмениваться носителями с двумерным каналом поверхностных состояний вблизи интерфейсов гетероструктур. При увеличении концентрации носителей в такой системе будут заполняться как состояния уровней Ландау двумерного слоя, так и приинтерфейсные состояния. Поскольку уровень Ферми связанных друг с другом подсистем должен быть одинаковым, его пининг локализованными приграничными обеспечивает прохождение уровня Ферми между уровнями Ландау в двумерном слое. Это является достаточным условием для того, чтобы холловской проводимость принимала квантованные значения (9.14). Поскольку для заполнения (опустошения) этих состояний требуется конечный интервал изменения концентрации (или магнитного поля) то квантованным значениям (9.14) на зависимостях $\sigma_{xy}(\nu)$ (величина ν на этих зависимостях теперь определяется полной

концентраций электронов в системе, т.е. это уже не кратность заполнения уровней Ландау!) будет теперь соответствовать не одна точка, а протяженный интервал, в пределах которого заполняются локализованные состояния между уровнями Ландау, как это и имеет место экспериментально. Такой механизм в той или иной степени присутствует, конечно, во всех реальных системах, однако, чтобы объяснить наблюдаемую ширину плато $\Delta\nu$ на зависимостях $\sigma_{xy}(\nu)$, требуются нереально высокие плотности пограничных состояний.

Таким образом, причину наблюдения протяженных плато нужно искать в рамках моделей, обеспечивающих локализацию носителей самого двумерного газа электронов в магнитном поле. Нетрудно видеть, что локализации электронов в двумерной системе способствует сама специфика движения носителей в скрещенных полях, а именно их движение (между актами рассеяния) вдоль эквипотенциальных линий. Уже на примере простого кулоновского потенциала заряженной примеси можно видеть, что достаточно сильное магнитное поле способствует образованию локализованных (не несущих тока) состояний. Так, если в отсутствии магнитного поля создать связанное состояние может только кулоновский центр, притягивающий носители, т.е. имеющий заряд противоположный по знаку заряду носителя, то в магнитном поле связать электрон может и отрицательно заряженный дефект! Действительно, поскольку носители двигаются по эквипотенциали, а для кулоновского потенциала это окружности, то электрон, вне зависимости от знака кулоновского центра, может быть захвачен на круговую орбиту. Противоположным знакам заряда дефекта будут соответствовать одинаковые траектории, отличающиеся только направлением вращения. Фактически, связанный с таким движением круговой ток есть холловский ток в кулоновском электрическом поле. Очевидно, что обусловленные такими орбитами (которые могут иметь макроскопически размеры) локальные токи не могут дать вклада в ток через образец, поскольку в любом сечении образца с такой траекторией связаны равные по величине и противоположные по знаку токи. В трехмерной системе описанный механизм локализации, очевидно, невозможен, поскольку носители могут свободно двигаться вдоль

магнитного поля и отталкивание между противоположными зарядами приведет к разрушению эквипотенциальной траектории, даже если она в какой-то момент и образовалась.

В реальных структурах потенциалы множества случайно расположенных кулоновских полей и других дефектов суммируются, формируя рельеф двумерного случайного потенциала, в поле которого и двигаются свободные носители двумерного электронного газа. Потенциальная энергия носителей меняется случайным образом в пространстве, так что энергетическое положение дна двумерных подзон в разных областях образца различно. В типичных случаях обусловленные беспорядком флуктуации потенциала не могут изменить локально энергетический спектр носителей, в том числе и в магнитном поле. В каждой точке пространства энергетическое положение уровней Ландау относительно дна зоны проводимости остается тем же что и в идеальной системе. Но поскольку энергия дна зоны проводимости флуктуирует в пространстве, то флуктуирует и энергетическое положение уровней Ландау. В результате, как и в случае короткодействующих рассеивателей, уровень Ландау размывается (правильнее говорить о размытии первоначально дискретных пиков в плотности состояний). Природа уширения уровней в этих двух случаях, однако, совершенно различна. Если столкновительное уширение определяется в среднем одинаковым по образцу временем "релаксации"и потому уровни уширены в каждой локальной точке двумерного пространства, притом примерно одинаково (однородное уширение), то уширение, обусловленное плавными флуктуациями потенциала, есть результат суммирования энергетических уровней по достаточно большой области пространства, причем локально уровни Ландау есть хорошо определенные (неразмытые) стационарные состояния (неоднородное уширение). Разным энергиям на неоднородно уширенном уровне Ландау соответствуют разные области образца.

Из простейших топологических соображений очевидно, что эквипотенциальные линии $V_s(x, y) = \varepsilon$ в случае двумерного флуктуирующего потенциала будут замкнутыми как в области малых, так и больших энер-

гий в полосе между максимальным и минимальным значениями $V_s(x, y)$.[5] В первом случае они окружают минимумы "долин" , во втором вершины потенциальных "холмов" . Такие состояния, как и в рассмотренном выше случае кулоновского потенциала, являются локализованными и не могут переносить тока. Рассмотренная локализации в магнитном поле отличается от локализации Андерсона в том же потенциале при $H = 0$. Даже состояния с разной энергией соответствуют разным эквипотенциалям, т.е. физически локализованы в разных областях образца. Данной энергии может соответствовать более одной замкнутой эквипотенциальной линии с близкими энергиями. Однако если эти траектории физически значительно удалены друг от друга, то между ними не будет никакой связи и вероятностью туннельного перехода с одной траектории на другую можно пренебречь. Вероятность же образования сетки таких траекторий, туннелирование между которыми может привести к переносу носителя вдоль всего образца, ничтожно мала, за исключением случая, соответствующего, состояниям вблизи цента уровня Ландау. Из условия симметрии флуктуаций потенциала (долины и вершины) следует, что наряду с локализованными эквипотенциалями возникает и пронизывающая всю систему эквипотенциаль, соответствующая единственной энергии в центре уровня Ландау.

Таким образом учет беспорядка в системе приводит с локализации состояний, соответствующих крыльям уровней Ландау. А поскольку с точки зрения заполнения эти состояния ничем не отличаются от подвижных состояний, уровень Ферми теперь может проходить посередине между токонесущими состояниями около центра уровня Ландау (фактора заполнения близи к целочисленным значениям), обеспечивая, тем самым, достаточное условие для наблюдения квантованных значений холловской проводимости. Действительно, если уровень Ферми при некоторой концентрации проходит через верхнее крыло i-го уровня Ландау (выше его середины), то все поступившие в систему электроны при увеличении

[5]Поскольку электрон при движении в скрещенных полях сосредоточен в области порядка циклотронного радиуса r_H вокруг эквипотенциальной линии, размер последней должен быть по крайней мере много больше r_H. В противном случае понятие траектории центра циклотронной орбиты теряет смысл. Это условие может быть выполнено всегда, если полагать $H \to \infty$. Отметим, что при этом условии любой потенциал может рассматриваться как плавный.

их концентрации будут заполнять локализованные состояния между i-м и $i+1$-м уровнями (*щель подвижности*), не изменяя числа электронов в делокализованных состояниях нижележащих уровней Ландау, обеспечивающих протекание тока в системе между внешними контактами к образцу[6]. Таким образом, этот режим соответствует i-му плато на зависимости $\sigma_{xy}(\nu)$. Диссипативная компонента в этой области близка к нулю, поскольку рана нулю плотность делокализованных состояний на уровне Ферми. При дальнейшем увеличении концентрации начнется заполнение подвижных состояний $i+1$-го уровня Ландау. В этом режиме продольная проводимость уже не будет иметь нулевого значения, а холловская будет возрастать ввиду формирования и заполнения электронами делокализованной эквипотенциали $i+1$-го уровня, что соответствует переходной области между i-м и $i+1$-м плато на кривой $\sigma_{xy}(\nu)$ (см. рис. 9.5). Конечно все сказанное остается в силе, если вместо увеличения концентрации уменьшать число состояний на уровнях Ландау, уменьшая магнитное поле, поскольку это одинаковым образом изменяет фактор заполнения.

Перколяционная модель, таким образом, дает объяснение наблюдаемым экспериментально плоским участкам зависимостей $\sigma_{xy}(\nu)$ с квантованными значениями холловской проводимости. Однако в этой картине квантового эффекта Холла возникает вопрос, почему квантованные значения σ_{xy} соответствуют формуле (9.14), полученной в предположении, что холловский ток создается всеми носителями, тогда как в рассмотренной перколяционной модели в переносе участвует только (небольшая) их доля, поскольку часть электронов уходит в локализованные состояния (показательно, правда, что концентрация из конечного результата (9.14) выпадает!). Ответ на этот вопрос нам уже известен. В разд. 8.1 мы видели, что при вырезании в двумерном газе произвольной полости величина холловского кондактанса не изменяется. Уменьшение числа носителей участвующих в переносе не отражается на кондактансе, поскольку в точности компенсируется увеличением их локальной скорости в обла-

[6]По-существу это бездиссипативный холловский ток, обусловленный движением в скрещенных магнитном и результирующем (внешнее поле плюс перпендикулярное делокализованной эквипотенциали поле флуктуирующего потенциала) электрическом полях.

стях, окружающих полости. Образование же областей с локализованными носителями (с циркулирующими локальными токами) в рассмотренной *перколяционной модели* квантового эффекта Холла полностью эквивалентно полному удалению таких областей из системы. Оставшиеся же электроны двигаются вдоль делокализованной эквипотенциали, которая проходит через области с наибольшим градиентом потенциала (линия, разделяющая долины и вершины). Как следствие, они дрейфуют вдоль эквипотенциалей со столь высокими скоростями, что переносимый ими ток равен току, который бы создали все носители в отсутствии локализации.

Мы видим, что беспорядок, обычно являющийся причиной разброса параметров разных образцов, выступает в квантовом эффекте Холла как раз определяющим фактором, обеспечивающим универсальный эффект - независящее от параметров беспорядка, одинаковое для всех структур и конкретных образцов квантование холловского кондактанса в единицах e^2/h. Отметим, что нечто подобное мы наблюдаем в сверхпроводимости, когда неизбежный источник рассеяния носителей - взаимодействие с фононами как раз является тем механизмом, который и обеспечивает образование куперовских пар.

Квантовый эффект Холла представляет собой весьма уникальное явление. С одной стороны мы имеем дело с системой макроскопических размеров, с неизбежным беспорядком, находящейся во внешнем электрическом и магнитном поле, воздействие которых носит достаточно сложный характер как в отношении траекторий, так и энергетического спектра. С другой, одна из важнейших характеристик такой системы-кондактанс, оказывается независим не только от каких-либо деталей, но и магнитного или электрического поля. Результат, содержащий только мировые константы, в некотором отношении носит даже более универсальный характер, чем характеристики такой простейшей микроскопической системы, как атом,в большинство из которых входит например масса электронов. Первое делает проблему квантового эффекта Холла чрезвычайно сложной для теоретического анализа. Простота же ответа, предполагает, что объяснение этого явления должно по-видимому стро-

иться на весьма общих принципах. [7] Соображения такого рода действительно существуют. Мы уже видели, что ключевым моментом в объяснении квантового эффекта Холла, является предположение о существовании щели подвижности в помещенном в магнитное поле двумерном электронном газе. Оказывается, что существование квантового эффекта Холла может быть доказано, основываясь только на этом предположении и самых общих топологических соображениях.[8]

В заключение раздела, подчеркнем, что в квантовом эффекте Холла квантуется именно кондактанс, а не локальная удельная проводимость. Рассматриваемая система чрезвычайно неоднородна с точки зрения протекающих в ней локальных токов. Мы видели, что в области пространства, соответствующей какой нибудь долине флуктуирующего потенциала, течет циркулирующий холловский ток, среднее значение которого по полному сечению этой области, а следовательно и проводимость, равны нулю. Однако если рассмотреть половину такого сечения, то ток и приводимость будут очень высокими. Однако, поскольку с формальной точки зрения холловский кондактанс и удельная проводимость в двумерной системе эквивалентны, можно, как это часто и делается, говорить о квантовании проводимости, понимая под последней усредненную по всему образцу величину. Тем более, что уменьшение размеров образца (измерение эффекта на любой его части) или их увеличение не сказывается на величине проводимости в области плато, если изменения размеров не меняют фундаментальных свойств рассматриваемой двумерной системы (например при уменьшение размеров до величин, сравнимых с магнитной длиной, или при пространственном ограничении в двумерной плоскости, приводящем к квантованию двумерного спектра[9]).

Современное объяснение квантования холловского сопротивления учитывает конечные размеры образца и наличие краевых состояний. Одна-

[7]Если конечно не полагать, что это новый закон природы, наподобие закона Кулона, который нельзя доказать; можно только попытаться "объяснить" его себе, исходя из интуитивных (но недоказуемых)соображений, например топологических.

[8]Подробно используемая при таком подходе аргументация изложена в оригинальной работе Лафлина: Р. Лафлин. Квантовая двумерная холловская проводимость. Сб. статей Квантовый эффект Холла. Под ред. Ю.В. Шмарцева. М.: Мир, 1986. с. 160.

[9]В следующем разделе мы, однако, увидим, что формируемые при таком ограничении одномерные подзоны также имеют квантованные значения проводимости.

ко, прежде чем перейти к этой модели, рассмотрим более подробно уже затягивающийся нами в разд. 8.4 вопрос об еще одном механизме квантового переноса - баллистическом переносе в квантовых нитях.

9.3 Кондактанс квантовой нити

9.3.1 Формула Ландауэра

В Гл. 8 мы видели. что в режиме баллистического переноса проводимость вырожденной одномерной электронной системы также не зависит от ее характеристик и с точностью до коэффициента определяется той же универсальной величиной e^2/h, что и холловская проводимость в режиме квантового эффекта Холла (см. формулу (8.51). В этом разделе мы вернемся к вычислению проводимости одномерного электронного газа. Однако в отличие от Гл. 8 мы рассмотрим не удельную проводимость, а кондактанс структуры с одномерным электронным газом. Пусть квантовая проволока соединяет два идеальных электронных резервуара (см.рис. 9.7, удовлетворяющих условиям:

1. Все падающие электроны поглощаются резервуарами безотносительно к их энергии и фазе.

2. Непрерывно обеспечивают поставку электронов с энергией ниже энергии Ферми. Энергии и фазы этих электронов независимы от таковых для падающих электронов.

Для простоты рассмотрим случай, когда заполнена одна одномерная подзона (мода), полагая сильное вырождение как в металлических резервуарах, так и в проволоке (низкие температуры). Поскольку в одномерной системе рассеивающий потенциал представляет собой одномерный потенциальный барьер или яму, то естественной характеристикой вероятности рассеяния является вероятность отражения от барьера (ямы), причем при упругом рассеянии, доминирующем при низких температурах, возможно только рассеяние из состояния p_x в состояние $-p_x$. Пусть

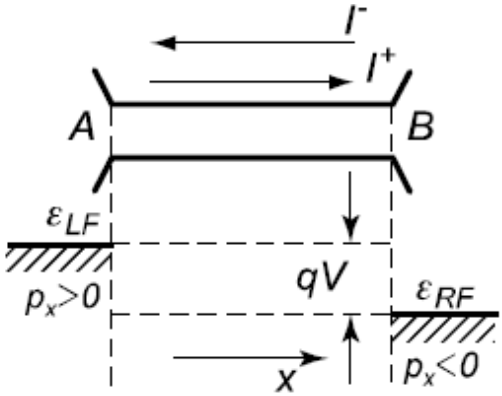

Рис. 9.6: К выводу формулы Ландауэра. E_{FL}, E_{FR} положение уровней Ферми в электронных резервуарах, между которыми приложена разность потенциалов $eV/$

μ_1 и μ_2 - хмпотенциалы резервуаров а T[10] и R соответственно вероятности электрону пройти через проволоку и отразиться назад. Из закона сохранения числа частиц эти величины связаны соотношением $T + R = 1$. Ток в одномерной проволоке (напомним , что в одномерном случае он формально совпадает с плотностью тока), а, следовательно, и ток через структуру, переносимый электронами, движущимися слева направо ($p_x > 0$) равен

$$I^{\rightarrow} = 2eT \int \frac{dp_x}{2\pi\hbar} v(p_x) f^+(\varepsilon) = 2e \int \frac{dp_x}{2\pi\hbar} \frac{d\varepsilon}{dp_x} f^+(\varepsilon) = \frac{2e}{h} T \int f^+(\varepsilon) \partial\varepsilon, \tag{9.16}$$

где во втором равенстве мы выразили скорость электрона через закон дисперсии. Здесь $f^+(\varepsilon)$ - фермиевская функция распределения электронов с положительной проекцией p_x. Обратите внимание, что это выражение получено без каких либо предположений о виде закона дисперсии и вообще не содержит концентрации электронов. Совершенно аналогично вычисляется ток электронов $I^{\leftarrow}$, текущих справа налево. Интегрирование по энергии при сильном вырождении, когда функция распределения близка к ступенчатой, тривиально и для результирующего тока получа-

[10]Мы используем общепринятое обозначение для коэффициента прохождения, совпадающее с также общепринятым обозначением температур. Однако из контекста всегда ясно о какой величине идет речь. То же относится к обозначениям для коэффициента отражения и сопротивления системы.

ем

$$I = \frac{2e}{h}(\mu_1 - \mu_2)T = \frac{2e}{h}V_{21}T, \tag{9.17}$$

где разность химпотенциалов выражена через падение напряжения между контактами (резервуарами) 1 и 2 $\mu_1 - \mu_2 = (-e)V_{21}$. Таким образом, двухконтактной кондактанс 1D проволоки определяется формулой

$$G_s = \frac{I}{V_{21}} = \frac{e^2}{\pi\hbar}T = \frac{2e^2}{h}T, \tag{9.18}$$

(*формула Ландауэра*). Как видно, кондактанс не становится бесконечно большим (сопротивление равным нулю) даже для идеального случая $T = 1$, хотя в нити полностью отсутствует рассеяние. Подчеркнем, что формула (9.18) определяет не кондактанс собственно проволоки, а кондактанс системы, включающей наряду с проволокой и контакты. В баллистическом режиме переноса ($R = 0$) конечную величину сопротивления такой системы следует скорее рассматривать как сумму двух одинаковых "контактных" сопротивлений

$$R_c = G_s^{-1}(T = 0) = G_c^{-1} = \frac{h}{4e^2} + \frac{h}{4e^2}. \tag{9.19}$$

Отметим, что подобное выражение, полученное нами в разд. 8.4 на основе качественных соображений из формулы для удельной одномерной проводимости в переделе баллистического переноса, также, по существу, характеризует не саму по себе квантовую проволоку, а систему, включающую ее границы. Действительно, к формуле (8.51) мы пришли на основании важного предположения о том, что в качестве рассеивателя носителей в баллистическом режиме выступают концы квантовой нити.

Оригинальная версия формулы Ландауэра относится к кондактансу самой проволоки и несколько усложнена. Кондактанс в обоих случаях есть, конечно, отношение тока к соответствующему падению напряжения. В рассмотренном нами случае системы резервуар + контакт + проволока это падение напряжения есть разница эффективных уровней Ферми в идеальных резервуарах. Когда $T = 0$, в системе нет тока и контакты (концы проволоки) находится в равновесии с резервуаром. Падение напряжения на проволоке в этом случае $eV = \mu_1 - \mu_2$. С другой

стороны, для идеальной проволоки ($T = 1$) мы имеем $eV = 0$, поскольку все напряжение падает на контактах. Очевидно, что в общем случае на проволоку приходится разность потенциалов $eV = (\mu_1 - \mu_2)R$. Тогда для ее кондактанса получаем выражение

$$G_w = \frac{2e^2}{h}\frac{T}{1-T}, \tag{9.20}$$

которое, собственно, и было получено Ландауэром. Нетрудно видеть, что величина, обратная кондактансу системы (9.18), т.е. ее сопротивление, может быть представлена в виде

$$G_s^{-1} = \frac{h}{2e^2T} = \frac{h}{2e^2}\frac{1-T}{T} + \frac{h}{2e^2} = G_w^{-1} + G_c^{-1}, \tag{9.21}$$

т.е. полностью согласуется с представлением о сопротивлении всей системы, как сумме сопротивлений контактов и одномерного проводника. Экспериментально обычно измеряется сопротивление проволоки вместе с контактами, поэтому на практике мы имеем дело с формулой (9.18), а не с оригинальным выражением Ландауэра (9.20). Обобщение на случай, если в одномерной системе имеется несколько заполненных одномерных подзон и носители этих подзон не взаимодействуют тривиально - их проводимости просто складываются. Если к тому же коэффициенты прохождения для разных каналов одинаковы, то во всех выписанных выше выражениях нужно просто заменить $2e^2/h$ на $N_{1D}2e^2/h$, где N_{1D} - число подзон (по аналогии с волноводами, эту величину часто называют числом мод).

9.3.2 Проводимость в конечной длины квантовой проволоке

Ранее мы уже неоднократно видели, что квантовое и классическое рассмотрение нередко приводят к близким результатам. Любопытно, что и формула Ландауэра может быть получена в рамках обычной модели Друде-Лоренца для динамической проводимости в одномерном конечном проводнике длины L, если вместо частоты внешнего поля использовать

частоту колебаний электрона, локализованного в такой ограниченной области

$$\omega_L = 2v_F/L = 2\hbar k_F/mL, \tag{9.22}$$

где v_F - скорость электрона на уровне Ферми. Роль периодического во времени внешнего поля играет в этом случае воздействие концов квантовой проволоки. В рамках друдевского рассмотрения мы исходим из уравнения движения

$$\dot{p} = eE(t) - p/\tau_{tr} \tag{9.23}$$

с зависящей от времени правой частью. Для гармонического возмущения $E(t) = eE_0 \exp(-i\omega t)$ реакция системы также будет гармонический, т.е. решение будет иметь вид

$$p(t) = p_0 \exp(-i\omega t).$$

Подстановка этих выражений в (9.23) приводит к линейному уравнению

$$-i\omega p + \frac{p}{\tau} = eE,$$

решение которого

$$p = e\tau E \frac{1}{1 - i\omega\tau}.$$

Тогда для плотности тока получаем

$$j = nev = \frac{nep}{m} = \frac{ne^2\tau}{m}\frac{1}{1 - i\omega\tau}E \equiv \sigma(\omega)E.$$

Проводимость, таким образом, равна

$$\sigma(\omega) = \frac{ne^2\tau}{m}\frac{1}{1 - i\omega\tau} \equiv \sigma_0 \left(\frac{1}{1 + (\omega\tau)^2} + i\frac{\omega\tau}{1 + (\omega\tau)^2} \right) = \sigma' + i\sigma'', \tag{9.24}$$

где

$$\sigma_0 = \frac{ne^2\tau}{m}$$

классическая (стационарная) проводимость Друде. По модулю динамическая проводимость равна

$$|\sigma(\omega)| = \frac{\sigma_0}{\sqrt{1+(\omega\tau)^2}}. \tag{9.25}$$

Комплексный характер динамической проводимости указывает на сдвиг фаз между током и электрическим полем – максимум плотности тока достигается позже, чем E.

В области малых частот $\omega\tau << 1$ доминирует действительная, диссипативная часть проводимости

$$\sigma' \approx |\sigma| = \sigma_0 \left[1-(\omega\tau)^2\right] \quad >> \sigma'' = \sigma_0\omega\tau \left[1-(\omega\tau)^2\right] \tag{9.26}$$

($\sigma'' \to 0$ при $\omega\tau \to 0$). Если же $\omega\tau >> 1$ (за один период внешнего поля электрон в среднем не испытывает столкновений, т.е. двигается баллистически), то основной вклад в модуль проводимости дает его мнимая часть

$$\sigma'' \approx \frac{\sigma_0}{\omega\tau}\left[1-(\omega\tau)^2\right] = \omega\tau\sigma' >> \sigma'.$$

Подставляя в эти выражения $\omega = \omega_L = 2v_F/L$, находим для случая $\omega_L\tau << 1$ (квантовая проволока длиной, намного превышающей длину свободного пробега $L >> v_F\tau = \ell$)

$$|\sigma| \approx \sigma' \approx \sigma_0, \quad (\sigma'' << \sigma'),$$

что совпадает с классическим выражением для статической удельной проводимости. В пределе же малых размеров $L << v_F\tau$ ($\omega_L\tau >> 1$) получаем выражение для σ, не зависящее от времени релаксации, но содержащее размер образца (!)

$$|\sigma| \approx \sigma'' \approx \frac{\sigma_0}{\omega\tau} = \frac{ne^2L}{2mv_F}, \quad (\sigma' << \sigma'').$$

Независимость σ от τ в общем-то естественна, поскольку в этом пределе рассеяния носителей на длине образца не происходит. Это полностью согласуется с поведением классической друдевской динамической проводимости при высоких частотах, когда из-за инерции электронов приобретаемая дрейфовая скорость определяется только частотой внешнего переменного поля, а не столкновениями. Именно благодаря пропорциональности $\sigma' \propto L$ кондактанс образца $G = \sigma/L$ для образцов малого

размера оказывается не зависящим от его размеров. Учитывая, что для одномерной системы

$$n = 2k_F/\pi = 4mv_F/h,$$

приходим к выражению для кондактанса, не содержащему каких-либо параметров материала и определяемому только мировыми константами, в полном соответствии с результатом, полученным выше в рамках подхода Ландауэра

$$G = \frac{2e^2}{h}.$$

Поскольку в пределе $L << v_F\tau$ в удельной электропроводности $|\sigma| \approx \sigma''$ доминирует мнимая часть, ток через образец является недиссипативным. Нетрудно видеть, что плотность мощности джоулевых потерь $P \propto \sigma'$ в области высоких частот в коротких образцах уменьшается с ростом частоты и времени релаксации, в противоположность стандартному случаю отсутствия эффектов, связанных с конечной длиной образца (низких частот) $\omega_L\tau << 1$, когда потери P растут с увеличением частоты столкновений.

В заключение этого раздела коснемся кратко вопроса об экспериментальном наблюдении предсказываемых теорией квантованных значений проводимости. В полупроводниковых низкоразмерных структурах, как мы знаем, глубиной и шириной потенциальной ямы, в которой формируется одномерный газ, можно управлять внешним электрическим полем. Следовательно, можно непрерывно изменять на том же образце как число одномерных подзон (каналов) в квантовой нити, так и степень их заполнения носителями. Очевидно, что при при последовательном формировании новых одномерных каналов (с углублением или уширением квантовой ямы) или заполнении уже имеющихся одномерных подзон с ростом концентрации (на практике чаще всего имеют место оба эффекта), проводимость систем с баллистическим характером переноса будет увеличиваться ступенчатым образом. В интервалах концентраций между заполнением соседних по номеру подзон, проводимость будет принимать квантованные значения $\sigma = 2N_{1D}e^2/H$ (N_{1D} - число уровней размерного квантования с энергиями ниже уровня Ферми) подобно тому, как

это имеет место на плато холловской проводимости в квантовом эффекте Холла (рис. 9.7). Следует однако отметить, что, поскольку режим

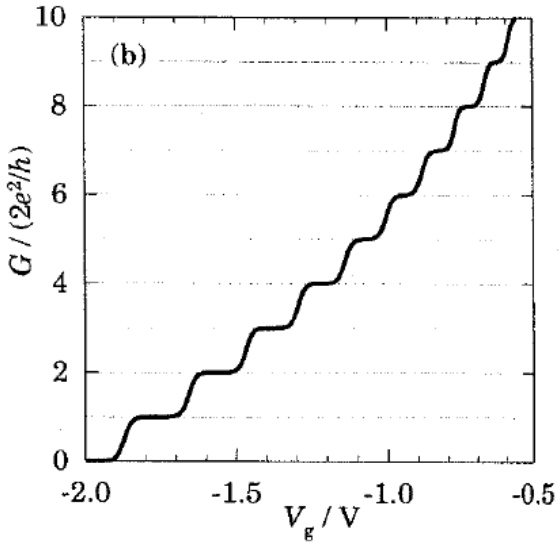

Рис. 9.7: Экспериментальная зависимость квантовой нити от напряжения на затворе, управляющем числом заполненных одномерных подзон.

баллистического переноса в структурах с квантовыми нитями легко нарушается дефектами ($T < 1$), точность квантования сопротивления в этой системе несравнимо меньше, нежели в квантовом эффекте Холла, где для подавления рассеяния, как мы видели, существуют фундаментальные причины, носящие строгий характер.

9.4 Краевая модель квантового эффекта Холла

Результаты предыдущего раздела вплотную приводят нас к механизму квантового эффекта Холла, альтернативного рассмотренной в разд. 9.2.1 модели. При изучении спектра в магнитном поле (разд. 7.1), мы видели, что вблизи краев ограниченной двумерной системы уровни Ландау (с учетом флуктуаций потенциала неоднородно уширенные) смещаются в сторону высоких энергии. Поскольку величина этого смещения определяется действием поля потенциального барьера на границе, т.е зависит от положения центра циклотронной орбиты относительного границы, спектр в магнитном поле в этой области становится непрерывным. В

результате, даже если в глубине двумерного слоя уровень Ферми расположен между центрами уровней Ландау (в щели подвижности), у поверхности образца он непременно будет пересекать все заполненные (в объеме образца) магнитные уровни (рис. 7.3). Пространственные размеры областей, соответствующих такому пересечению, определяются единственной характерной длиной системы в магнитном поле - магнитной длиной, а их расстояние от поверхности растет с увеличением номера уровня Ландау. Как мы видели, пограничным состояниям соответствуют "скачущие" траектории (рис. 7.1), причем направления пограничного дрейфа носителей и связанного с ним импульса для противоположных граней такой ограниченной системы взаимно противоположны. Такой перенос носителей вдоль границы, характеризуемый единственной, параллельной ей компонентой импульса, носит по существу одномерный характер. Согласно же результатам предыдущего раздела кондактанс, связанный с каждой такой одномерной граничной траекторией, а их число равно числу числу заполненных в глубине системы уровней Ландау, принимает в режиме баллистического переноса квантованные значения $2e^2/h$. Таким образом, если на таких траекториях нет рассеяния (T=1) (обязательное требование для объяснения чрезвычайно высокой точности квантования холловского кондактанса), то мы приходим к тому же результату, как и в модели перколяционного переноса квантовом эффекте Холла. Отличие соответствующих квантов проводимости в два раза, есть тривиальное следствие того, что проводимость нити вычислялась в отсутствии магнитного поля, когда состояния двукратно вырождены по спину. В присутствии магнитного поля это вырождение снимается (зеемановское расщепление имеет место в системе любой размерности).

Однако, в отношении рассеяния движение по граничным траекториям фундаментально отличается от рассмотренного ранее движения по делокализованным эквипотенциальным линиям в глубине системы. В последнем случае рассеяние, как мы знаем, сильно подавлено принципом Паули и существованием щели подвижности, тогда как рассеянию электронов в пограничных состояниях ничего, на первый взгляд, не препятствует. Действительно в первом случае мы имеем дело с обычной

проводимостью вырожденного (одномерного) электронного газа с уровнем Ферми, отделяющим занятые состояния и имеющиеся в непосредственной близости по энергии вышележащие пустые состояния (*спектр пограничных состояний непрерывен*!!). Имеется, однако, другое обстоятельство, учет которого полностью меняет картину переноса по пограничным состояниям. Как мы только что обсуждали, состояния каждого одномерного канала, принадлежащего одной стороне образца, могут иметь импульс, направленный строго в одну сторону, совпадающую с направлением движения по скачущей траектории. В одномерной же системе возможно только упругое рассеяние назад, т.е. как раз $\varepsilon(p) \to \varepsilon(-p)$ (неупругое рассеяние при низких температурах заморожено) требующее изменения направления импульса. Однако такое состояние существует только на противоположное границе образца. А поскольку мы имеем дело с системой макроскопических размеров, "географически" это состояние расположено на другой стороне "океана" , заполненного электронами в локализованных объемных состояниях. Вероятность такого перемещения в пространстве ничтожно мала, что и обеспечивает баллистический характер переноса по краевым состояниям и, как следствие, высокую точность квантования холловского кондактанса.

Может показаться, что в рамках краевой модели квантования холловского кондактанса эффект Холла является чисто поверхностным эффектом. Это совершенно не так! Все наиболее существенные моменты "объемной" перколяционной модели необходимы и в модели краевых состояний. Для формирования плато в холловском кондактансе уровень Ферми в объеме по-прежнему должен проходить в щели подвижности между уровнями Ландау, иначе поверхностные токи будут полностью шунтироваться большой объемной проводимостью состояний на незаполненных полностью уровнях Ландау. Нужно подчеркнуть, что в краевой модели магнитное поле играет двоякую роль. С одной стороны оно формирует спектр со щелями подвижности в глубине двумерной системы. С другой, он формирует одномерные приграничные каналы. Переход с одного квантового плато кондактанса на другое обусловлено опусканием очередного уровня объемного уровня Ландау под уровень Ферми, что

сопровождается образованием нового одномерного канала (заполнением электронами соответствующих этому уровню пограничных состояний). Пока уровень Ферми находится в области делокализованных состояний объема, поверхностные каналы, как только что обсуждалось, шунтируются проводимостью объема и наблюдается всплеск в продольной (диссипативной) проводимости системы. При входе уровня Ферми в щель подвижности в глубине образца, шунтирование устраняется и система переходит в режим, соответствующий новому плато холловского кондактанса. Диссипативная проводимость как объема так и одномерного канала становится малой, поскольку в объеме это соответствует переходу к режиму заполнения состояний на локализованных (замкнутых эквипотенциальных) орбитах, а поверхностные каналы теперь изолированы от объема (и следовательно от состояний на противоположной границе системы). Очевидно, что длина баллистического пробега в одномерном канале в режиме квантования холловского кондактанса (в равной мере это относится и к движению по делокализованным эквипотенциальным линиям в глубине образца в перколяционной модели) достигает значений, соответствующих размеру используемых образцов, т.е. лежит в области миллиметров. В то же время в слабых магнитных полях длина свободного пробега в двумерном электронном газе тех же структур на несколько порядков меньше.

9.5 Задача

1. Рассмотрите в двумерной плоскости со случайным плавным потенциалом область, ограниченную некоторой замкнутой эквипотенциальной линией. Система помещена в магнитное поле и заполнен только нижайший уровень Ландау. Пользуясь квазиклассическим приближением найдите концентрацию электронов, заключенных внутри этой области. Магнитная длина λ_H предполагается известной, а ее отношение к длине эквипотенциали сколь угодно мало.

   Ответ: $n = 1/2\pi\lambda_H^2 = n_H$.

9.6 Вопросы для самопроверки

1. Как ведет себя проводимость вырожденного двумерного электронного газа в отсутствии рассеяния при включении квантующего магнитного поля?

2. Может ли образоваться связанное состояние для носителя заряда на заряженном кулоновском центре того же знака в двумерной системе

   (a) в электрическом поле?

   (b) в перпендикулярном двумерному слою магнитном поле?

   (c) в параллельном двумерному слою магнитном поле?

   Отличаются ли ответы для трехмерного электронного газа?

3. С какой характерной длиной сравниваются радиусы действия рассеивающих свободные носители потенциалов при их классификации в магнитном поле на короткодействующие и дальнодействующие? Зависит ли эта длина от эффективной массы носителей?

4. Чему равна диссипативная проводимость вырожденного двумерного электронного газа в перпендикулярном квантующем магнитном поле в присутствии рассеивающих центров высокой концентрации с короткодействующим потенциалом? Зависит ли она от числа заполненных уровней Ландау? Как она изменяется при увеличении

   (a) магнитного поля?

   (b) концентрации рассеивателей?

5. Как изменится ответ на предыдущий вопрос, если радиус действия рассеивающего потенциала много больше магнитной длины?

6. Какова взаимная ориентация линий тока и силовых линий электрического поля при прохождении тока в двумерной системе

   (a) в отсутствии магнитного поля?

(b) помещенной в сильное параллельное двумерному слою магнитное поле?

(c) помещенной в сильное перпендикулярное двумерному слою магнитное поле?

7. К какому пределу стремятся средняя квантовомеханическая скорость и радиус циклотронной орбиты двумерного электрона при его движении в плавном электростатическом потенциале и перпендикулярном двумерному слою магнитном поле, при стремлении последнего к бесконечности?

8. По какой причине протекание тока в двумерном электронном газе в направлении приложенного электрического поля возможно только в присутствии рассеивателей, если система помещена в квантующее магнитное поле? Как можно качественно объяснить, что проводимость при этом ни от концентрации рассеивателей (при достаточно большой ее величине) ни от магнитного поля не зависит?

9. Пусть разность потенциалов между симметричными (несимметричными) холловскими потенциальными контактами на противоположных сторонах образца при протекании тока I равна V. Чему равна разность потенциалов между разными холловскими контактами на одной и той же стороне образца и между стоком и истоком, если недиагональная компонента тензора проводимости равна e^2/h, диагональная близка к нулю, а сопротивлением контактов можно пренебречь?

10. Какая величина квантуется в квантовом эффекте Холла - холловский кондактанс или холловскаыя удельная проводимость?

11. Кондактанс двумерного электронного газа в режиме квантового эффекта Холла (на плато проводимости) определяется переносом заряда только в узком канале вблизи делокализованной (пронизывающей весь образец) эквипотенциальной линии (или каналом вблизи границ образца в модели краевых состояний). Означает ли это отсутствие токов в других областях двумерной системы?

12. Может ли уровень Ферми в идеальном двумерном газе при $T = 0$, помещенном в магнитное поле, находиться между уровнями Ландау? Как ведет себя уровень Ферми с ростом магнитного поля, если концентрация электронов постоянна?

13. Могут ли рассеиваться электроны двумерного газа в режиме квантового эффекта Холла на примесях

    (a) при целочисленных значениях фактора заполнения уровней Ландау?

    (b) при полуцелых значениях фактора заполнения уровней Ландау?

14. Является ли квантовый эффект Холла в модели краевых состояний чисто поверхностным эффектом или это объемный эффект?

15. В краевой модели квантового эффекта Холла ток переносится расположенными вблизи уровня Ферми электронами одномерного канала. Принцип Паули (запрещающий рассеяние для лежащих под уровнем Ферми состояний в объеме) в этом случае не лимитирует рассеяние. Между тем, поскольку кондактанс канала равен кванту проводимости, рассеяние в канале все же отсутствует. По какой причине электроны в краевых состояниях не могут рассеиваться?

Глава 10

Эффекты релятивистского типа

Двумерный электронный газ в приповерхностных квантовых ямах кейновских полупроводников, особенно узкощелевых (УЩ) и бесщелевых (БЩ), обладает целым рядом специфических и интересных в физическом отношении особенностей. Часть из них (большие глубина и ширина самосогласованной квантовой ямы и, как следствие, образование и заполнение нескольких подзон размерного квантования; слабое влияние многоэлектронных эффектов и флуктуаций потенциала - ввиду больших значений энергии Ферми; вырождение электронного газа в объеме для обогащенных электронных слоев, приводящее к ряду эффектов, обусловленных вкладом электронов континуума в экранирование поверхностного потенциала) проистекает непосредственно из малости в этих материалах кейновской эффективной массы m. Наиболее же значительные особенности обусловлены многозонным характером описывающего их спектр гамильтониана. В рамках простейшей, однако, при малой ширине запрещенной зоны E_g адекватной в широкой области экспериментальных условий двухзонной аппроксимации $\mathbf{k} \cdot \mathbf{p}$ уравнение эффективной массы в пренебрежении спином аналогично уравнению Клейна-Гордона (KG) и электронные 2D системы на основе этих полупроводников могут рассматриваться как релятивистский аналог по отношению к традиционным 2D системам на основе материалов с параболическими зонами. Соответственно они сложнее для теоретического анализа, однако и существенно богаче в физическом отношении.

Благодаря квазирелятивистской непараболичности спектра эффективные массы электронов в двумерных подзонах зависят от двумерного волнового вектора $k = \left|\sqrt{k_x^2 + k_y^2}\right|$, номера подзоны i и глубины приповерхностной потенциальной ямы $\phi_s = \mu_s - \mu_b$ и подзонный спектр не может быть описан однопараметрическим дисперсионным соотношением. Из-за взаимозависимости движений вдоль двумерной плоскости и в направлении конфайнмента z протяженность волновой функции и сами условия связывания зависят от k, что приводит к обрыву дисперсионных подзонных ветвей со стороны малых k и может проявиться в межзонных оптических переходах, а в обогащенных слоях и в магнитоосцилляционных эффектах. Уже при небольших поверхностных концентрациях изгиб зон ϕ_s оказывается бóльшим E_g и эффекты релятивистского типа не только не могут быть игнорированы, но и не могут рассматриваться как поправки к параболическому приближению. Более того, для материалов с $|E_g| < 50 - 100$ мэВ практически во всем актуальном диапазоне подзонных концентраций n_i "энергией покоя"$ms^2 = |Eg/2|$ зачастую можно вообще пренебречь. Ввиду такого ультрарелятивистского характера движения электронов в приповерхностных слоях УЩ полупроводников следует ожидать масштабной инвариантности подзонных спектров относительно μ_s.

Существенная специфика имеет место в отношении спиновых эффектов в этих материалах, наиболее важные из которых -спиновое расщепление и "резонансный"сдвиг подзонных спектров за счет спин-межзонного перемешивания состояний поверхностным электростатическим потенциалом. В отличие от широкозонных полупроводников с параболическим спектром, в которых снятие спинового вырождения спектра за счет спин-орбитального взаимодействия (СОВ) описывается введением в трехмерный гамильтониан дополнительного члена, связанного с нарушением инверсионной симметрии на поверхности (модель Рашбы), в узкощелевых полупроводниках симметрия относительно пространственной инверсии нарушена в объеме и эффекты СОВ содержатся в исходном трехмерном гамильтониане. Возможная нестационарность состояний двумерного слоя- еще одна особенность 2D систем в приповерхностных слоях УЩ

полупроводников. Благодаря малости щели, отделяющей состояния электронной и дырочной ветвей спектра, а, с другой стороны, большим значениям глубины и ширины приповерхностной ямы, уже при небольших n_s дно 2D подзон в инверсионных слоях (а при достаточно больших n_s и в обогащенных) опускается ниже потолка валентной зоны в объеме, так что 2D состояния являются, по-существу, резонансными. В связи с этим не очевидна сама возможность двумеризации электронного газа при достаточно больших ϕ_s . Рассматриваемый класс материалов представляется идеальным объектом для изучения проблемы туннельного перемешивания состояний различных зон. Наиболее важно то обстоятельство, что при учете эффектов спинового (спинорного) типа, как будет показано ниже, существенно нарушается имеющая место в приближении KG идентичность параметров 2D подзон в материалах с прямой и инверсной структурой зон.

Роль отмеченных выше эффектов непосредственно определяется величиной E_g и ms^2 (и ожидается возрастающей при их уменьшении) и сравнительные исследования параметров 2D подзон в материалах с различной величиной и знаком щели представляют наибольший физический интерес.

10.1 Теоретическая модель

Для малой ширины запрещенной зоны E_g при описании подзонных спектров можно ограничиться шестизонной $\mathbf{k} \cdot \mathbf{p}$ моделью Кейна (точно учитывающей взаимодействие зон Γ_6 и Γ_8), пренебрегая влиянием удаленных зон, в том числе зоной Γ_7, и членом кинетической энергии с массой свободного электрона. Исключая из 6×6 матричного уравнения для шестикомпонентного спинора компоненты, относящиеся к тяжелой ветви зоны Γ_8, которая предполагается "бездисперсионной"(масса тяжелых дырок $m_{hh} = \infty$), задача для одномерного потенциала $V(z)$ сводится к системе линейных дифференциальных уравнений первого порядка для огибающих волновых функций

$$\frac{df_{1,4}}{dz} = Af_{2,5} \pm \frac{k}{2} f_{1,4}, \quad \frac{df_{2,5}}{dz} = Bf_{1,4} \mp \frac{k}{2} f_{2,5}, \tag{10.1}$$

где

$$A = -\sqrt{\frac{3}{2}} \frac{E_+}{s\hbar}, B = \sqrt{\frac{3}{2}} \frac{1}{s\hbar} \left(E_- - \frac{3}{4} (s\hbar k)^2 / E_+ \right), E_\pm = \varepsilon - V(z) \pm ms^2$$

Здесь и далее энергии отсчитываются от середины запрещенной зоны в объеме ($V(\infty) = 0$), штрихи означают дифференцирование по z. Квадрирование системы (10.1) приводит к четырем не связанным в явном виде уравнениям (в общем случае эти уравнения "сцеплены"самосогласованным поверхностным потенциалом $V(z) = V(f_{1,4}(z), f_{2,5}(z))$ и налагаемыми на компоненты спинора граничными условиями):

$$\left[\hat{k}_z^2 + i \left(\frac{A'}{A} \right) \hat{k}_z + AB + \frac{k^2}{4} \pm \left(\frac{A'}{A} \right) \frac{k}{2} \right] f_{1.4}(k, z) = 0, \tag{10.2}$$

$$\left[\hat{k}_z^2 + i \left(\frac{B'}{B} \right) \hat{k}_z + AB + \frac{k^2}{4} \mp \left(\frac{B'}{B} \right) \frac{k}{2} \right] f_{2,5}(k, z) = 0, \tag{10.3}$$

первые два из которых относятся к зоне Γ_6, вторые к легким ветвям зоны Γ_8. Для положительного изгиба зон (квантовая яма для носителей с отрицательным зарядом) спектр двумеризованных электронов зоны Γ_8 (далее "p-электроны") приповерхностного слоя в полупроводниках с $E_g < 0$ описывается уравнением для волновых функций $f_{2,5}$, тогда как состояния дырочного континуума в объеме (зоны Γ_6) уравнением для $f_{1,4}$. Для приповерхностных слоев в полупроводниках с $E_g > 0$ (двумерные "s-электроны"в приповерхностной квантовой яме) имеет место обратная ситуация.

В отличие от соответствующих уравнений Дирака, квадрированные уравнения Кейна (10.2,10.3) для электронной и "позитронной"(дырочной) ветвей спектра из-за вырождения зоны Γ_8 (даже при бесконечной массе тяжелой ветви Γ_8) не переходят друг в друга при операции зарядового сопряжения, т.е. существенно неэквивалентны. Следует подчеркнуть, что неэквивалентность сохраняется и в предельном случае полупроводника с нулевой щелью $E_g = 0$, т.е. уравнения различны для $E_g = +0$ и

$E_g = -0$. Симметрия восстанавливается лишь при игнорировании "спинорных"эффектов, описываемых вторыми (электрон-дырочное перемешивание электрическим полем dV/edz) и пятыми (спин-орбитальное взаимодействие) членами в уравнениях (2,3).

Уравнения (2,3) содержат первые производные волновой функции и не допускают простой физической интерпретации. Коэффициенты при операторе $\hat{k}_z$ в (10.2) и (10.3) имеют вид логарифмических производных, так что первые производные в этих уравнениях устраняются простой подстановкой

$$f_s^{\pm} = \sqrt{A} f_{1,4};\ f_p^{\pm} = \sqrt{B} f_{2,5}$$

В результате подзонные уравнения эффективной массы приобретают вид уравнений Шредингера (в уравнении для $f_p^{\pm}$ инвертированы зоны, т.е. проведена операция зарядового сопряжения $V \to -V$ и сменено направление отсчета энергии, что эквивалентно изменению знака E_g)

$$\left[\hat{k}_z^2 - \frac{2ms^2}{s^2\hbar^2}\left(E - U_{s,p}\right)\right] f_{s,p}^{\pm} = 0, \tag{10.4}$$

с эффективной энергией

$$E = \frac{\varepsilon^2 - m^2 s^4 - s^2\hbar^2 k^2}{2ms^2} \tag{10.5}$$

и зависящим от энергии эффективным потенциалом

$$U_{s,p}\left(\varepsilon, k, z\right) \equiv U = U_0 + U_{sp}, \tag{10.6}$$

в котором мы выделили независящий от "спина"и одинаковый для "s-"и "p-электронов"член

$$U_0 = (2V\varepsilon - V^2)/2ms^2 \tag{10.7}$$

и описывающий спинорные эффекты потенциал $U_{sp} = U_{so} + U_r$, в котором

$$U_{so} = \pm g a_{s,p} \frac{s^2\hbar^2}{4ms^2} \frac{dV/dz}{E_+} k \tag{10.8}$$

-член спин-орбитального взаимодействия, обусловленный перемешиванием электрическим полем различных "спиновых"состояний в каждой из зон,

$$U_r = \frac{s^2\hbar^2}{2ms^2}\left[\frac{3}{4}b_{s,p}\left(a_{s,p}\frac{dV/dz}{E_+}\right)^2 + \frac{1}{2}a_{s,p}\frac{d^2V/dz^2}{E_+}\right] \tag{10.9}$$

-"резонансный"член, порожденный "спин-межзонным"перемешиванием состояний зон Γ_6 и Γ_8 электрическим полем (второй член в уравнениях (2,3), который при стандартном рассмотрении интерпретируется, как ответственный за зинеровское межзонное (резонансное) туннелирование в электрическом поле. Отметим, что в рассматриваемой 2+1 мерной (пространственно) задаче входящий в выражение для E независящий от z кинетический член $s^2\hbar^2k^2/2ms^2$ является по-существу аналогом центробежного потенциала в центросимметричной трехмерной задаче и может быть формально включен в потенциал $U_0(z)$.

Уравнения для "s-"и "p-электронов"отличаются коэффициентами $a_{s,p}, b_{s,p}$ в выражениях для потенциалов U_r и U_{so}. Для кейновских "s-электронов "$a_s = b_s = g_s = 1$; для "p-электронов"$g_p = 1$,

$$a_p = \left(1 + \frac{3(s\hbar k)^2}{4E_-^2}\right) / \left(1 - \frac{3(s\hbar k)^2}{4E_-E_+}\right), \tag{10.10}$$

$$b_p = \frac{1}{3} + \frac{2}{3}\left[\frac{1}{1+E_-/E_+} + \frac{1}{a_p - 1}\right]^{-1} + \frac{2}{3a_p}\left[1 + 2\frac{(a_p-1)\,E_+/E_-}{a_p + E_-/E+}\right] \tag{10.11}$$

Уравнение для дираковских ("D-") электронов отличается от уравнения для "s-электронов"только значением $g = 2$. Нетрудно видеть, что в пределе $E_g \to \infty$ уравнение (10.4) переходит в уравнение Шредингера для исходного потенциала $V(z)$.

Клейн-гордоновский член U_0 в эффективном потенциале обладает существенной спецификой, проясняющей физические причины отмеченного выше стационарного характера 2D состояний в кейновских полупроводниках даже при нулевой щели (для (гипотетического) полупроводника со смыкающимися при $k = 0$ параболическими зонами локализованных поверхностных состояний в одномерном потенциале не должно существовать). Вхождению поверхностных электронных уровней в сплошной спектр дырочного континуума соответствуют энергии $\varepsilon < -ms^2$. При

малых z в потенциале U_0 доминирует релятивистский член $-V^2/2ms^2$, соответствующий (независимо от знака V) притяжению. Вдали от поверхности $U_0 \approx \varepsilon V(z)$, что для притягивающего потенциала $V(z) < 0$ соответствует (поскольку $\varepsilon < 0$) отталкиванию. Таким образом, при произвольной форме $V(z)$ в KG потенциале содержится барьер, отсутствующий в исходном потенциале $V(z)$. Нетрудно видеть, что U_0 достигает максимального значения $U_{0max} = \varepsilon^2/2ms^2 > 0$ в точке z_m, соответствующей $V(z_m) = \varepsilon$, а точки поворота $z_{1,2}$ в классически запрещенной области под барьером определяются условием $E = U_0(z)$, что соответствует $V(z_{1,2}) = \varepsilon \pm \sqrt{s^2\hbar^2k^2 + m^2s^4}$. Поскольку $E < U_{0max}$, то все 2D состояния отделены от объема барьером даже при $E_g = 0$, т.е. близки к стационарным (за исключением дна подзон $k = 0$ в материале с $m = 0$; однако из-за отражения от вершины барьера и в этом случае резонансный сдвиг и размытие не велики). Для состояний с $E < 0$, т.е. удовлетворяющих условию $\varepsilon < \sqrt{s^2\hbar^2k^2 + m^2s^4}$, всюду за точкой поворота z_1 (при $\varepsilon = \sqrt{s^2\hbar^2k^2 + m^2s^4}$ вторая точка поворота уходит на бесконечность) $E < U_0$, т.е. электроны в этих состояниях (а на них приходится более 80 % всех 2D электронов -см. разд.5) являются строго стационарными.

При учете спиновых эффектов строго стационарными оказываются все двумерные состояния. В спинорном потенциале U_{sp} из-за сингулярности для "s-"и "D-электронов"в точке $z = z_+$, соответствующей условию $E_+(z = z_+) = 0$ и для "p-электронов"в точке $z = z_a$, определяемой соотношением $V(z_a) = \varepsilon - \sqrt{3s^2\hbar^2k^2/4 + m^2s^4}$, имеется бесконечно высокий барьер. Существование такой непроницаемой потенциальной стенки диктует граничные условия для волновых функций в объеме $f^{\pm}_{s,p}(z_m) = 0$, и обеспечивает стационарный характер двумерных состояний при любых k и E_g, даже если они и в "резонансе"с состояниями валентной зоны.

Наличие резонансного члена U_r приводит к дополнительному, по отношению к невозмущенной спиновыми эффектами задаче, отталкиванию электронов со стороны больших z, т.е. к "выталкиванию"квантовых уровней из ямы (в пренебрежении U_{so}), уменьшающемуся, однако, с ростом двумерного квазиимпульса, поскольку при больших k в (10.4) доминирует "центробежный"член $s^2\hbar^2k^2/2ms^2$. Поскольку $a_p, b_p > 1$, то ука-

занный эффект (при $k \neq 0$) выражен более ярко для "p-электронов". Без учета спин-орбитального взаимодействия U_{so} подзонные спектры для "s-"и "D-электронов"одинаковы, величина же спин-орбитального расщепления из-за дополнительного множителя g=2 в последнем случае больше. Отметим, что это согласуется с разными значениями "эффективных"g-факторов для дираковских и кейновских электронов (отношение орбитального и спинового расщепления в магнитном поле для последних вдвое меньше). Спин-орбитальное расщепление для "p-электронов"также больше, чем для "s-электронов", поскольку наряду с линейным по k членом в эффективном гамильтониане имеется дополнительный (близкий к кубическому) член. Уравнения для "s-","p-"и "D-электронов"при учете спиновых эффектов совпадают только при описании дна двумерных подзон ($k = 0$) и в крайнем УР подзонном пределе $k \to \infty$.

В дальнейшем, как при описании поверхностного потенциала так и при квантовании спектра, мы как и в предыдущих главах, ограничимся квазиклассическим приближением, которое для рассматриваемой задачи является оптимальным компромиссом между точностью, ясностью физической интерпретации и трудоемкостью расчетов. При описании таких подзонных параметров, как заселенности подзон и циклотронные эффективные массы этот подход прекрасно согласуется с результатами численных самосогласованных расчетов. Подзонные законы дисперсии $\varepsilon_i(k)$ определяются из условия квантования Бора-Зоммерфельда

$$\int_0^{z(k_z=0)} k_z(z)\, dz = \pi \left(i + \frac{3}{4} \right), \tag{10.12}$$

где k_z, рассматриваемое как c-число, дается уравнением (10.4). Условие (10.12) с k_z из (10.4) является учитывающим спинорные эффекты обобщением правила квазиклассического квантования для релятивистских электронов[1]. Для самосогласованных потенциалов, описываемых уравнением Пуассона, (10.12), как мы видели в Гл. 5? удобно модифицировать, перейдя к интегрированию по потенциалу

[1]Изложение квазиклассического метода в рамках релятивисткого подхода можно найти в статье Г.К. Артимович, .В.И. Ритус, ЖЭТФ. **104**, 2912, (1993.

$$\int_{V(z=0)}^{V(k_z=0)} k_z\left(\varepsilon, k, V\right)\left(\frac{dV}{dz}\right)^{-1} dV = \pi\left(i+\frac{3}{4}\right), \tag{10.13}$$

поскольку при этом не требуется знать явную зависимость $V\left(z\right)$, первый же интеграл уравнения Пуассона dV/dz, как функция V, вычисляется элементарно. Отметим, что численное интегрирование уравнения (10.4) приводит к результатам, практически идентичным ВКБ приближению. Высокая точность квазиклассического подхода обусловлена в значительной степени тем обстоятельством, что носители основной подзоны, для которой квазиклассический подход наименее корректен, в рассматриваемой задаче локализованы, в основном, в области линейного хода потенциала (см.(10.14)), для которого ВКБ приближение дает результаты, близкие к точным решениям уравнения Эйри.

10.2 Квазиультрарелятивистское приближение

Особенности спектра, обусловленные релятивистским характером рассматриваемых систем, в том числе спиновые эффекты и их специфика для "s-","p-"и "D -электронов"наиболее ярко выражены. С другой стороны, в рамках этой аппроксимации (справедливой в широкой области актуальных экспериментальных условий) зачастую удается получить аналитическое описание измеряемых на опыте параметров 2D подзон (в случае квадратичного закона дисперсии аналитических выражений для параметров подзон при реалистичных ограничивающих потенциалах получить не удается).

В рассматриваемом пределе закон дисперсии линеен $E = s\hbar k$ (объемная дисперсия нейтринного типа) и локальная объемная концентрация электронов $n\left(z\right) = \left(\mu\left(z\right)/s\hbar\right)^3/3\pi^2$, что в рамках метода локальной плотности (в пренебрежении зарядом слоя обеднения или электронов континуума ($\mu_b = 0$)) дает:

$$\frac{dV}{dz} = -\frac{d\mu}{dz} = \frac{\beta V^2}{2\pi s\hbar};\ V\left(z\right) = -\mu\left(z\right) = \frac{\mu_s}{1+z/z_s}, \tag{10.14}$$

где характеристическая длина экранирования

$$z_s = 2\pi s\hbar/\beta\mu_s$$

, а параметр

$$\beta = \sqrt{(8\pi e^2/3s\hbar\epsilon)}$$

. Выражения для спиновых членов в эффективном потенциале принимают простой вид ($E_+ = E_- = \varepsilon - V$):

$$U_{so} = \pm\left(\frac{\beta}{2\pi}\right) g a_{s,p} \frac{s\hbar}{4ms^2}\frac{V^2}{E_+}k,$$

$$U_r = \left(\frac{\beta}{2\pi}\right)^2 \frac{1}{2ms^2}\left[\frac{3}{4}b_{s,p}\left(a_{s,p}\frac{V^2}{E_+}\right)^2 + a_{s,p}\frac{V^3}{E_+}\right],$$

$$b_p = 1 + \frac{4}{3}\frac{\left(1 - a_p^{-1}\right)}{(1 + a_p)} \simeq 1,$$

Следует обратить внимание, что, хотя в анализируемом случае $\left(ms^2 = 0\right)$ эффективная энергия E и потенциалы U_0, U_r и U_{so} сингулярны, уравнение эффективной массы (10.4) из-за множителя $2ms^2/s^2\hbar^2$ какой-либо особенности при $ms^2 = 0$ не содержит. В виду масштабной инвариантности УР предела относительно глубины приповерхностной квантовой ямы $\mu_s = -V(0)$ удобно использовать (безразмерные) эффективные энергии и потенциалы, нормированные на $\mu_s^2/2ms^2$.

10.3 Параметры двумерных подзон в пренебрежении спиновыми эффектами (приближение Клейна - Гордона

В приближении KG $U_{sp} = 0$, которое представляет как самостоятельный интерес, так и с точки зрения выяснения влияния спиновых эффектов, интеграл в (10.13) берется в элементарных функциях и двумерный спектр описывается трансцендентным уравнением:

$$\ln\left|\frac{\varepsilon + \mu_s + \sqrt{(\varepsilon + \mu_s)^2 - s^2\hbar^2k^2}}{s\hbar k}\right| - \frac{\sqrt{(\varepsilon + \mu_s)^2 - s^2\hbar^2k^2}}{\mu_s}$$

$$-\frac{\varepsilon J}{\sqrt{|\varepsilon^2 - s^2\hbar^2k^2|}} = \frac{\beta}{2}\left(i + \frac{3}{4}\right), \tag{10.15}$$

где

$$J = J_{nr} = \ln \frac{\varepsilon\left(\varepsilon + \mu_s\right) - s^2\hbar^2k^2 + \sqrt{\left[\left(\varepsilon + \mu_s\right)^2 - s^2\hbar^2k^2\right]\left(\varepsilon^2 - s^2\hbar^2k^2\right)}}{\mu_s s\hbar k}$$

соответствует состояниям с $|\varepsilon| < s\hbar k$, а

$$J = J_r = -\arcsin\frac{\varepsilon\left(\varepsilon + \mu_s\right) - s^2\hbar^2k^2}{\mu_s s\hbar k} - \frac{\pi}{2},$$

-состояниям с $|\varepsilon| > s\hbar k$. Уравнение (10.15) имеет решения при любых i, т.е. в сколь угодно мелкой яме имеется бесконечное число связанных состояний в соответствии с дальнодействующм характером УР потенциала (10.14) (при $z \gg z_s$ потенциал $V(z) \approx 2\pi s\hbar/\beta z$, т.е. кулоновского вида). Законы дисперсии $\varepsilon(k)$ (10.15) (для первых трех подзон приведены на рис. 10.1) масштабно инвариантны относительно μ_s, что является естественным следствием УР характера рассматриваемой системы. Из (10.15) непосредственно следует масштабная относительно μ_s инвариантность энергий связи

$$\varepsilon_i\left(k = 0\right) = -c_{\varepsilon i}\mu_s, \tag{10.16}$$

где масштабные коэффициенты $c_{\varepsilon i}$ определяются уравнением:

$$c_{\varepsilon i} - \ln c_{\varepsilon i} = \frac{\beta}{2}\left(i + \frac{3}{4}\right) \tag{10.17}$$

Масштабные коэффициенты для фермиевских квазиимпульсов

$$k_{Fi} = c_{ki}\mu_s/s\hbar \tag{10.18}$$

определяются из (10.15) при $\varepsilon = 0$, что приводит к уравнению:

$$\ln\left[1 + \sqrt{\left(1 - c_{ki}^2\right)/c_{ki}^2}\right] - \sqrt{\left(1 - c_{ki}^2\right)} = \frac{\beta}{2}\left(i + \frac{3}{4}\right), \tag{10.19}$$

ассимптотические решения которого

$$c_{ki} \approx c_0\left(\exp -\frac{\beta i}{2} + \frac{1}{3}\exp\frac{3\beta i}{2}\right),\ c_0 = 2\exp -\left(1 + \frac{3\beta}{8}\right) \approx 0.50, \tag{10.20}$$

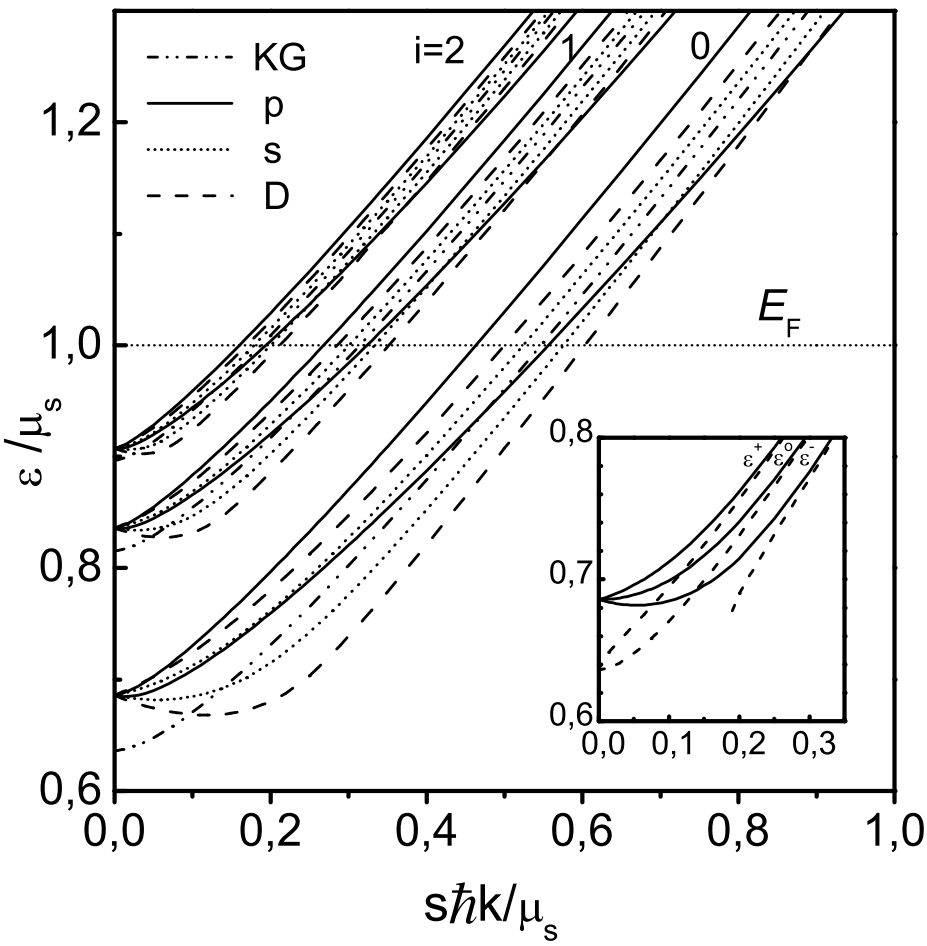

Рис. 10.1: Двумерный спектр для кейновских и дираковских электронов в УР пределе. На вставке приведены дисперсионные кривые вблизи дна основной подзоны, рассчитанные с учетом (сплошные линии) и без учета (пунктирные линии) резонансного члена в эффективном потенциале.

отличаются от точных менее, чем на 1%.

Поверхностная плотность носителей в подзонах

$$n_i = \frac{1}{2\pi}\left(c_{ki}\mu_s/s\hbar\right)^2, \tag{10.21}$$

а полная поверхностная плотность 2D -электронов

$$n_s = \sum n_i = \frac{1}{2\pi}\left(c_s\mu_s/s\hbar\right)^2,\ c_s = \sqrt{\sum_0^\infty c_{ki}^2} \approx 0,674\,(\text{при}\ \ \beta = 1)\,. \tag{10.22}$$

Двумерная концентрация n_s почти на 30% меньше поверхностной плотности электронов в неквантованной яме (УР модификация приближения

Томаса-Ферми) $n_s^{cl} = \mu_s^2/3\pi\beta\hbar^2 s^2$, что, на первый взгляд, указывает на заметную несамосогласованность рассматриваемой схемы расчета. Однако, приближение Томаса-Ферми не согласуется с нулевыми (или близкими к ним) граничными условиями для волновой функции на поверхности. Использование модифицированного метода локальной плотности[2], учитывающего затухание волновой функции на расстояниях от границы порядка половины длины волны де Бройля, приводит к значениям n_s^{cl} для классической ямы, фактически совпадающим с n_s, рассчитанным согласно (10.22). В соответствии с (10.21) и (10.22) заселенности подзон пропорциональны n_s

$$n_i = (c_{ki}/c_s)^2 \, n_s, \tag{10.23}$$

Для квазиклассической циклотронной массы приближение KG также приводит к простому выражению:

$$m_{ci} = \mu_s c_{ki} c_{mi}/s^2 = c_{mi} \frac{\hbar\sqrt{2\pi n_i}}{s}, \tag{10.24}$$

где коэффициент

$$c_{mi} = \frac{(\pi/2) - \arcsin c_{ki}}{\sqrt{1 - c_{ki}^2}} \simeq \pi/2 \tag{10.25}$$

слабо зависит от i и близок к $\pi/2$. Согласно (10.24) концентрационные зависимости m_{ci} в УР приближении спрямляются в координатах $m_{ci}^2 - n_s$ (рис. 10.2б).

Таким образом, в УР приближении параметры 2D подзон полностью определяются масштабными коэффициентами $c_{\varepsilon i}, c_{ki}, c_{mi}$, зависящими только от индекса подзон и параметра β. Поскольку величина последнего (как и кейновская скорость s) фактически одинакова для всех кейновских полупроводников (и в пределах нескольких процентов близка к единице - см. табл. 4.1), приведенные соотношения, равно как и само уравнение (10.15), фактически не содержат параметров материала, т.е. определяют универсальную для всех кейновских узкощелевых полупроводников структуру подзон. Представленные на рис. 10.2 зависимости

[2]Для знакомства с таким подходом можно рекомендовать статью J.P. Zollner, G. Paasch, G. Gobsch, and H. Ubensee, Phys. Stat. Sol.(b). **148**, 611, (1988).

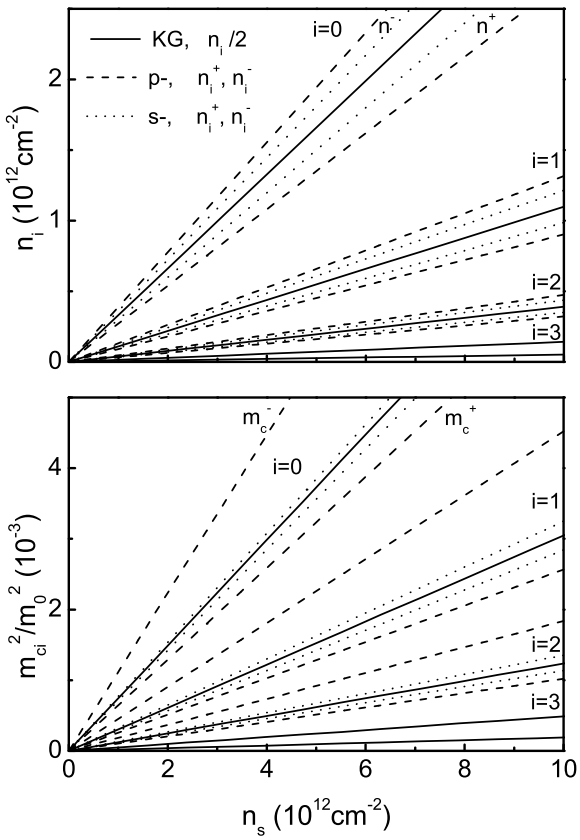

Рис. 10.2: Заселенности подзон и циклотронные массы в зависимости от n_s в УР аппроксимации для кейновских "s-"и "p-электронов"и в приближении Клейна-Гордона.

(10.23) и (10.24) ясно демонстрируют, что KG приближение в УР пределе находится в прекрасном согласии с экспериментальными данными как в отношении заселенностей, так и циклотронных масс.

Небольшие отклонения от УР теории обнаруживаются лишь в области предельно низких n_s. Ввиду малости, поправки к УР приближению могут быть получены по теории возмущений. В пренебрежении членами высшего порядка по ms^2/μ_s дисперсионное уравнение приводится к виду, полностью аналогичному (10.15), с точностью до замены $\mu_s \to \mu_s + ms^2$. Таким образом, в низшем по ms^2/μ_s приближении параметры подзон описываются теми же УР соотношениями (10.16), (10.18), (10.21)-(10.24), в которых, однако, нужно произвести указанную замену. Сохраняется и масштабная инвариантность (но по отношению к $\mu_s + ms^2$) с теми же УР масштабными коэффициентами.

Из сохранения масштабной инвариантности следует, что зависимости

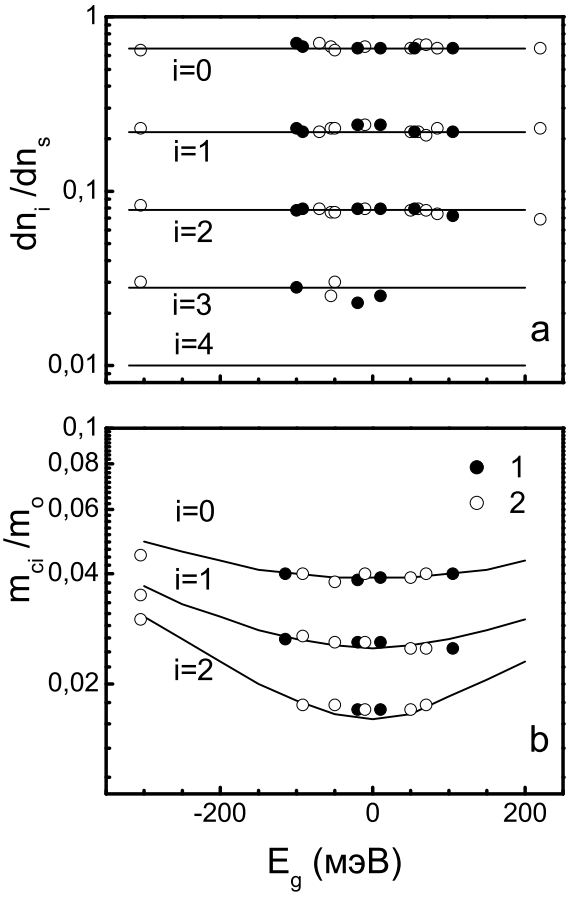

Рис. 10.3: Скорости заполнения dn_i/dn_s 2D подзон и циклотронные массы (в единицах массы свободного электрона, при $n_s = 2 \cdot 10^{12-2}$) для инверсионных (1) и обогащенных (2) слоев на слаболегированном $Hg_{1-x}Cd_xTe$ в зависимости от E_g. Точки - взятые из литературы экспериментальные данные.

$n_i\,(n_s)$, $m_{ci}\,(n_s)$, $\varepsilon_i\,(0) - n_s$ остаются теми же, что и в пределе $ms^2 = 0$. Учет членов второго порядка сводится к замене $k \to k\sqrt{1 + m^2s^4/s^2\hbar^2k^2}$, т.е. опять приходим к УР уравнению с перенормированным k. Это дает поправку второго порядка для циклотронной массы

$$m_{ci}\left(ms^2\right) = m_{ci}\,(0)\left[1 + m^2s^4/4\pi n_i s^2\hbar^2\right],$$

хорошо согласующуюся с результатами представленного на рис. 10.3б точного расчета. Для концентраций поправки равны нулю вплоть до третьего порядка. Таким образом, распределение носителей по подзонам $n_i\,(n_s)$, а это наиболее точно измеряемая экспериментальная величина, весьма слабо чувствительно к величине щели. Это и обеспечивает адек-

ватность УР аппроксимации при описании зависимостей $n_i\,(n_s)$ даже при изгибах зон $\mu_s \approx ms^2$, что фактически соответствует наименьшим экспериментально измеряемым подзонным заселенностям. При таких изгибах зон значительно более существенны поправки, обусловленные легированием. Эффективный KG потенциал для состояний основной подзоны

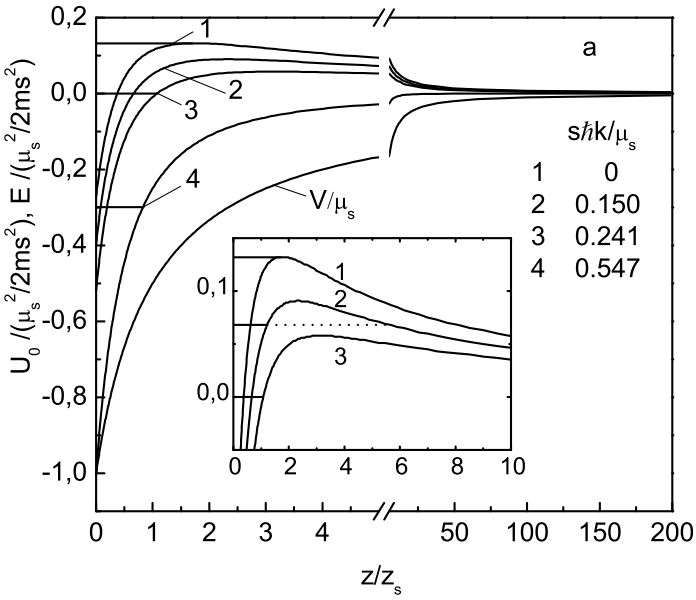

Рис. 10.4: Нормированные эффективные потенциалы и энергии (горизонтальные линии) для состояний с $k = 0$ (1), $k < k_r$ (2), $k = k_r$ (3) и $k = k_F$ (4). Основная подзона.

со спектром (10.15) приведен на рис. 10.4 для дна двумерных подзон ($k = 0$), состояний на уровне Ферми ($\varepsilon = 0$) и при двумерном волновом векторе, соответствующем $E = 0$, т.е. границе между квазистационарными и строго стационарными состояниями. При потенциале (10.14) точки поворота в эффективном KG потенциале U_0 и ширина классически запрещенной области для состояния $\varepsilon\,(k)$ определяются выражениями

$$z_{1,2} = z_s \left[\frac{\mu_s}{(-\varepsilon \pm s\hbar k)} - 1 \right], \;\; z_2 - z_1 = \frac{2 z_s \mu_s s\hbar k}{\varepsilon^2 - (s\hbar k)^2}. \tag{10.26}$$

Для состояния с $\varepsilon = -s\hbar k$ точка поворота z_2 уходит на бесконечность и проницаемость барьера становится экспоненциально малой (кривая 3 на рис. 10.4). Состояния, для которых $s\hbar k > |\varepsilon|$ и, следовательно, при всех $z > z_1$ потенциал $U_0 > E$, являются полностью стационарными.

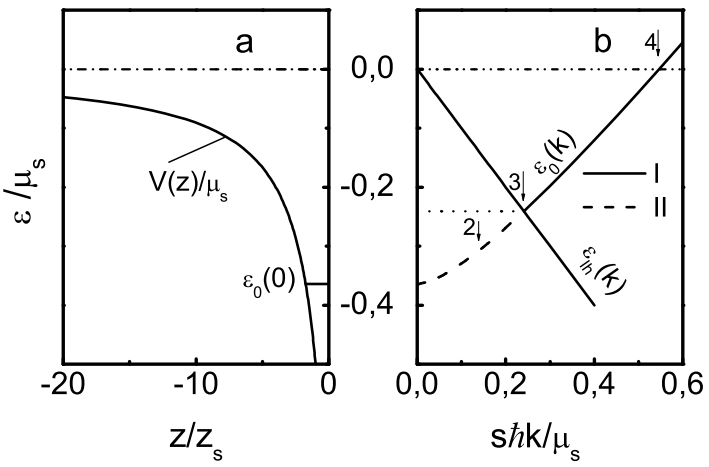

Рис. 10.5: Поверхностный потенциал (a) и дисперсионные кривые для 2D поверхностных состояний основной подзоны и легких дырок объема (b) при $E_g = 0$ в приближении KG. I -стационарные, II - квазистационарные состояния. Цифры у кривых соответствуют тем же значениям k, что и на рис. 10.4: $k < k_r$ (2), $k = k_r$ (3) и $k = k_F$ (4).

В связи с этим результатом интересно вернуться к законам дисперсии (10.15). Ясно, что двум математическим формам (10.15) соответствуют различные по характеру состояния: стационарные в случае $J = J_{nr}$ и квазистационарные в случае $J = J_r$ (см. рис. 10.4a). Пороговый волновой вектор k_r, соответствующий выполнению условия $s\hbar k = -\varepsilon$, ввиду масштабной инвариантности пропорционален μ_s

$$k_r = c_r \mu_s / s\hbar, \tag{10.27}$$

где масштабные коэффициенты c_r определяются из уравнения

$$\ln\left|\sqrt{\frac{1}{2c_r} - 1} + \sqrt{\frac{1}{2c_r}}\right| - \sqrt{1 - 2c_r} = \frac{\beta}{2}\left(i + \frac{3}{4}\right) \tag{10.28}$$

(для первых трех подзон это дает 0,241; 0,129; 0,074). Доля электронов в стационарных состояниях от общего числа заполненных 2D состояний равна

$$\frac{\delta n_{nri}}{n_i} = \frac{c_{ki}^2 - c_{ri}^2}{c_{ki}^2}, \tag{10.29}$$

и составляет более 80% (при $i \gg 1$ $\delta n_{ri}/n_i = 1 - exp{-}2 = 0,86$). Нетрудно видеть, что физической причиной непроницаемости барьера в U_0 для

состояний с $k > k_r$ является несохранение поперечного квазиимпульса для изоэнергетических переходов между 2D состояниями и состояниями объема. Действительно, поскольку закон дисперсии ветви легких дырок в объеме в УР пределе также "нейтринного"вида $\varepsilon_{lh}(k) = -s\hbar k_{lh}$, то для 2D состояний с $s\hbar k > |\varepsilon|$ двумерный электронный импульс больше дырочного и, тем более, его перпендикулярной направлению туннелирования компоненты (см. рис. 10.5b).

10.4 Эффекты спинорного типа

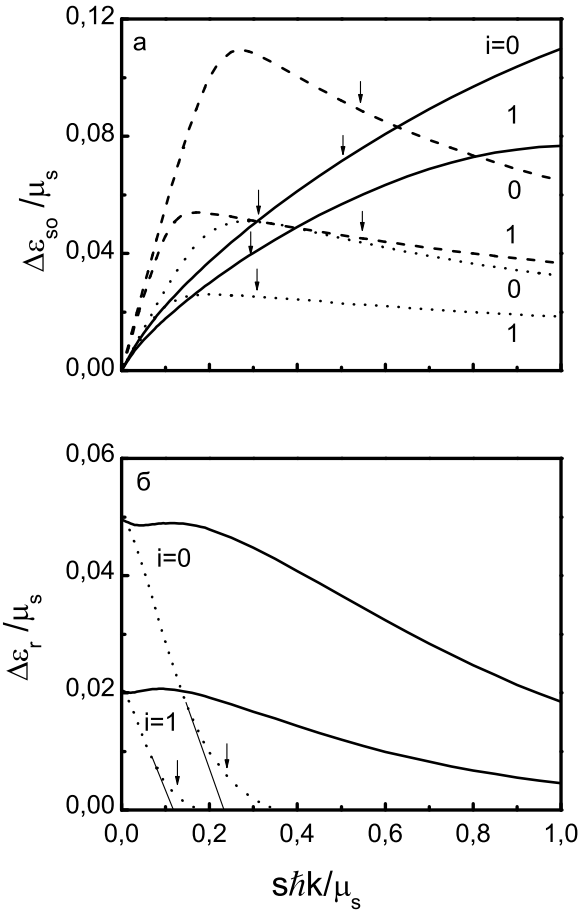

Рис. 10.6: Спин-орбитальное расщепление (a) и резонансный сдвиг (b) для первых двух подзон в УР пределе. Обозначения те же, что и на рис. 10.1. Стрелки на рис. 10.6a соответствуют значениям $k = k_{Fi}$, рассчитанным при $U_{so} = 0$, а на рис. 10.6b значениям $k = k_{ri}$.

Масштабная инвариантность УР приближения по отношению к μ_s

сохраняется и при учете спинорных эффектов, так что дисперсии в подзонах для потенциала $U = U_0 + U_{sp}$ также описываются универсальными соотношениями между нормированными энергией и квазиимпульсом (10.1). На рис. 10.6 представлены спин-орбитальное расщепление $\Delta\varepsilon_{so} = \varepsilon_i^+ - \varepsilon_i^-$ и поправки $\Delta\varepsilon_r = \varepsilon_i^0 - \varepsilon_i^{00}$ к энергиям ε_i^{00} , вычисленным в приближении KG, обусловленные резонансным потенциалом U_r (энергии ε_i^0 рассчитаны при $U_{so} = 0$). Эффективные потенциалы для "s-"и "p-электронов"с учетом спиновых эффектов представлены на рис. 10.7.

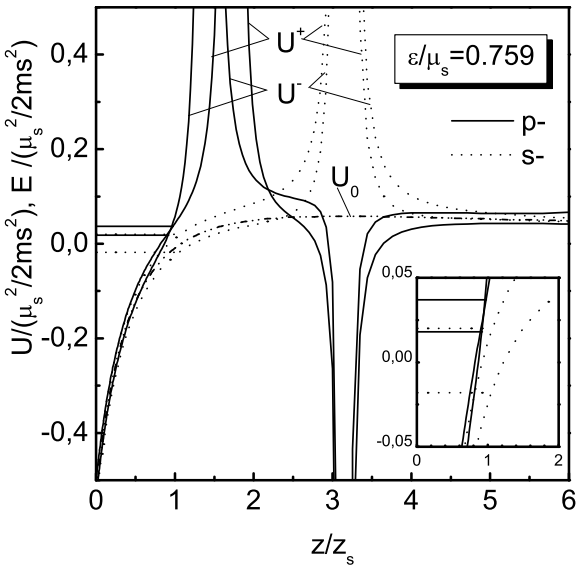

Рис. 10.7: Нормированные эффективные потенциалы и энергии (горизонтальные линии) в УР пределе для кейновских "s-"и "p-электронов"в состояниях с $\varepsilon/\mu_s = 0.759$ $(i = 0)$.

10.4.1 Резонансный сдвиг

Влияние резонансного члена U_r при больших k незначительно (в том числе для состояний на уровне Ферми для "s-"и "D-электронов"), однако весьма существенно вблизи дна подзон. При уменьшении k полюс в $U(z_+)$ для состояний $E < 0$ приближается к поверхности, тогда как точка поворота z_1 из-за квазирелятивистского смешивания движений в направлении конфайнмента и в двумерной плоскости удаляется. (В УР KG приближении протяженность волновой функции $f(z)$ для состояний на уровне Ферми $z_1(k_{Fi}) = z_s(1/c_{ki} - 1)$ почти вдвое меньше, чем на

дне подзон $z_1(k=0) = z_s(1/c_{\varepsilon i} - 1)$ -имеет место своеобразный "релятивистский конфайнмент"). В результате z_+ приближается к z_1 и положительная поправка к энергии от U_r возрастает. Полюс z_+ соответствует условию $V(z_+) = \varepsilon$, тогда как z_1 при малых k приближенному условию $V(z_1) = \varepsilon - s\hbar k$ (в приближении KG). Таким образом, при $k \to 0$ полюс z_+ максимально приближается к z_1 и резонансный сдвиг достигает максимального значения. Для важного случая состояний на уровне Ферми $V(z_+) = \varepsilon = 0$, т.е. полюс z_+ уходит на бесконечность и влияние U_r крайне слабо. Для "s-"и "D-электронов"квазиимпульсы, при которых поправки $\Delta\varepsilon_r$ становятся несущественными, хорошо коррелируют с пороговыми значениями k_r (10.27) (указаны на рис. 10.6б стрелками), выше которых межзонное зинеровское туннелирование в состояния объема запрещено законами сохранения поперечного импульса.

Для "p-электронов"положение полюса в U_r при $z = z_a$ определяется условием $V(z_a) = \varepsilon - \sqrt{3}s\hbar k/2$, так что интервал $z_1 - z_a$ почти не зависит от k вплоть до $k \approx 3k_F$ (для i=0 $z_1 - z_a \approx 0.6z_s$; полюс в U_r уходит на бесконечность лишь при $k \approx 10k_F$) и, таким образом, для больших k он существенно меньше чем для "s-"и "D-электронов"(рис. 10.7). В результате для "p-электронов"обусловленное U_r возмущение спектра для фермиевских состояний почти столь же существенно, как и для дна подзон (рис. 10.6), что в конечном счете обусловлено сильным взаимодействием "p-электронов"зоны Γ_8 в присутствии поверхностного потенциала с зоной тяжелых дырок.

Наиболее кардинально влияние резонансного члена на низкоэнергетическую ветвь $\varepsilon^-(k)$ двумерного спектра. Как видно из вставки на рис. 10.1, спектр для этой ветви, рассчитанный без учета U_r, обрывается со стороны малых квазиимпульсов, т.е. связанных состояний при малых k нет (формально это аналогично рассмотренному в предыдущей главе эффекту кинематического связывания в материалах с непараболическим спектром). Причина такого поведения может быть понята из вида эффективного потенциала при $U_r = 0$ (рис. 10.8). Из-за отрицательной сингулярности в потенциале U_{so} при $E_+(z_+) = 0$ в эффективном потенциале U для ветви $\varepsilon^-(k)$ имеется бесконечно глубокая потенциальная

яма (при $z \leq z_+ + 0$ для "s-"и при $z \leq z_a + 0$ для "p-электронов"), отделенная от приповерхностной квантовой ямы барьером. При достаточно больших k уровень энергии E лежит ниже вершины барьера, но по мере уменьшения k он поднимается и при некотором критическом значении k_t (для основной подзоны $s\hbar k_t/\mu_s = 0{,}19$, $0{,}29$ и $1{,}84$ соответственно для "s-", "p-"и "D-электронов") касается его вершины (рис. 10.8). (В квазиклассической картине этой точке соответствует одновременное обращение в нуль k_z и dk_z/dz). Начиная с этих значений k, уровень "проваливается"в яму с центром при $z = z_{+,a}$. По отношению к приповерхностной квантовой яме имеет место своеобразное "падение на центр". Если для "s-"и "D-электронов"критическое значение k_t приходится на область достаточно малых k, то для "p-электронов"низкоэнергетическая ветвь в спектре отсутствует во всем диапазоне энергий от дна подзон вплоть до уровня Ферми и выше.

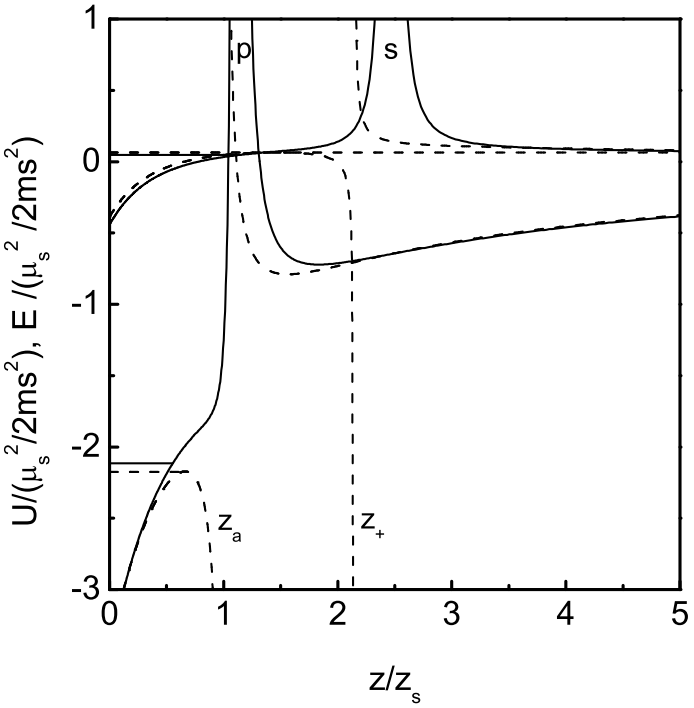

Рис. 10.8: Нормированные эффективные потенциалы и энергии (горизонтальные линии) в УР пределе для ветви ε^- основной подзоны вблизи ее "обрыва"($k = k_t$) с учетом (сплошные линии) и без учета (штриховые линии) резонансного члена.

Резонансный член U_r, имеющий в тех же полюсах положительную

сингулярность, устраняет образование ямы при $z = z_{+,a}$, а, следовательно, и обрыв дисперсионной ветви ε^-. Игнорирование U_r (или членов с первой производной в уравнениях (10.2),(10.2) при одновременном учете спин-орбитального потенциала может, таким образом, привести к эффектам, не имеющим в действительности места. Учет U_r важен еще в одном отношении. На рис. 10.9 приведены законы дисперсии вблизи дна подзон, рассчитанные без учета СОВ. Если законы дисперсии, вычисленные в приближении KG (одинаковые для "s-", "p-"и "D-электронов"), субквадратичны (квазиклассические эффективные массы на дне подзон $m_i\,(k \to 0) \to 0$) и не могут быть описаны квазирелятивистским дисперсионным соотношением

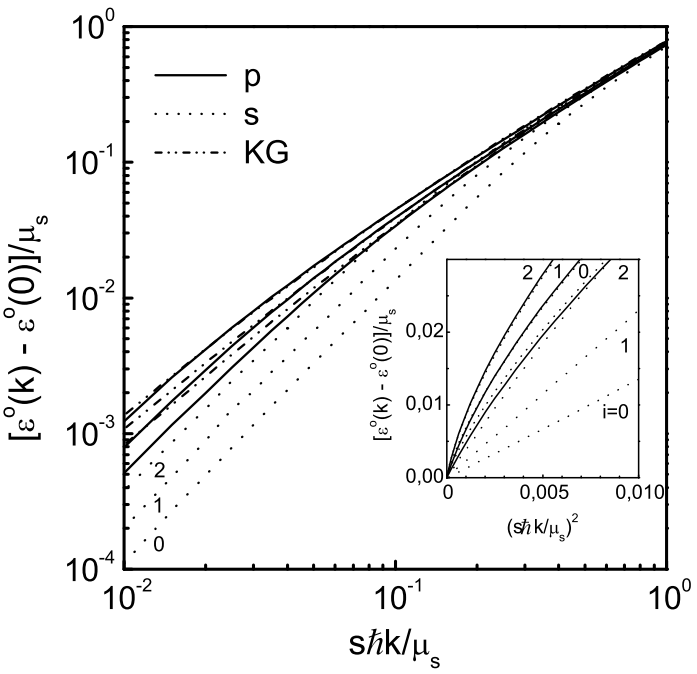

Рис. 10.9: УР подзонные дисперсии, вычисленные при $U_{so} = 0$.

$$\varepsilon_i^{\pm}(k) - \varepsilon_i^{\pm}(0) = \sqrt{(s\hbar k)^2 + m_i^2 s^4 \pm 2m_i s^2 \alpha_i k} - m_i s^2, \qquad (10.30)$$

(в приближении KG $\alpha_i = 0$), то при учете U_r спектр для "s-"и "D-электронов"параболичен вплоть до $k \approx k_F/2$ и аппроксимация (10.30) с использованием определенной в области малых k "энергией покоя"$m_i s^2 = s^2\hbar^2 k^2/2\varepsilon^{00}$ хорошо описывает подзонные спектры в широкой области k

от $k = 0$ до $k \gg k_F$. Для "p-электронов"учет U_r практически не влияет на законы дисперсии при малых k.

Как видно из рис. 10.6, величина резонансного сдвига (и спиновое расщепление -см. ниже) для всех трех типов электронов уменьшаются с ростом индекса подзоны, поскольку для возбужденных подзон, электроны которых локализованы в более широкой области пространства, эффективное электрическое поле, а, следовательно, и спиновые члены в эффективном потенциале меньше. Учет спинорных эффектов в области малых k, где резонансные сдвиги велики, приводит к существенному уменьшению межподзонных энергий по сравнению с их значениями, вычисленными в приближении KG. Следует, однако, отметить, что в оптических экспериментах (как и в обычных магнитоосцилляционных), дающих информацию о межподзонных энергиях, детектируются состояния с достаточно большими k, когда поправки $\Delta\varepsilon_r$ не значительны.

10.4.2 Снятие спинового вырождения поверхностным потенциалом

Как видно из рис. 10.1 и рис. 10.6а, связанное со спин-орбитальным членом в эффективном потенциале расщепление подзонных спектров (в отсутствие магнитного поля!) $\Delta\varepsilon_{so} = \varepsilon_i^+ - \varepsilon_i^-$ для "s-"и "D-электронов"при малых квазиимпульсах линейно по k, что соответствует феноменологической модели Рашбы для полупроводников со стандартными зонами [**?**].

$$\varepsilon_i^{\pm}(k) - \varepsilon_i^{\pm}(0) = \frac{\hbar^2 k^2}{2m_i} \pm \alpha_i k \tag{10.31}$$

Однако, уже при k приблизительно вдвое меньших его фермиевского значения $\Delta\varepsilon_{so}$ достигает максимума. Для "p-электронов"линейный участок практически отсутствует, а максимальное значение $\Delta\varepsilon_{so}$ достигается при $k > 3k_{Fi}$. С дальнейшим ростом квазиимпульса $\Delta\varepsilon_{so}$ во всех трех случаях медленно уменьшается (при $k \to \infty$ $\Delta\varepsilon_{so} \to 0$, как это и должно быть в крайнем УР пределе), что свидетельствует о неадекватности при описании спин-расщепленных подзонных законов дисперсии не только линейной (10.31), но и в дираковской аппроксимации (10.30), приводящей к насыщению величины спинового расщепления в области больших

k. Если параметр СОВ α_i в (10.30) рассматривать как функцию k, то его значения на уровне Ферми, определенные из аппроксимации (10.30) при использовании эффективных масс на дне подзон (рис. 10.9) и значений $\Delta\varepsilon_{so}(k_F)$ (рис. 10.6), во всех подзонах для "s-"и "D -электронов"вдвое меньше значений в области малых k, где действительно справедливо приближение Рашбы и использование параметра α_i имеет физические основания.

Величина спинового расщепления для "D-электронов"вдвое больше, чем для s-электронов в соответствии с дополнительным множителем $g = 2$ в спин-орбитальном члене, как это и следует ожидать ввиду малости эффектов СОВ, которые могут рассматриваться как возмущение к гамильтониану с $U_{so} = 0$, одинаковому для "s-"и "D-электронов". В случае "p-электронов"для состояний вблизи уровня Ферми расщепление $\Delta\varepsilon_{so}$ почти вдвое больше, чем для "s -электронов"и близко к его значению для дираковских электронов (рис. 10.1,10.6). Таким образом, в полном соответствии с экспериментом, в кейновских полупроводниках с инверсной структурой зон спин-орбитальные эффекты выражены более ярко, чем для материалов с положительной щелью.

10.4.3 Заселенности подзон и циклотронные массы

С точки зрения сопоставления с экспериментом наибольший интерес представляют значения фермиевских квазиимпульсов $k_{Fi}^{\pm}$, определяющие экспериментально измеряемые периоды осцилляций в магнитном поле для спин-расщепленных ветвей спектра (или периоды биений), а также циклотронные массы. Ввиду масштабной инвариантности УР спектра $k_{Fi}^{\pm}$ и подзонные циклотронные массы пропорциональны изгибу зон. Масштабные коэффициенты $c_{ki}^{\pm} = s\hbar k_{Fi}^{\pm}/\mu_s$ и, соответственно $k_{Fi}^{\pm}$ заметно различны для кейновских и дираковских электронов, в том числе для "p -электронов"они для обеих ветвей спектра меньше, чем в случае "s -электронов"(см. значения волновых векторов для состояний на уровне Ферми на рис. 10.1). В результате для одного и того же изгиба зон подзонные заселенности для "p-электронов"несколько ниже, чем для "s-"и "D-электронов". Однако соотношение между $c_{ki}^{\pm}$ для разных подзон для

всех трех случаев фактически одно и то же. Как следствие, при одной и том же значении n_s (чему, однако, для "p -электронов"соответствует изгиб зон на 6% больший, чем для "s-"и "D-электронов") практически одинаковы и суммарные подзонные заселенности $n_i = n_i^+ + n_i^-$ рис. 10.2 .

Зависимости циклотронных масс от n_s, вычисленные с учетом U_{sp} приведены на рис. 10.2б. Значения m_{ci}, вычисленные для одних и тех же n_i без учета СОВ (они близки к средним значениям $\left(m_{ci}^+ + m_{ci}^-\right)/2$), для "p-электронов"на 5-6% выше, чем для "s-"и "D-электронов", что, однако, фактически не выходит за рамки экспериментальных ошибок при определении m_{ci} из температурных зависимостей амплитуд осцилляций. Таким образом, хотя спиновые эффекты приводят к существенной и различной для "s-", "p-"и "D-электронов"перестройке 2D спектра, это не сказывается заметным образом на экспериментально исследуемых концентрационных зависимостях подзонных заселенностей и средних циклотронных масс. Те и другие описываются практически одинаковыми для "p-", "s-"и "D-электронов"универсальными зависимостями, совпадающими с зависимостями, полученными в рамках модели KG.

Для "s-"и "D-электронов"использование ВКБ метода позволяет и при учете спиновых эффектов получить в УР приближении (а при использовании теории возмущений и для случая $E_g \neq 0$) аналитические выражения для наиболее важных в экспериментальном отношении подзонных параметров. Масштабные коэффициенты $c_{ki}^{\pm}$ определяются трансцендентным уравнением

$$\sqrt{1+\eta_r}\ln\left|\frac{2\sqrt{1+\eta_r\mp\eta_s c_k - c_k^2}+2\left(1+\eta_r\right)\mp\eta_s c_k}{c_k\sqrt{4\left(1+\eta_r\right)+\eta_s^2}}\right| - \sqrt{1+\eta_r\mp\eta_s c_k - c_k^2}$$

$$\pm\frac{\eta_s}{4}\left[\pi - 2\arcsin\frac{2c_k\pm\eta_s}{\sqrt{4\left(1+\eta_r\right)+\eta_s^2}}\right] = \frac{\beta}{2}\left(i+\frac{3}{4}\right) \qquad (10.32)$$

где константа $\eta_r = (\beta/4\pi)^2$ обусловлена учетом резонансного, а $\eta_s = g\beta/4\pi$ спин-орбитального членов в (10.4) (при $\eta_r = \eta_s = 0$ (10.32) переходит в (10.19)). Поскольку $\eta_r \ll 1$, влияние "резонансных"эффектов на состояния вблизи уровня Ферми весьма незначительно (<1%). Спин-

орбитально "расщепленные"фермиевские квазиимпульсы $k_{Fi}^{\pm}$ могут быть, с другой стороны, найдены как поправки к значениям, полученным в KG приближении. Разлагая левую часть (10.32) в ряд по малому параметру $\eta_s c_{ki}$, легко показать, что поправка первого порядка сводится к замене "фазового"фактора 3/4 в правой части (10.19) (или в условии квазиклассического квантования (10.13)) на $3/4 \mp g\,(1/8 - c_{ki}/4\pi) \approx 3/4 \mp g/8$, т.е. к его изменению приблизительно на $\pm 1/4$ для "D-"и $\pm 1/8$ для "s-электронов". Соответствующие значения масштабных коэффициентов допускают при этом аппроксимацию

$$c_{ki}^{\pm} \approx c_{ki} \exp \mp \frac{g\beta}{16} \left(1 - \frac{2c_{ki}}{\pi}\right), \tag{10.33}$$

(c_{ki} -масштабные коэффициенты, вычисленные в приближении KG), что приводит к простому выражению для величины

$$\frac{\Delta n_i}{n_i} \approx \frac{g\beta}{8} \left(1 - \frac{2c_{ki}}{\pi}\right) \approx \frac{g\beta}{8}, \tag{10.34}$$

определяющей отношение основных периодов осцилляций к периодам биений. Аппроксимации (10.33) и (10.34) находятся в очень хорошем согласии с точными значениями (10.32).

10.4.4 Зависимость спинорных эффектов от величины кейновской щели и легирования

Рассмотренные выше качественные особенности УР спектра остаются в силе и для реальных 2D систем на основе УЩ и БЩ полупроводников, однако количественные соотношения могут заметно измениться в области небольших поверхностных плотностей носителей, когда глубина ямы сравнима с величиной щели или (и) велик вклад заряда слоя обеднения или электронов континуума в формирование поверхностного потенциала.

Что касается уровня легирования материала, то его влияние (здесь мы ограничимся спиновыми эффектами) существенно различно для инверсионных и обогащенных слоев. В первом случае дополнительное электрическое поле, связанное с зарядом слоя обеднения, увеличивает вклад

пропорциональных полю спиновых членов в эффективном потенциале и, как следствие, увеличивает как спин-орбитальное расщепление, так и "резонансный"сдвиг уровней. Последний эффект при стандартном рассмотрении интерпретируется как увеличение туннельного обмена между 2D слоем и состояниями объема за счет уменьшения толщины слоя обеднения с ростом легирования. В обогащенных слоях проявляется другая специфика узкощелевых полупроводников - малость критической для перехода Мотта концентрации доноров, в результате чего электронный газ в объеме полупроводиков n-типа практически всегда вырожден. Участие электронов континуума в экранировании приводит к таким специфичным эффектам, как автолокализация электронов в нулевом электрическом поле и кинематическое связывание . В отличие от инверсионных каналов, в обогащенных слоях спиновое расщепление спектра при уменьшении подзонных заселенностей, особенно в режиме кинематического связывания, значительно уменьшается (сколь-либо полное изложение относящихся сюда результатов требует, однако, специального рассмотрения).

В отличие от легирования, увеличение $|E_g|$, как в обогащенных, так и в инверсионных слоях приводит к уменьшению величины спиновых эффектов. Это непосредственно следует из вида эффективного потенциала в (10.4). Увеличение $|E_g|$ приводит к сдвигу "спиновой"сингулярности в эффективном потенциале при $z = z_{+,a}$, вглубь полупроводника, в результате чего для классически разрешенной области вклады спиновых членов, а, следовательно, и поправки $\Delta\varepsilon_r$ и $\Delta\varepsilon_{so}$ уменьшаются. Расчетные зависимости $\Delta n_i/n_i$ и $\Delta\varepsilon_r$ от нормированной щели E_g/μ_s при $\mu_b = 0$ ("обогащенные"слои на "нелегированном полупроводнике n-типа") приведены на рис. 10.10. В соответствии с большей спин-орбитальной связью для "p-электронов"в УР пределе, изменение с E_g (как и с увеличением уровня легирования) параметра $\Delta n_i/n_i$, характеризующего спин-орбитальное расщепление, значительно меньше для "p-", чем для "s-электронов"и для основной подзоны даже при $\mu_s = E_g$ не превышает 15% от его УР значения. Существенно более чувствительно к E_g расщепление в возбужденных подзонах, локализованных в более широкой

области пространства и испытывающих при $|E_g|$=0 более сильное влияние спинового потенциала.

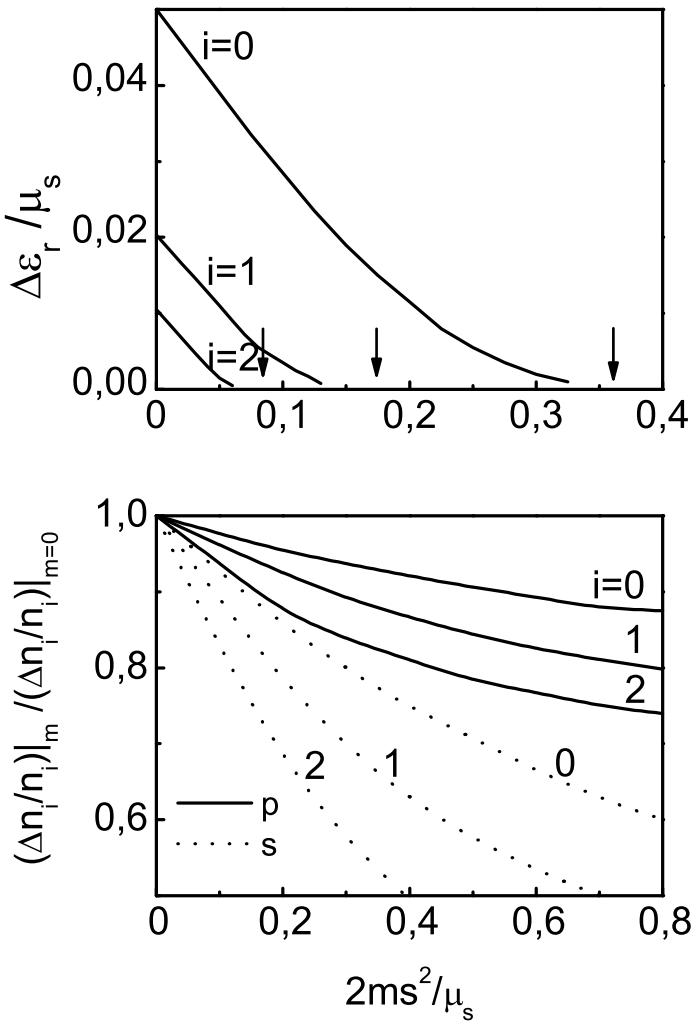

Рис. 10.10: Зависимости резонансного сдвига дна подзон и нормированной к ее значению при $E_g = 0$ относительной разности заселенностей спин-расщепленных подзон от нормированной кейновской щели. Стрелками указаны значения $|E_g|$, выше которых дно подзон выходит из "резонанса"с состояниями зоны легких дырок.

В рамках квазиклассического подхода могут быть легко получены аналитические поправки к УР приближению, обусловленные отличием $|E_g|$ от нуля. В первом порядке по ms^2/μ_s выражение для фермиевских импульсов легко преобразуется к виду

$$k_{Fi}^{\pm} = c_{ki}^{\pm}\left(ms^2\right)\left(\mu_s + ms^2\right)/s\hbar, \tag{10.35}$$

где масштабные коэффициенты $c_{ki}^{\pm}$ легко могут быть выражены через их значения (10.32,10.33) в УР пределе:

$$c_{ki}^{\pm}\left(ms^2\right) = c_{ki}^{\pm}\left(ms^2 = 0\right)\exp \pm\frac{g\beta}{16\pi}\left(\frac{2}{c_{ki}} - c_{ki}\right)\frac{ms^2}{\mu_s}. \qquad (10.36)$$

Для $\Delta n_i/n_i$ в том же приближении получаем:

$$\frac{\Delta n_i}{n_i} = \frac{\Delta n_i}{n_i}\,|_{ms^2=0}\left[1 - \frac{\left(2 - c_{ki}^2\right)}{c_{ki}\left(\pi - 2c_{ki}\right)}\frac{ms^2}{\mu_s}\right], \qquad (10.37)$$

что правильно описывает линейный участок представленных на рис. 10.10 результатов численного расчета для "s-"и "D- электронов". Соотношение (10.37), как и представленные на рис. 10.10 точные зависимости, характеризует изменение $\Delta n_i/n_i$ с E_g при одном и том же поверхностном потенциале. Нетрудно, однако, видеть, что они остаются в силе и для случая одинаковых подзонных заселенностей. Действительно, согласно (10.35), той же заселенности, что и в УР пределе при поверхностном потенциале μ_s в материалах с большими ms^2 соответствуют меньшие изгибы зон, а именно, $\mu_s\left(ms^2\right) = \mu_s(0) - ms^2$. Однако возникающая при этом поправка к УР значению $\Delta n_i/n_i$- второго порядка малости. Без учета спин-орбитального расщепления (g=0 в (10.36)) поправка по ms^2/μ_s к масштабным коэффициентам равна нулю вплоть до третьего порядка.

Более радикальные изменения с ростом E_g испытывают "резонансные"сдвиги $\Delta\varepsilon_r$ дна двумерных подзон (при $k = 0$ величина спиновых эффектов для "s-,p-"и "D-электронов", как уже указывалось выше, одинакова). Как видно из рис. 10.10, значения E_g, при которых резонансный сдвиг обращается в нуль, хорошо соответствуют величинам, при которых дно двумерных подзон попадает в щель между зонами Γ_6 и Γ_8 в объеме полупроводника (указаны стрелками), т.е. когда уровни с ростом E_g выходят из "резонанса"с состояниями легких дырок. Это также оправдывает интерпретацию члена U_r в эффективном потенциале, как ответственного за "зинеровское взаимодействие"(смешивание поверхностным потенциалом) разных "спиновых"состояний зон Γ_6 и Γ_8. Следует, однако, подчеркнуть, что в противоположность стандартной интерпретации

(зинеровское туннелирование), это взаимодействие приводит только к сдвигу, но не к уширению уровней. Благодаря непроницаемому барьеру, обусловленному потенциалом U_r и отделяющему поверхность от объема, 2D состояния не являются в обычном смысле резонансными, даже, если они совпадают по энергии с состояниями дырочного континуума, в том числе и при отсутствии щели между зонами Γ_6 и Γ_8 (в этом смысле член U_r правильнее бы назвать "антирезонансным"). Зинеровское туннелирование запрещено, по-видимому, благодаря существованию некоторой сохраняющейся величины, типа спиральности в уравнении Дирака. Оно возможно только при рассеянии и размытие уровней, таким образом, определяется не резонансным, а столкновительным уширением.